

A Calculus Approach to Energy-Efficient Data Transmission with Quality-of-Service Constraints

Murtaza Zafer, *Member, IEEE*, and Eytan Modiano, *Senior Member, IEEE*

Abstract—Transmission rate adaptation in wireless devices provides a unique opportunity to trade-off data service rate with energy consumption. In this paper, we study optimal rate-control to minimize the transmission energy expenditure subject to strict deadline or other quality-of-service (QoS) constraints. Specifically, the system consists of a wireless transmitter with controllable transmission rate and with strict QoS constraints on data transmission. The goal is to obtain a rate control policy that minimizes the total transmission energy expenditure while ensuring that the QoS constraints are met. Using a novel formulation based on cumulative curves methodology, we obtain the optimal transmission policy and show that it has a simple and appealing graphical visualization. Utilizing the optimal “offline” results, we then develop an online transmission policy for an arbitrary stream of packet arrivals and deadline constraints, and show, via simulations, that it is significantly more energy efficient than a simple head-of-line drain policy. Finally, we generalize the optimal policy results to the case of time-varying power-rate functions.

Index Terms—Energy efficiency, Delay constraints, Quality of Service, Rate control, Network calculus, Wireless.

I. INTRODUCTION

Services envisioned in modern communication systems extend beyond traditional voice communication to enhanced data applications such as video and real-time multimedia streaming, high-throughput data access and Voice-over-IP [1]. Invariably, meeting the quality-of-service (QoS) requirements for these applications translates into stricter packet-delay and throughput constraints. Wireless systems also generally have strict limitations on energy consumption thereby necessitating efficient utilization of this resource [2]. For example, minimizing energy consumption leads to improved battery utilization for mobile devices, increased lifetime for sensor nodes and ad-hoc networks, and better utilization of limited energy sources in satellites. Since in many scenarios transmission energy constitutes a significant portion of the total energy expenditure for wireless nodes [2], it is imperative to minimize this cost to achieve significant energy savings; henceforth, in this paper we will focus solely on transmission energy expenditure.

A. Motivation and Summary

Modern wireless devices are equipped with rate-adaptive capabilities [3], [4] which allows the transmitter to adjust

This work was supported by NSF ITR grant CCR-0325401, by DARPA/AFOSR through the University of Illinois grant no. F49620-02-1-0325, by ONR grant number N000140610064 and by NASA Space Communication Project grant number NAG3-2835.

M. Zafer (email: mzafer@us.ibm.com) is a Research Staff Member at IBM T. J. Watson Research center, NY, USA, and E. Modiano (email: modiano@mit.edu) is an Associate Professor at Massachusetts Institute of Technology, MA, USA.

the transmission rate over time. This is achieved in various ways that include adjusting the power level, symbol rate, coding scheme, constellation size and any combination of these approaches. Associated with a rate, there is a corresponding power expenditure which is governed by the power-rate function. Specifically, a power-rate function is a relationship which gives the amount of transmission power that would be required to transmit at a certain rate. Keeping the bit-error probability fixed, it is widely known that for most encoding schemes the required power is a convex function of the rate [7]–[10], [12], [14], [17]. This implies, from Jensen’s inequality, that transmitting data at a low rate and over a longer duration has less energy cost as compared to a fast rate transmission. However, with QoS constraints taken into consideration, a low rate transmission may not always be able to meet these constraints; thus, there is a tradeoff ! In this work, we seek to obtain the optimal rate-control policy that minimizes the transmission energy expenditure while also ensuring that the strict QoS constraints are met.

We consider a transmitter with data arrivals that have strict QoS constraints such as individual packet deadlines, finite buffer or other service constraints. We represent the arrivals as a cumulative curve (known as the arrival curve) and model the QoS constraints using the concept of a *minimum departure curve*. The minimum departure curve helps translate fairly general QoS constraints into a simple and graphical form. Using this model, we first consider a time-invariant power-rate function and obtain the optimal policy under the knowledge of the arrival curve. The optimal policy has a simple and appealing graphical visualization as discussed later. Using the optimal solution, a heuristic online policy is developed which does not require prior knowledge of the arrival process; the online policy is shown to be energy-efficient via simulations. Finally, in the latter half of the paper, we extend the results to a setup involving a time-varying power-rate function.

B. Related Work

Transmission rate adaptation/control is an active area of research in communication networks in various different contexts. Adaptive network control and scheduling have been studied in the context of network stability [16]–[19], average throughput [20]–[23], average delay [7], [14] and packet drop probability [15]. This literature considers “average metrics” that are measured over an infinite time horizon and hence do not directly apply for delay constrained and real-time data. Incorporating packet deadlines and other strict QoS constraints introduces new challenges and complexity in the problem; recent work in this direction includes [8]–[12].

The work in [8] studied the problem of a known stream of packet arrivals that must be transmitted by a common deadline using minimum energy. In [12], the problem was extended by allowing different energy functions for different packets, where, the authors proposed the MoveRight algorithm that eventually converges to the optimal solution; however, the actual analytical form of the optimal solution was not obtained. In [10], the authors considered batch arrivals and packet deadlines as the QoS metric and utilized filtering techniques to obtain an energy efficient transmission policy. In this paper, we provide a simple yet general framework for the QoS constrained energy minimization problem from which these earlier results can be recovered back as special cases (see Section III-C). The work in [11] has a different context wherein the transmitter can recover back partial energy lost while it is in the idle state, whereas in [9], the authors studied several data transmission problems using Dynamic Programming. The dynamic programming methodology, however, in most cases leads to numerical solutions without much insights.

Within a different context in [25], the problem of transmitting stored video file from a server to a client over a network was considered. Utilizing buffering at the client, the optimal policy was obtained that minimizes the bit-rate variability. Strikingly, the mathematical formulation in the work of [25] has similarities to that considered in this paper, hence, the optimal policies share various properties such as the shortest path feature (albeit with different contextual meanings). However, the solution methodology in [25] is based on a majorization technique which is only suitable for discrete data models. In contrast, our approach is based on continuous-time convex optimization, and applies to both discrete and fluid data models. Moreover, in this paper, we also develop the optimal solution for a time-varying power-rate relationship, which was not addressed in the stored video context. In another context, [26] studied the problem of job scheduling for dynamically variable voltage processor where similar properties (such as the shortest path property of the optimal solution) were observed. Finally, part of the results presented in this paper have appeared in our preliminary work in [6].

The rest of the paper is organized as follows. In Section II we present the system model; in Section III we present the optimal policy for the case of time-invariant power-rate function, and finally, in Section IV we generalize the setup and consider a time-varying power-rate function.

