

An Application of Stochastic Ordering to the Analysis of the Push-Out Mechanism*

J.M. Fourneau^a and N. Pekergin^{a,b} and H. Taleb^a

^aPRiSM, Université de Versailles-Saint-Quentin,
45, Av. des États-Unis 78000 Versailles, France

^bCERMSEM, Université de Panthéon-Sorbonne,
90, Rue de Tolbiac, 75013 Paris, France

In this paper, we compute the stochastic bounds on the cell loss rates in an ATM switch. The spatial priority in the buffer is controlled by the Push-Out mechanism, while the time priority is managed in the FIFO manner. We consider an i.i.d arrival process of cells and a constant switching time of a cell. Therefore, the system can be modelled by a discret-time Markov chain, however the size of the chain is approximately 2^B , where B is the buffer size. We propose a methodology based on the stochastic ordering to aggregate the underlying Markov chain to obtain a bounding Markov chain. In other words, the performance indices defined by the reward functions are bounded stochastically by the reward functions of the bounding Markov chain. We apply the methodology twice to have the bounding Markov chain reduced to B^2 states and finally to B states. Several bounds have been computed under various assumptions and they prove that the proposed methodology is numerically efficient.

Key Words Codes: C.2.0; I.6.3; G.3

Keywords: Computer-Communication networks, General; Simulation and Modeling, Application; Probability and Statistics

1. INTRODUCTION

Cell loss rates are one of the key problems for broadband ISDN. In an ATM switch, buffers have a finite capacity and may receive two flows of cells which have distinct cell loss requirements. One bit of the cell header (CLP bit) is used to indicate the cell loss priority level. Several buffer management schemes such as the Push-Out or the Partial Buffer Sharing mechanisms have been proposed to assure a low loss probability for high priority cells. In the sequel, we denote as high priority cells, the cells which have the highest requirements, i.e., the smallest loss ratio requirements.

Cells have a fixed size; thus service times are constant. The mixture of data, video and voice traffic is so bursty that Poisson arrivals assumption is not valid anymore. Processes like Markov Modulated Poisson Processes (MMPP) or Switched Batch Bernoulli Processes

*This work has been supported by CNET under grant 92 1B 178 on formal design of cooperative high-speed multimedia systems.

(SBBP) have been proposed to capture the bursty nature of traffic in ATM networks. However, these assumptions do not lead usually to exact analytical results.

These assumptions have been only partially handled by several authors. Usually, the results are obtained because of some simplification of the assumptions or after some approximations. Doshi and Heffes [1] have described and analyzed an overload control algorithm using the Push-Out scheme for the $M/M/1/N$ queue. In [2], Hébuterne and Gravey have obtained the loss probabilities assuming a Poisson arrival process, deterministic service time and LIFO replacement policy. Similarly, Kröner [3] has proposed a method to compute the loss rates of an $M_1, M_2/G/1/N$ Push-Out system with FIFO service discipline. Three different space priority mechanisms (Push-Out, Partial Buffer Sharing and separate route for each class) were considered. Finally, Cheng and Akyildiz [4], have analyzed a $M_1, M_2/G_1, G_2/N$ queue with different scheduling and Push-Out discipline. They have obtained an exact solution for loss probabilities of both classes. As an ATM switch is basically a discrete time system, a few works in discrete time queues without buffer management schemes [5, 6] have also been published.

Algorithmic results are also quite difficult to obtain (see [7] for a survey on the use of ad-hoc numerical methods to evaluate the cell loss rates). Finally, as the state spaces are so large, standard numerical analysis of Markov chains seems intractable.

So, we advocate the use of a new methodology based on stochastic ordering on Markov chains which leads to an efficient numerical solution. Assume that we have modelled the problem using a very large Markov chain. We need to compute its steady-state distribution in order to obtain reward functions (for instance, the cell loss rates for a finite capacity buffer). The key idea of the proposed methodology is to design a new chain such that the reward functions will be upper or lower bounds of the exact reward functions. This new chain is an aggregated model of the former one. These bounds and the aggregation criteria are based on some stochastic orderings applied to Markov and semi-Markov processes (see Stoyan [8] and other references therein). As we drastically reduced the state space, we may now use numerical methods to efficiently bound the problem. Note that bounding some reward functions is often sufficient while modeling ATM networks as we only need to verify the requirements for the QoS.

We apply this methodology to the analysis of a buffer management policy : the Push-Out mechanism. Since the switching time is constant, the study is performed in discrete-time where the time unit is equal to the cell switching time. We assume that the arrivals follow a Bernoulli Batch Process. Such a system is easily represented by a discrete-time Markov chain, however the size of the chain is approximatively 2^B where B is the buffer size.

To apply stochastic ordering, we must first find a natural ordering on the states of the Markov chain. The stochastic properties we may use are dependent of the properties of this ordering of the states. If it is a total order, then we may use the coupling technique to obtain strong stochastic ordering. If the order is only a partial order, then weaker stochastic orderings may be obtained.

In this paper, we only consider a simple arrival process (an i.i.d. batch arrival process). Indeed, Poisson arrival process and batch arrival process may lead to a totally ordered Markov chain whereas Markov arrival process or Markov modulated arrival process are associated to partial order on vector spaces. In this paper, we introduce the methodology

for totally ordered Markov chains. In a companion paper [9], we show how to analyze partially ordered Markov chains.

We apply the methodology twice. In the first step, we prove that the cell loss rates for high priority cells is stochastically bounded by the same reward function of the system with Head-Of-Line (HOL) service discipline. Such disciplines mixing time and space priorities have been considered in [10]. As the HOL service discipline is easier to represent than the FIFO discipline, the size of the Markov chain is reduced to B^2 . In the second step, we perform another aggregation of the states to obtain a chain with B states. In this aggregated chain we define the transition rates between states in order to obtain a bounding Markov chain. Note that the bounds we obtain are not trivial. Several bounds have been computed under various assumptions and they prove that the proposed methodology is numerically efficient.

