Performance and Fluid Simulations of a Novel Shared Buffer Management System

Krishnan Kumaran Debasis Mitra
Bell Laboratories, Lucent Technologies
Murray Hill, NJ - 07974
kumaran,mitra@lucent.com

Abstract— **We consider a switching system which has multiple ports that share a common buffer, in which there is a FIFO logical queue for each port. Each port may support a large number of flows or connections, which are approximately homogeneous in their statistical characteristics, with common QoS requirements in cell loss and maximum delay. Heterogeneity may exist across ports. Our first contribution is a buffer management scheme based on Buffer Admission Control, which is integrated with Connection Admission Control at the switch, and is at the same time fair, efficient and robust in sharing the buffer resources across ports. Our scheme is based on the resource-sharing technique of Virtual Partitioning. Our second major contribution is to advance the practice of discrete-event fluid simulations. Such simulations are approximations to cell-level simulations and offer orders of magnitude speed-up. A third contribution of the paper is the formulation and solution of a problem of optimal allocation of bandwidth and buffers to each port having specific delay bounds, in a lossless multiplexing framework. Finally, we report on extensive simulation results. The scheme is found to be effective, efficient and robust.**

Keywords— **Buffer Management, Fluid Simulations, Virtual Partitioning.**

I. Introduction

This paper considers a model of a switching node shown in Figure 1, which has multiple(N) ports with bandwidths $C_1, C_2, ..., C_N$ that share a common buffer of size B. In the shared buffer there is a FIFO logical queue implemented as a linked-list for each port. Each port may support a large number of flows or connections, which we envision to be more or less homogeneous in their statistical characteristics, with heterogeneity existing across ports. This view is compatible with the QoS requirements, which is that there is a common delay bound on all connections through a port and that the bound on the probability of either violating the delay bound or losing cells is specific to each port. While shared buffer management has been of interest for some time, see for instance [IRL78], [FGO83], [HBO93], [CHA93], [GCG95], [WMA95], [CHA96], the approach taken here is rather different. Our work shares objectives with recent work on Virtual Queueing (see [SWR97] and references therein), in which per-VC behavior is emulated on a shared FIFO buffer. However, the algorithms are different and further work is necessary to make the linkages explicit.

This work has three main contributions. First, we propose a shared buffer management scheme based on Buffer Admission Control (BAC), which is integrated with Connection Admission Control (CAC) at the switch and is at the same time fair, efficient and robust in sharing the buffer resource across multiple ports. Recent work on CAC for single port systems [EMW95], [EMI97], [LZT97], [LZI97], [RRR97] has used crucially the idea of trading bandwidth and buffer to arrive at designs which allocate specific amounts of bandwidth and buffer per connection. If this approach is to be followed for multiport, shared buffer systems, it is imperative that CAC and BAC work in tandem. The dual objectives of satisfying QoS, which requires protecting and isolating connections, and extracting multiplexing gains from sharing the buffer resources, need to be balanced in shared memory architectures. With this in mind we base our scheme on Virtual Partitioning [MZI96], [BMI97], [MZI97], which is a technique for fair, efficient and robust resource sharing.

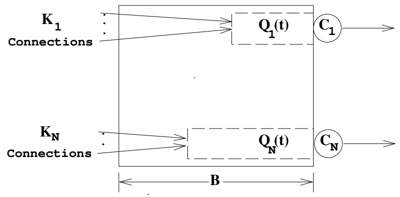


Fig. 1. A shared memory, multiport system with FIFO logical queues.

Our second main contribution is to advance the practice of discrete-event fluid simulations. While stochastic fluid models have come to be widely used in analysis because of their efficiency and natural correspondence to ATM cell switching and transport, fluid simulations have remained less familiar. This is in spite of the considerable speed advantage offered by fluid over cell-level simulations. The study of fluid simulations of sophisticated control schemes, such as the one in this paper, is important because the necessary techniques have been largely unexplored. An important feature of such fluid simulations is the collaboration with numerical calculations.

Our third main contribution is a new formulation and analysis of the problem of optimal allocation of bandwidth and buffers to ports, possibly with inhomogeneous QoS constraints in the system of Figure 1. This extends prior work in [EMW95], [LZT97] and [EMI97] on single-port systems.

Buffer Admission Control based on Virtual Partitioning has the following core ideas: First, at the time of design of CAC, each port is allocated a nominal buffer capacity, say

B_i for port i, where $\sum_i B_i \geq B$; at each point in time, ports that are using less than their nominal allocations, which may be called "underloaded", are accorded higher priority; conversely, ports that are exceeding their nominal allocations, which may be called "overloaded", are given lower priority; finally, the priority mechanism is implemented by a dynamic adaptation of the classical technique of trunk reservation [AKI84], [KEY90], [REI91], wherein cells for a port are admitted to the buffer only if the free capacity in the buffer exceeds a reservation parameter, which is specific to the port's loading status. The role of the reserved capacity is to protect underloaded ports and the dynamic reservation mechanism forces traffic to overloaded ports to back-off in the event that underloaded ports need to claim their allocated share. Note that the push-out mechanism is not explicitly used in this scheme for ease of implementation, but it may easily be added. Also note that the state information required to implement BAC for port i consists of just the current queue for port i and the current total buffer occupancy.

Traffic classes/ports are differentiated by QoS and also by the degree of isolation or protection required [CSZ93]. This is realized by the nominal allocations and the reservation parameters. For instance, a greedy "best effort" class/port will have relatively large reservation parameters working against it. Classical trunk reservation displays extraordinary robustness properties and, also, optimal reservation parameters scale slowly (eg. logarithmically or sublinearly) with the size of the problem [REI91], [MGH91]. We observe similar properties here. It suffices for the reserved capacities to be small and there are broad ranges of the reservation parameters that offer desirable performance. The former is key to the efficiency and the latter to the robustness of the scheme.

In a fluid simulation of our BAC the sample path of the total buffer occupancy $Q(t)$ exhibits persistence or "stickiness" at thresholds, which are determined by the nominal allocations and reservation parameters. These thresholds are points where the BAC is active and admits only fractional amounts of fluid generated by affected ports. The determination of these fractions is based here on a particular notion of fairness, wherein all affected ports are treated equally. Alternative notions of fairness may readily be substituted. These fractions, called "fairshare", are computed at the onset of a sticky period during simulations by solving an equation. We should point out that similar issues arise in Fair Queueing, Generalized Processor Sharing and Max-Min Fair ABR services, and the techniques for fluid simulation developed here are expected to extend to these applications.

The traffic source models considered in this work are adversarial to the extent permitted by Dual Leaky Bucket regulators, which are assumed to be typically present to police the sources. In the case of lossless multiplexing, the sources are permitted to collude so that in the worst case the sources synchronize to burst at their peak rates at the same time. In the contrasting, less conservative, statistical multiplexing approach, the sources are assumed to be non-collusive, so that their burst activity is asynchronous, i.e., their phases are independent, uniformly distributed random variables. In addition, we allow very small probabilities to bound the possibility that QoS requirements in delay and/or loss are not satisfied. Note that the design for statistical multiplexing relies on the design for lossless multiplexing, so that the complete treatment is the composite of phase 1 (lossless) and phase 2 (statistical).

This is the approach taken in [EMW95]. LoPresti et. al. [LZT97], in an important recent paper, formulated a general bandwidth-buffer trade-off problem in the lossless multiplexing framework, in addition to treating the buffers and bandwidth as two resources in the analysis of statistical multiplexing, in contrast to a conservative device in [EMW95] which allowed a reduction to a single-resource problem. Our extension of the resource allocation problem in lossless multiplexing to a shared buffer, multiport system integrates the notion of the *traffic class profile*, an input generated exogenously, possibly by market conditions, and also delay bounds for each traffic class. The solution to this general problem classifies bandwidth-buffer systems in a new way and sheds light on the desired mix of these resources to match traffic mixes. Le Boudec [LEB96] has also considered a problem of optimal resource allocation to inhomogeneous flows in the lossless framework in which the optimization criterion is quite different from ours.

It has been shown in [MMO95] that in a single-resource problem, the Chernoff estimate of the loss probability for asynchronous, independent sources is maximized by source rate waveforms which are on-off, periodic and extremal. This result was used in [EMI97] to formalize the choice of on-off waveforms as extremal in the bounding approach followed in [EMW95]. More recently, Rajagopal, Reisslein and Ross [RRR97], in an important work, have shown that the extremal source rate processes for a two-resource formulation of the statistical multiplexing problem are not on-off. Since the numerical results in [LZT97] and [RRR97] on system capacity with statistical multiplexing do not show a marked increase over that obtained by the approach in [EMW95] and [EMI97], which is based on a reduction to a single resource and on-off waveforms, we adopt here the latter, since it is substantially easier to implement.

Section 2 below gives the details of the system model and Buffer Admission Control. Section 3 formulates and solves a problem of optimally allocating buffers and bandwidth to ports in a lossless multiplexing framework. Section 4 describes the issues in fluid simulation of the BAC. Section 5 describes the numerical results, which are all based on statistical multiplexing. Concluding remarks are in Sec.6.

II. Description of System Model and Buffer Management

A. System Model

The N-port system is shown in Figure 1. The term "connection" is used interchangeably with "source". The shared buffer capacity is B and we let $C = \sum_i C_i$ denote the aggregate output bandwidth. Note that, while each port is work-conserving, work-conservation does not apply to the

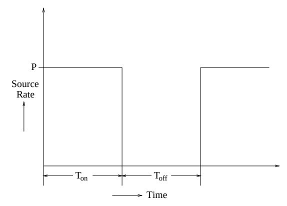


Fig. 2. Rate process $\Omega(t)$.

