Maximal Effective Bandwidth of Constrained Traffic

Cormac Walsh

INRIA Rocquencourt
B.P. 105 78153
Le Chesnay Cedex
France

March 14, 2002

Abstract

We investigate the worst possible behavior of a stationary traffic source when the traffic emanating from it is required to meet certain constraints. Specifically, the peak rate of the source is required not to exceed a level ρ and realizations must obey a "leaky bucket" constraint with bucket size β and leak rate σ. The worst case source is considered to be the one with the largest effective bandwidth, a concept which arises in the theory of Large Deviations and governs the asymptotic loss rate when a large number of sources send traffic to a single server queue. We conjecture the form of the worst case traffic in general and prove the conjecture for the special case when T, the time-scale parameter of the effective bandwidth, is less than both $\beta/(\rho-\sigma)$ and β/σ, the times taken respectively to fill and empty the leaky bucket.

Keywords: regulated traffic, worst case, statistical multiplexing, stationary independent, large deviations, effective bandwidth, leaky bucket, Markov decision procedure.

1 Motivation

Consider a telecommunications network to which traffic arrives from several sources. To guarantee sufficient quality of service, for example bounds on the cell delay or loss, the network must have a reasonable description of the characteristics of the offered traffic. This may be given at connection setup when a traffic contract is agreed. The user contracts that his traffic will conform to certain parameters, perhaps defining a policing algorithm that bounds properties of the traffic. In return the network contracts to carry this traffic with a particular quality of service. The traffic contract gives the network information that will bound the amount of network resources required to carry the call.

The network may have no additional information about user traffic and must therefore infer its characteristics from the descriptors specified in the contract. In any decisions the network makes concerning resource allocation or call acceptance, it would be prudent of the network to assume the worst case, that is that the user is adversarial to the maximum extent permitted by the policer.

These considerations motivate an optimization problem over the space of traffic processes. We must choose a constraint to represent the action of the policer and a functional to represent the performance of the network.

A constraint common in the literature is the "leaky bucket" mechanism described first in [16]. This consists of a counter for each connection that is incremented whenever a cell from that connection arrives and is decremented periodically. If the counter exceeds a fixed threshold, the cell responsible is declared a non-conforming cell. These cells are discarded, delayed, or tagged as a low priority cells. The user specifies both the rate σ at which the counter is incremented and the threshold β. We say that a traffic stream conforms to the leaky bucket constraint if none of its cells are non-conforming. The leaky bucket is the standard policing mechanism defined by the ATM Forum [1] under the name Generic Cell Rate Algorithm.

To simplify analysis, we will work with a fluid model. There is an obvious continuous version of the leaky bucket regulator. Here fluid flows into a buffer at a rate equal to that of the arriving traffic stream and is drained at a constant rate σ. A stream conforms to the leaky bucket constraint (σ, β) if the fluid never exceeds a level β.

With a single constraint of the form $(\rho, 0)$, the worst case source is trivial under any reasonable measure of network performance. Namely, it is the source that transmits constantly at rate ρ. Our concern will be with the next simplest case, that of two constraints: a peak rate constraint $(\rho, 0)$ and a mean rate constraint (σ, β).

There are many functionals that could represent the degree of congestion in a network. The simplest are those which arise when the source being optimized is the only one present. For many functionals of this type, the worst case is known. For example, Lee [10] shows that the worst case average queueing delay for a single source occurs when the source behaves in a periodic manner, transmitting firstly at the peak rate, then at rate σ, and finally at rate 0. Kesidis and Konstantopoulos [9] have a similar result for the source that maximizes the time the queue length spends above any given level.

More generally, one wishes to consider several constrained sources transmitting simultaneously. If the sources are not assumed to be stationary and independent, then in the worst case they conspire by synchronizing their bursts of traffic. The traffic patterns sent are similar to the single source worst case pattern [6]. In many situations however, independence is a natural assumption. Since it may greatly increase the multiplexing gain achieved, it would be advantageous to take it into account.

Examples of simple functionals incorporating independence include those arising from buffer-less systems. In general, the worst case sources with respect to these are "on-off", that is, alternate between transmitting at the peak rate and rate zero [7, 13]. With buffering, however, this is not true: counterexamples have been found which are worse than on-off sources [7, 14].

One approach to dealing with systems in which there is both independence and buffering is to find bounds on the functional of interest rather than attempt to calculate the exact worst case. This has been done by several authors [15, 11, 2].

In this paper, we calculate the exact worst case for the *effective bandwidth* functional, which is defined in the next section. This functional arises in theory of Large Deviations and is relevant in the limit as the number of sources becomes large. Kelly [4] has previously obtained results for a discrete-time packetized version of this problem.

2 The Problem

Let $\mathbb{D}$ be the set of non-negative, σ-finite measures on $[0,\infty)$. We interpret the elements of this set as the possible realizations of a traffic source. If $b \in \mathbb{D}$, then $b[s,t]$ gives the amount of traffic arriving in the time interval $[s,t]$. We denote by C those elements b of $\mathbb{D}$ for which

$$\begin{aligned} b[t_1,t_2] &\leq (t_2-t_1)\rho && (1)\\ \text{and} \quad b[t_1,t_2] &\leq (t_2-t_1)\sigma+\beta, && (2)\end{aligned}$$

for all $t_2 \geq t_1 \geq 0$. Condition (1) is referred to as the peak rate constraint and Condition (2) is referred to as the leaky bucket constraint. To avoid trivialities, we assume that $\rho > \sigma > 0$ and $\beta > 0$.

On $\mathbb{D}$, we use the σ-algebra generated by the topology of weak convergence. We identify traffic sources with probability measures on $\mathbb{D}$. Realizations are shifted forward in time with the operator $\Theta_h : \mathbb{D} \to \mathbb{D}$ defined, for each $h \geq 0$, by $(\Theta_h b)[B] := b[B+h]$ for all Borel sets $B \subset [0,\infty)$. When we talk of stationary sources we mean stationary with respect to this operator. In other words, a measure μ on $\mathbb{D}$ is stationary if $\mu[\Theta_h^{-1}\mathcal{B}] = \mu[\mathcal{B}]$ for all measurable subsets $\mathcal{B}$ of $\mathbb{D}$ and shifts $h \geq 0$. We are interested in the set of stationary sources that meet the constraints almost surely. These are represented by the set $\mathcal{C}$ of stationary probability measures μ on $\mathbb{D}$ for which $\mu[C] = 1$.

The *effective bandwidth* of a source μ is defined to be

$$\mathcal{E}_{\vartheta,T}(\mu) := \frac{1}{\vartheta T}\log \mathbb{E}_\mu e^{\vartheta b[0,T]}.$$

The two parameters $\vartheta \in \mathbb{R}$ and $T \in [0,\infty)$ are called the *space-scale* and the *time-scale* respectively. The effective bandwidth is exactly what is needed to calculate the asymptotics of the loss rate as the number of sources becomes infinite. Suppose there are J types of source, with type $j \in \{1,\dots,J\}$ having law μ_j. Let n_j be the proportion of sources of type j; thus $\sum_{j=1}^J n_j = 1$. Denote by $\mathbb{L}(S,B,N)$ the loss rate when a single server queue with service rate S and buffer size B is fed by N independent sources, where the proportion of each type is fixed by the n_j. As was shown in [5], $\mathbb{L}$ obeys the large deviations asymptotics

$$\lim_{N\to\infty}\frac{1}{N}\log \mathbb{L}(sN,bN,N) = -\inf_{T>0}\sup_{\vartheta\in\mathbb{R}}\Big[(b+sT)\vartheta - \vartheta T\sum_{j=1}^J n_j\mathcal{E}_{\vartheta,T}(\mu_j)\Big]. \tag{3}$$

These asymptotics are also obeyed by the proportion of time the buffer is full and by the probability that the queue length exceeds a level bN in an infinite buffer [3]. The values

T^* and θ^* where the infimum and supremum are attained define the operating point of the system. The time-scale parameter T^* will be the typical duration of the busy period before overflow. The space-scale parameter θ^* indicates the degree of statistical multiplexing obtained: for low values of θ, $\mathcal{E}_{\vartheta,T}(\mu)$ approaches the mean rate of source μ, while for high values it approaches the maximum rate that μ can sustain over an interval of length T. For an overview of the effective bandwidth function, see the paper of Kelly [8].

