Efficient Estimation of Cell Loss Probabilities for ATM Switches with Input Queueing via Light Traffic Derivatives

Young-Beom Kim and Jung Hur

Abstract

Under most system assumptions, closed form solutions of performance measures for ATM switches with input queueing are not available. In this paper, we present expressions and bounds for the derivatives of cell loss probabilities with respect to the arrival rate evaluated at a zero arrival rate. These bounds are used to give an approximation by Taylor expansion, thereby providing an economical way to estimate cell loss probabilities in light traffic.

I. Introduction

In implementing high--speed networks such as ATM networks, the bottleneck lies mainly in switching as a communication network can be viewed as consisting largely of transmission links and switching nodes. Several candidate architectures could possibly support high--speed packet switching. Noteworthy among them are the various space-division switching fabrics developed in the past decade for terrestrial ATM networks; in particular, we focus here on non-blocking ATM switches with input buffering. Having in mind satellite applications, we assume that the input buffers are finite, and typically small. In that context, the key performance measure we wish to evaluate is the cell loss probability (CLP).

To carry out this evaluation, we consider a simple discrete-time model for a synchronous $K \times K$ non-blocking ATM switch where the input queues are of finite size b. Cells arrive at each input port according to a Bernoulli process of rate λ; cells that find a full queue are rejected. Output contention manifests itself through head-of-the-line (HOL) blocking[1], and is resolved by a simple randomized arbitration mechanism. Despite their simplicity, these rules of operation produce a very complex queueing behavior as input queues become correlated over time. This explains why the performance analysis is possible only under special conditions like infinite switch size and saturation assumptions[1]. Therefore, under most model assumptions, closed form solutions of performance measures of interest are not available, nor can they be expected. Worse perhaps, when evaluating CLP, Monte-Carlo simulation techniques turn out to be of limited use owing to their vast computational cost as the desired CLP in ATM networks, being usually in the range of 10^{-6} to 10^{-12}, corresponds to rare events.

Manuscript received September 5, 1997; accepted November 15, 1997.
The Authors are with Department of Electronic Engineering, Kon-Kuk University, Seoul, Korea.

In this paper, we address the problem of evaluating in light traffic the cell loss probability(CLP) associated with $K \times K$ ATM switches with input queueing (in short, IQ switches). Light traffic here refers to the situation where the system is lightly loaded, i.e., the arrival rate λ to the system is very small. Throughout we use the notation that for any function $f: R \to R$, $f(0+) \equiv \lim_{x \to 0+} f(x)$ and $\frac{d^k}{dx^k} f(0+) \equiv \lim_{x \to 0+} \frac{d^k}{dx^k} f(x)$, $k=1,2,\ldots$, whenever the derivatives exist.

Let $P_b(\lambda)$ denote the CLP associated with a $K \times K$ IQ switch when the input queues are of finite size b and the arrival rate at each port is λ. We show how to evaluate the light traffic derivatives $\frac{d^k}{d\lambda^k} P_b(0+) \equiv \lim_{\lambda \to 0+} \frac{d^k}{d\lambda^k} P_b(\lambda)$ for various values of k. In particular, we show that $\frac{d^k}{d\lambda^k} P_b(0+)=0$ for $k=0,1\ldots,2b-1$, and spend most of our efforts on evaluating the first two non-zero derivatives. These formulae make use of the regenerative structure of the underlying Markov chain associated with the queueing model.

We then propose to approximate the CLP $P_b(\lambda)$, at least for small values of the arrival rate λ, by a Taylor series expansion of $P_b(\lambda)$ near the origin, which here takes the form

$$P_b(\lambda) \simeq \frac{1}{(2b)!} \lambda^{2b} \frac{d^{2b}}{d\lambda^{2b}} P_b(0+) + \frac{1}{(2b+1)!} \lambda^{2b+1} \frac{d^{2b+1}}{d\lambda^{2b+1}} P_b(0+), \quad \lambda \simeq 0. \quad \textbf{(1)}$$

This approximation works well for small values of λ where the CLP is expected to be very small, a situation often handled by variance reduction techniques such as importance sampling. The method proposed here thus provides a numerical alternative to these techniques in light traffic.

This paper is organized as follows: In next section, we provide the formulation of the model under consideration. Section 3 is devoted to establishing a suitable representation for the CLP using the notation given in Section 2; specifically we represent the CLP by numerator and denominator functions (in terms of the input rate λ), which are defined on a regeneration cycle of the system process. In Section 4, we provide some useful definitions and formulate the numerator and denominator functions in terms of *sample paths* of underlying system rvs. Section 5 is devoted to investigating several properties associated with numerator and denominator functions as well as to establishing a relationship between the light traffic derivatives of the numerator function and those of CLP. In Sections 6-7, we seek general solutions for the derivatives of CLP of order $2b$ and $2b+1$ evaluated near zero. We close this paper by experimenting our result with a 20×20 IQ switch.

II. The Model

The switching fabric of interest here is characterized by the number K of input ports, the number K of output ports, and the common buffer size b; the positive integers K and b are held fixed throughout the discussion.

The switch operates in a synchronous mode according to the following scenario: Time is divided into consecutive slots of equal duration; the length of a slot coincides with the transport time of a cell across the switching fabric. Each input port is equipped with a buffer which can contain at most b cells. At the beginning of each time slot, the switch controller mediates potential output contentions by randomly selecting one HOL cell amongst the HOL cells which have the same output address. The HOL cells selected for transmission are then removed from their buffer and start being transmitted across the fabric; this transmission is completed by the end of the time slot. In processing newly arriving cells, there could exist two transmission strategies, namely "gated" and "cut-through" strategies. With the gated transmission strategy, at the same time that transmission starts, new cells which arrive into the system during a time slot are enqueued by the end of the slot (if buffer space is available). With the cut-through strategy, if an input queue is empty at the beginning of a time slot, cells arriving at that queue during the time slot are eligible for possible transmission during the time slot. These steps are repeated from slot to slot. In this paper, we only consider the gated strategy- the results for the cut-through strategy can be similarly obtained.

