# Performance Modeling of Differential Fair Buffer Allocation*

V. Bonin†, O. Casals‡, B. Van Houdt♮ and C. Blondia♮

†TU Hamburg-Harburg
Digital Kommukationssysteme
D-21071 Hamburg - Germany

‡Polytech. Univ. of Catalonia
Dept. Computer Architecture
E-08071 Barcelona, Spain
olga@ac.upc.es

♮University of Antwerp
Math. and Computer Science
B-2610 Antwerp, Belgium
{vanhoudt,blondia}@uia.ua.ac.be

## Abstract

The Guaranteed Frame Rate (GFR) service is viewed as the most promising ATM service for carrying aggregate TCP/IP traffic over large distances. In this work, we develop an analytical model to assess the performance of TCP over the Differential Fair Buffer Allocation implementation suggested by the ATM Forum. We consider the problem of a single GFR VC fed by multiple TCP Reno sources. The proposed model distinguishes itself from prior work in two ways: it captures the behavior of aggregate TCP traffic and it explicitly allows for queue analysis. From a modeling point of view, our study shows that the reactive behavior of TCP in congestion avoidance can be approximated by a two-node Markov chain. In terms of performance measures, we quantify the impact of traffic aggregation on queueing performance. Among many other results, our model predicts that, although the mean queue length seems to reach a maximum at a certain aggregation level, the mean loss probability appears to increase linearly with the number of sources.

**Keywords:** Guaranteed Frame Rate, Differential Fair Buffer Allocation, TCP Flow and Congestion Control, Internet Backbones, Matrix-Geometric Methods.

## 1 Introduction

The *Guaranteed Frame Rate* service category (GFR, formerly UBR+) was proposed by the ATM Forum [3] as an enhancement to the UBR service. GFR aims to provide Minimum Cell Rate (MCR) guarantees and a fair share of the excess bandwidth without the cost of rate-based flow control. Service guarantees are provided in terms of complete AAL5 frames, making it particularly suitable for internetworks and other frame-based applications that were not designed to run over Quality of Service (QoS) networks. Two distinct conformance definitions coexist. GFR.1 specifies that the network must convey the CLP bit of the ATM cells transparently, and, therefore, that frame tagging is not allowed. In such case, each ATM node relies on its own resources to distinguish between minimum rate guarantee eligible and ineligible frames. The GFR.2 conformance definition allows tagging, and thus provides the network with an explicit marker for eligible and ineligible flows.

Three GFR implementations have been suggested by the ATM Forum [3]. A first implementation uses a single FIFO buffer ruled by a double Early Packet Discard (EPD) scheme and relies on frame tagging to discard QoS non-eligible packets (GFR.2). A second implementation requires per-VC queueing and uses a Weighted Fair Queueing (WFQ) scheduler to share the bandwidth

*The first author of this work was supported in part by the "Comissionat per a Universitats i Recerca" of the "Generalitat de Catalunya" under Project 1999SGR00126 and the Deutscher Akademischer Austauschdienst through the Technical University Hamburg-Harburg.

†Corresponding Author. Current Mailing Address: Institut für Neuroinformatik, ETHZ/UNIZH, Winterthurerstr. 190, Haus 55, 8057 Zurich, Switzerland. E-mail: vincent@ini.phys.ethz.ch

among the VCs. A third implementation, the Differential Fair Buffer Allocation (DFBA) scheme, uses a FIFO queue, dynamic thresholds, and probabilistic drops to provide approximate MCR guarantees. An alternative implementation was proposed by Cerdán et al. [11]. Other possible implementations are reviewed in [2].

Bonaventure showed that, in many circumstances, the FIFO with tagging implementation does not allow the TCP sources to make efficient use of the reserved and excess bandwidth [4, 5]. The main cause of this are undesirable interactions between the F-GCRA tagging and TCP flow control. TCP traffic is inherently bursty and the losses enforced by F-GCRA and EPD are highly correlated. Correlated losses cause TCP to timeout and throughput to drop substantially. The WFQ implementation is known to perform much better [4, 5]. However, its high implementation complexity conflicts with the objectives of simplicity pursued by GFR.

We have already investigated the fairness of DFBA with the help of simulations [6]. In this work, we develop an analytical model for the performance of TCP over DFBA. Modeling objectives are two-fold: to capture the behavior of aggregate TCP traffic, and to study the interaction between TCP flow control and the DFBA scheduling mechanisms.

Numerous attempts have been made to capture the behavior of TCP by the use of analytics. Misra et al. [23] modeled the behavior of the TCP congestion window when subjected to independent packets losses. Results are shown fairly accurate for the Early Random Drop (ERD) and the Random Early Detection (RED) packet dropping policies (see [13] and [20]). Kumar [19] designed a stochastic model to evaluate the performance of various TCP implementations when communicating over a wireless link. Padhye et al. developed [25] a closed-form formula for the end-to-end throughput of internet ftp transfers. In their study, internet-wide measurements show that such transfers are well modeled by connections in congestion avoidance.

None of these models allow for queue analysis, nor can they be resolved for more than a few TCP connections. A step in this direction was made by Casetti and Meo [10], who have isolated the source model from the network model and performed reciprocal tuning of the two models to obtain steady-state measures. However, in addition to not being aimed at modeling aggregated TCP traffic, their method disregards the correlation present at the output of the TCP sources, and thus cannot be applied to obtain the queue occupancy measures.

In this work, we consider the problematic of a single GFR DFBA queue fed directly by multiple TCP sources (no shared-media LAN). Two different models for TCP flow control are proposed. A first model captures in some details the behavior of the congestion window of a single TCP connection in congestion avoidance. It consists of a discrete-time Markov chain, where each effective window size is associated an independent state. Throughout this work, we will refer to this model as the *Detailed Source* (DS) model.

Then we simplify the DS model through a binary quantization of the window size. We map the states of the underlying process onto a two-node Markov chain and use heuristic to derive "equivalent" transition probabilities. We shall refer to this model as the *Approximate Source* (AS) model. We support our heuristic through a first-passage analysis of the Markov chain ruling the DS model.

Next we show how to multiplex one detailed source (if any) with an arbitrary number of approximate sources, and include an approximation of a DFBA queue in our analysis. The Markov chain obtained forms a level-dependent quasi-birth-and-death (LDQBD) process for which a matrix-geometric solution of the steady-state vector exists (see [15], [8] and [24]). The system modeling approach we take is inspired from Blondia et al. [9]. The matrix-geometric method was proposed by Wuyts et al. [26].

The structure of this paper is as follows: Section 2 introduces the reader to the Differential Fair Buffer Allocation implementation of the GFR service category. Section 3 discusses the modeling assumptions in details. The model is developed in Section 4, where we thoroughly cover the modeling of TCP, the traffic multiplexing and the queue analysis. We show some exemplary results in Section 5 and conclude our work in Section 6.

## 2 Differential Fair Buffer Allocation

Differential Fair Buffer Allocation was proposed by Goyal et al. as a possible implementation of the GFR service [3, 16]. Their scheme uses a single FIFO queue and relies on per-VC accounting to provide service guarantees. Alike to RED gateways [13], DFBA attempts to control the average queue length by probabilistically dropping packets. Unlike RED, DFBA's drop scheme is stateless (i.e. losses are geometric), and solely relies on instantaneous queue measurements.