The paper is organized as follows: the next section is devoted to the modelling of the Push-Out mechanism. In Section 3 and 4 the stochastic bounds are derived, and the results are presented in section 5. Finally, we introduce the concept of stochastic ordering and coupling method in the appendix.

2. THE MODEL

We consider a buffer policy which uses the Push-Out mechanism for the buffer management in ATM networks. We assume that there exists two types of cells with distinct loss rate requirements. In the Push-Out mechanism

- when the buffer is not full, low and high priority cells can be stored in the buffer
- when the buffer is full, an arriving low priority cell is lost, while an arriving high priority cell pushes out of the buffer a low priority cell if there is any in the buffer, otherwise the high priority cell is lost. The deletion discipline is LIFO.

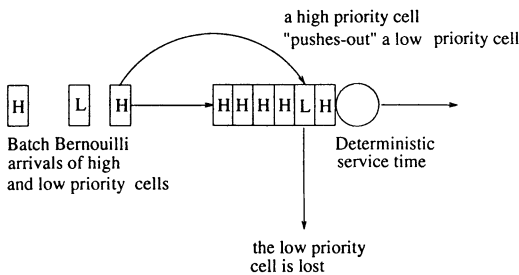


Figure 1. Push-Out mechanism description

We consider three service disciplines :

- [FIFO :] there is no time-priority between cells; so the service scheduling is First-In-

First-Out.

• [HOL :] We consider two service disciplines with Head-Of-Line priority (these mixed disciplines have been considered in [10] :

- [HOL1 :] in which, the low priority cells are scheduled before high priority cells,
- [HOL2 :] in which, the high priority cells are scheduled before low priority cells.

In an ATM network cells have a fixed size of 53 bytes, so their switching time is constant. Since we observe the system at the cell departure epochs, the time can be considered to be divided in slots, which are equal to the cell switching time. The cell arrivals are modelled by an i.i.d. *Batch Process*. We assume that the batch size varies between 0 and M , and the buffer size is B .

Let t_i be the departure time of the i th cell. For the Head-Of-Line disciplines, the state of the system at time t_i is presented by the vector

$$N_{HOL}(t_i) = (N1(t_i), N2(t_i)),$$

where $N1(t_i)$ is the total number of cells in the buffer and $N2(t_i)$ is the number of high priority cells in the buffer at time t_i .

Clearly, we must change the state description, if we consider FIFO service discipline. Since we must also know the state of the buffer to determine the next cell to be sent, we will add a third component, denoted as $B(t_i)$ representing the state of the buffer. As usual, this component is an ordered list of cells in the queue, i.e.,

$$N_{FIFO}(t_i) = (N1(t_i), N2(t_i), B(t_i)).$$

Note that the first two components are not necessary anymore; these information can be easily obtained from $B(t_i)$.

Let A_{i+1} ($A_{i+1}(high)$) be respectively the total number of cells (the number of high priority cells) arrived to the buffer in the time interval $[t_i - t_{i+1}[$. Note that this implies that the arrivals take place just after the service completion epochs.

Let $\mathbb{I}_{\{\text{condition}\}}$ be a binary valued $\{0, 1\}$ random variable, which equals 0 if the *condition* is not true, and which equals 1 if the condition is true. We denote $(x)^+$ for $\max\{x, 0\}$. The evolution equations of HOL1 and HOL2 service disciplines are :

- If the service priority is given to low priority cells (HOL1) :

$$N_{HOL1}(t_{i+1}) = (\min\{B, (N1_{HOL1}(t_i) - 1)^+ + A_{i+1}\}, \min\{B, N2_{HOL1}(t_i) - \mathbb{I}_{\{N1_{HOL1}(t_i)=N2_{HOL1}(t_i)\}} \mathbb{I}_{\{N1_{HOL1}(t_i)>0\}} + A_{i+1}(high)\}) \quad (1)$$

- If the service priority is given to high priority cells (HOL2):

$$N_{HOL2}(t_{i+1}) = (\min\{B, (N1_{HOL2}(t_i) - 1)^+ + A_{i+1}\}, \min\{B, (N2_{HOL2}(t_i) - 1)^+ + A_{i+1}(high)\}); \quad (2)$$

- there is no service priority; the cells are scheduled according to FIFO discipline :

$$N_{FIFO}(t_{i+1}) = (\min\{B, (N1_{FIFO}(t_i) - 1)^+ + A_{i+1}\}, \min\{B, N2_{FIFO}(t_i) - \mathbb{1}_{\{\text{head}(t_i)=\text{high}\}} + A_{i+1}(high)\}, B(t_i) \ominus \text{head}(t_i) \oplus A_{i+1}\}. \quad (3)$$

where $\mathbb{1}_{\{\text{head}(t_i)=\text{high}\}}$ is equal to 1, if the first cell in the buffer is a high priority cell at time t_i . $B(t_i) \ominus \text{head}(t_i)$ denotes the state reached from $B(t_i)$ after the departure of the first cell in the buffer (if it exists). Similarly, $B(t_i) \oplus A_{i+1}$ denotes the state reached from $B(t_i)$ due to the arrivals and the insertion in the queue of fresh customers. Remember that these transitions may be quite complex when the buffer is full because of the Push-Out mechanism.