II. SYSTEM MODEL

We consider a continuous-time model and assume that rate can be varied continuously in time. Such a model is an approximation of an actual system, but the assumption is justified, since in practice the communication slots over which rate-control can be done are of the order of 1 msec duration [3], and much smaller than packet delay requirements which are typically on the order of 100's of msec. The advantage of such a model is that it makes the problem that we consider mathematically tractable and also provides a simple and intuitive graphical visualization of the optimal solution.

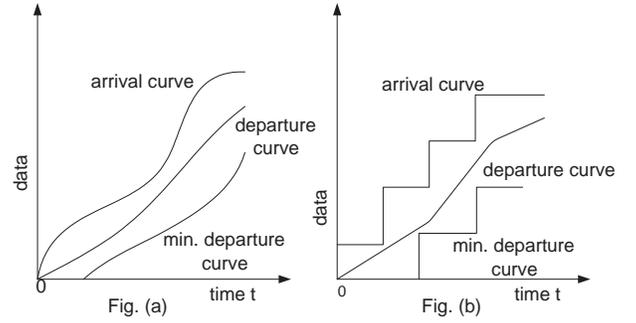


Fig. 1. Data flow model: (a) Fluid arrival model, (b) Packetized arrival model

The results thus obtained can then be applied to a discrete-time system in a straightforward manner by simply evaluating the solution at the discrete-time slot boundaries.

A. Data Flow Model

To describe the flow of data into the system, we utilize a cumulative curves methodology [13], [24]. This model applies to a general setting where data could arrive in packets (packetized model) or in a continuum of bits (fluid model). Let $A(t)$, $D(t)$ and $D_{min}(t)$ denote the arrival curve, departure curve and the minimum departure curve respectively; these are assumed to be right-continuous functions and are defined as follows.

Definition 1: (Arrival Curve) An arrival curve $A(t)$, $t \geq 0$, $t \in \mathbb{R}$, is the total number of bits that have arrived in time interval $[0, t]$.

Definition 2: (Departure Curve) A departure curve $D(t)$, $t \geq 0$, $t \in \mathbb{R}$, is the total number of bits that have departed (served) in time $[0, t]$.

In case of a fluid arrival model, $A(t)$ is a continuous function, whereas, for a packet arrival model it is a piecewise-constant function as depicted in Figure 1. To ensure that the transmitter does not transmit more than the data that has arrived to the queue, we require that $D(t) \leq A(t)$. We refer to this constraint as the *causality constraint*. Now, to model the quality-of-service constraints we introduce a new notion of a “minimum departure curve” which is defined as follows.

Definition 3: (Minimum Departure Curve) Given an arrival curve $A(t)$, a minimum departure curve $D_{min}(t)$ is a function such that $D_{min}(t) \leq A(t)$, $\forall t \geq 0$, and is defined as the cumulative minimum number of bits that if departed by time t would satisfy the quality-of-service requirements.

The function $D_{min}(t)$ can be viewed as the constraint function, such that in order to satisfy the QoS requirements the departure curve $D(t)$ must satisfy $D(t) \geq D_{min}(t)$. Thus, in a compact way the QoS and the causality constraints can be expressed as, $D_{min}(t) \leq D(t) \leq A(t)$, $\forall t$. Note that the definition of $D_{min}(t)$ hides the implicitly assumed service discipline (the order in which data is served), as the above model looks at the data flow only in a cumulative sense. As we show next, through a few illustrative examples, that a number

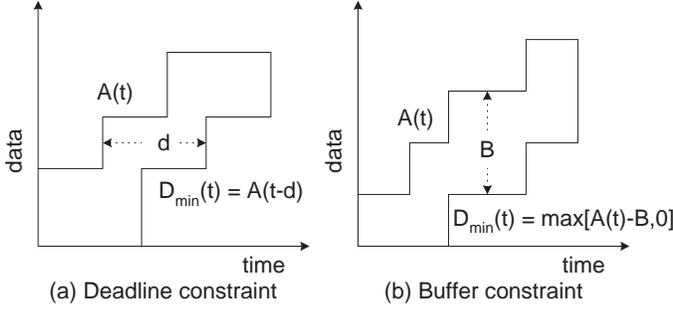


Fig. 2. QoS Examples: (a) Packet deadline constraint of d , (b) Buffer constraint of B .

of commonly used QoS constraints can be modelled within this framework.

Deadline Constraint: Consider a set of packet arrivals according to an arrival curve $A(t)$ and let d be the individual deadline constraint on the incoming data. To obtain $D_{min}(t)$ set, $D_{min}(t) = 0, t \in [0, d]$ and $D_{min}(t) = A(t - d), t \geq d$; now, following an earliest-deadline-first service discipline such that the departure curve satisfies $D_{min}(t) \leq D(t) \leq A(t), \forall t$, it is easy to see that the deadline constraints will all be met. Thus, here, $D_{min}(t)$ is simply a time-shifted version of $A(t)$ as shown in Figure 2(a). As a generalization, suppose that the data has variable deadlines and these deadlines are in the increasing order in which the bits arrive. Consider a packet arrival model and let $\{t_i\}$ denote the arrival epochs, $\{d_i\}$ the deadlines and $\{b_i\}$ the sizes of the data packets. Then, $D_{min}(t)$ is a piecewise constant function with jumps at times $\{t_i + d_i\}$ and the sizes of the jumps being $\{b_i\}$. Along similar lines as above, one can also obtain $D_{min}(t)$ for a fluid arrival model.

Buffer Constraint: Consider a buffer constraint of B , i.e. the queue size must not exceed $B, \forall t \geq 0$. For an arrival curve $A(t)$ and a departure curve $D(t)$ the buffer size at any time t is given by $b(t) = A(t) - D(t)$. Since $b(t) \leq B$, we have $D(t) \geq \max[A(t) - B, 0]$. Thus, we see that, following a first-come-first-serve service discipline, the minimum departure curve must be $D_{min}(t) = \max[A(t) - B, 0]$ as shown in Figure 2(b). It is easy to incorporate a time varying buffer constraint $B(t)$ as well.

Service-Curve Constraint: The notion of service curves forms an integral part of network calculus theory [24]. Given a service curve $\beta(t)$ and an arrival curve $A(t)$, the quantity $A(t) \otimes \beta(t)$ represents the minimum data that must flow out of the system, where \otimes is convolution in the min-plus algebra. Therefore, under network calculus theory, given any service curve $\beta(t)$, the minimum departure curve can be obtained as $D_{min}(t) = A(t) \otimes \beta(t)$.

Thus, we see that a wide variety of QoS constraints can be abstracted by constructing the appropriate minimum departure curve.

