A Conditional Probability Approach to Performance Analysis of Optical Unslotted Bus-Based Networks

Alexandre Brandwajn[1], Viet Hung Nguyen[2], Tülin Atmaca[3]

[1] *University of California at Santa Cruz, Baskin School of Engineering Santa Cruz, CA 95064*
Email: alexb@soe.ucsc.edu

[2] *Corresponding author, Institut National des Télécommunications, 9, Rue Charles Fourier, 91011, Evry, France*
Tel : +33 160 764 742

Fax : +33 160 764 291

Email : viet_hung.nguyen@int-evry.fr

[3] *Institut National des Télécommunications, 9, Rue Charles Fourier, 91011, Evry, France*
Email : tulin.atmaca@int-evry.fr

ABSTRACT

This paper presents a novel approach to the performance analysis of optical packet switching bus-based networks with unslotted Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol. Because of the interdependence among bus nodes, an accurate performance analysis of such networks has been an open question for a long time. We model the bus as a multiple-priority M/G/1 queuing system with preemptive-repeat-identical (PRI) service discipline. To solve this model, we use a recurrent level-by-level analysis technique where the interference from higher levels (upstream nodes) is taken into account in terms of reappearance and disappearance rates for the server. The key features of this method are as follows. First, it specifically accounts for the distribution of voids seen by a node (via the number of attempts before a successful transmission) as a function of the node's position and offered network load. Second, it approximately computes the queue-length distribution at each node, and thus provides a tool for buffer size selection so as to meet a loss rate criterion. A comparison of our approximate model solution with network simulations indicates that our model generally offers good accuracy in assessing the performance of the network, including in terms of the queue-length distribution. Occasionally, the results of our model may deviate from simulation results. A discussion of the causes of such deviations, most likely to occur when nodes are close to saturation, offers additional insight into the properties of the bus-based network and its approximate solution.

Keywords

Asynchronous Optical Bus-based Network, Unslotted CSMA/CA, Performance Analysis, Preemptive-Repeat-Identical Service, Recurrent Analysis.

1. INTRODUCTION

For many years, voice service was the main traffic in Metropolitan Area Networks (MANs). Since voice service does not tolerate jitter (or delay variation), traditional MANs are based on synchronous circuit-switched network technologies (e.g., SONET/SDH) that guarantee very high quality (no jitters) for any service they transport. Recent years have witnessed a dramatic increase in the volume of new multimedia and data traffic, resulting in new service and bandwidth demands. The inefficiencies in terms of bandwidth granularity associated with traditional circuit-switched

networks make the latter difficult to provision for these new demands. Therefore, a more efficient networking technology is required for modern MANs.

In this regard, the optical packet switching (OPS) technology is clearly the leading candidate for the next generation MANs. It offers bandwidth-on-demand service thanks to a high granularity of bandwidth provisioning. Additionally, in combination with new technologies such as circuit emulation [1] or GMPLS, it provides cost-effective network architectures that support both multimedia (voice, video) and data traffic. The bus topology appears as one of the best choices for next generation OPS MANs, because it inherits the reliable property of current SONET/SDH ring networks (viz., fast recovery from link failures). Furthermore, since bus topology allows many nodes to share the same transmission resource, it improves resource utilization thanks to statistical multiplexing of traffic flows. In order to efficiently share a transmission resource among nodes, OPS bus-based networks need an efficient medium access control (MAC) protocol. The Optical Unslotted Carrier Sense Multiple Access with Collision Avoidance (OU-CSMA/CA[1]) [2, 3] protocol appears an attractive solution. Its simplicity makes the network installation and management easier.

In spite of the above advantages, the OU-CSMA/CA protocol and the bus topology have several drawbacks. The bus topology may exhibit unfairness depending on the position of the nodes on the bus. For example, upstream nodes (i.e., the nodes closest to the beginning of the shared transmission resource) might monopolize the bandwidth and thus prevent downstream nodes from transmitting. Additionally, the asynchronous nature of the OU-CSMA/CA protocol may lead to an inefficient fragmentation of bandwidth, resulting in low resource utilization. Indeed, the asynchronous transmission of packets at upstream nodes may fragment the global bandwidth into small segments of bandwidth (voids) that are unusable for the transmission at downstream nodes.

Due to this interdependence among bus nodes, an accurate performance analysis of OPS bus-based networks using OU-CSMA/CA protocol has been an open question for a long time. There are performance analysis studies of packet-switched slotted-ring networks, such as [4, 5] for Fiber Distributed Data Interface, [6] for PLAYTHROUGH networks, [7, 8] for Distributed Queue Dual Bus, and [9, 10] for probabilistic p_i-persistent networks. These studies usually model a ring node as an M/G/1 vacation system with Bernoulli schedule [6, 9, 10] or with time-limited service discipline [4, 5], and obtain approximate values for the mean access delay and throughput at each node. Since the interdependence between nodes makes exact analysis intractable, most studies use the assumption of independent arrival of empty slots at a tagged node. For the Distributed Queue Dual Bus protocol, an exact analysis based on Markov chain model is provided in [7], but it does not appear to scale well for a larger network size.

For the CSMA with Collision Detection (CSMA/CD) protocol, a number of performance studies have been published for both slotted and unslotted ring. An analysis treating each node as an independent M/G/1 queuing system, which services fixed-length packets, is presented in [11]. This work takes into account the interference from other nodes in terms of the service time distribution; it also considers the propagation delay between two adjacent nodes. Another approach applying complex analysis of the packet output process of unslotted CSMA/CD is used in [12], under the assumption of variable length packets. Authors of [12] derive the Laplace-Stieltjes transform (LST) of the packet inter-departure time and of the packet delay.

The aforementioned studies are not easily and directly applicable to model the network studied in this work, mainly due to the difference in access schemes (e.g., slotted versus unslotted, collision detection versus collision avoidance, etc.). More recently, several authors used priority queuing

[1] Note that the OU-CSMA/CA discussed here is unrelated to the one that is used in wireless protocols such as 802.11

systems to analyze the performance of bus-based networks. In [13, 14, 15 and 16], authors have presented performance analysis of OPS bus-based networks employing *optical* CSMA/CA protocol (e.g., the DBORN architecture [2]). Using Jaiswal's [17] and Takagi's [18] results on M/G/1 priority queues, the authors derive mean packet delay at each node for both slotted [13, 14, 15] and unslotted [14, 15, 16] modes, which support both fixed and variable length packets. In particular, both [15] and [16] use the same approach to derive the upper and lower bounds of the packet delay at downstream nodes on the unslotted bus-based network.

The work presented in this paper focuses on the performance analysis of OPS bus-based network using OU-CSMA/CA protocol such as described in [2] or [3], supporting variable length packets (i.e., Ethernet-based network). The whole network can be modeled as a multiple-priority M/G/1 queuing system with Preemptive-Repeat-Identical (PRI) service discipline. Our approach to the analysis of this model provides an insight into the correlation between local transmission at a downstream node and transit traffic flowing from upstream nodes through the number of service interruptions before a successful transmission. In addition to the mean packet delay that was studied in [13, 14, 15 and 16], we are able to compute the queue length probability at each node via simple recurrent equations. This result may help network designers select buffer sizes so as to meet a loss rate criterion.

To solve the above model, we use a recurrent level-by-level (node-by-node) analysis technique, where the interference from upstream nodes (which causes service preemptions at downstream nodes) is taken into account in terms of reappearance and disappearance rates for the server (bandwidth for transmission), as well as in terms of change of service time distribution for preempted services. The solution for each level is based on conditional probabilities and a fixed point iterative method, and requires only a limited number of iterations. The advantage of this method, compared to classical methods for solving M/G/1 priority queue, is that it provides a computationally efficient approach to the evaluation of the stationary queue-length distribution.

This paper is organized as follows. Section 2 describes the network architecture and our analytical model. Section 3 presents the outline of the recurrent solution of this model (the detailed solution is given in the Appendix) yielding approximate queue-length distributions, as well as an approximate distribution of the number of transmission attempts at each node. Section 4 provides numerical results, and compares the solution of the proposed analytical model with simulation and with the model proposed in [13]. Finally, Section 5 is devoted to conclusions and discussion of future work.

2. NETWORK ARCHITECTURE AND ANALYTICAL MODEL

2.1 Network Architecture

The network considered consists of two unidirectional buses: a transmission (upstream) bus that provides a shared transmission medium for carrying traffic from several access nodes to a point of presence (POP) node; and a reception (downstream) bus carrying traffic from the POP node to all access nodes (e.g., the DBORN architecture [2]). Thus, an access node always "writes" to the POP node employing the transmission bus and "listens" for the POP node using the reception bus. The traffic emitted on the transmission bus by an access node is first received by the POP node, then is either switched to the reception bus to reach its destination node, or is routed to other MAN or backbone networks.

In this paper we only analyze access nodes performance on the transmission (upstream) bus where the problem of access control arises. The reception (downstream) bus is a simple broadcast or diffusion communication, which does not need to be deeply investigated. Therefore, we can describe the transmission bus under study as a unidirectional fiber connecting several nodes

(stations) to a POP node (Figure 1: Unidirectional OPS bus-based network). All nodes share a single wavelength (operating at 2.5 Gbps or 10 Gbps) to transmit local traffic to the POP node. For reasons of simplicity and cost-efficiency, bus nodes use passive optical technology that reduces the number of transceivers at each node. Specifically, there is no packet-drop operation at bus nodes; each node can only add/insert local traffic into the wavelength without intercepting transit traffic of upstream nodes.

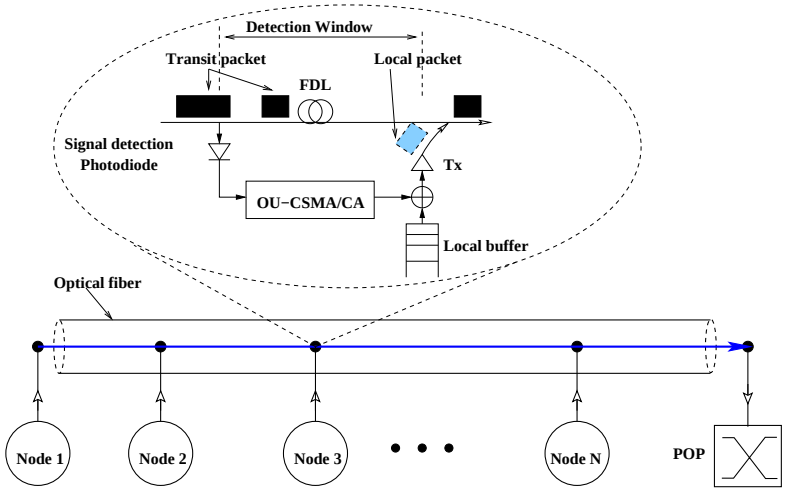


Figure 1. Unidirectional OPS bus-based network

Since the single wavelength is shared between many nodes, a MAC protocol is required to control the access of bus nodes to the wavelength. Thus the OU-CSMA/CA protocol (Figure 1), which is based on void (i.e., free segment of bandwidth) detection, is used to ensure an asynchronous collision-free packet insertion into the transmission wavelength. This protocol has two interesting characteristics: 1) It is a fully-distributed protocol that simplifies the implementation and management of the network; 2) Its asynchronous nature offers the capability to support data traffic with variable size packets in the network, without the need for complex segmentation/assembly process. Based on the second property, a mature technology like Ethernet is used for the data plan of the network. This means that the network supports variable size optical packets, each consists of one conventional Ethernet packet encapsulated by an optical overhead [2]. It is worth noting that the maximum packet length supported by the network is limited by the Maximum Transmission Unit (MTU) of Ethernet, which is around 1500 bytes.

