Boosting the Network Performance via Traffic Reshaping

Sergey Gorinsky
Department of Computer Sciences
The University of Texas at Austin
Austin, TX 78712

Sanjoy Baruah*
Department of Computer Science
The University of Vermont
Burlington, VT 05405

Alexander Stoyen†
CIS Department
NJIT
Newark, NJ 07102

Abstract

Traffic reshaping and its impact on providing deterministic guarantees of timely data delivery in a packet-switched virtual-circuit fixed-packet network are investigated. Two types of traffic smoothing are considered: global reshaping (when traffic on all network connections is smoothed) and local reshaping (when the traffic specification is changed only for a single connection). The conditions when reshaping is beneficial are derived and the optimal values of traffic model parameters are obtained. In particular, it is shown that, when one tries to minimize end-to-end delay bounds for leaky bucket constrained traffic, the traffic either should be made Constant Bit Rate (CBR) or should not be reshaped at all. It is proven that changing the leaky bucket specification to the dual leaky bucket specification is always able to yield better timeliness guarantees and utilization of network resources. Finally, local reshaping for the dual leaky bucket model is studied.

1 Introduction

Emerging multimedia applications, such as video-on-demand, require timely delivery of data over wide-area computer networks. Since even very small losses of transferred information can be unacceptable for such applications, deterministic guarantees of network behavior became a subject of intensive research [8]. Admission control employed to assure providing requested guarantees to all established connections is greatly affected by the model used to describe traffic. It has been shown that sending bursty data worsens resource utilization and weakens timeliness guarantees. Traffic *re*shaping refers to making the traffic conform to a smoother specification than one supplied by the user. Such transformation inside the network has a potential to yield better performance guarantees. Knightly and Rossaro [7] demonstrate that reshaping cannot improve delays for one-hop connections, but are likely to be beneficial for multi-hop connections. Georgiadis et al [2] prove that reshaping allows Earliest Deadline First (EDF) scheduling discipline to provide better performance than Generalized Processor Sharing (GPS). Exact analysis is usually complicated even for relatively simple traffic models, and attempts to experimentally study an impact of reshaping are being undertaken [6].

This paper studies traffic reshaping in packet-switched virtual-circuit fixed-packet networks and is organized as follows. Section 2 describes the network model. Section 3 investigates reshaping for the leaky bucket traffic model. Section 4 examines transforming leaky bucket constrained traffic to conform to a dual leaky bucket specification. For the case of local reshaping in a network employing the dual leaky bucket model, the optimal values of the model parameters are derived. Section 5 summarizes the results and outlines future work.

2 The network model

We consider packet-switched virtual-circuit networks of an arbitrary topology where all packets are of the same size. Our model can be viewed as a generalization of ATM networks. Network nodes are interconnected by links which may have different bandwidths. Point-to-point communication is supported by means of connections: data are transferred from a source node to a destination node along some fixed route in the network. Each connection is characterized by its traffic specification and a requested end-to-end delay bound. Connection establishment takes place before transmission of user's data and succeeds only if requested delay bounds can be provided. Traffic entering a node is demultiplexed and shaped to conform to the connection specification, and each packet is transferred to the queue of an appropriate output link (or an application process at the destination node). Every queue is associated with a First-Come First-Served scheduler which decides in which order packets are sent over the

*Research is supported in part by the NSF Grant CCR-9796028.
†Research is supported in part by the U.S. ONR Grant N00014-93-1-1047, NSF Grant CCR-9402827, and DARPA Grant DAPT-63-96-C-0044.

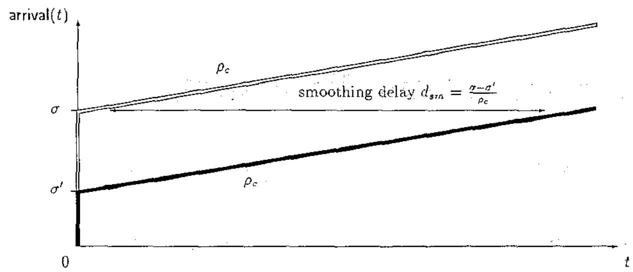

Figure 1: **Changing the traffic specification from (σ, ρ_c) to (σ', ρ_c).**

link. Using the same approach as in [2, 10], one can easily prove that shaping does not increase end-to-end delay bounds. Thus, shaping delays are ignored in the analysis below. We also disregard switching and alignment delays. A more detailed description of the network model can be found in [5].