One application of (3) is in Connection Admission Control. Here, a user requests admission to the network and the network operator must decide whether to admit the user. Suppose the network can be modeled as a single server queue with service rate S and buffer size B, and that the loss is required to be kept less than some value $e^{-\eta}$. Equation (3) may be used as the basis for an approximation: the loss rate if the user is admitted will be approximately $e^{-I(B,S)}$, where

$$I(B,S) := -\inf_{T>0}\sup_{\vartheta\in\mathbb{R}}\left[(B+ST)\vartheta - \vartheta T\Big(\sum_{n=1}^{N}\mathcal{E}_{\vartheta,T}(\mu_n) + \mathcal{E}_{\vartheta,T}(\mu_{\text{new}})\Big)\right]. \tag{4}$$

Here μ_{new} is the new user and $\{\mu_n : 1 \leq n \leq N\}$ are the users already in the system. We will assume that the effect of the new user will be small in comparison with that of those already present and therefore that the operating point will not change significantly. The criterion for admission will then be

$$\mathcal{E}_{\theta^*,T^*}(\mu_{\text{new}}) \leq \frac{B}{T^*} + S - \frac{\eta}{\vartheta^* T^*} - \sum_{n=1}^{N}\mathcal{E}_{\theta^*,T^*}(\mu_n).$$

The effective bandwidth of the current users and the operating point may be estimated using records of their behavior so far [12]. The problem remains, however, of how to predict the effective bandwidth of the new user. We adopt a conservative approach: use the largest effective bandwidth consistent with the traffic parameters it declares.

Thus we are led to consider the optimization program

$$\text{maximize } \mathcal{E}_{\vartheta,T}(\mu), \qquad \text{subject to } \mu \in \mathcal{C}. \tag{5}$$

For each γ in the range $[0,T]$, let w_γ be the source that transmits a periodic traffic pattern with uniformly distributed phase and whose sample paths are composed of following four sections:

- a peak rate burst of duration $t_\rho := \beta/(\rho-\sigma)$,
- an interval in which the source transmits at rate σ, of duration γ,
- a silent interval of duration $t_0 := \beta/\sigma$,
- another interval at rate σ, again of duration γ.

Note that t_ρ and t_0 are, respectively, just long enough to fill the leaky bucket from empty and to empty it from full. A source of this type is illustrated in Figure 1. We make the following conjecture.

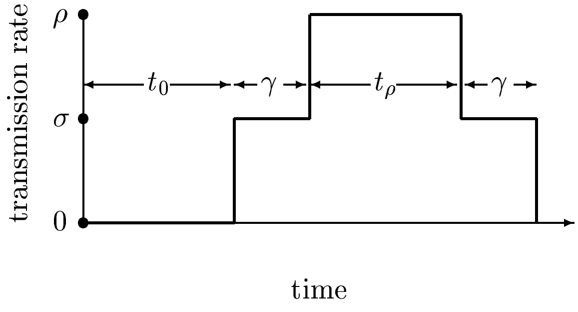

Figure 1: The "wedding cake" pattern, conjectured to have the worst case effective bandwidth.

Conjecture 1 *For each ϑ and T in $[0,\infty)$, the stationary traffic source that maximizes the effective bandwidth $\mathcal{E}_{\vartheta,T}$ under Constraints (1) and (2) is one of the sources $\{w_\gamma\}_{0\le\gamma\le T}$.*

Note that this conjecture allows different sources to be worst case for different values of ϑ and T. If the conjecture is true, the problem of finding the worst case effective bandwidth for any given ϑ and T reduces to a one-parameter optimization over γ, which we now describe.

Consider the source w_γ. Let b be one of its realizations and let $z(t) := b[t, t+T]$ for each $t \in [0,\infty)$. Then

$$\mathcal{E}_{\vartheta,T}(w_\gamma) = \frac{1}{\vartheta T}\log\frac{1}{p}\int_0^p e^{\vartheta z(x)}\,dx,$$

where $p := 2\gamma + t_0 + t_\rho$ is the period of w_γ. In this paper we consider the case when $T \le \min(t_0, t_\rho)$. For T in this range, z will be of the form shown in Figure 2. Using the identity

$$\int_s^t e^{\vartheta(c+mx)}\,dx = \frac{1}{m\vartheta}\left(e^{\vartheta(c+mt)} - e^{\vartheta(c+ms)}\right),$$

we may write down an explicit expression for the effective bandwidth of w_γ:

$$\mathcal{E}_{\vartheta,T}(w_\gamma) = \frac{1}{\vartheta T}\log\frac{1}{p}\sum_{n=0}^{3}\alpha_i e^{\vartheta A_i},$$

where

$$\begin{aligned}
\alpha_0 &= (t_0 - T) - \frac{2}{\vartheta\sigma}, & \alpha_1 &= \frac{2}{\vartheta\sigma} - \frac{2}{\vartheta\rho},\\
\alpha_2 &= \frac{2}{\vartheta\rho} - \frac{2}{\vartheta(\rho-\sigma)}, & \alpha_3 &= \frac{2}{\vartheta(\rho-\sigma)} + (t_\rho - T),
\end{aligned}$$

and

$$\begin{aligned}
A_0 &= 0, & A_1 &= \gamma\sigma,\\
A_2 &= \gamma\sigma + \rho(T-\gamma), & A_3 &= \rho T.
\end{aligned}$$

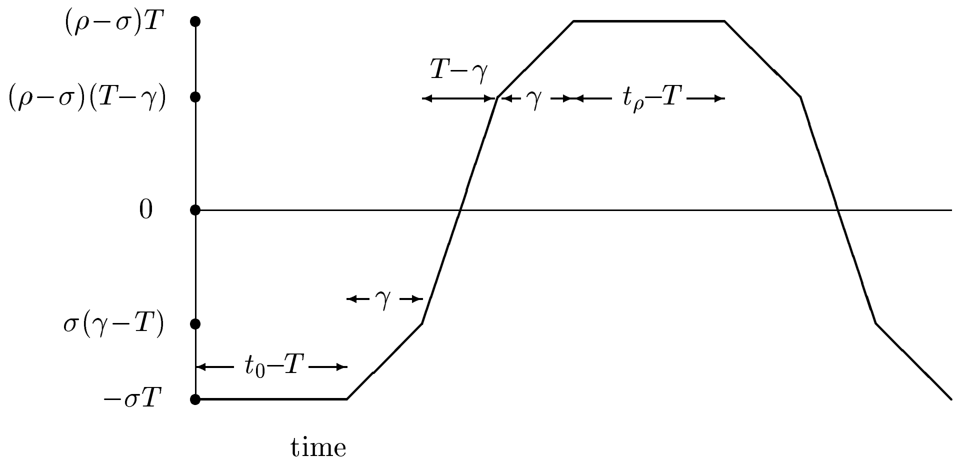

Figure 2: Graph of $z(t) := b[t, t+T]-\sigma T$ in the region $[0, t_0+t_\rho+2\gamma]$

The one-parameter optimization involved in maximizing this expression over γ is easy to perform numerically.

