Proceedings of the 32nd Conference
on Decision and Control
San Antonio, Texas • December 1993

An optimal flow control scheme that regulates the burstiness of traffic subject to delay constraints

Venkat Anantharam *
School of Electrical Engineering
Cornell University
Ithaca, NY 14853
ananth@camus.ee.cornell.edu

Takis Konstantopoulos †
Dept. of Electrical and Computer Engineering
University of Texas at Austin
Austin, TX 78712
takis@alea.ece.utexas.edu

Abstract

The problem of designing burst reducing flow controllers for traffic in an ATM network is considered. By requiring that the output flow obeys (σ, ρ) constraints, we show that an optimal design exists and can be easily implemented in real time.

1 Introduction

There is a need to rethink basic flow control and resource allocation issues to deal with the novel kinds of traffic encountered in a Broadband Integrated Services Digital Network (B-ISDN), that is, traffic with widely differing characteristics, such as audio, data and video. This traffic is expected to be of a highly bursty nature, making the traditionally employed models, such as Poisson or renewal processes, of doubtful relevance in deriving design insights. Several analyses of samples of traffic have pointed to the deficiency of traditional traffic models in the new applications context; see, for instance, Beran et al [4], and Jagerman and Melamed [8]. There is a serious need to incorporate explicit burstiness modelling in the traffic models used for performance analysis.

A simple and versatile class of models was introduced by Cruz [5], and has been further developed by Low and Varaiya, [10], [11]. Let A be a traffic process. Following Cruz, we say that A obeys (σ, ρ) constraints if there are positive constants σ, ρ, such that, if $A[s,t]$ denotes the amount of traffic over the time interval $[s,t]$ then, for all $s < t$,

$$A[s,t] \leq \sigma + \rho(t-s).$$

Intuitively, ρ poses a constraint on the long-term average rate and σ on the instantaneous bursts. A more precise characterization is that A obeys (σ, ρ) constraints iff when A is fed to a server with service rate ρ that is started empty at time s, and has an infinite buffer, the backlog in the buffer never exceeds σ at any time. There are a number of reasons why this class of flows is useful to consider. First of all, the above is an intuitively appealing characterization of burstiness, measured by the backlog induced in a simple queueing system, and as such is amenable to analysis and so to intelligent design choices. Secondly, it is a very simple characterization. Users may have to negotiate the parameters of the flows that they offer the network; the parameters σ and ρ are a simple pair with which to work, representing burstiness and average offered load.

Our focus in this paper is on the issue of flow control to ensure that a flow obeys burstiness constraints in the above sense. We examine the question of how a user, constrained to offer a flow obeying (σ, ρ) constraints can optimally control his source to ensure this requirement. Our key result is that there is a very simple answer to this question, involving recursively updating a quantity, which, for precision, we call the *virtual backlog*. This quantity can also be shown to work as a sufficient statistic to study various natural allocation problems withing a B-ISDN network,

*Research supported in part by NSF grant NCR 88-57731

†Research supported in part by NSF grant NCR 92-11343

as shown in [1].

One should mention that the leading contender for a standard for implementing the B-ISDN concept is Asynchronous Transfer Mode (ATM). In ATM traffic is segmented into fixed length cells (the standard calls for cells of length 53 bytes) which are routed through the network along virtual circuits. One of the most popular flow control schemes for traffic consisting of ATM cells is the "leaky bucket" flow control scheme; see, for example, Eckberg and Lucantoni [7]. This scheme has been shown in practice to be very effective in regulating burstiness and is also simple to implement. In [2], we analyzed a model of the leaky bucket and showed that it is burst-reducing, in the sense that it produces an output that induces delays within the network smaller than those that they would have been produced without the leaky bucket. A similar result was also obtained independently by Low and Varaiya, [10] and by Kuang, [9]. These results suggest that the leaky bucket is a good flow control scheme. It is interesting that an examination of the protocol shows that the output flow from bucket obeys (σ, ρ) constraints.

