Proceedings of the 10th Mediterranean Conference on Control and Automation, Lisbon, Portugal, July 2002

CONTROL OF DISCRETE COMMUNICATION NETWORKS: ASYMPTOTIC EFFICIENCY AND BUFFER OPTIMIZATION

Jacob Hammer

Department of Electrical and Computer Engineering
P.O. Box 116130
University of Florida
Gainesville, FL 32611, USA
hammer@mst.ufl.edu

Keywords: Communication Networks, Robust Nonlinear Control, Efficiency, Feedback Control.

Abstract

Asymptotic efficiency is an optimization measure aimed at maximizing the flow through large capacity digital communication networks. The present note addresses the issue of controlling the network so as to maximize asymptotic efficiency while overcoming the effects of unmodeled traffic uncertainties. It is shown that, in many common situations, feedback controllers can achieve asymptotic efficiency of 1 even in the presence of substantial unmodeled uncertainties.

1 Introduction

Discrete (or digital) communication networks are used to transfer data among computers or other digital devices. The flow of data through a discrete communication network is governed by a traffic control algorithm. The algorithm controls traffic by determining what data is admitted into the network and by adjusting the flow of traffic within the network. Issues related to the admission of data into a network are discussed in [10] and [12].

The traffic flow within a network can be adjusted by storing some of the transmitted data in buffers during high traffic conditions, and releasing the data back into the flow when the traffic subsides. The present note deals with the development of techniques for the optimization of buffer use. The main objective is to maximize the data transmitted through a network, especially when the data flow is subject to large unmodeled uncertainties.

We adopt a networking model whereby each data record is divided into small and equal segments. Each segment is combined with information identifying the data record to which the segment belongs and the position of the segment within the data record. The resulting combination is called a *cell* (e.g., [4]). The collection of all cells that pertain to one data record forms a *call*.

The flow rate of cells varies during a call, depending on the rate at which data is generated by the source. For example, in a digitized phone call there are periods of low cell transmission rate, corresponding to silence; in a digitized video call, a change of scenery may result in a substantial momentary increase in cell transmission rate. Each call category has its own typical cell flow pattern (or *waveform*).

The efficiency of network utilization depends on how well the waveforms of the various calls fit with one another. In [10] and [12] we have introduced the notion of complete families of calls. In a complete family of calls, the waveforms of the calls fit especially well together and are ideally suited for maximal efficiency.

In practice, the cell flow rate of many classes of calls is largely random. For such classes, the call waveforms change from one sample of a call to another, and are unpredictable. Moreover, in many cases (e.g., digitized video or other multimedia calls), there are no reliable comprehensive statistical models of call waveforms. Hence, it is important for traffic control algorithms to function properly in the presence of unmodeled uncertainties.

The present note examines some basic aspects of network optimization in the presence of unmodeled uncertainties. It shows that it is often possible to use feedback controllers to reshape call waveforms so as to reduce, and sometimes completely eliminate, the effects of unmodeled uncertainties on network efficiency. The ability to overcome the effect of uncertainties on network efficiency depends in a critical way on the delay tolerance of the transmitted calls (section 3 below).

The most costly part of a discrete communication network is usually its backbone - a long-distance high-capacity network link. Our discussion concentrates on maximizing the utilization of the backbone.

2 Basics

We represent a network by a discrete model, whereby the flow of cells corresponds to a sequence of integers. Each element of the sequence represents the number of cells that flow through a point of the network during a specified time interval $\Delta > 0$. The size of Δ is selected short compared to network time constants and delays. For an integer $k \geq 1$, the symbol v_k indicates the number of cells that flow through the point v of the network during the period $((k-1)\Delta, k\Delta]$. The interval $((k-1)\Delta, k\Delta]$ is referred to as *step* k.

A call has a finite duration of $T \geq 1$ steps, and is represented by a piecewise constant sequence. The interval $[1, T]$ is called the *call cycle*. The integer T may represent a common multiple of the durations of all calls of interest, so all calls become compatible with the call cycle. In the event of very long calls, T may indicate a convenient breakpoint of a call. The interval $[T+1, 2T]$ is then the *second call cycle*, and so on.

For an integer $q \geq 1$, we partition the call cycle into q disjoint sub-intervals $I_1 := [1, t_1]$, $I_2 := [t_1+1, t_2]$, ..., $I_q := [t_{q-1}+1, T]$, where $t_1 = T$ when $q = 1$. The sub-intervals $I_1, ..., I_q$ are called *segments*. The number of steps in the segment I_i is denoted by λ_i, and we assume that $\lambda_i \geq 1$ for all $i = 1, ..., q$; i.e., there are no segments of zero length. A list of q integers $\varphi(1), \varphi(2), ..., \varphi(q)$ defines a *piecewise constant sequence* φ over the set of segments $I_1, ... I_q$ by setting

$$\varphi_k := \begin{cases} 0 & \text{for} \quad k \leq 0 \\ \varphi(1) & \text{for} \quad 1 \leq k \leq t_1, \\ \varphi(2) & \text{for} \quad t_1+1 \leq k \leq t_2, \\ ..., & \\ \varphi(q) & \text{for} \quad t_{q-1}+1 \leq k \leq T, \\ 0 & \text{for} \quad T+1 \leq k. \end{cases}$$

The integers $t_1, t_2, ..., T$ are called the *switching times* of the sequence. By employing segments of length 1, every sequence with finite support can be represented in this form.