Consider a GFR queue shared by $N$ active connections (VCs), identified by $i$, such that $1 \leq i \leq N$. A VC is active when the queue contains at least one of its cells. Each VC is associated a weight $W_i$ equal to the ratio of the per-VC Minimum Cell Rate (MCR) allocation over the GFR capacity $\mathrm{MCR}_i/\mathrm{LINK}$. The state variable $X_i$ tracks the number of cells belonging to a single connection $i$ that are queued in the buffer. The total buffer occupancy is denoted by $X = \sum X_i$. As in other GFR implementations, DFBA defines two static thresholds, LBO and HBO. If $X$ is below LBO, all packets are accepted. Above HBO all packets are discarded. In the range LBO$\leq X \leq$HBO, all CLP=1 frames are discarded. The acceptance of CLP=0 frames depends on the connection's estimated fair share $X \cdot W_i/W$ of the buffer. If $X_i < X \cdot W_i/W$, the connection uses less than it's fair share and all CLP=0 frames are accepted. If $X_i > X \cdot W_i/W$, CLP=0 frames are dropped with probability

$$Z_i \cdot \left( \alpha \frac{X_i - X \cdot W_i/W}{X \cdot (1 - W_i/W)} + (1-\alpha) \frac{X - \mathrm{LBO}}{\mathrm{HBO} - \mathrm{LBO}} \right) \qquad (1)$$

where $W = \sum W_i$, and $Z_i$ defines the maximum drop probability enforceable by DFBA. Equation 1 is function of two additive terms. The left-hand term is a measure of relative fairness and requires per-VC accounting. It is known as the *fairness component* of the DFBA packet drop probability. The second term varies linearly with the total buffer occupancy if it lies between LBO and HBO. Since it attempts to keep the average buffer occupancy low, we will refer to it as the *efficiency component*. The parameter $\alpha$ determines the relative weights of the fairness and the efficiency components, and is typically set to 0.5 [16]. Goyal et al. showed that DFBA can provide MCR guarantees for TCP traffic for various buffer sizes and bandwidth allocations [16]. DFBA was also shown to favor connections with small MCR when sharing the link bandwidth. Setting $Z_i = (1 - W_i/W)^2$ can help to reduce the bias against large MCR connections [16].

DFBA differs from the other suggested GFR implementations in various aspects. Firstly, DFBA does not support the Maximum Burst Size (MBS) parameter in the GFR traffic contract, which means that bursts smaller than the MBS could still experience a relatively high loss probability. Secondly, the drop policy assumes that the traffic sources react to losses by lowering their transmission rate. DFBA does not provide strict isolation between the VCs, and a non-collaborative source could certainly use the bandwidth reserved to others. Moreover, there could be problems of fairness between sources using different flow control mechanisms.

In a previous paper, we investigated the fairness of DFBA using simulation [6]. Our main findings were:

- DFBA is not affected by the presence of high priority rt-VBR traffic in the ATM switch, but is unable to guarantee a fair share of the link to VCs with different MCR allocations and round-trip-times.

- The average buffer occupancy can be effectively controlled by the parameter $Z_i$, but $\alpha$ had virtually no effect on the fairness of DFBA.

Despite these flaws, the DFBA proposal remains very interesting, because it simply tackles the problem of providing (approximate) per-VC service guarantees for TCP over ATM. This study aims to provide some further insights on the interaction between aggregate TCP traffic and DFBA.

## 3 Modeling Assumptions

### 3.1 TCP Reno

The basis of the approach taken in [9] is, as a first step, to observe the system at time instants at which the traffic sources change their output rate. At this time scale, the nature of TCP traffic is solely determined by flow control mechanisms. Disregarding lower time scales reduces the state space considerably, since we do not need to keep track of the phases of the individual sources. As a second step, one may investigate the impact of the correlation structure at lower time scales on system performance (not included in this work).

For TCP sources, the time scale of interest is defined by the round-trip-time they experience. As in [25], we assume the round-trip-time experienced by all TCP connections to be of fixed and equal value $\mathrm{RTT}$. This value includes all transmission and queueing delays of the forward and return paths.

Since both TCP and GFR are defined at the packet level (AAL5 frames), no information is lost when disregarding the ATM cell level. We assume that TCP data packets are of constant size and equal to the TCP Maximum Segment Size (MSS). Each TCP packet is mapped directly onto one AAL5 CS Protocol Data Unit. We define the duration of a time slot as the necessary time to transmit one TCP packet over the GFR link. The GFR Peak Cell Rate is set equal to the GFR service rate, $\mathrm{LINK}$, and is assumed of constant value (not high-priority traffic), which fixes the duration of the time slot. All time measures will be expressed in slots. For instance, the system will be observed every $\mathrm{RTT}$ time slots.

Acknowledgments packets (acks) are assumed to be of negligible size, to consume no bandwidth and to suffer no losses. Acks are sent back to the sender $\mathrm{RTT}$ time slots after they have been served by the GFR queue. Thus the data flow considered is unidirectional.

All connections are assumed long-lived, that is, persistent and in the steady-state. All TCP sources are assumed greedy in the sense that they always have data to send. Theses assumptions generally hold for large data transfer (e.g. FTP or persistent HTTP connections), but do not necessarily hold for current

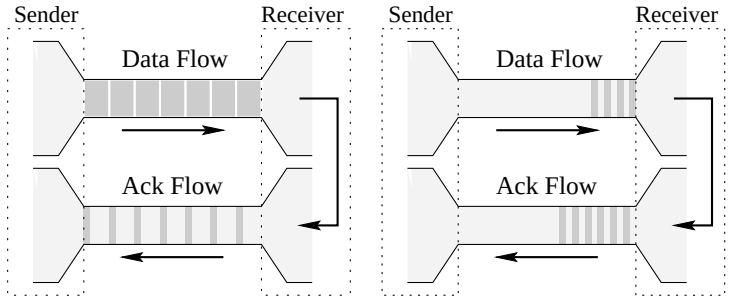



Figure 1: Self-clocking (links) vs. congestion avoidance (right).

web traffic. All sources are also assumed in congestion avoidance. More precisely, the relationship

$$\mathrm{MSS} \cdot \mathrm{cnwd} \ll \mathrm{RTT} \cdot \mathrm{LINK} \tag{2}$$

must hold. As shown in [25], this situation is typical of internet connections, where delays are large, losses are substantial and the average congestion window is small with respect to the bandwidth-delay product. Under this assumption, the TCP source emits close to an entire window of packets every round-trip-time. This behavior contrasts sharply with TCP's self-clocking property which is exhibited if, and only if, the average congestion window size is close to the bandwidth-delay product [18]. Figure 1 illustrates those two ranges of operation. Under Equation 2, the TCP traffic exhibits substantial correlation, that is that it tends to be emitted and received in bursts. Our model aims specifically at the operating range expressed by Equation 2.