In order to provide a precise model description, we begin with $3K$ mutually independent collections of rvs, namely $\{\alpha^k_{t+1},\ t=0,1,...\}$, $\{\beta^k_{t+1},\ t=0,1,...\}$, and $\{\mu^k_{t+1},\ t=0,1,...\}$, $k=1,...,K$. The following assumptions **(A1)-(A3)** are enforced during the discussion:

(A1) For each $k=1,...,K$, the rvs $\{\alpha^k_{t+1},\ t=0,1,...\}$ are i.i.d. rvs with

$P[\alpha^k_{t+1}=1]=1-P[\alpha^k_{t+1}=0]=\lambda,\ t=0,1,...;$

(A2) For each $k=1,...,K$, the rvs $\{\beta^k_{t+1},\ t=0,1,...\}$ are i.i.d. rvs which are uniformly distributed over the set $\{1,...,K\}$, i.e.,

$P[\beta^k_{t+1}=\ell]=\frac{1}{K},\ \ell=1,...,K,\ t=0,1,...;$

(A3) For each $k=1,...,K$, the rvs $\{\mu^k_{t+1},\ t=0,1,...\}$ are i.i.d. rvs which are uniformly distributed over the unit interval (0, 1).

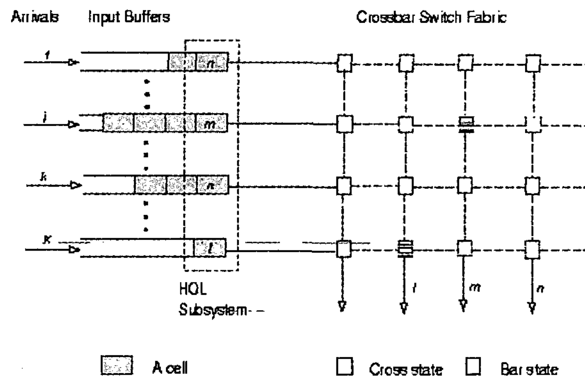


Fig. 1. Operational Diagram for ATM switches with Input Queueing.

These quantities will shortly be given an interpretation in the context of the input queueing system described earlier.

Fix $k=1,...,K$ and $t=0,1,....$ Let Q^k_t and V^k_t respectively denote the number of cells present in the k^{th} input queue and the destination of the HOL cell in that queue at the beginning of time slot $[t,t+1)$; by convention $V^k_t=0$ if $Q^k_t=0$. We write $\mathbf{Q}_t\equiv(Q^1_t,...,Q^K_t)$ and $\mathbf{V}_t\equiv(V^1_t,...,V^K_t)$. As will become apparent shortly, it is appropriate to view the pair $\mathbf{X}_t\equiv(\mathbf{V}_t, \mathbf{Q}_t)$ as the state of the system at the beginning of time slot $[t,t+1)$.

Contention resolution : For each $k=1,...,K$, we denote by D^k_{t+1} the rv indicating the departure of HOL cell from the k^{th} input buffer during the time slot $[t,t+1)$. In other words, the rv D^k_{t+1} indicates whether the HOL cell has been selected for transmission as the result of the contention resolution at the beginning of the slot $[t,t+1)$. Let G^ℓ_{t+1} denote the set of input ports whose HOL cell has destination address ell at the beginning of the slot $[t,t+1)$. According to transmission strategy adopted, G^ℓ_{t+1} can be defined as follows:

$$G^\ell_{t+1}\equiv\begin{cases}\{k\in\{1,...,K\}: Q^k_t=0\}, & \ell=0;\\ \{k\in\{1,...,K\}: V^k_t=\ell\}, & \ell=1,...,K.\end{cases}$$

Next, we define the G^ℓ_{t+1}-valued rv O^ℓ_{t+1} by

$$O^\ell_{t+1}=\arg\max\{i\in G^\ell_{t+1}: \mu^i_{t+1}\},\ \ell=1,...,K \qquad (2)$$

with a lexicographic tie-breaker; we use the convention $O'_{t+1}=0$ if G'_{t+1} is empty. If $|G'_{t+1}|>0$, we note that

$$P[O'_{t+1}=j|F_t]=|G'_{t+1}|^{-1},\ j\in G'_{t+1} \quad (3)$$

where $F_t\equiv\sigma\{\alpha_{s+1},\beta_{s+1},\mu_{s+1},\ s=0,1,\ldots,t-1\}$. In other words, as the rv O'_{t+1} selects an index in G'_{t+1} at random, it can be used to determine the input port (with index in G'_{t+1}), whose HOL cell will be transmitted to output ℓ during the time slot $[t,t+1)$. The binary rv D^k_{t+1} is then given by

$$D^k_{t+1}=\sum_{\ell=1}^{K}1[O'_{t+1}=k],\ k=1,\ldots,K. \quad (4)$$

Arrivals and blocking : According to the scenario outlined earlier, there are now $Q^k_t-D^k_{t+1}$ cells in the buffer at the k^{th} input port after contention has been resolved. New cells possibly arrive at the k^{th} input port with α^k_{t+1} cell arriving in slot $[t,t+1)$. An arriving cell is accepted into the buffer and put at the end of the line only if it finds the k^{th} input queue to be non-full, i.e., if $Q^k_t-D^k_{t+1}<b$; otherwise the cell is blocked and rejected. In short, blocking occurs at input port k during slot $[t,t+1)$ if a cell arriving during that time slot finds that

$$Q^k_t-D^k_{t+1}=b,\ \text{i.e.,}\ Q^k_t=b\ \text{and}\ D^k_{t+1}=0.$$

Therefore, at the end of the slot $[t,t+1)$, there remain $Q^k_t-D^k_{t+1}+1[Q^k_t-D^k_{t+1}<b]\alpha^k_{t+1}$ cells in the k^{th} input queue, and with $Q^k_0=0$, we have

$$Q^k_{t+1}=Q^k_t-D^k_{t+1}+\alpha^k_{t+1}1[Q^k_t-D^k_{t+1}<b],\ t=0,1,\ldots. \quad (5)$$

Addressing : The destination address of each newly arriving cell is assigned randomly and uniformly over the set $\{1,\ldots,K\}$; this assignment is performed independently over time across input ports, and independently of the generation of cell arrivals. Hence, there is no loss of generality in assuming that each cell declares its destination address immediately upon reaching the HOL and that it keeps its address until it begins transmission across the switching fabric. The destination of a cell newly reaching the HOL of the k^{th} input port is encoded in the rv β^k_{t+1}. With $V^{k_0}\equiv 0$, the following recursion

$$\begin{aligned}V^k_{t+1} &= (1-D^k_{t+1})(1[Q^k_t=0]\alpha^k_{t+1}\beta^k_{t+1}+1[Q^k_t>0]V^k_t)\\ &\quad +D^k_{t+1}(1[Q^k_t=1]\alpha^k_{t+1}+1[Q^k_t>1])\beta^k_{t+1},\ t=0,1,\ldots\end{aligned}$$

holds.