Our main result is the following.

Theorem 1 *Fix ϑ and T in $[0, \infty)$. Suppose that $T \leq \min[\beta/(\rho-\sigma), \beta/\sigma]$. Then the supremum of $\mathcal{E}_{\vartheta,T}$ over all sources in $\mathcal{C}$ is attained by a source w_γ, for some γ in the range $0 \leq \gamma \leq T$.*

The proof of this theorem will form the remainder of the paper.

3 Preliminaries

We say that a realization $b \in \mathbb{D}$ is periodic with period $p > 0$ if $\Theta^p b = b$. Given such a realization, we may construct a stationary traffic source μ by taking a uniform distribution over the phase. More precisely,

$$\mu[\mathcal{B}] := \frac{1}{p}\,\mathrm{Leb}\{t \in [0,p) : \Theta^t b \in \mathcal{B}\}, \qquad \text{for all measurable subsets } \mathcal{B} \text{ of } \mathbb{D}.$$

Denote by P_p the set of sources formed in this way from elements of C which are periodic with period p. We showed in [17] that

$$\sup\{\mathcal{E}_{\vartheta,T}(\mu) : \mu \in \mathcal{C}\} = \sup\{\mathcal{E}_{\vartheta,T}(\mu) : \mu \in P_p,\ p > 0\}. \tag{6}$$

We will use this result in the present paper. Note that the effective bandwidth of the source μ, formed from b in the manner above, is given by

$$\mathcal{E}_{\vartheta,T}(\mu) = \frac{1}{\vartheta T} \log \frac{1}{p} \int_0^p e^{\vartheta b[x,x+T]}\,dx. \tag{7}$$

We denote by C_p the subset of C consisting of those realizations which are periodic with period p and have an average transmission rate σ, the leak rate of the leaky bucket. The following lemma shows that we may restrict our attention to the set of sources constructed from elements of C_p by randomizing the phase.

Lemma 1 *Let $\mu \in \mathcal{C}$. Then, for some $p > 0$, there exists $c \in C_p$ such that $\mathcal{E}_{\vartheta,T}(\nu) \geq \mathcal{E}_{\vartheta,T}(\mu)$, where ν is the source constructed from c by choosing the phase uniformly.*

Proof. Since Equiation (6) holds, we may assume that $\mu \in P_p$. Let b be one of its realizations. Then b has an obvious extension to the whole of $\mathbb{R}$ defined by

$$\hat{b}[np, np+t] = b[0,t], \qquad t \in [0,p],\, n \in \mathbb{Z}.$$

The content of the leaky bucket at time t when fed by $\hat{b}$ is given by $q(t) := \sup_{s \leq t}[\hat{b}(s,t] - (t-s)\sigma]$. Now consider the realization defined by

$$c[s,t] := q(t) - q(s) + (t-s)\sigma, \qquad \text{for all } t \geq s \geq 0.$$

This is the traffic realization which has the same rate as b when the leaky bucket contains some fluid but has rate σ when it is empty. It is a simple exercise to show that c obeys Constraints (1) and (2) and that c is periodic with period p. Furthermore, the average rate of c is σ. Now, $c \geq b$ in the sense that $c[B] \geq b[B]$ for every Borel subset of $[0,\infty)$. The conclusion follows since the formula for the effective bandwidth in Equation (7) is isotone in b. □

In addition to satisfying Constraints (1) and (2), each $b \in C_p$ also satisfies

$$b[s,t] \geq (t-s)\sigma - \beta, \qquad \text{for all } s,t \in [0,\infty),\, t \geq s.$$

This leads to a certain duality in our lemmas.

Given the nature of the effective bandwidth function, it will prove useful to represent traffic realizations in the following form. For any $b \in \mathbb{D}$ we may define the function

$$z(t) := b[t, t+T] - \sigma T, \qquad \text{for all } t \in [0,\infty), \tag{8}$$

which specifies the amount of traffic in each block of length T in excess of σT, the maximum amount which could leak from the leaky bucket. Let Z_p be the set of functions obtained in this way from elements of C_p, that is

$$Z_p := \{z \in \mathcal{C}[0,\infty) : \text{there exists } b \in C_p \text{ such that } z(t) = b[t,t+T] - \sigma T \text{ for all } t \in [0,\infty)\}.$$

Note that, because of Constraint (1), each element of Z_p is Lipschitz with Lipschitz constant ρ. Our second lemma will give another consequence of the constraints on the realizations in C_p. This will take the form of a bound on the area under each $z \in Z_p$ in certain intervals. The bound depends only on the values taken by z at the end points of the interval.

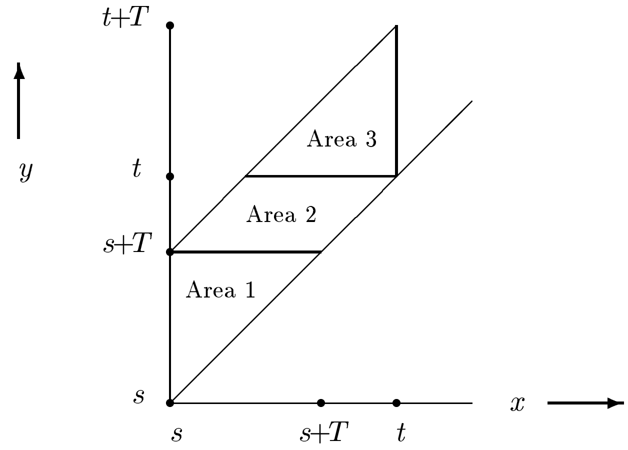

Figure 3: The range of integration of the double integral in the proof of Lemma 2.

Lemma 2 *Suppose $z \in Z_p$ and $s, t \in [0, \infty)$, $s \le t$. If $z(s) \ge 0$ and $z(t) \ge 0$, then*

$$\int_s^t z(x)\,dx \le \beta T - \frac{[z(s)]^2 + [z(t)]^2}{2(\rho - \sigma)}.$$

If $z(s) \le 0$ and $z(t) \le 0$, then

$$\int_s^t z(x)\,dx \ge -\beta T + \frac{[z(s)]^2 + [z(t)]^2}{2\sigma}.$$

Proof. We will only prove the first statement since the proof of the second is similar. Let $b \in C_p$ be related to z as in (8). Define the signed measure $\hat{b} := b - \sigma\mathrm{Leb}$ and the constants

$$\begin{aligned} l &:= \sup\left\{\hat{b}(x, s+T) : x \in [s, s+T]\right\}, \\ m &:= \hat{b}(s+T, t), \\ n &:= \sup\left\{\hat{b}(t, x) : x \in [t, t+T]\right\}. \end{aligned}$$

Here $\hat{b}(s+T, t)$ is interpreted to mean $-\hat{b}(t, s+T)$ if $s+T > t$. Since b is absolutely continuous with respect to the Lebesgue measure,

$$I := \int_s^t z(x)\,dx = \int_s^t \left(\int_x^{x+T} \hat{b}'(y)\,dy\right) dx.$$

We may rewrite this double integral as follows (see Figure 3):

$$I = \int_s^{s+T}\int_x^{s+T} \hat{b}'(y)\,dy\,dx + \int_{s+T}^t \int_{y-T}^y \hat{b}'(y)\,dx\,dy + \int_{t-T}^t \int_t^{x+T} \hat{b}'(y)\,dy\,dx. \qquad (9)$$

Define $V(x) := \min[l, (\rho-\sigma)(s+T-x)]$. We have that $\int_x^{s+T} \hat{b}'(y)\,dy \le V(x)$ for all $x \in [s, s+T]$, and so the first term on the right hand side of (9) is no greater than

$$\int_s^{s+T} V(x)\,dx = lT - \frac{l^2}{2(\rho-\sigma)}.$$

Similarly, the third term is less than $nT - n^2/2(\rho-\sigma)$. The middle term is equal to mT. Thus

$$I \le lT - \frac{l^2}{2(\rho-\sigma)} + mT + nT - \frac{n^2}{2(\rho-\sigma)}.$$

But $l+n+m \le \beta$ and $l \ge z(s)$ and $n \ge z(t)$. The result follows. $\square$

Motivated by this lemma, we define, for $x, y \ge 0$,

$$\begin{aligned} \overline{f}(x,y) &:= \beta T - \frac{x^2+y^2}{2(\rho-\sigma)} \\ \text{and} \qquad \underline{f}(x,y) &:= \beta T - \frac{x^2+y^2}{2\sigma}. \end{aligned}$$

If either x or y is negative, we use the convention that $\overline{f}(x,y)$ and $\underline{f}(x,y)$ both take the value $+\infty$.