The newest TCP implementations include various mechanisms to prevent timeouts as they occurred in TCP Tahoe. In TCP Reno, timeouts are either caused by a sudden variation in the measured RTT or by multiple losses within a single flight of packets [1]. Assuming a fixed RTT forbids the former, and the latter is easily solved. In absence of excessive network delays, the only

source of timeouts is a failure of the Fast Retransmit algorithm. Fast Retransmit is known to perform well under the assumption of independent losses [17], as the probability of multiple losses within a single flight is very small. As in RED, DFBA's drop scheme is probabilistic and achieves a certain level of independence. In our simulations with TCP Reno, timeouts typically accounted for around from 10 to 15 percent of the total number of losses [6]. TCP NewReno and SACK (see [14] and [22]) are expected to perform better [12]. In this study, we assume TCP timeouts to occur rarely and neglect them. This also frees us from modeling the complex adaptation behavior of the TCP retransmission timer (RTO).

Having assumed persistent connections and neglected timeouts, we do not need to model the TCP Slow Start phase. We assume the TCP source is always above the slow start threshold and neglect its adaptation. We also neglect the complex mechanics of the Fast Recovery algorithm, as they have little effect on the congestion window distribution [23] and on the achieved throughput [25]. In our model, a Fast Retransmit is immediately followed by the congestion avoidance phase.

### 3.2 Differential Fair Buffer Allocation

Recall that DFBA involves two types of thresholds: a set of static thresholds LBO and HBO, enforcing a low average total buffer occupancy, and a dynamic threshold $X \cdot W_i/W$, estimating the buffer's fair share belonging to VC $i$. Dynamic thresholding undoubtedly is an important feature of DFBA, as it allows per-VC guarantees to be provided. However, it is hardly feasible to track more than one buffer occupancy without jeopardizing our objective of modeling aggregate TCP traffic. Consequently, we choose to disregard dynamic thresholding and target the model at capturing the behavior of a single GFR VC. The fairness of DFBA was evaluated by means of simulation in a previous paper [6].

DFBA signals congestion to the TCP sources using probabilistic drops. Under the above assumptions, the packet drop probability $p$ increases linearly from 0 to $Z_i$ as the total buffer occupancy varies from LBO to HBO. We also assume that DFBA does not discard the packets until they are served, that is, dropped packets are nevertheless accounted for in the queue statistics. Assuming the traffic reactive, the overall loss probability is small. As this choice is also consistent with the measurements made in [16, 6], this assumption should have little impact on the accuracy of the results. We shall refer to this assumption as the *virtual losses* assumption. Finally, we assume that the probability of overflow of the DFBA queue is very close to zero. This allows us to disregard loss scenarios resulting from such overflow. This assumption holds as long as traffic is flow controlled traffic and the network is properly dimensioned.

## 4 A Matrix-Geometric Model for DFBA

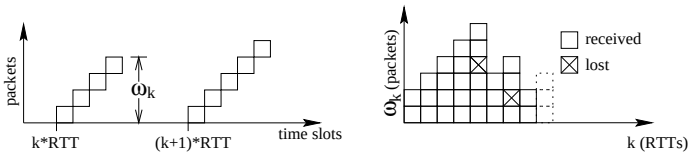
### 4.1 Modeling the TCP congestion window

According to the assumptions made in the previous section, the state of a TCP connection is fully described by the discrete-time discrete-value random variable $\{\omega_k \mid 0 \le k \le \infty\}$ representing the size of the congestion window at observation period $k$, that is the time slots $k \cdot \mathrm{RTT} \le t < (k+1) \cdot \mathrm{RTT}$. Each instance of $\omega_k$ draws its value from the finite set of states $\Omega := \{\omega \mid 1 \le \omega \le \omega_{\max}\}$, where the constant $\omega_{\max}$ represents the smallest of either the receiver or the sender maximum window size. In case all packets sent during period $k$ have been acknowledged, $\omega_{k+1}$ follows

$$\begin{aligned} \omega_{k+1} &:= \omega_k + 1 \quad \text{for } \omega_k < \omega_{\max} \\ &:= \omega_{\max}, \quad \text{otherwise}, \end{aligned} \tag{3}$$

which corresponds to the linear growth of the congestion window. In the case of at least one loss during period $k$, the window size is halved, that is

$$\begin{aligned} \omega_{k+1} &:= \lfloor \omega_k/2 \rfloor \quad \text{for } \omega_k > 1 \\ &:= 1 \quad \text{otherwise}, \end{aligned} \tag{4}$$




(a) Observation Period (b) Window Evolution

Figure 2: TCP Congestion Window Model

where the brackets stand for the floor operation. Figure 2a shows how packets are assumed to be sent in a batch at the beginning of an observation period. Figure 2b illustrates the two possible evolutions of the window size.

In order to derive the corresponding transition matrix, we need to introduce the loss function

$$\lambda(l,m,p) = \left( \begin{array}{c} m \\ l \end{array} \right) (1-p)^{m-l} p^l \,,$$

which describes the probability of independently loosing $l$ packets within a window of size $m$, given a loss probability $p$. We are looking for a set of transition matrices $\{\mathbf{C}_m(p), m \geq 0\}$ of size $\omega_{\max} \times \omega_{\max}$ which correspond to the state transitions causing $m$ packets to be accepted in the GFR queue during an observation period. Clearly, the matrix $\mathbf{C}(p) = \sum \mathbf{C}_m(p)$ is a stochastic transition matrix ruling the behavior of $\omega_k$. The transition probabilities solely depend on the loss probability, a feature which we shall exploit later. Under the virtual losses assumption, the transition matrix is expressed by

$$\begin{array}{rcll} \langle \mathbf{C}_m(p) \rangle_{ij} = & \lambda(0,m,p) & \text{for } i=m,\, j=m+1,\, 1 \leq m < \omega_{\max} \\ = & \lambda(0,\omega_{\max},p) & \text{for } i=j=\omega_{\max}=m \\ = & 1-\lambda(0,m,p) & \text{for } i=m,\, j=\lfloor i/2 \rfloor,\, 1 \leq m \leq \omega_{\max} \\ = & 0 & \text{otherwise}\,. \end{array} \tag{5}$$

Each instance of $\mathbf{C}_m(p)$ contains only two nonzero elements which suggests that sparse matrices can be used to reduce computational complexity.

Despite its apparent simplicity, this model hardly supports the multiplexing of more than a few sources, because the number of states required, $(\omega_{\max})^n$, grows exponentially with the number of sources multiplexed $n$.

## 4.2 Approximate Source Model

We subdivide the state space $\Omega$ into two disjoint subsets, $\Omega_0$ and $\Omega_1$, such that $\Omega_0 := \{\omega \mid 1 \leq \omega \leq \omega_{\max/2}\}$ and $\Omega_1 := \{\omega \mid \omega_{\max/2}+1 \leq \omega \leq \omega_{\max}\}$, and define a random variable $\hat{\omega}_k$, such that the process $\omega_k$ is mapped onto a two-node semi-Markov chain (arbitrary sojourn times). An example for $\omega_{\max} = 6$ is presented in Figure 3. Formally, the mapping follows

$$\begin{array}{rll} \hat{\omega}_k & := 0, & \omega_k \in \Omega_0 \\ & := 1, & \omega_k \in \Omega_1\,, \end{array} \tag{6}$$

which constitutes the process we want to approximate. If we inspect $\hat{\omega}_k$, we observe sojourn times that are not strictly geometrically distributed, and that can therefore hardly be captured by a regular Markov chain (geometrically distributed sojourn times). Nevertheless, we assume that the sojourn times do not differ substantially from the geometrical distribution, and attempt to map $\hat{\omega}_k$ onto a two-node Markov chain, as illustrated in Figure 4a. We need to derive expressions for the inter-sets transition probabilities $p_0$ and $p_1$, and for the per-state expected sending rates $c_0$ and $c_1$, such that the model behaves as

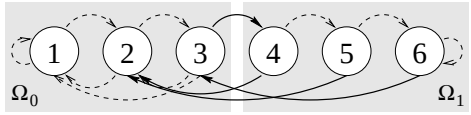



Figure 3: Mapping a six-node MC onto a two-node MC

closely as possible to $\hat{\omega}_k$. First, we use heuristic to obtain a simple approximation. Then we justify our choice through an first-passage characterization of the process $\hat{\omega}_k$.