Previous work on flow control for high-speed networks has been similar in that it has focused mainly on the performance evaluation and comparison between different proposed schemes. In this paper we take a harder look at the problem of designing an optimal flow control scheme. We pose the design problem as follows : we are given a target burstiness parameter pair (σ, ρ). The flow controller is allowed to delay incoming cells by at most K time units, $K \geq 0$. At each point of time the flow controller can either reject or transmit part of the delayed cells or the incoming traffic. The decision of the controller (whether and how much to transmit or reject) at each point of time t is allowed, in principle, to depend on the *entire past* of the input process up to time t. The only requirement is that the flow control scheme should create an output traffic stream that must satisfy the (σ, ρ) constraints. Every policy that obeys the above requirements is called a feasible policy. The goal is to find the best feasible policy in the sense that the rejected traffic is as small as possible. It turns out that there is an optimal policy that is rather simple in nature and can be implemented by recursively updating the quantity we call the virtual backlog. It appears to be very easy to implement in real time.

In what follows we define the problem in a more precise way (Section 2) and prove the optimality results (Section 3). The whole analysis is done in discrete time. For results that hold for arbitrary traffic processes in continuous time we refer to our forthcoming paper [3].

2 The discrete time flow controller

Messages arrive at times $n = 0, 1, 2, \ldots$. Let a_n be the amount of traffic arriving at time n. We take a_n to be a nonnegative real number. A flow controller maps $\{a_n, n \geq 0\}$ into $\{b_n, n \geq 0\}$. Thus b_n represents the amount of traffic transmitted at time n. Positive numbers σ and ρ are needed for the definition of the constraints. To set up the initial conditions we also consider an initial constraint $\sigma_0 \leq \sigma$. The sequence $\{b_n, n \geq 0\}$ is called <u>*(σ, ρ) constrained with initial constraint σ_0*</u> iff

$$\sum_{\ell=0}^{n} b_\ell \leq \sigma_0 + \rho(n+1), \; n \geq 0, \qquad (1)$$

$$\sum_{\ell=m}^{n} b_\ell \leq \sigma + \rho(n-m+1), \; n \geq m \geq 1. \qquad (2)$$

Consider then a (σ, ρ) constrained sequence $\{b_n, n \geq 0\}$. Fix now an index n. By isolating the last term b_n from inequalities (2) we obtain

$$b_n \leq \sigma + \rho(n-m+1) - \sum_{\ell=m}^{n-1} b_\ell, 1 \leq m \leq n. \quad (3)$$

Similarly, (1) yields

$$b_n \leq \sigma_0 + \rho(n+1) - \sum_{\ell=0}^{n-1} b_\ell, \; n \geq 0 \,. \quad (4)$$

Combining the latter inequalities (3), (4), we further obtain

$$b_n \leq \rho + \sigma_n, \; n \geq 0, \qquad (5)$$

where σ_n is defined by

$$\sigma_n = \min \left\{ [\sigma_0 + \rho n - \sum_{\ell=0}^{n-1} b_\ell], \right.$$

$$\min_{1\le m\le n}[\sigma+\rho(n-m)-\sum_{\ell=m}^{n-1}b_\ell]\Big\} \qquad (6)$$

We refer to σ_n as the *virtual backlog* at time n. Observe that σ_n depends on σ_0 and the past $\{b_m, 0\le m\le n-1\}$ only. What is important is the observation that the virtual backlog can be *updated recursively*:

$$\begin{aligned}\sigma_0 &= \sigma_0\\ \sigma_{n+1} &= \min\{\sigma,\ \sigma_n+\rho-b_n\},\ n\ge 0 \quad (7)\end{aligned}$$

The proof of this recursion follows easily from the defining relation (6). We have thus proved that if $\{b_n\}$ is (σ,ρ) constrained with initial constraint σ_0 then (5) holds with σ_n obtained from the recursion (7). Conversely, given a traffic process $\{b_n\}$, with $\{\sigma_n\}$ computed from the recursion (7), then $\{b_n\}$ can be seen to be (σ,ρ) constrained with initial constraint σ_0.