3 The leaky bucket model

This section deals with a network where all connections are characterized by the leaky bucket model. Traffic on every connection is described through a pair of parameters (σ, ρ): a maximum burst of size σ is allowed over the maximum average rate ρ. The amount of data having arrived from the connection during any period of duration $t \geq 0$ can not exceed:

$$\mathsf{arrival}(t) = \sigma + \rho t.$$

First, consider the case when all connections are characterized by the same burstiness σ. Let us globally reshape traffic and bring the burstiness of each connection down to σ' where $\sigma \geq \sigma' \geq 0$ (see Figure 1). According to [5], the minimum end-to-end delay bound that can be guaranteed to connection C after the reshaping is given by:

$$d_c{}' = d_c + (\sigma - \sigma')\left(\frac{1}{\rho_c} - \sum_{k \in S} \frac{N_k}{l_k}\right)$$

where d_c is the minimum end-to-end delay bound before the reshaping; S is a set of links used by connection C; ρ_c is the maximum average rate of this connection; l_k is a bandwidth of link k, and N_k is the number of connections employing link k.

This result is somewhat surprising since it shows that the degree of reshaping , i.e. parameter $(\sigma - \sigma')$, does not affect the fact whether the reshaping decreases or increases the delay bound; this parameter just determines how much the bound is improved or worsened. The qualitative effect of reshaping is decided by the term $(\frac{1}{\rho_c} - \sum_{k \in S} \frac{N_k}{l_k})$. If

$$\rho_c \geq \frac{1}{\sum_{k \in S} \frac{N_k}{l_k}}, \qquad (1)$$

the reshaping is beneficial, and to obtain the minimum possible delay bound, one should minimize the value of allowed bursts, i.e. make $\sigma' = 0$. If

$$\rho_c < \frac{1}{\sum_{k \in S} \frac{N_k}{l_k}}, \qquad (2)$$

the reshaping increases the delay bound and therefore should not be applied.

To sum up, depending on network load, connection traffic either should be made Constant Bit Rate (CBR) or should not be reshaped at all.

Note that the global reshaping is advantageous for all the connections only if condition (1) holds for every connection. Otherwise, delay bounds for some connections are increased. For some applications, an increase of the end-to-end delay bound is acceptable as far as the bound does not exceed the value requested by the user. Such increase can be desirable, since reducing burstiness of a connection diminishes delays on other connections. To preserve the promised delay guarantees, it is necessary to ensure that for each connection C satisfying condition (2), the end-to-end delay bound d_c' is not larger than the requested bound D_c. According to [5], the minimum value that may be selected for σ' equals to:

$$\sigma' = \mathsf{door}\left(\max_{\rho_c < \frac{1}{\sum_{k \in S} \frac{N_k}{l_k}}} \left\{\frac{\frac{\sigma}{\rho_c} - D_c}{\frac{1}{\rho_c} - \sum_{k \in S} \frac{N_k}{l_k}}\right\}\right)$$

where door function is defined as follows: $\mathsf{door}(x) = x$ if $x \geq 0$ and $\mathsf{door}(x) = 0$ when $x < 0$.

Now, let different connections have different burstiness. Let us locally reshape traffic on connection C by reducing its burstiness from σ to σ', where $\sigma' \geq 0$, while keeping the burstiness of the other connections unchanged. According to [5], the the minimum end-to-end delay bound for connection C after the reshaping is equal to:

$$d_c{}' = d_c + (\sigma - \sigma')\left(\frac{1}{\rho_c} - \sum_{k \in S} \frac{1}{l_k}\right)$$

Once again, the qualitative result of reshaping does not depend on the degree of reshaping.

If $\rho_c \geq \frac{1}{\sum_{k \in S} \frac{1}{l_k}}$, the reshaping is advantageous, and to minimize the end-to-end delay bound, $\sigma' = 0$ should be selected, i.e., connection traffic should be transformed to conform to the CBR specification. Apart from improving the guarantees given to connection C, this will reduce delay bounds for the other connections by $\sigma \sum_{k \in S} \frac{1}{l_k}$.