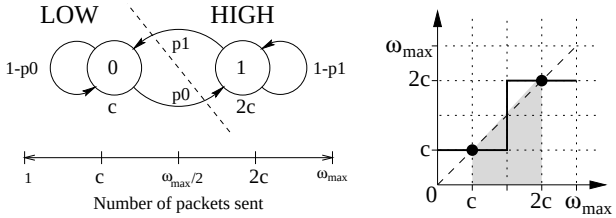
#### 4.2.1 Heuristic-based Approximation

To obtain the per-state sending rates, we perform the binary quantization of the congestion window illustrated in Figure 4b. We set the quantization step to $c = \omega_{\max}/4$, and fix to $c_0 = c$ the rate at which $\hat{\omega}_k$ emits packets if in state 0, and to $c_1 = 2c$ if in state 1.

Looking at Figure 3, we see that $p_1$ must account for all possible $\Omega_1 \rightarrow \Omega_0$ transitions, whereas $p_0$ depends on a single path. We define $p_1$ as the probability of having at least one independent loss in a window size of $c_1$ packets, that is

$$p_1 := 1 - (1-p)^{2c}\,, \tag{7}$$

assuming independent losses of constant probability of $p$. This behavior is consistent with the one of a TCP source, since the rate of the source is halved when a loss occurs. To derive $p_0$ we must account for the fact that the growth from a window size of $c_0$ to $c_1$ is linear. Again, under the independent loss assumption, the probability of transmitting consecutively without losses $c$ rounds of $c, c+1, ..., 2c-1$ packets respectively, is $(1-p)^{3(c^2-c)/2}$ (i.e. the shaded area in Figure 4b ). We account for the linear growth, as well as for the fact that the probability needs to be applied every round (and not every $c$ rounds) by setting




(a) Underlying Markov Chain (b) Rate Quantization

Figure 4: Approximate model for TCP's reactive behavior

$$p_0 := \frac{(1-p)^{3(c^2-c)/2}}{c}. \tag{8}$$

#### 4.2.2 Stationary Probabilities and First Passage Times

Let $\boldsymbol{\pi} := (\boldsymbol{\pi}_0, \boldsymbol{\pi}_1)$, where $\boldsymbol{\pi}_0 = (\pi_1, \pi_2, ..., \pi_{\omega_{\max/2}})$ and $\boldsymbol{\pi}_1 = (\pi_{\omega_{\max/2+1}}, ..., \pi_{\omega_{\max}})$, be the stationary probability vector corresponding to the transition matrix $\mathbf{C}(p)$, that is, it fulfills $\boldsymbol{\pi}\mathbf{C}(p) = \boldsymbol{\pi}$.

From the steady-state vector, we can compute the conditional expectations $\mathrm{E}(\omega \mid \omega \in \Omega_0)$ and $\mathrm{E}(\omega \mid \omega \in \Omega_1)$, from which we can assert whether our assumptions with regards to the rate quantization hold. The upper part of Figure 5 shows them as a function of the loss probability $p$. The bottom part plots the ratio $\mathrm{E}(\omega \mid \omega \in \Omega_1)/\mathrm{E}(\omega \mid \omega \in \Omega_0)$, which is shown to equal 2 for a wide range of loss probabilities.

A formal way to obtain $p_0$ and $p_1$ is to characterize the sojourn times of the semi-Markov chain $\hat{\omega}_k$. Since the sets $\Omega_0$ and $\Omega_1$ are disjoint, this problem can be treated as a first passage problem. Let $i$ be a recurrent state and $j$ be

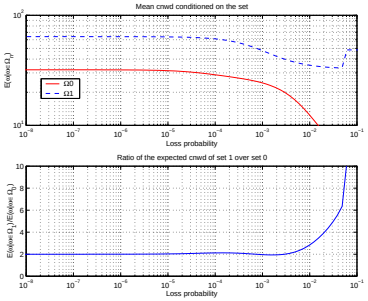



Figure 5: Per-set Mean Congestion Window vs. Loss Prob.

absorbing, we define $\tau_{ij} = \langle \boldsymbol{\tau} \rangle_{ij}$ as the time to absorption in state $j$, conditioned on $\omega_k$ starting in state $i$. We subdivide the transition matrix according to $\Omega_0$ and $\Omega_1$, that is

$$\mathbf{C}(p) := \begin{pmatrix} \boldsymbol{\delta}_{00} & \boldsymbol{\delta}_{01} \\ \boldsymbol{\delta}_{10} & \boldsymbol{\delta}_{11} \end{pmatrix}$$

and define an associated transient process $\boldsymbol{\Delta}_0$, such that all states belonging to $\Omega_1$ are absorbing, whereas those from $\Omega_0$ are maintained recurrent. In mathematical terms, this means that

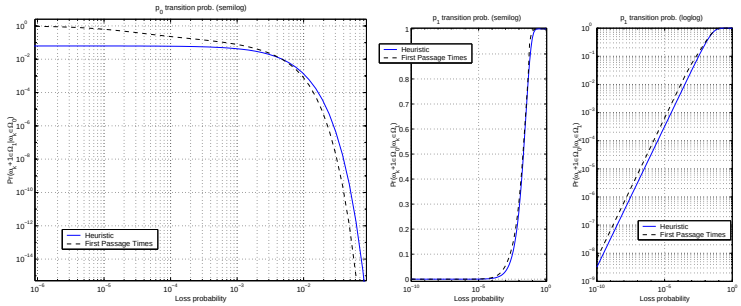
$$\boldsymbol{\Delta}_0 := \begin{pmatrix} \boldsymbol{\delta}_{00} & \boldsymbol{\delta}_{01} \\ \mathbf{0} & \mathbf{1} \end{pmatrix},$$

and it can be shown that the corresponding mean absorption times follow

$$\mathrm{E}(\tau_{ij}) \quad = \quad (\mathbf{I} - \boldsymbol{\delta}_{00})^{-1}, \text{ for } i \in \Omega_0,\ j \in \Omega_1\,.$$

The metric $\mathrm{E}(\tau_{ij})$ represents the mean time spent in $\Omega_0$ before absorption occurs in $j$, conditioned on $\hat{\omega}_k$ starting in state $i$. Therefore, the expression

$$\hat{\tau}_0 = \boldsymbol{\pi}_0 (\mathbf{I} - \boldsymbol{\delta}_{00})^{-1} \mathbf{e} / \boldsymbol{\pi}_0 \mathbf{e}$$




(a) Candidates for $p_0$ (b) Candidates for $p_1$

Figure 6: Deriving the transition probabilities

is the mean time spent by $\hat{\omega}_k$ in state 0 before entering state 1 for the first time ($\mathbf{e}$ is a column vector with unit values). A similar development can be performed for $1 \to 0$ transitions and leads to

$$\hat{\tau}_1 = \boldsymbol{\pi}_1 (\mathbf{I} - \boldsymbol{\delta}_{11})^{-1} \mathbf{e} / \boldsymbol{\pi}_1 \mathbf{e}\,.$$

From the sojourn times, we can obtain a second approximate of the transition probabilities, that is $\hat{p}_0 = 1/\hat{\tau}_0$ and $\hat{p}_1 = 1/\hat{\tau}_1$ .