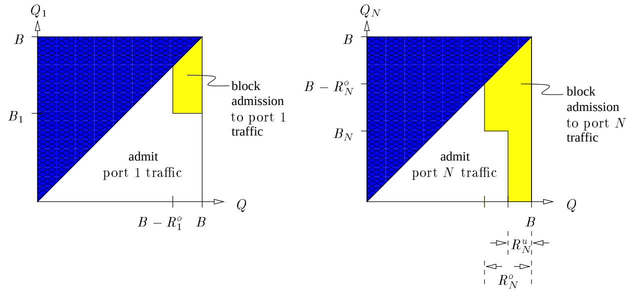


Fig. 3. Illustration of the Buffer Admission Control for an N-port system where ports 1 and N are carrying "guaranteed service" and "best effort" traffic, respectively.

aggregate output bandwidth. That is, unlike Generalized Processor Sharing (GPS), it is possible for some ports to be not fully utilized, while other ports are backlogged.

Each port has a logical or virtual FIFO queue, which is implemented as a linked-list, with cell addresses in a data RAM [EPA97]. Note that situations may arise where the buffer is full, but there is no queue for one or more ports. In this case, the incoming data for a port without a queue may access the port directly and not be blocked. We let $Q_i(t)$ denote the logical queue for port i at time t, and $Q(t) = \sum_i Q_i(t)$ denotes the total buffer occupancy at t.

We assume that there are K_i connections for port i. A tacit assumption made here is that the connections for any port share common QoS requirements and are policed by Dual Leaky Bucket regulators with similar, if not identical, specifications. In the analysis of the next section it is assumed that the connections for port i are homogeneous with Dual Leaky Bucket regulators specified by $(r_i, B_{T,i}, P_i)$. In the succeeding section on fluid simulations we allow the additional generality of heterogeneous traffic classes for each port. The QoS requirement on the probability of loss for connections going through port i is L_i, and the delay bound is D_i.

In the Dual Leaky Bucket regulator with specifications (r, B, P), r is the mean sustainable rate or the token rate, B_T is the burst tolerance or the token buffer size, and P is the peak rate. The connection traffic rate which is regulated by such a device is assumed to be on-off, periodic and extremal in the sense that when the source is on its rate is the maximum permitted rate, P. The amount of data Θ generated in an on period is also maximum, and is given by

$$\Theta = \frac{B_T P}{P - r}, \qquad (1)$$

with the mean rate in each period being r. This extremal rate process, denoted by $\Omega(t)$, is shown in Figure 2, where

$$T_{on} = B_T/(P - r) \qquad \textit{and} \qquad T_{off} = B_T/r. \qquad (2)$$

B. Description of the Buffer Admission Control

We let B_i denote the "nominal allocation" of buffer capacity to port i. Assume that B_i is calculated by some technique appropriate for single port systems, such as [EMW95], [LZT97] or [RRR97], which takes into account C_i, K_i, L_i and D_i. Thus, the process of nominal buffer allocation and CAC are intimately coupled. Typically, $\sum_i B_i/B \geq 1$, where the extent of buffer over-allocation is a subject of design influenced by issues such as multiplexing gain across ports, the degree of isolation desired for ports from other ports which are misbehaving due to, say, the breakdown of policing regulators, and the presence of "best effort" traffic through some ports. Some of these issues are probed in Section 5 on numerical investigations.

We define the "status" of a port i as *overloaded* if $Q_i(t) \geq B_i$ and *underloaded* if $Q_i(t) < B_i$. For each port i we define the pair of "reservation parameters" (R_i^u, R_i^o), with the superscripts u and o denoting "underloaded" and "overloaded", respectively. These parameters are chosen so that $R_i^u < R_i^o$ for all i. Typically, $R_i^u = 0$, as we shall see, and the main reason for having $R_i^u > 0$ is to give traffic for such a port lower priority in accessing the buffer. Our Buffer Admission Control (BAC) algorithm is as follows: *Admit traffic for port i into the buffer if and only if one of the following conditions is satisfied*

- *Port i is underloaded, i.e. $Q_i < B_i$ and $Q < B - R_i^u$.*
- *Port i is overloaded, i.e. $Q_i \geq B_i$ and $Q < B - R_i^o$.* (3)

From the fluid flow point of view, the BAC is not complete until the policy at the global thresholds $Q = B - R_i^s$, for each s and i, is specified. This is an important aspect of the BAC. The principle we adopt, which is based on a particular notion of fairness, admits fluid to affected ports at rates which correspond to equal fractions of the rate of total fluid generated by the sources for each port. This principle may readily be substituted by another such as admitting fractions proportional to prespecified weights, or, as in the cell-level control scheme of [RWM95], treating the admitted fractions as control variables to meet QoS requirements.

We illustrate the application of the BAC in a scenario where port 1 is carrying "guaranteed service" traffic, while port N is carrying "best effort" traffic. The tacit assumption is that the traffic through port 1 is tightly regulated and admission-controlled, while the traffic through port N is loosely regulated, perhaps only its peak rate is policed, and the QoS constraints for CAC are slack. Hence the flow for port 1 is given priority in accessing the buffer over port N. One level of control in buffer management is reflected in the selection of the nominal allocations B_1 and B_N. Yet another level of control is exercised through reservation parameters. We set $0 = R_1^u < R_N^u$ and $0 < R_1^o < R_N^o$. Figure 3 illustrates the BAC for the two ports. Consider an instant in time when port N is overloaded, while port 1 is underloaded and its sources have just become active.Besides the reserved buffer capacity being made available to port 1, the other mechanism working in its favor is that every departing cell of port N is systematically replaced by an arriving cell for port 1.

III. Bandwidth and Buffer Allocations: The Design Problem

In the first phase, the design problem is considered in the lossless multiplexing framework, and in the second phase with statistical multiplexing. The antecedents of this work are [EMW95], [LZT97] and [EMI97]. The lossless multiplexing problem is as follows. We are given (see Figure 1) B and C, the shared buffer capacity and aggregate output bandwidth, respectively. The connections for port i are subject to Dual Leaky Bucket regulation with parameters $(r_i, B_{T,i}, P_i)$, and furthermore, the sources are on-off periodic and extremal (see Figure 2). We are also given the *traffic class profile* $\boldsymbol{\eta} \triangleq (\eta_1, \eta_2, ..., \eta_N)$, $\sum_i \eta_i = 1$, which represents the relative mix of connections of the various traffic classes at the desired operating point. That is, the desired operating point is $(K_1, K_2, ..., K_N) = K\boldsymbol{\eta}$. The scalar K is hence a proxy for the capacity of the system for the specified traffic mix. The QoS requirements are zero loss and delay bound D_i for class/port i. The problem is to obtain the buffer-bandwidth allocations (B_i, C_i) for port i, $(i = 1, 2, ..., N)$, such that the capacity K is maximized subject to the resource constraints $\sum B_i \leq B$ and $\sum C_i \leq C$.

It will be shown shortly that the above problem is equivalent to the following.

Lossless Multiplexing Design Problem $(B, C, \boldsymbol{\eta}, \mathbf{D})$

$$\max \quad K \tag{4}$$

subject to

$$\sum B_i \leq B, \qquad \sum C_i \leq C \tag{5}$$

$$\begin{aligned} K_i\Theta_i - T_{on,i}C_i &\leq B_i & (i = 1, 2, ..., N) \quad (6)\\ B_i &\leq C_iD_i & (i = 1, 2, ..., N) \quad (7) \end{aligned}$$

where $K_i = K\eta_i$.

The burst size and time, Θ_i and $T_{on,i}$, are given in (1) and (2). The solution to this problem is given in the following subsection.

A. Lossless Multiplexing

We begin by briefly reviewing the notion of a "virtual buffer/trunk system" for a single source [EMW95]. Consider a single source of class i with periodic, on-off rate process $\Omega(t)$, which supplies an initially empty, infinite buffer with an output port of bandwidth c_i, where $c_i > r_i$ for stability. Let b_i denote the maximum buffer content over time. Then it is easy to see that

$$\Theta_i - T_{on,i}c_i = b_i \tag{8}$$

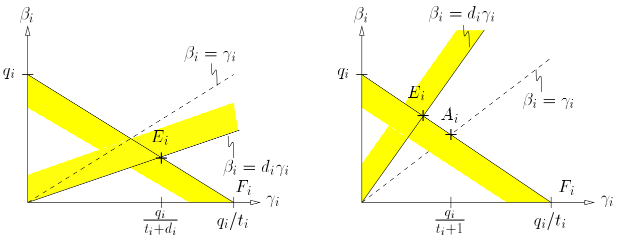


Fig. 4. Two cases for class i in the min-max design problem.

When K_i sources of class i are connected to port i, the constraint implied by lossless multiplexing is

$$K_i\Theta_i - T_{on,i}C_i \leq B_i, \tag{9}$$

where

$$C_i = K_ic_i \qquad \textit{and} \qquad B_i = K_ib_i. \tag{10}$$

The constraint implied by the bound on the maximum delay is

$$B_i \leq C_iD_i. \tag{11}$$

Thus we obtain the design problem stated in (4)-(7).

