Worst Case Traffic from Policed Traffic Sources

Cormac Walsh

Abstract

We address the problem of maximising the effective bandwidth of a traffic source over the space of all stationary traffic sources which have realisations meeting certain constraints. Specifically, we ask that the peak rate of the source does not exceed ρ and that realisations obey a "leaky bucket" constraint with bucket size β and leak rate σ.

We present two results as a step towards a solution. Firstly, we reduce the optimisation over all stationary processes to an optimisation over periodic processes, that is sources which emit traffic in a periodic manner with uniformly distributed phase. Secondly, we show that the realisations of a worst case source must obey the following conditions: at each time, the transmission rate must be one of 0, σ, or ρ; the transmission rate may only be σ when the leaky bucket is either empty or full; each burst of activity must either start with the leaky bucket empty or end with it full.

1 Introduction

Let $\mathbb{D}^+$ be the set of non-negative, σ-finite measures on $[0, \infty)$. We interpret the elements of this set as possible realisations of a traffic source. In other words, if $a \in \mathbb{D}^+$, then we take $a(\mathcal{B})$ to be the amount of traffic arriving in the Borel set of instants $\mathcal{B}$. Suppose $\rho > \sigma > 0$ and $\beta > 0$ are given. We denote by C those elements a of $\mathbb{D}^+$ for which

$$a(t_1, t_2) \leq (t_2 - t_1)\rho \tag{1}$$

$$\text{and} \qquad a(t_1, t_2) \leq (t_2 - t_1)\sigma + \beta, \tag{2}$$

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. These constraints may be reformulated as follows. For $a \in \mathbb{D}^+$, $q > 0$, $s > 0$, and $t \geq 0$, we define

$$Q_s(q, a, t) := \max\Big(q + a[0, t] - ts, \sup_{t_1 \in [0,t]} \big(a[t_1, t] - (t - t_1)s\big)\Big).$$

This is the queue length at time t in a buffer if a is the arrivals, s is the service rate, and q is the initial queue length. It is easy to see that (1) is equivalent to $Q_\rho(0, a, t) = 0$ for all $t \geq 0$ and (2) is equivalent to $Q_\sigma(0, a, t) \leq \beta$ for all $t \geq 0$. The fact that these constraints may be expressed in this form is the reason for the term "leaky bucket". Such constraints are very popular in the engineering literature and are part of the standard for Asynchronous Transfer Mode technology [1].

We identify traffic sources with probability measures on $\mathbb{D}^+$. Realisations are shifted forward in time with the operator $\Theta_h : \mathbb{D}^+ \to \mathbb{D}^+$ defined, for each $h \geq 0$, by $(\Theta_h a)[\mathcal{B}] := a[\mathcal{B} + h]$ for all Borel sets $\mathcal{B} \subset [0, \infty)$. 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 $\mu[\Theta_h^{-1} B] = \mu[B]$ for all measurable subsets 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

$$\mathcal{E}_{\vartheta,T}(\mu) := \frac{1}{\vartheta T} \log \mathbb{E}_\mu e^{\vartheta X_T},$$

where $X_T : \mathbb{D}^+ \to \mathbb{R}^+ : a \mapsto a[0, T]$ evaluates the amount of traffic up to time T. The two parameters $\vartheta \in \mathbb{R}$ and $T \in [0, \infty)$ are called the *space scale* and the *time scale* respectively. See [2] for discussion of their interpretation. The effective bandwidth is exactly what is needed to calculate the asymptotics of the loss rate as the number of sources becomes infinite. 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, each distributed with the same law μ. As was shown in [3], $\mathbb{L}$ obeys the large deviations asymptotics

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{L}(sN, bN, N) = -\inf_{T>0} (f_T^{(\mu)})^*(b + sT).$$

where $f_T^{(\mu)}(\vartheta) := \vartheta T \mathcal{E}_{\vartheta,T}(\mu)$ and the *Legendre-Fenchel* transform of any convex function f is defined to be $f^*(y) := \sup_{x\in\mathbb{R}}[xy - f(x)]$.

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 Nb in an infinite buffer [4]. For an overview of the many applications of the effective bandwidth function see the paper of Kelly [2].

The asymptotics above motivate the optimisation program:

$$\text{for fixed } \vartheta \text{ and } T, \qquad \text{maximise } \mathcal{E}_{\vartheta,T}(\mu), \qquad \text{subject to } \mu \in \mathcal{C}.$$

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 (see Figure 1):

- 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. We make the following conjecture.

Conjecture 1 *For each ϑ and T, the stationary traffic source that maximises the effective bandwidth $\mathcal{E}_{\vartheta,T}$, under the peak rate and leaky bucket constraints 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 optimisation over γ.

This conjecture was proved in [5] for the case when T is less than both t_0 and t_ρ. In that paper the methods used were specific to the effective bandwidth functional. However, there are other functionals on the set of sources which it may be of interest to maximise. Examples include the loss rate when a fixed number of independent copies of the source pass trough a buffer, or the average queue length when the source passes alone through a buffer which is being served according to an independent stationary stochastic service process. These functionals share many of the properties of the effective bandwidth functional that are exploited in the present paper and so there is much more scope for extending these methods than those of [5].

Many of these functionals, including the effective bandwidth, have been considered previously [6, 7, 2]. Results have been obtained for the simplest functionals such as the unbuffered loss rate [8, 9] and the average queue length of a single source [10]. A common approach with more difficult functionals is to find a tractable upper bound on the quantity of interest and to minimise this [11, 12].

The rest of the paper is organised as follows. In Section 2, we reduce the optimisation problem to an optimisation over processes with periodic realisations. In Section 3, we define a linear structure on the set of periodic realisations with fixed period and show that the worst case source must have realisations which are extreme points of C. In Section 4, we characterise these points. In Section 5, we define an alternative linear structure on the same set and show that the realisations of a worst case source must be extreme points in this linear structure also. Finally, in Section 6, we characterise these points.

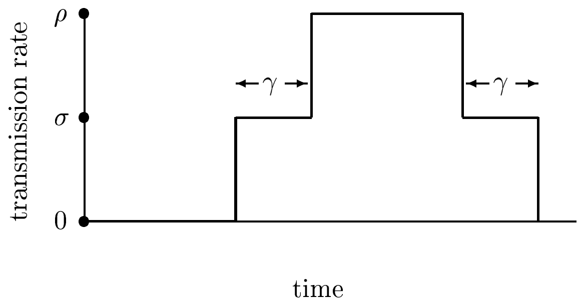

Figure 1: The wedding cake pattern, conjectured to have the worst case effective bandwidth.

2 Periodicity

In this section we will reformulate the optimisation problem as a deterministic Markov Decision Procedure (MDP). Continuous time MDPs generally have a state space which is a differentiable manifold. The paths take the form of absolutely continuous trajectories through this space and the instantaneous reward is given by a function L, called the Lagrangian, on the tangent bundle. Thus the reward of a path x over a time interval $[0, h]$ is

$$\int_0^h L[x(t), \dot{x}(t)]\, dt.$$

In our MDP the Lagrangian will be purely a function of the state: it will not depend on its second parameter. This will allow us to dispense with an investigation of the differential structure of our space. Also we do not require that paths be absolutely continuous, merely that they be measurable.

We define $\mathbb{D}_p^+$ to be the set of non-negative, finite measures on $[0, p]$. We wish to consider sections of realisations which meet the constraints. However, the content of the leaky bucket at the start of the section plays an important role in determining whether the constraints are met. With this in mind we define

$$C_p := \{(q, a) \in [0, \beta] \times \mathbb{D}_p^+ : Q_\sigma(q, a, t) \leq \beta \text{ and } Q_\rho(0, a, t) = 0, \text{ for all } t \in [0, p]\}.$$

The MDP we wish to consider has state space $S := C_T$. Consider the following construction. Let $(q, a) \in C_{h+T}$ for some $h > 0$. Define the path $x(t) := (q_t, x_t)$, where $q_t := Q(q, a, t) \leq \beta$ for $t \in [0, h]$ and x_t is the measure defined on $[0, T]$ by $x_t[\mathcal{B}] := a[\mathcal{B} + t]$ for all $t \in [0, h]$ and Borel subsets $\mathcal{B}$ of $[0, T]$. It is not hard to verify that x is indeed a path through S. We call the paths formed in this way *valid* paths of length h and denote them V_h. Note that the correspondence between V_h and C_{h+T} is a bijection. It is important to note also that periodic paths correspond to periodic realisations with the same period.