III. Cell Loss Probabilities

In this section, we establish some suitable expressions for CLP; of special interest is the formulation of CLP in terms of input rate λ as the purpose of this section is to estimate the derivatives of CLP with respect to λ in light traffic regime.

Let $P_b(\lambda)$ denote the CLP associated with a $K\times K$ IQ switch when the input queues are of finite size b and the arrival rate at each port is λ. Throughout the discussion we fix $b=1,2,\ldots$. We denote by $P^n_b(\lambda)$ the probability that the n^{th} cell arriving at the first input port will be blocked; this quantity is simply

$$P^n_b(\lambda)=P[Q^1_{t_n}=b],\quad n=1,2,\ldots$$

with t_n representing the left boundary of the time slot during which the n^{th} arriving cell falls in.

The steady-state probability that an arbitrary cell arriving at the first input port will be blocked is given by

$$\begin{aligned}P_b(\lambda) &= \lim_{n\to\infty}P[Q^1_{t_n}=b]\\ &= P[Q^1=b],\end{aligned}$$

where in the last step we have used BASTA. Standard regenerative arguments yield the steady-state measure $P_b(\lambda)$ as a ratio of two expected values, namely

$$P_b(\lambda)=\frac{E\left[\sum_{t=0}^{\tau-1}1[Q^1_t=b]\ \middle|\ Q_0=0\right]}{E[\tau|\ Q_0=0]}$$

with the first passage time τ to the empty state 0 given by

$$\tau\equiv\inf_{t>0}\{t:\ Q_t=0\}.$$

If $Q_0=0$, then τ can be interpreted as the length of a regeneration cycle. The two quantities in (3.1) are functions of input rate λ; for the sake of convenience we denote by $\Phi(\lambda)$ and $\Psi(\lambda)$, respectively, the numerator and denominator in the expression (3.1), respectively, so that (3.1) is now rewritten as

$$P_b(\lambda)=\frac{\Phi(\lambda)}{\Psi(\lambda)}. \quad (6)$$

The estimation of CLP is now equivalent to that of $\Phi(\lambda)$ and $\Psi(\lambda)$.

IV. Formulation of $\Phi(\lambda)$ and $\Psi(\lambda)$

In this section, we provide some needed definitions and formulate $\Phi(\lambda)$ and $\Psi(\lambda)$ in terms of sample paths of the underlying system rvs.

As the system is driven by mutually independent rvs $\{\alpha_{t+1},\beta_{t+1},\mu_{t+1},\ t=0,1,\ldots\}$, the rvs $\{X_t,\ t=0,1,\ldots\}$ representing the system states are uniquely determined by the rvs $\{\alpha_{t+1},\beta_{t+1},\mu_{t+1},\ t=0,1,\ldots\}$. To facilitate the presentation, we define the space M by

$$M \equiv \{0,1\}^K \times \{1,\ldots,K\}^K \times \{0,1,\ldots,K\}^K, \quad (7)$$

and throughout we use the convention

$s \equiv (s^1, s^2, s^3)$, where $s^r \equiv (s^{r,1},\ldots,s^{r,K})$, $r=1,2,3$, $s \in M$. We define M-valued rvs ξ_{t+1}, $t=0,1,\ldots$, by

$$\xi_{t+1} \equiv (\alpha_{t+1}, \beta_{t+1}, O_{t+1}), \quad t=0,1,\ldots,$$

For each $t=1,2,\ldots$, given the rvs $\{\xi_1,\ldots,\xi_t\}$, there exist two functions $\sigma_t: M^t \to \{0,\ldots,t\}^K$ and $\gamma_t: M^t \to \{0,1,\ldots,K\}^K$ such that

$$Q_t = \sigma_t(\xi_1,\ldots,\xi_t) \text{ and } V_t = \gamma_t(\xi_1,\ldots,\xi_t).$$

Let Ξ denote the random vector consisting of rvs $\{\alpha_{t+1}, \beta_{t+1}, O_{t+1}, t=0,\ldots,\tau-1\}$ within a cycle, i.e., $\Xi \equiv (\xi_1,\ldots,\xi_\tau)$.

For each $n=1,2,\ldots$, we denote by S_n the set of all possible realizations of the random vector Ξ with cycle length n, i.e.,

$$S_n = \{ s \equiv (s_1,\ldots,s_n) \in M^K: \sigma_t(s_1,\ldots,s_t) \neq 0,\ t=1,\ldots,n-1, \text{ and } \sigma_n(s) = 0\}.$$

For each $n=1,2,\ldots$, the set S_n is finite, and the set $S \equiv \bigcup_{n=1}^{\infty} S_n$ now represents the collection of all possible sample paths. For any sample path s in S, we denote by $\ell(s)$ the cycle length associated with s, and by $a(s)$ the total number of cells fed to the switch, pertaining to s, i.e.,

$$a(s) \equiv \sum_{t=1}^{\ell(s)} \sum_{k=1}^{K} s_t^{1,k}, \quad s \in S.$$

Under the independence assumptions enforced on the arrival processes $\{a_{t+1}^k, k=1,\ldots,K;\ t=0,1,\ldots\}$, the probability $P[\Xi = s]$ that a cycle is realized along the sample path s (in S) can be expressed by

$$\begin{aligned} P[\Xi = s] &= P[\xi_t = s_t, t=1,\ldots,\ell(s)] \\ &= c(s)\lambda^{a(s)}(1-\lambda)^{(K\ell(s)-a(s))} \end{aligned}$$

where we have set

$$c(s) \equiv P[(\beta_t, O_t) = (s_t^2, s_t^3), t=1,\ldots,\ell(s) \mid \alpha_t = s_t^1,\ t=1,\ldots,\ell(s)].$$