If $\rho_c < \frac{1}{\sum_{k \in S} \frac{1}{l_k}}$, reshaping increases the delay bound and therefore should not be applied.

Thus, depending on the requested rate, amount and bandwidths of links constituting the virtual circuit, connection traffic either should be made CBR or should not be reshaped at all.

Sometimes, an increase of the end-to-end delay bound is acceptable as far as the bound does not exceed the value D_c requested by the user. Since reshaping improves performance guarantees for the rest of connections, such increase can be even desirable. According to [5], if connection C is such that $\rho_c < \frac{1}{\sum_{k\in S}\frac{1}{l_k}}$, then the minimum value that may be selected for σ' is:

$$\sigma' = \mathsf{door}\Big(\sigma - \frac{D_c - \sum_{k\in S}\big(\frac{1}{l_k}\sum_{m=1}^{N_k}\sigma_{k,m}\big)}{\frac{1}{\rho_c} - \sum_{k\in S}\frac{1}{l_k}}\Big).$$

4 The dual leaky bucket model

When the constant bit rate service cannot guarantee acceptable end-to-end delays, reshaping should not be applied to leaky bucket restrained traffic. It indicates limited usefulness of the reshaping. This discouraging conclusion is not necessary true for other traffic models. Below we study an impact of reshaping when a multirate model is employed. A multirate specification was shown to be useful for efficient description of VBR data [9]. The corresponding shaping of traffic can be performed by a series of leaky buckets [1]. We consider the dual leaky bucket model which was accepted as a standard by the ATM Forum [3]: traffic is described through a tuple of parameters $(\sigma, \rho', \rho, \tau)$ where σ is the size of the maximum instantaneous burst on the connection, ρ' is the maximum peak rate, ρ is the maximum average rate ($\rho \le \rho'$), and τ is the maximum peak duration. The amount of data having arrived from the connection during any period of length $t \ge 0$ is at most:

$$\mathsf{arrival}(t) = \begin{cases} \sigma + \rho' t & \text{if } 0 \le t \le \tau, \\ \sigma + \rho'\tau + \rho(t-\tau) & \text{otherwise.} \end{cases}$$

First, we consider a network where all connections are characterized by the leaky bucket model with the same burstiness σ and possibly different rates ρ.

Let us globally reshape traffic so that the traffic specification for every connection is changed from (σ, ρ) to $(\sigma', \rho', \rho, \frac{\sigma-\sigma'}{\rho'})$ dual leaky bucket specification where $\sigma \ge \sigma' \ge 0$ and $\rho' \ge \rho$ (see Figure 2).

According to the schedulability conditions given in [4], the minimum end-to-end delay bound that can be guaranteed to a connection C without reshaping equals to:

$$d_c = \Big(\sum_{k\in S}\frac{N_k}{l_k}\Big)\sigma. \qquad (3)$$

When the reshaping is applied, delay of sending data over link k cannot exceed:

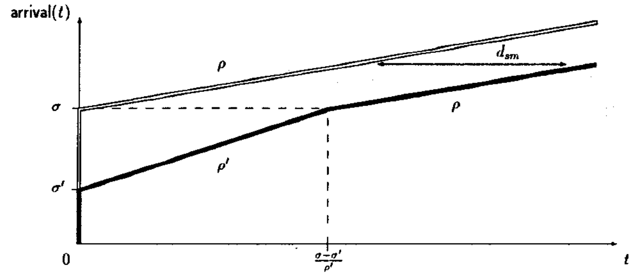

Figure 2: **Changing the traffic specification from (σ, ρ) to $(\sigma', \rho', \rho, \frac{\sigma-\sigma'}{\rho'})$.**

$$d'_k = \max\Big\{\frac{N_k\sigma'}{l_k}, \frac{N_k\sigma}{l_k} - \frac{\sigma-\sigma'}{\rho'}\Big\}.$$

Smoothing delay at the entrance to the network is limited by

$$d_{sm} = \frac{\sigma-\sigma'}{\rho'}.$$

Therefore, the minimum end-to-end delay bound becomes:

$$\begin{aligned} d_c{}' &= d_{sm} + \sum_{k\in S} d'_k \\ &= \frac{\sigma-\sigma'}{\rho'} + \sum_{k\in S}\max\Big\{\frac{N_k\sigma'}{l_k}, \frac{N_k\sigma}{l_k} - \frac{\sigma-\sigma'}{\rho'}\Big\}. \end{aligned}$$

