Formal Proof of Some Heuristics Used in the Importance Sampling Model to Estimate Rare Cell Delay Threshold Pobabilities

Ahmet A. Akyamaç
J. Keith Townsend
Center for Advanced Computing and Communication
North Carolina State University
Raleigh, N.C.

Abstract

This reported is intended as a followup and provides more formal proofs of the key theoretical developments used in the Importance Sampling model to estimate rare cell delay threshold probabilities in [1].

1 Introduction

The Importance Sampling model used in [1] to estimate rare cell delay threshold probabilities was developed based on theoretical arguments as well as extensive empirical arguments. This technical report provides formal and concise proofs of these arguments.

For sake of completeness, we restate the problem, the simulation system and the model. We then follow with the proofs.

2 System Description and Simulation Model

The model we use for the ATM switch is shown in Fig. 1. The switch has N_P input ports. Each connection into the network is routed to one of the N_P output ports through the non-blocking switch fabric. We assume homogeneous traffic in that all inputs have identical characteristics. An example of a non-blocking switch architecture is a shared bus. For this case, the shared bus operates at a speed of $N_P\hat{\lambda}$ so that all the output ports are given access to the shared bus. The output buffers have a finite size of K cells and use a FIFO queueing discipline. The cells that are not blocked (or lost), due to the fact that they encounter a full buffer upon arrival, exit the buffer at a service rate of $\hat{\mu}$.

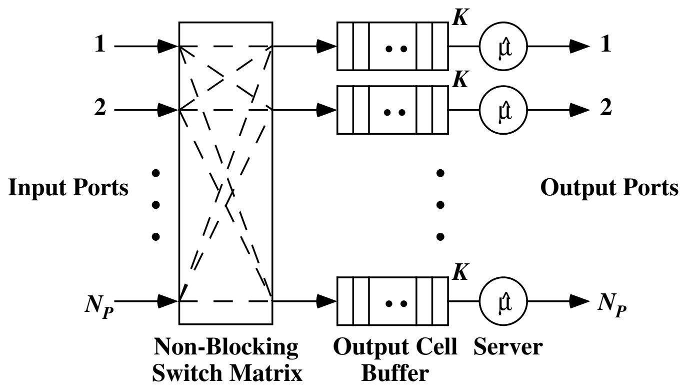

Figure 1: ATM switch.

Each connection into the ATM switch is characterized by the following connection traffic descriptors standardized by the ATM Forum User-Network Interface (UNI) Specification [2]:

$$\hat{\lambda}, \text{ peak cell rate (PCR)} \tag{1}$$

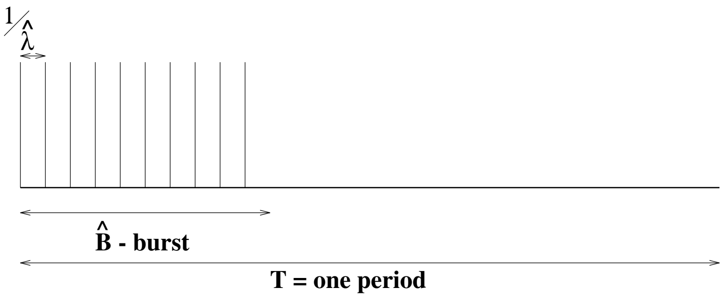

Figure 2: Greedy ON/OFF traffic.

$$\bar{\lambda}\text{, sustained (average) cell rate (SCR)} \tag{2}$$

$$\hat{B}\text{, maximum burst length in cells at the peak cell rate} \tag{3}$$

We refer to the collection of these parameters as the *triplet*. Hence, the homogeneous traffic assumption implies that all traffic sources have identical triplets.

Our goal is to determine the probability that a cell exiting the switch experiences an end-to-end delay greater than a specified threshold τ. We call this the *delay threshold probability*. Note that for a cell that gets blocked (due to the fact that it encounters a full buffer upon arrival), end-to-end delay is not well-defined. Thus, delay statistics should not be collected for blocked cells. The threshold τ can be specified either as a time unit or in reference to the queue length, K_m. In the latter approach, K_m is a percentage of the queue length K, $K_m = \alpha K$. Here, we assume that the routing of a cell to one of the output ports is instantaneous in that it has negligible effect on the cell's end-to-end delay. The traffic we consider is essentially worst-case UPC compliant traffic, which we consider to be the greedy ON/OFF periodic traffic pattern shown in Fig. 2. For a discussion of the worst-case UPC compliant traffic, refer to [1, 3, 4].

We assume uniform traffic routing to each of the N_P output ports so that we analyze only one "tagged" output buffer, which has a performance characteristic to all the output buffers. We also assume that the number of ports N_P is equal to the number of connections N_C routed through the output buffer, since this results in a maximum amount of congestion per output buffer after routing.

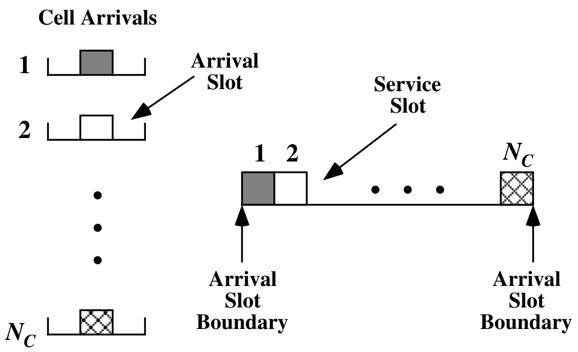

Figure 3: Arrival slot and service slot relationship.