Defining a function $\phi: S \to N$ by

$$\phi(s) \equiv \sum_{t=0}^{\ell(s)-1} 1[\sigma_t^1(s) = b], \quad s \in S,$$

the measures $\Phi(\lambda)$ and $\Psi(\lambda)$ now can be rewritten as

$$\begin{aligned} \Phi(\lambda) &= \sum_{s \in S} P[\Xi = s]\phi(s) \\ &= \sum_{s \in S} \phi(s)c(s)\lambda^{a(s)}(1-\lambda)^{(K\ell(s)-a(s))}. \end{aligned} \quad (8)$$

and

$$\begin{aligned} \Psi(\lambda) &= \sum_{s \in S} P[\Xi = s]\ell(s) \\ &= \sum_{s \in S} \ell(s)c(s)\lambda^{a(s)}(1-\lambda)^{(K\ell(s)-a(s))}. \end{aligned} \quad (9)$$

Finally, for future use we define

$$T_n \equiv \{ s \in S: a(s) = n\}, \quad n=0,1,\ldots.$$

V. Light Traffic Derivatives

We first show that the first (2b-1) light traffic derivatives of CLP are all zero and that in view of (III.2) the first two non-zero light traffic derivatives are equal to the corresponding derivatives of the numerator $\Phi(\lambda)$, i.e.,

$$\frac{d^{2b}}{d\lambda^{2b}} P_b(0+) = \frac{d^{2b}}{d\lambda^{2b}} \Phi(0+) \quad \text{and} \quad \frac{d^{2b+1}}{d\lambda^{2b+1}} P_b(0+) = \frac{d^{2b+1}}{d\lambda^{2b+1}} \Phi(0+).$$

We then proceed to evaluate the $(2b)^{th}$ and $(2b+1)^{th}$ derivatives of $\Phi(\lambda)$.

We begin with a preliminary lemma: This lemma shows that in order to make the first queue full at least once during a regeneration cycle, i.e., $\phi(\cdot) > 0$, at least $2b$ cells must be generated during the cycle and the cycle length should be no shorter than $2b$. Let $S_\star$ denote the set of all sample paths in S along which at least once the first queue becomes full, namely

$$S_\star \equiv \{ s \in S: \phi(s) > 0\}.$$

Lemma 1 : *For any sequence* $s \in S_\star$, *we have*

$$\ell(s) \geq 2b \text{ and } a(s) \geq 2b \quad (9)$$

and furthermore,

$$\phi(s) = 0 \text{ and } \ell(s) = 2b, \quad s \in S_\star \cap T_{2b}. \quad (10)$$

Proof : Consider a sample path $s \in S_\star$. First note that each time slot at most one cell is fed to each input port and at most one cell can be transmitted from each queue. As a result, starting from an empty system, at least b time slots are required until the first queue becomes full. Similarly, once the first queue is full, at least b time slots must elapse for the system to become empty. Therefore, we have $\ell(s) \geq 2b$.

For any sample path $s \in S_\star$, $\phi(s) > 0$ implies that at least b cells out of $a(s)$ should be fed to the first input queue. In order to keep these b cells in the first queue, each time slot there must always exist at least one cell residing in the other queues, playing the role of blocking the HOL cell in the first queue, and we can conclude that $a(s) \geq 2b$. Finally, for any $s \in S_\star$ such that $a(s) = 2b$, it is plain that $\phi(s) = 1$ and $\ell(s) = 2b$. □

Lemma 2 : *For* $b=1,2,\ldots$, *we have*

(1) $\frac{d^k}{d\lambda^k}\Phi(0+)=0,\quad k=1,\ldots,2b-1;$

(2) $\frac{d^{2b}}{d\lambda^{2b}}\Phi(0+) = (2b)!\sum_{s\in S_*\cap T_{2b}} c(s);$

(3) $\frac{d^{2b+1}}{d\lambda^{2b+1}}\Phi(0+)=(2b+1)!\left(\sum_{s\in S_*\cap T_{2b+1}} c(s)\phi(s)-2b(K-1)\sum_{s\in S_*\cap T_{2b}} c(s)\right).$

Proof : By virtue of Lemma 1, for each $k=1,2,\ldots$, we have

$$\begin{aligned}\frac{d^k}{d\lambda^k}\Phi(\lambda) &= \sum_{n=2b}^{\infty}\sum_{s\in T_n} c(s)\phi(s)\frac{d^k}{d\lambda^k}(\lambda^n(1-\lambda)^{K\ell(s)-n}) \\ &= \sum_{n=2b}^{\infty}\sum_{s\in T_n} c(s)\phi(s)\sum_{i=0}^{k}\binom{k}{i}\left(\frac{d^i}{d\lambda^i}\lambda^n\right)\left(\frac{d^{k-i}}{d\lambda^{k-i}}(1-\lambda)^{K\ell(s)-n}\right)\end{aligned} \quad (11)$$

and it is now straightforward to check that Claim 1-Claim 2 hold. For $k=2b+1$, (11) yields

$$\begin{aligned}\frac{d^{2b+1}}{d\lambda^{2b+1}}\Phi(0+)= &-\sum_{s\in T_{2b}} c(s)\phi(s)(2b+1)(2b)!(K\ell(s)-2b)+ \\ &\sum_{s\in T_{2b+1}} c(s)\phi(s)(2b+1)!,\end{aligned}$$

and Claim 3 then follows by making use of (10). □

Lemma 3 : *For* $b=1,2,\ldots$, *we have*

(1) $\Psi(0+)=1;$

(2) $\frac{d}{d\lambda}\Psi(0+)=0.$

Proof : (Claim 1) Claim 1 immediately follows from the fact that $\Psi(0+)=\sum_{s\in T_0}\ell(s)c(s)$ and that $\ell(s)=1$ for any $s\in T_0$.