A call c is a piecewise constant sequence, composed of the sum of a deterministic part and an uncertain part:

$$c = \chi + \upsilon.$$

Here, χ represents the deterministic (or *nominal*) part of the call, while υ represents the uncertain part. Both χ and υ are piecewise constant functions over the partition $\{I_1, ..., I_q\}$ of $[1, T]$. The only information available about the uncertain part υ is an amplitude bound $\rho \geq 0$:

$$0 \leq \upsilon \leq \rho. \tag{1}$$

No further information is available about the statistics of the uncertain part. In this sense, υ represents an unmodeled traffic uncertainty. Condition (1) characterizes all *permissible* uncertain parts.

The signal model just introduced forms the basis of the theory of sturdy traffic control ([10], [11], [12]). In brief terms, sturdy traffic control deals with the control of network traffic under conditions of large and unmodeled uncertainties. An important requirement of sturdy traffic control is to provide lossless network transmission. In other words, no cells may be lost during transfer through the network. The requirement of lossless transmission sets sturdy traffic control somewhat apart from the more traditional statistical traffic control methods. The latter do permit some cell losses during certain rare traffic events (see for example [1], [9], and [7]). By using sturdy traffic control algorithms, we demonstrate in section 3 below that it is often possible to achieve full asymptotic efficiency with no cell loss, despite call uncertainties.

The traffic passing through the network consists of calls belonging to m call classes $C^1, ..., C^m$. Each call class has its own nominal call waveform and its own service requirements. The main service requirement of interest to us here is the maximal buffering delay, namely, the maximal time a cell of the class may spend in the network buffering system.

Some call classes are more tolerant of cell delays than others. The class of computer data file transfers, for example, allows substantial buffering delays, while classes such as streaming multimedia have a relatively low tolerance of delays. Let $\tau(r)$ be the maximal buffering delay permitted for cells of the class C^r. We assume that

$$\tau(r) \geq 1 \text{ for all } r = 1, ..., m,$$

namely, that each call class allows a buffering delay of at least one step. Maximal buffering delay is, of course, just one of many service requirements (e.g., [4], [13], [16]). It is one of the most critical service requirements from a network control point of view.

A call of the class C^r is of the form

$$c^r = \chi^r + \upsilon^r, r = 1, ..., m,$$

where χ^r is the nominal part and υ^r is the uncertain part. All calls of a class C^r share the same nominal call waveform χ^r, but their uncertain part υ^r may vary from one call sample to another. For the sake of notational simplicity, we assume that all uncertain parts have the same amplitude bound $\rho \geq 0$, so that

$$0 \leq \upsilon^r \leq \rho \text{ for all } r = 1, ..., m.$$

As mentioned earlier, the amplitude bound ρ is the only a-priori information available about the uncertain parts of the calls. No statistical model of the uncertainties is available. No restrictions are imposed on the magnitude of the uncertainty bound ρ; it is not necessarily small when compared to call amplitude.

We refer to the set $F := \{c^1, ..., c^m\}$ as the *family of calls* passing though the network. The family F induces the family $F(\chi) := \{\chi^1, ..., \chi^m\}$ consisting of the deterministic parts of the calls. The members of $F(\chi)$ are piecewise constant sequences over the partition $\{I_1, ..., I_q\}$ of the call cycle $[1, T]$.

The waveforms of the calls may change as the calls are processed by the buffering system before entering the backbone. Let c^i_b be the waveform of call class C^i when it enters the backbone. As before, we decompose this waveform into a deterministic part χ^i_b and an uncertain part υ^i_b:

$$c^i_b = \chi^i_b + \upsilon^i_b, i = 1, ..., m.$$

At a given time, let $\alpha_i \geq 0$ be the number of calls of the class C^i entering the backbone, $i = 1, ..., m$. The non-negative integers $\alpha_1, ..., \alpha_m$ are called the *call populations*. Let $c^i_b(k)$ be the number of cells of call i entering the backbone at the step k. Then, the total number of cells entering the backbone at the step k is given by

$$z_k = \Sigma_{i=1}^m \alpha_i c^i_b(k).$$

Now, let ϕ be the maximal number of cells the backbone can transmit in one step. We refer to ϕ as the *backbone capacity*. The call populations can be regarded as functions $\alpha_1(\phi), ..., \alpha_m(\phi)$ of ϕ, as they are obviously determined based on backbone capacity. To make the entire notation consistent, define the quantity

$$z_k(\phi) = \Sigma_{i=1}^m \alpha_i(\phi) c^i_b(k),$$

which is equal to the total number of cells injected into the backbone at the step k; it characterizes the *backbone flow*. Clearly, one must have

$$0 \leq z_k(\phi) \leq \phi. \quad (2)$$

We refer to $z(\phi)$ as the *traffic control algorithm*; it is determined by the call populations $\alpha_1(\phi), ..., \alpha_m(\phi)$ and by the waveforms $c_b^1, ..., c_b^m$.