Next we systematically normalize with respect to B, C and K. Let

$$\beta_i \triangleq \eta_ib_i/B \quad \textit{and} \quad \gamma_i \triangleq \eta_ic_i/C \quad (i = 1, 2, ..., N) \tag{12}$$

The resource constraints in (5) translate to

$$\sum_i \beta_i \leq 1/K \qquad \textit{and} \qquad \sum_i \gamma_i \leq 1/K \tag{13}$$

Also, (9) and (11) give

$$q_i - t_i\gamma_i \leq \beta_i \leq d_i\gamma_i \qquad (i = 1, 2, ..., N) \tag{14}$$

where

$$q_i \triangleq \eta_i\Theta_i/B, \quad t_i \triangleq T_{on,i}C/B \quad \textit{and} \quad d_i \triangleq D_iC/B \tag{15}$$

Note that $\{q_i\}$, $\{t_i\}$ and $\{d_i\}$ are dimensionless parameters.

Hence the design problem is equivalent to the following min-max problem:

$$\min \quad K^{-1} \tag{16}$$

where

$$K^{-1} = \max \quad (\sum_{i=1}^{N} \beta_i, \ \sum_{i=1}^{N} \gamma_i) \tag{17}$$

subject to

$$q_i - t_i\gamma_i \leq \beta_i \leq d_i\gamma_i \qquad (i = 1, 2, ..., N) \tag{18}$$

To analyze the problem it will help to first consider two cases for each class: $d_i \leq 1$ and $d_i > 1$. For reasons which will become clear shortly, the former may be called "bandwidth constrained" and the latter "balanced" (see Figure 4). In case (i) the delay requirement necessitates a disproportionate amount of bandwidth to be expended.

Clearly, the point E_i is locally the most attractive allocation point. For case (ii), the delay requirement is not so stringent, which makes the balanced allocation point A_i locally the most attractive. However, as we shall see, global considerations will lead to allocations different from the local optima.

For the min-max problem (16)-(18), a little thought shows that, for each class/port i, the globally optimum allocation (β_i, γ_i) will lie on the segment E_iF_i shown in Figure 4. Hence, the constraint (18) may be replaced by the following: for $i = 1, 2, ..., N$

$$\beta_i = q_i - t_i\gamma_i \tag{19}$$

$$\frac{q_i}{t_i + d_i} \leq \gamma_i \leq \frac{q_i}{t_i} \tag{20}$$

Now, using (19), we may eliminate $\{\beta_i\}$ from the problem to obtain the following version:

$$\min \quad K^{-1}(\gamma) \tag{21}$$

where

$$K^{-1}(\gamma) = \begin{cases} \sum_i \gamma_i & \text{if } \sum_i\{(t_i+1)\gamma_i - q_i\} \geq 0 \\ \sum_i(q_i - t_i\gamma_i) & \text{if } \sum_i\{(t_i+1)\gamma_i - q_i\} < 0 \end{cases} \tag{22}$$

subject to

$$\frac{q_i}{t_i + d_i} \leq \gamma_i \leq \frac{q_i}{t_i} \qquad (i = 1, 2, ...N) \tag{23}$$

The two cases in (22) correspond to whether the limiting resource is bandwidth or buffer, respectively. Note too that a feasible solution always exists.

There are now two cases to consider depending upon the location of the corner point $\{q_i/(t_i + d_i)\}$ with respect to the separating hyperplane in γ-space, which is in (22). In the first, simpler, case

$$E \triangleq \sum_{i=1}^{N}(t_i + 1)\{\frac{q_i}{t_i + 1} - \frac{q_i}{t_i + d_i}\} \leq 0 \tag{24}$$

This may be referred to as the "globally bandwidth-constrained" case in analogy to the previously discussed classification for the single class. Certainly this case holds if each class is bandwidth constrained, i.e., $d_i \leq 1$ for all i. If (24) holds then it is easy to see that the solution to the min-max problem is

$$\gamma_i = q_i/(t_i + d_i) \quad \text{and} \quad \beta_i = d_i\gamma_i \quad (i = 1, 2, ...N) \tag{25}$$

Now consider the alternative to (24), namely $E > 0$, which may be termed the "globally balanced" case. In this case, it may be deduced that the solution to the min-max problem as stated in (21)-(23) must lie on the separating hyperplane where the buffer and bandwidth are simultaneously exhausted., i.e.,

$$\sum_{i=1}^{N}[(t_i + 1)\gamma_i - q_i] = 0 \tag{26}$$

It then follows that the min-max problem reduces to

$$\min \quad \sum_{i=1}^{N}\gamma_i \tag{27}$$

subject to (26), and the inequality constraints (23); also $K^{-1} = \sum \gamma_i$.

This problem is an LP with a simple solution, which is also exploited in [LZT97]. The solution is incorporated in the summary given below to the design problem (4)-(7), which is stated in terms of the original variables.

Proposition 1: Let

$$E = \frac{1}{B}\sum_{i=1}^{N}\eta_i\Theta_i\left[1 - \frac{T_{on,i} + B/C}{T_{on,i} + D_i}\right] \tag{28}$$

(i) In the "globally bandwidth-constrained" case, $E \leq 0$. In this case the maximum capacity is

$$K = \frac{C}{\sum_{i=1}^{N}\eta_i\Theta_i/(T_{on,i} + D_i)}, \tag{29}$$

which is obtained for the following buffer and bandwidth allocations.

$$B_i = \frac{K_i\Theta_iD_i}{T_{on,i} + D_i} \quad \text{and} \quad C_i = \frac{K_i\Theta_i}{T_{on,i} + D_i} \quad (i = 1, 2, ..., N) \tag{30}$$

where $K_i = K\eta_i$.

(ii) In the "globally balanced" case, $E > 0$. Let the classes be indexed in decreasing order of their on periods, i.e.,

$$T_{on,1} \geq T_{on,2} \geq ... \geq T_{on,N} \tag{31}$$

and let k be the smallest integer such that

$$E - \sum_{i=1}^{k-1}\frac{\eta_i\Theta_i}{B}\frac{D_i(T_{on,i} + B/C)}{T_{on,i}(T_{on,i} + D_i)} \leq \frac{\eta_k\Theta_k}{B}\frac{D_k(T_{on,k} + B/C)}{T_{on,k}(T_{on,k} + D_k)} \tag{32}$$

The capacity K is then maximized by the following allocations.

$$C_i = \begin{cases} K\eta_iP_i & i = 1, 2, ..., k-1 \\ K\eta_i\Theta_i/(T_{on,i} + D_i) & i = k+1, k+2, ..., N \\ \\ K[\frac{EB}{T_{on,k}+B/C} + & \\ \sum_{j=1}^{k-1}\eta_j\Theta_j\frac{T_{on,j}+B/C}{T_{on,k}+B/C}\{\frac{1}{T_{on,j}} - \frac{1}{T_{on,j}+D_j}\} + & \\ \frac{\eta_k\Theta_k}{T_{on,k}+D_k}] & i = k \end{cases} \tag{33}$$

The maximum capacity K is obtained from the relation $\sum_i C_i = C$. The buffer allocations are obtained from the relation

$$B_i = K_i\Theta_i - T_{on,i}C_i \qquad (i = 1, 2, ..., N) \tag{34}$$

which satisfies $\sum_i B_i = B$. •

The differences between the two cases in the above result are noteworthy. In the globally bandwidth-constrained case the sum of the allocated bandwidths equals the given aggregate bandwidth, but that is not the case with the buffers. In contrast, in the globally balanced case, both buffers and bandwidth are totally consumed by the allocations.

This dichotomy has broad implications, which we briefly touch upon now. Suppose that in the former case, class/port N is "best effort" with generous delay requirements, say $D_N \geq B/C$. Clearly, by giving greater weight to the best effort class in the traffic-class profile $\boldsymbol{\eta}$, i.e., by increasing η_N, perhaps induced by tariffs, the quantity E in (28) can be made to change from negative to positive, which would make the system globally balanced, resulting in full utilization of both resources with proper allocations.

B. Statistical Multiplexing

The objective here is to take the buffer-bandwidth allocation (B_i, C_i) for port i as given, and maximize the number of connections K_i by exploiting statistical multiplexing, such that the loss probability $P_{loss} \leq L_i$. In particular, as mentioned in Section 1, here we do not consider the sharing of buffer resources across ports as governed by the BAC. The technique given below for computing the loss probability for any given K_i is based on the Chernoff asymptotic approximation technique and is due to [EMW95].

A key observation is that the allocation (B_i, C_i) is based on the behavior of the "virtual buffer/trunk system" for a single source, where the peak buffer utilization is b_i and the peak bandwidth utilization is c_i, with (see 10),

$$\frac{b_i}{c_i} = \frac{B_i}{C_i} \tag{35}$$

This relation is key to the reduction to a single-resource, which we may now take to the bandwidth. The bandwidth utilization by a single connection k is denoted by u_{ik}. On account of the assumption that the sources have phases which are independent and uniformly distributed over the period $T_i = T_{on,i} + T_{off,i}$, we have

$$Pr(u_{ik} = c_i) = 1 - Pr(u_{ik} = 0) = w_i \tag{36}$$

where $w_i = r_i/c_i$. The loss probability for port i is

$$P_{loss,i} = Pr(U_i > C_i) \tag{37}$$

where the bandwidth load on port i is $U_i = \sum_{k=1}^{K_i} u_{ik}$. Since the distribution of the random variables is binomial, it is straightforward to obtain the following Chernoff asymptotic approximation to $P_{loss,i}$.