(Claim 2) Because

$$\begin{aligned}\frac{d}{d\lambda}\Psi(\lambda) &= \sum_{s\in S}\ell(s)c(s)a(s)\lambda^{a(s)-1}(1-\lambda)^{K\ell(s)-a(s)} \\ &-\sum_{s\in S}\ell(s)c(s)(K\ell(s)-a(s))\lambda^{a(s)}(1-\lambda)^{(K\ell(s)-a(s)-1)},\end{aligned}$$

we obtain

$$\frac{d}{d\lambda}\Psi(0+)=\sum_{s\in T_1}\ell(s)c(s)-\sum_{s\in T_0}Kc(s).$$

Hence, upon using the fact $\sum_{s\in T_1}\ell(s)c(s)=\sum_{s\in T_0}Kc(s)=K$, the desired result readily follows. □

From Lemmas 2-3, we have

Proposition 1 : *For* $b=1,2,\ldots,$

(1) $\frac{d^k}{d\lambda^k}P_b(0+)=0,\ k=0,1,\ldots,2b-1;$

(2) $\frac{d^{2b}}{d\lambda^{2b}}P_b(0+)=\frac{d^{2b}}{d\lambda^{2b}}\Phi(0+);$

(3) $\frac{d^{2b+1}}{d\lambda^{2b+1}}P_b(0+)=\frac{d^{2b+1}}{d\lambda^{2b+1}}\Phi(0+).$

Proof : Because for $k=1,2,\ldots,$

$$\frac{d^k}{d\lambda^k}P_b(\lambda)=\sum_{i=0}^{k}\binom{k}{i}\frac{d^i}{d\lambda^i}\Phi(\lambda)\frac{d^{k-i}}{d\lambda^{k-i}}\left(\frac{1}{\Psi(\lambda)}\right),$$

by making use of Lemma 5.2 (Claim 1) and Lemma 5.3 (Claim 1), we readily get Claim 1 and Claim 2, respectively. Finally, for $k=2b+1$, we get

$$\begin{aligned}\frac{d^{2b+1}}{d\lambda^{2b+1}}P_b(0+) &= (2b+1)\left(\frac{d^{2b}}{d\lambda^{2b}}\Phi(\lambda)\right)\left(\frac{d}{d}\lambda\frac{1}{\Psi(\lambda)}\right)\Bigg|_{\lambda=0+} \\ &+\left(\frac{d^{2b+1}}{d\lambda^{2b+1}}\Phi(\lambda)\right)\frac{1}{\Psi(\lambda)}\Bigg|_{\lambda=0+}\end{aligned}$$

and Claim 3 follows upon making use of Claim 2 of Lemma 3.□

VI. Calculation of the First Non-zero Derivatives

In this and the next sections, we evaluate the $(2b)^{th}$ and $(2b+1)^{st}$ derivatives of CLP, which by Lemma 2 and Proposition 1 is equivalent to evaluating $\sum_{s\in S_*\cap T_{2b}} c(s)$ and $\sum_{s\in S_*\cap T_{2b+1}} c(s)\phi(s)l\ !d.$ In this section, we focus on the evaluation of the quantity $\sum_{s\in S_*\cap T_{2b}} c(s)$; the calculation of $\sum_{s\in S_*\cap T_{2b+1}} c(s)\phi(s)$ is discussed in the next section. For the sake of convenience, throughout we assume that each cell is assigned its destination upon arrival, and keeps the address until it departs the switch; this assumption is equivalent to that provided in Section 2.

As should become clear in this section, for large values of b, it is not easy to find closed-form solutions for these two quantities: In order to cope with this difficulty, we provide upper bounds on these non-zero derivatives; these bounds turn out to be very tight when K is much bigger than b. Throughout we assume $K>b$.

Set

$$A\equiv\left\{(x_1,\ldots,x_b)\in\{0,1,\ldots,b\}^b:\sum_{t=1}^{b}x_t=b,\ \sum_{t=1}^{r}x_t-r\ge 0,\ r=1,\ldots,b\right\}.$$

Proposition 1 : For each $b=1,2,\ldots$, we have

$$\frac{d^{2b}}{d\lambda^{2b}}P_b(0+)\le\frac{(2b)!}{K^b}\sum_{x\in A}\prod_{t=1}^{b}\binom{K-1}{x_t}\frac{\sum_{r=1}^{t}x_r-t+1}{\sum_{r=1}^{t}x_r-t+2}. \quad (12)$$

Proof : We first establish a set of constraints that each element of $S_*\cap T_{2b}$ should satisfy. Consider a sample path s in the set $S_*\cap T_{2b}$. Given s, the system state at the beginning of time slot $[t,t+1)$ is given by

$$(\sigma_t(s_1,\ldots,s_t),\ \gamma_t(s_1,\ldots,s_t)),\quad t=1,\ldots,b-1$$

with the convention $\sigma_t(s_1,\ldots,s_t)=\gamma_t(s_1,\ldots,s_t)\equiv 0$ if $t=0$. During the first b time slots, of the $2b$ cells involved in s, b cells should be assigned to the first input port (i.e., one cell per each time slot) while the remaining b cells to the other input ports, i.e.,

$$s_{t+1}^{1,1} = 1,\ t=0,\dots,b-1; \qquad \sum_{t=0}^{b-1}\sum_{k=2}^{K} s_{t+1}^{1,k} = b. \tag{13}$$

In order for the first queue to become full through b time slots, no cell fed to the first queue should be transmitted and thus at least one cell should be present in the other queues, blocking the HOL cell in the first queue (we call the HOL cell in the first queue, the *tagged cell*). Therefore, the sample path s satisfies

$$\sum_{k=2}^{K} \sigma_t^k(s_1,\dots,s_t) + \sum_{k=2}^{K} s_{t+1}^{1,k} > 0,\quad t=0,\dots,b-1, \tag{14}$$

and because all the cells arriving to input ports $2,\dots,K$ (these cells are termed *blocking cells*) contend with the tagged cell for the same output at least once, the blocking cells should have the same address as the tagged cell, i.e., for $t=0,\dots,b-1$ and $k=2,\dots,K$,

$$1[\sigma_t^k(s_1,\dots,s_t)>0]\,\gamma_t^k(s_1,\dots,s_t) + 1[\sigma_t^k(s_1,\dots,s_t)=0,\ s_{t+1}^{1,k}>0]\,s_{t+1}^{2,k} = s_1^{2,1}. \tag{15}$$