With a capacity of ϕ, the backbone can transfer a maximum of ϕT cells during one call cycle. Regarding ϕT as the *backbone volume*, it follows that the fraction of backbone volume utilized by the traffic control algorithm $z(\phi)$ is given by

$$\eta(z(\phi)) := \frac{\Sigma_{k=1}^{T} z_k(\phi)}{T\phi}.$$

In view of (2),

$$0 \leq \eta(z(\phi)) \leq 1. \quad (3)$$

We refer to $\eta(z(\phi))$ as the *efficiency* of $z(\phi)$; it describes the backbone efficiency induced by the traffic control algorithm $z(\phi)$.

In applications, backbones usually have very large capacity, so it is of interest to consider the efficiency in the limit as $\phi \to \infty$. This leads to the notion of *asymptotic efficiency* of the traffic control algorithm $z(\phi)$, which is defined by (see also [10], [12])

$$\eta_\infty(z) := \lim_{\phi\to\infty} \eta(z(\phi)).$$

Due to (3), the asymptotic efficiency is always between zero and one. An asymptotic efficiency of one means that, for large capacity backbones, the flow induced by the traffic control algorithm $z(\phi)$ fills (almost) the entire backbone volume. Our objective is to devise traffic control algorithms that achieve the highest possible asymptotic efficiency. The highest achievable asymptotic efficiency depends on the properties of the family F of calls moving through the backbone.

The family $F(\chi_b) := \{\chi_b^1, ..., \chi_b^m\}$ is a *complete family* of calls if there are integers $\alpha_1, ..., \alpha_m \geq 0$ such that the linear combination $\Sigma_{i=1}^m \alpha_i \chi_b^i = c$ is a non-zero constant over the interval $[1, T]$. Consider for a moment the case of purely deterministic calls, namely, the case where each call c_b^i injected into the backbone is entirely equal to its deterministic part χ_b^i, $i = 1, .., m$. Then, complete families are the only families of calls for which one can achieve asymptotic efficiency of 1, as indicated by the following statement ([10], [12]).

THEOREM 4. Let $F(\chi_b) := \{\chi_b^1, ..., \chi_b^m\}$ be a family of piecewise constant deterministic calls over the partition $\{I_1, ..., I_q\}$. Then, the following two statements are equivalent.

(i) There is a traffic control algorithm with asymptotic efficiency of 1 for the family $F(\chi_b)$.

(ii) The family $F(\chi_b)$ is a complete family.

Moreover, asymptotic efficiency of 1 is achieved only with constant backbone flow. ♦

The network optimization process developed in this note is a global process, taking into account traffic flow over a period of time. It is a functional optimization approach to dynamic network control, refining and highlighting some results of [11]. Other approaches to network control are exposed in [8], [15], [6], as well as in many other sources.

The traffic control algorithms developed in this note depend on the use of feedback control. In this sense, the present discussion continues a long tradition of feedback use in traffic control algorithms (e.g., [14], [5], [2] and [3]).

3 Calls With Uncertainties: Complete Nominal Families

We turn now to the problem of maximizing the asymptotic efficiency of a backbone carrying calls with amplitude uncertainties. We show that, in some common cases, it is possible to eliminate entirely the effects of the uncertainties on network efficiency. In order to demonstrate the principles with minimal complications, we restrict our attention here to the special case where the family $F(\chi)$ of deterministic parts forms a complete family of calls.

Let then $F(\chi) = \{\chi^1, ..., \chi^m\}$ be a complete family of calls. Denote by $V^m(\{\chi^1, ..., \chi^m\})$ the set of all integers $\alpha_1, ..., \alpha_m \geq 0$ for which $\Sigma_{i=1}^m \alpha_i \chi^i = c$ is a non-zero constant over the interval $[1, T]$. Note that the definition of a complete family guarantees that the set $V^m(\{\chi^1, ..., \chi^m\})$ is not an empty set.

Returning now to the family $F = \{c^1, ..., c^m\}$ of calls entering the network, recall that $c^i = \chi^i + \upsilon^i$, where χ^i is the nominal (deterministic) part and υ^i is the uncertain part, $i = 1, ..., m$. The calls are defined over the partition $\{I_1, ..., I_q\}$ of the call cycle $[1, T]$. For the sake of simplicity, we assume that the following are valid throughout the remaining part of this note.

ASSUMPTIONS 5.

(i) All segments $I_1, ..., I_q$ have the same length $\lambda_i = \lambda \geq 1$, $i = 1, ..., q$.

(ii) The nominal parts $\chi^1, ..., \chi^m$ form a complete family of calls.

(iii) All calls have the same uncertainty bound $\rho \geq 0$, i.e., $0 \leq \upsilon^i \leq \rho$ for all $i = 1, ..., m$.