As A_{i+1} and $A_{i+1}(high)$ are i.i.d. random sequences, the three processes $N_{HOL1}(t_i)$, $N_{HOL2}(t_i)$ and $N_{FIFO}(t_i)$ are obviously discrete-time Markov chains. The size of the Markov chain for FIFO policy is equal to $(2^{B+1} - 1)$ while it is equal to $(B+1) * (B+2)/2$ for Head-Of-Line policies. The size of the Markov chain increases exponentially with the buffer capacity, for the FIFO service discipline. So we easily exceed the size of Markov chains which may be numerically solved.

A naive bound on the loss probabilities may be easily computed. Indeed, the total number of cells is modelled by a Markov chain T with $B+1$ states. The evolution equation of T is

$$T(t_{i+1}) = \min\{B, (T(t_i) - 1)^+ + A_{i+1}\}. \quad (4)$$

Obviously, the total loss probability is an upper bound for the loss probability for high priority (or low priority) cells.

3. STOCHASTIC BOUNDS FOR HIGH PRIORITY CELLS

In this section, we stochastically compare the loss rates of high priority cells for these three service disciplines. We define the loss rates of cells as the expected number of lost cells per slot. First, we prove that the HOL1 policy provides a stochastic upper bound while the HOL2 policy provides a stochastic lower bound for the FIFO scheduling. Assume that we have obtained the steady-state probability distributions for these three models. We denote by $\Pi(n_{model})$ the probability that the state vector of the considered model has the value n_{model} at the steady-state. In the sequel, N_{model} denotes the state vector of the considered model, while n_{model} denotes a particular realization of this vector ($N_{model} = n_{model}$).

First, let us define the rewards functions which may be used to compute the loss rates of high priority for these three service discipline. Assume that there are m high priority cells at the buffer and during a slot, k high priority cells arrive to the buffer; the probability of this event is $Prob\{A(high) = k\}$ and will be denoted as p_k^{high} . The first event which will

occur in the system is the service completion of the cell, so the number of high priority cells becomes n (n may be m or $m - 1$ according to the state and the service discipline). Then, the admission mechanism takes place and we lose $(n + k - B)^+$ high priority cells. So, the loss rates for the three disciplines are :

$$R_{FIFO}^1 = \sum_{n_{FIFO}} \sum_{k=0}^M \Pi(n_{FIFO}) p_k^{high} (n_2 + k - B - \mathbb{1}_{\{\text{head}=\text{high}\}})^+$$

$$R_{HOL1}^1 = \sum_{n_{HOL1}} \sum_{k=0}^M \Pi(n_{HOL1}) p_k^{high} (n_2 + k - B - \mathbb{1}_{\{n_1=n_2\}} \mathbb{1}_{\{n_1>0\}})^+$$

$$R_{HOL2}^1 = \sum_{n_{HOL2}} \sum_{k=0}^M \Pi(n_{HOL2}) p_k^{high} (n_2 + k - B - \mathbb{1}_{\{n_2>0\}})^+$$

3.1. Stochastic upper bound for FIFO policy

We apply the coupling method in order to compare these three service policies stochastically. Since the state spaces for HIOI policies, and FiFO policy are not the same, we define two mappings φ and ψ to project them into the same state space. This method is usually known as stochastic comparison by images or by state functions [11]. We denote by \mathcal{S} the lexicographic ordering by considering first the second component and then the first one.

Definition 1 Let us define the following many to one mapping φ which maps a random vector

$(N1(t_i), N2(t_i), B(t_i))$ into a non negative random integer z which is the index of the realization of the first two components of the vector i.e., $(N2(t_i), N1(t_i))$ according to the order \mathcal{S} .

Definition 2 Similarly, let us define the following one to one mapping ψ which maps a random vector $(N2(t_i), N1(t_i))$ into a non negative integer z which is the index of the realization according to the order \mathcal{S} .

Proposition 1

$$\mathbb{1}_{\{N2_{HOL2}(t_i)>0\}} \geq \mathbb{1}_{\{\text{head}(t_i)=\text{high}\}} \geq \mathbb{1}_{\{N1_{HOL1}(t_i)=N2_{HOL1}(t_i)\}} \mathbb{1}_{\{N1_{HOL1}(t_i)>0\}} \quad (5)$$

Proof : If the first random variable is equal to 0, then there is no high priority cell in the buffer. So, the first cell in the buffer cannot be a high priority cell and the second random variable is equal to 0. But if the first cell is not a high priority one, then either the buffer is empty, or there exists only low priority cells in the buffer. Thus, the third random variable is equal to 0.

Assume now that the third random variable is equal to 1. So, there is no low priority cell in the buffer (the first two components are equal to each other), but there are some cells (the first component is positive). Thus, the first cell in the buffer is a high priority one and the second random variable is equal to 1. And the third random variable is equal to 1 as there exists high priority cells in the buffer. \square

Let us define now $P_{FIFO}(t)$ as $\varphi(N_{FIFO}(t))$. Similarly $P_{HOL1}(t) = \psi(N_{HOL1}(t))$ and $P_{HOL2}(t) = \psi(N_{HOL2}(t))$

Theorem 1 *If $P_{HOL2}(0) \leq_{st} P_{FIFO}(0) \leq_{st} P_{HOL1}(0)$, then*

$$P_{HOL2}(t_i) \leq_{st} P_{FIFO}(t_i) \leq_{st} P_{HOL1}(t_i), \quad \forall t_i > 0$$

Proof : First, couple the initial values of $P(0)$.

$$P_{HOL2}(0) \leq P_{FIFO}(0) \leq P_{HOL1}(0) \tag{6}$$

Then, the proof is by induction on t_i to establish the deterministic relation

$$P_{HOL2}(t_i) \leq P_{FIFO}(t_i) \leq P_{HOL1}(t_i), \quad \forall t_i > 0 \tag{7}$$

Assume that the inequality holds for t_i . By construction, we have :

$$P_x(t_i) < P_y(t_i) \iff \begin{cases} N2_x < N2_y \\ \text{or} \\ N2_x = N2_y \text{ and } N1_x < N1_y \end{cases} \tag{8}$$

$$P_x(t_i) = P_y(t_i) \iff N1_x = N1_y \text{ and } N2_x = N2_y$$

where x and y represent the considered service disciplines.