Suppose first that we are interested in the case where the flow controller takes instantaneous decisions. Thus if a_n is the amount of traffic arriving at time n, the flow controller has to immediately decide the amount $b_n\le a_n$ to be transmitted; the remaining amount is instantaneously rejected. Furthermore, the decision at time n should be based only on information up to this time only. The above discussion shows that a feasible policy must produce an output $\{b_n\}$ such that $b_n\le a_n$, for all n, and (5) holds.

Before proceeding to the solution, let us generalize to the situation at which traffic arriving at time n is allowed to wait at most K time units before it is transmitted. At each point of time n let m_n^i represent traffic that arrived at time $n-i$ $(i=1,\ldots,K)$ but has not yet been released. A given policy produces b_n that is at most equal to $a_n+m_n^1+\cdots+m_n^K$. Consequently, a feasible policy is one that (i) is non-anticipative, and (ii) produces an output $\{b_n\}$ such that

$$b_n\le\min\{\sigma_n+\rho, a_n+m_n^1+\cdots+m_n^K\}, \quad (8)$$

where σ_n is recursively obtained via (7).

3 Optimality results

Among all feasible flow controllers, as described in Section 2, we are interested in those that maximize the rate of cells transmitted. We show that this is possible by explicitly constructing one such flow controller. This is made precise in the following theorem.

Theorem 1 *Given a nonnegative integer K, nonnegative numbers σ_0, $\sigma\ge\sigma_0$, ρ, $m_0^1,\ldots,m_0^K$, and an arrival sequence $\{a_n, n=0,1,\ldots\}$, let $\sigma_0^*=\sigma_0$, $m_0^{*i}=m_0^i$, $1\le i\le K$. Next consider a policy which, at time n, releases an amount of traffic b_n^*, recursively defined through*

$$b_n^*=\min\{\sigma_n^*+\rho, a_n+m_n^{*1}+\cdots+m_n^{*K}\},$$

$$\sigma_{n+1}^*=\min\{\sigma,\sigma_n^*+\rho-b_n^*\},$$

$$m_{n+1}^{*1}+\cdots+m_{n+1}^{*i}=$$
$$=a_n+\min\{m_n^{*1}+\cdots+m_n^{*i-1}, m_n^{*1}+\cdots+m_n^{*K}-b_n^*\},$$
$$i=1,\ldots,K.$$

for $n=0,1,\ldots$. Then $\{b_n^*\}$ is (σ,ρ) constrained with initial constraint σ_0, and for any feasible policy $\{b_n\}$, we have

$$\sum_{\ell=0}^{n}b_\ell^*\ge\sum_{\ell=0}^{n}b_\ell,\ \textit{for all } n\ge 0. \qquad (9)$$

Proof. The fact that $\{b_n^*\}$ is (σ,ρ) constrained with initial constraint σ_0 is an immediate consequence of the discussion of Section 2. We shall show the validity of (9) by induction. Consider an arbitrary feasible policy that produces the output process $\{b_n\}$. Let $B_n=\sum_{\ell=0}^n b_\ell$, $B_n^*=\sum_{\ell=0}^n b_\ell^*$, $M_n=m_n^1+\cdots+m_n^K$, $M_n^*=m_n^{*1}+\cdots+m_n^{*K}$. Next consider the following inequalities.

$$\begin{aligned}B_n^* &\ge B_n &(10)\\ B_{n-1}^*+\sigma_n^* &\ge B_{n-1}+\sigma_n &(11)\\ B_{n-1}^*+M_n^* &\ge B_{n-1}+M_n. &(12)\end{aligned}$$

Clearly, they all hold for $n=0$; indeed, $b_0^*=\min\{\sigma_0+\rho, a_0+M_0\}\ge b_0$ (see (8)), $\sigma_0^*=\sigma_0$, and $M_0^*=M_0$. Now assume that they hold when n is replaced by $n-1$:

$$\begin{aligned}B_{n-1}^* &\ge B_{n-1} &(13)\\ B_{n-2}^*+\sigma_{n-1}^* &\ge B_{n-2}+\sigma_{n-1} &(14)\\ B_{n-2}^*+M_{n-1}^* &\ge B_{n-2}+M_{n-1}. &(15)\end{aligned}$$

Inequalities (13) - (15) form our induction hypothesis. We shall show that (10) - (12) hold true.