4 A Comparison Lemma

In what follows, the symbol I denotes the indicator function, in other words $\mathrm{I}_A(t)$ is 1 if condition A holds at t, and 0 otherwise.

Lemma 3 *Suppose that $D_1, D_2 \subset \mathbb{R}$ are Borel sets and that $z_1 : D_1 \to \mathbb{R}^+$ and $z_2 : D_2 \to \mathbb{R}^+$ are non-negative measurable functions such that*

$$\int_{D_1} z_1(t)\,\mathrm{I}_{z_1(t)\le l}\,dt \ge \int_{D_2} z_2(t)\,\mathrm{I}_{z_2(t)\le l}\,dt$$

for all $l \ge 0$, and that $\int_{D_1} z_1(t)\,dt = \int_{D_2} z_2(t)\,dt$. Suppose also that Leb$\{t \in D_2 : z_2(t) = 0\} = 0$ and that $\zeta \ge 1$. Then, for any $\vartheta \in \mathbb{R}$,

$$\int_{D_1} (e^{\vartheta z_1(t)} - \zeta)\,dt \le \int_{D_2} (e^{\vartheta z_2(t)} - \zeta)\,dt.$$

Proof. Since $\int_{D_1} (e^{\vartheta z_1(t)} - \zeta)\,\mathrm{I}_{z_1(t)=0}\,dt = (1-\zeta)\,\mathrm{Leb}\{t \in \mathbb{R} : z_1(t) = 0\} \le 0$, we may assume without loss of generality that $\mathrm{Leb}\{t \in \mathbb{R} : z_1(t) = 0\} = 0$. For each $l > 0$, let

$$G_i(l) := \int_{D_i} z_i\,\mathrm{I}_{z_i(t)\le l}\,dt,$$

where i can be either 0 or 1. Then G_i is a distribution function on $(0,\infty)$ with an associated measure μ_i. By hypothesis, $G_1 \ge G_2$. Equivalently, $\mu_1 \le \mu_2$ in the usual stochastic ordering. (Although μ_1 and μ_2 will not necessarily be probability measures, we have that $\mu_1(0,\infty) = \mu_2(0,\infty)$ and so there is no difficulty in comparing them). The function

$$k : (0,\infty) \to \mathbb{R} : x \mapsto \frac{e^{\vartheta x} - \zeta}{x}$$

is non-decreasing and therefore $\int k\, d\mu_1 \leq \int k\, d\mu_2$ since both integrals exist. But, for any function $f : \mathbb{R} \to \mathbb{R}$, we have that $\int f\, d\mu_i = \int z_i(t).(f \circ z_i)(t)\, dt$. The conclusion follows after applying this to k. □

The proof of Theorem 1 will involve comparing each section of $z \in Z_p$ that lies either entirely above or entirely below the x-axis with one of a set of functions which are in some sense optimal. We now introduce these functions.

5 Definition of a Family of Functions

On the interval $[0, T + t_\rho]$, define the function

$$\overline{m}(x) := (\rho - \sigma) \min(x,\, T,\, T + t_\rho - x).$$

We wish to define a two-parameter family $\{\overline{z}_{u,v} : u, v \in [0, \beta T], u \geq v\}$ of non-negative, continuous functions. Let $x(u)$ be such that

$$\int_0^{x(u)} \overline{m}(t)\, dt - \frac{\overline{m}^2(x(u))}{2\rho} = \beta T - u$$

and $y(v)$ be such that

$$\int_{y(v)}^{T+t_\rho} \overline{m}(t)\, dt - \frac{\overline{m}^2(y(v))}{2\rho} = v.$$

Our definition of $\overline{z}_{u,v}$ proceeds differently according to whether $x(u)$ is less than $y(v)$ or not. In the former case we define $\overline{z}_{u,v}$ so that

- $\overline{z}_{u,v}$ agrees with $\overline{m}$ on the interval $[x(u), y(v)]$,
- $\overline{z}_{u,v}$ has slope ρ to the left of $x(u)$ and slope $-\rho$ to the right of $y(v)$.

On the other hand, for those u and v for which $x(u) > y(v)$ we make an alternative definition of $\overline{z}_{u,v}$: it is the unique continuous function defined on $[0, 2\sqrt{(u-v)/\rho}]$ that passes through the origin and has slope ρ over the first half of its domain and $-\rho$ over the second half. Functions of the first and second type are shown in Figures 4 and 5 respectively.

In both cases, we consider the domain of $\overline{z}_{u,v}$, which we denote by $\overline{D}(u,v)$, to be the maximal interval over which this function is defined and non-negative. Also, we define

$$\overline{x}(u,v) := \sup\{t \in \overline{D}(u,v) : \overline{z}'_{u,v}(t) = \rho\} \tag{10}$$

and

$$\overline{y}(u,v) := \inf\{t \in \overline{D}(u,v) : \overline{z}'_{u,v}(t) = -\rho\}, \tag{11}$$

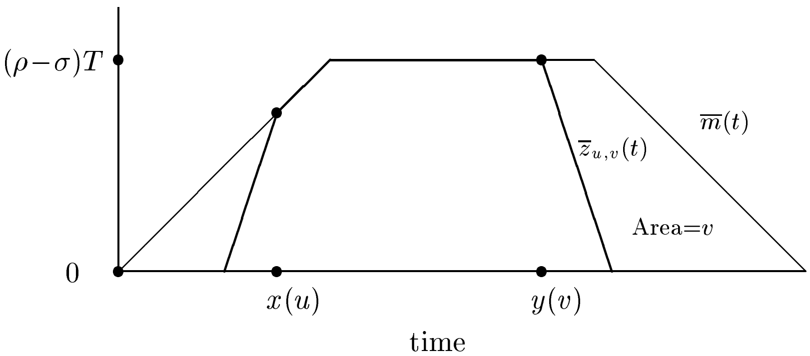

Figure 4: Graph of $\overline{z}_{u,v}$ when $x(u) \leq y(v)$.

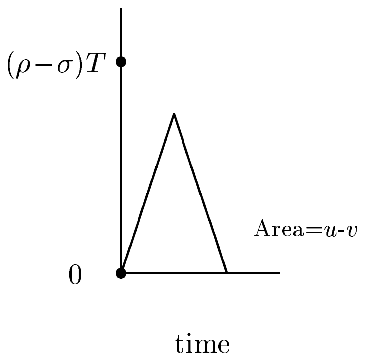

Figure 5: Graph of $\overline{z}_{u,v}$ when $x(u) > y(v)$.

where the dash denotes differentiation. Note that if $\overline{z}_{u,v}$ is of the first type then $\overline{x}(u,v) = x(u)$ and $\overline{y}(u,v) = y(v)$, while if it is of the second type then $\overline{x}(u,v) = \overline{y}(u,v)$ and the common value is where $\overline{z}_{u,v}$ attains its maximum.