We take the Lagrangian of our MDP to be $L : S \to \mathbb{R}^+ : (q, a) \mapsto \exp(a[0, T])$ and define the reward of a path $x : [0, h] \to S$ to be

$$\mathcal{R}_h(x) := \begin{cases} \int_0^h L[x(t)]\, dt, & x \in V_h \\ -\infty, & \text{otherwise.} \end{cases}$$

Note that if x is constructed from initial data q and a as above then

$$\mathcal{R}_h(x) = \int_0^h e^{a[t, t+T]}\, dt.$$

We denote the space of continuous real-valued functions on S by $C(S)$. For each $h \geq 0$, we define an operator $R^h : C(S) \to C(S)$ as follows:

$$(R^h f)(u) := \sup_{x \in V_h} \Big\{ \mathcal{R}_h(x) + f(x(h)) : x(0) = u \Big\}.$$

This operator has kernel

$$r^h(u,v) = \sup_{x\in V_h} \Big\{ \mathcal{R}_h(x) : x(0)=u,\, x(h)=v \Big\}. \tag{3}$$

By this we mean that, for $h \geq 0$,

$$(R^h f)(u) = \sup_{v\in S} \left(r^h(u,v) + f(v) \right).$$

Any operator that can be represented in this form is called a Bellman operator. Observe that the operator R^0 is the identity operator on $C(S)$ and has kernel

$$r^0(u,v) = \begin{cases} 0, & u = v \\ -\infty, & u \neq v. \end{cases}$$

The operators $\{R^h\}_{h\geq 0}$ together form a semigroup, that is for any $s,t \geq 0$ we have $R^{s+t} = R^s R^t = R^t R^s$. The identity element of this semigroup is R^0.

A function $f \in C(S)$ is said to be a $(\max,+)$-eigenfunction of a Bellman operator B with eigenvalue $\lambda \in \mathbb{R}$ if $Bf = f + \lambda$. A Bellman operator on $C(S)$ is said to be compact if S is a compact metric space and the operator's kernel is jointly continuous in its two entries and does not take the value $-\infty$. The justification for calling these operators compact is that they are compact in the usual sense of the term: they take bounded subsets of $C(S)$ to precompact ones. A well known result in the theory of Markov Decision Procedures is that each compact Bellman operator possesses a unique eigenvalue [13]. The eigenfunction is not necessarily unique however.

The topology we use on $\mathbb{D}_p^+$ is the one which in Probability Theory is known as the topology of weak convergence. This is the weak* topology induced on $\mathbb{D}_p^+$ by regarding it as the dual of $C[0,p]$, the space of continuous functions on $[0,p]$ endowed with the uniform topology. The following facts concerning this topology will be useful. Firstly, Helly's theorem states that a set of signed measures on a compact set is precompact if and only if the total variation of the measures is bounded. Secondly a sequence $\{a_n\}_{n\in\mathbb{N}}$ in $\mathbb{D}_p^+$ converges to $a \in \mathbb{D}_p^+$ if and only if $a_n[\mathcal{B}] \to a[\mathcal{B}]$ for each Borel set $\mathcal{B}$ for which $a[\partial\mathcal{B}] = 0$. Thirdly, this topology is metrisable and complete.

Since V_h can be identified with a subset of $[0,\beta] \times \mathbb{D}_{h+T}^+$, we choose to take on V_h the topology inherited from this set.

Lemma 1 *The reward $\mathcal{R}_h$ is continuous on the set V_h for each $h > 0$.*

Proof. Suppose that $\{x_n\}_{n\in\mathbb{N}}$ is a sequence of paths in V_h that converges to some $x \in V_h$. For each $n \in \mathbb{N}$, let $a_n \in C_{h+T}$, be the realisation associated with x_n, and let $a \in C_{h+T}$ be that associated with x. Consider the sequence of functions

$$f_n : [0,h] \mapsto \mathbb{R} : t \to e^{\vartheta a_n[t,t+T]}.$$

Then $f_n(t)$ converges to $f(t) = \exp(\vartheta a[t,t+T])$ pointwise. Each f_n is bounded above by $\exp(\vartheta a_n[0,h+T])$ which converges to $\exp(\vartheta a[0,h+T])$. This means that for n large enough the functions f_n are uniformly bounded and we can apply the Bounded Convergence Theorem to conclude that

$$\int_0^h f_n(t)\,dt \longrightarrow \int_0^h f(x)\,dx \qquad \text{as } n \to \infty.$$

Thus $\mathcal{R}_h(x_n) \to \mathcal{R}_h(x)$ as $n \to \infty$. □

Lemma 2 *The functional $\mathcal{Q}_s : \mathbb{R}^+ \times \mathbb{D}_h^+ \to \mathbb{R} : (q,a) \mapsto \sup_{t\in[0,h]} Q_s(q,a,t)$ is jointly continuous in q and a for each $s > 0$.*

Proof. Let $a_n \to a$ and $q_n \to q$. Write $v_n := a_n - s\,\mathrm{Leb} + q_n\delta_0$ and $v := a - s\,\mathrm{Leb} + q\delta_0$. Let $\mathcal{G}$ be the set of intervals which are open in $[0,h]$. Firstly,

$$\begin{aligned} \liminf_{n\to\infty} \mathcal{Q}_s(q_n,a_n) &= \liminf_{n\to\infty} \sup_{G\in\mathcal{G}} v_n[G] \\ &\geq \sup_{G\in\mathcal{G}} \liminf_{n\to\infty} v_n[G] \\ &\geq \sup_{G\in\mathcal{G}} v[G] \\ &= \mathcal{Q}_s(q,a). \end{aligned}$$

Now let $H_N := \{ih/N : i \in \{0, \dots, N\}\}$. For each $t_1, t_2 \in [0, h]$, we can find $t_1', t_2' \in H_N$ such that $v_n[t_1, t_2] \leq v_n[t_1', t_2'] + 2h\sigma/N$, for all $n \in \mathbb{N}$. Then

$$\mathcal{Q}_s(q_n, a_n) \leq \sup_{t_1', t_2' \in H_N} v_n[t_1', t_2'] + 2h\sigma/N.$$

Thus

$$\limsup_{n \to \infty} \mathcal{Q}_s(q_n, a_n) \leq \sup_{t_1', t_2' \in H_N} v[t_1', t_2'] + 2h\sigma/N.$$

By making N large enough we can make the right hand side of this inequality as close as we want to $\mathcal{Q}_s(q, a)$. □

We conclude that C_h is a closed subset of $[0, \beta] \times \mathbb{D}_h^+$. It is also compact since for any $(q, a) \in C_h$, the total variation of a is bounded above by ρh. In particular both the state space S and the set of valid paths V_h are compact.

Unfortunately, for small values of h, our operator R^h is not compact. However, it will suffice that R^h be compact for some $h > 0$.

Lemma 3 *The operator R^k is compact for $k = T + \beta / \min(\sigma, \rho - \sigma) + 1$.*

Proof. Let $u = (q_u, a_u)$ and $v = (q_v, a_v)$ be two states in S. For any state $x := (q_x, a_x)$, define

$$\overline{q}_x := Q_\sigma(q_x, a_x, T).$$

This may be interpreted as the amount of fluid in the leaky bucket at the end of the time interval represented by the state.

We first show that r^k does not take the value $-\infty$. Let $a \in \mathbb{D}_{k+T}^+$ be defined as follows. Suppose $\mathcal{B}$ is a Borel subset of $[0, k+T]$. Then $\mathcal{B}$ can be written as the disjoint union of Borel sets $\mathcal{B}_u$, $\mathcal{B}_m$, and $\mathcal{B}_v$, where $\mathcal{B}_u \subset [0, T]$, $\mathcal{B}_m \subset (T, k)$, and $\mathcal{B}_v \subset [k, k+T]$. We define

$$a[\mathcal{B}] := a_u[\mathcal{B}_u] + \left(\frac{q_v - \overline{q}_u}{k - T} + \sigma\right) \mathrm{Leb}\,[\mathcal{B}_m] + a_v[\mathcal{B}_v - k],$$

where Leb is the Lebesgue measure. Repeating the construction of valid paths above, we define $q_t := Q_\sigma(q_u, a, t)$ and $x_t(\mathcal{B}) := a(\mathcal{B} + t)$ for all Borel subsets $\mathcal{B}$ of $[0, T]$ and $t \in [0, h]$. Then the path $x(t) := (q_t, x_t)$ is a valid path of length k from u to v, and so $r^k(u, v) > -\infty$.