Taking (3) into account, we can rewrite this expression as:

$$d_c{}' = d_c - \frac{\sigma-\sigma'}{\rho'}\Big(\sum_{k\in S}\min\Big\{\frac{N_k\rho'}{l_k}, 1\Big\} - 1\Big). \qquad (4)$$

Formula (4) shows that the qualitative result of reshaping is not affected by the value of the maximum burst size σ'. When reshaping is beneficial, the parameter should be minimized, i.e. $\sigma' = 0$ should be chosen.

Without loss of generality, let us assume that $\rho'_{min} = \min_{k\in S}\frac{l_k}{N_k} = \frac{l_n}{M_n}$ for link n. If $\rho' = \rho'_{min}$, then $\min\Big\{\frac{M_n\rho'}{l_n}, 1\Big\} = \frac{M_n\rho'}{l_n} = 1$ and the end-to-end delay is limited by:

$$\begin{aligned} d_{min} &= d_c - \frac{\sigma-\sigma'}{\rho'_{min}}\Big(\sum_{k\in S}\frac{N_k\rho'_{min}}{l_k} - 1\Big) \\ &= d_c - (\sigma-\sigma')\sum_{k\,\in\, S-\{n\}}\frac{N_k}{l_k} \le d_c. \end{aligned}$$

If $\rho' \ge \rho'_{min}$, the end-to-end delay bound equals to:

$$\begin{aligned} d_c{}' &= d_c - \frac{\sigma-\sigma'}{\rho'}\Big(\sum_{k\in S}\min\Big\{\frac{N_k\rho'}{l_k}, 1\Big\} - 1\Big) \\ &= d_c - \frac{\sigma-\sigma'}{\rho'}\sum_{k\,\in\, S-\{n\}}\min\Big\{\frac{N_k\rho'}{l_k}, 1\Big\}. \end{aligned}$$

Since $d_c - \frac{\sigma-\sigma'}{\rho'}\sum_{k\,\in\, S-\{n\}}\min\Big\{\frac{N_k\rho'}{l_k}, 1\Big\} \ge d_c - (\sigma - \sigma')\sum_{k\,\in\, S-\{n\}}\frac{N_k}{l_k}$, we conclude that $d'_c \ge d_{min}$. In general, function $d'_c(\rho')$ is nondecreasing for $\rho' \ge \rho'_{min}$.

Therefore, making the peak rate larger than ρ'_{min} does not yield a decrease in the delay bound.

On the other hand, when $\rho' < \rho'_{min}$, the end-to-end delay bound equals:

$$\begin{aligned} d_c{}' &= d_c - (\sigma - \sigma')(\sum_{k \in S} \frac{N_k}{l_k} - \frac{1}{\rho'}) \\ &> d_c - (\sigma - \sigma')(\sum_{k \in S} \frac{N_k}{l_k} - \frac{1}{\rho'_{min}}) = d_{min}. \end{aligned}$$

Thus, reducing ρ' below ρ'_{min} is not advantageous also.

Taking into account that $\rho' \geq \rho$, we can conclude that

$$\rho'_{opt} = \max\left\{\rho,\ \min_{k \in S} \frac{l_k}{N_k}\right\} \tag{5}$$

is the optimal value for the peak rate.

Therefore, the traffic specification should be changed from (σ, ρ) to $(0, \rho'_{opt}, \rho, \frac{\sigma}{\rho'_{opt}})$ where $\rho'_{opt} = \max\left\{\rho,\ \min_{k \in S} \frac{l_k}{N_k}\right\}$.

Example 1. Let network traffic be described using the leaky bucket specification with burstiness $\sigma =$ 100 KB. Consider a connection C that has the maximum average rate $\rho = 3\ \frac{Mb}{s}$ and traverses five links. The link bandwidths and numbers of connections employing these links are given in Figure 3. According to (3), the end-to-end delay is bounded by $d_c = 233\ ms$. Reshaping the traffic to another leaky bucket specification is not advantageous. For instance, reshaping to the CBR specification increases the end-to-end delay bound to $d_{\text{CBR}} = 267\ ms$. At the same time, making the traffic conform to a dual leaky bucket specification can be beneficial. According to (4) and (5), the end-to-end delay bound reaches its minimal value $d_{min} = 80\ ms$ when $\rho' = 10\ \frac{Mb}{s}$. In other words, changing the connection specification from (100 KB, 3 $\frac{Mb}{s}$) to (0, 10 $\frac{Mb}{s}$, 3 $\frac{Mb}{s}$, 80 ms) reduces the minimum end-to-end delay bound almost thrice (see Figure 3).