With OU-CSMA/CA protocol, a node begins transmitting a packet if and only if it detects a void on the transmission wavelength that is at least equal to the packet size (including an optical overhead if necessary). Generally speaking, voids seen by a node are free segments of bandwidth between transit packets coming from upstream nodes. Therefore, the most upstream node that begins the shared wavelength is always able to transmit, since it always has available bandwidth at any time (i.e., infinite void length).

2.2 Network Model

The transmission bus uses only one wavelength shared by the N access nodes. From modeling perspective, it is convenient to view the operation of the OU-CSMA/CA protocol as follows. A

node begins transmitting a local packet as soon as it detects a void. The transmission is interrupted when a transit packet arrives from an upstream node (i.e., the void is not large enough), and the packet returns to the queue waiting for the next void. At detection of the next void, the node starts again transmitting the local packet whose transmission was interrupted. This process is repeated until the packet is successfully transmitted (i.e., a large enough void is found). Thus, for performance modeling purposes, a transmission of a packet in the OU-CSMA/CA protocol may be viewed as a number of interrupted (preempted) transmission attempts followed by one successful transmission.

The behavior of the real network is very similar to that of a priority queuing system, in which a single server (i.e. the shared wavelength) services N queues (i.e. the N bus nodes) with Preemptive-Repeat-Identical (PRI) priority service discipline. Each node in this system defines a separate priority level according to its position on the bus. Nevertheless, the queuing system with PRI priority discipline does not exactly match the operation of the real network under study. Indeed, in the queuing system with PRI discipline, a low level (a downstream node) can start transmitting if and only if there is no client packet at higher levels (upstream nodes). This means that the server (bandwidth) viewed by a low level client remains occupied until all higher level clients have been successfully serviced. But in the real network, the bandwidth viewed by a downstream node is occupied only during the *successful transmission periods* of client packets at any upstream node, and the bandwidth remains available during *interrupted transmission periods* of client packets at any upstream node (i.e. client packets at downstream nodes may be serviced even if upstream nodes are attempting to transmit their client packets). More precisely, when an upstream node detects a void and cannot use it for transmission (i.e. transmission attempt is interrupted) because the void length is not long enough, this void (or, in other words, the time elapsed from the moment where the node attempts to transmit the packet till the transmission is interrupted) may be used by a downstream node for transmitting smaller packets whose lengths fits this void. Thus, from queuing model perspective, the real network behavior corresponds to a queuing system with a special priority service discipline, which appears more complicated than the PRI discipline. Note, however, that if the network supports only fixed length packets, then the real network behavior perfectly matches that of the queuing system with PRI discipline, because in this case a void shorter than the packet length is unusable for any node.

In this paper, we approximately represent the OPS bus-based network supporting variable length packets as a queuing system with the PRI discipline. The case with more complex priority discipline is left for future work. Starting with most upstream (and highest priority) node, we number the nodes 1 through N so that an upstream node i has priority over a downstream node j, $1 \le i < j \le N$. We assume that each node has an infinite buffer, and client packets stored in the buffer are serviced in First-Come-First-Serve (FCFS) order.

Since the network under consideration is supposed to be a future metro network receiving a high aggregation of traffic coming from high-speed access networks such as Fiber To The Home (FTTH), we can reasonably assume that local packets at node i arrive according to a Poisson process with rate λ_i, and that their service times are mutually independent and have a general distribution with known finite mean (m_i) and variance (Var_i). Thus our network model is an M/G/1 system with N priority levels and PRI FCFS service discipline.

To solve this model, we propose a new approach, different from classical approaches for solving the M/G/1 priority queue, which allows us to approximately compute not only the mean queue length but also the steady-state queue length distribution at each node. In our method, we analyze bus nodes one by one, and we use a specific state description to represent the interference from upstream nodes (if any). In the following sections, we present a model of the general service distribution at each node, highlighting the effect of PRI discipline on the service distribution at downstream nodes.

2.2.1 Modeling of Original Service Time Distribution

To model the service time distribution at each node, we use a Coxian distribution [19] with the minimum number *k* of exponential stages needed to match the first two moments of the service time distribution as shown in Figure 2 (k-stages Coxian system modeling service time distributions at each node). The resulting form of the Coxian depends on the coefficient of variation *Cv* of the distribution being represented. In the case where $Cv \geq 1$, we use a two-stage Coxian distribution ($k = 2$ in Figure 2) with three parameters μ_1, μ_2 and p_2. For $1/\sqrt{k} \leq C_v < 1/\sqrt{(k-1)}$, we use a *k*-stage hypoexponential distribution [20] ($p_2 = 1$ in Figure 2) with two parameters μ_1 and μ_2. As the number of stages k increases, the *Cv* of this distribution tends to zero, which corresponds to the case of fixed-length packets. The parameters of these Coxian distributions are readily derived from the first and second moments of the original service-time distributions as follows.

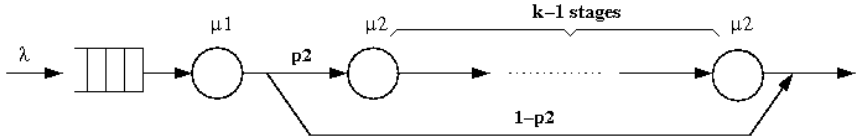


Figure 2. k-stages Coxian system modeling service time distributions at each node

Consider the case with $Cv \geq 1$. We use a two-stage Coxian distribution (Figure 2 with $k = 2$). Let $t_1 = 1/\mu_1$ and $t_2 = 1/\mu_2$ be the mean service times at each stage, we have the following equations:

$$m_i = t_1 + p_2 t_2 ,$$

$$Var_i = t_1^2 + p_2(2 - p_2)t_2^2 = m_i^2 C_v^2 . \qquad (1)$$

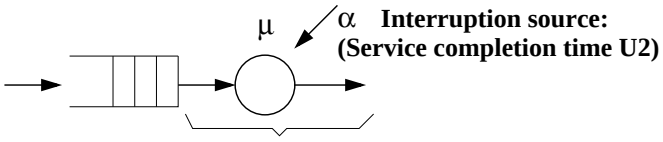
Set $t_1 = \gamma m_i$, where γ $(0 < \gamma < 1)$ is a constant. We are able to obtain the values for the Coxian distribution parameters

$$p_2 = \frac{2(1-\gamma)^2}{C_v^2 + (1-\gamma)^2 - \gamma^2}, \quad \text{and} \quad t_2 = \frac{m_i(1-\gamma)}{p_2}, \qquad (2)$$

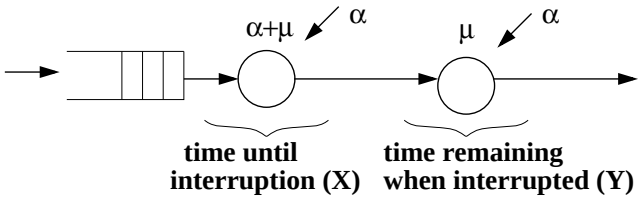
In practice, we may choose γ such that $t_1 < t_2$ (i.e. $0 < \gamma \leq 0.5$). Similarly, we can derive the parameters for the other cases with $1/\sqrt{k} \leq C_v < 1/\sqrt{(k-1)}$.

2.2.2 Modeling of Interrupted Service Time Distribution

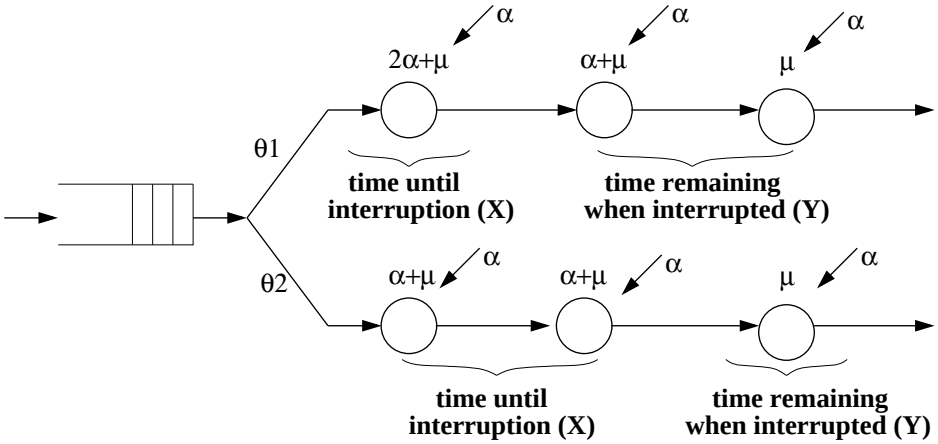
As stated earlier, the service at nodes $i > 1$ may be interrupted due to arrivals of client packets at upstream nodes. Thus, a node may have to attempt the transmission of a packet several times, each time the transmission being interrupted by transit traffic from higher priority upstream nodes until a long enough void comes along (no interruption). Clearly, after every interrupted transmission, the node reattempts the transmission of the same packet at the next void. Somewhat paradoxically, to describe the fact that it is the same packet whose transmission is reattempted, we need to represent a potentially different packet length (and hence service time) distribution (for packets whose transmission got interrupted) at each consecutive attempt.



a. System with original service time distribution



b. Service time distribution on the second transmission attempt



$\theta 1 \sim \alpha/(2\alpha+\mu)$, $\theta 2 \sim [(\alpha+\mu)/(2\alpha+\mu)][\alpha/(\alpha+\mu)]$, $\theta 1 + \theta 2 = 1$.

c. Service time distribution on the third transmission attempt

Figure 3. Evolution of service time distribution following interruptions

To understand what is going on, perhaps the simplest example is to examine the case where the original service times are exponentially distributed with parameter μ and interruptions arrive from a Poisson source with rate α.