We use a slotted-time simulation model where one arrival slot corresponds to the time necessary for the transmission or arrival of a 53-byte ATM cell. Our worst case assumption is a burst of $\hat{B}$ cells at the peak rate of $\hat{\lambda}$. Due to the fact that compliant traffic cannot exceed the average rate, which is the sustained rate $\bar{\lambda}$, the behavior of the input traffic due to one of the connections is periodic. Thus, the cells arrive according to ON/OFF periods T of $\hat{B}$ length bursts. This period is given by:

$$T = \hat{B}\hat{\lambda}/\bar{\lambda}, \text{ arrival slots} \tag{4}$$

Due to this cyclic behavior, the randomness in the simulation model is found in the connection starting-slot positions, which are each characterized by a discrete time random variable, uniform on $[0, T-1]$. The effective service rate per slot is given by:

$$\mu = \hat{\mu}/\hat{\lambda}, \text{ cells/arrival slot} \tag{5}$$

The speed of the shared bus can also be expressed as N_C cells per arrival slot. This corresponds to N_C service slots per arrival slot, as in Fig. 3, since the bus has to be able to route N_C cells to their corresponding output buffer(s) in one arrival slot. If at an arrival slot, the number of arrivals is $N < N_C$, then we assume that these arrivals occupy the first N service slots. This is also a worst-case assumption in terms of congestion. In each service slot, one cell can be loaded into the output buffer and μ/N_C cells can be serviced. If the output buffer contains less than μ/N_C cells, then all of these cells are serviced. Cell arrivals occur before any service (early arrivals) and service occurs for the cells in the buffer even if there are no cell arrivals in a service slot. The queue length behavior is accurately given using the service slot approach. An arrival slot approach gives the queue length changes for every N_C service slots. The queue length with respect to the service slots coincides with the queue length with respect to the arrival slots at the arrival slot boundaries [1, 3].

3 Problem Description

The problem is to find the delay threshold probabilities in the case of the finite buffer, $\Pr(D_K > \tau)$. In [1, 3], we conclude that $\Pr(D_K > \tau)$ is upper bounded by the difference of the delay threshold probability assuming an infinite buffer, $\Pr(D_\infty > \tau)$, and the cell loss probability, Pr(cell loss), as follows:

$$\Pr(D_K > \tau) \leq \Pr(D_\infty > \tau) - \Pr(\text{cell loss}) \tag{6}$$

Hence, we proceed to find the upper bound for the delay threshold probability, $\overline{\Pr(D_K > \tau)}$, given by:

$$\overline{\Pr(D_K > \tau)} = \Pr(D_\infty > \tau) - \Pr(\text{cell loss}) \tag{7}$$

Due to the fact that an analytical solution of this problem is not possible and also to the fact that extremely lengthy exhaustive runs throughout the sample space spanned by the connection starting-slot vectors are required to generate an exhaustive solution ([1, 3]), we use simulation to estimate the desired probabilities. The estimation problem for cell loss, Pr(cell loss), was considered in [4]. We consider the problem of estimating the delay threshold probability, $\Pr(D_\infty > \tau)$ for the infinite buffer. Throughout the remainder of this report, the delay threshold probability is denoted by P_{TH} for conciseness.

Following the development in [1, 3], the Monte Carlo simulation estimate for P_{TH} is given by:

$$\hat{P}_{TH} = \frac{\sum_{j=1}^{D_{max}} j\hat{p}_{TH_j}}{N_C\hat{B}} \tag{8}$$

where D_{max} is the maximum number of cells that can exceed the threshold for any given connection starting-slot configuration.

Here, $\hat{P}_{TH}$ is obtained by estimating the individual probabilities $\hat{p}_{TH_j}$. These individual estimates $\hat{p}_{TH_j}$ are found by running N_{MC} simulation runs, with one starting-slot vector drawn in every run, and observing the number of connection starting-slot vectors that cause j cells to exceed the threshold in steady state. After each of the N_{MC} independent runs, the number of cells that exceeded the threshold is stored for the corresponding indicator function. That is to say, the estimates $\hat{p}_{TH_j}$ are formed by:

$$\hat{p}_{TH_j} = \frac{1}{N_{MC}} \sum_{s=1}^{N_{MC}} I_j(s) \tag{9}$$

where $I_j(\cdot)$ is the indicator function for j cells exceeding the threshold τ:

$$I_j(\cdot) \;=\; \begin{cases} 1, \text{ if } j \text{ cells have } D_\infty > \tau \text{ in s.s.} \\ 0, \text{ otherwise} \end{cases} \quad \text{for } 1 \leq j \leq D_{max} \tag{10}$$

We refer to the individual probability structure $\hat{p}_{TH_j}$ as being composed of bins numbered from 0 to D_{max}, where each bin j represents the number of vectors that result in j cells exceeding the threshold.

Note that, for a given threshold τ, there is a minimum number of connections, $N_{C_0}(\tau)$, such that if there are less than $N_{C_0}(\tau)$ connections, no cells will exceed the threshold for any given starting-slot configuration. Also, there is a number of connections, N_{C_g} such that if there are more than N_{C_g} connections, all cells will eventually exceed the threshold in steady state. Simulations are performed in the range $N_{C_0}(\tau) \leq N_C \leq N_{C_g}$.

The expressions for D_{max}, $N_{C_0}(\tau)$ and N_{C_g} are crucial to the simulation process and were given in [1, 3]. In the next sections, we provide more formal derivations or proofs for these expressions.

Due to the extremely rare delay threshold probabilities in ATM networks (on the order of 10^{-9} to 10^{-12}), Monte Carlo simulation is not feasible since, to obtain sufficiently accurate estimates, typically 10^{11} to 10^{14} runs are required. In [1, 3], we have developed a framework based on Importance Sampling to efficiently simulate very low probabilities.