Now, consider a network where all connections are characterized by the dual leaky bucket model. We allow different connections to possess different burstiness σ, peak rates ρ', and average rates ρ. At the same time, one value of the maximum peak duration τ is shared by all connections.

Let us reshape traffic on connection C by reducing its burstiness from σ to σ^*, where $\sigma^* \geq 0$, and decreasing the peak rate from ρ' to ρ^*, where $\rho^* \geq \rho$, (see Figure 4) while keeping the parameters of the other connections unchanged.

According to [4], the minimum delay bound that can be guaranteed over link k without reshaping is equal to:

(a)

Link number, k	1	2	3	4	5
Link bandwidth, l_k, $\frac{Mb}{s}$	60	100	155	100	40
Number of connections using the link, N_k	2	10	12	3	2
$\frac{l_k}{N_k}$, $\frac{Mb}{s}$	30	10	12.9	33.3	20

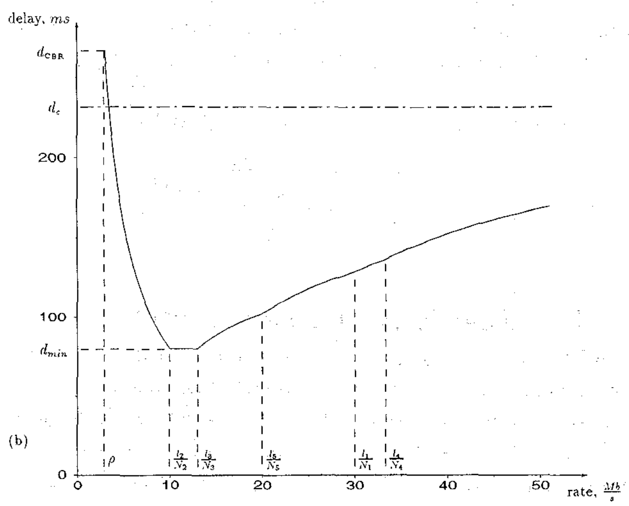

Figure 3: **Global reshaping from (σ, ρ) traffic specification to $(0, \rho', \rho, \frac{\sigma}{\rho'})$: (a) information about the links forming connection C; (b) the minimum end-to-end delay bound $d_c{}'$ as a function of the peak rate ρ'.**

$$\begin{aligned} d_k &= \max\left\{\frac{1}{l_k}\sum_{m=1}^{N_k} \sigma_{k,m}, \frac{1}{l_k}\sum_{m=1}^{N_k}(\sigma_{k,m} + \rho'_{k,m}\tau) - \tau\right\} \\ &= \frac{\sum_{m=1}^{N_k} \sigma_{k,m}}{l_k} + \tau \text{door}(\frac{\sum_{m=1}^{N_k} \rho'_{k,m}}{l_k} - 1) \end{aligned}$$

where $\sigma_{k,m}$ and $\rho'_{k,m}$ are the maximum burst and the maximum peak rate of the m-th connection on link k.

Then the end-to-end delay for connection C without reshaping is bounded by:

$$\begin{aligned} d_c &= \sum_{k \in S} d_k \\ &= \sum_{k \in S} \Big(\frac{\sum_{m=1}^{N_k} \sigma_{k,m}}{l_k} + \tau \text{door}(\frac{\sum_{m=1}^{N_k} \rho'_{k,m}}{l_k} - 1) \Big). \end{aligned} \tag{6}$$

When reshaping is employed, the delay of sending data over link k is limited by:

$$\begin{aligned} d_k^* &= \frac{\sum_{m=1}^{N_k} \sigma_{k,m}}{l_k} - \frac{\sigma - \sigma^*}{l_k} \\ &+ \tau \text{door}(\frac{\sum_{m=1}^{N_k} \rho'_{k,m}}{l_k} - \frac{\rho' - \rho^*}{l_k} - 1). \end{aligned}$$