(iv) The delay bound $\tau(i)$ of the call class C^i satisfies $1 \leq \tau(i) < 2\lambda$ for all $i = 1, ..., m$. ♦

Now, let $\mu(\cdot)$ denote the unit step function, i.e., $\mu(t) = 1$ when $t \geq 0$, and $\mu(t) = 0$ when $t < 0$. For a list of integers $\alpha_1, ..., \alpha_m \in V^m(\{\chi^1, ..., \chi^m\})$, define the following quantities.

$$\begin{aligned} &\tau^* := \min\{\tau(r) : \alpha_r \neq 0, r = 1, ..., m\} \\ &\theta(i,k) := \min\{k-\tau(i), T, 0\}. \end{aligned} \tag{6}$$

$$\sigma(k) := (k - T)\Sigma_{i=1}^m \alpha_i\theta(i,k)/\lambda, \quad k = T+1, ..., T+\lambda. \tag{7}$$

$$\varepsilon_k = \frac{\Sigma_{r=1}^m \Sigma_{i=\max\{1,\theta(i,k)+1\}}^k \alpha_r\chi_i^r}{\Sigma_{r=1}^m (k-\tau(r))\mu(k-\tau(r))\,\alpha_r}, \quad k = \tau^*+1, ..., T, \quad \text{if } \tau^*+1 \leq T, \tag{8}$$

$$\varepsilon_k = \frac{\Sigma_{r=1}^m \Sigma_{i=\max\{1,\theta(i,k)+1\}}^T \alpha_r\chi_i^r}{[\Sigma_{r=1}^m \theta(i,k)\mu(\theta(i,k))\alpha_r] - \sigma(k)}, \quad k = T+1, ..., T+\lambda, \tag{9}$$

$$\varepsilon_{T+\lambda+1} := \frac{\lambda\Sigma_{i=1}^m \alpha_i\chi^i}{T(\Sigma_{i=1}^m \alpha_i)}, \tag{10}$$

where, for any $k = 1, ..., T+\lambda$, set $\varepsilon_k := \infty$ if the denominator of the expression for ε_k is less than or equal to zero.

Finally, define the function

$$\varepsilon(\alpha_1, ..., \alpha_m) := \min\{\varepsilon_k : k = 1, ..., T+\lambda+1\}$$

and note that it is always non-negative. A direct examination shows that the value of $\varepsilon(\alpha_1, ..., \alpha_m)$ is determined by the delay bounds $\tau(1), ..., \tau(m)$, by the waveforms of the nominal parts $\chi^1, ..., \chi^m$ of the calls, and by the coefficients $\alpha_1, ..., \alpha_m$. It does not depend on the uncertain parts of the calls, and consequently can be evaluated from a-priori information. The following statement shows that $\varepsilon(\alpha_1, ..., \alpha_m)$ provides a tight bound on the maximal uncertainty amplitude for which asymptotic efficiency of 1 can be achieved for the backbone of the network (see also [11]). It is the main result of our current discussion.

THEOREM 11. Let $F = \{c^1, ..., c^m\}$ be a family of calls, where χ^i is the nominal part of the call c^i and $\rho > 0$ is the amplitude of its uncertain part, $i = 1, ..., m$. Assume that Assumptions 5 are valid. Then, the following two statements are equivalent.

(i) There is a traffic control algorithm that yields backbone asymptotic efficiency of 1 for F.
(ii) There are integers $\alpha_1, ..., \alpha_m \in V^m(\{\chi^1, ..., \chi^m\})$ such that $\rho \leq \varepsilon(\alpha_1, ..., \alpha_m)$.

An outline of the proof of Theorem 11 is provided in the Appendix below. As we can see from the Theorem, asymptotic efficiency of 1 can be achieved in certain cases, despite the presence unmodeled call uncertainties. The only requirement is that the uncertainty amplitude bound ρ must not exceed the value $\varepsilon(\alpha_1, ..., \alpha_m)$. This value can be relatively large, as demonstrated in Example 12 below. Accordingly, sturdy network control techniques can handle substantial traffic uncertainties without impairing asymptotic efficiency.

The proof of Theorem 11 contains a traffic control algorithm that achieves asymptotic efficiency of 1 under the conditions of the Theorem. This algorithm is based on the use of feedback, and is described in more detail in [11]. When employing this algorithm, no cells are delayed beyond their limit and no cells are lost in transfer.

REMARK. Condition (ii) of Theorem 11 remains a sufficient condition for the achievement of asymptotic efficiency of 1 even when restriction (iv) of Assumptions 5 is released. It can be turned into a necessary condition in that case by somewhat modifying the formula of $\varepsilon(\alpha_1, ..., \alpha_m)$ (see [11] for details). ♦

EXAMPLE 12. To demonstrate the significance of Theorem 11, consider the following special, but relatively common, case. Assume that one of the call classes, say C^m, allows relatively large cell delays. A common example of such a call class is the class of computer file transfers. Assume that, for the class C^m, the cell delay bound satisfies $\tau(m) \geq T+\lambda$. Let

$$\sigma^m := \min \{\chi_1^m, ..., \chi_q^m\} \tag{13}$$

be the minimal value of the nominal waveform χ^m of c^m. If C^m is the class of computer file transfers, then its nominal waveform is usually constant, and σ^m is equal to that constant value.

Under these circumstances, condition (ii) of Theorem 11 remains a sufficient condition for achieving asymptotic efficiency of 1, and one can derive from it the following simplified form (see [11] for details). If the uncertainty amplitude ρ satisfies

$$\rho \leq \frac{\alpha_m \sigma^m}{\Sigma_{r=1}^m \alpha_r} \tag{14}$$

for some integers $\alpha_1, ..., \alpha_m \in V^m(\{\chi^1, ..., \chi^m\})$, then the backbone can be operated with asymptotic efficiency of 1.