We have stated previously that the randomness in our model arises from the connection starting-slots, which are independently and uniformly distributed on the interval $[0, \cdots, T -$

1]. We can represent this randomness as a joint probability density function (pdf) $f_{\underline{V}}(\underline{v})$. The basic idea behind IS is to modify, or "bias", this pdf $f_{\underline{V}}(\underline{v})$ to a new pdf $f^*_{\underline{V}^*}(\underline{v}^*)$ such that the estimate is formed using this new pdf. We denote the new estimate by $\hat{P}^*_{TH}$. We require that the variance of the new estimate $\hat{P}^*_{TH}$ be reduced for a given number of simulation runs, or equivalently, the number of runs required be reduced for a given estimator variance. Each important event, identified by the end-to-end delay of a cell exceeding the threshold τ, must be appropriately weighted, or "unbiased", so that the new estimate $\hat{P}^*_{TH}$ remains unbiased. The only condition required on the new joint pdf $f^*_{\underline{V}^*}(\underline{v}^*)$ is that $f^*_{\underline{V}^*}(\underline{v}^*) > 0$ whenever the indicator function of the cell delay exceeding τ is nonzero.

The key to using IS is to find an appropriate pdf $f^*_{\underline{V}^*}(\underline{v}^*)$. In [1, 3], we have derived sets of d-distances for a given number of connections N_C. These d-distances are such that, if the bursts for the N_C connections do not all arrive within d arrival slots of each other, then no cells will exceed the threshold. The d-distances provide the biasing portion of the Importance Sampling scheme. A formal derivation of these d-distances will also be provided in the subsequent sections.

4 Proof of the AAZ Case

In this section, we prove that the AAZ case results in the maximum number of cells that exceed any given threshold τ. We also show that there is no other configuration that results in more cells that exceed the threshold.

Consider a system with a service rate of μ cells per arrival slot, with N_C connections each with a burst length of $\hat{B}$ cells and a threshold of K_m in queue length units. In this configuration, each cell contibutes by one to the overall queue length. In each service slot, μ/N_C cells are served. Also, arrivals occur prior to service in each service slot.

Note that for the system to be stable, we require that the sum of the mean arrival rates be less than the service rate, $\sum_{i=1}^{N_C} \bar{\lambda} \leq \hat{\mu}$. When translated to arrival slots this condition is equivalent to $\sum_{i=1}^{N_C} \bar{\lambda}/\hat{\lambda} \leq \mu$. Noting that $T = \hat{\lambda}\hat{B} \leq \bar{\lambda}$, the stability condition can be rewritten as $\sum_{i=1}^{N_C} \hat{B} < T\mu$. Since $\sum_{i=1}^{N_C} \hat{B}$ is the total number of cells that arrive in a period and $T\mu$ is the total service in a period, this signifies that for a stable system, the queue length must drop to 0 at least once in the overall period T. Note also that due to periodicity, the cell topology is translationally invariant in that shifting the starting-slots of all connections by the same amount does not alter the queue behavior or the total number of cells that exceed the threshold. Hence, we start observing the queue behavior at a certain arrival slot where the queue length is initially 0 and consider this slot to be slot 0.

Let us assume that F service slots must lapse until the first cell exceeds the threshold. Let N be the total number of cells required in the support of F service slots such that the next cell, which arrives at service slot $F+1$ exceeds the threshold. Then, we necessarily have:

$$N - \frac{F\mu}{N_C} + 1 > K_m \tag{11}$$

Hence, the minimum number of cells required so that the cell at service slot $F+1$ exceeds the threshold is given by:

$$N_{min} = \left\lfloor K_m - 1 + \frac{F\mu}{N_C} \right\rfloor + 1 \tag{12}$$

Denote the maximum number of cells that can exceed the threshold for any given starting-slot configuration by D_{max}. From Eq. (12), D_{max} is upper bounded as follows:

$$D_{max} \leq N_C \hat{B} - \left\lfloor K_m - 1 + \frac{F\mu}{N_C} \right\rfloor - 1 \tag{13}$$

An important question that has to be answered here is whether or not the absolute upper bound for D_{max} can be achieved for any starting-slot configuration. A starting-slot configuration that will always result in the absolute maximum for D_{max} is shown in Fig. 4. In this configuration, after N_{min} cell arrivals occur in the first F service slots, the remaining

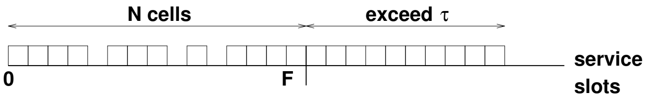

Figure 4: A configuration that is guaranteed to achieve the absolute upper bound for D_{max}.

$N_C \hat{B} - N_{min}$ cells arrive contiguously starting from service slot $F+1$. Since the cell at slot $F+1$ exceeds the threshold and the queue length is increasing after slot F, then all of the $N_C \hat{B} - N_{min}$ cells exceed the threshold. This is precisely the absolute upper bound for D_{max}.

Having established that, for a given F, there is a cell arrival configuration that results in the absolute upper bound for D_{max}, we now maximize this upper bound with respect to F. This maximization problem is given by:

$$\begin{aligned} & \max_{\{F\}} \left\{ N_C \hat{B} - \left\lfloor K_m - 1 + \frac{F\mu}{N_C} \right\rfloor - 1 \right\} \\ \equiv\ & \min_{\{F\}} \left\{ \left\lfloor K_m - 1 + \frac{F\mu}{N_C} \right\rfloor \right\} \\ \equiv\ & \min_{\{F\}} \{F\} \end{aligned} \tag{14}$$

We observe that maximizing D_{max} is equivalent to minimizing F. Since N cells arrive in the total support of F service slots and since no more than one cell can arrive in each service slot, we conclude that D_{max} is maximized at:

$$F^* = N \tag{15}$$

Incorporating F^* into Eq. (11), we obtain:

$$N - \frac{N\mu}{N_C} + 1 > K_m \tag{16}$$

Hence, we obtain the following:

$$N_{min} = \left\lfloor \frac{K_m - 1}{1 - \frac{\mu}{N_C}} \right\rfloor + 1 \tag{17}$$