Figure 6 compares the heuristic-based $p_0$ and $p_1$ and the first passage-based $\hat{p}_0$ and $\hat{p}_1$. We see that the heuristic-based transition probabilities provide a good approximation of the first-passage metrics. They also have the advantage of being well-defined for high loss probabilities, whereas we can hardly inverse $\mathbf{C}$ for $p > 0.1$ ($\mathbf{C}(1.0)$ is non-recurrent). The probability $p_0$ was ceiled to $1/c$ as $p \to 0$, because the AS model tends to be greedier than the DS model. In the rest of this work, we shall use the heuristic-based approach when the AS model is employed.

## 4.3 Traffic Multiplexing

### 4.3.1 $N$ Approximate Sources

Let us first consider the multiplexing of $N$ approximate sources as defined in the previous section. The state of such a system at a given time instant is completely described by the random variable $\{n \mid 0 \leq n \leq N\}$, which counts the number of sources in state 1. According to our quantization scheme, the number of packets at the input of the queue per observation period follows $(n+N)c$.

The associated state transition matrix $\mathbf{D}(p)$ of size $(N+1) \times (N+1)$ can be obtained by defining two discrete random variables, $U(p)$ and $V(p)$, as the number of $0 \rightarrow 1$ resp. $1 \rightarrow 0$ transitions occurring during the observation period. The evolution of the variable is a Markov chain $\{n_k \mid k \geq 0\}$, where $n_{k+1} = n_k + U(p) - V(p)$ and

$$\begin{aligned} \Pr(U(p)=u) &= \binom{N-n}{u} p_0^u (1-p_0)^{N-n-u} \\ \Pr(V(p)=u) &= \binom{n}{u} p_1^u (1-p_1)^{n-u} . \end{aligned}$$

Assuming virtual losses, the matrix $\mathbf{D}(p)$ with elements $\langle \mathbf{D}(p) \rangle_{ij} = \Pr(n_{k+1} = j \mid n_k = i)$ follows

$$\langle \mathbf{D}(p) \rangle_{ij} = \sum_{u=\max(0,j-i)}^{\min(j,N-i)} \Pr(U(p)=u) \cdot \Pr(V(p)=u-j+i) , \quad (9)$$

for $0 \leq i,j \leq N+1$.

### 4.3.2 One Detailed Source + $N$ Approximate Sources

Now consider the problem of multiplexing one detailed source and $N$ approximate sources. In this case, the Markov chain is ruled by $\{(\omega,n) \mid 1 \leq \omega \leq \omega_{\max}, 0 \leq n \leq N\}$. The two random variables are mutually independent, and the transition probabilities are expressed by

$$\begin{aligned} \Pr(\omega_{k+1}=j, n_{k+1}=l \mid \omega_{k+1}=i, n_{k+1}=k) &= \Pr(\omega_{k+1}=j \mid \omega_{k+1}=i) \times \\ &\quad \Pr(n_{k+1}=l \mid n_{k+1}=k) , \end{aligned}$$

where the probabilities right-hand terms were defined in Equations 5 and 9. The corresponding transition matrix is expressed by

$$\mathbf{E}(p) = \mathbf{C}(p) \otimes \mathbf{D}(p) ,$$

where the binary operator $\otimes$ denotes the Kronecker product of two square matrices.

## 4.4 Buffer Analysis

In this section we perform the buffer analysis for the two configurations described in the previous section. We assume a finite queue of size $Q^*$ ruled by an acceptance mechanism which depends solely on the instantaneous buffer occupancy.

### 4.4.1 $N$ Approximate Sources

Assuming a service rate of 1 packet per time slot, the variation in the queue index is expressed by the Markov chain $\{m_k \mid k \geq 0\}$ such that

$$m_k = (n_k + N) - \mathrm{RTT}/c$$

and is bound by $S \leq m \leq T$, where $S = N - \mathrm{RTT}/c$ and $T = 2N - \mathrm{RTT}/c$. Since the input rate is quantized in steps of $c$, we quantize the queue accordingly and select a value of RTT such that it is a multiple of $c$. The queue occupancy at the end of round $k$ follows the recursive equation $\{q_k \mid k \geq 0\}$ such that

$$q_k = \min(Q, \max(0, q_{k-1} + m_k)) .$$

The parameter $Q$ is the maximum buffer index and is set to $Q^*/c$. We define $\mathbf{D}_m(p)$ as the transition matrix describing the source transitions leading to a variation of $m$ in queue size (analogous to Equation 5). Negative $m$ values are interpreted as decreases in queue size during the observation period. The buffer transition matrix $\mathbf{P}$ is of the form

$$\mathbf{P} = \begin{pmatrix} \sum_{m=S}^{0} \mathbf{D}_m^{(0)} & \mathbf{D}_1^{(0)} & \cdots & \mathbf{D}_T^{(0)} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots \\ \sum_{m=S}^{-1} \mathbf{D}_m^{(1)} & \mathbf{D}_0^{(1)} & \cdots & \mathbf{D}_{T-1}^{(1)} & \mathbf{D}_T^{(1)} & \cdots & \mathbf{0} & \mathbf{0} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ \mathbf{D}_S^{(-S)} & \mathbf{D}_{S+1}^{(-S)} & \cdots & \mathbf{D}_{S+T}^{(-S)} & \mathbf{D}_{S+T+1}^{(-S)} & \cdots & \mathbf{D}_T^{(-S)} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{D}_S^{(-S+1)} & \cdots & \mathbf{D}_{S+T-1}^{(-S+1)} & \mathbf{D}_{S+T}^{(-S+1)} & \cdots & \mathbf{D}_{T-1}^{(-S+1)} & \mathbf{D}_T^{(-S+1)} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \end{pmatrix} \tag{10}$$

for the upper part and

$$\mathbf{P} = \begin{pmatrix} \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \cdots & \mathbf{D}_S^{(Q-T)} & \mathbf{D}_{S+1}^{(Q-T)} & \cdots & \mathbf{D}_{S+T}^{(Q-T)} & \mathbf{D}_{S+T+1}^{(Q-T)} & \cdots & \mathbf{D}_T^{(Q-T)} & \mathbf{0} \\ \cdots & \mathbf{0} & \mathbf{D}_S^{(Q-T+1)} & \cdots & \mathbf{D}_{S+T-1}^{(Q-T+1)} & \mathbf{D}_{S+T}^{(Q-T+1)} & \cdots & \mathbf{D}_{T-1}^{(Q-T+1)} & \mathbf{D}_T^{(Q-T+1)} \\ \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_S^{(Q-1)} & \mathbf{D}_{S+1}^{(Q-1)} & \cdots & \mathbf{D}_0^{(Q-1)} & \sum_{m=1}^{T} \mathbf{D}_m^{(Q-1)} \\ \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{D}_S^{(Q)} & \cdots & \mathbf{D}_{-1}^{(Q)} & \sum_{m=0}^{T} \mathbf{D}_m^{(Q)} \end{pmatrix} \tag{11}$$