We also wish to define $\underline{z}_{u,v}$ for $u, v \in [0, \beta T]$, $u \geq v$. The simplest way is to repeat the above procedure with σ replacing $\rho - \sigma$ and t_0 replacing t_ρ in the definition of $\overline{m}$, and then to multiply the resulting function by -1. The domain of $\underline{z}_{u,v}$, which we denote by $\underline{D}(u,v)$, is considered to be the maximal interval over which this function is defined and non-positive. In general, when we introduce a quantity denoted by an overlined symbol, we will take it for granted that a similar definition applies for the same symbol underlined.

We are interested in the following function:

$$\overline{R}(u,v) := \int_{\overline{D}(u,v)} e^{\vartheta \overline{z}_{u,v}(t)}\, dt.$$

This function is differentiable with respect to both arguments:

$$\frac{\partial}{\partial u}\overline{R}(u,v)) = \frac{e^{\vartheta \overline{j}(u,v)}}{\overline{j}(u,v)}, \qquad \text{and} \qquad \frac{\partial}{\partial v}\overline{R}(u,v)) = -\frac{e^{\vartheta \overline{k}(u,v)}}{\overline{k}(u,v)}, \tag{12}$$

where $\overline{j}(u,v) := \overline{z}_{u,v}(\overline{x}(u,v))$ and $\overline{k}(u,v) := \overline{z}_{u,v}(\overline{y}(u,v))$. We also have that

$$\frac{\partial}{\partial u}\mathrm{Leb}\overline{D}(u,v) = \frac{1}{\overline{j}(u,v)}, \qquad \text{and} \qquad \frac{\partial}{\partial v}\mathrm{Leb}\overline{D}(u,v) = -\frac{1}{\overline{k}(u,v)}. \tag{13}$$

Suppose now that $T < \min(t_0, t_\rho)$. Observe that if b is a realization of the source w_γ and z is constructed from b according to (8), then the area under the positive section of z over a single period is $a := \beta T - (T-\gamma)^2(\rho - \sigma)\sigma/\rho$. Moreover, this section is identical to $\overline{z}_{u_a,v_a}$ with $u_a := (\beta T + a)/2$ and $v_a := (\beta T - a)/2$. Similarly, the negative section of z is identical to $\underline{z}_{u_a,v_a}$. This motivates the following definition:

$$\lambda := \sup_{a \in (0,\beta T]} \frac{\overline{R}(u_a,v_a) + \underline{R}(u_a,v_a)}{\mathrm{Leb}\overline{D}(u_a,v_a) + \mathrm{Leb}\underline{D}(u_a,v_a)} \tag{14}$$

Clearly, λ is related to the one-parameter optimization problem described in the Introduction.

Lemma 4 *If $T < \min(t_0, t_\rho)$ and $\vartheta > 0$, then $\lambda > 1$ and*

$$\sup_{\gamma \in [0,\beta T]} \mathcal{E}_{\vartheta,T}(w_\gamma) = \sigma + \frac{\log \lambda}{\vartheta T}.$$

Proof. We need to show that a^*, the point at which the supremum in (14) is attained, lies in the range $[\eta, \beta T]$, where $\eta := \beta T - T^2(\rho - \sigma)\sigma/\rho$. Let $r(a)$ and $p(a)$ be the numerator and denominator respectively of the fraction appearing in the definition of λ. The derivatives of r and p may be computed using (12) and (13) and the corresponding expressions involving $\underline{R}(u,v)$ and $\underline{D}(u,v)$. It will be noticed that $r'(a)/p'(a)$ is the ordinate of the intersection with the y-axis of the line segment joining the two points

$$\left(\overline{\jmath}(u_a,v_a), e^{\vartheta \overline{\jmath}(u_a,v_a)}\right) \qquad \text{and} \qquad \left(\underline{j}(u_a,v_a), e^{\vartheta \underline{j}(u_a,v_a)}\right).$$

As such, it increases whenever $\overline{\jmath}(u_a,v_a)$ or $-\underline{j}(u_a,v_a)$ do. But both of these are non-decreasing in a in the range $(0,\eta]$, and thus $r'(a)/p'(a)$ is also non-decreasing in this range. It follows that so also is $r(a)/p(a)$. This proves the second part of the lemma. The first part follows from the fact that $\lim_{a\to 0} p(a)/r(a) = 1$. □

Let $\overline{h} := \overline{\jmath}(u_{a^*}, v_{a^*})$ and $\underline{h} := \underline{j}(u_{a^*}, v_{a^*})$. The following relation between these values and λ will be useful. Differentiating $r(a)/p(a)$ and setting the derivative equal to zero, we find that $p(a^*)r'(a^*) = r(a^*)p'(a^*)$. Thus

$$\lambda = \frac{r(a^*)}{p(a^*)} = \frac{\overline{h}e^{\vartheta \underline{h}} - \underline{h}e^{\vartheta \overline{h}}}{\overline{h} - \underline{h}}. \tag{15}$$

6 Optimality of the Functions

The following properties of $\overline{z}_{u,v}$ are important. Fix $l > 0$ and define

$$m := \inf\{t \in \overline{D}(u,v) : \overline{z}_{u,v}(t) \geq l\} \tag{16}$$

$$\text{and} \qquad n := \sup\{t \in \overline{D}(u,v) : \overline{z}_{u,v}(t) \geq l\}. \tag{17}$$

We have that, for each $u, v \in [0, \beta T]$, $u \geq v$

- the area under $\overline{z}_{u,v}$ over its domain $\overline{D}(u,v)$ is $u - v$,
- $\overline{z}_{u,v}$ is unimodal,
- if $m \leq \overline{x}(u,v)$ then the area under $\overline{z}_{u,v}$ to the left of m is $l^2/2\rho$. If $m > \overline{x}(u,v)$ and $l \leq \sup_{\overline{D}(u,v)} \overline{z}_{u,v}$ then the area under $\overline{z}_{u,v}$ to the right of m is $\overline{f}(l,0) - v$,
- similarly, if $n \geq \overline{y}(u,v)$ then the area under $\overline{z}_{u,v}$ to the right of n is $l^2/2\rho$. If $n < \overline{y}(u,v)$ and $l \leq \sup_{\overline{D}(u,v)} \overline{z}_{u,v}$ then the area under $\overline{z}_{u,v}$ to the left of n is $\overline{f}(0,l) - \beta T + u$,

We will use these properties to prove the following lemma. But first we introduce some notation: for all $x \in \mathbb{R}$, $x^+ := \max(x,0)$ and $x^- := \min(x,0)$. Also, if z is a real valued function on $[0,\infty)$, then $z[s,t] := \int_s^t z$.

Lemma 5 *Suppose $z \in Z_p$ is such that $z(s) = z(t) = 0$ for some $s, t \in [0,\infty)$, $s \leq t$, and that $z[s,t] \geq 0$. Then, letting $a := z[s,t]$ and $v := \inf_{\tau\in[0,t]} (\overline{f}(z(\tau),0) - z[\tau,t])$, we have that*

$$\int_s^t z^+(\tau)\, \mathrm{I}_{z^+(\tau)\leq l}\, d\tau \geq \int_{\overline{D}(v+a,v)} \overline{z}_{v+a,v}(\tau)\, \mathrm{I}_{\overline{z}_{v+a,v}(\tau)\leq l}\, d\tau, \tag{18}$$

for any $l > 0$.

