TECHNICAL
RESEARCH
REPORT

SYSTEMS
RESEARCH
CENTER

Supported by the
National Science Foundation
Engineering Research Center
Program (NSFD CD 8803012),
Industry and the University

Monotonicity Properties of the Leaky Bucket

by L. Kuang

TR 92-27

CDC31

MONOTONICITY PROPERTIES OF THE LEAKY BUCKET

Lei Kuang[1]

Department of Electrical Engineering
and Systems Research Center
University of Maryland at College Park
College Park, MD 20742

March 3, 1992

[1]Supported by NSF grants NSFD CDR-88-03012 and NCR-88-14566-02.

Abstract In this paper, we present monotonicity properties of the Leaky Bucket (LB) input rate regulation scheme. We show that the asymptotic version of the inter-departure time decreases in the convex ordering as the size of the token pool decreases or as the token generation period increases. When measured by the coefficients of variations of the inter-departure times, the burstiness of the output traffic from the LB increases with the size of the token pool, and decreases with the token generation period. These results may prove useful in designing the LB.

1 Introduction

Congestion control in high speed networks is one of the challenges raised by the emerging of the Broadband ISDN. There are two kinds of congestion controls techniques: reactive congestion control and preventive congestion control. Reactive congestion control has been successfully used in conventional networks such as X.25 packet switching networks, where appropriate actions are taken whenever congestion is detected inside the network. It is widely accepted, however, that preventive congestion control is more suitable for high speed networks. Because of the high speed, actions taken after congestion occurs may already be too late to avoid performance degradation. Instead, precautions must be taken to prevent congestion from arising, and the Leaky Bucket (LB) input rate regulation scheme is one of the most prominent preventive congestion control methods. The original LB was proposed in [10] for monitoring the traffic into the network and keeping its rate within a certain range. It has, however, little effect on reducing the burstiness of the traffic. To increase the LB's ability of reducing traffic burstiness, Eckberg et al [4] proposed a variation of the LB in [10] by adding an input buffer to it.

A lot of studies have been done for finding numerically tractable solutions to assess the performance of the LB [2, 1, 8]. Fluid models have been used to get approximation results [3]. Structural properties, however, have received little attention.

In this paper, we study the monotonicity properties of the LB with respect to various parameters using a method proposed in [6], where the author showed that the output traffic from the LB is less bursty than the input traffic. It turns out that the results in [6] can be viewed as a special case of the results presented here.

We shall describe the LB in detail in section 2, and present the main results in section 3. We then sketch the proof in section 4 by exploring the burst structure of the output traffic from the LB.

2 The Leaky Bucket

The LB studied in this paper is composed of an input buffer, a token pool, and a mechanism for generating tokens into the token pool at a constant rate (Fig. 1). If the capacity of the token pool is reached while a new token is generated, the newly generated token is discarded. Cells (i.e., short, fixed

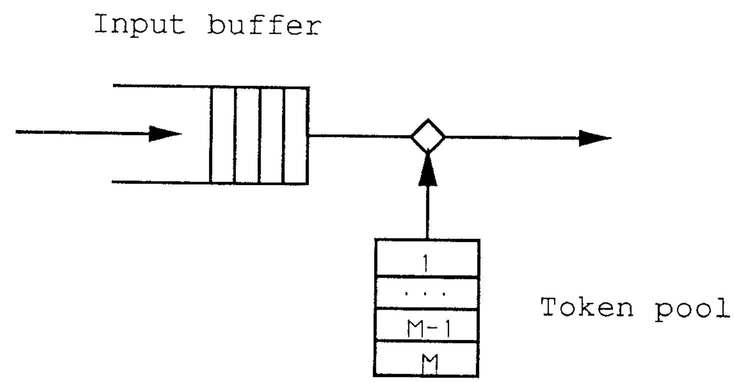

Figure 1: The leaky bucket.

length packets) must first obtain tokens from the LB before entering the network. Available tokens will be distributed to the cells in a FCFS manner. Cells which have obtained tokens enter the network instantaneously, and consume the tokens, whereas cells which cannot get a token upon arrival have to wait in the input buffer. In this paper, we only consider the case where the input buffer has infinite capacity. So there will be no cell loss in the LB.