for the lower part. Since we assume that sources are reactive and that the network is properly dimensioned, Equation 11 only accounts for the losses enforced by DFBA and disregards the scenario of buffer overflow. The set of matrices $\{\mathbf{D}_m^{(q)} \mid S \leq m \leq T, 0 \leq q \leq Q\}$ is obtained by setting $\mathbf{D}_m^{(q)} = \mathbf{D}_m(p)$ where the loss probability follows

$$\begin{array}{rcll} p := & 0 & \text{for } 0 \leq qc < \text{LBO} \\ := & \frac{qc - \text{LBO}}{\text{HBO} - \text{LBO}} Z & \text{for LBO} \leq qc \leq \text{HBO} \\ := & Z & \text{for HBO} < qc \leq Q^* \,. \end{array} \tag{12}$$

The constant LBO and HBO are the DFBA lower and upper thresholds. The parameter $Z$ defines the maximum drop probability enforceable by the acceptance scheme. The matrix defined by Equations 10 and 11 is highly structured and allow for matrix-geometric solutions [24]. For instance, if we set $\mu = |S| = |T|$ and group the sub-matrices appropriately, a level-dependent quasi-birth-and-death (LDQBD) process is obtained, that is

$$\mathbf{P} = \begin{pmatrix} \mathbf{A}^{(0)} & \mathbf{\Lambda}^{(0)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{M}^{(1)} & \mathbf{A}^{(1)} & \mathbf{\Lambda}^{(1)} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^{(2)} & \mathbf{A}^{(2)} & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \mathbf{\Lambda}^{(Q/\mu-1)} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{M}^{(Q/\mu)} & \mathbf{A}^{(Q/\mu)} \end{pmatrix},$$

for which a simple solution of the steady-state solution exists [15].

#### 4.4.2 One Detailed Source + $N$ Approximate Sources

To study the interaction between the DS and the AS models, one can include a detailed source in the analysis. In such case, the variations in queue occupancy follow $\{m_k \mid k \geq 0\}$ such that

$$m_k = \lfloor (\omega_k + (n_k + N)c - \text{RTT})/d + 0.5 \rfloor$$

where $d$ is a quantization factor introduced to limit the size of the LDQBD matrices. A development similar to the previous section also leads to a LDQBD process.

### 4.5 Delayed Acknowledgements

The DS model can easily be modified to account for delayed acknowledgements (one ack for $b$ packets received). Equation 3 can be redefined such that

$$\begin{aligned} \Pr(\omega_{k+1} = \omega_k + 1) &= \lambda(0, \omega_k, p)/b \\ \Pr(\omega_{k+1} = \omega_k) &= \lambda(0, \omega_k, p) \cdot (1 - 1/b) \end{aligned}$$

for $\omega_k < \omega_{\max}$. The corresponding expected increment is $1/b$ for rounds where no losses have occurred. In the case of the AS model, $0 \rightarrow 1$ transition probability is simply set to

$$p_0 = \frac{(1-p)^{3(c^2-c)/2}}{cb}.$$

### 4.6 Performance Metrics

Many performance measures can be obtained from the current model. Let $f_a(x)$ and $F_a(x)$ be the steady-state probability mass and probability distribution functions of an arbitrary random variable $a$. The DS model allows us to look at the congestion window distribution $F_\omega(x)$ or at the mean throughput achieved $\bar{\omega} = \mathrm{E}(\omega)$ packets per RTT. The AS model is mainly limited to the mean throughput achieved by a single source, $\tilde{\omega} = (\mathrm{E}(n) + N)c/N$ packets per RTT. At the queue side, we can either look at the buffer occupancy complementary distribution function (CDF) $G_q(x) = 1 - F_q(x)$ or at the mean queue length $\bar{q} = \mathrm{E}(q)$ packets and the associated variance $\sigma_q^2 = \mathrm{Var}(q)$. To facilitate comparison, we shall plot the standard deviation $\sigma_q$ along with $\bar{q}$. From $F_q(x)$ and from the mapping expressed in Equation 12 we can also compute the average loss probability $\bar{p}$.

## 5 Performance Results

In this section we provide some results obtained from the proposed model and discuss their impact on network performance. Firstly, we investigate the interaction between the detailed source and the approximate source models when their traffic is multiplexed in the same queue. Section 5.3 shortly compares the results to simulation. Section 5.4 evaluates the impact of the parameters LBO and $Z$ on the queue and congestion window distributions. Section 5.5 investigates the impact of delayed acknowledgements on TCP throughput. Finally, the last section assesses the impact of traffic aggregation on queue statistics.

### 5.1 Model Parameters

The maximum congestion window size $\omega_{\max}$ was set large enough to allow the source to use its share of the available bandwidth, but not necessarily to the bandwidth-delay product. A smaller $\omega_{\max}$ reduces the sizes of the QBD matrices and has little effects on the results as long as the probability of reaching a congestion window of $\omega_{\max}$ is small.

The default value for $c$ is $\omega_{\max}/4$ but we varied it to assess the sensitivity of the model with respect to that parameter. The parameter $\mathrm{RTT}$ was set to keep the system symmetric with respect to the queue increments and decrements. This constraint is not required to obtain a QBD but the structure of the LDQBD matrices becomes hard to track if we let the difference $T - S$ vary. Setting $T = -S$ fixes the structure of $\mathbf{M}^{(r)}$, $\mathbf{A}^{(r)}$ and $\mathbf{\Lambda}^{(r)}$. This constraint has little influence on the results, because the sources are flow controlled, which causes large queue decrements or increments to have a small probability. Finally, we varied the quantization grain $d$ such that the maximum queue increment $T$ is fixed. This fixes the sizes of the matrices $\mathbf{M}^{(r)}$, $\mathbf{A}^{(r)}$ and $\mathbf{\Lambda}^{(r)}$.