Proof. We first establish two inequalities which the constraints and the hypothesis impose on z. Firstly, let $\epsilon > 0$. Then there exists $\tau \le t$ such that $\overline{f}(z(\tau), 0) - z[\tau, t] \le v + \epsilon$. Now, for each $x \in [s, t]$, $z[\tau, x]$ is bounded above by $\overline{f}(z(\tau), z(x))$ if $\tau \le x$, and by $-z^-[x, t]$ if $\tau > x$. Thus, in either case it is bounded above by the sum of these two quantities. Therefore

$$\begin{aligned} z[s,x] &= z[\tau,x] + z[s,t] - z[\tau,t] \\ &\le \overline{f}(z(\tau),z(x)) - z^-[x,t] + a + v + \epsilon - \overline{f}(z(\tau),0). \end{aligned}$$

Using the identity $\overline{f}(\zeta, 0) - \overline{f}(\zeta, \eta) = \beta T - \overline{f}(\eta, 0)$ for all $\zeta, \eta \ge 0$, and the fact that ϵ is arbitrary, we conclude

$$z[s,x] \le \overline{f}(z(x),0) - z^-[x,t] - \beta T + v + a \tag{19}$$

for all $x \in [s, t]$.

Secondly, we have the inequality $z[x, t] \le \overline{f}(z(x), 0) - v$ for all $x \in [s, t]$.

To lighten notation, we drop the subscripts of $\overline{z}_{v+a,v}$. Also, we write $\overline{D}(v + a, v)$, the domain of $\overline{z}$, as $[d_0, d_1]$

Since the left hand side of (18) is non-decreasing in l and the right hand side reaches its maximum at $l = \max_{x \in \overline{D}} \overline{z}(x)$, we need only prove the result for l less than this value. If z never rises above the level l in the interval $[s, t]$ then we are done. If z does rise above l, then define

$$\begin{aligned} & q &:=& \inf\{x \in [s,t] : z(x) \ge l\} \\ \text{and} \qquad & r &:=& \sup\{x \in [s,t] : z(x) \ge l\}. \end{aligned}$$

We also define m and n as in Equations (16) and (17).

Suppose $m \le \overline{x}(u, v)$. Then $\overline{z}[d_0, m] = l^2/2\rho \le z[s, q] \le z^+[s, q]$, where the first inequality uses the fact that z is Lipschitz. On the other hand, if $m > \overline{x}(u, v)$, then $z^+[q, t] + z^-[s, t] \le z[q, t] \le \overline{f}(l, 0) - v = \overline{z}[m, d_1]$. But $z[s, t] = \overline{z}[d_0, d_1] = a$. We thus have, again, that $\overline{z}[d_0, m] \le z^+[s, q]$.

Similarly, using Equation (19), we can show that $\overline{z}[n, d_1] \le z^+[r, t]$.

Now

$$\begin{aligned} & \int_s^t z^+(\tau)\, \mathrm{I}_{z^+(\tau) \le l}\, d\tau &\ge& z^+[s,q] + z^+[r,t], \\ \text{and} \qquad & \int_{d_0}^{d_1} \overline{z}(\tau)\, \mathrm{I}_{\overline{z}(\tau) \le l}\, d\tau &=& \overline{z}[d_0,m] + \overline{z}[n,d_1]. \end{aligned}$$

The conclusion follows. □

We use the following notation:

$$\overline{Q}(u,v) := \int_{D(u,v)} (e^{\vartheta \overline{z}_{u,v}(x)} - \lambda)\, dx.$$

We also define $\underline{Q}(u, v)$ in an analogous manner.

Corollary 1 *Let $z \in Z_p$ for some $p > 0$. Suppose that $z(s) = z(t) = 0$ for some $s, t \in [0, \infty)$, $s \leq t$, and that z is non-negative on (s, t). With a and v defined as above, we have that $\int_s^t (e^{\vartheta z(\tau)} - \lambda)\, d\tau \leq \overline{Q}(v + a, v)$.*

Proof. This is a simple application of Lemma 3. □

The following lemma and corollary are proved in an identical manner.

Lemma 6 *Suppose $z \in Z_p$ is such that $z(s) = z(t) = 0$ for some $s, t \in [0, \infty)$, $s \leq t$, and that $z[s, t] \leq 0$. Then, letting $a := -z[s, t]$ and $d := \inf_{\tau \in [0,t]} (\underline{f}(z(\tau), 0) + z[\tau, t])$, we have that*

$$\int_s^t z^-(\tau)\, \mathrm{I}_{z^-(\tau) \geq -l}\, d\tau \leq \int_{\underline{D}(d+a,d)} \underline{z}_{d+a,d}(\tau)\, \mathrm{I}_{\underline{z}_{d+a,d}(\tau) \geq -l}\, d\tau,$$

for any $l > 0$.

Corollary 2 *Let $z \in Z_p$ for some $p > 0$. Suppose that $z(s) = z(t) = 0$ for some $s, t \in [0, \infty)$, $s \leq t$, and that z is non-positive on (s, t). With a and d defined as in Lemma 6, we have that $\int_s^t (e^{\vartheta z(\tau)} - \lambda)\, d\tau \leq \underline{Q}(d + a, d)$.*

7 Markov Decision Procedures

Recall that a deterministic Markov Decision Procedure (MDP) is a set of states along with a reward function on the transitions between states. We write the transition from a state x to a state y as $x \to y$. A cycle of length n through an MDP is a sequence of n states along with the transitions between consecutive states, including the transition from the last back to the first. If c is a cycle, we write $c = c_1 \to \cdots \to c_n \to c_1$. It is often useful to adopt the convention that $c_0 := c_n$.

For each $\alpha \in [0, \beta T]$ we define the Markov Decision Procedure $\mathcal{M}_\alpha$ as follows:

- *State Space.* $S_\alpha := [0, \alpha]$,
- *Reward.* The transition $u \to v$ has reward

$$W_\alpha(u \to v) := \begin{cases} \overline{Q}(u, v), & \text{if } u \geq v, \\ \underline{Q}(\alpha - u, \alpha - v), & \text{if } u < v. \end{cases}$$

The motivation for these definitions comes from the following lemma.

Lemma 7 *Let $z \in Z_p$ for some $p > 0$. Then there exists $\alpha \in [0, \beta T]$ and a cycle $c_1 \to \cdots \to c_n \to c_1$ through $\mathcal{M}_\alpha$, of some length $N \geq 2$, such that*

$$\int_0^p (e^{\vartheta z(x)} - \lambda)\, dx \leq \sum_{n=1}^N W_\alpha(c_{n-1}, c_n).$$

Proof. Our first step will be to decompose a single period of z into a finite number N intervals $\{[t_{n-1}, t_n)\}_{1\le n\le N}$ in such a way that, for each $n \in \{1,\ldots,N\}$, both $z(t_n) = 0$ and one of the following hold:

(A) z is non-negative on $[t_{n-1}, t_n)$,

(B) z is non-positive on $[t_{n-1}, t_n)$,

(C) $\underline{h} \le z \le \overline{h}$ on $[t_{n-1}, t_n)$, and moreover $z^+[t_{n-1}, t_n) \le \beta T$ and $z^-[t_{n-1}, t_n) \ge -\beta T$.

The following procedure shows how this may be accomplished. We may assume without loss of generality that $z(p) = z(2p) = 0$ and may choose $t_0 := p$ and $t_N := 2p$. First, remove all sub-intervals of $[p, 2p)$ that are of type either (A) or (B) and have length greater than

$$L := \min\left(\frac{2\overline{h}}{\rho}, -\frac{2\underline{h}}{\rho}, \frac{\beta T}{\overline{h}}, -\frac{\beta T}{\underline{h}}\right).$$

There will only be a finite number of these. Each of the remaining intervals may be partitioned into a finite number of subintervals of length less than L in such a way that z takes the value 0 at the boundary of each. Because of the length restriction, each of these intervals is of type (C). Thus the required decomposition is accomplished.