B. Transmission Model

Let $P(t)$ denote the required transmission power to reliably transmit at rate $r(t)$ at time t . We assume the following power-

rate relationship,

$$P(t) = g(r(t), t) \quad (1)$$

where the function $g(r, t)$ is a convex increasing function with respect to the first argument (rate) and $g(r, t) \geq 0$ for $r \geq 0, \forall t$. The relationship in (1) is a general transmission model for most encoding schemes and has been widely studied in the literature in various forms [7]–[12], [14], [15]. As an example, the well-known Shannon formula for the power per gives the following relationship [8], $P = N_0 W (2^{r/W} - 1)$; in case of other coding schemes the Shannon formula gives a lower bound on the power per bit.

Given the relationship in (1), the transmission energy expenditure of a departure curve $D(t)$ over time interval $[0, T]$ is given by,

$$\mathcal{E}(D(t)) = \int_0^T g(D'(t), t) dt \quad (2)$$

where $D'(t)$ is the derivative¹ at time t ; it gives the transmission rate at that instant and the term $g(D'(t), t)$ gives the instantaneous transmission power.

Throughout the paper, our focus will be on the time interval $[0, T]$ for some finite T , and with finite deadline/QoS constraints. Thus, we deal with energy minimization over a finite time interval rather than considering an infinite time horizon, as done in much of the literature on power-rate adaptation which studies average performance metrics. Since a departure curve specifies the transmission rate and vice-versa, we will use the terms departure curve and transmission policy interchangeably.

III. TIME-INVARIANT POWER-RATE FUNCTION

We first consider the case of a time-invariant power-rate function and assume in this section that $P(t)$ is only a function of $r(t)$, i.e. $P(t) = g(r(t))$. Such an assumption models a static channel or a slow fading wireless channel where over $[0, T]$ the channel gain does not change appreciably over time. This is a good model for wireless LAN or fixed wireless network scenarios.

A. Problem Formulation

Consider an arrival curve $A(t)$ and assume that this curve is known over the interval $[0, T]$. Based on the QoS requirements, one can construct the minimum departure curve $D_{min}(t)$ as discussed in Section II. Now given $A(t)$ and $D_{min}(t)$, a departure curve $D(t)$ which represents how data is transmitted is said to be *admissible* if it satisfies both the causality and the QoS constraints; i.e. $D_{min}(t) \leq D(t) \leq A(t), t \in [0, T]$. The energy minimization problem is to obtain the admissible departure curve with the least energy expenditure. Mathematically, this can be stated as follows,

$$\begin{aligned} \min_{D(t) \in \Gamma} \quad & \mathcal{E}(D(t)) = \int_0^T g(D'(t)) dt \quad (3) \\ \text{subject to} \quad & D_{min}(t) \leq D(t) \leq A(t), t \in [0, T] \end{aligned}$$

¹Throughout the paper, at points of non-differentiability, $D'(t)$ is taken as the right-derivative, and the right-derivative is assumed to exist for all t .

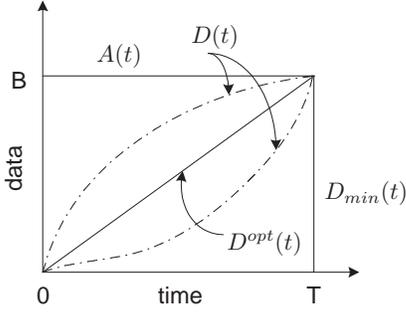


Fig. 3. Cumulative curves for the BT -problem.

Without loss of generality, we take $D_{min}(0) = 0, D(0) = 0$, and $D_{min}(T) = A(T)$, where the last equality simply states that all the data must depart by T . For the above problem, we also require that $D(t)$ belongs to the set Γ , where Γ consists of all non-decreasing, continuous functions with bounded right-derivative for all $t \in [0, T]$ and with $D(0) = 0$. The non-decreasing assumption follows from the cumulative nature of the departure curves, the continuity assumption is natural as any discontinuity would imply instantaneous transmission of non-zero amount of data which is practically infeasible and finally, the bounded right-derivative assumption ensures that the rate and the energy cost in (3) are finite. Furthermore, if one makes the natural assumption that there is no data that arrives and needs to be transmitted instantaneously, then, admissible departure curves exist.

B. Optimality Properties

In this section, we present the optimality criterion and the various properties of the optimal departure curve. To motivate the discussion, consider the following simple example: the transmitter has B units of data that must be transmitted by deadline T using minimum energy. We refer to this as the “ BT -problem”. This example sheds important insights into the problem and also serves as a building block for the general problem.

BT -problem: The two curves $A(t)$ and $D_{min}(t)$ for this problem are as follows. Since there are no new arrivals and the queue has B units of data at the beginning, the arrival curve is $A(t) = B, \forall t \in [0, T]$. Furthermore, there is no minimum data transmission requirement until the deadline T , at which point all the data must be transmitted; hence, $D_{min}(t) = 0, t \in [0, T]$ and $D_{min}(T) = B$. The admissibility criterion specialized to this case thus becomes $0 \leq D(t) \leq B$ and $D(T) = B$. Figure 3 is a schematic diagram of these curves which also depicts a few admissible departure curves.

We claim that the optimal policy is constant rate transmission at rate B/T , i.e. $(D^{opt})'(t) = \frac{B}{T}$ and $D^{opt}(t) = \frac{Bt}{T}, t \in [0, T]$, where $D^{opt}(t)$ denotes the optimal departure curve. To see why this is true consider the following integral version of Jensen’s inequality [27].

Lemma 1: Let $f(t), p(t)$ be two functions defined for $a \leq t \leq b$ such that $\alpha \leq f(t) \leq \beta$ and $p(t) > 0$, with $p(t) \not\equiv 0$. Let $\phi(u)$ be a convex function defined on the interval $\alpha \leq u \leq \beta$;

then

$$\phi \left(\frac{\int_a^b f(t)p(t)dt}{\int_a^b p(t)dt} \right) \leq \frac{\int_a^b \phi(f(t))p(t)dt}{\int_a^b p(t)dt} \quad (4)$$

with strict inequality if $\phi(\cdot)$ is strictly convex and $a \neq b, \alpha \neq \beta$.

Proof: See [27]. \blacksquare

Now, consider an admissible departure curve $D(t)$ and make the following substitution in the above lemma, $p(t) = 1, \phi(\cdot) = g(\cdot), f(\cdot) = D'(\cdot), a = 0$ and $b = T$. This gives,

$$g \left(\frac{\int_0^T D'(t)dt}{\int_0^T dt} \right) \leq \frac{\int_0^T g(D'(t))dt}{\int_0^T dt} \quad (5)$$

$$g \left(\frac{D(T) - D(0)}{T} \right) T \leq \int_0^T g(D'(t))dt \quad (6)$$

$$g(B/T)T \leq \int_0^T g(D'(t))dt \quad (7)$$

The left hand side in (7) is the total energy cost of the constant rate transmission policy at rate B/T , while, the right hand side is the total cost of any other admissible departure curve. The inequality in (7) thus proves the optimality claim.