Upper semicontinuity. Let $l > -\infty$. Suppose that u_i and v_i are two sequences of states, converging to u and v respectively, such that $r^k(u_i, v_i) \geq l$ for all $i \in \mathbb{N}$. Then, given any $\epsilon > 0$, there exists a sequence of valid paths $x_n : [0, k] \to S$ such that $x_n(0) = u_n$, $x_n(k) = v_n$, and $\mathcal{R}_h(x_n) \geq l - \epsilon$ for all $n \in \mathbb{N}$. Since the set V_k of valid paths of length k is compact, some subsequence x_{n_k} converges to a path $x \in V_k$. From the continuity of $\mathcal{R}_h$ on V_k,

$$r(u, v) \geq \mathcal{R}_h(x) = \lim_{k \to \infty} \mathcal{R}_h(x_{n_k}) \geq l - \epsilon.$$

Since ϵ is arbitrary, $r(u, v) \geq l$. Therefore, the upper level set is closed for each level l and so r is upper semicontinuous.

Lower semicontinuity. Let $x : [0, k] \to S$ be a path from u to v such that $\mathcal{R}_h(x) > r^k(u, v) - \epsilon$. Let $(q_u, a) \in C_{k+T}$ be the initial leaky bucket state and realisation associated with this path.

We may assume that $a(T, t) + \overline{q}_u \geq (t - T)\sigma$ for all $t \in [T, k)$. For if this is not the case then define b to be such that $b[T, t] := q_t - \overline{q}_u + (t - T)\sigma$ for $t \in [T, k)$, and such that b agrees with a outside this interval. The following facts are easily verified. Firstly, b obeys constraints (1) and (2). Secondly, it obeys the inequality above. Thirdly, the path y corresponding to (q_u, b) is the same length as x, between the same two states. Finally, $b \geq a$, which implies that $\mathcal{R}_h(y) \geq \mathcal{R}_h(x)$.

We will perform a construction which assigns to any pair of states (u', v') near (u, v) a valid path from u' to v' near x. Let $\Delta := q_{v'} - q_v - \overline{q}_{u'} + \overline{q}_u$ where $(q_u, a_u) := u$ and $(q_v, a_v) := v$. Condition (1) implies that a is absolutely continuous, and so it equals the integral of its own derivative, Da.

For δ in the range $[-1, 1]$, we define the set of maps $z_\delta : [0, \rho] \to [0, \rho]$ by

$$z_\delta(x) := \begin{cases} (1+\delta)x, & \delta < 0 \\ (1-\delta)x + \delta\rho, & \delta \geq 0. \end{cases}$$

Consider the mapping $w : [-1,1] \to \mathbb{D}^+_{k+T}$ defined by

$$\Big(w(\delta)\Big)[\mathcal{B}] := \begin{cases} a_{u'}[\mathcal{B}], & \text{if } \mathcal{B} \subset [0,T] \\ \displaystyle\int_{\mathcal{B}} z_\delta\big(Da(x)\big)\,dx, & \text{if } \mathcal{B} \subset (T,k) \\ a_{v'}[\mathcal{B}-k], & \text{if } \mathcal{B} \subset [k,k+T]. \end{cases}$$

Clearly, for each $\delta \in [-1,1]$, $w(\delta)$ obeys the peak rate constraint. Also, w is a continuous function of δ.

It can not be that $Da(t) = \rho$ at almost all $t \in [T,k]$, for if this were the case then q_v would equal $\overline{q}_u + (k-T)(\rho - \sigma)$ which is strictly greater than β, and a would not obey Constraint (2). Neither can $Da(t) = 0$ at almost all $t \in [T,k]$ for this would contradict the fact that $a(T,k) + \overline{q}_u \geq (k-T)\sigma$. Thus the function

$$\mathcal{Z} : [-1,1] \to \mathbb{R} : \delta \mapsto (w(\delta))[0,k] - a[0,k]$$

is continuous and strictly increasing in δ. Hence we can find δ^* such that $w(\delta^*)[0,k] = q_{v'}$. However, $w(\delta^*)$ may not obey the leaky bucket constraint. Let w' be the measure on $[0,k+T]$ such that

$$w'[0,t] := (\beta + \sigma t - q_{u'}) \wedge w(\delta^*)[0,t] \vee (\sigma t - q_{u'}), \qquad \text{for all } t \in [0,k+T].$$

Then $(q,w') \in C_{k+T}$ and $Q_\sigma(q_{u'}, w', k) = w(\delta^*)[0,k] = q_{v'}$. Thus the pair $(q_{u'}, w')$ corresponds to a valid path x' from u' to v'. It is not hard to see that this path is continuous in the initial data u' and v', and that if $u' = u$ and $v' = v$ then the construction gives $x' = x$. □

To exploit this result we will have to consider discrete paths (sequences of states). Let D be the set of discrete paths and K be the subset of these that are periodic. We define the long term average reward of $y \in D$ to be

$$\mathcal{G}(y) := \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} r^k(y_{i-1}, y_i).$$

From the compactness of R^k we get that

$$\sup_{y\in D} \mathcal{G}(y) = \sup_{y \in K} \mathcal{G}(y) = \Gamma,$$

where Γ is the $(\max,+)$-eigenvalue of R^k.

A realisation $a \in \mathbb{D}^+$ is said to be periodic with period $p > 0$ if $\theta^p a = a$ and p is the least positive number for which this is true. Given any periodic realisation a with period p, we may construct a stationary traffic source μ by taking a uniform distribution over the phase. More precisely,

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

We denote by P_p the set of measures formed in this way from elements of C which are periodic with period p. The following is the main result of this section.

Theorem 1 *We have that* $\sup\{\mathcal{E}_{\vartheta,T}(\mu) : \mu \in \mathcal{C}\} = \sup\{\mathcal{E}_{\vartheta,T}(\mu) : \mu \in P_p,\, p > 0\}$

Proof. Let $\mu \in \mathcal{C}$ be a stationary traffic source that meets the constraints. The continuous time Ergodic Theorem implies that the function

$$\overline{f}(a) := \lim_{t\to\infty} \frac{1}{t} \int_0^t e^{\vartheta a[x,x+T]}\,dx$$

exists μ-almost surely and has the same expectation with respect to μ as $f(a) := \exp(\vartheta a[0,T])$.

Denote by L the set of realisations at which $\overline{f}$ is defined. Since $\mu[C \cap L] = 1$, we have that

$$\mathbb{E}_\mu f = \mathbb{E}_\mu \overline{f} \leq \sup_{a \in C\cap L} \overline{f}(a).$$

Let $a \in C \cap L$. Then $\overline{f}(a) = \lim_{h\to\infty} h^{-1}\mathcal{R}_h(x)$, where x is the corresponding path through the MDP. Recall that $x(t) := (q_t, x_t)$, where $q_t := Q_\sigma(0,a,t)$ and $x_t[\mathcal{B}] := a[\mathcal{B}+t]$ for all $t \in [0,\infty)$ and Borel subsets $\mathcal{B}$

of $[0,T]$. Let $y_i := x(ki)$ be the discrete path obtained from x by taking samples spaced k apart, where k was defined in Lemma 3. We have that $\lim_{h\to\infty} h^{-1}\mathcal{R}_h(x) \le \mathcal{G}(y)/k \le \lambda/k$. Thus $\mathbb{E}_\mu f \le \lambda/k$, which provides an upper bound on the worst case effective bandwidth.

To finish the proof, we will show that we can find a periodic source μ such that $\mathbb{E}_\mu f$ is a close as we like to λ/k. By Lemma 3, R^k is compact, and so it has a unique $(\max,+)$-eigenvalue λ. As discussed above, there are periodic discrete paths with long term average reward per step arbitrarily close to λ. Let $y \in K$ be such a path and let y_i and y_{i+1} be any two consecutive states. Then, since r^k is given by (3), we can find a continuous time path $z \in V_k$ of length k such that $z(0) = y_i$, $z(k) = y_{i+1}$, and $\mathcal{R}_k(z)$ is as close as we want to $r^k(y_i, y_{i+1})$. By pasting together such paths we can construct a continuous time periodic path w whose long term average reward is as close as we want to $\mathcal{G}(y)/k$.

This path will be in V_∞, the set of valid paths of infinite length. Let $(q,a) \in C$ be the corresponding initial leaky bucket state and traffic realisation. Then a is periodic and $\overline{f}(a) = \lim_{h\to\infty} h^{-1}\mathcal{R}_h(w)$.

The final step is to construct a stationary probability measure μ from a in the manner described above. Clearly, $\mathbb{E}_\mu f = \overline{f}(a)$ can be made as close as we want to λ/k. □

3 Optimisation with Fixed Period

In this section we consider the problem of maximising $\mathcal{E}_{\vartheta,T}(\mu)$ over $\mu \in P_p$ for fixed p. Recall that P_p is the set of traffic sources formed by taking a single periodic realisation of period p and choosing the phase uniformly. We first develop some notation that will be helpful in this setting.

The realisations of any source in P_p are shifted copies of some realisation which is periodic with period p and meets the constraints. Therefore we need only concern ourselves with the representative time interval $[0,p]$. The reader may consider it more natural to identify points 0 and p, in other words to consider the set of measures on the set $[0,p)$ endowed with the topology of the circle. However the advantages of the current approach will become clear in Section 5 when we relate the current representation of realisations to the one there.