As a numerical example, consider the case where the class C^m contributes, say, half of the call population of the backbone. Then, (14) becomes

$$\rho \leq \frac{1}{2}\sigma^m,$$

i.e., half of the flow amplitude of computer file transfers. This shows that asymptotic efficiency of 1 can be achieved even when the uncertainty amplitude is at a rather significant level. ♦

To conclude, we have presented some aspects of the theory of sturdy control of discrete communication networks. The main advantage of this theory is that it provides traffic control algorithms that handle uncertainties without requiring statistical models and without incurring transmission losses. In the present note, we have concentrated on the special case where the deterministic parts of the calls form a complete family. We have seen that even in cases of substantial uncertainty amplitudes, it is often possible to achieve asymptotic efficiency of 1 without any cell losses.

Appendix

This appendix contains an outline of the proof of Theorem 11. A more complete theory related to this proof is developed in [11].

First note that Assumption 5(iv) implies that all cells that approach the backbone within the time $[1, T]$ must clear the buffers by the end of the time period $[1, T+\lambda]$. Indeed, since all segments of the partition $\{I_1, ..., I_q\}$ are of length λ, and since all call waveforms must be constant over each partition segment by definition, it follows that the flow over the segment $[T+\lambda+1, T+2\lambda]$ must be constant. Consequently, if any cells are left in the buffers after the step $T+\lambda$, it will take until $T+2\lambda$ to clear them all out. This violates Assumption 5(iv). Thus, all cells of the first call cycle must exit the buffers during the time $[1, T+\lambda]$. In accord with this observation, the proof below extends the transmission cycle to $[1, T+\lambda]$ for incoming calls of the first call cycle $[1, T]$.

Now, consider a set of integers $\alpha_1, ..., \alpha_m \in V^m(\{\chi^1, ..., \chi^m\})$. The linear combination

$$z := \Sigma_{i=1}^m \alpha_i \chi^i \tag{15}$$

of the deterministic parts is then constant and nonzero over $[1, T]$. The corresponding linear combination of the actual incoming calls is

$$z(c) := \Sigma_{i=1}^m \alpha_i c^i.$$

Assume that $\beta > 0$ copies of $z(c)$ have been admitted into the network. There are then $\beta\alpha_i$ samples of the call c^i, i = 1, ..., m. Label the samples of c^i by

$$c^{i,j} = \chi^i + \upsilon^{i,j},\ j = 1, ..., \beta\alpha_i.$$

The uncertain parts $\upsilon^{i,j}$ may vary from one sample of the call to another, and are subject to the restriction $0 \le \upsilon^{i,j} \le \rho$.

To simplify our notation, we adopt the (somewhat unusual) convention of setting

$$\Sigma_{i=h}^{j} s_i := 0 \text{ whenever } j < h,$$

for any sequence $s_0, s_1, s_2, ...$

Now, consider a step $k \in \{1, ..., T+\lambda\}$. Recalling that $\tau(i)$ is the maximal delay allowed for cells of the class C^i and using the notation of (6), it follows the number of cells of the call $c^{i,j}$ that must leave the buffers by the end of the step k is

$$\Sigma_{t=1}^{\theta(i,k)} c_t^{i,j} = \Sigma_{t=1}^{\theta(i,k)} (\chi_t^i + \upsilon_t^{i,j}).$$

Consequently, the total number of cells that must leave the buffers by the end of step k is

$$N(k) := \Sigma_{i=1}^{m} \Sigma_{j=1}^{\beta\alpha_i} \Sigma_{t=1}^{\theta(i,k)} (\chi_t^i + \upsilon_t^{i,j}). \tag{16}$$

Define the quantities

$$\sigma^{i,j} := \Sigma_{t=1}^{T} \upsilon_t^{i,j} \le T\rho.$$

$$\sigma(i,\beta\alpha_i) = \Sigma_{j=1}^{\beta\alpha_i} \sigma^{i,j} \le \beta\alpha_i T\rho,\ i = 1, ..., m.$$

$$\sigma(\beta) := \Sigma_{i=1}^{m} \sigma(i,\beta\alpha_i) \le \beta T\rho(\Sigma_{i=1}^{m} \alpha_i) =: \delta(\beta).$$

Then, $\sigma(\beta)$ is the total number of cells included in the uncertain parts of all admitted calls.

Using the linear combination z of (15), define a rational function $\psi(\beta)$ over the interval $[1, T+\lambda]$ by setting

$$\begin{aligned}\psi_k(\beta) &:= \beta z,\ k = 1, ..., T,\\ \psi_k(\beta) &= \sigma(\beta)/\lambda,\ k = T+1, ..., T+\lambda.\end{aligned} \tag{17}$$

Note that $\psi(\beta)$ is constant over the segments $[1, T]$ and $[T+1, T+\lambda]$. The value of $\psi(\beta)$ over $[T+1, T+\lambda]$ is determined by the uncertain parts of the calls.