A LB can then be characterized by two parameters M and D, where M is the size of the token pool, and D is the token generation period which is assumed constant. Tokens are generated at times $\{kD, k = 0, 1, \cdots\}$. Cells are tagged upon arrival in the order of their arrival, and we assume that the first cell arriving at time $t = 0$ finds an empty buffer. These assumptions are made only for notational convenience and do not affect the results obtained in this paper. A LB with parameters D and M is denoted by $LB(D, M)$.

The output traffic from the LB can be easily described by a sequence of

$\mathbb{R}_+$-valued random variables (rvs) $\boldsymbol{\delta} \triangleq \{\delta_n, n = 1, 2, \cdots\}$. We interpret δ_1 as the first departure epoch, and for $n = 1, 2, \cdots$, we interpret δ_{n+1} as the inter-departure time between the n^{th} and the $(n+1)^{st}$ cells. The purpose of this paper is to establish relations between the parameters of the LB and the burstiness of this sequence.

Before presenting our main results, we need the following notion of stability. For any $\mathbb{R}_+$-valued sequence of rvs $\boldsymbol{\zeta} \triangleq \{\zeta_n, n = 1, 2, \cdots\}$, we say that $\boldsymbol{\zeta}$ is *convexly stable* if there exists an integrable $\mathbb{R}_+$-valued rv ζ (i.e., $\mathbb{E}[\zeta] < \infty$) such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \phi(\zeta_i) = \mathbb{E}[\phi(\zeta)] \qquad a.s. \tag{1}$$

for any convex function $\phi : \mathbb{R} \to \mathbb{R}$, in which case we call ζ the asymptotic version of $\boldsymbol{\zeta}$. Note that the convex stability of $\boldsymbol{\zeta}$ implies

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \zeta_i = \mathbb{E}[\zeta] \qquad a.s. \tag{2}$$

In particular if $\boldsymbol{\zeta}$ represents the inter-arrival times, then under the assumption of convex stability, the (long-run) cell arrival rate λ is well defined by $1/\,\mathbb{E}[\zeta]$. We shall say that a LB $LB(D, M)$ is *stable* to the input traffic of rate λ, if

$$\lambda D < 1. \tag{3}$$

In this paper we consider the case where the LBs are stable, although the intermediate transient results hold without this condition.

The main results of the paper are the asymptotic monotonicity of the inter-departure times with respect to the parameters of the LB in the sense of convex ordering. We recall [9] that an $\mathbb{R}$-valued rv X is smaller than another $\mathbb{R}$-valued rv Y in the sense of convex ordering, and we write $X \leq_{cx} Y$, if

$$\mathbb{E}[\phi(X)] \leq \mathbb{E}[\phi(Y)] \tag{4}$$

for all convex functions $\phi : \mathbb{R} \to \mathbb{R}$ whenever the expectations exist in (4). Note that for any $\mathbb{R}_+$-valued rvs all moments exist (though possibly infinite), so that $X \leq_{cx} Y$ implies

$$\mathbb{E}[X] = \mathbb{E}[Y]$$

and

$$\mathbb{E}[X^k] \leq \mathbb{E}[Y^k], \qquad k = 1, 2, \cdots.$$

Consequently,

$$c^2(X) \leq c^2(Y),$$

where $c^2(X)$ is the coefficient of variation of the $\mathbb{R}_+$-valued rv X which is defined by

$$c^2(X) \triangleq \frac{Var(X)}{(\mathbb{E}[X])^2}.$$

We are now ready to present our main results.

3 Monotonicity in D and M

From now on we assume that all LBs we are considering are fed with the same input traffic of rate λ. We first consider the monotonicity of the inter-departure times with respect to M. As introduced in section 2, $\boldsymbol{\delta} = \{\delta_n, n = 1, 2, \cdots\}$ denotes the sequence of inter-departure times from $LB(M, D)$, with $M \geq 0$ and $D \geq 0$. For $0 \leq \hat{M} \leq M$, let $\hat{\boldsymbol{\delta}} = \{\hat{\delta}_n, n = 1, 2, \cdots\}$ denote the sequence of inter-departure times from $LB(D, \hat{M})$. Then we have our first theorem.