On the first transmission attempt of any client packet, the service time distribution is the original exponential distribution with parameter μ (Figure 3.a: System with original service time distribution). This service may be interrupted by the Poisson source with rate α. Note that on the first attempt, we are dealing with the whole population of client packets.

On the second attempt, we are dealing with a subset of client packets whose transmissions got interrupted for the first time (i.e., we exclude all client packets that have been successfully transmitted at the first attempt). To derive the service time distribution of this subset of client packets, we consider the time until the interruption (*X*) and the time remaining when the interruption occurred (*Y)*. Consider a small interval of time $(t, t+\Delta t]$. The probability that a first service interruption will happen during this interval can be expressed as $\Delta t \alpha e^{-\alpha t} e^{-\mu t} + o(\Delta t)$ where $o(\Delta t)/\Delta t \underset{\Delta t \to 0}{\to} 0$. The overall probability that a first service will be interrupted is simply the probability that an exponentially distributed process with parameter α (the Poisson interruption arrivals) finishes before the exponentially distributed service process with parameter μ, which is readily seen to be $\frac{\alpha}{\alpha+\mu}$. Hence, the conditional density of the time to interruption given that the service is interrupted is $(\alpha+\mu)e^{-(\alpha+\mu)t}$, i.e., the time until interruption *X* is exponentially distributed with parameter $\alpha + \mu$. Because of the memoryless property of the original service time distribution, the remaining service time *Y* at the point of interruption is exponentially distributed with the original parameter μ. Therefore the resulting service time distribution of the subset of client packets after the first interruption (or, in other words, on the second transmission attempt) is the hypoexponential distribution shown in Figure 3.b (Service time distribution on the second transmission attempt).

On the third attempt, we are dealing with a subset of client packets whose transmissions got interrupted for the second time (i.e., we exclude all client packets that have been successfully transmitted on the first and the second attempt). In other words, we are dealing with the two-stage hypoexponential distribution in Figure 3.b interrupted by a Poisson arrival process with rate α. This interruption could have taken place while the service was in the first or the second stage of the two-stage hypoexponential. Thus, as shown in Figure 3.c (Service time distribution on the third transmission attempt), with probability θ_1, the interruption could have taken place while the service was in the first stage of the two-stage hypoexponential. This results in an exponentially distributed time to interruption with parameter $2\alpha + \mu$, followed by an exponentially distributed residual of the first stage (parameter $\alpha + \mu$) and the full second stage. With probability θ_2 the interruption could have taken place while the service was in the second stage. Then the time before interruption consists of the full first stage, followed by the part of the second stage preceding the interruption (exponential with parameter $\alpha + \mu$), and the time after the interruption is the exponentially distributed residual with parameter μ. An analogous process continues at subsequent interruptions.

In a similar way we can derive and represent the distribution of service times at each interruption when $Cv > 1$ and $Cv < 1$. We observe that, with the obvious exception of a constant packet length (hence, service time), for all distributions of packet lengths, the mean increases while the coefficient of variations decreases on each subsequent transmission attempt. In our exponential service example, after the first interruption the mean nearly doubles when α is small. We also observe that

both the increase in the mean and the decrease in the coefficient of variation slow down at each subsequent attempt. This makes perfect physical sense: as transmissions are attempted, shorter packets are more likely to be successfully transmitted and longer packets need more attempts. The elimination of shorter packets accounts for both the increase in the mean and the decrease in variability of the packet length of the subsets of client packets. In the limit, we expect the packet length to tend to the maximum packet length (MTU) at subsequent attempts at the given node with variance tending to zero.

Figure 4 (Mean and square coefficient of variation of the service time as a function of the number of transmission attempts) illustrates the evolution of the mean and the squared coefficient of variation of the service time at each transmission attempt for an initial distribution with an initial squared coefficient of variation of 3.3. For the example considered, at the second attempt the mean service time is close to 4 times the initial average while the squared coefficient of variation is less than half the initial value. At the third attempt, the mean is over 7 times the original value while the squared coefficient of variation drops to less than 20% of the original.

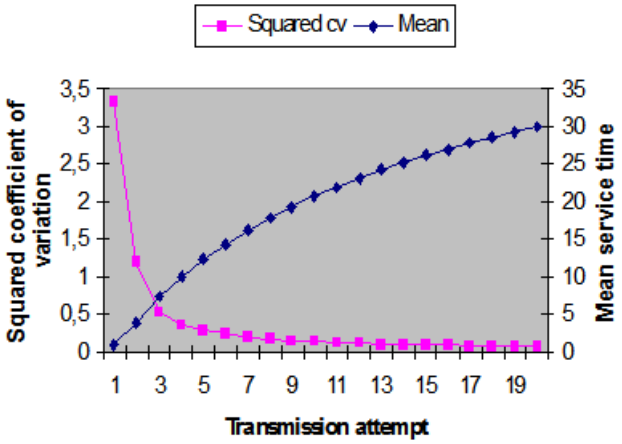


Figure 4. Mean and square coefficient of variation of the service time as a function of the number of transmission attempts

3. OUTLINE OF MODEL SOLUTION

In order to obtain a tractable approximate solution to our model in the steady state, we analyze the bus nodes one by one, starting with node *1* for which there is no upstream transit traffic. We focus on node i ($i = 1 \ldots N$), and, to simplify our notation, we omit the node subscript i whenever this does not create confusion.

3.1 Solution for Node 1

Balance equations derivation

Since node *1* always "sees" the bus bandwidth free, we simply describe the equilibrium behavior of this node by the joint steady-state probability $p(n,l)$, where n ($n \geq 1$) is the number of packets at this node and l refers to the stage of service of the Coxian service time distribution shown in Figure 2. We denote by $p(n)$ the marginal distribution for the number of packets at the node, and by $p(l|n)$ the corresponding conditional probability for l given n. Using the fact that $p(n,l) = p(n)p(l|n)$, we are able to readily obtain the equations for $p(l|n)$ and $p(n)$.

Fixed-point iteration method for solution of the balance equations

We detail in the Appendix A a simple recurrent solution using fixed point iteration method for these equations. This solution is based on the general notion of state equivalence [21], and its specific application to M/G/1-like queues [22]. This solution allows us to compute the conditional rate at which packets are served (i.e., effectively transmitted) given *n*, which we denote by *u(n)*, and, hence, *p(n)*.

The computation scheme can be described by the following pseudo code:

factor = sum = 1.0;

mean_occupancy = 0.0;

for (n=1; n < n_max; n++) {

Solve *equations for conditional probabilities p(l|n);*

Compute *u(n);*

*factor *= λ / u(n);*

sum += factor;

*mean_occupancy += n*factor;*

if ($|u(n)-u(n-1)<\varepsilon|$*)* **break***;*

}

Complete *computation of "infinite part" of sum and mean_occupancy;*

sum = 1.0/sum;

mean_occupancy **=** *sum;* */* normalize */*

prob_node_idle = sum;

Compute *server reappearance rate for node 2, i.e.* β_2*;*

In our computation, we used $\varepsilon = 10^{-10}$ for the test of convergence of *u(n)* to its limiting value. The computation of the infinite part of the normalizing constant, as well as the mean node occupancy, is straightforward given the geometric-series form of the tail of the node occupancy distribution [22].

3.2 Solution for Node *i > 1*

Balance equations derivation

As stated earlier, in our PRI model, a node $i > 1$ may find the server (bandwidth) available or occupied. In our model, the server is available if and only if there are no client packets at all nodes *1 ... i-1*, and is occupied otherwise. When the server is available, it serves client packets with a constant rate *R* which is the wavelength bit rate. Thus, viewed from node $i > 1$, an available server becomes occupied whenever a client packet arrives at an upstream node $j < i$, hence interrupting the service at node *i*; and an occupied server becomes available whenever the last client packet at the upstream nodes *i-1* has been successfully transmitted (recall that, in our model, client packets at node *i-1* are serviced if and only if all upstream node queues are empty).

Let α_i and β_i respectively be the disappearance and reappearance rates of the server viewed by node $i > 1$. Since the arrivals to each node come from a Poisson source, the disappearance rate α_i of the server is exactly given by $\sum_{j<i} \lambda_j$. The reappearance rate β_i can be expressed in terms of the conditional transmission rate $u(n)$. As an example, for node 2 we have $\beta_2 \approx u_1(1)p_1(1)/[1-p_1(0)]$ (see Appendix D for more details).

Since the service distribution changes at consecutive transmission attempts in our model, to represent the Preemptive-Repeat-Identical discipline at node $i > 1$, we choose a state description that explicitly accounts for possible service interruptions and retrials at the node. The parameters of the service time distribution change with each transmission attempt as discussed above. We denote by k_j the number of exponential stages required to represent the service time distribution at the j-th transmission attempt ($j = 1,2...$). We describe the state of node i by the triple (n, j, l) where n is the current number of packets at the node, j is the transmission attempt, and l is equal to the current number of the service stage at this attempt ($1,..., k_j$) or 0 if the server (bandwidth) is unavailable. Recall that, in our model, the server becomes unavailable with rate α_i , and available again with rate β_i.

Fixed point iteration method for solution of the balance equations

Using the fact that $p(n,j,l) = p(n)p(j,l|n)$, we are able to transform the balance equations for $p(n,j,l)$ into equations for the conditional probability $p(j,l|n)$. We then derive a recurrent solution using a fixed-point iteration method for increasing values of $n \geq 1$. To limit the size of the state space for each n, we explicitly compute the parameters of the service time distributions at transmission attempts up to a certain value j^*, and we use "catch all" average values for the parameters of the service time distribution at transmission attempts above j^*. As stated above, the mean value of the service time distribution increases and its coefficient of variation decreases as the number of transmission attempts increases. Theoretically, for a Coxian distribution this mean value might increase to infinity. But, in our real network, this mean value is naturally limited by the service time of the maximum transmission unit (MTU) of the transmission protocol used. Thus, if j^* is chosen large enough, all service time distributions at transmission attempts above j^* may be replaced by a constant service which is the service time of the maximum length packet MTU. We also limit the number of stages in the Coxian distribution to k^* (so that the minimum value of Cv in our model is $1/\sqrt{k^*}$).

From the conditional probability $p(j,l|n)$ computed using the above recurrent solution, we readily obtain the conditional rate of transmission completions $u(n)$, and, hence, the marginal probability of the number of packets at the node $p(n)$, as well as an approximate value for β_{i+1}, the rate at which the server becomes idle, i.e. available for downstream nodes. We give in the Appendices B and C additional details of our solution, and in Appendix D an outline of the estimation of the value of β_{i+1}.