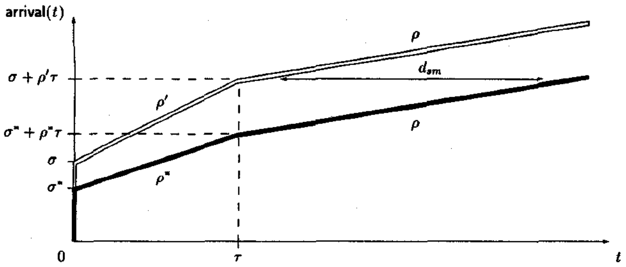

Figure 4: **Changing the traffic specification from $(\sigma, \rho', \rho, \tau)$ to $(\sigma^*, \rho^*, \rho, \tau)$.**

Smoothing delay at the entrance to the network cannot exceed:

$$d_{sm} = \frac{\sigma - \sigma^*}{\rho} + \frac{\rho' - \rho^*}{\rho}\tau.$$

Thus, the minimum end-to-end delay bound with reshaping is:

$$d_c{}^* = d_{sm} + \sum_{k\in S} d_k^*$$

$$= \frac{\sigma - \sigma^*}{\rho} + \frac{\rho' - \rho^*}{\rho}\tau + \sum_{k\in S}\Big(\frac{\sum_{m=1}^{N_k}\sigma_{k,m}}{l_k} - \frac{\sigma - \sigma^*}{l_k} + \tau \mathsf{door}\big(\frac{\sum_{m=1}^{N_k}\rho'_{k,m}}{l_k} - \frac{\rho' - \rho^*}{l_k} - 1\big)\Big).$$

Taking (6) into account, we can rewrite this expression as:

$$d_c{}^* = d_c - (\sigma - \sigma^*)\big(\sum_{k\in S}\frac{1}{l_k} - \frac{1}{\rho}\big) - \tau\Big(\sum_{k\in S}\Big(\mathsf{door}\big(\frac{\sum_{m=1}^{N_k}\rho'_{k,m}}{l_k} - 1\big) - \mathsf{door}\big(\frac{\sum_{m=1}^{N_k}\rho'_{k,m}}{l_k} - \frac{\rho' - \rho^*}{l_k} - 1\big)\Big) - \frac{\rho' - \rho^*}{\rho}\Big). \quad (7)$$

Let us divide set S of connection links into subsets S_1, S_2, and S_3:

- $l_k \geq \sum_{m=1}^{N_k}\rho'_{k,m}$ for each link k from set S_1,
- $\sum_{m=1}^{N_k}\rho'_{k,m} - (\rho' - \rho^*) < l_k < \sum_{m=1}^{N_k}\rho'_{k,m}$ for each link k from set S_2,
- $l_k \leq \sum_{m=1}^{N_k}\rho'_{k,m} - (\rho' - \rho^*)$ for each link k from set S_3.

Such division implies that:

$$\begin{cases} \rho^* < \rho' - \max_{k\in S_2}\Big\{\sum_{m=1}^{N_k}\rho'_{k,m} - l_k\Big\} \\ \rho^* \geq \rho' - \min_{k\in S_3}\Big\{\sum_{m=1}^{N_k}\rho'_{k,m} - l_k\Big\} \end{cases} \quad (8)$$

Then expression (7) can be rewritten as:

$$d_c^* = d_c - (\sigma - \sigma^*)\big(\sum_{k\in S}\frac{1}{l_k} - \frac{1}{\rho}\big) - (\rho' - \rho^*)\tau\big(\sum_{k\in S_3}\frac{1}{l_k} - \frac{1}{\rho}\big) - \tau\sum_{k\in S_2}\big(\frac{\sum_{m=1}^{N_k}\rho'_{k,m}}{l_k} - 1\big).$$

If $\rho \geq \frac{1}{\sum_{k\in S}\frac{1}{l_k}}$, decreasing the burstiness of connection C is advantageous and $\sigma^* = 0$ should be chosen. If $\rho < \frac{1}{\sum_{k\in S}\frac{1}{l_k}}$, the burstiness should not be decreased, i.e. $\sigma^* = \sigma$ should be selected.