Furthermore, one of the blocking cells must win the first b contentions with the tagged cell, and the sample path {bold s} also satisfies

$$s_{t+1}^{3,s_1^{2,1}} \neq 1,\ t=0,\dots,b-1, \tag{16}$$

while

$$s_{t+1}^{3,k}=0,\ k\neq s_1^{2,1},\ t=0,\dots,b-1.$$

At the end of each time slot, of the first b time slots, exactly one cell, chosen among the non-empty input queues $2,\dots,K$, departs the switch. Therefore, we have

$$\sum_{k=2}^{K} \sigma_{t+1}^k(s_1,\dots,s_{t+1}) = \sum_{k=2}^{K} \sigma_t^k(s_1,\dots,s_t) + \sum_{k=2}^{K} s_{t+1}^{1,k} - 1,\ t=0,\dots,b-1,$$

and thus

$$\sum_{k=2}^{K} \sigma_t^k(s_1,\dots,s_t) = \sum_{r=0}^{t-1}\sum_{k=2}^{K} s_{r+1}^{1,k} - t,\quad t=1,\dots,b. \tag{17}$$

By combining (13), (14) and (17), we obtain the constraint

$$\sum_{r=0}^{t-1}\sum_{k=2}^{K} s_r^{1,k} - t \geq 0,\ t=1,\dots,b-1 \text{ with } \sum_{t=0}^{b-1}\sum_{k=2}^{K} s_{t+1}^{1,k} = b. \tag{18}$$

Fix $t=0,\dots,b-1$. According to the constraints (15)-(17), we define the sets D, D_{t+1}', and D_{t+1}'' as follows:

$$D \equiv \left\{ x \equiv (x_1,\dots,x_b) \in \{0,1\}^{Kb}:\ x_t^1 = 1,\ t=1,\dots,b;\ \left(\sum_{k=2}^{K} x_1^k,\dots,\sum_{k=2}^{K} x_b^k\right) \in A \right\};$$

given $\sigma_t(s_1,\dots,s_t)$, $\gamma_t(s_1,\dots,s_t)$, and s_{t+1}^1,

$$D_{t+1}' \equiv \{ y \in \{1,\dots,K\}^K : y^k = s_1^{2,1} \text{ if } s_{t+1}^{1,k}>0,\ k=2,\dots,K\};$$ and given $\sigma_t(s_1,\dots,s_t)$, $\gamma_t(s_1,\dots,s_t)$, s_{t+1}^1, and s_{t+1}^2,

$$D_{t+1}'' \equiv \{ z \in \{0,1,\dots,K\}^K :\ z^{s_1^{2,1}} \neq 1 \}.$$

We are now ready to compute $\frac{d^{2b}}{d\lambda^{2b}} P_b(0+)$ as

$$\begin{aligned}\frac{d^{2b}}{d\lambda^{2b}} P_b(0+) &= (2b)! \sum_{s\in S_\star\cap T_{2b}} P[(\beta_t, O_t)=(s_t^2,s_t^3),\ t=1,\dots,b \mid \alpha_t = s_t^1,\ t=1,\dots,b] \\ &= (2b)! \sum_{(s_1^1,\dots,s_b^1)\in D} \prod_{t=0}^{b-1} \sum_{s_{t+1}^2\in D_{t+1}'} P[\beta_{t+1}=s_{t+1}^2] \\ &\quad \times \sum_{s_{t+1}^3\in D_{t+1}''} P\left[O_{t+1}=s_{t+1}^3 \,\middle|\, \begin{array}{c}(\alpha_r, \beta_r, O_r)=s_r,\ r=1,\dots,t, \\ (\alpha_{t+1}, \beta_{t+1}) = (s_{t+1}^1, s_{t+1}^2)\end{array} \right],\end{aligned}$$

and developing this equation further, we get

$$\begin{aligned}\frac{d^{2b}}{d\lambda^{2b}} P_b(0+) &= (2b)! \sum_{(s_1^1,\dots,s_b^1)\in D} \prod_{t=0}^{b-1} \frac{1}{K^K} K^{K-\sum_{k=2}^{K} s_{t+1}^{1,k}} \times \sum_{s_{t+1}^3\in D_{t+1}''} \left(\sum_{k=2}^{K} 1[\sigma_t^k(s_1,\dots,s_t)+s_{t+1}^{1,k}>0]+1 \right)^{-1} \\ &= (2b)! \sum_{(s_1^1,\dots,s_b^1)\in D} K^{-\sum_{t=0}^{b-1}\sum_{k=2}^{K} s_{t+1}^{1,k}} \prod_{t=0}^{b-1} \left| \frac{\sum_{k=2}^{K} 1[\sigma_t^k(s_1,\dots,s_t)+s_{t+1}^{1,k}>0]}{\sum_{k=2}^{K} 1[\sigma_t^k(s_1,\dots,s_t)+s_{t+1}^{1,k}>0]+1} \right|.\end{aligned}$$

Because $\sum_{k=2}^{K} 1[\sigma_t^k(s_1,\dots,s_t)+s_{t+1}^{1,k}>0] \leq \sum_{k=2}^{K} \sigma_t^k(s_1,\dots,s_t) + \sum_{k=2}^{K} s_{t+1}^{1,k},\ t=0,\dots,b-1,$ we have

$$\begin{aligned}\frac{d^{2b}}{d\lambda^{2b}} P_b(0+) &\leq (2b)! \sum_{(s_1^1,\dots,s_b^1)\in D} K^{-\sum_{t=0}^{b-1}\sum_{k=2}^{K} s_{t+1}^{1,k}} \prod_{t=0}^{b-1} \left| \frac{\sum_{k=2}^{K} (\sigma_t^k(s_1,\dots,s_t)+s_{t+1}^{1,k})}{\sum_{k=2}^{K} (\sigma_t^k(s_1,\dots,s_t)+s_{t+1}^{1,k}+1)} \right| \\ &= \frac{(2b)!}{K^b} \sum_{(s_1^1,\dots,s_b^1)\in D} \prod_{t=0}^{b-1} \left| \frac{\sum_{r=0}^{t}\sum_{k=2}^{K} s_{r+1}^{1,k} - t}{\sum_{r=0}^{t}\sum_{k=2}^{K} s_{r+1}^{1,k} - t+1} \right|\end{aligned}$$

where in the last step, we have used (17) and the fact $\sum_{t=0}^{b-1}\sum_{k=2}^{K} s_{t+1}^{1,k} = b$. The desired result (12) is now immediate by incorporating the set A into the last equation. □

It is worth pointing out that when $K \gg b$, the size of the set $\{ s\in S_\star\cap T_{2b}: \sum_{k=2}^{K} 1[\sigma_t^k(s_1,\dots,s_t)+s_{t+1}^{1,k}>0] < \sum_{k=2}^{K} \sigma_t^k(s_1,\dots,s_t) + \sum_{k=2}^{K} s_{t+1}^{1,k} \}$ is relatively very small compared to that of the set $\left\{ s\in S_\star\cap T_{2b}: \sum_{k=2}^{K} 1\left[\sigma_t^k(s_1,\dots,s_t)+s_{t+1}^{1,k}>0\right] = \sum_{k=2}^{K} \sigma_t^k(s_1,\dots,s_t) + \sum_{k=2}^{K} s_{t+1}^{1,k} \right\}$; indeed the upper bound is tight enough as should be clear from Fig. 2.