The computation scheme can be described by the following pseudo code:

Compute *rate of server disappearance for this node, i.e.* α_i*;*

Compute $p(l=0|0) = \alpha_i / [\alpha_i + \beta_i + \lambda]$*;*

factor = sum = 1.0;

mean_occupancy = 0.0;

for (n=1; n < n_max; n++) {

Solve *equations for conditional probabilities* $p(j,l \mid n)$ *;*

{

let $x = p(j = 0, l = 1 \mid n)$ *and* $y = u(n)$ *;*

For all $j = 1,..., j^* - 1$, **Express** $p(j,l \mid n)$

as $p(j,l \mid n) a_{j,l} x + b_{j,l} y$ *;*

Use

$$\sum_{j \geq 1} \sum_{l=0}^{k_j} \left[a_{j,l} x + b_{j,l} y \right] = 1$$

$$y = \sum_{j \geq 1} \sum_{l=0}^{k_j} \left[a_{j,l} x + b_{j,l} y \right] u_l^j (1 - \overline{q}_l^{\,j})$$

to **determine** $p(j = 0, l = 1 \mid n)$ *and* $u(n)$ *;*

}

*factor *= $\lambda / u(n)$;*

sum += factor;

*mean_occupancy += n*factor;*

if ($\left| u(n) - u(n-1) < \varepsilon \right|$*)* **break***;*

}

Complete *computation of "infinite part" of sum and mean_occupancy;*

sum = 1.0/sum;

mean_occupancy = sum; */* normalize */*

prob_node_idle = sum;

Compute *server reappearance rate for next node (if applicable);*

In the next section, we present numerical results obtained from our model, and compare them with results obtained from a detailed network simulation, as well as from another analytical model [13].

4. RESULTS

4.1 Model Accuracy

In our approach to the solution of a model of an optical bus, outlined in the preceding section, we are able to approximately analyze each node one by one. At node $i > 1$, the presence of upstream nodes is represented as the server (bandwidth) becoming busy with rate α_i and then available with

rate β_i. Since we are only able to compute the rate β_i approximately, this is one potential source of inaccuracy of the model.

Another potentially important point is the fact that we match only the first two moments of the distribution in our representation of the service times. For the M/G/1 Preemptive Resume, as well as for Head-of-Line Non-Preemptive priority queue, it is well known (e.g. [23]) that the expected number of customers of each class in the system depends on the service time distribution only through its first two moments. To assess the effect of higher moments in our Preemptive-Repeat-Identical model, we focus on the simple case of the first two nodes.

Table 1. Influence of high moments of service time distribution in PRI model

Packet rate ($\lambda_1 = \lambda_2$)	**Model**	**Mean packets number at node 2 (Distribution I)**	**Mean packets number at node 2 (Distribution II)**
0.06733	Full two-node model	0.1073 ± 0.0002	0.1058 ± 0.0002
	Model of node 2 alone	0.1043 ± 0.0003	0.1023 ± 0.0003
0.13466	Full two-node model	0.4179 ± 0.0091	0.4292 ± 0.0123
	Model of node 2 alone	0.4009 ± 0.0026	0.4038 ± 0.0038
0.20199	Full two-node model	2.5726 ± 0.3549	1.6053 ± 0.0750
	Model of node 2 alone	2.6417 ± 0.4280	1.4507 ± 0.1157

We consider a full two-class PRI queuing model, as well as a model of node 2 with the approximate server reappearance rate β_2 computed from the recurrent solution of node 1. In this way, in addition to the potential influence of higher moments, we are able to study the effect of the approximate computation of the rate with which the server becomes available (by comparing the results for node 2 in both models). We use discrete-event simulations of both models for two different random two-stage Coxian distributions with the same mean and variance. The parameters of these distributions are given as follows. The parameters for Distribution I are $\mu_1 = 1.9606$, $\mu_2 = 0.4915$, $p_2 = 0.328$. For Distribution II the corresponding parameters are $\mu_1 = 9.8573$, $\mu_2 = 0.6316$, $p_2 = 0.5802$. Both distributions have a mean of 1.0201 and a variance of 2.0755, but Distribution I has a third moment of 17.4364 (and hence a skewness of 3.3521), while Distribution II has a third moment of 14.787 for a skewness of 2.4594. Table 1 (Influence of high moments of service time distribution in PRI model) illustrates the results for the expected number of packets at node 2 obtained as the rate of packet arrivals to both nodes increases. All simulation results in Table 1 include confidence interval estimated at 95% confidence level using 7 independent replications of 800,000 successful packet transmissions each.

We observe that the inaccuracy caused by the approximate server reappearance rate β_2 seems to be quite limited (on the order of a few percent, and, in several cases, the confidence intervals for both models overlap). As the packet arrival rate increases, the shape of the distribution of service times beyond the first two moments appears to have a much greater influence: over 25% relative difference in the expected number of packets at node 2 for the example considered.

Table 2. Two random different packet length distributions with the same mean and variance

Distribution	Mean (μs)	C_v^2	Packet length distribution
III	2.56	0.4375	63.64% packets of 400 bytes 36.36% packets of 1500 bytes
IV	2.56	0.4375	36.36% packets of 100 bytes 63.64% packets of 1200 bytes

Table 3. Influence of higher moments of service time distribution in real network

Mean packet arrival rate (packets/μs)	Mean response time (μs) at node 4	
	Distribution III	Distribution IV
0.058	20.04 ± 0.1316	19.50 ± 0.0849
0.068	51.39 ± 0.6901	47.78 ± 0.3457
0.078	1633.00 ± 204.2	965.80 ± 90.43

Interestingly, we have also observed the effect of higher moments of service time (packet length) distribution in the real network environment. For example, we simulate the network with 4 nodes transmitting on one wavelength at 2.5Gbps, and all nodes are subjected to the same packet arrival process. We consider two different random packet length distributions with the same mean and variance as shown in Table 2 (Two different packet length distributions with the same mean and variance). Table 3 (Influence of higher moments of service time distribution in real network) shows the mean response time observed at the last node (node 4) as the packet arrival rates at all nodes increase. Clearly, in the real network, the shape of the packet length distribution (or service time distribution) beyond the first two moments also has a significant impact on the mean response time at bus nodes when the packet arrival rate increases (hence, the network load increases). In this example, the relative difference in mean response time at node 4 is about 40%.

There are also some potential sources of inaccuracy of the model related to numerical computation. The use of a "catch all" average service time distribution for transmission attempts $j > j^*$ is one potential source. This one is more likely to have an effect when a larger number of interruptions can be expected (e.g. low priority node, heavier bandwidth utilization, service time distributions with higher variability …). Our limitation on the maximum number of stages k^* in the Coxian representation of the service time at a given transmission attempt may also introduce some inaccuracies, notably when C_v becomes very small (e.g. $C_v < 1/\sqrt{k^*}$).

4.2 Performance Evaluation

We now attempt to analyze the accuracy of our PRI model in evaluating the performance of the OPS unslotted bus-based network discussed earlier. We use discrete-event network simulator tool (NS 2.1b8 [24]) to simulate the network with 8 nodes transmitting on one wavelength at 2.5 Gbps. The simulation results are then compared to the results obtained from our PRI model, and from the model proposed in [13] using Jaiswal's results on priority queues (we refer to it as the C-H model). All mean values in our simulation results are estimated with confidence intervals of no more than a

few percent wide around the midpoint at 95% confidence level using Batch Means method [25] (i.e., each mean response time value is computed by collecting at least 7 batches of 800,000 successful packet transmissions each). In the first set of results, illustrated in Figures 5 and 6, we assume that all nodes share the same rate of packet arrivals and service time distribution (uniform traffic profile at all bus nodes).

Table 4. Original service time distributions used for performance study

Distribution	**Mean (*μs*)**	**C_v^2**	**Packet length distribution**
1	2.13	0.125	67% packets of 500 bytes 33% packets of 1000 bytes
2	0.8418	0.7295	53% packets of 50 bytes 47% packets of 500 bytes
3	1.434	1.182	45% packets of 50 bytes 40% packets of 500 bytes 15% packets of 1500 bytes
4	1.02	1.994	64% packets of 50 bytes 26% packets of 500 bytes 10% packets of 1500 bytes
5	1.23	2.524	77% packets of 50 bytes 23% packets of 1500 bytes

Table 4 (Original service time distributions used for performance study) describes service time distributions (packet size mixes) used in this study. The square coefficients of variation of these distributions vary from close to zero to higher than 1, implying the use of both two-stage Coxian and hypoexponential distributions to model the service time distribution in our PRI model.

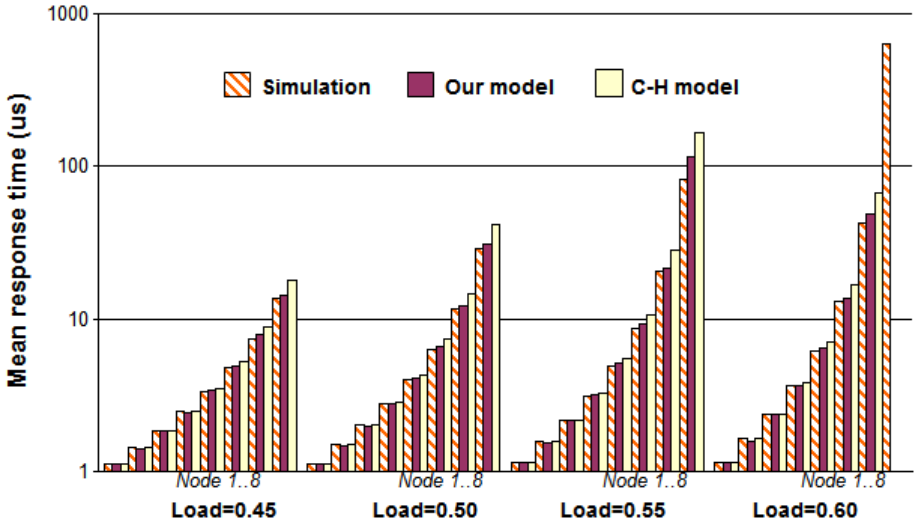


Figure 5. Mean response time at each node as a function of offered network load for service time distribution 4

Notice that distribution 4 is closest to the "real life" service time distribution of Internet traffic (e.g. [26]). Figure 5 (Mean response time at each node as a function of offered network load for service time distribution 4) shows the mean response time of the system at each bus node for packet length distribution 4 as the offered network load increases. The offered network load is defined as the ratio of the sum of traffic volume offered to all nodes to the network transmission capacity. The mean response time at a node is defined as the expected time elapsed from the moment when a client packet arrives at the queue of the node until it is successfully transmitted on the bus. For this experiment, we first observe that both simulation and our analytic model capture the expected behavior of mean response time in OPS bus-based networks: the mean response time is likely to increase rapidly as the node's priority decreases. Moreover, the mean response time tends to "explode" at the lowest priority nodes as the offered network load increases. For instance, for packet size mix distribution 4, simulation results show that the mean response time at node 8 is about 13.7 μs with offered network load of 0.45, and some 640 μs with offered network load of 0.60. An explanation for these results is that the transmission at low priority nodes may be delayed (in our model, interrupted) by the arrivals of packets from higher priority nodes, thus a successful transmission at low priority nodes takes on average a longer time than at higher priority nodes. The number of service interruptions becomes more and more important as the offered network load increases, leading to excessive response time at the lowest priority nodes.