### 5.2 Interactions between the DS and the AS models

Table 1 presents throughput results for various configurations. The AS model appears to fairly share the bandwidth with the DS model. Note also the relative insensitivity of the DS model to the $c$ parameter.

| | $\omega_{\max}$ | $N$ | $c$ | $d$ | $Q$ | LBO | HBO | RTT | $\bar{\omega}$ | $\tilde{\omega}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| a | 60 | 1 | 20 | 8 | 1000 | 500 | 900 | 60 | 29.90 | 29.46 |
| b | 60 | 2 | 20 | 10 | 1000 | 500 | 900 | 90 | 29.57 | 29.52 |
| c | 60 | 2 | 15 | 10 | 500 | 250 | 450 | 75 | 25.33 | 23.35 |
| d | 60 | 2 | 25 | 11 | 500 | 250 | 450 | 105 | 31.39 | 35.41 |
| e | 90 | 3 | 30 | 18 | 500 | 250 | 450 | 190 | 36.61 | 41.66 |

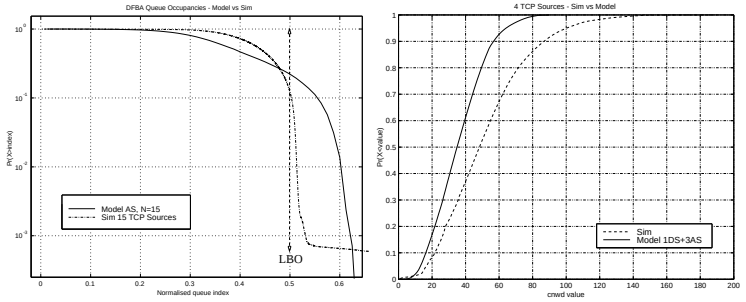
Table 1: Multiplexing the foreground and background source models

## 5.3 Analytic vs. simulation results

This section is an attempt to validate the model using simulation. Figure 7a compares two queue occupancy CDFs in which the queue occupancy is normalized with respect to the queue size. The straight line plots $G_q(x)$ for the multiplexing of 15 ASs. The dashed-dotted line is the result of a simulation obtained for 15 TCP Reno sources multiplexed into a DFBA buffer. The simulation setup employed is described in detail in [6] and [11]. 15 VCs have been setup, each of them VC carrying the traffic of a single TCP source. The parameter $\alpha$ was set to 0.0 to limit the impact of the fairness mechanisms of DFBA. .

The analytical and simulation results qualitatively have the same behavior, but the model is far from being exact. Both predict an exponential decay of $G_q(x)$ for indices larger than LBO. Below LBO, the Reno sources are greedier than the analytical model. Above LBO, the model is not as reactive to packet drops as TCP Reno.

Figure 7b compares the congestion window distribution obtained from configuration e in Table 1 to a simulation involving four TCP sources multiplexed in a GFR queue. Again, the curves have a similar outlook but the analytical model can hardly be used to predict the congestion window of a single TCP source. The main obstacle resides in the fact that the model holds for TCP sources in congestion avoidance. To obtain such state with only four sources sharing a high-speed link, the GFR buffer must be under-dimensioned (we



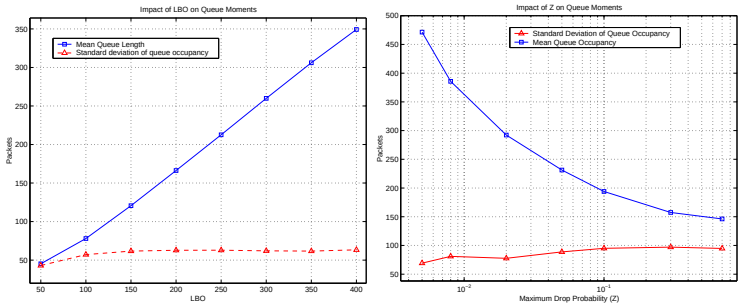

(a) Queue CDF (b) Congestion Window Distribution

Figure 7: Four TCP sources - Model vs. simulations

used 2000 cells in this simulation). In that range of operation, the congestion window distribution strongly depends on the queue size, and the gap observed between the model and the simulation has little meaning. Nevertheless, the figure shows that the DS model not only captures the qualitative behavior of the congestion window, but also that the assumptions made for the AS model have little impact on the steady-state behavior of the DS source.

## 5.4 Impact of LBO and $Z$

The LBO and $Z$ parameters provide control over the average queue length (see [16, Chapter 6] and [6]). Figure 8a shows the mean queue length and the standard deviation of the buffer occupancy as a function of LBO. Clearly, the average queue length is a linear function of LBO. In addition, the variance does seem to depend on LBO. In terms of throughput, static thresholds have virtually no impact on the congestion window distribution (not shown). This holds as long as the buffer is well dimensioned [16, Section 4.8].




(a) Impact of LBO on Queue Moments (b) Impact of Z on Queue Moments

Figure 8: Impact of Drop Probability Function on Queue Distribution

We know that LBO can be varied without affecting the throughput. So what is constraining its dimensioning? The threshold LBO defines a portion of the buffer whose access is completely unrestricted. The larger its value, the burstier the traffic generated by TCP. This is an important issue, given that the mean queue length strongly depends on the level of aggregation (see Section 5.6). We also know that $G_q(x)$ decays exponentially above LBO. A small LBO with respect to the buffer size keeps the sources under tight control, but makes little use of the GFR buffer. Conversely, setting LBO too close to HBO may prevent DFBA to control the average queue length efficiently. In case the sources are non-cooperative, the results may be catastrophic. The area between LBO and HBO must be large enough to accommodate sources ranging from the tamest to the greediest behavior. A source of variability in terms of greediness is the use of delayed acknowledgements (see Section 5.5).

Another way to control the average queue is through the maximum drop probability $Z$. Firstly, we found that the parameter $Z$ has marginal impact on the congestion window distribution (not shown). Figure 8b shows that it does not influence the the variance of the buffer occupancy. In contrast, the mean queue occupancy varies non-linearly with $Z$. The main difference between the use of $Z$ or LBO to control the average queue length is that, with $Z$, the size of the unrestricted portion of the buffer is preserved. This is advantageous since increasing LBO reduces protection against non-collaborative sources.

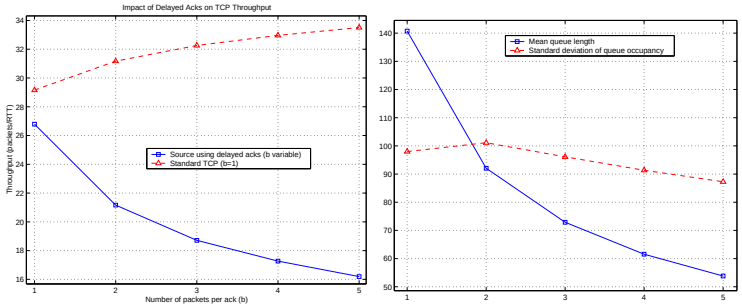
## 5.5 Impact of Delayed Acknowledgements

To evaluate the impact of delayed acknowledgements, we consider two configurations: a GFR queue fed by three sources (one DS and two ASs) or by 15 sources. In the first configuration, the detailed source uses delayed acknowledgements, whereas the approximate sources do not. Figure 9a plots the achieved throughput for each source type as a function of the number of acks per packet $b$. Clearly, TCP connections using delayed acks are disfavored by DFBA.

Delayed acks also have significant impact on the queue statistics. Figure 9b shows the mean queue length and the standard deviation for a buffer fed by 15 approximate sources. We see that delayed acks can reduce queueing delays, but that they have a limited impact on delay variations.