Consider first (11). We have

$$\begin{aligned}B^*_{n-1}+\sigma^*_n &= B^*_{n-1}+\min\{\sigma,\sigma^*_{n-1}+\rho-b^*_{n-1}\}\\ &= \min\{B^*_{n-1}+\sigma, B^*_{n-2}+\sigma^*_{n-1}+\rho\}\\ &\geq \min\{B_{n-1}+\sigma, B_{n-2}+\sigma_{n-1}+\rho\}\\ &= B_{n-1}+\min\{\sigma,\sigma_{n-1}+\rho-b_{n-1}\}\\ &= B_{n-1}+\sigma_n;\end{aligned}$$

we used (13), (14) of the induction hypothesis to obtain the inequality in the third step above. Thus (11) holds.

Next consider (12). To prove that this holds we will make use of the following lemma.

Lemma 1 *Consider the policy defined in Theorem 1. If, for some n, $m^{*K}_n > 0$ then $m^{*i}_n = a_{n-i}$, for $i = 1,\ldots,K-1$.*

Proof of Lemma 1. From the evolution equations of the policy defined in Theorem 1 we readily obtain

$$m^{*i}_n = \min\{M^*_{n-1},\sum_{j=1}^{i-1} m^{*j}_{n-1}\} -$$

$$-\min\{M^*_{n-1},\sum_{j=1}^{i-2} m^{*j}_{n-1}\}, i=2,\ldots,K. \quad (16)$$

$$m^{*1}_n = \min\{a_{n-1}+0, M^*_{n-1}\}. \quad (17)$$

Suppose that, for some $2 \leq i \leq K$, $m^{*i}_n > 0$. Then, from (16), $M^*_{n-1} > \sum_{j=1}^{i-2} m^{*j}_{n-1}$. (For if $M^*_{n-1} \leq \sum_{j=1}^{i-2} m^{*j}_{n-1}$ then (16) yields $m^{*i}_n = 0$.) In particular, $M^*_{n-1} > 0$, and (17) yields $m^{*1}_n = a_{n-1}$. Also, from (16) again, $m^{*i}_n = \sum_{j=1}^{i-1} m^{*j}_{n-1} - \sum_{j=1}^{i-2} m^{*j}_{n-1} = m^{*i-1}_{n-1}$, and thus $m^{*i-1}_{n-1} > 0$. So we can replace i, in the above argument, by $i-1$ and n by $n-1$, and thus Lemma 1 has been shown by induction.

Using the evolution equations of the theorem statement, we now have:

$$B^*_{n-1}+M^*_n =$$

$$B^*_{n-1}+a_{n-1}+\sum_{j=1}^{K-1} m^{*j}_{n-1}+$$

$$+\min\{0, m^{*K}_{n-1}-b^*_{n-1}\}. \quad (18)$$

We consider two cases. First, if $b^*_{n-1} \geq m^{*K}_{n-1}$, then (18) gives

$$\begin{aligned}B^*_{n-1}+M^*_n &=\\ &= B^*_{n-1}+a_{n-1}+\sum_{j=1}^{K-1} m^{*j}_{n-1}+m^{*K}_{n-1}-b^*_{n-1}\\ &= a_{n-1}+B^*_{n-2}+M^*_{n-1}\\ &\geq a_{n-1}+B_{n-2}+M_{n-1}\\ &= B_{n-1}+M_{n-1}+a_{n-1}-b_{n-1}\\ &= B_{n-1}+M_n,\end{aligned}$$

where we used inequality (15) of the induction hypothesis in the fourth step above; the last equality is obvious. Thus, in the case $b^*_{n-1} \geq m^{*K}_{n-1}$, (12) holds. Next, suppose $b^*_{n-1} < m^{*K}_{n-1}$. In this case, (18) gives

$$B^*_{n-1}+M^*_n = B^*_{n-1}+a_{n-1}+\sum_{j=1}^{K-1} m^{*j}_{n-1}.$$

Since $m^{*K}_{n-1} > 0$, the condition of Lemma 1 holds, and so we can write

$$\begin{aligned}B^*_{n-1}+M^*_n &=\\ &= B^*_{n-1}+a_{n-1}+a_{n-2}+\cdots+a_{n-K}\\ &\geq B_{n-1}+a_{n-1}+a_{n-2}+\cdots+a_{n-K}\\ &\geq B_{n-1}+M_n;\end{aligned}$$

here we used (13) of the induction hypothesis and the fact that M_n is, for a feasible policy, at most equal to the amount of traffic arrived between $n-K$ and $n-1$.