Additionally, we observe in this experiment that when the offered network load is low, the results obtained with our analytical model are very close to those obtained with simulation. For instance, the difference between analytical and simulation results is on the order of only a few percent for all offered network loads below 0.55. But this difference becomes more significant at the last bus nodes as the network load increases (e.g. some 28% relative difference in the mean response time at node 8 when the offered network load is 0.55). We also notice that in comparison with simulation results, our model provides remarkably better results than those obtained with C-H mode at most downstream nodes (node 5 to 8). On the contrary, the C-H model provides results lightly closer to simulation results than our model at first upstream nodes (node 1 to 4).

Note that, with the network load at 0.6, the precise shape of the service time distribution (in terms of moments higher than the first two) starts playing an important role at the most downstream node where the bandwidth (server) is close to saturation. While the network simulation indicates a mean response time of some 640 μs at node 8, both analytical models peg the node as unstable. Interestingly, a direct simulation of the PRI system with the same service time distribution as in the analytical model (i.e., matching only the first two moments of the service time distribution in the network model) also shows that the node 8 becomes overloaded.

It is worth noting that the saturation at the most downstream node is not due to the lack of physical bandwidth (server) capacity, because the bus is actually loaded at merely 60% of its transmission capacity. As stated at the beginning of the paper, this saturation is mostly due to the fact that the physical bandwidth has been fragmented into small segments of bandwidth (voids) between asynchronous transmissions of packets at upstream nodes. Those voids are unusable for most downstream nodes to insert big client packets (i.e., in our model, this is equivalent to a large number of interruptions during the service of big client packets), leading to the "head-of-the-line" (HOL) blocking phenomenon. Clearly, the effect of this phenomenon on most downstream nodes depends not only on the network load, but also on the shape of the packet length distribution. We now specifically study the impact of packet length distribution (or, equivalently, service time distribution) on the network performance.

To assess the behavior of our PRI model with respect to the service time distribution, we focus on the analysis of the mean response time obtained with the same offered network load but with different service time distributions. Figure 6 (Mean response time at each node as a function of service time distribution for offered network load of 0.55) illustrates the mean response time at each node with the offered network load of 0.55 as a function of service time distribution. In this study, we set the offered network load at 0.55 because, as shown by the results of our preceding experiment, beyond this load level the stability condition might not be satisfied for some nodes. As before, the workload is uniform, i.e., statistically identical for all nodes.

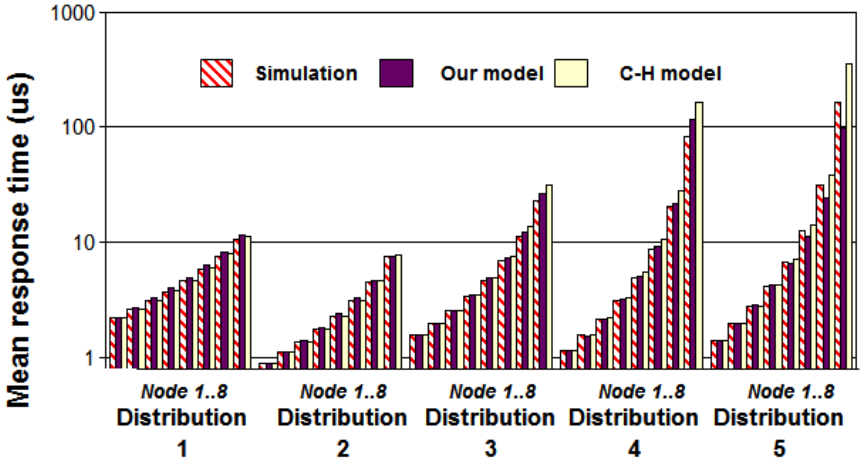


Figure 6. Mean response time at each node as a function of service time distribution for offered network load of 0.55

We observe for the uniform workload considered that the first few nodes on the bus experience little queuing time and few interruptions (for nodes other than the first one). This means that the server viewed by the first few upstream nodes is highly available, servicing clients at those nodes rapidly.

Therefore, the mean response time (which is the sum of mean queuing time and mean service time) at these nodes depends mostly on the mean of the service time (and not higher moments). Indeed, in Figure 6, we notice that for the first upstream nodes (up to node 4), distribution 1 with highest mean of service time (but lowest variance) provides highest mean response time, followed by distribution 3, 4 and 2. For lower priority nodes, the mean response time becomes clearly more sensitive to higher moments of the service time distribution. In particular, when the variance of the service time distribution is high, we observe very high response time at the lowest priority nodes compared to the response time of higher priority nodes. For example, in Figure 6, for distribution 5 with highest C_v^2 of 2.524 (but not highest mean), the simulation shows that the mean response time at node 8 is the highest compared to other distributions, and is some five times higher than the mean response time at node 7, and nearly 120 times higher than the response time at node 1.

The above effect of service time distribution with high variability on the mean response time at most downstream nodes is readily explained by the bandwidth fragmentation phenomenon. The high variability of service time distributions in our experiments means that there is an important percentage of small/medium packets (e.g. 50/500 bytes) and a smaller percentage of big packets (e.g. 1500 bytes) in the offered traffic (see Table 4). From physical perspective, this translates into the fact that the available and usable bandwidth for low priority nodes is strongly reduced because it becomes considerably fragmented into small voids due to the asynchronous insertion of small/medium packets at higher priority nodes. In reality, when an upstream node detects a void, it may insert a packet at the beginning, at the middle or at the end of this void depending on whether a packet is available in the queue at that moment. The insertion of a small/medium packet into a big void will break the void into two small voids, which may be unusable for the transmission of bigger packets at lower priority nodes. Thus the high variability of service time distribution leads to high probability of the HOL blocking phenomenon at the most downstream node, resulting in excessive mean response time at that node. Notice that HOL blocking may not occur if downstream nodes have small packets only.

As far as the accuracy of our model is concerned, we observe in Figure 6 that the difference between our model and simulation results is limited to a few percent when the service time is not highly variable (e.g. distribution 1 and 2), but it becomes larger as the service time becomes more variable (e.g. 11% and 28% relative difference for distribution 3 and 4 respectively). However, our model still provides significantly better result than the C-H model for highly variable service times. For instance, the relative difference between the C-H model and simulation results is about 26% and 50% for distribution 3 and 4 respectively. As mentioned earlier, part of reason for the behavior of these analytical models (which only match the first two moments of the service time distribution) may be due to the influence of moments higher than the first two.

In the numerical results shown in Figure 7 (Mean response time at each node as a function of traffic patterns at offered network load of 0.55), we study the effect of varying patterns of the offered traffic on the network performance. We consider the "real life" service time distribution 4, and we set the offered network load to 0.55. In addition to the uniform traffic considered before, we include the case where the traffic increases uniformly as we move downstream on the bus, the case where traffic decreases uniformly as the node priority decreases, the case where the highest priority node carries 70% of the traffic the remaining nodes sharing uniformly the remaining 30% of the load, and, finally, the case where node 4 carries 70% of the traffic while other network nodes share the remaining 30%. We show the mean response time estimated from the network simulation, obtained from our approximate PRI model, as well as from the C-H model.

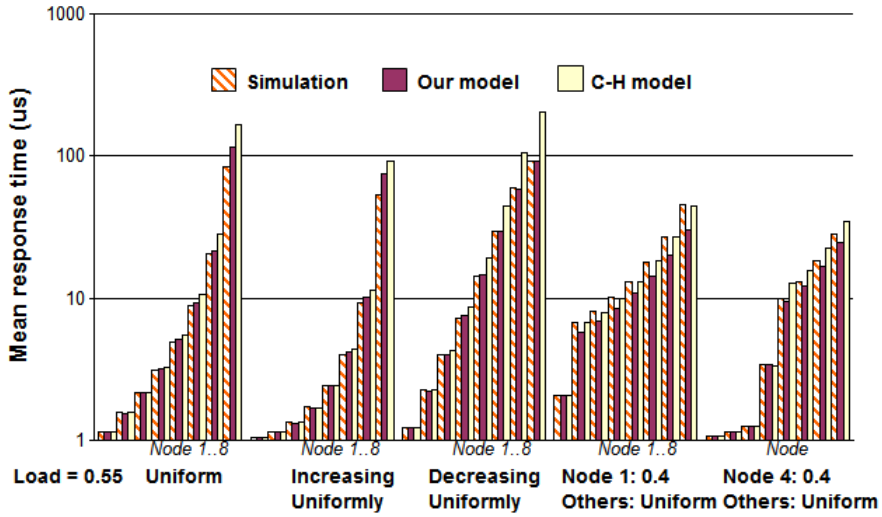
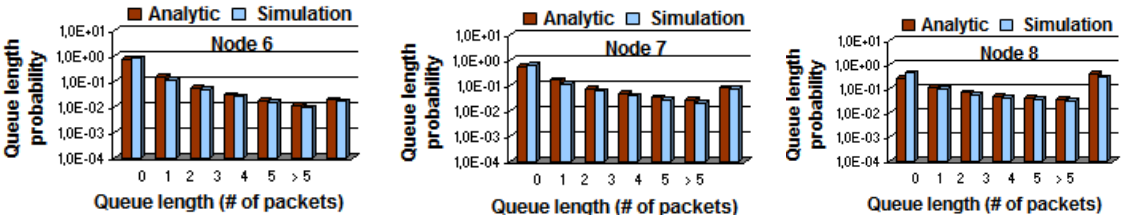
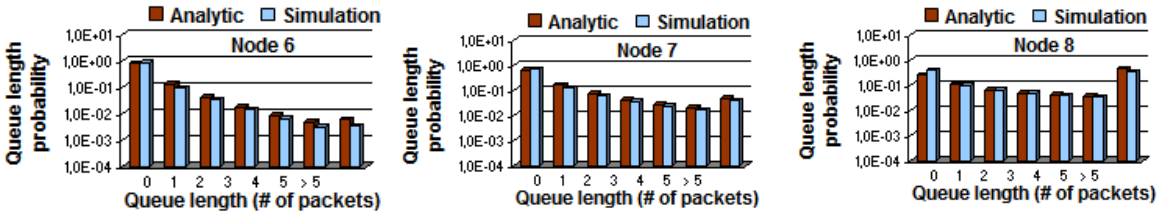