It will be convenient to define, for $t_1, t_2 \in [0,p]$,

$$[t_1,t_2]_p := \begin{cases} [t_1,t_2], & \text{if } t_1 \le t_2 \\ [t_1,p] \cup [0,t_2], & \text{otherwise.} \end{cases}$$

If either t_1 or t_2 is outside the range $[0,p]$, then $[t_1,t_2]_p$ is interpreted to mean $[t_1 (\mathrm{mod}\ p), t_2 (\mathrm{mod}\ p)]_p$. We define $(t_1,t_2]_p := [t_1,t_2]_p - \{t_1\}$. The sets $(t_1,t_2)_p$ and $[t_1,t_2)_p$ are defined similarly. Another useful piece of notation will be $t_2 \ominus t_1 := \mathrm{Leb}[t_1,t_2]_p$.

The set of periodic sources of period p may be identified with the set $\mathbb{D}_p^+$ modulo the equivalence relation $\sim$, where $a \sim b$ if there exists $h \in \mathbb{R}$ such that $a(t_1,t_2)_p = b(t_1+h, t_2+h)_p$ for all $t_1,t_2 \in [0,p]$. Motivated by this identification, we define the queue length at time $t \in [0,p]$ to be

$$W_s(a,t) := \sup_{t_1 \in [0,p]} \Big(a[t_1,t]_p - (t \ominus t_1)s \Big). \tag{4}$$

Constraints (1) and (2) translate into the condition that an element of $\mathbb{D}_p^+$ be in the set

$$B_p := \{a \in \mathbb{D}_p^+ : W_\sigma(a,t) \le \beta \text{ and } W_\rho(a,t) = 0, \text{ for all } t \in [0,p]\}.$$

If μ is the periodic source identified with the equivalence class of $a \in \mathbb{D}_p^+$, then the effective bandwidth of μ is $(\vartheta T)^{-1} \log E(a)$, where

$$E(a) := \frac{1}{p} \int_0^p e^{\vartheta a[t,T+t]_p}\, dt. \tag{5}$$

Note that if $a \sim b$ then $E(a) = E(b)$. The problem of maximising the effective bandwidth over P_p is the same as maximising $E(a)$ over $a \in B_p$.

To discuss convexity, we need a linear space. We take the linear completion $\mathbb{D}_p := \mathbb{D}_p^+ - \mathbb{D}_p^+ = \{a - b : a,b \in \mathbb{D}_p^+\}$. The elements of $\mathbb{D}_p$ are the finite signed measures on $[0,p]$. For each $s \ge 0$, the functional

$\sup_{t\in[0,p]} W_s(a,t)$ is convex on $\mathbb{D}_p^+$ since it can be written as the supremum of a family of affine functions. Using arguments similar to those in Lemma 2, it can be shown that these functionals are also continuous on $\mathbb{D}_p^+$. It follows that the set B_p is closed and convex since it is the intersection of a level set of two of these functionals. Again, compactness follows from the peak rate constraint.

By the Krein-Milman theorem, B_p is the closed convex hull of ξ_p, the set of its extreme points. Recall that a point p is said to be an extreme point of a set S if $p \in S$ and p is not an interior point of any line segment in S.

As the mean of a family of convex functionals, the functional E is also convex. Its continuity may be established using means similar to those in the proof of Lemma 1.

Theorem 2 $\sup\{\mathcal{E}_{\vartheta,T}(\mu) : \mu \in C\} = \sup\{(\vartheta T)^{-1} \log E(a) : a \in \xi_p,\, p > 0\}$

Proof. Since E is convex on B_p, its supremum over ξ_p the same as its supremum over the convex hull of ξ_p. Since E is continuous, this second supremum is the same as the supremum of E over the closure of this set. But we have seen above that the closure of the convex hull of ξ_p is ξ_p itself. The proof is complete when we apply Theorem 1. □

4 The Extreme Points of the Constraint Set

Recall the definition of W given in (4). We also define $U(a,t) := \sup_{t_1\in[0,p]} \big(a[t,t_1] - (t_1 \ominus t)s\big)$, for $t \in [0,p]$, $a \in B_p$, and $s > 0$.

Theorem 3 *A point $a \in B_p$ is an extreme point if and only if for almost all $t \in [0,p]$ one of the following hold:*

(X1) a is differentiable at t and $Da(t) = \rho$,

(X2) a is differentiable at t and $Da(t) = 0$,

(X3) $W(a,t) = \beta$,

(X4) $U(a,t) = \beta$.

Proof. a is extreme $\Rightarrow$ condition holds. Let $a \in B_p$. Suppose that there is a set $S \subset [0,p]$ such that none of the four conditions *(X1–4)* hold at any point in S and that $\mathrm{Leb}[S] > 0$. We shall construct $\overline{a}$ and $\underline{a}$ in B_p such that $a = (\overline{a} + \underline{a})/2$.

Since $W(a,\cdot)$ and $U(a,\cdot)$ are continuous on $[0,p]$, the set $\{t \in [0,p] : W(a,t) < \beta, U(a,t) < \beta\}$ is open (in $[0,p]$). We use the fact that every open set of $[0,p]$ can be expressed as a countable disjoint union of open intervals. The intersection of a least one of these intervals with S must have positive Lebesgue measure, for otherwise the countable union, which contains S, would have Lebesgue measure zero. Let G be this interval.

In what follows it will be convenient to work on a compact interval. It is obvious that there exists a compact interval $K \subset G$ such that $\mathrm{Leb}[K \cap S] > 0$. Furthermore, we can find two disjoint compact intervals $K_1, K_2 \subset K$, such that both $\mathrm{Leb}(K_1 \cap S)$ and $\mathrm{Leb}(K_2 \cap S)$ are positive. We may take K_1 to be on the left of K_2 in K, that is if $t_1 \notin K$ and $[t_1,t]_p$ contains a point of K_2, then $K_1 \subset [t_1,t]_p$. Since K is compact and $W(a,\cdot)$ and $U(a,\cdot)$ are continuous, these two functions attain their supremum over K. Therefore $T := \sup_{t\in K} W(a,t)$ and $B := \sup_{t\in K} U(a,t)$ are both strictly less than β.

We shall perturb a in the following way. For each $\delta \in (0,1)$ define the following two maps from $[0,\rho] \to [0,\rho]$:

$$\overline{z}_\delta(x) \;:=\; \begin{cases} (1+\delta)x, & \text{if } x \le \rho/2 \\ (1-\delta)x + \delta\rho, & \text{if } x > \rho/2, \end{cases}$$

$$\underline{z}_\delta(x) \;:=\; \begin{cases} (1-\delta)x, & \text{if } x \le \rho/2 \\ (1+\delta)x - \delta\rho, & \text{if } x > \rho/2. \end{cases}$$

Note that $\overline{z}_\delta(x) \ge x$ and $\underline{z}_\delta(x) \le x$ for all $x \in [0,\rho]$, and that $(\overline{z}_\delta + \underline{z}_\delta)/2 = \mathrm{Id}$, where Id is the identity function. Both maps leave 0 and ρ unchanged. Let

$$k_1(\delta) \;:=\; \int_{K_1} ((\overline{z}_\delta - \mathrm{Id}) \circ Da)(x)\, dx$$

$$\text{and} \qquad k_2(\delta) \;:=\; \int_{K_2} ((\overline{z}_\delta - \mathrm{Id}) \circ Da)(x)\, dx.$$

Clearly both of these functions are continuous and strictly increasing, and therefore they are both invertible in neighbourhood of 0. Choose ϵ in this neighbourhood such that $0 < \epsilon < \min(\beta - T, \beta - B)$. Then we can find $\delta_1, \delta_2 \in (0,1)$ such that $k_1(\delta_1) = k_2(\delta_2) = \epsilon$.

Write $K' := [0,p] - (K_1 \cup K_2)$ and consider the measures $\overline{a}$ and $\underline{a}$ defined by

$$\overline{a}[\mathcal{B}] \;:=\; \int_{\mathcal{B}} \Big(Da.\mathrm{I}_{K'} + (\overline{z}_{\delta_1} \circ Da).\mathrm{I}_{K_1} + (\underline{z}_{\delta_2} \circ Da).\mathrm{I}_{K_2} \Big)(x)\, dx \tag{6}$$

$$\text{and} \qquad \underline{a}[\mathcal{B}] \;:=\; \int_{\mathcal{B}} \Big(Da.\mathrm{I}_{K'} + (\underline{z}_{\delta_1} \circ Da).\mathrm{I}_{K_1} + (\overline{z}_{\delta_2} \circ Da).\mathrm{I}_{K_2} \Big)(x)\, dx. \tag{7}$$

Clearly $a = (\overline{a} + \underline{a})/2$. Since $\overline{z}_\delta(x) > x$ for all $x \in (0, \rho)$ and K_1 and K_2 each contain a set of positive measure in which $0 < Da < \rho$, we have that $\overline{a}[K_1] > a[K_1]$ and that $\underline{a}[K_1] < a[K_1]$. Thus a, $\overline{a}$, and $\underline{a}$ are distinct. and a is in the interior of the line segment $[\overline{a}, \underline{a}]$. We will show that both $\overline{a}$ and $\underline{a}$ are in B_p.