Proposition 2:

$$P_{loss,i} \sim \frac{e^{-F(s^*)}}{s^* \sigma(s^*)\sqrt{2\pi}} \tag{38}$$

where

$$a_i = \frac{C_i}{K_i}\frac{1}{c_i}, \qquad s^* = \frac{1}{c_i}\log\left[\frac{a_i}{1-a_i}.\frac{1-w_i}{w_i}\right],$$

$$F(s^*) = K_i \left[a_i \log\frac{a_i}{w_i} + (1-a_i)\log\frac{1-a_i}{1-w_i}\right]$$

and

$$\sigma^2(s) = K_i w_i (1-w_i) c_i^2 \{\exp(sc_i)\}/\{1 - w_i + w_i \exp(sc_i)\}^2$$

$\bullet$

The capacity $K_{max,i}$ is the value of K_i for which

$$\frac{e^{-F(s^*)}}{s^* \sigma(s^*)\sqrt{2\pi}} = L_i. \tag{39}$$

IV. Fluid Simulations

In this section, we investigate the technical issues in a fluid simulation of the proposed Buffer Admission Control scheme. For the simulation, we also allow the additional generality of possibly several classes of connections through each port, with each class having a corresponding Dual Leaky Bucket regulator. Note that the fluid flow representation is an approximation to cell-level behavior in which cell-level details in the behavior of sources, trunks and buffers are smoothed into a steady rate of fluid flow from the sources causing piecewise linear changes in buffer occupancy. Events such as admission/removal of connections and starts/ends of bursts in connections are key events in fluid simulations, which coincide with jumps in the fluid rates. In addition to these source related events, there are buffer and control related events. For example, when the buffer becomes empty or full, the rate of flow out of the buffer undergoes a discontinuous change. In a high speed system such as ATM, the advantage of fluid simulation accrues from the fact that these "events" occur on a much slower time-scale as compared to cell arrivals and departures, thereby requiring much less simulation time to cover a given amount of actual system time. Obviously, fluid simulations must be event-driven in order to realize this advantage.

In event-driven simulations, the event set in the system determines the logical complexity; the larger the event set, the more complex the simulation, since it is necessary to determine the earliest next event from among a larger number of possibilities.

In the simulations described here, the source model for traffic of an individual connection of class j going into port i consists of on-off, periodic rate waveforms, which are extremal with respect to the Dual Leaky Bucket regulator for the class (see Figure 2) and have phases that are independent random variables uniformly distributed over the period of the waveform. The rate process of the aggregate traffic offered to port i is the superposition of the rate processes of all connections through the port.

An important feature of the fluid simulations described here, which is a reflection of the BAC, is that the logical queues $\{Q_i\}$ for the ports are computed and monitored, of course, but not the queues for the individual connections.

The set of events relevant to our fluid simulation is of two types - (i) Source events and (ii) Buffer events. Buffer events, which are threshold crossings, can further be subdivided into (i) port-specific and (ii) global. We now proceed to enumerate and describe these events.

• **Type 1: Source state changes** The state changes are discrete events corresponding to the discontinuities of the piece-wise constant, aggregate rate process for each port.
• **Type 2: Port threshold crossings** When the logical queue for any port empties or the port uses up its nominal allocation, i.e., $Q_i(t) = 0$ or B_i, some of the incoming fluid is treated differently by the control scheme. Hence the rate of change of the port buffer occupancy, dQ_i/dt, encounters a discontinuity, as a result of which $dQ/dt = d(\sum_i Q_i)/dt$ is discontinuous as well. We postpone the detailed discussion of this event, except to note that the change in port status caused by this event causes changes in the admitted fluid rates for all classes feeding into the port.
• **Type 3: Global threshold crossings** These events are characterized by $Q(t) = B - R$ where $R = 0$ or B or $R = R^s_{ij}$ for some class j and status indicator s. When one of these thresholds is encountered, there is a jump in dQ/dt, since when $R = R^s_{ij}$, traffic of class j is not admitted into the buffer, and $dQ/dt = 0$ immediately after $Q = 0 \, or \, B$. We later discuss in detail the behavior of the system following this event.

Event types 2 and 3 correspond to instants when the BAC is active. There is a "stickiness" associated with these thresholds, i.e., the thresholds behave like attractors, holding the queue fixed until the next source state change. A similar feature is shown to exist and analyzed in the loss priority control scheme of [EMI94]. The stickiness is a result of our control policy at the thresholds, which requires admitting only a fraction of the fluid belonging to the affected classes. The exact composition of admitted fluid of the various classes can be controlled in different ways depending on QoS considerations. Here, motivated by fairness, equal fractions of each of these flows are admitted.

A. *Characterization and Analysis of Events*

We now proceed to describe how the aforementioned events are processed in the simulation. Each of the above events triggers computations that determine acceptance, buffer usage and loss rates for each port and source class till the next event occurs. The scheduling of future events is also dependent on these computations. The set of ports are grouped into subsets based on status.

• $\mathcal{U}_0 \equiv \{i : Q_i = 0\}$ denotes ports with zero occupancy.
• $\mathcal{U} \equiv \{i : Q_i > 0 \; and \; Q_i < B_i\}$ denotes underloaded ports with nonzero occupancy.
• $\mathcal{B} \equiv \{i : Q_i = B_i\}$ denotes "border" ports, i.e. these ports are at their overload thresholds.
• $\mathcal{O} \equiv \{i : Q_i > B_i\}$ denotes strictly overloaded ports.

A.1 Single Threshold System

For the benefit of the reader, we first consider the simple case where the reservation parameters are common to all ports, i.e., $R^u_i = 0$ and $R^o_i = R$ for all i. Thus, there is only one global threshold $B - R$ in the buffer occupancy. However, even in this case, inhomogeneity exists on account of the diverse character and number of connections through the ports and in the QoS specifications. This diversity is reflected in the possibly different nominal allocations B_i for different ports.

Let $\lambda_i(t)$ be the piece-wise constant rate of fluid generated for port i. We then have the following cases:
1. $0 \le Q < B - R$: No control is active in this case and all fluid going into the system is accepted, i.e.,

$$\frac{dQ}{dt} = \sum_i \frac{dQ_i}{dt} = \sum_{i \in \mathcal{U}_0} (\lambda_i(t) - C_i)^+ + \sum_{i \in \mathcal{U},\mathcal{B},\mathcal{O}} (\lambda_i(t) - C_i)$$

where $x^+ = max(x, 0)$.
2. $Q = B - R$: In this case, the BAC blocks flows to overloaded ports. However, ports at the border between overload and underload, i.e., those with $Q_i(t) = B_i$, must be treated carefully in the fluid system. To start with, we avoid this complication and assume that there are no such *border* ports with $Q_i(t) = B_i$ at instant t. We then define the following rate function $f(\beta)$ in terms of the fraction $\beta \in [0, 1]$ of admitted fluid from overloaded ports.

$$\begin{aligned} f(\beta) &= \sum_{i \in \mathcal{U}_0} (\lambda_i(t) - C_i)^+ + \sum_{i \in \mathcal{U}} (\lambda_i(t) - C_i) \\ &\quad + \sum_{i \in \mathcal{O}} (\beta \lambda_i(t) - C_i) \end{aligned} \tag{40}$$

Note that $f(\beta)$ monotonically increases with β. We then have the following cases:
(a) $f(0) \ge 0$: In this case, the underloaded ports by themselves cause a net increase in buffer occupancy and hence no fluid from overloaded ports is admitted, i.e., $\beta = 0$ and $dQ/dt = f(0)$.
(b) $f(1) \le 0$: All fluid coming into the system is admitted and no control is exercised, i.e., $\beta = 1$ and $dQ/dt = f(1)$.
(c) $f(0) < 0 < f(1)$: In this case, the operating value of β satisfies $f(\beta) = 0$ and we have $dQ/dt = f(\beta) = 0$, i.e., the net occupancy exhibits stickiness at $Q = B - R$.
Now consider the case in which the queue for one of the ports is at nominal occupancy, say $Q_i(t) = B_i$. All fluid going into this port is admitted if the port subsequently becomes underloaded, which is the case if $\lambda_i(t) - C_i \le 0$. Alternatively, if the "fairshare" $(\beta \lambda_i(t) - C_i)$ causes the port to stay overloaded, only the fairshare is admitted. If neither of these conditions is true, the occupancy of this port stays unchanged at $Q_i(t) = B_i$. This port-level stickiness is induced by the notion of port status, which undergoes a discontinuous change when the condition $Q_i(t) = B_i$ is crossed. Thus, the contribution of this port to the net drift is given by

$$\frac{dQ_i}{dt} = (\lambda_i(t) - C_i)^- + (\beta \lambda_i(t) - C_i)^- \tag{41}$$

where $x^- = min(x, 0)$. For this case, the modified rate function is

$$\begin{aligned} f(\beta) &= \sum_{i \in \mathcal{U}_0} (\lambda_i(t) - C_i)^+ + \sum_{i \in \mathcal{U}} (\lambda_i(t) - C_i) + \\ &\quad \sum_{i \in \mathcal{B}} [(\lambda_i(t) - C_i)^- + (\beta \lambda_i(t) - C_i)^+] + \end{aligned}$$

$$\sum_{i\in\mathcal{O}}(\beta\lambda_i(t) - C_i). \qquad (42)$$

This rate function is used instead of (40) to determine β and the net rate of change of the buffer occupancy, dQ/dt.
3. $B-R < Q(t) < B$: In this case, no fluid from ports with $Q_i(t) > B_i$ is admitted. However, a fraction of fluid from ports with $Q_i(t) = B_i$ is admitted so as to keep $dQ_i/dt \le 0$. Thus, we have

$$\begin{aligned}\frac{dQ}{dt} &= \sum_{i\in\mathcal{U}_0}(\lambda_i(t) - C_i)^+ + \sum_{i\in\mathcal{U}}(\lambda_i(t) - C_i) + \\ &\quad \sum_{i\in\mathcal{B}}(\lambda_i(t) - C_i)^- \end{aligned} \qquad (43)$$

4. $Q(t) = B$: When the entire buffer is occupied, the uniform fraction β of fluid going into underloaded and border ports that is admitted is such that the net drift is non-positive. We define the rate function

$$\begin{aligned}f(\beta) &= \sum_{i\in\mathcal{U}_0}(\beta\lambda_i(t) - C_i)^+ + \sum_{i\in\mathcal{U}}(\beta\lambda_i(t) - C_i) + \\ &\quad \sum_{i\in\mathcal{B}}(\beta\lambda_i(t) - C_i)^- \end{aligned} \qquad (44)$$

which clearly satisfies $f(0) < 0$. We have the cases

(a) $f(1) \le 0$: All fluid from underloaded ports and border ports is admitted, i.e. $\beta = 1$.