Figure 7. Mean response time at each node as a function of traffic patterns at offered network load of 0.55

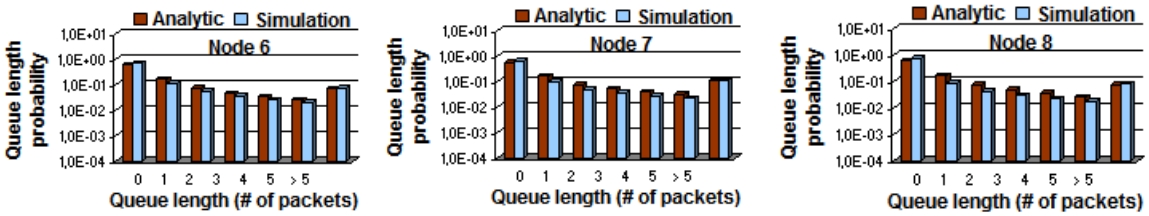
We observe that uniformly decreasing traffic (lower priority nodes carry less traffic) and the uniform traffic pattern appear most penalizing in terms of mean response time at the lowest priority nodes. Interestingly, uniformly increasing traffic and the case where the middle node (node 4) dominates the network seem to fare best. We also observe that, with the possible exception of the last node, the results of our model tend to closely track simulation results, and are in most cases closer to simulation than those of the C-H model (in this experiment, the C-H model provides results closer to simulation results than our model only in the case where the first node dominates the network).



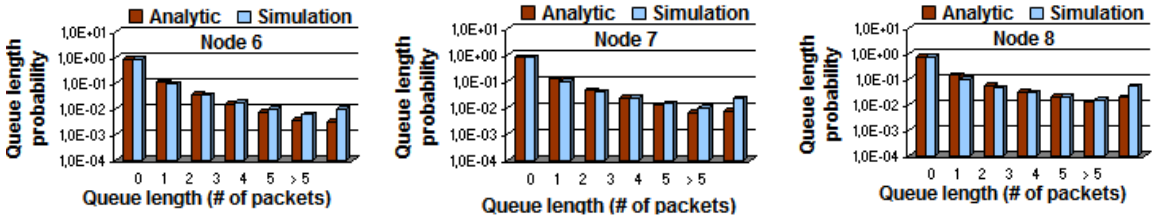
a. Uniform



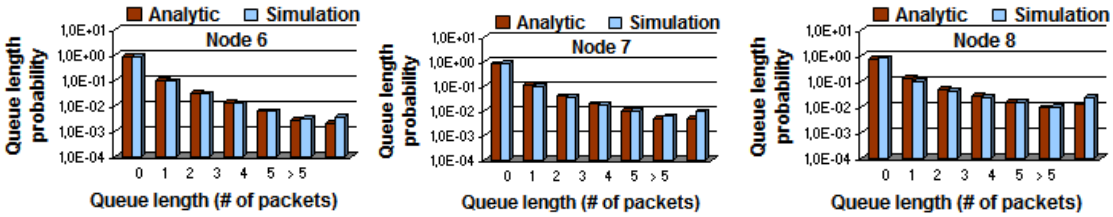
b. Increasing uniformly



c. Decreasing uniformly



d. Node 1: 0,7; Others: Uniform



e. Node 4: 0,7; Others: Uniform

Figure 8. Queue length distribution at the three lowest priority nodes for different traffic patterns at offered network load of 0.55.

One of the advantages of our approach is that it produces the approximate marginal distribution for the number of packets at each node in the simple product form akin to that of an M/M/1-like queue with a state-dependent service time. Such a distribution can then be used to dimension buffers at each node, as well as to assess packet loss ratios. In our approach, we analyze nodes one by one, representing upstream nodes through rates at which the server vanishes and reappears. Clearly, one might be concerned that for lower priority nodes, accumulated approximations might excessively distort the queue-length distribution. In Figures 8.a to 8.e (Queue length distribution at the three lowest priority nodes for different traffic patterns at offered network load of 0.55), we compare the distribution of the number of packets at the three lowest priority nodes for the different traffic patterns considered in Figure 7. We observe that, even for the lowest priority node, our model produces results remarkably close to those obtained from network simulation.

Overall, we think that our model correctly captures the performance characteristics of an OPS bus-based network, including the shape of the stationary distribution of the number of packets at each node. The results of our model may on occasion deviate from simulation results (typically, close to node saturation). As discussed earlier in this section, possible reasons for the observed differences include approximation errors, as well as sensitivity to higher moments of the service time

distribution. It is well known that near saturation in an open queue, even a small difference in service times can amount to a large relative difference in mean response times.

5. CONCLUSION

We have presented an approach to the performance analysis of optical packet switching bus-based networks employing the OU-CSMA/CA protocol and supporting variable length packets. For modeling purposes, we approximately view the bus as a multiple-priority M/G/1 queuing system with preemptive-repeat-identical (PRI) service discipline. We have proposed an approximate solution for this model, in which we apply a recurrent level-by-level analysis. Each level corresponds to a single bus node, and the bandwidth usage by upstream nodes is represented through server disappearance and reappearance rates. To model the PRI discipline, we use different service time distributions at consecutive transmission attempts. The solution to each level is based on conditional probabilities and a fixed point iteration method, which tends to require only a small number of iterations. As a result, we are able to compute not only the mean response time but also the steady-state queue length distribution at each level.

We have used our model to study the expected response time at the nodes of such a bus-based network for several packet length mixes, as well as for several patterns of offered traffic. Our results indicate that a uniform or uniformly decreasing traffic pattern appears more taxing on the network in terms of mean response time at lower priority nodes, while a pattern where the middle node dominates the network traffic seems to fare significantly better. Additionally, for higher traffic levels, the network performance at lower priority nodes is sensitive to the form of the service time distribution as represented by moments higher than the first two.

Comparisons with network simulation results indicate that our model correctly captures the performance characteristics of an OPS unslotted bus (i.e., unfairness property and bandwidth fragmentation phenomenon causing low bandwidth usage and low performance at downstream nodes). In addition, our model is able to provide the shape of the steady-state distribution of the number of packets at each node that closely tracks simulation results, even for the lowest priority node. Compared with other models proposed in the literature such as the C-H model [13], our model in most cases provides better results (i.e., closer to simulation results) than those obtained with the C-H model.

Occasionally, the results of our model may deviate from simulation results. This appears most likely close to node saturation when the service time distribution is highly variable. We have identified approximation errors, as well as sensitivity to higher moments of the service time distribution as possible causes for the observed differences.

In our model, we assume that packets arrivals come from a Poisson source, each node has unlimited buffer space, and we match only the first two moments of the service time distribution. Future work includes an improved matching of the distribution of packet lengths, as well as a possible extension of our approach to different packet arrival patterns and finite buffer sizes.

APPENDIX A. Recurrent Solution for Node 1

Balance equations derivation

With the service time distribution represented by the Coxian distribution of Figure 2, the conditional rate of packet transmissions given the number of packets at node 1 can be expressed as

$$u(n) = p(1 \mid n)\mu_1(1 - p_2) + p(k \mid n)\mu_k , \qquad \text{(Eq.A.1)}$$

where k is the total number of exponential stages in the Coxian distribution, $p(l \mid n)$ is the conditional probability for the current service stage given n, for $l = 1,...,k$ and $n = 1,...$

It is not difficult to show that the steady-state probability distribution for n can be expressed as (where G is a normalizing constant):

$$p(n) = \frac{1}{G}\prod_{i=1}^{n} \lambda / u(i) . \qquad \text{(Eq.A.2)}$$

We show the proof of (Eq.A.2) in Appendix C. Using (Eq.A.2) and the fact that $p(n,l) = p(n)p(l \mid n)$ in the balance equations for $p(n,l)$, we obtain the following equations for $p(l \mid n)$ when $n = 1$.

$$p(2 \mid n)[\lambda + \mu_2] = p(1 \mid n)\mu_1 p_2 ,$$

$$p(l \mid n)[\lambda + \mu_l] = p(l-1 \mid n)\mu_{l-1} , \; l = 3,...,k .$$

Hence, for $n = 1$ we have

$$p(2 \mid n) = p(1 \mid n)\mu_1 p_2 / [\lambda + \mu_2] ,$$

$$p(l \mid n) = p(l-1 \mid n)\mu_{l-1} / [\lambda + \mu_l] , \text{ for } l = 3,...,k , \qquad \text{(Eq.A.3)}$$

where $p(1 \mid n)$ is readily determined from the normalizing condition $\sum_{l=1}^{k} p(l \mid n) = 1$ that must hold for all values of $n = 1,...$ For values of $n > 1$, we have

$$p(2 \mid n)[\lambda + \mu_2] = p(1 \mid n)\mu_1 p_2 + p(2 \mid n-1)u(n) ,$$

$$p(l \mid n)[\lambda + \mu_l] = p(l-1 \mid n)\mu_l + p(l \mid n-1)u(n) , \qquad \text{(Eq.A.4)}$$

for $l = 3,...,k$.

Fixed-point iteration method for solution of the balance equations

Starting with the known solution for $n = 1$, together with the normalizing condition $\sum_{l=1}^{k} p(l \mid n) = 1$, we can solve (Eq.A.4) as recurrence for increasing values of $n = 2,...$. In theory, since there is no upper limit to the values of n, there would be an infinite number of equations to solve. In practice, for the service time distributions considered, the conditional probabilities $p(l \mid n)$ and the conditional rate of packet transmissions $u(n)$ quickly reach limiting values as n increases. In the examples explored, convergence was typically achieved for values of n on the order of a few tens. Clearly, knowing $u(n)$ we can use (Eq.A.2) to compute $p(n)$ and any performance measures derived from it.