Next, for an integer $k \in \{1, ..., T+\lambda\}$, define

$$\sigma(k,\beta) := \beta\rho\Sigma_{i=1}^{m} \alpha_i\theta(i,k)\mu(\theta(i,k)),$$

where $\mu(\cdot)$ is the unit step function and $\theta(\cdot,\cdot)$ is from (6). Then,

$$\beta\Sigma_{i=1}^{m} \Sigma_{t=1}^{\theta(i,k)} \alpha_i(\chi_t^i + \rho) = [\beta\Sigma_{i=1}^{m} \Sigma_{t=1}^{\theta(i,k)} \alpha_i\chi_t^i] + \sigma(k,\beta). \tag{18}$$

Combining this with (16) and the fact that $0 \le \upsilon^{i,j} \le \rho$ for all i and j, we obtain

$$N(k) \le \sigma(k,\beta) + \beta\Sigma_{i=1}^{m} \Sigma_{t=1}^{\theta(i,k)} \alpha_i\chi_t^i, \tag{19}$$

$k = 1, ..., T+\lambda$, for all possible uncertainties.

Recalling that z is constant over $[1, T]$, define

$$\begin{aligned}\varphi_k(j,\beta) &:= \beta z,\ k = 1, ..., T\\ \varphi_k(j,\beta) &= \sigma(j,\beta)/\lambda,\ k = T+1, ..., T+\lambda,\end{aligned}$$

$j = 1, ..., T+\lambda$. The following is an auxiliary technical result (compare to [11]).

LEMMA 20. The inequalities

$$N(k) \le \Sigma_{i=1}^{k} \psi_i(\beta),\ k = 1, ..., T+\lambda, \tag{21}$$

are valid for all permissible uncertainties if and only if

$$\beta\Sigma_{r=1}^{m} \Sigma_{i=1}^{\theta(r,k)} \alpha_r(\chi_i^r + \rho) \le \Sigma_{i=1}^{k} \varphi_i(k,\beta), \quad k = 1, ..., T+\lambda. \tag{22}$$

Proof. The Lemma is a consequence of the following facts.

(i) For $k = 1, ..., T$, the right sides of (21) and (22) are identical.

(ii) For $k = 1, ..., T$, formulas (18) and (19) imply that (21) is valid whenever (22) is valid.

(iii) For $k = T+1, ..., T+\lambda$, note that the inclusion of an extra cell in the uncertain part of the left side of (21) cannot increase the right side of (21) by more than one, since $\lambda \ge 1$. Thus, the worst case of the inequalities (21) for each $k = T+1, ..., T+\lambda$ occurs when the uncertain parts included on the left side are at their maximal level ρ. The value of the left side of (21) in that case is given by the left side of (22).

(iv) Fix a value $k \in \{T+1, ..., T+\lambda\}$. For a given value of the left side of (21), the smallest value of the right side of (21) occurs when the uncertain parts are zero at steps that are not included in the sum on the left side of (21). At the step k, such uncertain parts yield $\psi_i(\beta) = \varphi_i(k,\beta)$, $i = T+1, ..., T+\lambda$. Consequently, in this worst case, the right side of (21) becomes equal to the right side of (22).

In conclusion, (22) describes the worst case of (21), and whence, if (22) holds, (21) will be valid for all permissible uncertain parts. ♦

The next statement provides necessary and sufficient conditions for achieving asymptotic efficiency of 1, despite call uncertainties.

THEOREM 23. Let $c^i = \chi^i + \upsilon^i$, $i = 1, ..., m$, be a family of call classes over the partition $\{I_1, ..., I_q\}$ of the interval $[1, T]$, where the uncertain parts satisfy $0 \le \upsilon^i \le \rho$, $i = 1, ..., m$, $\rho > 0$. Assume that the deterministic parts $\chi^1, ..., \chi^m$ form a complete family of call classes over $[1, T]$. Let $\tau(i)$ be the cell delay limitation of the class c^i, $i = 1, ..., m$. Assume that all partition segments $I_1, ..., I_q$ have the same length $\lambda \ge 1$, and that $\tau(i) < 2\lambda$, $i = 1, ..., m$. Let $\psi(1)$ be the rational function defined in (17) for $\beta = 1$. Then, the following two statements are equivalent.

(i) There is a traffic control algorithm that achieves asymptotic efficiency of 1 for the family $F = \{c^1, ..., c^m\}$.

(ii) There is a set of integers $\alpha_1, ..., \alpha_m \in V^m(\{\chi^1, ..., \chi^m\})$ for which the following conditions hold:

(a) $T\rho(\Sigma_{i=1}^m \alpha_i)/\lambda \le \Sigma_{i=1}^m \alpha_i\chi^i$, and

(b) $\Sigma_{r=1}^m \Sigma_{i=1}^{\theta(r,k)} \alpha_r(\chi_i^r + \rho) \le \Sigma_{i=1}^k \varphi_i(k,1)$,
$k = 1, ..., T+\lambda$.

Proof. Recall from the second paragraph of the Appendix that the requirement $\tau(i) < 2\lambda$, $i = 1, ..., m$, implies that all cells of the first cycle $[1, T]$ must leave the buffer system by the end of the step $T+\lambda$. In other words, all cells that remain stored at the end of the first call cycle $[1, T]$ must be transmitted during the segment $I_{q+1} := [T+1, T+\lambda]$.