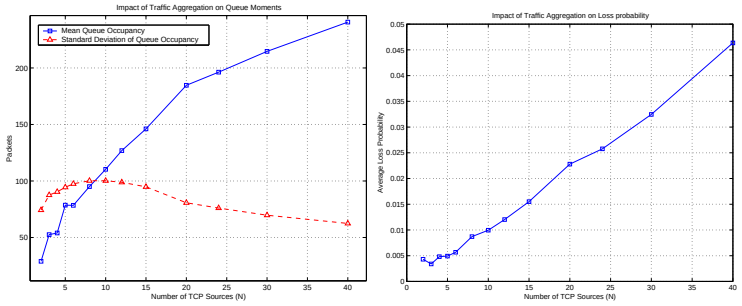
## 5.6 Impact of aggregation level

In this experiment, we fix the product $Nc$ and assess the influence of increased traffic aggregation on the queue and throughput statistics. Figure 10a plots on the same scale the mean and the standard deviation of the queue occupancy as a function of the number of sources $N$. We see that if, on one hand, the variance or equivalently, the delay variation, seems to go down as the number of sources increases (i.e. a multiplexing gain), the buffer requirements themselves increase substantially. From Figure 10a, we would expect the mean queue length to reach a maximum for a certain aggregation level, but this level




(a) On Throughput (b) On Queue Statistics

Figure 9: Impact of Delayed Acknowledgements




(a) On Queue Moments (b) On Loss Probability

Figure 10: Impact of traffic aggregation

is beyond the reach of our analysis. Also, there appears to be a certain number of sources for which the variance is maximum.

Figure 10b presents the mean loss probability $\bar{p}$ as a function of the number of sources. In this case, the mean loss probability reaches 5% for 40 sources, while the slope of the curve does not yet show any signs of falling down. This raises serious questions with respect to DFBA's ability to handle the massive amount of TCP traffic carried over the Internet. Despite the probabilistic nature of the drops beyond a certain average loss probability, real TCP sources would start to timeout and the overall throughput would sink dramatically.

## 6 Conclusion

In this paper, we developed an analytical model for assessing the performance of TCP over DFBA. We have shown that the reactive behavior of a single TCP source in congestion avoidance can be approximated a by two-node Markov chain. Simulation results also showed the behavior of the proposed model to be consistent with TCP Reno.

In terms of performance results, we quantified the relationship between the parameters LBO and $Z$ and the steady-state buffer occupancy measures. We also quantified the influence of delayed acks on the switch performance, and showed that sources using delayed acks are disfavored by the packet dropping mechanism.

We have also assessed the impact of traffic aggregation on queueing performance. Our results show that the average queue length strongly depends on the number of sources. They predict that the mean queue length should reach a maximum at a certain aggregation level, but also that the average loss probability could grow without bounds with the number of TCP connections. This raises the question whether the DFBA proposal can effectively support the large number of simultaneous connections carried on Internet backbones.

## 7 Acknowledgements

The authors would like to thank Fernando Cerdán for lending his ATM simulator, and Anne Goldbeck for proofreading the manuscript.

## References

[1] Allman, M., V. Paxson and W. Stevens, "TCP Congestion Control", *RFC 2581*, April 1999.

[2] Andrikopoulos, I., A. Liakopoulos, G. Pavlou and Z. Sun, "Providing Rate Guarantees for Internet Application Traffic Across ATM Networks", *IEEE Communications Surveys*, vol. 2 no. 3, `http://www.comsoc.org/pub/surveys`.

[3] ATM Forum, "Traffic Management Specification Version 4.1", *AF-TM-0121.000*, March 1999.

[4] Bonaventure, Olivier, "A Simulation Study of TCP with the GFR Service Category", In *High Performance Networking for Multimedia Applications*, Kluwer Academic Publishers, 1998.

[5] Bonaventure, Olivier, "Integration of ATM under TCP/IP to provide services with guaranteed minimum bandwidth", Ph.D. Thesis, Université de Liège, March 1999.

[6] Bonin, Vincent, Fernando Cerdán and Olga Casals, "A Simulation Study of Differential Fair Buffer Allocation", Joint IEEE ATM Workshop 2000 and 3rd International Conference on ATM, Heidelberg, Germany, June 2000.

[7] Bonin, Vincent, "Performance Modeling of Differential Fair Buffer Allocation", M.Sc. Thesis, Technische Universität Hamburg-Harburg, February 2000.

[8] Bright, L. P. Taylor, "Calculating the equilibrium distribution in level-dependent quasi-birth-and-death processes", *Stochastic Models*, 11(3):497-525, 1995.

[9] Blondia, C. , O. Casals and B. Van Houdt, "Buffer and Throughput Analysis of the Explicit Rate Congestion Control Mechanism for the ABR Service Category in ATM Networks", Proceedings of PICS'98, Lund.

[10] Casetti, C. and M. Meo, "A New Approach to Model the Stationary Behavior of TCP Connections", INFOCOM'2000, Tel Aviv, Israel, March 2000.

[11] Cerdán, Fernando, O. Casals, "Performance of Different TCP Implementations over the GFR Service Category". ICON, January 2000.

[12] Fall, K. and S. Floyd, "A Simulation-Based Comparisons of Tahoe, Reno and SACK TCP", *Computer Communication Review*, vol. 26, no. 3, July 1996.

[13] Floyd, Sally, Van Jacobson, "Random Early Detection Gateways for Congestion Avoidance", IEEE/ACM Trans. on Networking, August 1993.

[14] Floyd, Sally, T. Henderson, "The NewReno Modification to TCP's Fast Recovery Algorithm", *RFC2582*, 1999.

[15] Gaver, D.P., P.A. Jacobs and G. Latouche, "Finite Birth-and-death Models in Randomly changin environments.", Adv. Appl. Prob. 16, 715-731, 1984.

[16] Goyal Rohit, "Traffic Management for TCP/IP over Asynchronous Transfer Mode Networks" PhD Dissertation, Department of Computer and Information Science, The Ohio State University, 1999. `http://www.cis.ohio-state.edu/~jain/papers/rohitphd.htm`.

[17] Hoe, Janey C., “Improving the Start-up Behavior of a Congestion Control Scheme for TCP”, *ACM SIGCOMM*, 1996.

[18] Jacobson, Van and Michael J. Karels, “Congestion Avoidance and Control”, *Computer Communication Review*, Vol. 18, No. 4, pp. 314-329, August 1998.

[19] Kumar, Anurag, “Comparative Performance Analysis of Versions of TCP in a Local Network with a Lossy Link”, IEEE/ACM Transactions on Networking, Vol. 6, No. 5, August 1998.

[20] Labrador, Miguel A. and Sujata Banerjee, “Packet Dropping Policies for ATM and IP Networks”, *IEEE Communications Surveys*, vol. 2 no. 3, `http://www.comsoc.org/pub/surveys`.

[21] Latouche, Guy, V. Ramaswami, “A Logarithmic Reduction Algorithm for Quasi-Birth-Death Processes”, J. Appl. Prob. 30, pp. 650-674, 1993.

[22] Mathis, M., J. Mahdavi, S. Floyd and A. Romanow, “TCP Selective Acknowledgement Options”, *RFC 2018*, October 1996.

[23] Misra, Archan, Teunis J. Ott, “The Window Distribution of Idealized TCP Congestion Avoidance with Variable Packet Loss”, IEEE INFOCOM’99.

[24] Neuts, Marcel F., *Matrix-Geometric Solutions in Stochastic Models*, John Hopins University Press, 1981.

[25] Padhye, J., V. Firoiu, D. Towsley and J. Kurose, “Modeling TCP throughput: A simple model and its empirical validation. Proceedings of SIGCOMM’98.

[26] Wuyts, K., B. Van Houdt, R.K. Boel, and C. Blondia, “Matrix Geometric Analysis of Discrete Time Queues with Batch Arrivals and Batch Departures with Applications to B-ISDN”, ITC-16, 1999.