Since neither $\overline{z}_\delta$ nor $\underline{z}_\delta$ take values outside $[0,\rho]$, we have that $0 \le D\overline{a}(x) \le \rho$ for almost all $x \in [0,p]$. We conclude that $\overline{a}$ is a non-negative measure on $[0,p]$ and satisfies the peak rate constraint: $\overline{a}_\epsilon[y,z] \le (z \ominus y)\rho$ for all $y, z \in [0,p]$. The same holds for $\underline{a}$.

Let $t_1, t_2 \in [0,p]$ and write $\kappa := (t_2 \ominus t_1)\sigma$. If $t_2 \notin K$, then $\overline{a}[t_1,t_2]_p \le a[t_1,t_2]_p \le \beta + \kappa$. Otherwise $t_2 \in K$ and so $a[t_1,t_2]_p \le T + \kappa$. Therefore $\overline{a}[t_1,t_2]_p \le a[t_1,t_2]_p + \epsilon \le T + \kappa + \epsilon < \beta + \kappa$. Thus $\overline{a}$ obeys the leaky bucket constraint.

An identical argument with time reversed shows that $\underline{a}$ also obeys the leaky bucket constraint.

condition holds $\Rightarrow$ a is extreme. Suppose that a obeys at least one of *(X1–4)* at almost every time $t \in [0,p]$ and that $a = (\overline{a} + \underline{a})/2$ for two traffic realisations $\overline{a}$ and $\underline{a}$ in B_p. Since a, $\overline{a}$, and $\underline{a}$ are non-negative measures, the derivatives of all three exist almost everywhere. Let $t \in [0,p]$ be such that a obeys one of the conditions *(X1–4)* at t, and $\overline{a}$ and $\underline{a}$ are both differentiable at t. Depending on which condition a obeys at t, we have the following four cases.

- If $Da(t) = 0$ then $D\overline{a}(t) = D\underline{a}(t) = 0$, since $Da(t) = [D\overline{a}(t) + D\underline{a}(t)]/2$ and both $D\overline{a}(t)$ and $D\underline{a}(t)$ lie in the interval $[0, \rho]$.
- Similarly, if $Da(t) = \rho$ then $D\overline{a}(t) = D\underline{a}(t) = \rho$.
- If $W(a,t) = \beta$ then $W(\overline{a},t) = W(\underline{a},t) = \beta$ since $W(a,t) \le (W(\overline{a},t) + W(\underline{a},t))/2$ and both $W(\overline{a},t)$ and $W(\underline{a},t)$ are less than or equal to β.
- Similarly, if $U(a,t) = \beta$ then $U(\overline{a},t) = U(\underline{a},t) = \beta$.

We have shown that if a obeys any one of the four conditions at t, then $\overline{a}$ and $\underline{a}$ also obey the same condition at t. Since either of the last two conditions imply that the derivative is the leaky bucket service rate σ, we conclude that $Da = D\overline{a} = D\underline{a}$ almost everywhere. Applying the Fundamental Theorem of Calculus and using the absolute continuity of a, $\overline{a}$, and $\underline{a}$, we find that $a = \overline{a} = \underline{a}$. Thus, a is not in the interior of any line segment contained in B_p and is therefore an extreme point of B_p. □

Note that if $a \in B_p$ obeys either *(X3)* or *(X4)* at an instant t and a is differentiable there, then $Da(t) = \sigma$. Thus, at any instant t, any source with realisations obeying the condition of Theorem 3 may only transmit at one of three rates: 0, the peak rate ρ, or the leaky bucket service rate σ. The source can transmit at rate 0 or ρ at any time, but may transmit at rate σ only when the leaky bucket buffer is full, or is empty and will fill before any service goes unused. Figure 2 shows a typical extreme point of B_p. The figures 0 and β above the diagram indicate the times when the leaky bucket is full and when it is empty.

We see that B_p has many extreme points. In fact we will show that the extreme points form a dense subset of B_p. Heuristically, any source behaviour can be approximated by the source transmitting at rates 0 and ρ. All that is needed is for the source to switch between these two rates sufficiently quickly and to spend the appropriate proportion of time transmitting at each rate. This result means that we did not need to use the convexity of the functional E to show that it approaches its supremum on the set ξ_p, all we needed was its

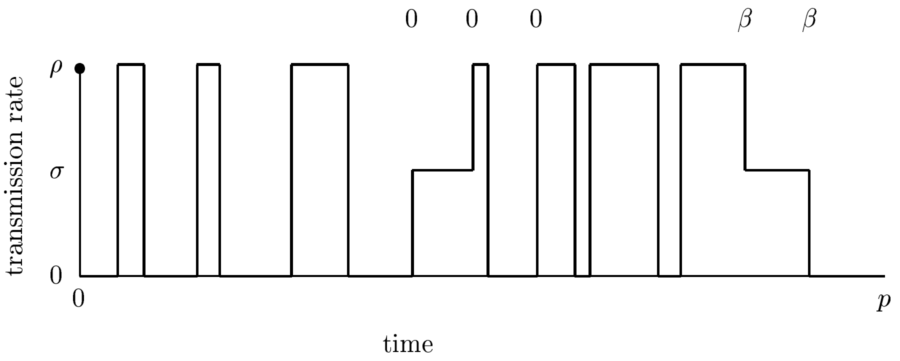

Figure 2: A typical extreme point of C. The values above the figure give the content of the leaky bucket at that time.

continuity. Unfortunately, we have not yet reduced the domain of our optimisation significantly. In the next section we will see how, in another linear structure on the space of traffic processes, convexity is crucial and enables us to make a more meaningful reduction in the domain of our optimisation problem.

Theorem 4 *The constraint set B_p is the closure of the set of its extreme points ξ_p.*

Proof. If w is a signed measure on $[0,p]$, we write $||w|| := \sup_{t_1,t_2\in[0,p]} |w(t_1,t_2)_p|$. Let $a \in B_p$. For some $\eta \in (0,1)$, let

$$b := (1-\eta)a.$$

Then $||b-a|| = \eta a[0,p]$. Now choose $N \in \mathbb{N}$ and let c be the unique element of $\mathbb{D}_p^+$ such that, for each $i \in 0,\dots,N-1$,

$$c[\frac{ip}{N},x] = \min\Big(b_i,(x-\frac{ip}{N})\rho\Big), \qquad \text{for } \frac{ip}{N} \le x < \frac{(i+1)p}{N},$$

where $b_i := b[ip/N,(i+1)p/N)$. Heuristically, we are breaking the interval $[0,p]$ into N partitions and rearranging the traffic within each so that each partition contains a single burst at the peak rate. Observe that $||c-b|| \le 2\rho p/N$. Thus $c[t_1,t_2]_p \le (1-\eta)\beta + 2\rho p/N + (t_2 \ominus t_1)\sigma$ for all $t_1,t_2 \in [0,p]$. If $\beta\eta \ge 2\rho p/N$, then c will obey the leaky bucket constraint. Thus, by making N large and η small while ensuring that this condition is met, we can find an element of B_p as close as we like to a. □

5 Another Linear Structure

In this section we consider another way of representing traffic realisations. Here the time taken for a certain amount of traffic to arrive is specified rather than the amount of traffic that arrives within each interval. This representation leads to a linear structure on the set of arrival paths which is different from the one considered in the previous section. It turns out that the functional E considered previously is convex in this linear structure also. Moreover, the set of extreme points of the constraint set in this linear structure is strictly smaller. This enables us to derive a stronger result, restricting even further the set over which the optimisation must be performed.

Let

$$\mathbb{D}_{p,q}^+ := \{a \in \mathbb{D}_p^+ : a[0,p] = q\}.$$

We define the map $\triangleleft : \mathbb{D}_{p,q}^+ \to \mathbb{D}_{q,p}^+$ by

$$(\triangleleft a)(c) := \sup\{t \in [0,p] : a[0,t] \le c\}.$$

Note that $\triangleleft$ is a bijection and is its own inverse. If the distribution function of a, which is given by $a[0,\cdot]$, is strictly increasing and continuous, then so is the distribution function of $\triangleleft a$ and these functions are inverses.