Finally, D_{max} is upper bounded as in the following expression:

$$D_{max} \leq N_C \hat{B} - \left\lfloor \frac{K_m - 1}{1 - \frac{\mu}{N_C}} \right\rfloor - 1 \tag{18}$$

We conclude from Fig. 4 and the accompanying discussion that one way to achieve the absolute upper bound for D_{max} with $F = F^*$ is given in Fig. 5. The cell configuration shown

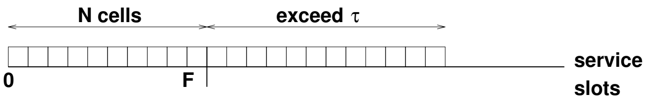

Figure 5: A configuration that is guaranteed to achieve the absolute upper bound for D_{max} at $F = F^*$.

in Fig. 5 corresponds to the AAZ case when the time slots are converted to arrival slots as seen in Fig. 6. Note also that the total support of the AAZ case, $\hat{B}$ is the minimum possible

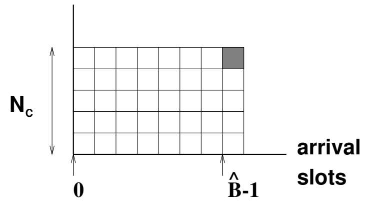

Figure 6: The aligned at zero (AAZ) case.

support achievable when cells arrive in bursts of length $\hat{B}$. The queue length at the end of the AAZ support is given by $N_C\hat{B} - \mu\hat{B}$ and is strictly greater than the queue length at the end of the support of any other connection starting-slot configuration.

From the above development, we can conclude the following:

- The maximum number of cells that can exceed the threshold for any given connection starting-slot configuration is given by:

$$D_{max} = N_C \hat{B} - \left\lfloor \frac{K_m - 1}{1 - \frac{\mu}{N_C}} \right\rfloor - 1 \tag{19}$$

- The AAZ case is guaranteed to produce D_{max} cells that exceed the threshold.

- There is no other connection starting-slot combination that can produce more than D_{max} cells that exceed the threshold.

- For any connection starting-slot configuration that produces D_{max} cells that exceed the threshold, there must be a contiguous arrival of F^* cells (in service slots) somewhere in the total support of all cells.

- Among all connection starting-slot configurations that produce D_{max} cells that exceed the threshold, the AAZ case results in the maximum queue length at the end of its support.

- Following from the above item, if no cells exceed the threshold for the AAZ case, then no cells can exceed the threshold for any other connection starting-slot configuration.

5 Proof of $N_{C_0}(\tau)$

We have shown in the previous section that the AAZ case generates the maximum number of cells that exceed the threshold (D_{max}) and also corresponds to the connection starting-slot configuration that results in the maximum queue length at the end of its support. Using these two facts, we find the minimum number of connections necessary so that at least one cell can exceed the threshold. We denote this number of connections by $N_{C_0}(\tau)$.

For the AAZ case, since all cells arrive contiguously in the service slots, the queue length is strictly increasing from the beginning of arrival slot 0 to the end of arrival slot $\hat{B} - 1$ (the total support of the AAZ case is $\hat{B}$ cells). Then $N_{C_0}(\tau)$ is such that $N_{C_0}(\tau) - 1$ connections correspond to the maximum number of connections possible such that no cells exceed the threshold for the AAZ case. Note that, as established in the previous section, if no cells exceed the threshold for the AAZ case, then no cells will exceed the threshold for any other connection starting-slot configuration.

Consider the AAZ case (as in Fig. 6) for $N_C = N_{C_0}(\tau) - 1$ connections. Each cell contributes by 1 to the overall queue length and in each service slot the amount of service is $\mu/(N_{C_0}(\tau) - 1)$. Arrivals occur before any service in a service slot. Then, we must have the following:

$$\hat{B}(N_{C_0}(\tau) - 1 - \mu) \leq K_m - \frac{\mu}{N_{C_0}(\tau) - 1} \tag{20}$$

This guarantees that the shaded cell in Fig. 6 does not exceed the threshold, and hence that none of the other cells exceed the threshold either. After some algebraic manipulations, we obtain the following second order equation:

$$(N_{C_0}(\tau) - 1)^2 - (\mu + \frac{K_m}{\hat{B}})(N_{C_0}(\tau) - 1) + \frac{\mu}{\hat{B}} \leq 0 \tag{21}$$

Since $N_{C_0}(\tau) - 1$ connections correspond to the maximum number of connections possible such that no cells exceed the threshold for the AAZ case, we extract the larger of the roots (or the respective floor if the root is not an integer):

$$N_{C_0}(\tau) - 1 = \left\lfloor \frac{\mu + \frac{K_m}{\hat{B}} + \sqrt{(\mu + \frac{K_m}{\hat{B}})^2 - \frac{4\mu}{\hat{B}}}}{2} \right\rfloor \tag{22}$$

Thus, $N_{C_0}(\tau)$ is given by:

$$N_{C_0}(\tau) = \left\lfloor \frac{\mu + \frac{K_m}{\hat{B}} + \sqrt{(\mu + \frac{K_m}{\hat{B}})^2 - \frac{4\mu}{\hat{B}}}}{2} \right\rfloor + 1 \tag{23}$$

Note that for the AAZ case at $N_{C_0}(\tau)$ connections, the cell that arrives in the final service slot in the support (shown by the shaded cell in Fig. 6) is guaranteed to exceed the threshold. Some cells that arrive prior to the shaded cell may also exceed the threshold. If there are such other cells that exceed the threshold, then there will be other vector combinations that result in D_{max} cells that exceed the threshold.