VII. Calculation of the Second Non-zero Derivatives

The evaluation of $\sum_{s\in S_\star\cap T_{2b+1}} c(s)\phi(s)$ is more complicated but similar to that of $\sum_{s\in S_\star\cap T_{2b}} c(s)$. While in the previous section, we needed to consider only one case where b cells (resp. b cells) are fed to the first queue (resp. input ports $2,\dots,K$) during the first b time slots and all HOL cells have the same address, in this section several cases arise in assigning ($2b$+1) cells to K input ports as well as their destination addresses; the set $S_\star\cap T_{2b+1}$ can be partitioned into the seven sets $S_1^*,\dots,S_7^*$ described below. For any $s\in S_\star\cap T_{2b+1}$, let $\hat{\ell}(s)$ denote the number of time slots required until the first queue becomes full for the first time.

(S_1^*) During the first b time slots, $(b, b+1)$ cells are fed to the

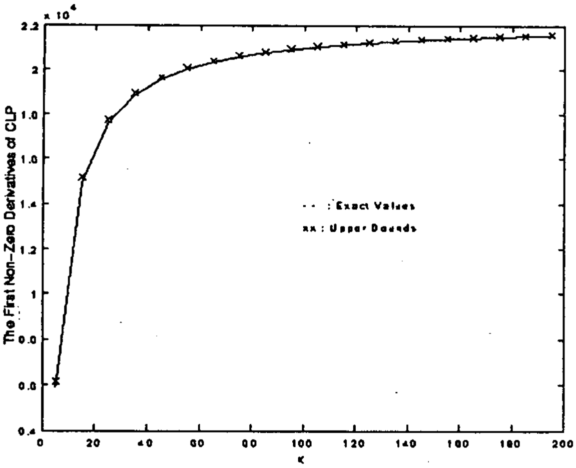


Fig. 2. Comparison of Exact Values and Upper Bounds(12) for $2b^{th}$ derivatives in IQ Switches with b=4.

first input port and the other ports, respectively. The addresses of the blocking cells are the same as that of the tagged cell, and $\hat{\ell}(s)=b$;

(S_2^*) During the first b time slots, $(b, b+1)$ cells are fed to the first input port and the other ports, respectively. All the addresses of the blocking cells are the same as that of the tagged cell, and $\hat{\ell}(s)=b+1$;

(S_3^*) During the first b time slots, $(b, b+1)$ cells are fed to the first input port and the other ports, respectively. All the addresses of the blocking cells are the same as that of the tagged cell except only one of the $(b+1)$ blocking cells, and $\hat{\ell}(s)=b$;

(S_4^*) During the first b time slots, (b, b) cells are fed to the first input port and the other ports, respectively, and one more cell arrives to the first port in time slot $[b, b+1)$. All the addresses of the blocking cells are the same as that of the tagged cell, and $\hat{\ell}(s)=b$;

(S_5^*) During the first b time slots, (b, b) cells are fed to the first input port and the other ports, respectively, and one more cell arrives to the first port in time slot $[b, b+1)$. The addresses of blocking cells are the same as that of the tagged cell. One of the first b cells in the first queue wins the contention, and $\hat{\ell}(s)=b+1$;

(S_6^*) During the first b time slots, (b, b) cells are allocated to the first port and the other ports, respectively. The b cells in the queues $2,\ldots,K$, have the same addresses as that of the tagged cell, and $\hat{\ell}(s)=b$. One more cell arrives into one of the input ports, $1,\ldots,K$, during one of the time slots $[b, b+1)$, $t=b+1,\ldots,2b-1$;

(S_7^*) During the first b time slots, (b, b) cells are allocated to the first input port and the other ports, respectively. The b cells in the queues $2,\ldots,K$, have the same addresses as that of the tagged cell, and $\hat{\ell}(s)=b$. One more cell arrives into one of the input ports, $2,\ldots,K$, during the time slot $[b, b+1)$ and its address is different from that of the tagged cell.

We define the sets A_k, $k=1,\ldots,b+1$, B_{b+1}, and C_{b+1} by

$$A_k \equiv \left\{ x\in\{0,1,\ldots,b\}^k : \sum_{i=1}^{k} x_i = k,\ \sum_{i=1}^{t} x_i - t \ge 0,\ t=1,\ldots,k,\ k=1,\ldots,b+1 \right\};$$

$$B_{b+1} \equiv \left\{ x\in\{0,1,\ldots,b\}^{b+1} : x_1 \ge 2,\ \sum_{i=1}^{b+1} x_i = b+1,\ \sum_{i=1}^{t} x_i - t \ge 0,\ t=1\ldots,b+1 \right\};$$

$$C_{b+1} \equiv \left\{ x\in\{0,1,\ldots,b\}^b : \sum_{i=1}^{b} x_i = b+1,\ \sum_{i=1}^{t} x_i - t \ge 0,\ t=1,\ldots,b \right\}.$$