Assume now that part (i) of the Theorem is valid, and let ϕ be the backbone capacity. Let $\Omega(k)$ be the number of cells entering the backbone at the step k. Then, in view of Theorem 4, asymptotic efficiency of 1 implies that $\Omega(k)$ must be constant over $[1, T]$, say $\Omega(k) = \Omega'$, $k = 1, ..., T$. Further, since I_{q+1} is a partition segment, all waveforms, including $\Omega(k)$, must be constant over it by definition. Let that constant value be $\Omega(k) := \Omega''$, $k = T+1, ..., T+\lambda$. Note that Ω' and Ω'' do not have to be equal, since new calls can be admitted at the step $T+1$.

Due to the unpredictable nature of the uncertain parts $\upsilon^1, ..., \upsilon^m$, the backbone can be asymptotically filled over the interval $[1, T]$ only if the flow during $[1, T]$ is a combination of the deterministic parts $\chi^1, ..., \chi^m$. In view of Theorem 4, this implies that there are integers $\beta_1, ..., \beta_m \in V^m(\{\chi^1, ..., \chi^m\})$ such that $\Omega' = \Sigma_{i=1}^m \beta_i\chi_k^i$, $k = 1, ..., T$.

Let $\beta > 0$ be an integer greatest common divisor of $\beta_1, ..., \beta_m$. Define integers $\alpha_1, ..., \alpha_m \ge 0$ by setting $\beta_i = \beta\alpha_i$, $i = 1, ..., m$. A slight reflection shows that $\alpha_1, ..., \alpha_m \in V^m(\{\chi^1, ..., \chi^m\})$. The number of stored cells at the end of the interval $[1, T]$ is then given by $\sigma(\beta) := \beta\,\Sigma_{k=1}^T \Sigma_{i=1}^m \alpha_i\upsilon_k^i$. Since all stored cells must be released during the segment I_{q+1} whose length is λ, it follows that $\sigma(\beta) = \lambda\Omega''$.

For each $i \in \{1, ..., m\}$, extend now the call class $c^i = \chi^i + \upsilon^i$ of the family F to the interval $[1, T+\lambda]$ by setting $\chi_k^i := 0$ and $\upsilon_k^i := 0$ (whence $c_k^i := 0$) for all $k = T+1, ..., T+\lambda$. We continue to use the symbol F to denote the resulting family of extended call classes. Denoting $z(c) := \Sigma_{i=1}^m \alpha_i c^i$, it follows by our discussion so far that

$$\beta z(c) = \Omega(k), k \in [1, T+\lambda]. \tag{24}$$

Next, consider a step $k \in [1, T+\lambda]$. The fact that cells of a class c^r cannot be delayed by more the $\tau(r)$ steps requires the following: all cells of the class c^r that entered the buffer at the step $k-\tau(r)$ or earlier must leave the buffer by the step k. In view of (24), this yields the inequality

$$\beta\Sigma_{r=1}^{m}\,\Sigma_{i=1}^{k-\tau(r)}\,\alpha_r c_i^r \le \Sigma_{i=1}^{k}\,\Omega_i,\ k = 1, ..., T+\lambda$$

for all $0 \le \upsilon^i \le \rho$. Condition (ii)(b) follows then from Lemma 20, since $\varphi(k,1) = (1/\beta)\varphi(k,\beta)$.

To show that (ii)(a) is also required, recall the constant flow $z := \Sigma_{i=1}^{m}\,\alpha_i\chi^i$, and denote

$$\delta := T\rho(\Sigma_{i=1}^{m}\,\alpha_i)/\lambda - z.$$

Assume, by contradiction, that $\delta > 0$. Consider the special case where the uncertain parts are at their maximum, i.e., $\upsilon^i = \rho$ for all $i = 1, ..., m$. Then, $\sigma(\beta) = \beta T\rho(\Sigma_{i=1}^{m}\,\alpha_i)$, and the flow amplitude Ω'' during the segment I_{q+1} satisfies

$$\Omega'' = \sigma(\beta)/\lambda = \beta T\rho(\Sigma_{i=1}^{m}\,\alpha_i)/\lambda = \beta z + \beta\delta. \qquad (25)$$

Now, in order to fill the backbone asymptotically over the interval $[1, T]$, we have to select β as the integer determined by the division algorithm $\phi = \beta z + \varpi$, where $0 \le \varpi < z$; this guaranties that the largest possible number of packages z are transmitted (see [10] for more details). Substituting into (25), we obtain $\Omega'' = \phi - \varpi + \beta\delta$. This yields $\Omega'' > \phi$ whenever $\beta > z/\delta$, which violates the capacity limit of the backbone for large values of β, a contradiction. Thus, we must have $\delta \le 0$, and condition (ii)(a) must be valid. Together with the earlier part of the proof, this shows that (i) implies (ii).