We stated in Section 4 that the queue length must drop to 0 at some slot for a stable system, hence we must have $N_C\bar{\lambda} \leq \hat{\mu}$. Otherwise, the queue length will keep inceasing until, in steady state, all cells exceed the threshold. Then, N_{C_g}, as defined in Section 3 is given by:

$$N_{C_g} = \left\lfloor \frac{\hat{\mu}}{\bar{\lambda}} \right\rfloor \tag{24}$$

Thus, simulations are performed in the range $N_{C_0}(\tau) \leq N_C \leq N_{C_g}$.

6 Proof of the d-distances

Given a number of connections, we define a d-distance as the *maximum* distance in arrival slots between the connection starting-slots of the first and last bursts such that at least one cell can exceed the threshold. Thus, a d-distance is such that, unless all connections start their bursts within d slots of each other, no cells will exceed the threshold.

The d-distances depend on the relative values of the service rate μ and the number of connections N_C. In the following, we develop the d-distances for the three different cases. Without loss of generality, we assume that one burst is always fixed at 0. This assumption is valid due to the translational invariance of the connection starting-slot vectors discussed above.

6.1 Case 1: $\mu < 1$

Let d_r denote the d-distance for $N_C = N_{C_0}(\tau) + r$ connections. We first consider the case where the number of connections is $N_{C_0}(\tau)$. For a cell to exceed the threshold, all $N_{C_0}(\tau)$ connections must arrive within d_0 slots of each other.

For this case, the service rate per arrival slot is less than 1. This implies that one cell alone in each arrival slot can cause the queue length to increase for that arrival slot. Hence, $d_0 > \hat{B}$.

To find d_0, we consider the configuration shown in Fig. 7. In this starting-slot configuration, one burst is fixed at slot 0 and another burst is fixed at slot d_0. The remaining $N_{C_0}(\tau) - 2$ bursts are distributed over the interval $[0, \cdots, d_0]$. By definition, all bursts in

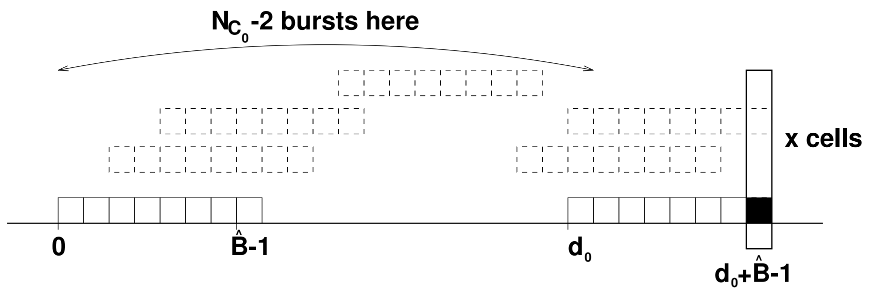

Figure 7: Calculation of d_0 for Case 1.

this configuration start within d_0 slots of each other. In the configuration in Fig. 7, we force one cell to exceed the threshold. Note that since, by definition, $N_{C_0}(\tau)$ cells are required for at least one cell to exceed the threshold, the queue length cannot drop to 0 in the interval $[0, \cdots, d_0]$. Also, since $\mu < 1$, if one cell in a burst exceeds the threshold, the remaining cells in that burst will also exceed the threshold. Hence, to find d_0, we force the last arriving cell of the connection fixed at slot d_0 to exceed the threshold. Then, d_0 is found by maximizing the distance obtained by varying the starting-slot positions of the $N_{C_0}(\tau) - 2$ bursts distributed over the interval $[0, \cdots, d_0]$.

Note also that the queue length will not drop to zero throughout the entire support of the bursts. Hence, the effects of the $N_{C_0}(\tau) - 2$ bursts can be quantified by how many cells of these bursts arrive at slot $d_0 + \hat{B} - 1$, as shown by the vertical rectangle in Fig. 7. We assume x cells arrive at slot $d_0 + \hat{B} - 1$ including the black cell in Fig. 7, where $1 \leq x \leq N_{C_0}(\tau) - 1$. Forcing the black cell to exceed the threshold, we have:

$$N_C \hat{B} - x - \mu(d_0 + \hat{B} - 1) + (x - 1)\left(1 - \frac{\mu}{N_C}\right) + 1 > K_m \tag{25}$$

After some algebraic manipulations, we obtain

$$d_0 < \frac{\hat{B}(N_{C_0}(\tau) - \mu) - K_m}{\mu} + \frac{1}{N_{C_0}(\tau)}(1 - x + N_{C_0}(\tau)) \tag{26}$$

The expression in Eq. (26) is maximized when $x = 1$, corresponding to only 1 cell in the last slot (slot $d_0 + \hat{B} - 1$). With the condition that d_0 is the largest possible integer, we find:

$$d_0 = \left\lceil \frac{\hat{B}(N_{C_0}(\tau) - \mu) - K_m}{\mu} + 1 \right\rceil - 1 \tag{27}$$

Next, we assume that the number of connections is increased by 1 to $N_C = N_{C_0}(\tau) + 1$. At $N_{C_0}(\tau) + 1$ connections, a cell may exceed the threshold either due to $N_{C_0}(\tau)$ bursts arriving within d_0 slots of each other or $N_{C_0}(\tau) + 1$ bursts arriving within d_1 slots of each other.