The result for the BT -problem is fairly intuitive given the convexity property of the power-rate function. Its practical implication is interesting as it says that employing a complex variable-rate policy does not provide any gains in the energy expenditure; in fact, a constant-rate policy suffices. Another observation is that when $g(\cdot)$ is strictly convex, the inequality in (7) is strict for any admissible $D(t)$ other than the constant-rate policy. Hence, in this case the constant-rate policy is the unique optimal solution. On the other hand, for the case when $g(\cdot)$ is linear, there is equality in (7) and all policies have the same energy cost.

General Case: We now consider the general setup and assume without loss of generality that $A(t) > D_{min}(t), 0 < t < T$. Otherwise, if at some time t_e there is equality, the problem can be divided into two sub-problems over time intervals $[0, t_e]$ and $[t_e, T]$ and each can be solved independently. The first result, Theorem I, is a generalization of the result for the BT -problem and it gives the criterion for the optimality of a departure curve.

Theorem 1: (Optimality Criterion) Let $D(t)$ be an admissible departure curve and $L(t)$ be a straight line segment over $[a, b]$ that joins points $D(a)$ and $D(b), 0 \leq a < b \leq T$. If $L(t)$ satisfies $D_{min}(t) \leq L(t) \leq A(t)$, and, $L(t) \not\equiv D(t)$, the new departure curve $D^{new}(t)$ constructed as,

$$\begin{aligned} D^{new}(t) &= D(t), \quad t \in [0, a) \\ &= L(t), \quad t \in [a, b] \\ &= D(t), \quad t \in (b, T] \end{aligned}$$

satisfies, $\mathcal{E}(D^{new}(t)) \leq \mathcal{E}(D(t))$, where the inequality is strict if $g(\cdot)$ is strictly convex.

The above theorem states that, if there exists any two points on the curve $D(t)$ that can be joined by a straight line without violating the admissibility constraints, replacing that part of $D(t)$ with the straight line can only lower the energy cost. The implication of this is that whenever admissible, it is optimal

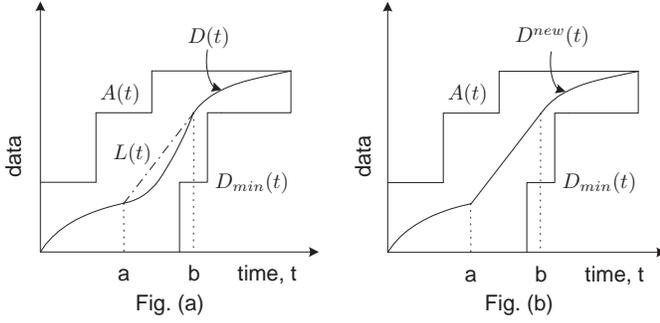


Fig. 4. Figure for Theorem I: (a) an admissible departure curve $D(t)$ and (b) the new curve $D^{new}(t)$.

to transmit at a constant rate. A schematic diagram depicting this is given in Figure 4. Henceforth, the criterion that along a departure curve there does not exist any two points that can be joined by a distinct admissible straight line will be referred to as the “Optimality Criterion”.

Proof: First note that since $L(t)$ is admissible, the new curve $D^{new}(t)$ is also admissible. Consider,

$$\mathcal{E}(D^{new}(t)) - \mathcal{E}(D(t)) = \mathcal{E}(L(t)) - \int_a^b g(D'(t))dt \quad (8)$$

Over the interval $[a, b]$, we know from the result of the BT-problem that $L(t)$ has the least energy cost among all departure curves that would transmit $(D(b) - D(a))$ amount of data in time $(b - a)$. Hence, from (5)-(7), we get, $\mathcal{E}(L(t)) - \int_a^b g(D'(t))dt \leq 0$ and the result follows. ■

Remark 1: (Linear power-rate function) An interesting special case arises when the power-rate relationship is linear, i.e. $P = \kappa r$ where $\kappa > 0$ is a constant. In this case, the integral value in (3) is the same for all admissible departure curves and hence, all departure curves have the same energy cost. Thus, with a linear power-rate curve it does not matter, in terms of the energy cost, how data is transmitted as long as the causality and the QoS constraints are met. However, even in this special case of linear power-rate function, we will see next that the departure curve that satisfies the optimality criterion has appealing properties that make it a good candidate transmission policy.

Henceforth, we consider the more interesting case of strictly convex $g(\cdot)$ function. The next result shows that the optimal departure satisfying the optimality criterion is unique.

Theorem II: (Uniqueness) Consider the optimization problem in (3) with $g(\cdot)$ being strictly convex. Let $\tilde{D}(t)$ be an admissible departure curve that satisfies the optimality criterion, then, $\tilde{D}(t)$ is unique and it minimizes the energy cost in (3).

Proof: See Appendix A. ■

Throughout now, we will denote the admissible departure curve satisfying the optimality criterion as $D^{opt}(t)$ and later in Section III-C, give an algorithm for constructing $D^{opt}(t)$. We now characterize the points in time at which the optimal rate changes, i.e. points at which the slope (or the right-derivative where non-differentiable) of $D^{opt}(t)$ changes, either

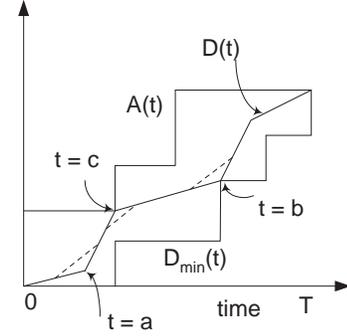


Fig. 5. Example showing violation of Lemmas 2-4. The dotted lines show that $D(t)$ does not meet the optimality criterion.

continuously or in a discrete step. Denoting any such point as t_0 , the following results are obtained².

Lemma 2: At t_0 , $D^{opt}(t)$ either intersects $A(t)$ or it intersects $D_{min}(t)$; i.e. we have $D^{opt}(t_0) = A(t_0)$ or $D^{opt}(t_0) = D_{min}(t_0)$. Note, if there is a discontinuity in $A(t)$ at t_0 (jump point for packetized data) then $D^{opt}(t_0) = A(t_0^-)$.

Lemma 3: Suppose that at t_0 we have $D^{opt}(t_0) = D_{min}(t_0)$, then, the slope change must be negative.

Lemma 4: Suppose that at t_0 we have $D^{opt}(t_0) = A(t_0)$ (or $A(t_0^-)$) then the change in slope must be positive.

The proofs of the above lemmas are straightforward and omitted for brevity. They can be easily understood from Figure 5. Point $t = a$ corresponds to a point of rate change and it violates Lemma 2. It is easy to see that around $t = a$ the optimality criterion is violated since an admissible straight line segment exists (the dotted segment around $t = a$ in the figure). Similarly, points $t = b$ and $t = c$ correspond to a violation of Lemmas 3 and 4 respectively.