For $0 \le n \le N$, let

$$\begin{aligned} u_n &:= \inf_{t\in[0,t_n]} \left(\overline{f}(z(t), 0) - z[t, t_n]\right) \\ \text{and} \qquad d_n &:= \inf_{t\in[0,t_n]} \left(\underline{f}(z(t), 0) + z[t, t_n]\right). \end{aligned}$$

Now, $u_n + d_n$ is non-increasing with n, and $u_N + d_N = u_0 + d_0$. Therefore $u_n + d_n$ is a constant, which we take as our value of α. Clearly, $u_{n-1} - u_n = d_n - d_{n-1} = z[t_{n-1}, t_n)$ for each $n \in \{1,\ldots,N\}$. We will show that the cycle $u_0 \to u_1 \to \cdots \to u_N$ has the property required in the statement of the lemma.

Suppose z obeys condition (A) in the interval $[t_{n-1}, t_n)$. Then $a := z[t_{n-1}, t_n)$ is non-negative. Applying Corollary 1, we have that

$$\begin{aligned} \int_{t_{n-1}}^{t_n} (e^{\vartheta z(x)} - \lambda)\, dx &\le \overline{Q}(u_n + a, u_n) \\ &= W_\alpha(u_{n-1} \to u_n). \end{aligned}$$

A similar result holds if z obeys condition (B) in the interval.

Now suppose it is Condition (C) that holds in the interval. Without loss of generality, assume that $a := z[t_{n-1}, t_n)$ is non-negative. Define $D_1 := \{t \in [t_{n-1}, t_n) : z(t) < 0\}$ and

$$\eta := \inf\left\{l > 0 : \int_{t_{n-1}}^{t_n} z(t)\, \mathrm{I}_{0\le z(t)\le l}\, dt \ge a\right\}.$$

Let $m := \int_{t_{n-1}}^{t_n} z(t)\, \mathrm{I}_{0\le z(t)<\eta}\, dt$. Then the subset of $[t_{n-1}, t_n)$ on which z takes the value η has Lebesgue measure no less than $(a - m)/\eta$. Choose a subset of this set that has measure $(a-m)/\eta$ and take its union with $\{t \in [t_{n-1}, t_n] : 0 \leq z(t) < \eta\}$ to form a set D_2. Let $D_3 := [t_{n-1}, t_n) - D_2 - D_1$.

We thus arrive at a decomposition of $[t_{n-1}, t_n)$ into sets D_1, D_2, and D_3 in which z takes values in the ranges $[\underline{h}, 0)$, $[0, \eta]$, and $[\eta, \overline{h}]$ respectively. Furthermore $z[D_2] = a$. Let $c := -z[D_1] = z[D_3]$.

We compare z on D_3 with the function $\tilde{z} : [0, c/\overline{h}] \to \{\overline{h}\}$. The conditions of Lemma 3 are satisfied and we deduce that

$$\int_{D_3} (e^{\vartheta z(t)} - \lambda)\, dt \leq \int_0^{c/\overline{h}} (e^{\vartheta \tilde{z}(t)} - \lambda)\, dt = \frac{(e^{\vartheta \overline{h}} - \lambda)c}{\overline{h}}.$$

Similarly,

$$\int_{D_1} (e^{\vartheta z(t)} - \lambda)\, dt \leq -\frac{(e^{\vartheta \underline{h}} - \lambda)c}{\underline{h}}.$$

Therefore the sum of these two terms is non-positive, using the expression for λ given in Equation (15).

For $l \in (0, \eta)$, we apply Lemma 5 and obtain

$$\int_{D_2} z(t)\, \mathrm{I}_{z(t) \leq l}\, dt \geq \int_{\overline{D}(u_n + a, u_n)} \overline{z}_{u_n+a,u_n}(t)\, \mathrm{I}_{\overline{z}_{u_n+a,u_n}(t) \leq l}\, dt.$$

For $l \geq \eta$, the same inequality still holds since, in this case, the left hand side is equal to a which cannot be less than the right hand side.

Since z is non-negative on D_2, all the conditions of Lemma 3 are satisfied and we deduce that

$$\int_{D_2} (e^{\vartheta z(t)} - \lambda)\, dt \leq \overline{Q}(u_n + a, u_n).$$

Now, $[t_{n-1}, t_n)$ is the disjoint union of D_1, D_2, and D_3, and so

$$\begin{aligned} \int_{t_{n-1}}^{t_n} (e^{\vartheta z(t)} - \lambda)\, dt &\leq \overline{Q}(u_n + a, u_n) \\ &= W_\alpha(u_{n-1} \to u_n). \end{aligned}$$

We have proved that this inequality holds regardless of whether interval $[t_{n-1}, t_n)$ is of type (A), (B), or (C). Adding the inequalities for each $n \in \{1, \ldots, N\}$, we obtain the result. □

To proceed, we need to bound the total reward over any cycle through the MDP $\mathcal{M}_\alpha$ for each $\alpha \in [0, \beta T]$. This is accomplished by the following lemma.

Lemma 8 *For each $\alpha \in [0, \beta T]$ and cycle $c_1 \to \cdots \to c_n \to c_1$ through $\mathcal{M}_\alpha$,*

$$\sum_{i=1}^{n} W_\alpha(c_{i-1} \to c_i) \leq 0.$$

Proof. Our first objective will be to show that the average reward per transition over any cycle will be dominated by that over a cycle of length two.

For $v < u$, we have that

$$g_u(v) := \frac{\partial W_\alpha(u,v)}{\partial v} = -\frac{e^{\vartheta \overline{k}(u,v)} - \lambda}{\overline{k}(u,v)}. \tag{20}$$

Since $\overline{k}(u,v)$ is non-decreasing in u for fixed v, so too is $-g_u(v)$.

Suppose that w, x, y, z are states such that $w > x \geq y > z$. Then

$$\begin{aligned} W_\alpha(w,y) &= W_\alpha(w,z) + \int_z^y g_w(t)\,dt \\ \text{and} \qquad W_\alpha(x,y) &= W_\alpha(x,z) + \int_z^y g_x(t)\,dt. \end{aligned}$$

Using the monotonicity of $g_u(v)$ in u, we have that $W_\alpha(w,z) + W_\alpha(x,y) \geq W_\alpha(w,y) + W_\alpha(x,z)$.

Now let c be a cycle in $\mathcal{M}_\alpha$. Without loss of generality, we may assume that the transitions are alternately upwards and downwards and that the first is downwards. This means we may write $c = u_1 \to v_1 \to u_2 \to v_2 \to \cdots \to v_n \to u_1$, with $u_i > v_i$ and $v_{i-1} < u_i$ for all $1 \leq i \leq n$. Here we are using the convention that $u_0 := u_n$. Consider the set $J_0 := \{u_i \to v_i : 1 \leq i \leq n\}$ of downward transitions in c. We perform the following operation on J_0. If J_0 contains one or more pairs of transitions $w \to y$ and $x \to z$ such that $w > x \geq y > z$, then we choose one such pair and replace it with the pair $w \to z$ and $x \to y$. If there are no such pairs then we do nothing. Call the new set of transitions J_1. We iterate this procedure to form sets $J_2, J_3, \ldots$ and so on. The procedure terminates when no more changes can be made. For any set of transitions $K \subset S_\alpha \times S_\alpha$, define

$$\mathcal{W}(K) := \sum_{(w\to y)\in K} W_\alpha(w \to y).$$

From the considerations above, $\mathcal{W}(J_n) \leq \mathcal{W}(J_{n+1})$. Since at each stage of the procedure the quantity

$$\sum_{(u\to v)\in J_n} (v-u)^2$$

must strictly increase, we must reach a stable set J of transitions after a finite number of steps. This set will not contain any pair of transitions $w \to y$ and $x \to z$ such that $w > x \geq y > z$. But there is a unique pairing of the starting states $\{u_i\}_{1\leq i\leq n}$ and the destination states $\{v_i\}_{1\leq i\leq n}$ that has this property. This shows that the end result of the algorithm described above does not depend on the order in which transitions are replaced.