It follows from the evolution equations of the three systems (see equation 1, 2, 3), that the first components of the state vectors have the same evolution in all systems. However, the binary valued random variables ($\mathbb{1}_{\{\text{condition}\}}$) are different in the second components of the state vectors according to the considered system. We have the inequality 5 between these random variables from the Proposition 1, and they are subtracted in the evolution equations. Because of the vector ordering (equation 8), the orders between the first components and between the second components at time t_i are conserved at time t_{i+1} . Therefore, the inequality 7 holds at time t_{i+1} and by induction it holds for all $t_i > 0$. it follows from the coupling theorem that

$$P_{HOL2}(t_i) \leq_{st} P_{FIFO}(t_i) \leq_{st} P_{HOL1}(t_i), \quad \forall t_i > 0 \quad \square$$

Proposition 2 *The following relations hold for the reward functions :*

$$R_{HOL2}^1 \leq R_{FIFO}^1 \leq R_{HOL1}^1$$

Proof : Consider the FIFO reward function.

$$R_{FIFO}^1 = \sum_{n_{FIFO}} \sum_{k=0}^M \Pi(n_{FIFO}) p_k^{high} (n_2 + k - B - \mathbb{1}_{\{\text{head}=\text{high}\}})^+$$

Because of the relation 5, we have

$$R_{FIFO}^1 \geq W = \sum_{n_{FIFO}} \sum_{k=0}^M \Pi(n_{FIFO}) p_k^{high} (n_2 + k - B - \mathbb{1}_{\{n_2>0\}})^+$$

W is defined as an expectation on the stationary distribution of N_{FIFO} . To compare the expectations, we must apply the mapping φ .

$$W = \sum_z \sum_{n_{FIFO} | \varphi(n_{FIFO})=z} \Pi(n_{FIFO}) \sum_{k=0}^M p_k^{high} (n_2 + k - B - \mathbb{1}_{\{n_2 > 0\}})^+$$

The function $\sum_{k=0}^M p_k^{high} (n_2 + k - B - \mathbb{1}_{\{n_2 > 0\}})^+$ has the same value for all the states of N_{FIFO} which are mapped on the same value z ; therefore,

$$W = \sum_z \sum_{k=0}^M p_k^{high} (n_2 + k - B - \mathbb{1}_{\{n_2 > 0\}})^+ \sum_{n_{FIFO} | \varphi(n_{FIFO})=z} \Pi(n_{FIFO})$$

As the size of the batch is smaller than B , (i.e., $k \leq M < B$) the function $\sum_{k=0}^M p_k^{high} (N_2 + k - B - \mathbb{1}_{\{N_2 > 0\}})^+$ is a positive increasing function. It follows from the fundamental property of the strong stochastic ordering (see definition 4 of the appendix) and theorem 1 that

$$W \geq V = \sum_z \sum_{k=0}^M p_k^{high} (n_2 + k - B - \mathbb{1}_{\{n_2 > 0\}})^+ \sum_{n_{HOL2} | \psi(n_{HOL2})=z} \Pi(n_{HOL2})$$

As ψ is a one to one mapping,

$$V = \sum_{n_{HOL2}} \Pi(n_{HOL2}) \sum_{k=0}^M p_k^{high} (n_2 + k - B - \mathbb{1}_{\{n_2 > 0\}})^+ = R_{HOL2}^1$$

Thus, the first inequality is proved. The proof of the second inequality is omitted as it is exactly the same proof which begins by the second part of the relation 5.

$$R_{FIFO}^1 \leq \sum_{n_{FIFO}} \sum_{k=0}^M \Pi(n_{FIFO}) p_k^{high} (n_2 + k - B - \mathbb{1}_{\{n_1 = n_2\}} \mathbb{1}_{\{n_1 > 0\}})^+ \quad \square$$

3.2. Stochastic upper bounds on HOL1 policy

In the second step, we perform an aggregation of the states to obtain a chain with $O(B \times F)$ states where F is a factor which is used during the aggregation process. F is denoted as the aggregation factor. We define a macro-state (n_1, n_3) , which gathers the states having the same value of N_1 and for which the value of N_2 , i.e., the number of high priority cells is less than $n_1 - F$. All the other states are kept unchanged in the aggregated chain (see figure 2). Note that the states of the initial chain where cells can be lost are not aggregated.

More precisely, let (n_1, n_3) be a state of the aggregated chain :

- if $n_3 = n_1 - F$ then the state is a macro-state which contains all the states (n_1, n_2) such that $n_2 \leq n_1 - F$. So, N_3 is an upper bound of the number of high priority cells in the buffer.
- if $n_3 > n_1 - F$ then the state contains only one state (n_1, n_3) where $n_3 = n_2$. So, n_3 represents exactly the number of high priority cells in the buffer.

Therefore $n3 \in \{n1 - F, \dots, n1\}$ These states and these transitions are defined to provide a stochastic upper bound of the Head-Of-Line service discipline (HOL1). Intuitively, they are defined by considering the worst case from the point of view of the loss rate for high priority cells.

The new system is denoted as $HOL1_{agg}$, and its evolution equations are :

$$\begin{cases} N1(t_{i+1}) = \min\{B, (N1(t_i) - 1)^+ + A_{i+1}\} \\ N3(t_{i+1}) = \max\{N1(t_{i+1}) - F, N4(t_{i+1})\} \\ \text{where } N4(t_{i+1}) = \min\{B, N3(t_i) - \mathbb{1}_{\{(N1(t_i)=N3(t_i)) \mathbb{1}_{\{N1(t_i)>0\}}\}} + A_{i+1}(high)\} \end{cases}$$

$N4(t_{i+1})$ represents the number of high priority cells in the buffer after the service and the arrival process. And the max operation in the second evolution equation models the aggregation process into the macro-state.