Theorem 3.1 *(Monotonicity in M) Assuming $\boldsymbol{\delta}$ and $\hat{\boldsymbol{\delta}}$ to be convexly stable and (3) to hold, we have*

$$\hat{\delta} \leq_{cx} \delta, \tag{5}$$

where δ and $\hat{\delta}$ are the asymptotic versions of $\boldsymbol{\delta}$ and $\hat{\boldsymbol{\delta}}$, respectively, and therefore, $c^2(\hat{\delta}) \leq c^2(\delta)$.

Theorem 3.1 says that the smaller the token pool size is, the less bursty the output traffic from the LB will be as would be expected.

The monotonicity of the inter-departure times with respect to D is considered next. We first assume $M = \infty$. For $0 \leq \tilde{D} \leq D$, let $\tilde{\boldsymbol{\delta}} = \{\tilde{\delta}_n, n = 1, 2, \cdots\}$ denote the sequence of inter-departure times from $LB(\tilde{D}, M)$. In close parallel with the monotonicity results in M, we have the following

Theorem 3.2 *(Monotonicity in D) Assuming* $\boldsymbol{\delta}$ *and* $\tilde{\boldsymbol{\delta}}$ *to be convexly stable and (3) to hold, we have*

$$\delta \leq_{cx} \tilde{\delta}, \tag{6}$$

where δ *and* $\tilde{\delta}$ *are the asymptotic versions of* $\boldsymbol{\delta}$ *and* $\tilde{\boldsymbol{\delta}}$, *respectively, and therefore,* $c^2(\delta) \leq c^2(\tilde{\delta})$.

So in the case where the token pool has infinite capacity, the larger the token generation period is, the less bursty the output traffic from the LB will be.

The assumption $M = \infty$ in Theorem 3.2 can be relaxed. In fact, the sample path property leading to Theorem 3.2 will be retained as long as the tilde LB can hold the extra tokens it generates due to its faster token generation rate. Combining the previous results and the comments just made, we have the following results.

Let $\bar{\boldsymbol{\delta}} = \{\bar{\delta}_n, n = 1, 2, \cdots\}$ denote the sequences of inter-departure times from $LB(\bar{D}, \bar{M})$, with $0 \leq \bar{D} \leq D$ and $0 \leq M < \bar{M}$, then we have

Theorem 3.3 *(Monotonicity in D and M) Assuming* $\boldsymbol{\delta}$ *and* $\bar{\boldsymbol{\delta}}$ *to be convexly stable and (3) to hold, then we have*

$$\delta \leq_{cx} \bar{\delta}, \tag{7}$$

where δ *and* $\bar{\delta}$ *are the asymptotic versions of* $\boldsymbol{\delta}$ *and* $\bar{\boldsymbol{\delta}}$, *respectively, and therefore,* $c^2(\delta) \leq c^2(\bar{\delta})$.