(b) $f(1) > 0$: The admitted fraction β satisfies $f(\beta) = 0$ and the net occupancy sticks at $Q = B$ till the next event.

A.2 General Case

We now extend the above analysis of events to the general case where the reservation parameters are arbitrary and there are a number of classes of connections indexed by $j = 1, ..., J$ through each port. Let $\lambda_{ij}(t)$ denote the piece-wise constant rate of fluid of class j offered to port i at time t. Then, the total rate of fluid generated for port i at time t is given by

$$S_i(t) = \sum_j \lambda_{ij}(t).$$

The strictly admissible, or "legal", component of this rate is given by

$$S_i^l(t) = \sum_{j:Q(t)<B-R_{ij}^s} \lambda_{ij}(t) \qquad (45)$$

where the superscript s in R_{ij}^s denotes "status" of port i at time t.

Type 1 Events are most simply processed since they merely involve changing the values of the legal fluid rates. Assume that, at time t, the total occupancy is not at a global threshold, i.e., $Q(t) \ne B - R$ where R is any one of the reservation parameters R_{ij}^s, and the port occupancies are not at their nominal values, i.e. $Q_i(t) \ne B_i$. Then the rates of change for these quantities at this instant are expressed simply as

$$\begin{aligned}\frac{dQ_i}{dt} &= I(Q_i(t) = 0)\,(S_i^l(t) - C_i)^+ + \\ &\quad (1 - I(Q_i(t) = 0))\,(S_i^l(t) - C_i) \\ \frac{dQ}{dt} &= \sum_{i\in\mathcal{U}_0}(S_i^l(t) - C_i)^+ + \sum_{i\in\mathcal{U}}(S_i^l(t) - C_i) + \\ &\quad \sum_{i\in\mathcal{O}}(S_i^l(t) - C_i)\end{aligned} \qquad (46)$$

where $I(.)$ is an indicator function, which is 1 if the argument is true and 0 otherwise. Note that $\mathcal{B} = \emptyset$ since we have assumed that no port queue is at its nominal value. We will later modify equations (46) to include terms for these ports, but with this assumption, (46) determines the evolution of port occupancies till an event of type 2 or type 3 occurs, i.e. one of the thresholds is encountered.

Type 2 Events occur when one or more port queues are at nominal allocations $\{B_i\}$. In this case, the set $\mathcal{B}$ is non-empty and we need to include these ports in the rate calculations. We assume initially that the total occupancy $Q(t)$ is not at a global threshold. For ports in $\mathcal{B}$, we define the following aggregate rate from "border"[1] flows,

$$S_i^{bp}(t) = \sum_{j:B-R_{ij}^o<Q(t)<B-R_{ij}^u} \lambda_{ij}(t) \qquad (47)$$

The "legal" flow rates for these ports is computed using (45) and the o superscript on the reservation parameters. We then have the following cases

1. $S_i^l(t) > C_i$. In this case, $dQ_i/dt = S_i^l(t) - C_i$ is positive and none of the border fluid is admitted, since the port becomes strictly overloaded after this instant.
2. $S_i^l(t) + S_i^{bp}(t) < C_i$. In this case, $dQ_i/dt = S_i^l(t) + S_i^{bp}(t) - C_i$ is strictly negative and all the border fluid is admitted, since the port becomes underloaded after this instant.
3. $S_i^l(t) \le C_i \le S_i^l(t) + S_i^{bp}(t)$. This case demonstrates "stickiness" at the port level. The fluid system has $dQ_i/dt = 0$, while the cell behavior experiences fluctuations about $Q_i = B_i$ with a long-term adherence to this condition. In this case, the contention is between the various classes of fluid going into port i. Following our notion of fairness, we propose that a fraction α_i of the border fluid of each border class be admitted. This fraction is determined by the equation

$$\frac{dQ_i}{dt} = S_i^l(t) + \alpha_i S_i^{bp}(t) - C_i = 0$$

which gives

$$\alpha_i = (C_i - S_i^l(t))/S_i^{bp}(t)$$

These cases can be combined into the following pair of equations that determine dQ_i/dt and α_i for ports in $\mathcal{B}$

[1] We would like to distinguish this "port-level border" by superscript bp, from other "globally border" flows to be defined later on.

$$\frac{dQ_i}{dt} = (S_i^l(t) - C_i)^+ \quad + \quad (S_i^l(t) + S_i^{bp}(t) - C_i)^-$$
$$S_i^l(t) + \alpha_i S_i^{bp}(t) - C_i \;=\; 0 \quad 0 \le \alpha_i \le 1 \tag{48}$$

With this information, the system evolution is determined, until the next event, by the equation

$$\begin{aligned}\frac{dQ}{dt} &= \sum_i \frac{dQ_i}{dt} = \sum_{i\in\mathcal{U}_0}(S_i^l(t) - C_i)^+ + \sum_{i\in\mathcal{U}}(S_i^l(t) - C_i) + \\ &\sum_{i\in\mathcal{B}}[(S_i^l(t) - C_i)^+ + (S_i^l(t) + S_i^{bp}(t) - C_i)^-] + \\ &\sum_{i\in\mathcal{O}}(S_i^l(t) - C_i)\end{aligned} \tag{49}$$

Type 3 Events are the most computationally involved and also have the most impact in the control scheme. For simplicity, we begin with the case $\mathcal{B} = \emptyset$ again. Assume that, at instant t, $Q(t) = B - R$ where $R = R_j^s$ is one of the reservation parameters. Then, for each port, we define the "globally border" flow rate S_i^{bg} as

$$S_i^{bg}(t) = \sum_{j:Q=B-R_{ij}^s} \lambda_{ij}(t) \tag{50}$$

where, as before, superscript s represents port status(u or o). The possibilities for system evolution are then as follows
1. $dQ/dt > 0$: This is possible if and only if the legal flows by themselves cause the total occupancy to increase, since S_i^{bg} would be strictly disallowed after t. Thus we require

$$\begin{aligned}\frac{dQ}{dt}^{(1)} &= \sum_{i\in\mathcal{U}_0}(S_i^l(t) - C_i)^+ + \sum_{i\in\mathcal{U}}(S_i^l(t) - C_i) + \\ &\sum_{i\in\mathcal{O}}(S_i^l(t) - C_i) > 0.\end{aligned} \tag{51}$$

2. $dQ/dt < 0$: This condition occurs when admitting the legal and the border flows still causes the total occupancy to decrease. This implies

$$\begin{aligned}\frac{dQ}{dt}^{(2)} &= \sum_{i\in\mathcal{U}_0}(S_i^l(t) + S_i^{bg}(t) - C_i)^+ + \\ &\sum_{i\in\mathcal{U}}(S_i^l(t) + S_i^{bg}(t) - C_i) + \\ &\sum_{i\in\mathcal{O}}(S_i^l(t) + S_i^{bg}(t) - C_i) < 0.\end{aligned} \tag{52}$$

3. $dQ/dt = 0$: This case occurs when $dQ/dt^{(1)} \le 0$ and $dQ/dt^{(2)} \ge 0$. Then, as before, we admit a fraction β of the border fluid that would keep the total occupancy unchanged. This fraction is determined by the equation

$$\frac{dQ}{dt} = f(\beta) = 0 \tag{53}$$

where

$$\begin{aligned}f(\beta) &= \sum_{i\in\mathcal{U}_0}(S_i^l(t) + \beta S_i^{bg}(t) - C_i)^+ + \\ &\sum_{i\in\mathcal{U}}(S_i^l(t) + \beta S_i^{bg}(t) - C_i) + \\ &\sum_{i\in\mathcal{O}}(S_i^l(t) + \beta S_i^{bg}(t) - C_i)\end{aligned} \tag{54}$$

The form of equation (53) is generic to the computation of "fairshare" in various contexts, such as Fair Queueing, Generalized Processor Sharing and Max-Min Fairness. A solution procedure is described in the Appendix.
When $\mathcal{B} \neq \emptyset$, we must make further distinctions between the flows into ports with $Q_i = B_i$ for type 3 events. We divide these flows into the following four classes,

$$\begin{aligned}S_i^l &= \sum_{j:Q<R_{ij}^o} \lambda_{ij} \quad ; \quad S_i^{bg1} = \sum_{j:Q=R_{ij}^o} \lambda_{ij} \\ S_i^{bp} &= \sum_{j:R_{ij}^o<Q<R_{ij}^u} \lambda_{ij} \quad ; \quad S_i^{bg2} = \sum_{j:Q=R_{ij}^u} \lambda_{ij}\end{aligned} \tag{55}$$

The rate equations must now include these ports. Admitting a fraction β of border fluid results in the following rate changes.

$$\begin{aligned}\frac{dQ_i}{dt} &= \begin{cases} (S_i^l + \beta S_i^{bg} - C_i)^+ & i \in \mathcal{U}_0 \\ (S_i^l + \beta S_i^{bg} - C_i) & i \in \mathcal{U} \text{ or } i \in \mathcal{O} \\ \\ (S_i^l + \beta S_i^{bg1} - C_i)^+ + \\ (S_i^l + S_i^{bg1} + S_i^{bp} + \beta S_i^{bg2} - C_i)^- & i \in \mathcal{B} \end{cases} \\ \frac{dQ}{dt} &= \sum_i \frac{dQ_i}{dt} = f(\beta)\end{aligned} \tag{56}$$

The possibilities are as follows.
1. $f(0) > 0$ implies $dQ/dt = f(0) > 0$, i.e. only legal flows are admitted.
2. $f(1) < 0$ implies $dQ/dt = f(1) < 0$. Border flows have unrestricted access.
3. $f(0) < 0 < f(1)$ implies $dQ/dt = 0$. The fraction β of border flows admitted is determined $f(\beta) = 0$. This equation is solved using the procedure outlined in the Appendix.