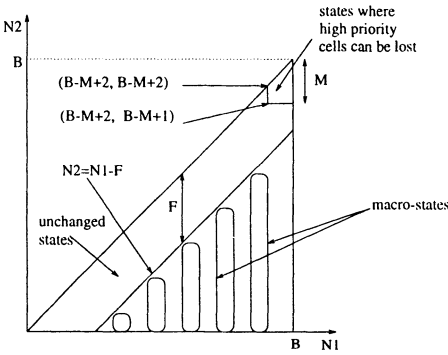


Figure 2. The aggregated chain

Clearly, if the value of F is large, then the bound is tight (if F is equal to B , then the Markov chain is not aggregated). And if F is too small the bound is quite similar to the naive bound presented in section III. So a good value of F is a trade-off between computational complexity and the accuracy of the results. The size of the aggregated Markov chain is $\sum_{i=0}^F (B + 1 - i) = O(B \times F)$. In the following, we assume that F is greater than the maximum batch size M . Therefore the states where cells are lost are represented explicitly (they are original states, not the aggregated ones).

To compare these policies, we first map the states from the HOL1 model into the states of the aggregated chain. More precisely, for the HOL1 model, for a given $n1$, all the states such that $N2$ is in the set $\{0, 1, 2, \dots, (n1 - F)^+\}$ will be mapped to the state denoted as $(n1, (n1 - F)^+)$.

Definition 3 Let us define two mappings ϕ and ψ which maps the states of the two Markov chains into the same subset of the integers such that

1. ψ is a one to one mapping which maps a random vector $N_{HOL1_{agg}}$ into a non negative integer z which is the index of the realization according to the order \mathcal{S} .
2. ϕ is a many to one mapping such that for all states N of N_{HOL1} , we have $\phi(N) = \psi(NA)$ where NA is the macro-state in the aggregated chain which contains the state N .
3. Then, we define $Q_{HOL1}(t) = \phi(N_{HOL1}(t))$ and $Q_{HOL1_{agg}}(t) = \psi(N_{HOL1_{agg}}(t))$

Theorem 2 If $Q_{HOL1}(0) \leq_{st} Q_{HOL1_{agg}}(0)$, then

$$Q_{HOL1}(t_i) \leq_{st} Q_{HOL1_{agg}}(t_i), \quad \forall t_i > 0$$

Proof : The proof is by induction on the cell departure epochs and is quite similar to the proof of Theorem 1. First we couple the initial values of $Q(0)$. Then we establish the deterministic relation

$$Q_{HOL1}(t_i) \leq Q_{HOL1_{agg}}(t_i), \quad \forall t_i > 0 \quad (9)$$

Assume that the inequality holds at t_i . The first component which is the total number of cells in the buffer evolves in the same manner in both systems for all t_i . It follows from the evolution equations, the second component is greater in the aggregated system, so the orders on $N1(t_i)$ $N2(t_i)$ at time t_i are preserved at time t_{i+1} . As a result of the construction of the mappings ψ, ϕ , and the order \mathcal{S} , $Q_{HOL1}(t_{i+1}) \leq Q_{HOL1_{agg}}(t_{i+1})$, so by induction the inequality 9 holds. It follows from the coupling theorem that

$$Q_{HOL1}(t_i) \leq_{st} Q_{HOL1_{agg}}(t_i), \quad \forall t_i > 0 \quad \square$$

The reward function for the HOL1 discipline is already defined. The reward function for the aggregated chain ($R_{HOL1_{agg}}$) is the expectation of the same function on a reduced state space.

$$R_{HOL1_{agg}}^1 = \sum_{n_{HOL1_{agg}}} \sum_{k=0}^M \Pi(n_{HOL1_{agg}}) p_k^{high} (n_2 + k - B - \mathbb{1}_{\{n_1=n_2\}} \mathbb{1}_{\{n_1>0\}})^+$$

Proposition 3 The following relation holds for the reward functions.

$$R_{HOL1}^1 \leq R_{HOL1_{agg}}^1$$

The proof is similar to the proof of Proposition 2 and is omitted.

3.3. Stochastic lower bound for FIFO policy

It follows from the proposition 2 that for high priority cells, the cell loss rates with HOL2 policy gives the lower bound on the cell loss rates with FIFO policy. Since the reward function defining the loss rates with HOL2 policy depends upon only the number of high priority cells, we can easily compute this reward function by using a Markov chain of $B + 1$ states. In this representation, each state $Y(t_i)$ represents the number of high priority cells in the buffer at time t_i and the evolution of the system is following :

$$Y(t_{i+1}) = \min\{B, (Y(t_i) - 1)^+ + A_{i+1}(high)\}$$

Note that there is no aggregation of states, so the computed reward function gives the exact results of the cell loss rates for high priority cells with HOL2 policy.

4. STOCHASTIC BOUNDS FOR LOW PRIORITY CELLS

In this section, we consider the loss rates of low priority cells, and prove that the HOL2 policy provides a stochastic upper bound and the HOL1 policy provides a stochastic lower bound for FIFO scheduling. First, we define the reward functions in order to compute the loss rates of low priority cells with these service policies.

We consider the pseudo-state of the buffer after the completion of service and after the arrivals of both types of cells, just before the admission mechanism takes place. Let this pseudo state be (m, q) where m is the total number of cells and q is the number of low priority cells. As we consider pseudo-states before the admission control, we may have $m > B$ and $q > B$. Clearly, the number of low priority cells rejected by the Push-Out mechanism is $\min((m - B)^+, q)$.