Among other properties, the optimal departure curve $D^{opt}(t)$ uses the least maximum transmission-power and has the shortest length metric. The minimal maximum-power property of $D^{opt}(t)$ states that among all admissible departure curves, if we look at the maximum instantaneous power requirement over time, $D^{opt}(t)$ curve has the least such requirement. This is summarized in the theorem below.

Theorem III: (Minimal Maximum Power) Given any admissible departure curve $D(t)$, the optimal departure curve $D^{opt}(t)$ satisfies,

$$\max_{t \in [0, T]} (D^{opt})'(t) \leq \max_{t \in [0, T]} D'(t) \quad (9)$$

Equivalently, $\max_{t \in [0, T]} P^{opt}(t) \leq \max_{t \in [0, T]} P(t)$, where $P(\cdot)$ denotes the power expenditure over time.

Proof: See Appendix B. ■

Remark 2: The above theorem is very significant if we impose an additional maximum power constraint in the optimization problem in (3). In this case, the problem is first solved without the power constraint. If the optimal solution

²The notation $f(x^+)$ means $\lim_{n \rightarrow \infty} f(x + \epsilon_n)$ and $f(x^-)$ means $\lim_{n \rightarrow \infty} f(x - \epsilon_n)$ with $\epsilon_n > 0, \epsilon_n \rightarrow 0$.

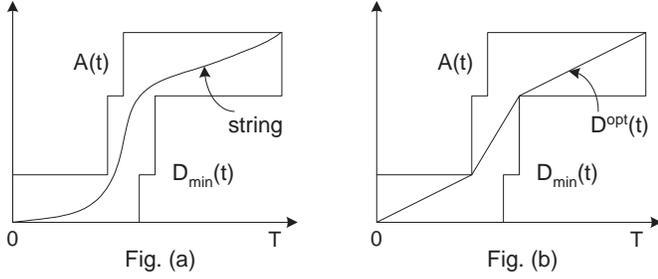


Fig. 6. String visualization for the optimal curve, (a) string lying between $A(t)$ and $D_{min}(t)$; (b) $D^{opt}(t)$ as taut string.

satisfies the maximum power constraint, we are done; otherwise from Theorem III it follows that there does not exist any other admissible departure curve that can satisfy the power constraint and the constrained optimization problem has no solution. Thus, we see that $D^{opt}(t)$ is the unique curve that satisfies the QoS constraints with both the least total energy cost and the least maximum power requirement.

As mentioned earlier, $D^{opt}(t)$ also has the shortest length among admissible departure curves. More specifically, for any continuous, piecewise differentiable curve, its total length using standard geometrical result is given as $\int_0^T \sqrt{1 + (D'(t))^2} dt$. The result below states that $D^{opt}(t)$ minimizes this metric.

Theorem IV: (Shortest Length) *The optimal departure curve $D^{opt}(t)$ has the shortest length among all admissible departure curves. Specifically, it minimizes the metric,*

$$\text{len}(D(t)) \triangleq \int_0^T \sqrt{1 + (D'(t))^2} dt \quad (10)$$

Proof: Since $D^{opt}(t)$ minimizes the integral in (3) for a convex increasing function $g(\cdot)$, the result follows by replacing $g(r)$ with $g(r) = \sqrt{1 + r^2}$. ■

C. Optimal Policy

In the last section, we presented the optimality criterion and the various properties of the optimal curve. We now construct the optimal departure curve $D^{opt}(t)$. However, before giving the algorithmic description, it is instructive to consider a very insightful visualization. This graphical picture provides a simple and intuitive way to understand $D^{opt}(t)$ and is described next.

String Visualization: *Consider a string restricted to lie between $A(t)$ and $D_{min}(t)$ (i.e. visualize $A(t), D_{min}(t)$ curves as hard boundaries for the string). Tie one end of the string at the origin and pass the other end through $D_{min}(T)$. If we now make the string tight, its trajectory gives the optimal departure curve³.*

Figure 6 is an illustration showing a general $A(t)$ and $D_{min}(t)$ curve and the corresponding $D^{opt}(t)$ visualized as a tight string. Intuitively, when the string is in the tight condition it cannot be made tighter between any two points along its

³This observation was pointed out by Rene L. Cruz

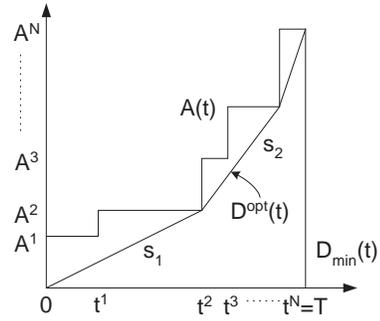


Fig. 7. Curves $A(t), D_{min}(t)$ and $D^{opt}(t)$ for Example 1.

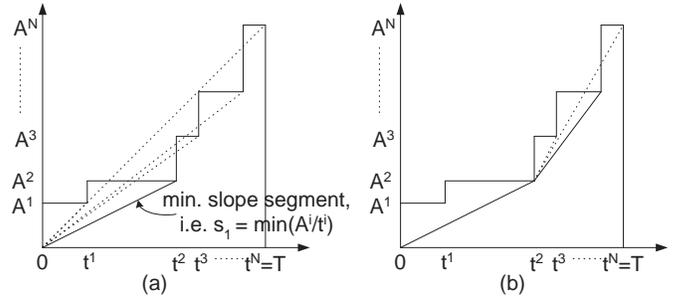


Fig. 8. (a) Construction of the first segment of $D^{opt}(t)$ for Example 1, (b) Construction of the second segment.

trajectory. This means that the optimality criterion must be satisfied, because otherwise, the construction in Theorem I would make the string tighter, thereby leading to a contradiction. By the uniqueness result, it then follows that this must be the optimal curve. Note that depending on the shape of $A(t)$ and $D_{min}(t)$ curves, the curve $D^{opt}(t)$ consists of segments of constant-rate transmission and/or segments where the rate is varying continuously over time; see for example Figure 10(b), where over time $[a, b]$ and $[c, d]$ the curve $D^{opt}(t)$ has a continuous rate change.

Examples: Using the above string visualization, we now present a few illustrative examples for which the optimal solution can be obtained in closed-form. Among these, the first two examples have been studied earlier in the literature [8], [10], and their solutions were obtained using a discrete-optimization approach that was mathematically tedious. By reformulating the problems within our framework, the solutions can be obtained more easily from the graphical picture.