V. Numerical Investigation

In this section we present simulation results of case studies on the efficacy of the BAC and the influence of the scheme's parameters on system performance. All cases studied involve statistical multiplexing. In fact, as we shall see, in some cases the number of connections considered are significantly more than can be supported if there was no buffer sharing among ports. Also, the cell loss ratios (CLR) given in the results are the ratios of lost cells to cells offered to the buffer. We have also computed the fraction of time during which the losses occur but these are not reported. Finally, we consider two classes of connections with rather different characteristics. All results reported are in duplicate, one set for each class. The results are grouped in five subsections, with Sec. 5.1 giving background data and benchmarks for the whole section.

A. Background Data and Benchmarks

In this study the standard sources or connections are either of class 1 or 2 with the specifications given in Table I. These classes have been studied earlier in [EMW95], [LZT97] and [RRR97]. Note that non-regulated sources, "rogue" and "best-effort" are also considered in the study. Table II contains basic data on

Class	r(Mbps)	P(Mbps)	Θ(cells)	T_{on}(ms)	T(ms)
1	0.15	6	25	1.7	70.6
2	0.15	15	250	70.6	706.6

TABLE I
DUAL LEAKY BUCKET PARAMETERS.

statistical multiplexing gain in capacity that may be expected from sharing the buffer. Single port and 4-port systems are considered. In all cases the CLR is in the range $L = 10^{-4} - 10^{-5}$, which was a choice made to keep the burden on the simulations acceptable.

The buffer management for the multiport system of Table II is *Complete Sharing*, which is the case of Virtual Partitioning is all reservation parameters are null. Thus, in the multiport system, no consideration is given to selective isolation and protection for purposes of QoS, and hence the multiplexing gain reported is optimistic. In Table II,

Class	Single port system	4-port system	Sharing gain
1	$K_{max} = 90, B = 333.3 cells$	$K_i \equiv 90, B = 800 cells$	1.66
2	$K_{max} = 54, B = 333.3 cells$	$K_i \equiv 54, B = 700 cells$	1.9

TABLE II
BUFFER REQUIREMENTS IN SINGLE AND MULTI-PORT SYSTEMS WITH COMPLETE SHARING. $L \approx 10^{-4}$ AND PORT SPEED=15MBPS.

K_{max} is the maximum admissible number of connections in the single port system for given L, $B = 333$ *cells* and $C = 15Mbps$. Thus, K_{max} takes into account statistical multiplexing across connections through each port.

In the following experiments, we take the multiplexing advantage between ports into account only when all connections in the system belong to the guaranteed service category. We do not assume any multiplexing advantage between ports when the buffer is shared with "best effort" or greedy "rogue" connections.

B. BAC Performance under Homogeneous Loading

Here we provide basic performance data for our BAC. The case considered has four ports ($N = 4$) with equal numbers of connections through each port (see Figure 5. The reservation parameters are common to all ports:

$$R_i^u = 0, \quad R_i^o = R \tag{57}$$

and the parameter R is varied. In Figures 5(a) and 6(a), there are $K_i = 90$ class 1 connections and in Figures 5(b) and 6(b), there are $K_i = 54$ class 2 connections through each port.

The figure plots unconditional and conditional CLR. Of great importance is the "CLR conditional on port underload" for each port, which is the ratio of cells lost to cells offered during periods when the port is underloaded. This quantity, above all, should be small if the BAC is working effectively in conjunction with CAC.

Figures 5 and 6 show the effect of the reservation parameters on the conditional and unconditional CLR values for two different nominal buffer allocations B_i. In Figure 5, B_i are set aggressively at 333 cells per port, while in Figure 6, the B_i are set conservatively so that $\sum B_i = B$. For class 1 connections, which have a high value of P/r of 40, the reservation mechanism helps to substantially reduce the CLR conditional on underload only for the conservative setting of $\{B_i\}$, while for class 2 connections with a more benign P/r of 10, reservation is uniformly effective in restricting underloaded CLR. In both cases, the CLR conditional on underload is uniformly lower than the unconditional CLR. The reservation effectively trades off overloaded CLR for underloaded CLR, while the unconditional CLR increases very slowly with R. Thus, at the cost of a small amount of throughput, the reservation mechanism in the BAC enforces compliance to resource allocations in CAC, even while allowing effective sharing of buffers across ports.

C. Inhomogeneity among Ports/Classes

The objective here is to investigate the capabilities of the reservation mechanism to support diverse QoS to the individual ports (see Figure 7), where the single variable R controls the reservation parameters of each of the four ports:

$$R_i^u = 0 \qquad (1 \le i \le 4)$$
$$R_1^o = R, \quad R_2^o = 2R, \quad R_3^o = 3R, \quad R_4^o = 4R \tag{58}$$

The nominal allocations to the ports are set conservatively at $B_i = B/4$.

For Figure 7 the number of connections through ports, K_i, in each case, have been selected such that, if the ports were isolated with buffer allocations matching the nominal allocations, the QoS experienced by the ports would be substantially different ($L \approx 10^{-3} - 10^{-6}$). However, note that, in Figure 7(a) at $R = 0$, the intended difference in CLR is completely diluted and the CLR experienced by the ports is roughly similar, while Figure 7(b) shows a similar dilution to a lesser extent. The main observation on this experiment is that a small value of R (eg. $R = 10$ *cells*) introduces a wide separation in QoS, again at the cost of a small amount of throughput. To minimize this slowly varying difference in throughput, the design should select the smallest value of R that satisfies the CLR of all the ports. This selection is usually determined by the most stringent CLR requirement.

D. Effect of "Best-Effort" Traffic

Here we consider the case where one of the ports, port 4, in a 4-port system is dedicated to carrying traffic not regulated by Dual Leaky Buckets. The traffic through port-4, called "best-effort", has been designed to be detrimental to the other ports carrying regulated traffic, from the multiplexing point of view. We assume that the aggregate source traffic for port 4 is offered to the buffer at a constant rate, λ_4, which is slightly in excess of the bandwidth of the port, C_4. Note that, if $\lambda_4 \leq C_4$, then $Q_4(t) = 0$ if the queue is initially empty. Indeed, a good implementation of the "best-effort" service would restrict the peak rate not to exceed the port speed. Here, however, our objective is to test the BAC under the presence of such a greedy source that takes up all of the buffer space permitted to it. We chose

$$\lambda_4 = 1.0001\, C_4 \qquad \text{for Fig. 8(a)} \tag{59}$$
$$\lambda_4 = 1.01\, C_4 \qquad \text{for Fig. 8(b)} \tag{60}$$

The reservation here is applied only against the best effort port, i.e., $R_4^u = 0, R_4^o = R$, while for the other ports $R_i^u = R_i^o = 0$.

The results in Figure 8 show that the scheme hence provides substantial protection and isolation for the regulated sources, while carrying substantial "best effort" traffic.

E. Effect of Misbehaving Connection

In Figure 9, we show the effect of a "rogue" connection through port 4 in a 4-port system. Under ideal conditions, the system is as described in Section 5.2 with the same reservation parameters and nominal allocations. However, we do not expect any multiplexing advantage across ports in such a scenario, and we set $B = 4B_i = 1333\ cells$. The BAC here is only intended to protect the well-behaved ports 1-3.

Our model of the misbehaving guaranteed service connection is that of a connection with a defective Dual Leaky Bucket regulator that fails to regulate the mean rate. Hence, the connection pumps in fluid into the system steadily at it's peak rate. This case differs from the best-effort case of Section 5.4 in that the reservation is applied uniformly to all ports, and also the port with the misbehaving connection has other well-behaved connections through it. Thus, while the reservation restricts well-behaved ports as well, the rogue port is not always backlogged on account of the multiplexing in bandwidth between connections through port 4.

The results in Figure 9 once again reveal the benefits of the reservation mechanism. The ports carrying regulated traffic are effectively protected from the effects of the rogue connection. The difference between cases (a) and (b) is due in part to the fact that in the former the peak rate of one source is a significant fraction (6/15) of the port speed.

VI. Conclusions

This paper has proposed and evaluated a shared buffer management scheme in which the Buffer Admission Control is coupled to the Connection Admission Control. Also, a problem of optimal allocation of buffers and bandwidth to ports in a shared memory, multiport system has been formulated and solved in a lossless multiplexing framework. The scheme is found to be efficient and robust.