The values of m and q are obtained easily from the former state description, the service discipline and the arrival process. The probability of the arrivals of k high priority cells and l low priority cells during one slot is $\text{Prob}\{A = k + l \text{ and } A(\text{high}) = k\}$. The probability of this event will be denoted as $p_{k,l}$. Thus, the reward function for FIFO service discipline is :

$$R_{FIFO}^2 = \sum_{n_{FIFO}} \sum_{k=0}^M \sum_{l=0}^{M-k} \Pi(n_{FIFO}) p_{k,l} \min\{(m - B)^+, q\},$$

where $m = (n1 - 1)^+ + k + l$ and $q = n1 - n2 - \mathbb{1}_{\{\text{head}=\text{low}\}} + l$.

In order to compute the bounds for the loss rates of low priority cells, we use the bounds derived for the high priority cells. We denote by R^{1+2} the reward function defining the total cell loss rates which can be computed from the naive bound (see equation 4):

$$R^{1+2} = \sum_{n=0}^B \sum_{k=0}^M \Pi(n) ((n - 1)^+ + k - B)^+ p_k$$

where $p_k = \text{Prob}\{A = k\}$, denotes the probability of k cells arrivals.

If we subtract from the total loss rates the lower bound on the loss rates for high priority cells, we derive the upper bound on loss rates for low priority cells. In the same way, when we subtract the upper bound on the loss rates for high priority cells, then we obtain the lower bound on the loss rates for low priority cells. Therefore, we have these inequalities :

$$R^{1+2} - R_{HOL1}^1 \leq R_{FIFO}^2 \leq R^{1+2} - R_{HOL2}^1$$

5. RESULTS

In this section we give some typical results for the loss rates of high priority and low priority cells. The steady state distribution are computed using the GTH algorithm because of its accuracy [12] . As the chain are quite small, the results needs only few seconds of computation, on a SPARC station SUN.

We assume that the maximum batch size is $M = 3$ or $M = 4$. We consider three models for the arrival process. The first two models allow us to shorten the number of parameters. The third model is the general batch model (with maximum size 3) and we

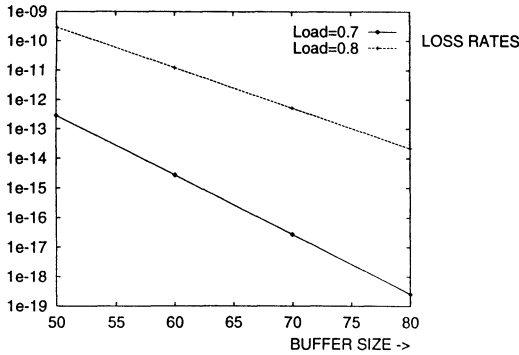


Figure 3. High priority cells : loss rates versus buffer size

have to give the probability of the arrival of $i - j$ high priority cells and j low priority cells. This event is denoted as $e_{i,j}$.

In the first model, we assume that the batch process is represented by $c1$ the ratio of high priority cells and by the vector $(p_i)_{i=0,4}$ where p_i is the probability that the batch size is i . Each cell of a batch may be a high priority cell following an i.i.d. Bernoulli process with probability $c1$.

In the second model, we assume that only few batches are feasible. We assume that the batches of size 1 or 2 contain only high priority cells whereas the batches of size 3 or 4 contain only low priority cells. We also denote, in this model, $(p_i)_{i=0,4}$ as the distribution of the batch size.

In the first figure, we present loss rates of the high priority cells versus buffer size. Two experiments are considered using the third model of arrival.

- the load is 0.7; and the probability of the events are : $p(e_{0,0}) = 0.65$, $p(e_{1,0}) = 0.05$, $p(e_{1,1}) = 0.05$, $p(e_{2,0}) = 0.075$, $p(e_{2,1}) = 0.075$, $p(e_{2,2}) = 0$, $p(e_{3,0}) = 0.05$, $p(e_{3,1}) = 0.05$, $p(e_{3,2}) = 0$
- the load is 0.8; and the probability of the events are : $p(e_{0,0}) = 0.6$, $p(e_{1,0}) = 0.05$, $p(e_{1,1}) = 0.05$, $p(e_{2,0}) = 0.1$, $p(e_{2,1}) = 0.1$, $p(e_{2,2}) = 0$, $p(e_{3,0}) = 0.05$, $p(e_{3,1}) = 0.05$, $p(e_{3,2}) = 0$

In figure 4, we present the loss rates for high priority cells in a buffer of size 80. The load increase from 0.75 to 0.9 but the load due to high priority cells is kept constant to 0.6. This is obtained using the second arrival model with the following parameters : $p_1 = 0.4$, $p_2 = 0.1$, $p_3 = 0.02 * n$, $p_4 = 0.01 * n$ where n is a parameter varying from 1 to 3.