If we now consider the set of upward transitions $I_0 := \{v_i \to u_i : 1 \leq i \leq n\}$ we may employ a similar procedure to arrive at a set I for which $\mathcal{W}(I_0) \leq \mathcal{W}(I)$. But again this set is independent of the order of substitution and in fact $I = \{v \to u : u \to v \in J\}$. In other words I is the same as J but with the direction of the transitions reversed. So,

$$\begin{aligned} \frac{1}{2n}(\mathcal{W}(J) + \mathcal{W}(I)) &= \frac{1}{2n} \sum_{(u \to v) \in J} [W_\alpha(u \to v) + W_\alpha(v \to u)] \\ &\leq \frac{1}{2} \max_{(u \to v) \in J} [W_\alpha(u \to v) + W_\alpha(v \to u)]. \end{aligned}$$

This shows that every cycle has a lower mean reward than some cycle of length two.

Note that the reward of each transition $\hat{u} \to \hat{v}$ in $\mathcal{M}_\alpha$ is the same as that of some transition $u \to v$ in $\mathcal{M}_{\beta T}$, with $u - v = \hat{u} - \hat{v}$. We will now maximize $W_{\beta T}(u \to v)$, keeping $a := u - v$ constant. Suppose for the moment that a is positive. The partial derivative of $W_{\beta T}(u \to v)$ with respect to v was given in Equation (20). The partial derivative with respect to u is

$$\frac{\partial W_{\beta T}(u \to v)}{\partial u} = \frac{e^{\vartheta \overline{j}(u,v)} - \lambda}{\overline{j}(u,v)}.$$

It is clear that the maximum will only occur when $\overline{j}$ and $\overline{k}$ are equal, in which case $\overline{z}_{u,v}$ is a shifted copy of $\overline{z}_{u_a,v_a}$, where $u_a := (\beta T + a)/2$ and $v_a := (\beta T - a)/2$. We reach a similar conclusion when we maximize $W_{\beta T}(u \to v)$, again keeping $u - v$ constant, but this time negative. We conclude that

$$W_\alpha(u \to v) + W_\alpha(v \to u) \leq \overline{R}(u_a, v_a) + \underline{R}(u_a, v_a) - \lambda\Big(\mathrm{Leb}\overline{D}(u,v) + \mathrm{Leb}\underline{D}(u,v)\Big),$$

for all $\alpha \in [0, \beta T]$ and $u, v \in [0, \alpha]$. But the definition of λ in Equation (14) implies that the expression on the right hand side is no greater than zero. □

8 The Main Result

We now have the tools necessary to prove the main result.

Proof of Theorem 1. Let λ be defined as in Equation (14). Then, by Lemma 4,

$$\sup_{\gamma \in [0, \beta T]} \mathcal{E}_{\vartheta, T}(w_\gamma) = \sigma + \frac{\log \lambda}{\vartheta T}.$$

We must show that the quantity on the right hand side is an upper bound on the effective bandwidth of every source in $\mathcal{C}$. In light of Lemma 1, it will be sufficient to show that, for each realization $b \in C_p$, where $p > 0$,

$$\frac{1}{p} \int_0^p e^{\vartheta b[x, x+T]}\, dx \leq \lambda e^{\sigma \vartheta T}.$$

To each such b there is an associated $z \in Z_p$, defined according to Equation (8). By Lemma 7, there exists an $\alpha \in [0, \beta T]$ and a cycle c through the Markov Decision Procedure $\mathcal{M}_\alpha$ such that

$$\int_0^p (e^{\vartheta z(x)} - \lambda)\, dx \leq \sum_{i=1}^{n} W_\alpha(c_{i-1} \to c_i),$$

where n is the length of c. But every cycle of $\mathcal{M}_\alpha$ has non-positive reward by Lemma 8. Thus

$$\int_0^p e^{\vartheta z(x)}\, dx \leq \lambda p,$$

from which the required inequality follows.

□

References

[1] ATM Forum. Traffic management specification version 4.0. Technical Report 95-0013R10, ATM Forum, February 1996.

[2] R. Boorstyn, A. Burchard, J. Liebeherr, and C. Oottamakorn. Statistical service assurances for traffic scheduling algorithms. *IEEE Journal for Selected Areas in Communications*, 18(12):2651–2664, 2000.

[3] D. D. Botvich and N. G. Duffield. Large deviations, the shape of the loss curve, and economies of scale in large multiplexers. *Queueing Systems*, 20:293–320, 1995.

[4] C. Courcoubetis, F. Kelly, and R. Weber. Measurement-based usage charges in communication networks. *Operations Research*, 48:535–548, 2000.

[5] Costas Courcoubetis and Richard Weber. Buffer overflow asymptotics for a buffer handling many traffic sources. *Journal of Applied Probability*, 33:886–903, 1996.

[6] Rene L. Cruz. A calculus for network delay, part I: Network elements in isolation. *IEEE Transactions on Information Theory*, 37(1):114–131, January 1991.

[7] B. T. Doshi. Deterministic rule based traffic descriptors for broadband ISDN: Worst case behavior and connection acceptance control. In *Proceedings of the 14th ITC*, pages 591–600, Antibes Juan-les-Pins, June 1994.

[8] Frank P. Kelly. Notes on effective bandwidths. In Kelly, Zachary, and Ziedins, editors, *Stochastic Networks: Theory and Applications*, pages 141–168. Oxford University Press, 1996.

[9] G. Kesidis and T. Konstantopoulos. Extremal shape-controlled traffic patterns in high-speed networks. *IEEE Trans. Comm.*, 48(5):813–819, May 2000.

[10] Daniel Chonghwan Lee. Effects of leaky bucket parameters on the average queueing delay: Worst case analysis. In *Proceedings of IEEE Infocom'94*, pages 482–489, Toronto, Canada, June 1994.

[11] F. LoPresti, Z. Zhang, D. Towsley, and J. Kurose. Source time scale and optimal buffer/bandwidth trade-off for regulated traffic in an ATM node. In *Proceedings of IEEE Infocom '97*, pages 347–355, Kobe, Japan, April 1997.

[12] B. McGurk and C. Walsh. Investigations of the performance of a measurement-based connection admission control algorithm. In *Proceedings of ATM '97 - Fifth IFIP Workshop on Performance Modelling and Evaluation of ATM Networks*, Ilkely, UK, July 1997.

[13] Debasis Mitra and John A. Morrison. Multiple time scale regulation and worst case processes for ATM network control. In *Proceedings of the 34th Conference on Decision & Control*, pages 353–358, New Orleans, LA, December 1995.

[14] Philippe Oechslin. In search of the worst case arrivals of independent leaky bucket constrained sources. In *Proceedings UKPEW*, pages 100–106, Bradford, July 1997.

[15] M. Reisslein, K. W. Ross, and S. Rajagopal. Packet multiplexers with adversarial regulated traffic. In *Proceedings of IEEE Infocom '98*, pages 347–355, San Francisco, April 1998.

[16] Jonathan S. Turner. New directions in communications (or which way to the information age?). *IEEE Communications Magazine*, 24(10):8–15, October 1986.

[17] Cormac Walsh. Worst case traffic from policed traffic sources. Technical Report HPL-BRIMS-01-01, HP Labs, 2001. www.stp.dias.ie/~walsh/effband.ps.