To find d_1, we consider a scenario as in Fig. 7, but where one burst starts at slot 0, one burst starts at slot d_1 and $N_{C_0}(\tau) - 1$ bursts are distributed over the interval $[0 \cdots d_1]$. We assume x cells arrive at slot $d_1 + \hat{B} - 1$ including the black cell as in Fig. 7, where for $N_C = N_{C_0}(\tau) + 1$ connections, $1 \leq x \leq N_{C_0}(\tau)$. We make a further assumption that there is no combination of $N_{C_0}(\tau)$ cells that arrive within d_0 slots of each other. Forcing the black cell as in Fig. 7 to exceed the threshold, we have:

$$(N_{C_0}(\tau) + 1)\hat{B} - x - \mu(d_1 + \hat{B} - 1) + (x - 1)\left(1 - \frac{\mu}{N_{C_0}(\tau) + 1}\right) + 1 > K_m \qquad (28)$$

After some algebraic manipulations, we obtain:

$$d_1 < \frac{\hat{B}(N_{C_0}(\tau) + 1 - \mu) - K_m}{\mu} + \frac{2 + N_{C_0}(\tau) - x}{N_{C_0}(\tau) + 1} \qquad (29)$$

The expression in Eq. (29) is again maximized when $x = 1$, which corresponds to only 1 cell in the last slot (slot $d_1 + \hat{B} - 1$). With the condition that d_0 is the largest possible integer, we find:

$$d_1 = \left\lceil \frac{\hat{B}(N_{C_0}(\tau) + 1 - \mu) - K_m}{\mu} + 1 \right\rceil - 1 \qquad (30)$$

From Eq.'s (27) and (30), it is easily inferred that for the general case of $N_C = N_{C_0}(\tau) + r$ connections,

$$d_r = \left\lceil \frac{\hat{B}(N_{C_0}(\tau) + r - \mu) - K_m}{\mu} + 1 \right\rceil - 1 \qquad (31)$$

6.2 Case 2: $\mu \geq 1$, $N_{C_0}(\tau) - 1 \geq \lceil \mu \rceil$

For this case, at least $\lfloor \mu \rfloor + 1$ cells per arrival slot are required for the queue length to increase for that arrival slot. We first consider $N_C = N_{C_0}(\tau)$ connections.

A physical constraint for this case is that $d_0 \leq \hat{B}$. If $d_0 \geq \hat{B} + 1$, this would imply that at least two bursts are separated by a distance greater than the burst length. Consider

$N_{C_0}(\tau) - 1$ bursts aligned to start at the same slot and one burst that arrives at a distance of $d_0 \geq \hat{B} + 1$. By definition, no cells will exceed the threshold in the block of $N_{C_0}(\tau) - 1$ bursts. Additionally, no cells will exceed the threshold in any burst a distance of $d_0 \geq \hat{B} + 1$ arrival slots away. This is due to the fact that at least one arrival slot will have no arrivals and will have a service of $\mu > 1$.

Thus, to find d_0, we start with the AAZ case and cancel the effects of cells pertaining to one of the bursts such that they no longer contribute to the final queue length (in essence, we remove the effects of the cells on the queue length, or simply "remove" the cells). We "remove" cells from the last burst starting from the rightmost support and proceed towards slot 0 or starting from the leftmost support and proceed towards slot $\hat{B} - 1$, as seen in Fig. 8. The rationale behind removing cells from the last burst and not any other burst is that,

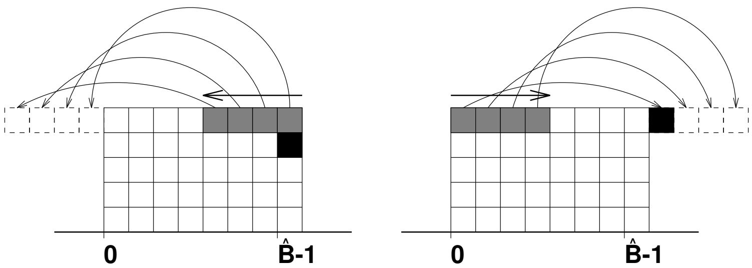

Figure 8: Calculation of d_0 for Case 2.

by definition, $N_{C_0}(\tau) - 1$ bursts alone are not sufficient for a cell to exceed the threshold. Burst $N_{C_0}(\tau)$ provides the cells required to increase the queue length beyond the threshold.

Let f be the number of cells removed. We find the maximum possible value of f such that one cell is guaranteed to exceed the threshold. The cell that exceeds the threshold varies depending on the direction of the removal process. For both cases, the one cell that is guaranteed to exceed the threshold is shown in black in Fig. 8.

Let y be the number of cells that arrive in the arrival slot in which the black cell arrives. Then $y \in \{N_{C_0}(\tau) - 1, N_{C_0}(\tau)\}$. Let x be a variable depicting the direction of the removal, where $x = 1$ corresponds to removing from left to right and $x = 0$ corresponds to the opposite.

Forcing the black cell to exceed the threshold, we have the following equation that accounts for both cases:

$$N_{C_0}(\tau)(\hat{B}-1)-\mu(\hat{B}-1)-(f-(N_{C_0}(\tau)-y))+(y-1)\left(1-\frac{\mu}{N_{C_0}(\tau)}\right)+$$
$$x(N_{C_0}(\tau)-(y-1))\left(1-\frac{\mu}{N_{C_0}(\tau)}\right)+1>K_m \tag{32}$$

Solving for f yields:

$$f<\hat{B}(N_{C_0}(\tau)-\mu)-K_m+\frac{\mu}{N_{C_0}(\tau)}(N_{C_0}(\tau)-(y-1))+x(N_{C_0}(\tau)-(y-1))\left(1-\frac{\mu}{N_{C_0}(\tau)}\right) \tag{33}$$

Thus,

$$x=\begin{cases} 0, & f<\hat{B}(N_{C_0}(\tau)-\mu)-K_m+\frac{\mu}{N_{C_0}(\tau)}(N_{C_0}(\tau)-(y-1)) \\ 1, & f<\hat{B}(N_{C_0}(\tau)-\mu)-K_m+1 \end{cases} \tag{34}$$