Figure 5 shows the importance of the aggregation factor F . The same models were analyzed with two values of F ($F = 3$ and $F = 10$). Clearly, there are between 1 and

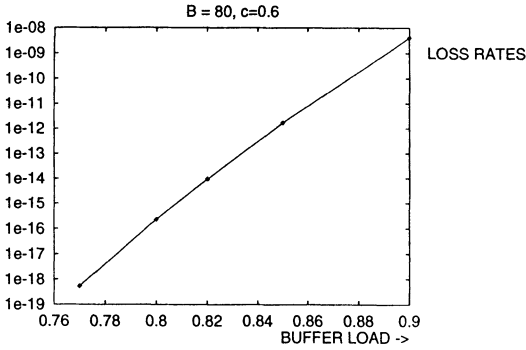


Figure 4. High priority cells : loss rates versus buffer load

3 orders of magnitude between the two bounds. The buffer size is 60. We consider the second model of arrival with a maximum batch size equal to 3. The ratio of high priority cells is 0.1. And the points are obtained for the following values of the arrival process : (Point 1 : $p_1 = 0.2, p_2 = 0.1, p_3 = 0.1$) (Point 2 : $p_1 = 0.3, p_2 = 0.1, p_3 = 0.1$) (Point 3 : $p_1 = 0.2, p_2 = 0.2, p_3 = 0.1$)

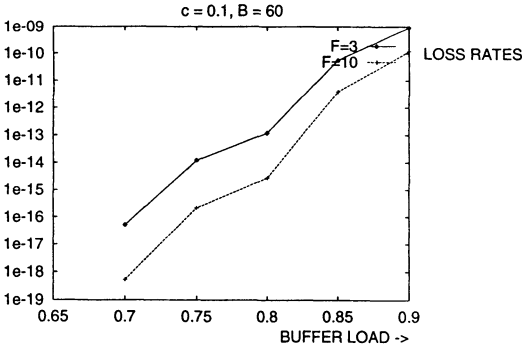


Figure 5. High priority cells : loss rates versus buffer load for two aggregation factor F

This is also illustrated in Figure 6. Furthermore, we have plotted in this figure the naive bound presented in section III. Clearly, our approach gives in these experiments much better results than the naive bound. The buffer size is 60. The maximum size of

the batch is 3. We use the second model of arrivals. The loads were obtained using the following probabilities of arrival : (load 0.7 : $p_1 = 0.2, p_2 = 0.1, p_3 = 0.1$), (load 0.75 : $p_1 = 0.25, p_2 = 0.1, p_3 = 0.1$), (load 0.8 : $p_1 = 0.25, p_2 = 0.125, p_3 = 0.1$), (load 0.85 : $p_1 = 0.3, p_2 = 0.1, p_3 = 0.1$), (load 0.9 : $p_1 = 0.4, p_2 = 0.1, p_3 = 0.1$).

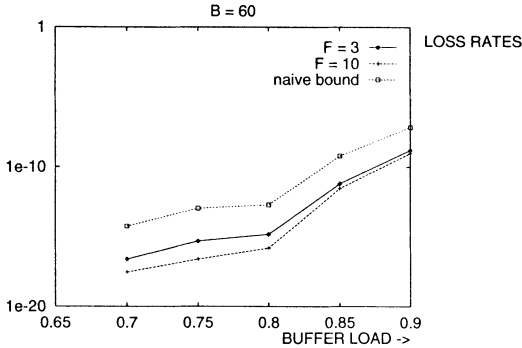


Figure 6. High priority cells : loss rates versus load

Figure 7 shows the difference between the lower and the upper bound : HOL1 policy provides a good upper bound, but HOL2 policy does not provide a good lower bound. The loads were obtained using the following probabilities of arrival : (load 0.8 : $p_1 = 0.1, p_2 = 0.2, p_3 = 0.1$), (load 0.85 : $p_1 = 0.15, p_2 = 0.2, p_3 = 0.1$), (load 0.9 : $p_1 = 0.2, p_2 = 0.2, p_3 = 0.1$), (load 0.95 : $p_1 = 0.25, p_2 = 0.2, p_3 = 0.1$), and with a ratio of high priority cells which is 0.8.

However, for the low priority cells, the naive bound provides good results. For instance, in figure 8, we present the loss rates for low priority cells versus the load. The loads were obtained using arrival probabilities: $p_0 = 0.5, p_1 = 0.2, p_2 = 0.2, p_3 = 0.1$ where p_i denotes the probability of i cell arrivals. The ratio of high priority cells is 0.8. In this case, our bound is so close, in the logarithmic scale, to the naive bound that only one curve is depicted. This is quite natural because, if the Push-Out mechanism is efficient, then almost surely the lost cells are low priority cells. Therefore, the naive bound is very close to the exact result and our bound lies somewhere in between.

6. CONCLUSIONS

In this paper, we evaluate cell loss rates in an ATM switch with Push-Out mechanism and FIFO service discipline. The considered system is modelled by a discrete-time Markov chain. As the state space of the underlying chain increases exponentially with the buffer size, the analysis with standard numerical methods seems intractable.

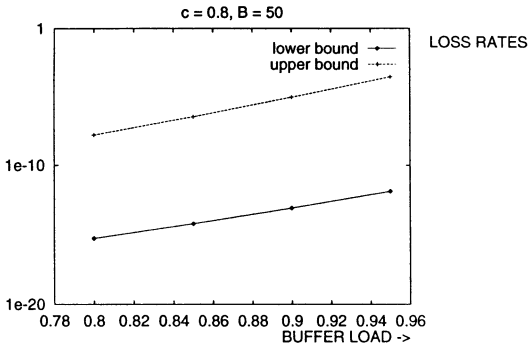


Figure 7. Upper and lower bounds for high priority cells versus buffer load

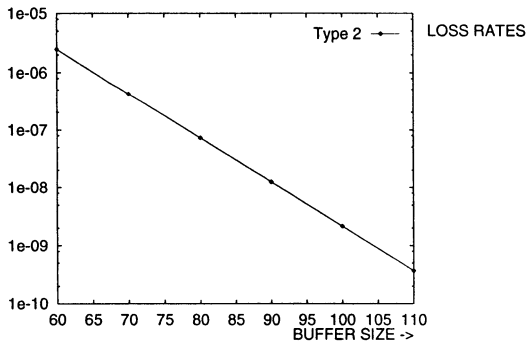


Figure 8. Low priority cells : loss rates versus buffer load

We propose a methodology based on stochastic ordering to obtain a bounding chain of smaller size. It means that the performance indices defined by the reward functions are bounded by the reward functions computed from the bounding chain. The advantage of the method of stochastic bounds is that we keep important information about the evaluated system.