Extensions to the work include, first, an analytical technique to evaluate the benefits of statistically multiplexing the buffer resource across ports and second, the development of fluid simulation techniques for the Generalized Processor Sharing discipline in a multiport system.

(a).

(b).

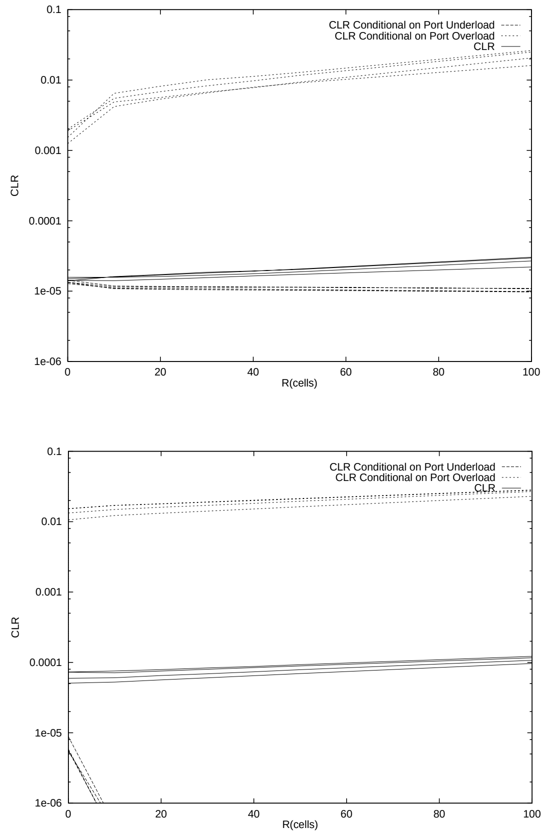


Fig. 5. Effect of reservation parameter on unconditional and conditional cell loss ratios in 4-port system. See Sec. 5.2. (a) $K_i = 90$ class 1 connections in each port, $B = 800\ cells$, $B_i = 333\ cells$, $C_i = 15Mbps$. (b) $K_i = 54$ class 2 connections in each port, $B = 700\ cells$, $B_i = 333\ cells$, $C_i = 15Mbps$.

References

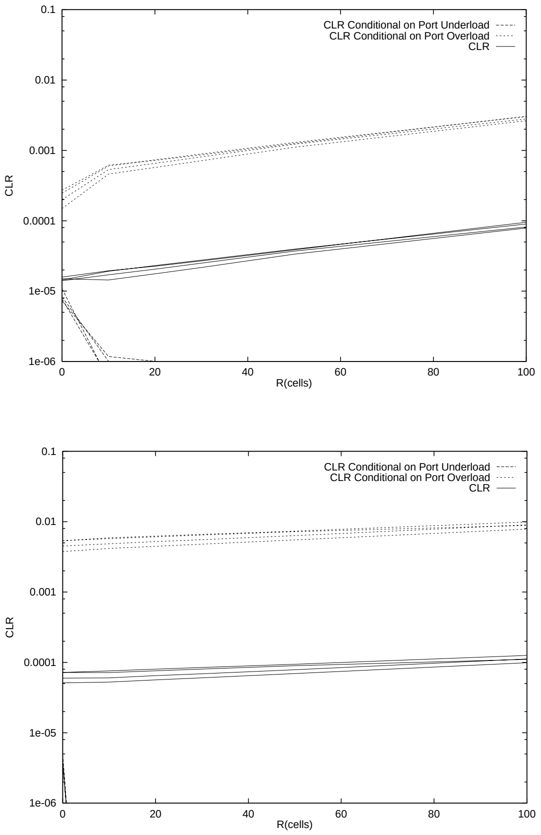
[AKI84] J. M. Akinpelu. The Overload Performance of Engineered networks with Non-Hierarchical and Hierarchical Routing. *AT&T Bell Labs Technical Journal*, 63:1261–1281, 1984.

[BMI97] S. Borst and D. Mitra. Virtual Partitioning for Resource Sharing by State-Dependent Priorities: Analysis, Approximations and Performance for Heterogeneous Traffic. In *Teletraffic Contributions for the Information Age, Proc. ITC-15*, ed. V. Ramaswami and P. E. Wirth, pages 1457–1468. Elsevier, 1997.

[CHA93] A. K. Choudhury and E. L. Hahne. Space Priority Management in a Shared Memory ATM Switch. In *Proc. IEEE GLOBECOM'93*, volume 3, pages 1375–1383, Dec. 1993.

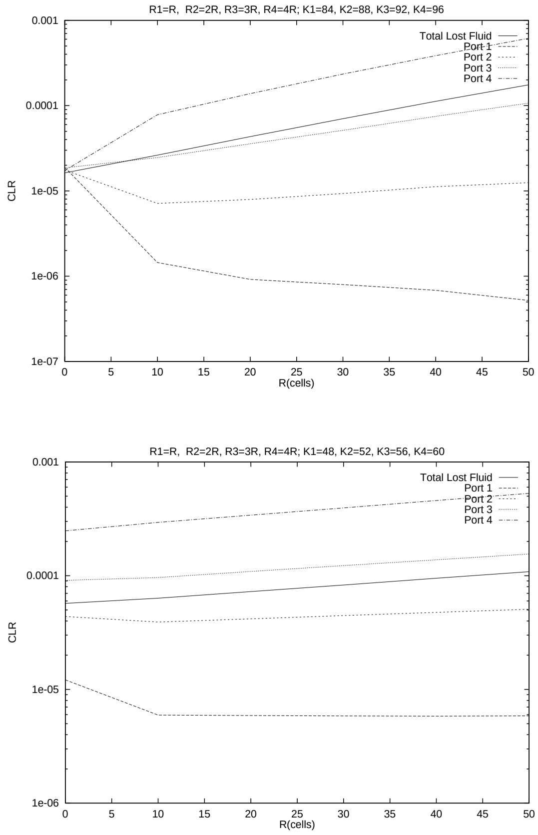
[CHA96] A. K. Choudhury and E. L. Hahne. Dynamic Queue Length Thresholds in a Shared Memory ATM Switch. In *Proc. IEEE INFOCOM'96*, volume 3, pages 679–687, March 1996.

[CSZ93] D. Clark, S. Shenker, and L. Zhang. Supporting Real-Time Applications in an Integrated Services Packet Net-



(a). (b).

Fig. 6. The parameters are identical to Figure fig:case1, (a) and (b), except that $B_i = B/4$, $i = 1, 2, 3, 4$, for (a) and (b), respectively.



(a). (b).

Fig. 7. Inhomogeneity among ports/classes in 4-port system. See Sec. 5.3. (a) Class 1 connections in all ports, $B = 800$ *cells*, $B_i = 333$ *cells*, $C_i = 15Mbps$. (b) Class 2 connections in all ports, $B = 700$ *cells*, $B_i = 333$ *cells*, $C_i = 15Mbps$.

work: Architecture and Mechanism. In *Proc. ACM SIGCOMM'93*, pages 14–26, 1993.

[EMI94] A. Elwalid and D. Mitra. Statistical Multiplexing with Loss Priorities in Rate-Based Congestion Control of High-Speed Networks. *IEEE Trans. Communications*, 42(11):2989–3002, Nov. 1994.

[EMI97] A. Elwalid and D. Mitra. Traffic Shaping at a Network Node: Theory, Optimum Design, Admission Control. In *Proc. IEEE INFOCOM'97*, pages 444–454, 1997.

[EMW95] A. Elwalid, D. Mitra, and R. H. Wentworth. A New Approach for Allocating Buffers and Bandwidth to Heterogenous, Regulated Traffic in an ATM Node. *IEEE Journal on Selected Areas in Communications*, 13(6):1115–1127, Aug. 1995.

[EPA97] K. Y. Eng and M. A. Pashan. Advances in Shared-Memory Designs for Gigabit ATM Switching. *Bell Labs Technical Journal*, 2(2):175–187, Spring 1997.

[FGO83] G. J. Foschini and B. Gopinath. Sharing Memory Optimally. *IEEE Trans. Commun.*, 3:352–360, March 1983.

[GCG95] L. Georgiadis, I. Cidon, R. Guerin, and A. Khamisy. Optimal Buffer Sharing. *IEEE J. Select. Areas Commun.*, 13:1229–1240, Sept. 1995.

[HBO93] E. J. Hernandez-Valencia and F. G. Bonomi. Simulation of a Simple Loss/Delay Priority Scheme for Shared Memory ATM Fabrics. In *Proc. IEEE GLOBECOM'93*, volume 3, pages 1389–1394, Dec. 1993.

[IRL78] M. I. Irland. Buffer Management in a Packet Switch. *IEEE Trans. Commun.*, 26:328–337, March 1978.

[KEY90] P. B. Key. Optimal Control and Trunk Reservation in Loss Networks. *Prob. Eng. Inf. Sci.*, 4:203–242, 1990.

[LEB96] J.-Y. Le Boudec. Network Calculus Made Easy. EPFL Technical Report 96/218, Dec. 1996.

[LZI97] J.-Y. Le Boudec and A. Ziedins. A CAC Algorithm for VBR Connections over a VBR Trunk. In *Teletraffic Contributions for the Information Age, Proc. ITC-15*, ed. V. Ramaswami and P. E. Wirth, pages 59–70. Elsevier, 1997.