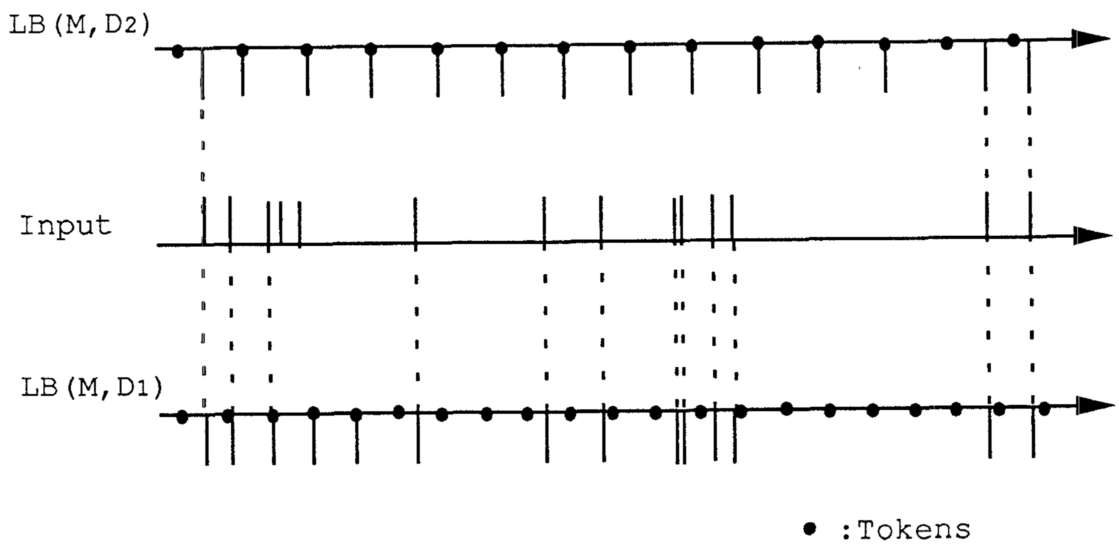

Figure 2: Burst structures of the output traffic from two LBs with different token generation periods.

Taking $\bar{D} = 0$, it is easily seen that $\bar{\boldsymbol{\delta}}$ is just the inter-arrival times to the LB, thus the results obtained in [6] is simply a special case of theorem 3.3.

4 Discussions

The results presented in section 3 are direct consequences of the burst structure of the traffic which can be characterized by means of majorization [7]. An example of this burst structure is depicted in Fig. 2. Full details are available in [5].

For any vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ in $\mathbb{R}^n$ we denote by $x_{[i]}$ the i^{th} largest element of $\boldsymbol{x}$, $i = 1, \cdots, n$, i.e.,

$$x_{[1]} \geq x_{[2]} \geq \cdots \geq x_{[n]}. \tag{8}$$

For vectors $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$ such that

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \tag{9}$$

we say that $\boldsymbol{x}$ *majorizes* $\boldsymbol{y}$, and write $\boldsymbol{y} \prec \boldsymbol{x}$, if

$$\sum_{i=k}^{n} y_{[i]} \geq \sum_{i=k}^{n} x_{[i]}, \qquad k = 1, \cdots, n. \tag{10}$$

To simplify the notation, we use the following convention. For any sequence $\boldsymbol{\xi} = \{\xi_n, n = 1, 2, \cdots\}$ of $\mathbb{R}_+$-valued rvs, we define the $\mathbb{R}_+^{m-n+1}$-valued rvs $\boldsymbol{\xi}_{m,n}$, $m \leq n, n = 1, 2, \cdots$ by

$$\boldsymbol{\xi}_{m,n} \triangleq (\xi_m, \cdots, \xi_n).$$

Let $Z^{(1)}$ be the collection of tags of cells which obtain tokens immediately upon their arrival. The following theorem establishes the transient orderings of the sequences of inter-departure times from LBs studied in section 3.

Theorem 4.1 *The following results hold.*

(i) *For any $n \in \hat{Z}^{(1)}$ with respect to $LB(D, \hat{M})$, we have*

$$\hat{\boldsymbol{\delta}}_{1,n} \prec \boldsymbol{\delta}_{1,n}. \tag{11}$$

(ii) *For any $n \in Z^{(1)}$ with respect to $LB(D, \infty)$, we have*

$$\boldsymbol{\delta}_{1,n} \prec \hat{\boldsymbol{\delta}}_{1,n}. \tag{12}$$

(iii) *For any* $n \in Z^{(1)}$ *with respect to* $LB(D, M)$, *we have*

$$\boldsymbol{\delta}_{1,n} \prec \bar{\boldsymbol{\delta}}_{1,n}. \tag{13}$$

To see why the theorem might be true, we focus on (i) for instance: If $n \in \hat{Z}^{(1)}$, then $n \in Z^{(1)}$, because $LB(D, M)$ will never have fewer tokens than $LB(D, \hat{M})$. Since cells with tags in $\hat{Z}^{(1)}$ leave the LB instantaneously, they have the same departure epochs in both LBs. Thus $\hat{Z}^{(1)}$ induces a partition on both sequences $\hat{\boldsymbol{\delta}}$ and $\boldsymbol{\delta}$. We showed in [6] that the majorization ordering holds for each corresponding segments of such partition of both sequences. So (i) follows from the closure property of majorization under concatenation [7, Proposition 5.A.7, p. 121]. Similar comments apply for (ii) and (iii).