On the other hand, it is not always easy to prove that the modified system provides stochastic bounds on the analyzed system and generally the bounds are not tight. In this paper, we use the coupling method to derive stochastic bounds. First, we compare stochastically different service disciplines for the Push-Out mechanism. In the second step, we perform an aggregation of states by conditioning on the events which occur in the system.

We analyzed the Push-Out mechanism with Bernoulli Batch arrivals using the strong ordering. In the case where the arrivals are assumed to be modulated by a Markov chain, the strong ordering cannot be applied since the evolution of the system in different phases is not the same. So weaker orders must be defined.

7. APPENDIX

In this appendix, we give the essential definitions for strong stochastic orderings and the coupling method (see [8] for further information).

Definition 4 Let X and Y be two random variables on the same space ϵ , we say that X is stochastically lower in the sense of the strong ordering than Y , denoted by $X \leq_{st} Y$

$$X \leq_{st} Y \quad \text{if and only if} \quad E(f(X)) \leq E(f(Y))$$

for all increasing functions f on ϵ , if the expectations exist.

This definition is equivalent to the next one in terms of the distribution functions (F_X, F_Y):

Definition 5 $X \leq_{st} Y$ if and only if $F_X(a) \geq F_Y(a), \quad \forall t \in \epsilon$

The strong stochastic order is a sample path ordering, so from the Strassen's theorem [13] it follows that :

Theorem 3 (coupling theorem) *The following conditions are equivalent :*

- i . $X \leq_{st} Y$;
- ii . there exists a probability measure λ on $R \times R$ with support in $K = \{(x, y) \in R \times R, \text{ such that } x \leq y\}$, with first marginal F_X and with second marginal F_Y ;
- iii . there exist random variables X^* and Y^* , having the same distribution of X and Y and such that $X^* \leq Y^*$ almost surely.

Definition 6 Let X_i and Y_i be discrete-time Markov processes (Markov chains). $X_i \leq_{st} Y_i$ if and only if the initial stochastic order $X_0 \leq_{st} Y_0$ is preserved for all i :

$$X_i \leq_{st} Y_i, \quad \forall i > 0$$

The coupling theorems are generalized in order to compare the probability measures by their images [11]. We state here only the theorem for comparison of Markov chains.

Theorem 4 (coupling by images) *Let X_i (resp. Y_i) be Markov chain on E (resp. F) with probability transition matrices $P1$ (resp. $P2$); G be a totally ordered space; and φ (resp. ψ) be a one to many mapping from E (resp. F) to G . The image of X_i on G , $\varphi(X_i)$ is stochastically lower in the sense of the strong ordering than the image of Y_i on G , $\psi(Y_i)$ if and only if*

$$\varphi(X_0) \leq_{st} \psi(Y_0) \implies \varphi(X_i) \leq_{st} \psi(Y_i), \quad \forall i > 0$$

The following conditions are equivalent :

i . $\varphi(X_i) \leq_{st} \psi(Y_i), \forall i \geq 0;$

ii . *there exists a Markov chain $\{X_i^*, Y_i^*\}$ having the probability transition matrix P with first marginal $P1$, with second marginal $P2$ and with support in $K = \{(x, y) \in E \times F, \text{ such that } \varphi(x) \leq \psi(y)\}$.*

REFERENCES

1. B.T. Doshi and H. Heffes, Overload Performance of Several Processor Queueing Disciplines For the M/M/1 Queue, IEEE Trans. on Communications, Vol. COM-34, No. 6 (1986) 538 546.
2. G. Hébuterne and A. Gravey, A space priority queueing mechanism for multiplexing ATM channels, ITC Specialist Seminar, Computer Network and ISDN Systems, Vol. 20 (Dec. 90) 37 43.
3. H. Kroner, Comparative Performance Study of Space Priority Mechanisms for ATM Networks, IEEE INFOCOM'90, San Francisco, (June 1990) 1136 1143.
4. X. Cheng and I. F. Akyildiz, A Finite Buffer Two Class Queue with Different Scheduling and Push-Out Schemes, IEEE INFOCOM'92, Florence, (1992) 231 241.
5. O. Hashida and Y. Takahaski, A Discrete-time Priority Queue with Switched Batch Bernoulli Process Inputs and Constant Service Time, Proc. of 13th International Teletraffic Congress, (Copenhagen 1991) 521 526.
6. Z. Zhang, Finite Buffer Discrete-Time Queues with Multiple Markovian Arrivals and Services in ATM Networks, IEEE INFOCOM, Florence, (1992) 2026 2035.
7. W.J. Stewart, On the Use of Numerical Methods for ATM Models, Proceedings of the First International Workshop of the Modelling and Performance Evaluation of ATM technology, La Martinique, North Holland, 1993.
8. D. Stoyan, Comparison Methods for Queues and Other Stochastic Models, Wiley, New York, 1983.
9. J.M. Fourneau, N. Pekergin and H. Taleb, Weak Ordering for Markovian Arrival Processes, 1995, submitted for publication.
10. A. Gravey and G. Hébuterne, Mixing Time and Loss Priorities in a Single Server Queue, in Workshop on ATM, 13th International Teletraffic Congress, Copenhagen, 1991.
11. M. Doisy, Comparison de Processus Markoviens, Ph-D thesis, Univ. de Pau et des Pays de l'Adour, 1992.

12. D.P. Heymann, Further comparisons of direct methods for computing stationary distributions of Markov chains, *Siam J. Alg. Disc. Meth.*, Vol. 8, N 2, (April 1987) 226-232.
13. T. Lindvall, *Coupling Method*, Wiley New York 1992.