APPENDIX B. Recurrent Solution for Nodes *i > 1*

Balance equations derivation

We describe the state of a downstream node by the triple (n, j, l) where n is the current number of client packets at the node, j is the transmission attempt, and l is equal to the current number of the service stage at this attempt $(1,..., k_j)$, or equal to 0 if the server (bandwidth) is unavailable. Let α be the rate with which the server becomes unavailable, and β be the rate with which the server returns from unavailability. The balance equations for $p(n, j, l)$ are readily derived.

$$p(n, j = 1, l = 0)[\lambda + \beta] = p(n-1, j = 1, l = 0)\lambda ;$$

$$p(n, j, l = 0)[\lambda + \beta] = \sum_{l=1}^{k_{j-1}} p(n, j-1, l)\alpha + p(n-1, j, l = 0)\lambda, \quad j = 2,.., j^* - 1;$$

$$p(n, j, l = 1)[\lambda + \mu_1^j + \alpha] = p(n, j-1, l = 0)\beta + p(n-1, j, l = 1)\lambda ;$$

$$p(n, j = 1, l)[\lambda + \mu_l^j + \alpha] = p(n, j = 1, l-1)\mu_{l-1}^j \overline{q}_{l-1}^j + p(n-1, j = 1, l)\lambda , \; l > 1 ;$$

$$p(n, j^*, l = 0)[\lambda + \beta] = \sum_{l=1}^{k_{j^*-1}} p(n, j^* - 1, l)\alpha + \sum_{l=1}^{k_{j^*}} p(n, j^*, l)\alpha + p(n-1, j^*, l = 0)\lambda ;$$

$$p(n, j^*, l = 1)[\lambda + \mu_1^{j^*} + \alpha] = p(n, j^* - 1, l = 0)\beta + p(n, j^*, l = 0)\beta + p(n-1, j^*, l = 1)\lambda ;$$

$$p(n, j^*, l)[\lambda + \mu_l^{j^*} + \alpha] = p(n, j^*, l-1)\mu_{l-1}^{j^*} \overline{q}_{l-1}^{j^*} + p(n-1, j^*, l)\lambda , \quad l > 1 .$$

In the above equations, we denote by μ_l^j the parameter of stage l of the Coxian representation of the service time at transmission attempt j, and by $\overline{q}_l^j$ the corresponding probability that stage l will be followed by another service stage, where $l = 1,..., k_j$. For the "catch all" value j^*, we use "average" service parameters values set up so as to maintain the proper average number of transmission attempts.

For $n = 1$, the first balance equation becomes

$$p(n = 1, j = 1, l = 0)[\lambda + \beta] = p(n = 0, l = 0)\lambda ,$$

where $p(n = 0, l = 0)$ is the probability that there are no packets at the node and the server is unavailable. In the remaining equations for $n = 1$, the term involving $n - 1$ is simply absent.

The conditional rate of packet transmissions given the number of packets at the node can be expressed as:

$$u(n) = \sum_{j \geq 1} \sum_{l=1}^{k_j} p(j, l \mid n)\mu_l^j (1 - \overline{q}_l^j) . \qquad \text{(Eq.B.1)}$$

As for node 1, the steady-state probability distribution for n can be expressed as (G is a normalizing constant):

$$p(n) = \frac{1}{G}\prod_{i=1}^{n} \lambda / u(i)\,. \qquad \text{(Eq.B.2)}$$

Using (Eq.B.2) together with the fact that $p(n,j,l) = p(n)p(j,l \mid n)$, we transform the balance equations for $p(n,j,l)$ into equations for the conditional probability $p(j,l \mid n)$:

$$p(j=1,l=0 \mid n)[\lambda+\beta] = p(j=1,l=0 \mid n-1)u(n)\,;$$

$$p(j,l=0 \mid n)[\lambda+\beta] = \sum_{l=1}^{k_{j-1}} p(j-1,l \mid n)\alpha + p(j,l=0 \mid n-1)u(n),\ \ j=2,..,j^*-1\,;$$

$$p(j,l=1 \mid n)[\lambda+\mu_1^j+\alpha] = p(j-1,l=0 \mid n)\beta + p(j,l=1 \mid n-1)u(n)\,;$$

$$p(n,j=1,l)[\lambda+\mu_l^j+\alpha] = p(n,j=1,l-1)\mu_{l-1}^j \overline{q}_{l-1}^j + p(n-1,j=1,l)\lambda\,,\ l>1\,;$$

$$p(j^*,l=0 \mid n)[\lambda+\beta] = \sum_{l=1}^{k_{j^*-1}} p(j^*-1,l \mid n)\alpha + \sum_{l=1}^{k_{j^*}} p(j^*,l \mid n)\alpha + p(j^*,l=0 \mid n-1)u(n)\,;$$

$$p(j^*,l=1 \mid n)[\lambda+\mu_1^{j^*}+\alpha] = p(j^*-1,l=0 \mid n)\beta + p(j^*,l=0 \mid n)\beta + p(j^*,l=1 \mid n-1)u(n)\,;$$

$$p(j^*,l \mid n)[\lambda+\mu_l^{j^*}+\alpha] = p(j^*,l-1 \mid n)\mu_{l-1}^{j^*}\overline{q}_{l-1}^{j^*} + p(j^*,l \mid n-1)u(n)\,,\ \ l>1.$$

For $n=1$, the first equation becomes:

$$p(j=1,l=0 \mid n=1)[\lambda+\beta] = p(l=0 \mid 0)u(1)\,,$$

where $p(l=0 \mid 0)$ is the conditional probability that the server is unavailable given that there are no packets to be transmitted at the node. Note that we must have

$$\sum_{j\geq 1}\sum_{l=0}^{k_j} p(j,l \mid n) = 1\,, \qquad \text{(Eq.B.3)}$$

for all values of $n \geq 1$.

For $n=0$, the only possible states correspond to the availability of the server, and we easily get for the probability that the server is unavailable

$$p(l=0 \mid 0) = \alpha/[\alpha+\beta+\lambda]\,. \qquad \text{(Eq.B.4)}$$

Fixed-point iteration method for solution of the balance equations

Armed with the known value of $p(l=0 \mid 0)$, we consider the above equations for $p(j,l \mid n)$ for increasing values of $n = 1,2,...$ For each n, we express all $p(j,l \mid n)$ in terms of $p(j=0,l=1 \mid n)$ and $u(n)$. Then, these two unknowns are determined from the normalizing condition (Eq.B.3) and from the definition of $u(n)$ in (Eq.B.1). As was the case for node 1, although in theory there is an infinite number of values of n (and hence an infinite number of equation sets to solve), in practice, the conditional probabilities $p(j,l \mid n)$ and the conditional rate of packet transmissions $u(n)$ quickly reach limiting values as n increases.

Knowing $u(n)$ we readily obtain $p(n)$ from (Eq.B.2). The steady-state probability distribution for the number of transmission attempts at node i can then be expressed as

$$r(j) = \frac{1}{H}\sum_{n} p(n)\sum_{l} p(j,l \mid n)\mu_l^j (1-\bar{q}_l^{\,j}),$$

where H is a normalizing constant.

The expected number of packets at the node is given by $\sum_{n=1}^{\infty} np(n)$, and the expected number of interruptions per transmission (due to void too small, in actual network) at node other than the first one is approximately $\sum_{n>0} p(n)\sum_{j}\sum_{l>0} p(j,l)\alpha/\lambda$.

APPENDIX C. Proof of Product-Form for Coxian Distribution

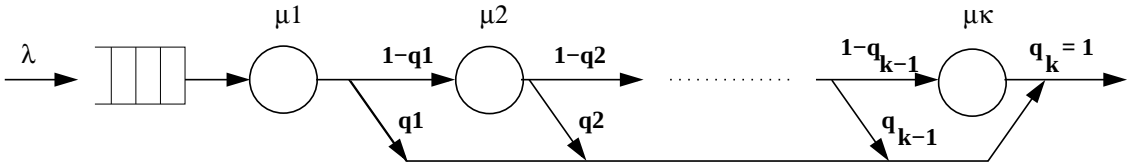


Figure 9. General Coxian distribution

Consider the general Coxian system in Figure 9 (General Coxian distribution). We describe the equilibrium behavior of this system by the joint steady state probability $p(n,l)$, where n $(n \geq 1)$ is the number of packets in the system and $l = 1...k$ refers to the stage of service of the Coxian distribution. The balance equations for $p(n,l)$ are readily derived as follows:

$$p(n,1)[\lambda + \mu_1] = p(n-1,1)\lambda + \sum_{l=1}^{k} p(n+1,l)\mu_l q_l, \qquad \text{(Eq.C.1)}$$

$$p(n,l)[\lambda + \mu_l] = p(n-1,l)\lambda + p(n,l-1)\mu_{l-1}(1-q_{l-1}), \text{ for all } l = 2, 3, ..., k. \qquad \text{(Eq.C.2)}$$

Using the definition of conditional probability $p(n,l) = p(n)p(l \mid n)$ in (Eq.C.1) and (Eq.C.2), and then summing equations (Eq.C.1) and (Eq.C.2) for all $l = 1, ..., k$ while taking into account the normalizing condition $\sum_{l=1}^{k} p(l \mid n) = 1$ for all $n = 1, 2, \ldots$, we obtain:

$$p(n)[\lambda + \sum_{l=1}^{k} p(l \mid n)\mu_l] = p(n-1)\lambda + p(n+1)\sum_{l=1}^{k} p(l \mid n+1)\mu_l q_l + p(n)\sum_{l=1}^{k-1} p(l \mid n)\mu_l(1-q_l).$$

Simplifying the above equation, we have:

$$p(n)[\lambda + \sum_{l=1}^{k} p(l \mid n)\mu_l q_l] = p(n-1)\lambda + p(n+1)\sum_{l=1}^{k} p(l \mid n+1)\mu_l q_l. \qquad \text{(Eq.C.3)}$$

Set $u(n) = \sum_{l=1}^{k} p(l \mid n)\mu_l q_l$ for *n = 1, 2, ...*, (Eq.C.3) becomes:

$$p(n)[\lambda + u(n)] = p(n-1)\lambda + p(n+1)u(n+1). \qquad \text{(Eq.C.4)}$$

We observe that equation (Eq.C.4) is identical to the steady state balance equation of an M/M/1 queuing system with arrival rate λ and service rate *u(n)*. Thus, we are able to readily obtain the product-form for the general Coxian system: $p(n) = \frac{1}{G}\prod_{i=1}^{n} \lambda / u(i)$, where G is a normalizing constant: $G = 1 + \sum_{n \geq 1} \prod_{i=1}^{n} \lambda / u(i)$.

Note that this result is obviously applicable for the k-stage Coxian distribution shown in Figure 2 since it is a special case of the general Coxian distribution.