If $\rho \geq \frac{1}{\sum_{k\in S_3}\frac{1}{l_k}}$, reducing the peak rate as much as possible minimizes the end-to-end delay. Taking (8) into account, one should choose $\rho^* = \max\Big\{\rho,\ \rho' - \min_{k\in S_3}\Big\{\sum_{m=1}^{N_k}\rho'_{k,m} - l_k\Big\}\Big\}$. If $\rho < \frac{1}{\sum_{k\in S_3}\frac{1}{l_k}}$, decreasing the peak rate does not reduce the end-to-end delay bound. According to (8), $\rho^* = \max\Big\{\rho,\ \rho' - \max_{k\in S_2}\Big\{\sum_{m=1}^{N_k}\rho'_{k,m} - l_k\Big\}\Big\}$ should be selected[1]. As a result, $\rho^* = \rho'$ should be chosen when $\rho < \frac{1}{\sum_{k\in S_2\cup S_3}\frac{1}{l_k}}$.

Finding the optimal peak rate ρ^* includes determination of the best split of the connection links between sets S_2 and S_3. To accomplish it, we number the links in $S_2 \cup S_3$ in the decreasing order of $(\sum_{m=1}^{N_k}\rho'_{k,m} - l_k)$ and detect the minimum q such that $\rho \geq \frac{1}{\sum_{k=1}^{q}\frac{1}{l_k}}$. According to the previous paragraph, $\rho^* = \max\Big\{\rho,\ \rho' - (\sum_{m=1}^{N_q}\rho'_{q,m} - l_q)\Big\}$ should be selected as the peak rate.

To sum it up, reshaping should not be applied if $\rho < \frac{1}{\sum_{k\in S}\frac{1}{l_k}}$. When $\rho \geq \frac{1}{\sum_{k\in S-S_1}\frac{1}{l_k}}$, the traffic specification should be changed from $(\sigma, \rho', \rho, \tau)$ to $(0, \max\Big\{\rho,\ \rho' - (\sum_{m=1}^{N_q}\rho'_{q,m} - l_q)\Big\}, \rho, \tau)$. Otherwise, traffic should be reshaped to conform to $(0, \rho', \rho, \tau)$ model.

Example 2. Consider a network where all connections are characterized by the dual leaky bucket model. Let us locally reshape the traffic on a connection C which uses (100 KB, 10 $\frac{Mb}{s}$, 3 $\frac{Mb}{s}$, 80 ms) specification and consists of five links. The bandwidths, ag-

[1] If set S_2 is empty, then $\rho^* = \rho'$ is selected.

(a)

Link bandwidth, l_k, $\frac{Mb}{s}$	5	8	10	12	8
Aggregate burstiness, $\sum_{m=1}^{N_k} \sigma_{k,m}$, KB	100	100	100	200	100
Aggregate peak rate, $\sum_{m=1}^{N_k} \rho'_{k,m}$, $\frac{Mb}{s}$	11	13	11	11	11
Link set	$S - S_1$	$S - S_1$	$S - S_1$	S_1	$S - S_1$
$\sum_{m=1}^{N_k} \rho'_{k,m} - l_k$, $\frac{Mb}{s}$	6	5	1	-1	3
Link number, k	1	2	4	—	3
$\frac{1}{\sum_{j=1}^{k} \frac{1}{l_j}}$, $\frac{Mb}{s}$	5.0	3.1	1.8	—	2.2
$r_k = \max\left\{\rho,\ \rho' - (\sum_{m=1}^{N_k} \rho'_{k,m} - l_k)\right\}$, $\frac{Mb}{s}$	4	5	9	—	7

(b)