In this setting we wish to maximise the functional $\overline{E}(d) := E(\triangleleft d)$ subject to $\triangleleft d \in B_p$.

We restrict our attention to sources which average transmission rate σ, the maximum allowed. The constraint set of interest is therefore $F_q := \{d \in \mathbb{D}^+_{q,q/\sigma} : \triangleleft d \in B_{q/\sigma}\}$. Note that now elements of the constraint set may have atoms. These represent time intervals time intervals during which the source is silent.

It is shown in [14](Lemma A4) that the map $\triangleleft$ is a homeomorphism between $\mathbb{D}^+_{p,q}$ and $\mathbb{D}^+_{q,p}$. An immediate corollary is that the functional $\overline{E}$ is continuous in $\mathbb{D}^+_{q,p}$. Establishing convexity was trivial in the linear structure of the previous two sections, but here it is considerably more difficult. We will first prove the convexity of $\overline{E}$ in the space of *simple* realisations and then extend the result to $\mathbb{D}^+_{q,q/\sigma}$ using a continuity argument. By a simple realisation, we mean one that can be written

$$a = \sum_{i=1}^{n} m_i \delta_{t_i},$$

with $n \in \mathbb{N}$, $t_i \in [0, q]$, and $m_i \in [0, p]$ with $\sum_{i=1}^{n} m_i = q$. We denote by $A_{q,p}$ the set of elements of $\mathbb{D}^+_{q,p}$ which have this property. Note that the if a is in $A_{p,q}$ then $\triangleleft a$ is in $A_{q,p}$. Also note that $A_{q,p}$ is a convex subset of $\mathbb{D}^+_{q,p}$.

Simple realisations are useful because they may be used to approximate any other realisation.

Lemma 4 *For each $p, q > 0$, the set $A_{q,p}$ is dense in $\mathbb{D}^+_{q,p}$.*

Proof. Let $d \in \mathbb{D}^+_{q,p}$ Define the sequence of simple realisations

$$d_n[0, c] := \sum_{i=1}^{n} d\left[\frac{(i-1)p}{n}, \frac{ip}{n}\right) \delta_{ip/i},$$

for $n \in \mathbb{N}$. Clearly, $d_n \in A_{q,p}$. We have that $\left|d[0, c] - d_n[0, c]\right| \leq p/n$ for all $c \in [0, q]$, and so $d_n \to d$ as $n \to \infty$. □

Lemma 5 *The functional $\overline{E}$ is convex on $A_{q,p}$.*

Proof. Write $g(x) := \exp(\vartheta x)$. The only property of g we will use is its convexity. To show that $\mathcal{E}_{\vartheta,T}$ is convex on $A_{q,p}$ it will suffice to show that the one-dimensional function $f(\alpha) : [0, 1] \to \mathbb{R} := \overline{E}((1-\alpha)d_1 + \alpha d_2)$ is convex for any two points d_1 and d_2 in $A_{q,p}$.

Fix $d_1, d_2 \in A_{q,p}$. For $\alpha \in (0, 1)$, let a_α be the realisation in $A_{p,q}$ such that $\triangleleft a_\alpha := (1-\alpha)d_1 + \alpha d_2$. This realisation may be expressed as $a_\alpha = \sum_{i=1}^{n} m_i \delta_{t_i(\alpha)}$, where the m_i are constant and the t_i are affine in α. We write v_i for the slope of t_i.

Denote by J_α the set of ordered pairs (c_1, c_2) such that $t_{c_2}(\alpha) \ominus t_{c_1}(\alpha) = T$ and $v_{c_1} \neq v_{c_2}$. Also, denote by K the set of ordered pairs (c_1, c_2) such that $t_{c_2}(\alpha) \ominus t_{c_1}(\alpha) = T$ and $v_{c_1} = v_{c_2}$. The set K needs no subscript since it does not change with α. There are only a finite number of values of α for which J_α is nonempty. For any other value of α, let $\eta \in (0, T)$ be such that $2 \max_c |v_c \eta|$ is less than both

$$\min_{(c_1,c_2) \notin K} \Big[t_{c_2}(\alpha) \ominus (t_{c_1}(\alpha) + T)\Big] \qquad \text{and} \qquad \min_{1 \leq c_1, c_2 \leq n} \Big[t_{c_2}(\alpha) \ominus t_{c_1}(\alpha)\Big].$$

Consider first the case when K is empty. Then, for each set $[x, x+T]_p$, there may be at most one $i \in 1, \ldots, n$ such that $t_i(\gamma)$ passes across the boundary of the set as γ ranges from α to $\alpha + \eta$. We conclude that

$$f(\alpha + \eta) - f(\alpha) = \sum_{c=1}^{n} v_c \eta \Big[g(z_c + m_c) - g(z_c) + g(y_c) - g(y_c + m_c)\Big], \tag{8}$$

where $z_i(\alpha) := a(t_i, t_i + T)_p$ and $y_c(\alpha) := a(t_i - T, t_i)_p$. Here we are suppressing the dependence on α for clarity.

Now suppose that K is non-empty and contains a pair (c_1, c_2). In this case $z_{c_1}(\alpha) = y_{c_2}(\alpha)$ for all $\alpha \in (0, 1)$ and $v_{c_1} = v_{c_2}$. Suppose v_{c_1} is positive (the case when it is negative is similar). Then the contribution made by the intervals $\{[x, x+T]_p; t_{c_1} \le x \le t_{c_1} + v_{c_1}\eta\}$ to the integral (5) at $\gamma = \alpha$ is $v_{c_1}\eta\, g(z_{c_1} + m_{c_2})$ and at $\gamma = \alpha + \eta$ is $v_{c_1}\eta\, g(z_{c_1} + m_{c_1})$. The increase is equal to

$$v_{c_1}\eta\big[g(z_{c_1} + m_{c_1}) - g(z_{c_1})\big] + v_{c_2}\eta\big[g(y_{c_2} + m_{c_2}) - g(y_{c_2})\big].$$

So we see that expression (8) holds in case also.

We conclude from the considerations above that f is piecewise affine and that changes in its slope occur at those values of α for which J_α is non-empty. Now fix α at such a value. For each pair $(c_1, c_2) \in J_\alpha$, let $z_{c_1 c_2}(\alpha) := z_{c_1}(\alpha) = y_{c_2}(\alpha)$. Then the change in the derivative of f at α is given by

$$\begin{aligned}
\Delta f'(\alpha) &= \sum_{(c_1,c_2)\in J_\alpha} v_{c_1}\,\mathrm{sgn}(v_{c_1} - v_{c_2})\Big[g(z_{c_1c_2} + m_{c_1} + m_{c_2}) - g(z_{c_1c_2} + m_{c_1}) - g(z_{c_1c_2} + m_{c_2}) + g(z_{c_1c_2})\Big] \\
&\quad + \sum_{(c_1,c_2)\in J_\alpha} v_{c_2}\,\mathrm{sgn}(v_{c_1} - v_{c_2})\Big[g(z_{c_1c_2} + m_{c_1}) - g(z_{c_1c_2}) - g(z_{c_1c_2} + m_{c_1} + m_{c_2}) + g(z_{c_1c_2} + m_{c_2})\Big] \\
&= \sum_{(c_1,c_2)\in J_\alpha} |v_{c_1} - v_{c_2}|\Big[g(z_{c_1c_2} + m_{c_1} + m_{c_2}) - g(z_{c_1c_2} + m_{c_1}) - g(z_{c_1c_2} + m_{c_2}) + g(z_{c_1c_2})\Big],
\end{aligned}$$

where $\mathrm{sgn}(x)$ is 1, 0, or -1, depending on whether x is positive, zero, or negative. From the convexity of g we have that

$$g(z_{c_1c_2} + m_{c_1}) \le \frac{m_{c_1}}{m_{c_1} + m_{c_2}} g(z_{c_1c_2} + m_{c_1} + m_{c_2}) + \frac{m_{c_2}}{m_{c_1} + m_{c_2}} g(z_{c_1c_2})$$

and

$$g(z_{c_1c_2} + m_{c_2}) \le \frac{m_{c_2}}{m_{c_1} + m_{c_2}} g(z_{c_1c_2} + m_{c_1} + m_{c_2}) + \frac{m_{c_1}}{m_{c_1} + m_{c_2}} g(z_{c_1c_2}).$$

Therefore $\Delta f'(\alpha) \ge 0$, and the convexity of f follows. □

Lemma 6 *Let S be a convex subset of a metrisable topological vector space and let $T \subset S$ be convex and dense in S. If $f : S \to \mathbb{R}$ is a continuous function and its restriction to T is convex, then f is also convex on S.*

Proof. Let $p, q \in S$ and let $\alpha \in [0,1]$. Then there exist sequences p_n and q_n in T such that $p_n \to p$ and $q_n \to q$. Consider the sequence $r_n := (1-\alpha)p_n + \alpha q_n$ which lies in T since this set is convex. The operations of addition and multiplication by a scalar are continuous, and so this sequence must approach the limit $r := (1-\alpha)p + \alpha q$. Now the continuity of f implies that $f(p_n) \to f(p)$, that $f(q_n) \to f(q)$, and that $f(r_n) \to f(r)$. Furthermore, the convexity of f implies that $f(r_n) \le (1-\alpha)f(p_n) + \alpha f(q_n)$ for all $n \in \mathbb{N}$. It follows that $f(r) \le (1-\alpha)f(p) + \alpha f(q)$ and therefore f is convex on S. □

We can now extend the convexity result to the whole of $\mathbb{D}^+_{q,p}$.