Since we want to maximixe f, then $y=N_{C_0}(\tau)-1$ for $x=0$. Hence, combining the two cases for $x=0$ and $x=1$, we have:

$$f<\hat{B}(N_{C_0}(\tau)-1)-K_m+\max\left\{1,\frac{2\mu}{N_{C_0}(\tau)}\right\} \tag{35}$$

For this case, because $\mu>1$, the maximum number of arrival slots necessary to cancel the effects of f cells, thus keeping the queue length unchanged, is f. Hence, d_0 satisfies the following:

$$d_0<\hat{B}(N_{C_0}(\tau)-\mu)-K_m+\max\left\{1,\frac{2\mu}{N_{C_0}(\tau)}\right\} \tag{36}$$

Here, d_0 is chosen as the largest possible integer, yielding:

$$d_0=\left\lceil\hat{B}(N_{C_0}(\tau)-\mu)-K_m+\max\left\{1,\frac{2\mu}{N_{C_0}(\tau)}\right\}\right\rceil-1 \tag{37}$$

Next, we consider $N_C=N_{C_0}(\tau)+1$ connections. We know from the above analysis that $N_{C_0}(\tau)$ bursts that start within d_0 slots of each other can cause at least one cell to exceed the threshold. Hence, d_1 should be such that, assuming no combination of $N_{C_0}(\tau)$ bursts arrive within d_0 slots of each other, $N_{C_0}(\tau)+1$ bursts arriving within d_1 slots of each other can cause at least one cell to exceed the threshold (note that at $N_C=N_{C_0}(\tau)+1$ connections, cells can exceed the threshold either as a result of $N_{C_0}(\tau)$ bursts that arrive within d_0 slots of each other or as a result of $N_{C_0}(\tau)+1$ bursts that arrive within d_1 slots of each other).

Since $d_0 \leq \hat{B}$, we must necessarily have $d_1 > \hat{B}$. This is because $N_{C_0}(\tau)$ connections are sufficient for a cell to exceed the threshold when $N_{C_0}(\tau)$ bursts start within d_0 slots of each other. When $\hat{B}$ more cells are added with burst $N_{C_0}(\tau)+1$, the number of "removable" cells from the AAZ case is greater than $\hat{B}$. To find d_1, we again consider the removal of f cells

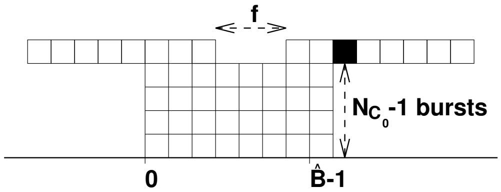

Figure 9: Calculation of d_1 for Case 2.

such that one cell exceeds the threshold. The assumption is that no combination of $N_{C_0}(\tau)$ bursts arrive within d_0 slots of each other. Thus, the cell that exceeds the threshold does so due to $N_{C_0}(\tau)+1$ bursts that arrive within d_1 slots of each other. The resulting removal scheme is shown in Fig. 9. Note that since $d_1 > \hat{B}$, the configurations in which cells from both bursts are removed towards the right or left simultaneously are actually cases in which a cell exceeds the threshold due to $N_{C_0}(\tau)$ cells arriving within d_0 slots of each other and hence do not aid us in finding d_1.

When the black cell in Fig. 9 is forced to exceed the threshold, we have:

$$N_{C_0}(\tau)\hat{B} - \mu\hat{B} - f + 1 > K_m \tag{38}$$

Solving for f, we obtain:

$$f < \hat{B}(N_{C_0}(\tau) - \mu) - K_m + 1 \tag{39}$$

Due to the configuration in Fig. 9, we have $d_1 = f + \hat{B}$. Thus,

$$d_1 = \left\lceil \hat{B}(N_{C_0}(\tau) + 1 - \mu) - K_m + 1 \right\rceil - 1 \tag{40}$$

Generalizing to the case of $N_C = N_{C_0}(\tau) + r$ connections, $r \geq 1$, we have:

$$d_r = \left\lceil \hat{B}(N_{C_0}(\tau) + r - \mu) - K_m + 1 \right\rceil - 1 \tag{41}$$

6.3 Case 3: $\mu \geq 1$, $N_{C_0}(\tau) - 1 \leq \lfloor \mu \rfloor$

Once again, we initially consider $N_C = N_{C_0}(\tau)$ connections. This case is different than the previous case in that when a cell is removed from a slot in which there are $N_{C_0}(\tau)$ cells, the remaining $N_{C_0}(\tau)-1$ cells can no longer contribute to the queue length since $N_{C_0}(\tau)-1 \leq \mu$. Hence, at least $N_{C_0}(\tau)$ cells are required in an arrival slot for the queue length to increase for that arrival slot. This situation is depicted as downward pointing arrows in Fig. 10. Note that, as in Case 2, there are two scenarios depending on the direction in which cells are removed. The physical properties also result in the constraint $d_0 \leq \hat{B}$ as in Case 2. Again,

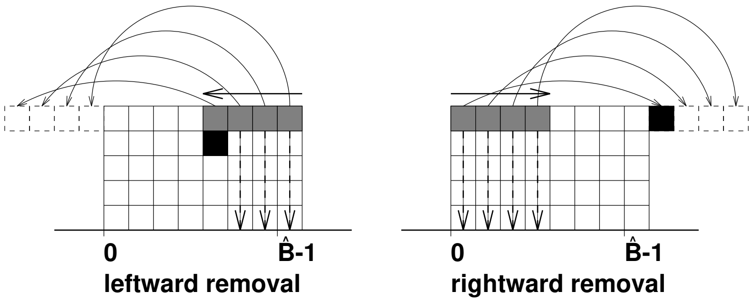

Figure 10: Calculation of d_0 for Case 3.

we find the maximum possible value of f such that the last cell (shown by the black cell in Fig. 10 for both leftward and rightward removal) exceeds the threshold.