APPENDIX D. Computation of Server Disappearance Rate (β) at a Nodes *i > 1*

As stated earlier, the disappearance rate β_i of the server viewed by a downstream node *i* is computed approximately in our solution. For the node 2, this rate can be expressed in terms of conditional probabilities as follows:

$$\beta_2 = p(n_1 = 1 \mid n_2, U_2) \sum_{l=1}^{k} \mu_l q_l p(l \mid n_1 = 1, n_2, U_2), \qquad \text{(Eq.D.1)}$$

where n_i is the current number of packets at node *i*, U_i (A_i respectively) indicates the server at node *i* is unavailable (available respectively), and k, l, μ_l, q_l are the parameters of the service time distribution as shown in Figure 9. Using the fact that $p(U_2) = p(n_1 \geq 1) = 1 - p(n_1 = 0)$ and $u(n) = \sum_{l=1}^{k} p(l \mid n)\mu_l q_l$, equation (Eq.D.1) can be approximately computed based on known parameters (i.e. parameters computed from the solution of the preceding node) as follows:

$$\beta_2 \approx p(n_1 = 1 \mid U_2) \sum_{l=1}^{k} \mu_l q_l p(l \mid n_1 = 1) \approx \frac{p(n_1 = 1)u(n_1 = 1)}{1 - p(n_1 = 0)}. \qquad \text{(Eq.D.2)}$$

Regarding nodes *i* > 2, the exact expression of β_{i+1} is:

$$\begin{aligned} \beta_{i+1} &= p(n_i = 1 \mid n_{i+1}, U_{i+1}) \sum_{l=1}^{k} \mu_l p_l p(l \mid n_i = 1, n_{i+1}, U_{i+1}) \\ &+ p(n_i = 0 \mid n_{i+1}, U_{i+1}) p(U_i \mid n_i = 0, n_{i+1}, U_{i+1}) \beta_i(n_i). \end{aligned} \qquad \text{(Eq.D.3)}$$

We notice in the right hand side of this expression that in addition to the first term similar to the one in Eq.D.1, we introduce the second term that represents the case where the server at the preceding node *i* has been occupied and its queue was empty. This case does not exist for node 2 because the node 1 (which is the preceding node of node 2) always finds server available. Similar to the case for node 2, we can compute the value of β_{i+1} approximately:

$$\beta_{i+1} \approx p(n_i = 1 \mid U_{i+1}) \sum_{l=1}^{k} \mu_l p_l p(l \mid n_i = 1) + p(n_i = 0 \mid U_{i+1}) p(U_i \mid n_i = 0) \beta_i$$
$$\approx \frac{p(n_i = 1) u(n_i = 1) + p(n_i = 0) p(U_i \mid n_i = 0) \beta_i}{p(U_{i+1})} \quad . \qquad \text{(Eq.D.4)}$$

Here, the probability that the server is unavailable at node i+1 (i > 1) is computed as follows:

$$p(U_{i+1}) = 1 - p(n_i = 0, A_i) = 1 - p(A_i \mid n_i = 0) p(n_i = 0) \, .$$

Using the fact that $p(A_i \mid n_i = 0) = 1 - p(U_i \mid n_i = 0)$ with $p(U_i \mid n_i = 0)$ computed from (Eq.B.4), we are able to determine the value of β_{i+1} from (Eq.D.4):

$$\beta_{i+1} \approx \frac{p(n_i = 1) u(n_i = 1) + p(n_i = 0) p(U_i \mid n_i = 0) \beta_i}{1 - \{1 - p(U_i \mid n_i = 0)\} p(n_i = 0)} \, . \qquad \text{(Eq.D.5)}$$

REFERENCES

[1] Nguyen, V.H., Ben Mamoun, M., Atmaca, T., et al. Performance evaluation of Circuit Emulation Service in a metropolitan optical ring architecture. *In Proceedings of the Telecommunications and Networking – ICT 2004*. LNCS 3124 vol., pp. 1173-1182, Fortaleza – Brazil, August 2004.

[2] Le Sauze, N., Dotaro, E., Dupas, A., et al. DBORN: A Shared WDM Ethernet Bus Architecture for Optical Packet Metropolitan Network. In *Proceedings of Photonic in Switching Conference*. July 2002.

[3] White I.M., A New Architecture and Technologies for High-Capacity Next Generation Metropolitan Networks, *Ph.D. Dissertation*, Department of Electrical Engineering of Stanford University, CA, August 2002.

[4] LaMaire, R.O. An M/G/1 Vacation Model of an FDDI Station. *IEEE Journal on Selected Areas in Communications*. Vol.9, Issue 2, Feb. 1991, pp.257 – 264.

[5] Rubin, I., and WU, J.C.H. Analysis of an M/G/1/N Queue with Vacations and its Application to FDDI Asynchronous Timed-Token Service Systems. In *Global Telecommunications Conference (GLOBECOM '92)*. Communication for Global Users, vol.3, pp.1630 - 1634.

[6] Ghafir, H.M., and Silio, C.B. Performance Analysis of a Multiple-Access Ring Network. *IEEE Transactions on Communications*. Vol. 41, Issue 10, Oct. 1993, pp.1494 – 1506.

[7] Mukherjee, B., and Banerjee, S. Alternative Strategies for Improving the Fairness in and an Analytical Model of the DQDB Network. *IEEE Transactions on Computers*. Vol. 42, Issue 2, Feb. 1993, pp. 151 – 167.

[8] Starvrakakis, I., and Landry, R. Delay Analysis of the DQDB MAN based on a Simple Model. *IEEE International Conference on Communications*. Vol.1, June 1992, pp. 154 - 158.

[9] Mukherjee, B., and Meditch, J. The p_i-Persistent Protocol for Unidirectional Broadcast Bus Networks. *IEEE Transactions on Communications*. VoL.36, Issue 12, Dec. 1988, pp. 1277 – 1286.

[10] Miller, G.J., and Paterakis, M. A Dynamic Bandwidth-Allocation-Based Priority Mechanism for the p_i-Persistent Protocol for MAN's. *IEEE Journal on Selected Areas in Communications*. Vol.11, No. 8, October 1993.

[11] Takine, T. Y., Takahashi, and Hasegawa, T. An Approximate Analysis of a Buffered CSMA/CD. *IEEE Transactions on Communications*. Vol.36, Issue 8, Aug. 1988, pp. 932 – 941.

[12] Matsumoto, Y., Takahashi, Y., and Hasegawa, T. The Effects of Packet Size Distributions on Output and Delay Processes of CSMA/CD. *IEEE Transactions on Communications*. Vol. 38, Issue 2, Feb. 1990, pp. 199 – 214.

[13] Castel, H., and Hébuterne, G. Performance Analysis of an optical MAN ring for variable length packet traffic. In *Proceedings of Photonic in Switching Conference*. 2003.

[14] Castel, H., Chaitou, M., and Hébuterne, G. Preemptive Priority Queues for the Performance Evaluation of an Optical MAN Ring. In *Proceedings of Performance Modeling and Evaluation of Heterogeneous Networks (Het-Net'05)*. 2005.

[15] Hu, G., Gauger, C.M., and Junghans, S. Performance Analysis of the CSMA/CA MAC Protocol in the DBORN Optical MAN Network Architecture. In the *Proceedings of the 19th International Teletraffic Congress (ITC 19)*. 2005.

[16] Bouabdallah, N., Beylot, A.L., Dotaro, E., and Pujolle, G. Resolving the Fairness Issue in Bus-Based Optical Access Networks. *IEEE Journal on Selected Areas in Communications*. Vol. 23, No. 8. August 2005.

[17] Jaiswal. *Priority Queues*. New York: Academic 1966.

[18] Takagi, H. *Queuing Analysis*. Vol.1, pp. 365-373. North-Holland, 1991.

[19] Cox, D. R., and Smith, Walter L. *Queues*. John Wiley, New York, 1961.

[20] Trivedi, K.S. *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*. Prentice-Hall, Inc., Englewood Cliffs, NJ 07632.

[21] Brandwajn, A. Equivalence and Decomposition in Queueing Systems - A Unified Approach. *Performance Evaluation*, vol. 5, pp. 175-186, 1985.

[22] Brandwajn, A and Wang, H. A Conditional Probability Approach to M/G/1-like Queues. Submitted for publication, available as a technical report, 2006.

[23] Allen, A.O. *Probability, Statistics, and Queuing Theory*. Academic Press, 2nd edition, 1990.

[24] Network Simulator. Available at HTTP: http://www.isi.edu /nsnam/ns/.

[25] McDougall, M.H. *Simulating Computer Systems: Techniques and Tools*. The MIT Press, 1987.

[26] CAIDA, "IP packet length distribution", [Online document] 2000, Available at HTTP: http://www.caida.org/analysis/AIX/plen_hist.

Alexandre Brandwajn holds a Ingénieur Civil des Télécommunications degree from the Ecole Nationale Supérieure des Télécommunications in Paris, France, and a Docteur ès-Sciences degree from the University

of Paris. After spending a few years as a researcher at INRIA near Paris, he was Associate Professor and later Professor of Computer Science at ENST in Paris. He then worked for close to six years for Amdahl Corporation in Sunnyvale, CA, USA, first as a Senior Computer Architect, and then as Manager of the Systems Analysis Department, before joining the Computer Engineering Department of the University of California at Santa Cruz. Dr. Brandwajn is the author of over 20 journal publications and numerous conference presentations. His current research interests include efficient solutions of systems with large state space, priority systems, models of I/O subsystems, virtualized systems and models of optical networks.

Viet Hung NGUYEN received the Master of Science degree in distribubted computer systems from University of Paris XI (Orsay) in 2002, and the Ph.D. degree in computer science from the Institut National des Telecommunications at Evry, France in 2006.

He is currently researcher at the Institut National des Telecommunications in France. His primary research area is performance modelling and evaluation of telecommunication networks. In this area, Dr. Nguyen has been working on performance study of optical packet switching networks, with an emphasis on access protocols (MAC) and quality of service (QoS) mechanisms. He is involved in numerous European and French national research projects on optical networking: EuroNGI, ROM-EO, Carriocas, EcoFrame, etc.

Tülin ATMACA received the PhD degree in Computer Science from The University of Paris XI (ORSAY), in France in 1987. From 1986 to 1988 she has taught in the Department of Computer Science at the University of Paris XI. From 1989 to 1991, she was Assistant Professor in the Computer Science Department at the University of Paris VI. Since January 1992, she has been Associate Professor at National Institute of Telecommunications (INT) in France, in the networking and service department. She was a Visiting Professor in the Computer Science Department of the North Carolina State University, as well as the department of Industrial Engineering of the Rutgers University. Her research interests are in the areas of performance evaluation of telecommunication networks (ATM, FR, Optical Networks, TCP/IP,...), of traffic and congestion control, and Quality of Services aspects in these networks. She is involved in several national and international research projects in the field of the optical packet switching network and their performance (ROM, ROM-EO, EuroNGI, Industrial research projects). She served on the program committees and was co-chair of various well-known international conferences (AICT'06, ICDT'06). Her recent research interests include the all-optical multi-service networks: WAN and MAN and the performance of optical access networks.