Example 1: [8]: Consider a sequence of N packets arriving to the system in time $[0, T)$ with the first packet arrival at time 0 and the rest arriving at times $\{t^i\}_{i=1}^{N-1}$. The deadline constraint is that all the packets must depart by time T (common deadline), where $T > t^{N-1}$. The curves $A(t)$ and $D_{min}(t)$ for this problem are depicted in Figure 7 (where we have set $t^N = T$). From the string visualization it is easy to see that the optimal departure curve consists of piecewise linear segments with increasing slopes and the points at which the slope changes, the optimal policy just empties the buffer (see Fig.7).

Algorithmically, to construct the linear segments of $D^{opt}(t)$

proceed as follows. Let A^i denote the cumulative amount of data arrived to the queue just before time t^i (the total data in the first i packets). Starting at time 0, consider the straight line segments that join the points $(0, 0)$ (origin) and (t^i, A^i) (jump points of $A(t)$). From among these choose the segment with the minimum slope, i.e. the segment having slope equal to the minimum over i of $\left(\frac{A^i}{t^i}\right)$, as shown schematically in Figure 8(a). Denoting the minimizing index as π , the first segment of $D^{opt}(t)$ is then constant-rate transmission with rate $s_1 = \min_i \left(\frac{A^i}{t^i}\right)$, from $t = 0$ until $t = t^\pi$. Starting at t^π , the procedure is repeated by shifting the origin to the new point (t^π, A^π) , as shown in Figure 8(b). Thus, the slopes of the linear segments denoted as $\{s_1, \dots, s_q\}$ can be computed recursively as follows. Take $l_1 = 1, t^0 = 0, A^0 = 0$ and initialize $m = 1$; we then have,

$$s_m = \min_{i \in \{l_m, \dots, N\}} \left(\frac{A^i - A^{(l_m-1)}}{t^i - t^{(l_m-1)}} \right) \quad (11)$$

$$l_{m+1} = 1 + \arg \min_{i \in \{l_m, \dots, N\}} \left(\frac{A^i - A^{(l_m-1)}}{t^i - t^{(l_m-1)}} \right) \quad (12)$$

The above iteration stops when $l_{m+1} = N + 1$. Intuitively, the optimal policy follows a constant rate transmission until points where the future arrivals are such that relative to the deadline constraint, the transmission rate must be higher.

Example 2: [10]: Consider M data packets in the transmitter buffer at time 0, with individual packets having a deadline by which they must be transmitted. Let the j^{th} packet have b_j units of data and a deadline t_j , $j = 1, \dots, M$. The packets in the queue are served in the earliest-deadline-first order and for this case, the $A(t)$ and $D_{min}(t)$ curves can be obtained as shown in Figure 9(a). Note that the structure of this problem is the reverse of Example 1 and in some loose sense one can regard these problems as “duals” of each other. From the string interpretation, we see that the optimal policy is a piecewise linear curve as shown in the figure, and as compared to Example 1, the slopes of the linear segments are now monotonically decreasing in time.

To obtain the segments of $D^{opt}(t)$ proceed as follows. Let $B_j = \sum_{l=1}^j b_l$, where B_j denotes the cumulative data in the first j packets. Starting at time 0, consider the straight line segments that join the points $(0, 0)$ (origin) and (t_j, B_j) (jump points of $D_{min}(t)$). From among these choose the segment with the maximum slope, i.e. the segment having slope equal to the maximum over j of $\left(\frac{B_j}{t_j}\right)$. Denoting the maximizing index as π , the first segment of $D^{opt}(t)$ is constant-rate transmission with rate $\frac{B_\pi}{t_\pi}$ from $t = 0$ until $t = t_\pi$. Starting at t_π , the procedure is repeated by shifting the origin to the new point (t_π, B_π) . Specifically, the slopes denoted as $\{s_1, \dots, s_q\}$ are obtained as follows. Take $l_1 = 1, t_0 = 0, B_0 = 0$ and initialize $m = 1$, we then have,

$$s_m = \max_{j \in \{l_m, \dots, M\}} \left(\frac{B_j - B_{(l_m-1)}}{t_j - t_{(l_m-1)}} \right) \quad (13)$$

$$l_{m+1} = 1 + \arg \max_{j \in \{l_m, \dots, M\}} \left(\frac{B_j - B_{(l_m-1)}}{t_j - t_{(l_m-1)}} \right) \quad (14)$$

The above iteration stops when $l_{m+1} = M + 1$.

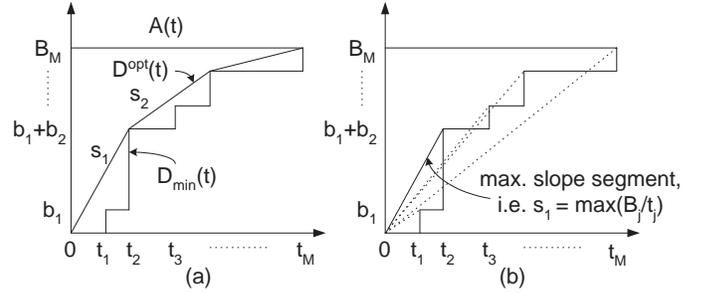


Fig. 9. (a) Curves $A(t), D_{min}(t)$ and $D^{opt}(t)$ for Example 2, (b) Construction of the first segment of $D^{opt}(t)$.

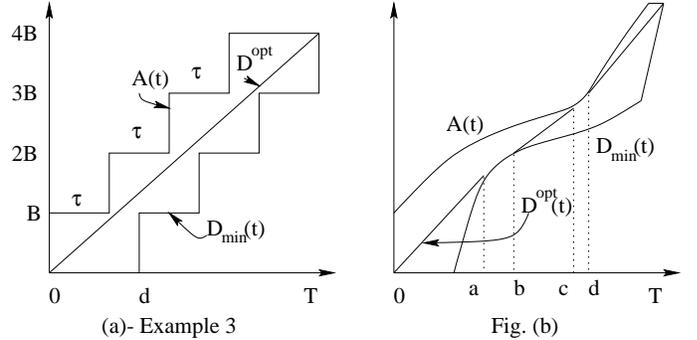


Fig. 10. Curves $A(t), D_{min}(t)$ and $D^{opt}(t)$ for (a) Example 3 and (b) Continuous data flow.

Example 3: Consider a stream of N packet arrivals of size B units with a constant inter-arrival time τ . Each packet has a deadline d before which it must depart (Figure 10(a)). Such an arrival stream is a good model for applications which generate packets at regular times (or with a small variance), e.g. voice data. The optimal minimum energy curve is shown in the figure and is given as follows. If $d < \tau$, the solution is trivial and the packet must be transmitted before the next arrival. If $d \geq \tau$, the optimal curve is a straight line with slope $NB/(d + (N - 1)\tau)$.

The intuition gained from the examples above can now be utilized to obtain $D^{opt}(t)$ for the general setting and this is presented next. For simplicity, however, we restrict our attention to only piecewise-constant $A(t)$ and $D_{min}(t)$ curves (i.e. staircase functions corresponding to the packet data model). The algorithm for the more general case with continuous curves is a direct extension of the arguments presented here and can be found in [5].