We define the parameters x and y as in Case 2. Here, $y = N_{C_0}(\tau) - 1$ for $x = 0$ and $y = N_{C_0}(\tau)$ for $x = 1$. Forcing the black cell to exceed the threshold, we have the following equation that accounts for both cases:

$$N_{C_0}(\tau)(\hat{B} - f - x) - \mu(\hat{B} - f - x) + x\left(1 - (N_{C_0}(\tau) - y + 1)\frac{\mu}{N_{C_0}(\tau)}\right) + 1 > K_m \quad (42)$$

Solving for f, we find:

$$f(N_{C_0}(\tau) - \mu) < \hat{B}(N_{C_0}(\tau) - \mu) - K_m + \frac{\mu}{N_{C_0}(\tau)}(1-y) + y + x\left(1 - (N_{C_0}(\tau) - y + 1)\frac{\mu}{N_{C_0}(\tau)}\right) \quad (43)$$

Thus,

$$x = \begin{cases} 0, & f < \hat{B} - \frac{K_m - 1}{N_{C_0}(\tau) - \mu} + \frac{N_{C_0}(\tau) - 2}{N_{C_0}(\tau)} \\ 1, & f < \hat{B} - \frac{K_m - 1}{N_{C_0}(\tau) - \mu} \end{cases} \quad (44)$$

Case	d_0
$\mu < 1$	$\left\lceil \frac{\hat{B}(N_{C_0}(\tau)-\mu)-K_m}{\mu} + 1 \right\rceil - 1$
$\mu \geq 1, N_{C_0}(\tau) - 1 \geq \lfloor \mu \rfloor + 1$	$\left\lceil \hat{B}(N_{C_0}(\tau) - \mu) - K_m + \max\left\{1, \frac{2\mu}{N_C}\right\} \right\rceil - 1$
$\mu \geq 1, N_{C_0}(\tau) - 1 \leq \lfloor \mu \rfloor$	$\left\lceil \hat{B} - \frac{K_m-1}{N_{C_0}(\tau)-\mu} + \frac{N_{C_0}(\tau)-2}{N_{C_0}(\tau)} \right\rceil - 1$
Case	**$d_r, r \geq 1$**
$\mu < 1$	$\left\lceil \frac{\hat{B}(N_{C_0}(\tau)+r-\mu)-K_m}{\mu} + 1 \right\rceil - 1$
$\mu \geq 1, N_{C_0}(\tau) - 1 \geq \lfloor \mu \rfloor + 1$	$d_r = \left\lceil \hat{B}(N_{C_0}(\tau) + r - \mu) - K_m + 1 \right\rceil - 1$
$\mu \geq 1, N_{C_0}(\tau) - 1 \leq \lfloor \mu \rfloor$	$d_r = \left\lceil r\hat{B} - \frac{K_m-1}{N_{C_0}(\tau)-\mu} \right\rceil - 1$

Table 1: Table listing the d-distances for the different cases.

We observe that for Case 3, the case for $x = 1$ is a subset of the case for $x = 0$. Hence, we have:

$$f < \hat{B} - \frac{K_m - 1}{N_{C_0}(\tau) - \mu} + \frac{N_{C_0}(\tau) - 2}{N_{C_0}(\tau)} \tag{45}$$

By similar arguments as in Case 2, we have $d_0 = f$. Choosing d_0 as the largest possible integer results in:

$$d_0 = \left\lceil \hat{B} - \frac{K_m - 1}{N_{C_0}(\tau) - \mu} + \frac{N_{C_0}(\tau) - 2}{N_{C_0}(\tau)} \right\rceil - 1 \tag{46}$$

For the general case of $N_C = N_{C_0}(\tau) + r$ connections, proceeding as in Case 2, we obtain the following for $r \geq 1$:

$$d_r = \left\lceil r\hat{B} - \frac{K_m - 1}{N_{C_0}(\tau) - \mu} \right\rceil - 1 \tag{47}$$

6.4 Results

Table 1 lists the results of the proofs of the d-distances.

7 Conclusion

This report is intended as a followup of [1, 3]. We started by restating the problem of the simulation of the ATM switch for the rare delay threshold probabilities and proceeded to give a brief outline of the solution technique we developed in [1, 3]. The simulation technique developed in [1, 3] relied on a number of factors.

The first of these was the fact that the configuration that resulted in the maximum number of cells that exceed the threshold is the aligned-at-zero (AAZ) case. Here, we formally proved this fact and demonstrated some other useful properties of the AAZ case.

Second, we formally derived the minimum number of connections required for at least one cell to exceed the threshold, $N_{C_0}(\tau)$ and the maximum number of connections N_{C_g}, such that if there are more than N_{C_g} connections, all cells will eventually exceed the threshold in steady state. The formal proof here is similar but more detailed than the proof given in [1, 3].

Finally, we formally proved the d-distances, which were previously derived in a manner based on theoretical as well as empirical arguments.

References

[1] A. A. Akyamac and J. K. Townsend. Efficient Simulation of Delay Threshold Probabilities in ATM Networks with ON/OFF Arrivals Using Connection Traffic Descriptors. Technical Report TR-96/03, Center for Adv. Comput. and Comm., North Carolina State University, Feb. 1996.

[2] ATM User-Network Interface Specification, Version 3.0. The ATM Forum, September 10, 1993. The ATM Forum.

[3] A. A. Akyamac. Efficient Simulation of Delay Threshold Probabilities in ATM Networks with ON/OFF Arrivals Using Connection Traffic Descriptors. Master's thesis, North Carolina State University, Nov. 1995.

[4] J. A. Freebersyser. *Efficient Simulation of Cell Loss Probability in ATM Networks Using Connections Traffic Descriptors.* PhD thesis, North Carolina State University, October 1995.