[LZT97] F. LoPresti, Z. Zhang, D. Towsley, and J. Kurose. Source Time Scale Optimal Buffer/Bandwidth Trade-off for Regulated Traffic in an ATM Node. In *Proc. IEEE INFOCOM'97*, pages 675–682, 1997.

[MGH91] D. Mitra, R. J. Gibbens, and B. D. Huang. Analysis and Optimal Design of Aggregated-Least-Busy-Alternative Routing in Symmetrical Lossless Networks with Trunk Reservations. In *Teletraffic and Datatraffic, Proc. ITC-13*, ed. A. Jensen and V. B. Iversen, pages 477–482. North Holland, 1991.

[MMO95] D. Mitra and J. A. Morrison. Multiple Time Scale Regulation and Worst Case Processes for ATM Network Control. In *Proc. 34th Conference on Decision and Control*, pages 353–358, 1995.

[MZI96] D. Mitra and I. Ziedins. Virtual Partitioning by Dynamic Priorities: Fair and Efficient Resource Sharing by Several Services. In *Broadband Communications, Proceedings of the International Zurich Seminar on Digital Communications*, ed. B. Plattner, pages 173–185. Springer, 1996.

[MZI97] D. Mitra and I. Ziedins. Hierarchical Virtual Partitioning: Algorithms for Virtual Private Networking. *Proc. IEEE GLOBECOM'97*, pages 1784–1791, 1997.

[REI91] M. I. Reiman. Optimal Trunk Reservation for a Critically Loaded Link. In *Teletraffic and Datatraffic, Proc. ITC-13*, ed. A. Jensen and V. B. Iversen, pages 247–252. North Holland, 1991.

[RRR97] S. Rajagopal, M. Reisslein, and K. W. Ross. Packet Multiplexers with Adversarial Regulated Traffic. *preprint*, 1997.

[RWM95] M. I. Reiman, J. Wang, and D. Mitra. Dynamic Call Admission Control of an ATM Multiplexer with On/Off Sources. In *Proceedings of the 34th Conference on Decision and Control*, pages 1382–1388, 1995.

[SWR97] K. Y. Siu, Y. Wu, and W. Ren. Virtual Queueing Techniques for UBR+ Service in ATM with Fair Access

(a).

(b).

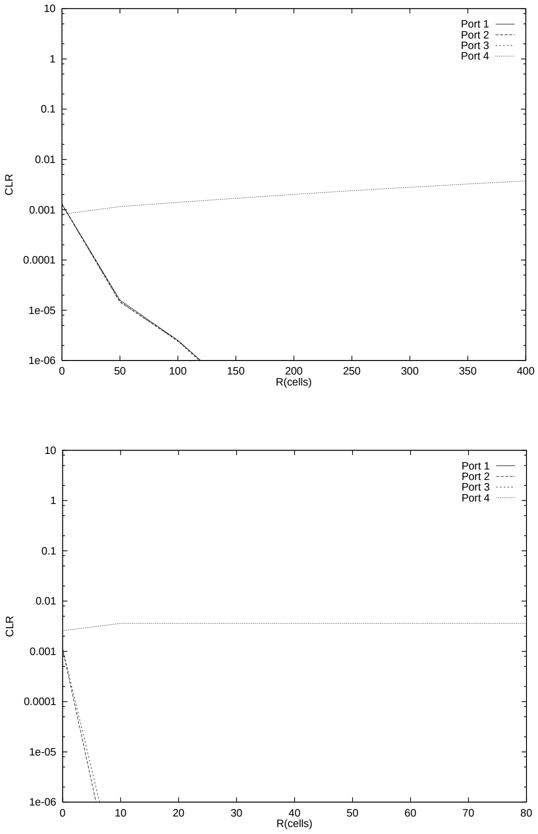


Fig. 8. (b). Port 4 carries "best-effort traffic, while ports 1,2,3 carry exclusively class 1 or class 2 traffic in (a) and (b), respectively. See Sec.5.4. (a) $B = 1333\ cells$, $B_i = B/3$ for $i = 1, 2, 3$, $B_4 = 0$, $C_i = 15Mbps$. (b) $B = 1333\ cells$, $B_i = B/3$ for $i = 1, 2, 3$, $B_4 = 0$, $C_i = 15Mbps$.

(a).

(b).

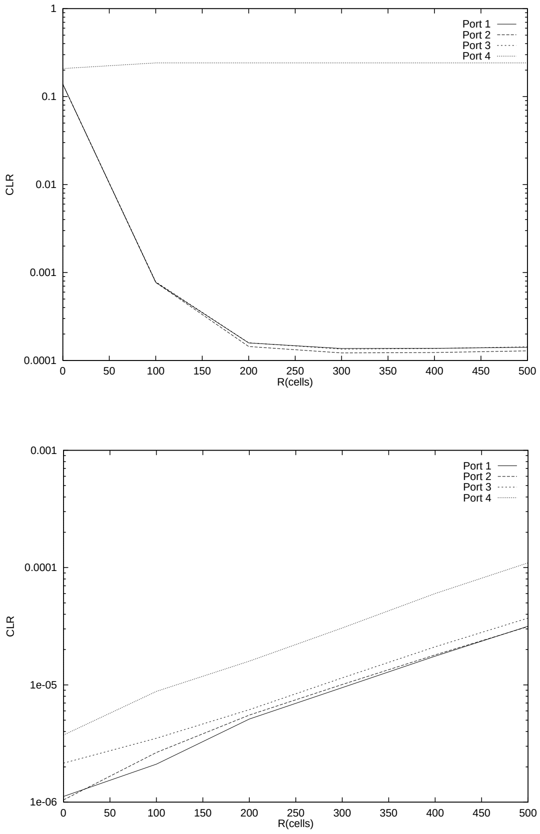


Fig. 9. Effect of rogue connections in port 4. See Sec.5.5. (a) $K_i = 90$ class 1 connections in each port, $B = 1333\ cells$, $B_i = 333\ cells$, $C_i = 15Mbps$. (b) $K_i = 54$ class 2 connections in each port, $B = 1333\ cells$, $B_i = 333\ cells$, $C_i = 15Mbps$.

and Minimum Bandwidth Guarantee. In *Proc. IEEE GLOBECOM'97*, pages 1081–1085, 1997.

[WMA95] G. L. Wu and J. W. Mark. A Buffer Allocation Scheme for ATM Networks: Complete Sharing Based on Virtual Partition. *IEEE/ACM Trans. Networking*, 3(6):660–670, Dec. 1995.

Appendix: Solution to the Fairshare Equation

Consider the equation

$$
\begin{aligned}
f(\beta) \quad = \quad & \sum_{i\in\mathcal{U}_0} (S_i^l + \beta S_i^{bg} - C_i)^+ + \sum_{i\in\mathcal{U}} (S_i^l + \beta S_i^{bg} - C_i) \\
& + \sum_{i\in\mathcal{O}} (S_i^l + \beta S_i^{bg} - C_i) = 0
\end{aligned}
$$

$f(\beta)$ is piece-wise linear, monotonic and convex, assuming all the parameters in the equation are positive. Assume also, that $f(0) < 0$ and $f(1) > 0$, as otherwise there is no valid solution $\beta \in [0, 1]$. The following algorithm solves for β by sorting the points of slope change and searching within each linear piece.

1. Start with $X = \sum_{i\in\mathcal{U},\mathcal{O}} (C_i - S_i^l)$, $Y = \sum_{i\in\mathcal{U},\mathcal{O}} S_i^{bg}$ and $\beta = X/Y$.

2. The points of slope change are given by

$$
R_i = \quad (C_i - S_i^l)/S_i^{bg} \quad i \in \mathcal{U}_0 \ .
$$

Sort the list of R_i's, denoted as $\mathcal{R}$, while making the same rearrangements to the list $\mathcal{U}_0$.

3. Starting with the minimum value in $\mathcal{R}$, if $\beta > R_i$, set

$$
\begin{aligned}
X \rightarrow X + (C_i - S_i^l), \qquad & Y \rightarrow Y + S_i^{bg} \\
i \rightarrow i + 1, \qquad & \beta \rightarrow X/Y.
\end{aligned}
$$

Else, terminate.

A variation of the above method can be employed to solve for β when one or more of the individual ports is simultaneously at nominal buffer occupancy. Then $f(\beta)$ has the form

$$
\begin{aligned}
f(\beta) \quad = \quad & \sum_{i\in\mathcal{U}_0} (S_i^l + \beta S_i^{bg} - C_i)^+ + \sum_{i\in\mathcal{U}} (S_i^l + \beta S_i^{bg} - C_i) \\
& + \sum_{i\in\mathcal{B}} [(S_i^l + \beta S_i^{bg1} - C_i)^+ \\
& + (S_i^l + S_i^{bg1} + S_i^{bp} + \beta S_i^{bg2} - C_i)^-] \\
& + \sum_{i\in\mathcal{O}} (S_i^l + \beta S_i^{bg} - C_i).
\end{aligned}
$$

$f(\beta)$, while being piece-wise linear and monotonic as before, is however no longer convex. We hence need to sort and scan through two arrays of turning points.

$$
\begin{aligned}
R_i^+ \quad &= \quad \begin{cases} (C_i - S_i^l)/S_i^{bg} & i \in \mathcal{U}_0 \\ (C_i - S_i^l)/S_i^{bg1} & i \in \mathcal{B} \end{cases} \\
R_i^- \quad &= \quad (C_i - S_i^l - S_i^{bg1} - S_i^{bp})/S_i^{bg2} \quad i \in \mathcal{B}
\end{aligned}
$$