Setting

$$P_n \equiv \sum_{s\in S_n^*} c(s)\phi(s),\ n=1,\ldots,7,$$

we have $\sum_{s\in S_*\cap T_{2b-1}} c(s)\phi(s) = \sum_{n=1}^{7} P_n$. The calculation of $P_1,\ldots, P_7$ proceeds similarly as was done in Section 6 and the details are omitted for the sake of brevity. By Claim 3 of Lemma V.2, we have

$$\begin{aligned}\frac{d^{2b+1}}{d\lambda^{2b+1}}\Phi(0+) &= (2b+1)!\left(\sum_{i=1}^{7} P_i - 2b(K-1)\sum_{s\in S_*\cap T_{2b}} c(s)\right) \\ &= (2b+1)!\left(P_1 + P_2 + P_5 + \left(\frac{2b+1}{K} - (b+1)\right)\sum_{s\in S_*\cap T_{2b}} c(s)\right),\end{aligned}$$

where setting $C_k \equiv \sum_{s\in S_*\cap T_{2k}} c(s), k=1,\ldots,b$, the parameters P_1, P_2 and P_5 are given by

$$P_1 \simeq \frac{3(K-1)}{2K} C_b + \frac{3}{2K^{b+1}} \sum_{x\in C_{b+1}} \prod_{l=1}^{b} \binom{K-1}{x_l} \frac{\sum_{i=1}^{l} x_i - l + 1}{\sum_{i=1}^{l} x_i - l + 2};$$

$$P_2 \simeq \frac{1}{K^{b+1}} \sum_{x\in B_{b+1}} \binom{K-1}{x_1} \prod_{l=2}^{b+1} \binom{K-1}{x_l} \frac{\sum_{i=1}^{l} x_i - l + 1}{\sum_{i=1}^{l} x_i - l + 2} + \frac{b-1}{K^{b+1}} \sum_{x\in A_{b+1}} \prod_{l=1}^{b+1} \binom{K-1}{x_l} \frac{\sum_{i=1}^{l} x_i - l + 1}{\sum_{i=1}^{l} x_i - l + 2};$$

and

$$P_5 \simeq \sum_{k=1}^{b} \frac{1}{K^{k+1}} \sum_{x\in A_k} \sum_{j=1}^{k} \prod_{l=1}^{j-1} \binom{K-1}{x_l} \frac{\sum_{i=1}^{l} x_i - l + 1}{\sum_{i=1}^{l} x_i - l + 2} \binom{K-1}{x_j} \frac{1}{\sum_{i=1}^{j} x_i - j + 2}$$
$$\times \prod_{l=j+1}^{k} \binom{K-1}{x_l} \frac{\sum_{i=1}^{l} x_i - l + 2}{\sum_{i=1}^{l} x_i - l + 3} \cdot \frac{1}{2} C_{b-k} + \sum_{k=1}^{b-1} C_k C_{b-k}.$$

VII. An Example

We have applied our result to the estimation of CLP in IQ switches with $K=20$ and $b=3$(See Fig. 3). The simulation results were obtained using the variance reduction technique called *importance sampling* in order to speed up the simulation because the CLP in this light traffic regime is very small and a plain Monte Carlo simulation takes a great amount of time.

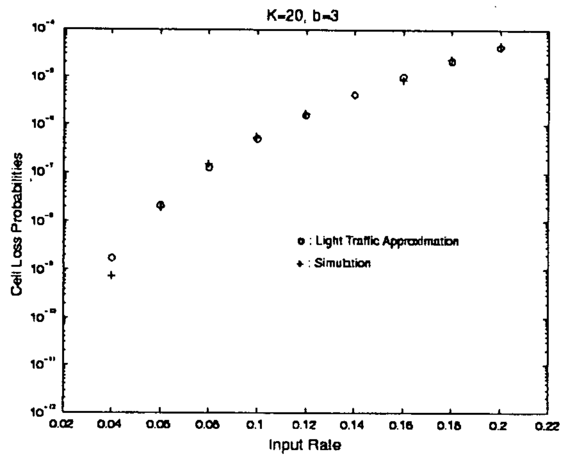


Fig. 3. Comparisons between Light Traffic Approximation(1.1) and Monte Carlo Simulation Results.

IX. Conclusions

In this paper, we have derived expressions for the non-zero derivatives of CLP with respect to the arrival rate evaluated in light traffic. These light traffic derivatives obtained are incorporated into an approximation of the CLP via a Taylor expansion, thereby providing an economical way to get a quick evaluation of CLP in light traffic. This result may be further extended by interpolating heavy and medium traffic values to yield a global configuration of input rate versus CLP.

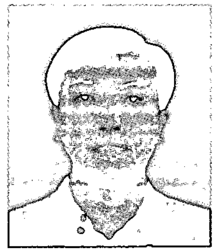
References

[1] Karol, M. J., M. Hluchyj, and S. P. Morgan "Input Versus Output Queueing on a Space-Division Packet Switch," in *IEEE Transactions on Communications*, COM-35(12): 1347-1356.

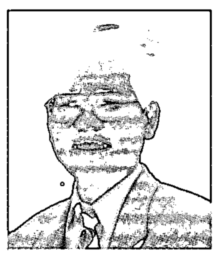
[2] Kim, Y.-B. and A. M. Makowski, "Cell Loss Probabilities in Input Queueing Crossbar Switches: Light Traffic Interpolations," Technical Report CCDS-TR-94-5, Center for Satellite and Hybrid Communication Networks, University of Maryland, College Park, MD.

[3] Reiman, M. I. and B. Simon "An Interpolation Approximation for Queueing Systems with Poisson Input," in *Operations Research*, 36: 454-469.

[4] Reiman, M. I. and B. Simon "Open Queueing Systems in Light Traffic," in *Operations Research*, 14: 26-59.



Young-Beom Kim recevied B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea in 1984 and 1986, respectively, and Ph.D. degree in electrical engineering from University of Maryland, College Park USA, in 1996. From 1986 to 1988, he was with Korea Telecom, Seoul, Korea. From 1988 to 1989, he was with the school of electrical engineering and computer science, Korea Institute of Technology, and from 1996 to 1997, he was with TelPerion Network Systems, Virginia, USA, as a member of technical staff. He is currently an assistant professor at Kon-Kuk University, Seoul, Korea. His research interests include traffic modeling and congestion control in ATM networks; ATM switch fabric architecture and performance evaluation, and inter-networking over ATM.



Jung Hur received B.S., M.S., and Ph.D. degrees from Seoul National University, Seoul, Korea, in 1981, 1983, and 1991, respectively, all in electronic engineering. From 1983 to 1984, he was with Electronics and Telecommunications Research Institute (ETRI), Taejon, Korea, as a researcher. He is currently an associate professor at Department of Electronics Eng., Kon-Kuk University, Seoul, Korea. His research interests include antenna analysis and propagation modelling in mobile communicatios, wireless and terrestrial ATM networks.