Finally, we show that inequality (10) holds as well.

$$\begin{aligned}B^*_n &= B^*_{n-1}+b^*_n\\ &= B^*_{n-1}+\min\{\sigma^*_n+\rho, a_n+M^*_n\}\\ &= \min\{B^*_{n-1}+\sigma^*_n+\rho, a_n+B^*_{n-1}+M^*_n\}\\ &\geq \min\{B_{n-1}+\sigma_n+\rho, a_n+B_{n-1}+M_n\},\end{aligned}$$

where we used (11) and (12) to establish the latter inequality. This concludes the proof of the theorem.

4 Conclusions and extensions

We investigated flow controllers whose goal is to reduce the burstiness of traffic by imposing a target burstiness parameter pair (σ,ρ), while, at the same time, imposing delay constraints on the incoming traffic. We posed the problem in discrete time and showed that the constraints can be optimally satisfied by the flow controller defined by the recursions of the statement of Theorem 1. A closer examination of this optimal flow controller shows its similarity to the leaky bucket. The results can be extended in continuous time where the flow control problem can be posed for arbitrary traffic processes, namely processes that have a continuous (fluid) part as well as instantaneous jumps (point process part). It is worthwhile to notice that the implementation of the scheme that we proposed can be done in a very simple way and in real time.

References

[1] Anantharam, V. (1992) An approach to the control of ATM Networks. Talk presented at IEEE International Symposium on Information Theory, Bahia, Brazil, July 1992.

[2] Anantharam, V. and Konstantopoulos, T. (1991). Burst reduction properties of the leaky bucket flow control scheme in ATM networks. *Proc. 29th Allerton Conf.*, 302-309. Subm. to *IEEE Tr. Comm.*

[3] Anantharam, V. and Konstantopoulos, T. (1993). An optimal flow control scheme that regulates the burstiness of traffic. *Preprint.*

[4] Beran, J., Sherman, R., Taqqu, M., and Willinger, W. (1992). Variable-bit-rate Video traffic and Long-range dependence. *Preprint*, Bell Communications Research, Morristown, NJ.

[5] Cruz, R.L. (1991). A calculus for network delay, part I: network elements in isolation. *IEEE Tr. Inf. Th.*, **37**, 114-131.

[6] Cruz, R.L., and Chuah, M.C. (1991). A minimax approach to a simple routing problem (1991). *IEEE Tr. Aut. Contr.*, **36**, 1424-1435.

[7] Eckberg, A.E. and Lucantoni, D.M. (1988). A traffic/performance analysis of the bandwidth management throughput-burstiness filter. *Proc. 29th IEEE Conf. Dec. Contr.*, 2118-2123.

[8] Jagerman, D., and Melamed, B. (1992). The Transition and Autocorrelation Structure of TES Processes. Part I : General Theory; Part II : Special Cases. *Communications in Statistics and Stochastic Models*, Vol. 8, No. 1, pp. 194 -219; Vol. 8, No. 3, pp. 499 -527

[9] Kuang, L. (1992) On the Variance Reduction Property of Buffered Leaky Bucket. *Preprint.*

[10] Low, S. and Varaiya, P. (1991). A simple theory of traffic and resource allocation in ATM. *Proc. Globecom.*

[11] Low, S. (1992). *Traffic management of ATM networks : Service provisioning, Routing, and Traffic shaping.*, PhD. Dissertation, University of California, Berkeley.

[12] Sasaki, G. (1991). Input buffer requirements for round robin polling systems with nonzero switchover times. *Proc. 29th Allerton Conf.*, 603-612.