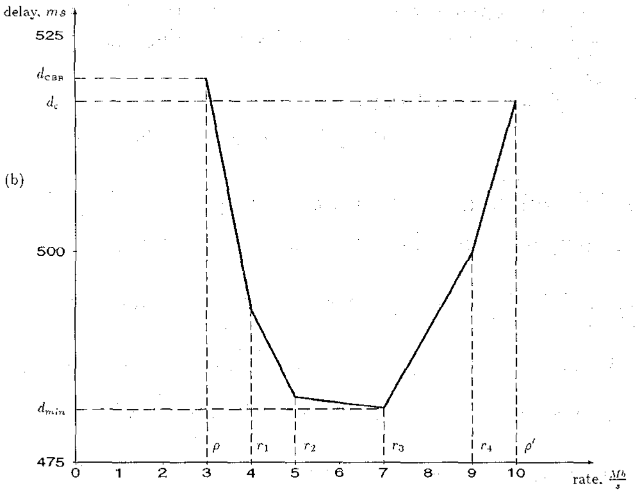

Figure 5: **Local reshaping for the dual leaky bucket model: (a) information about the links traversed by connection C; (b) the minimum end-to-end delay bound $d_c{}^*$ as a function of the peak rate ρ^*.**

gregate burstiness, and aggregate peak rates of these links are given in Figure 5. According to (6), the end-to-end delay for connection C without reshaping is bounded by $d_c = 517\ ms$. Reshaping the traffic to to the CBR specification increases the end-to-end delay bound to $d_{\text{CBR}} = 520\ ms$. The proposed algorithm allows us to detect that the end-to-end delay bound reaches its minimal value $d_{min} = 481\ ms$ when $\sigma^* = 0$ and $\rho^* = 7\ \frac{Mb}{s}$ (see Figure 5). In other words, the traffic on the connection C should be reshaped to conform to $(0,\ 7\ \frac{Mb}{s},\ 3\ \frac{Mb}{s},\ 80\ ms)$ specification.

5 Conclusions

This paper studied an impact of traffic reshaping inside the network on providing deterministic guarantees of timely data delivery in packet-switched virtual-circuit fixed-packet networks.

We considered global and local reshaping in networks where all connections were characterized by the leaky bucket model. If end-to-end delay bounds should be minimized, then, depending on network load, traffic either should be made CBR or should not be reshaped at all. For the case when reshaping increased the end-to-end delay bounds, we obtained the minimum burstiness which preserved the guarantees requested by the user.

We proved that global reshaping of leaky bucket constrained traffic to a dual leaky bucket specification is always able to reduce the end-to-end delay bounds. For the case of local reshaping in a network employing the dual leaky bucket model, the optimal values of the model parameters were derived.

While transforming one dual leaky bucket specification to another, we did not permit the peak rate to exceed its initial value. Section 4 indicated that increasing the peak rate can yield better timeliness guarantees. In future, we plan to consider the general case when the peak rate ρ^* is allowed to take values between ρ and $\rho' + \frac{\sigma - \sigma^*}{\tau}$.

We are also going to study traffic reshaping in networks employing other scheduling disciplines.

References

[1] R.L. Cruz, "A Calculus for Network Delay, Part II: Network Analysis," *IEEE Transactions on Information Theory,* 37(1):132-141, January 1991.

[2] L. Georgiadis, R. Guerin, V. Peris, and K.N. Sivarajan, "Efficient Network QoS Provisioning Based on Per Node Traffic Shaping," *IEEE/ACM Transactions on Networking,* 4(4):482-501, August 1996.

[3] W. Goralski, *Introduction to ATM Networking,* McGraw-Hill, 1995.

[4] S. Gorinsky, S. Baruah, T. J. Marlowe, and A. D. Stoyenko, "Exact and Efficient Analysis of Schedulability in Fixed-Packet Networks: A Generic Approach," *Proceedings of INFOCOM'97,* pp. 585-592, April 1997.

[5] S. Gorinsky, S. Baruah, A. Stoyenko, "Traffic Reshaping in Packet-Switched Virtual-Circuit Fixed-Packet Networks," *Proceedings of Workshop on Resource Allocation Problems in Multimedia Systems,* December 1996.

[6] R. Guerin, *Personal correspondence,* September 1996.

[7] E. Knightly and P. Rossaro, "Improving QoS through Traffic Smoothing," *Proceedings of IFIP Workshop on Quality of Service*, March 1996.

[8] C. Partridge, *Gigabit Networking,* Addison-Wesley, 1993.

[9] D. Saha, S. Mukherjee, S. Tripathi, "Multirate Scheduling for Guaranteed and Predictive Services in ATM Networks," *Proceedings of RTSS'96,* pp. 155-164, December 1996.

[10] H. Zhang and D. Ferrari, "Rate-Controlled Service Disciplines," *Journal of High Speed Networks,* 3(4):389-412, 1994.