Theorem 5 *The functional $\overline{E}$ is convex on $\mathbb{D}^+_{q,p}$.*

Proof. We have shown that $\overline{E}$ is continuous on $\mathbb{D}^+_{q,p}$ and convex on a dense convex subset $A_{q,p}$ of $\mathbb{D}^+_{q,p}$. The conclusion follows from an application of the previous lemma. □

The following theorem, combined with Theorem 7, is the main result of this paper.

Theorem 6 *The worst case effective bandwidth $\sup\{\mathcal{E}_{\vartheta,T}(\mu) : \mu \in \mathcal{C}\}$ is equal to*

$$\sup\{(\vartheta T)^{-1} \log \overline{E}(d) : q > 0, d \in \varphi_q\},$$

where φ_q is the set of extreme points of F_q.

Proof. The proof of this result is analogous to that of Theorem 2. □

6 Extreme Points in the Alternative Representation

We will need the following definitions. For $c_1, c_2 \in [0,q]$, write $z(c_1,c_2)_q := c_2 - c_1 - \sigma d(c_1,c_2)_q$. Clearly, $d \in F_q$ if and only if $z(c_1,c_2)_q \le \beta$ for all $c_1, c_2 \in [0,q]$. For $d \in \mathbb{D}_{q,p}$ and $c \in [0,q]$, we define

$$\begin{aligned} W^*(d,c) &:= \sup_{c_1 \in [0,q]} z(c_1,c)_q \\ \text{and} \quad U^*(d,c) &:= \sup_{c_1 \in [0,q]} z(c,c_1)_q. \end{aligned}$$

Let $\mathcal{W}$ and $\mathcal{U}$ be defined by

$$\begin{aligned} \mathcal{W} &:= \{y \in [0,q] : z(x,y)_q = \beta \text{ for some } x \in [0,q]\} \\ \mathcal{U} &:= \{x \in [0,q] : z(x,y)_q = \beta \text{ for some } y \in [0,q]\}. \end{aligned}$$

In contrast to the elements of C_p, the elements of F_q may have a singular part. Recall [15] that every measure d can be decomposed into the sum of a singular measure d_s and an absolutely continuous measure d_a. Also, the support of a measure is the intersection of all closed subsets of full measure.

We have the following characterisation of extreme points of F_q.

Theorem 7 *$d \in F_q$ is an extreme point if and only if both the following conditions hold:*

(Y) for almost every $c \in [0,q]$, one of the following hold

- *(Y1) d is differentiable at c and $Dd(c) = 1/\rho$,*
- *(Y2) $W^*(d,c) = \beta$,*
- *(Y3) $U^*(d,c) = \beta$.*

(Z) for each pair of distinct points c_1 and c_2 in the support of d_s, either $\mathcal{U} \cap [c_1,c_2)_q$ or $\mathcal{W} \cap (c_1,c_2]_q$ is non-empty.

We will postpone the proof of this theorem. Instead, we will now show that the set of extreme points of F_q is strictly smaller than that of C_p.

Theorem 8 *If d is an extreme point of F_q then $\lhd d$ is an extreme point of C_p where $p := q/\sigma$.*

Proof. Define $a := \lhd d$ and let $g : [0,p] \to \mathbb{R} : t \mapsto a[0,t]$ be its distribution function. Since

$$\begin{aligned} d[0,c] &= \sup\{t \in [0,p] : g(t) \le c\} \\ &= \operatorname{Leb} g^{-1}[0,c], \end{aligned}$$

we have that $d[\mathcal{B}] = \operatorname{Leb}[g^{-1}\mathcal{B}]$ for all Borel subsets $\mathcal{B}$ of $[0,q]$.

We define the following subsets of $[0,p]$:

$$\begin{aligned} Z &:= \{t \in [0,p] : Da(t) \text{ exists and equals } 0\}, \\ P &:= \{t \in [0,p] : Da(t) \text{ exists and is strictly positive}\}. \end{aligned}$$

Since Da exists almost everywhere, $\operatorname{Leb} Z + \operatorname{Leb} P = p$. Note that g is a bijection from P to $g(P)$. If $t \in P$ and $c := g(t)$ then $g'(t) = Da(t) = 1/(Dd(c))$. We also have that $W(a,t) = W^*(d,c)$ and that $U(a,t) = U^*(d,c)$. (Recall that W and U were defined in Sections 3 and 4.) It follows that a obeys $(X1)$, $(X3)$, or $(X4)$ at t if and only if d obeys, respectively, $(Y1)$, $(Y2)$, or $(Y3)$ at c.

It remains to show that $\operatorname{Leb}[g^{-1}(K)] = 0$, where

$$K := \{c \in g(P) : \text{none of } (Y1\text{–}3) \text{ hold for } d \text{ at } c\ \}.$$

Let d_a and d_s be the absolutely continuous and singular parts respectively of d. Since K has Lebesgue measure zero, $d_a[K] = 0$. It is well known [15] that any singular measure μ has infinite derivative μ-almost everywhere.

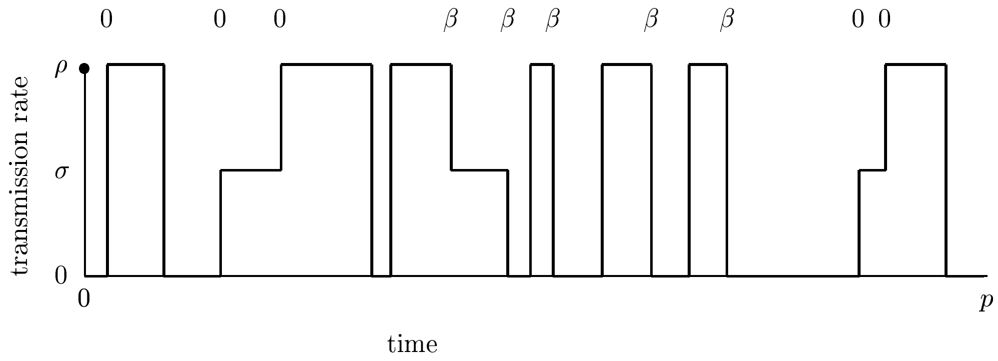

Figure 3: A typical extreme point of F in this linear structure. The values above the figure give the content of the leaky bucket at that time. Note that each burst must either start when the leaky bucket is empty or end when it is full.

Thus $d_s[g(P)] = 0$ since Dd is finite on this set. But $K \subset g(P)$ and so we conclude that $d[K] = 0$. Thus $\mathrm{Leb}\,[g^{-1}K]$ is indeed zero. $\square$

Realisations which are extreme in $C_{q/\sigma}$ but not in F_q are easy to find. For example, if $\rho = 2\sigma = 1$, $q = 5$, and $\beta > 1/4$, then such a realisation is given by

$$a[\mathcal{B}] := \mathrm{Leb}\left[\mathcal{B} \cap \bigcup_{i=0}^{9} [i, i + \frac{1}{2}]\right], \qquad \text{for all Borel subsets } \mathcal{B} \text{ of } [0, 10].$$

This realisation switches on at integer times and switches off at half integer times.

We break the proof of Theorem 7 into the following three lemmas.