Construction of the Optimal Departure Curve: As is the case in Examples 1 and 2, the main idea behind constructing the optimal curve $D^{opt}(t)$ is to obtain its segments in a recursive fashion. From Example 1, we see that with $A(t)$ constraints the minimum-slope line segments are chosen; while from Example 2, we see that with $D_{min}(t)$ constraints the maximum-slope line segments are chosen. Thus, intuitively, in the general case we would need to combine these two ideas and this is done more formally in the discussion below.

To proceed, consider any generic point (t_0, α) , where $0 \leq t_0 < T$ and $D_{min}(t_0) \leq \alpha \leq A(t_0)$. Starting at this point,

consider straight lines with non-negative slopes. Among these, choose those lines that starting at (t_0, α) remain admissible for some finite duration. In other words, consider straight lines $L(t)$ for which there exists an $\epsilon > 0$ (ϵ could depend on the chosen $L(t)$), such that $L(t)$ is admissible for $t \in [t_0, t_0 + \epsilon)$, i.e. $D_{min}(t) \leq L(t) \leq A(t)$, for $t \in [t_0, t_0 + \epsilon)$. Denote this set as \mathcal{F} . Intuitively, the slopes of the lines in \mathcal{F} give the possible admissible slopes that $D^{opt}(t)$ can have at that point. Note that the set \mathcal{F} depends on the point (t_0, α) but to make the notations simple we drop the explicit dependence.

Now, consider $L(t) \in \mathcal{F}$, then clearly, $L(t)$ eventually either intersects $A(t)$ or it intersects $D_{min}(t)$, where we use the following definition of *intersection*.

Definition 4: Starting at t_0 , $L(t)$ intersects $D_{min}(t)$ if for some point $\tilde{t} > t_0$, called the point of intersection, one of the following holds: (a) either $L(\tilde{t}) = D_{min}(\tilde{t})$ or, (b) the function $L(t) - D_{min}(t)$ changes sign at \tilde{t} (here \tilde{t} is a discontinuity point).

Intuitively, the above definition means that $L(t)$ crosses the curve $D_{min}(t)$ at \tilde{t} . A similar definition holds for intersection with $A(t)$. Let the set \mathcal{F} be partitioned into a set of lines that intersect $A(t)$ first and those that intersect $D_{min}(t)$ first. Denote these sets as \mathcal{F}_A and \mathcal{F}_{D_m} respectively. The following intuitive result states that the slope of the lines in \mathcal{F}_A (those that intersect $A(t)$ first) is greater than the slope of the lines in \mathcal{F}_{D_m} (those that intersect $D_{min}(t)$ first).

Lemma 5: (a) Let $L_D(t) \in \mathcal{F}_{D_m}$ then any $L(t) \in \mathcal{F}$ that has slope less than L'_D intersects $D_{min}(t)$ first. (b) Let $L_A(t) \in \mathcal{F}_A$ then any $L(t) \in \mathcal{F}$ that has slope greater than L'_A intersects $A(t)$ first.

Proof: See Appendix C. ■

Let \mathcal{S}_A and \mathcal{S}_{D_m} denote the slopes of the lines in \mathcal{F}_A and \mathcal{F}_{D_m} respectively. Consider the line, which we denote as L_o , with slope β_o at the boundary of the two intervals, i.e.,

$$\beta_o = \inf \mathcal{S}_A = \sup \mathcal{S}_{D_m} \quad (15)$$

If either \mathcal{S}_A or \mathcal{S}_{D_m} is empty above, it is neglected. We call β_o the *optimal slope* and the line L_o the *optimal line*. Thus, in simple terms, L_o is the least slope line that intersects $A(t)$ first, or the maximum slope line that intersects $D_{min}(t)$ first (note the similarities with Examples 1 and 2). Using this line L_o , we can now obtain an algorithm for constructing the optimal departure curve as illustrated next.

To begin with, we have $D^{opt}(0) = 0$; thus, the starting point is $(0, 0)$. Let t_0 denote a generic time instant, where $t_0 = 0$ in the first iteration.

- 1) Obtain β_o as in (15) and the optimal line L_o .
- 2) Obtain the first instant t_1 such that, (a) $L_o(t_1) = D_{min}(t_1)$, or, (b) $L_o(t_1) = A(t_1)$ or $L_o(t_1) = A(t_1^-)$. Set $D^{opt}(t) = L_o(t)$, $t \in (t_0, t_1]$.

If $t_1 = T$ terminate; else repeat the above steps with the new starting point as $(t_1, D^{opt}(t_1))$. The correctness and optimality of the above algorithm is shown in Appendix D.

As an example, consider $A(t)$ and $D_{min}(t)$ shown in Figure 11 for which the algorithm executes as follows. Start at the origin $(0, 0)$ and note that L_1 is the optimal line as defined above and t_1 is the first instant at which it equals

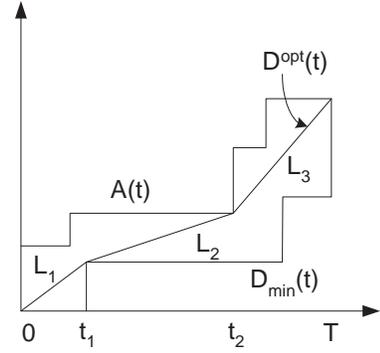


Fig. 11. Example depicting $A(t)$ and $D_{min}(t)$ curves, and the constructed $D(t)$ curve.

$D_{min}(t)$. Thus, segment L_1 from $t = [0, t_1]$ is the first part of the optimal curve. Note that lines with slope greater than L'_1 intersect $A(t)$ first and lines with slope less than L'_1 intersect $D_{min}(t)$ first. The line L_1 is the one with slope at the boundary (as defined in (15)). Next, starting from the new point $(t_1, D_{min}(t_1))$, L_2 is the optimal line and t_2 is the first instant such that $L_2(t_2) = A(t_2^-)$. The segment L_2 from $t = [t_1, t_2]$ forms part of the optimal curve. The segment L_3 is also obtained in a similar fashion and it is the last segment as $t = T$ is reached.

D. Online Policy Without Arrival Information

In the previous sections, we obtained a fundamental understanding of the energy minimization problem by assuming that the data arrival information was known in advance. In this section, we utilize those results to consider the more realistic case when there is an arbitrary stream of packet arrivals to the queue and there is no information, statistical or otherwise, of the packet arrival process. Each arriving packet has a distinct deadline by which it must be served, and the goal as before is to minimize the total energy expenditure. To address this problem, we present an online transmission policy, referred to as the “*Backlog-Adaptive*” (BA) policy, and give numerical results comparing the energy cost of the BA policy with the head-of-line drain policy.