Applying Proposition 4.B.1 in [7, p. 108], Theorem 4.1 then implies

Theorem 4.2

(i) *For any* $n \in \hat{Z}^{(1)}$ *with respect to* $LB(D, \hat{M})$*, we have*

$$\frac{1}{n}\sum_{i=1}^{n} \phi(\hat{\delta}_i) \leq \frac{1}{n}\sum_{i=1}^{n} \phi(\delta_i). \tag{14}$$

(ii) *For any* $n \in Z^{(1)}$ *with respect to* $LB(D, \infty)$, *we have*

$$\frac{1}{n}\sum_{i=1}^{n} \phi(\delta_i) \leq \frac{1}{n}\sum_{i=1}^{n} \phi(\tilde{\delta}_i). \tag{15}$$

(iii) *For any* $n \in Z^{(1)}$ *with respect to* $LB(D, M)$, *we have*

$$\frac{1}{n}\sum_{i=1}^{n} \phi(\delta_i) \leq \frac{1}{n}\sum_{i=1}^{n} \phi(\bar{\delta}_i). \tag{16}$$

Notice that Theorems 4.1 and 4.2 hold without any condition. The stability conditions are needed only in obtaining steady state results. Under condition (3), $Z^{(1)}$ can be seen to contain infinitely many tags. Thus by taking limits along such subsequence, Theorems 3.1–3.3 follow directly from Theorem 4.2 and the assumptions of convex stability.

5 Conclusion

The results obtained in this paper will shed light on the design of the parameters of a LB. Larger D or smaller M will usually reduce the burstiness of the traffic which in turn may lead to smaller network delay. It introduces, however, more buffering delay at the LB. Trade-off between burstiness and buffering delay at the LB is then an important issue to the network designers. Together with bandwidth allocation, this issue will be addressed further in future research.

6 Acknowledgement

The author would like to thank Professor A. M. Makowski for many fruitful discussions with him, and for his valuable comments and suggestions in preparing this manuscript.

References

[1] *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 3, April 1991.

[2] U. Briem, T. H. Theimer, and H. Kroner, "A general discrete-time queueing model: analysis and applications," In *Proc. ITC-13*, 1991, pp. 13–19.

[3] M. Butto, E. Cavallero, and A. Tonietti, "Effectiveness of the "leaky bucket" policing mechanism in ATM networks," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 3, pp. 335–342, April 1991.

[4] A. E. Eckberg, D. T. Luan, and D. M. Lucantoni, "Bandwidth management: a congestion control strategy for broadband packet networks – characterizing the throughput-burstiness filter," In *Proc. ITC Specialist Seminar,* Adelaide, Sept. 1989, paper 4.4.

[5] L. Kuang, Ph.D. thesis, Electrical Engineering Department, University of Maryland at College Park.

[6] L. Kuang, "On the variance reduction property of buffered leaky bucket," submitted to *IEEE Trans. Communications.* Also see SRC technical report TR 91–90, University of Maryland at College Park.

[7] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York, 1979.

[8] M. Sidi, W. Liu, I. Cidon, and I. Gopal, "Congestion control through input rate regulation," In *Proc. GLOBECOM'89*, 1989, paper 49.2.

[9] D. Stoyan, *Comparison Methods for Queues and Other Stochastic Models*, J. Wiley & Sons, Berlin, 1983.

[10] J. S. Turner, "New directions in communications (or which way to the information age?)," *IEEE Communication Magazine*, vol. 24, no. 10, pp. 8–15, Oct. 1986.