Lemma 7 *If d is an extreme point of F_q, then condition (Y) holds.*

Proof. The proof of this lemma follows that of Theorem 3 and we will only describe the main differences. Firstly, $W^*(d, \cdot)$ and $U^*(d, \cdot)$ are not continuous, only upper-semicontinuous. This does not significantly alter the proof however. Another difference is that the functions $\overline{z}_\delta$ and $\underline{z}_\delta$ should now map the interval $[1/\rho, \infty)$ to itself. A suitable choice of maps is

$$\begin{aligned} \overline{z}_\delta(x) &:= (1+\delta)x - \delta/\rho \\ \text{and} \qquad \underline{z}_\delta(x) &:= (1-\delta)x + \delta/\rho. \end{aligned}$$

Finally, analogues of expressions (6) and (7) give only absolutely continuous measures. To find $\overline{d}$ and $\underline{d}$ such that $d = (\overline{d} + \underline{d})/2$, we must add the singular parts of d. $\square$

Lemma 8 *If d is an extreme point of F_q, then condition (Z) holds.*

Proof. Suppose $d \in F_q$ is such that, for some c_1 and c_2 in the support of d_s, both

$$\sup_{c \in (c_1, c_2)_q} W^*(d, c) \qquad \text{and} \qquad \sup_{c \in (c_1, c_2)_q} U^*(d, c)$$

are strictly less than γ, where $\gamma < \beta$. In the following manner, we will construct $\overline{d}, \underline{d} \in F_q$ such that $\overline{d}$, d, and $\underline{d}$ are distinct and $(\overline{d} + \underline{d})/2 = d$.

If d_s has no atom at c_1, then both $W^*(d, \cdot)$ and $U^*(d, \cdot)$ are continuous there and so there is some open interval I_1 containing c_1 such that both $W^*(d, c_1)$ and $U^*(d, c_1)$ are less than γ for all $c \in I_1$. In this case, define a positive measure Δ_1 such that $\Delta_1[\mathcal{B}] := d_s[I_1 \cap \mathcal{B}]$ for all Borel subsets $\mathcal{B}$ of $[0, q]$.

If, on the other hand, d_s has an atom at c_1, then define Δ_1 to be the measure on $[0,q]$ with total mass 1, concentrated at c_1. In this case, $I_1 := \{c_1\}$.

Replacing c_1 with c_2 in the definition above, we similarly define an interval I_2 and a positive measure Δ_2. Note that we may choose I_1 and I_2 in such a manner that they do not overlap.

Choose $\delta_1 > 0$ and $\delta_2 > 0$ such that $\delta_1\Delta_1[0,q] = \delta_2\Delta_2[0,q]$ and such that this number is less than $\beta - \gamma$. We now define

$$\begin{aligned} \overline{d} &:= d - \delta_1\Delta_1 + \delta_2\Delta_2 \\ \text{and} \quad \underline{d} &:= d + \delta_1\Delta_1 - \delta_2\Delta_2. \end{aligned}$$

Let $x, y \in [0,q]$. Suppose y is in the interior of $I_1 \cup (c_1,c_2) \cup I_2$. Then, $z(x,y)_q \le W^*(d,y) \le \gamma$ and we have that $\overline{z}(x,y)_q \le z(x,y)_q + \beta - \gamma \le \beta$. Suppose, on the other hand, that y is not in the interior of that set. Then $\overline{z}(x,y)_q \le z(x,y)_q \le \beta$. Either way, we conclude that $\overline{d}$ is in F_q.

A symmetrical argument (with time reversed) shows that $\underline{d}$ is also in F_q.

We have shown that if d is an extreme point of F_q, then either $\sup_{c\in(c_1,c_2)} W^*(d,c)$ or $\sup_{c\in(c_1,c_2)} U^*(d,c)$ equals β. Suppose that the former is the case. (The latter case is treated similarly.) Since $W^*(d,\cdot)$ is upper semi-continuous, it attains its maximum over every compact interval $[c_1,c_2]_q$. Suppose $W^*(d,c_1) = \beta$. Note that if d obeys $(Y2)$ or $(Y3)$ at a point c and d is differentiable there, then $Dd(c) = 1/\sigma$. By the previous lemma, d_a, the absolutely continuous part of d, obeys $d_a \le \sigma^{-1}\mathrm{Leb}$. It follows that if $W^*(d,c_2) < \beta$, then $d_s(c_1,c_2)_q > 0$. Thus d_s has a point of support x in the set $(c_1,c_2)_q$. By the above, $\sup_{c\in(x,c_2)} W^*(d,c) = \beta$ and this supremum is attained by some point in $[x,c_2]_q$, which is a subset of $(c_1,c_2]_q$. □

Lemma 9 *If $d \in F_q$ obeys conditions (Y) and (Z), then d is an extreme point of F_q.*

Proof. Assume that d obeys the given conditions and that $d = (\overline{d} + \underline{d})/2$ for $\overline{d}, \underline{d} \in F_q$. Let $c \in [0,q]$ be such that d obeys one of $(Y1)$, $(Y2)$, or $(Y3)$ at c and $\overline{d}$ and $\underline{d}$ are both differentiable at c. The argument proceeds as in the proof of Theorem 3. We show that if d obeys one of the three sub-conditions of (Y) at c, then both $\overline{d}$ and $\underline{d}$ obey the same condition. Since both $(Y2)$ and $(Y3)$ imply that the derivative at c is $1/\sigma$, we conclude that $Dd = D\overline{d} = D\underline{d}$ almost everywhere. Thus the absolutely continuous parts of d, $\overline{d}$, and $\underline{d}$ are equal.

Note that if $x \in \mathcal{U}$ and $y \in \mathcal{W}$ then $z(x,y)_q = \beta$. To see this, observe that there exists $x' \in \mathcal{U}$ and $y' \in \mathcal{W}$ such that $z(x,y')_q = z(x',y)_q = \beta$. Then, since $z[y',x']_q \ge -\beta$,

$$\beta \ge z(x,y)_q = z(x,y')_q + z[y',x']_q + z(x',y)_q \ge \beta.$$

We define $\overline{z}$ and $\underline{z}$ analogously to z. Then $z = (\overline{z} + \underline{z})/2$. Let $x \in \mathcal{U}$ and $y \in \mathcal{W}$. Since neither $\overline{z}$ nor $\underline{z}$ can exceed β on $(x,y)_q$, we conclude that $\overline{z}(x,y)_q = \underline{z}(x,y)_q = \beta$. This implies that d, $\overline{d}$, and $\underline{d}$ agree on $(x,y)_q$. Since the absolutely continuous parts of these measures are everywhere equal, their singular parts, d_s, $\overline{d}_s$, and $\underline{d}_s$, must also be equal on this set.

We denote the support of d_s by S and define

$$\Pi := \{S \cap (x,y)_q : x \in \mathcal{U},\, y \in \mathcal{W}\}.$$

Let Π^* be the set of all subsets of $[0,q]$ which can be expressed as the difference of two elements of Π. Then d_s, $\overline{d}_s$, and $\underline{d}_s$ agree on all elements of Π^*.

To show that these three measures agree on every Borel subset of S, it will be sufficient to show that they agree on every open subset. Moreover, each open subset of S can be expressed as a countable disjoint union of sets of the form $S \cap (a,b)_q$, with $a, b \in [0,q]$. Hence we need only consider sets of this form. Let G be such a set. We will define an increasing sequence of sets P_i such that $P_i \cap S$ is contained in Π^* and $\bigcup_i P_i \cap S = G$.

Suppose there exists a sequence v_i in S which decreases strictly to a. Let a_i be the point given by Condition (Z) applied to the two points v_i and v_{i+1}. This will be the left boundary point of P_i. Take P_i to be closed on the left if a_i is in $\mathcal{W}$ and open if it is in $\mathcal{U}$.

Suppose, on the other hand, that there is no such sequence v_i. Then we can find $l, r \in S$ such that $a \in [l,r)_q$ and $(l,r)_q$ contains no point of S. Let a' be the point given by Condition (Z) applied to l and r. We now define the left boundary point of P_i to be a' for all $i \in \mathbb{N}$. Again, we take P_i to be closed on the left if a' is in $\mathcal{W}$ and open if this point is in $\mathcal{U}$.

The right boundary point of P_i is defined, for each $i \in \mathbb{N}$, in an analogous way. It may be easily verified that each member of the sequence $\{S \cap P_i : i \in \mathbb{N}\}$ is in Π^*, that this sequence is increasing, and that the union of all its members is equal to G. $\square$

References

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

[2] 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.

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

[4] 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.

[5] Cormac Walsh. Worst case traffic from sources constrained by policed sources. Technical report, BRIMS, HP Labs, 2000.

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

[7] S. Shioda. Worst-case performance of ATM multiplexer. *J. Oper. Res. Soc. Japan*, 41(1):3–20, 1998.

[8] 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.

[9] 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.

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

[12] T. Kesidis, G.; Konstantopoulos. Worst-case performance of a buffer with independent shaped arrival processes. *IEEE Communication Letters*, 4(1):26–28, 2000.

[13] V. P. Maslov and S. N. Samborskiĭ, editors. *Idempotent Analysis*, volume 13 of *Advances in Soviet Mathematics*. American Mathematical Society, 1992.

[14] A. Puhalskii. Large deviation analysis of the single server queue. *Queueing Systems*, 21:5–65, 1995.

[15] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 1987.