Conversely, assume that condition (ii) of Theorem 23 holds. It follows then by Lemma 20 that inequalities (21) are valid for all possible uncertain parts $0 \le \upsilon^i \le \rho$, $i = 1, ..., m$. Since $\psi(1)$ is a rational function with at most two values, there is an integer $\beta > 0$ such that $\beta\psi(1)$ is an integer valued function. Define the cell flow

$$\Omega := \beta\psi(1).$$

It follows then by inequality (21) that the following is true for all uncertainties $0 \le \upsilon^i \le \rho$, $i = 1, ..., m$: the flow Ω can transfer the call package $\beta z(c)$ over the interval $[1, T+\lambda]$ without violating the maximal delay restriction $\tau(r)$ of any of the call classes C^r, $r = 1, ..., m$. The fact that the flow Ω is constant over the interval $[1, T]$ indicates that asymptotic efficiency of 1 over $[1, T]$ can be achieved. Finally, the fact that Ω is constant over I_{q+1} and satisfies condition (ii)(b) shows that, with the possible addition of new calls, it can achieve asymptotic efficiency of 1 over I_{q+1} as well. This completes our proof. ♦

Proof (of Theorem 11). We show that the requirement $\rho \le \varepsilon(\alpha_1, ..., \alpha_m)$ is equivalent to conditions (ii) of Theorem 23. Recall that $\varphi_i(k,1) := \Sigma_{r=1}^{m}\,\alpha_r\chi_i^r$ for $i = 1, ..., T$. By moving the term

$$\Sigma_{r=1}^{m}\,\Sigma_{i=1}^{\theta(r,k)}\,\alpha_r\chi_i^r$$

from the left side of the inequality to its right side, condition (ii)(b) of Theorem 23 can be rewritten in the form

$$\Sigma_{r=1}^{m}\,\Sigma_{i=1}^{\theta(r,k)}\,\alpha_r\rho \le \Sigma_{r=1}^{m}\,\Sigma_{i=\max\{1,\theta(r,k)+1\}}^{k}\,\alpha_r\chi_i^r \qquad (26)$$
$$\text{for } k = 1, ..., T,$$

and

$$\Sigma_{r=1}^{m}\,\Sigma_{i=1}^{\theta(r,k)}\,\alpha_r\rho \le$$
$$\Sigma_{r=1}^{m}\,\Sigma_{i=\max\{1,\theta(r,k)+1\}}^{T}\,\alpha_r\chi_i^r + \rho\sigma(k) \qquad (27)$$
$$\text{for } k = T+1, ..., T+\lambda.$$

Moving the terms with ρ in (27) to the left, we get

$$\rho[\Sigma_{r=1}^{m}\,\Sigma_{i=1}^{\theta(r,k)}\,\alpha_r - \sigma(k)] \le$$
$$\Sigma_{r=1}^{m}\,\Sigma_{i=\max\{1,\theta(r,k)+1\}}^{T}\,\alpha_r\chi_i^r \qquad (28)$$
$$\text{for } k = T+1, ..., T+\lambda.$$

Note that (27) is valid for any $\rho > 0$ when the left side of (28) is negative or zero. Dividing (26) and (28) by the coefficient of ρ, yields inequalities whose right sides are given by the right sides of (8) and (9), respectively. Finally, observing that condition (ii)(a) of Theorem 23 can be rewritten in the form $\rho \le \varepsilon_{T+\lambda+1}$, we obtain the conditions of Theorem 11. This concludes the proof. ♦

References

[1] J.D. ATKINS, "Path control: the transport network of SNA," IEEE Transactions on Communications, COM 28, 4, pp. 527–538, (1980).

[2] ATM FORUM, "Closed loop rate based traffic management," Document number 94–0438R2, September 1994.

[3] ATM FORUM, "Traffic management 4.0," document af–tm–0056.0000, (1996)

[4] ATM FORUM, "Technical specifications: approved ATM forum specifications," (1997).

[5] J.C. BOLOT and A.V. SHANKAR, "Dynamical behavior of rate based flow control mechanisms," ACM Computer Communication Review, pp. 35–49, (1990).

[6] CCITT, "Traffic and congestion control in B–ISDN, I.," CCIT Study Group XVIII, Geneva, (1992).

[7] C.S. CHANG, "Stability, queue length, and delay of deterministic and stochastic queueing networks." IEEE Trans. on Automatic Control, AC 39, 5, pp. 913–931, (1994).

[8] M. DECINA and T. TONIATTI, "On bandwidth allocation to bursty virtual connections in ATM networks," Proceedings of the IEEE ICC, Atlanta, GA, (1990).

[9] S.J. GOLESTANI, "A framing strategy for congestion management," IEEE JSAC, SAC 9.7, pp. 1064–1077, (1991).

[10] J. HAMMER, "Sturdy control of discrete communication networks, part I: population control," submitted for publication, (2000).

[11] J. HAMMER, "Sturdy control of discrete communication networks, part II: call reshaping," submitted for publication, (2000).

[12] J. HAMMER, "Asymptotic efficiency and admission control for discrete communication networks", Proceedings of the American Control Conference, June 2001.

[13] R. HANDEL, M.N. HUBER, and S. SCHRODER, "ATM networks, Concepts, Protocols, Applications," second edition, Addison–Wesley, Wokingham, England, (1994).

[14] K.K. RAMAKRISHNAN and R. JAIN, "A binary feedback scheme for congestion avoidance in computer networks," ACM Transactions on computer systems, 8, 2, pp. 158–181, (1988).

[15] E.P. RATHGEB, "Modeling and performance comparison of policing mechanisms for ATM networks," IEEE JSAC, SAC 9, 3, pp. 325–334, (1991).

[16] M. SCHWARTZ, "Broadband integrated networks," Prentice Hall, N.J., (1996).