To understand the BA policy, let us first revisit Example 2 in Section III-C which we summarize here again. Suppose that the transmitter has M packets with individual deadlines on the packets; there are no new arrivals to the system and the goal is to empty the buffer with minimum energy. The optimal policy for this case is shown in Figure 9, but to highlight the dynamic nature of the policy and for computational simplicity we rephrase it as follows. Denote the state of the system as (t, D) where the notation means that at time t , the cumulative amount of data that has been transmitted is D , i.e. $D(t) = D$. Assuming an admissible system state, i.e. $D_{min}(t) \leq D \leq A(t)$ and $t < T$, the optimal transmission rate for this state is obtained as follows. First, visualize the origin at point (t, D) , then, it is easy to see that the optimal rate is the maximum value among the slopes $\left(\frac{B_j - D}{t_j - t}\right)$, corresponding to the straight line segments that connect the points (t, D) and (t_j, B_j) for

all $\{j : B_j \geq D, t_j \geq t\}$. Specifically, let $r^*(t, D)$ denote the optimal rate, we then have,

$$r^*(t, D) = \max_{j: (B_j \geq D, t_j \geq t)} \frac{B_j - D}{t_j - t} \quad (16)$$

The above function is an alternate way to state the optimal policy shown in Figure 9 for Example 2; it provides a convenient way for implementation. The transmitter simply keeps track of the cumulative amount of data that has been transmitted and at time t , it computes the rate at that instant as given in (16) by a simple max operation. Note that the policy in (16) applies for a static buffer that already contains packets with deadlines; we now extend it to incorporate packet arrivals to the queue that are unknown in advance.

Consider arbitrary packet arrivals to the queue with each packet having a distinct deadline associated with it. Assume that the arrivals occur at discrete time instances. Clearly, at the instant immediately following an arrival, the transmitter queue consists of (a) earlier remaining packets with their deadlines and (b) the new packet with its own deadline. Re-arranging the data in the earliest-deadline-first order we can view the queue as consisting of a total amount B of data with variable deadlines. This is identical to the problem mentioned earlier of emptying the data in the buffer with minimum energy, hence, we can use the transmission policy given in (16). Now, as this policy is followed, at the next packet arrival instance we simply repeat the above procedure by rearranging the data and taking the new packet into account. We refer to the above policy as the Backlog-Adaptive (BA) policy and it can be summarized as follows,

BA-policy: *Transmit the data in the queue with the rate as given in (16); at every packet arrival instant re-arrange the data in the earliest-deadline-first order to obtain a new set of $B_j t_j$ values by including the new packet and its deadline; re-initialize D and t to zero and follow (16) thereafter.*

Note that the BA policy is not based on any specific arrival process, hence it is robust to changes in the arrival statistics and can even accommodate multiple deadline classes of packet arrivals to the queue.

E. Simulation Results

In this section, we present illustrative simulation results comparing the performance of the Backlog-Adaptive policy with the ‘‘Head-of-Line Drain’’ (HLD) policy. In HLD policy, the data in the queue is arranged in the earliest-deadline-first order and the packets are served in that order. At time t , let H_t be the amount of data left in the head-of-the-line packet and T_H be the amount of time until its deadline, then, under HLD policy the rate is chosen as $r_t = \frac{H_t}{T_H}$. Thus, the transmitter serves the first packet in queue at a rate to transmit it out by its deadline, then moves to the next packet in line and so on. At every packet arrival instant, the data in the queue is re-arranged in the earliest-deadline-first order and the above policy is repeated with the new packet taken into account.

The simulation setup is as follows. The transmitter has Poisson packet arrivals and each packet has a deadline associated with it. On each simulation run, the total time over which the

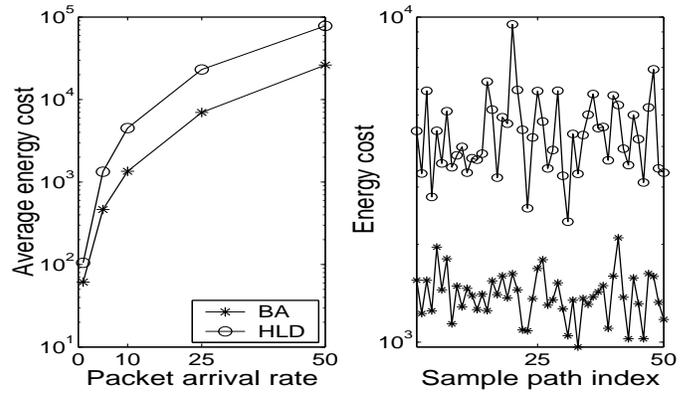


Fig. 12. Energy cost comparison for Poisson arrival process for (a) different arrival rate, (b) different sample paths.

packets arrive and the system is operated is taken as $L = 10$ seconds. This interval $[0, 10]$ is partitioned into 10,000 slots, thus each slot is of duration $dt = 1$ msec, and for simplicity, the packet arrivals take place at the slot boundaries. For both the BA and the HLD policies, the transmission rate chosen for a slot is obtained by evaluating the respective policies at the time corresponding to the start of that slot. We take $g(r) = r^2$, hence the energy cost per slot is $r^2 dt$. The total expected energy cost is obtained by taking an average of the total cost over multiple sample runs of the system.

We first consider the setup where each packet has 1 unit of data and a deadline of 200 msec. Figure 12(a) is a plot of the energy cost averaged over the sample paths and plotted on a logarithmic scale versus the packet arrival rate. As is evident from the plot, the BA policy has a much lower energy cost compared to the HLD policy and as the arrival rate increases the difference between the two increases. This can be intuitively explained as follows. When the arrival rate is low, most of the time the queue has at most a single packet, hence, both policies choose a rate based on the head-of-line packet. As the arrival rate increases and due to the bursty nature of the poisson process, the queue tends to have more packets. The BA policy then adapts the rate based on the backlog and the deadlines of all the packets in the queue, whereas, the HLD policy chooses a rate based solely on the head-of-line packet. In Figure 12(b), we set the arrival rate as 10 packets/second and plot the energy cost for the first 50 sample paths. As evident from the plot, the BA policy has lower energy cost not just in an average sense but even on most individual sample paths.

In Figure 13(a), we set the arrival rate as 10 packets/second and plot the average energy cost by varying the packet size. Clearly, as seen in the figure, the energy cost increases as the packet size increases since there is more data that needs to be transmitted, however, the BA policy has a much lower energy cost as compared to the HLD policy. In Figure 13(b), we plot the average energy cost by varying the packet deadlines and a similar trend is observed. The energy cost decreases as the packet deadline increases since lower transmission rates are required to meet the deadlines, and here as well, the BA policy has a significant lower energy cost as compared to HLD policy.

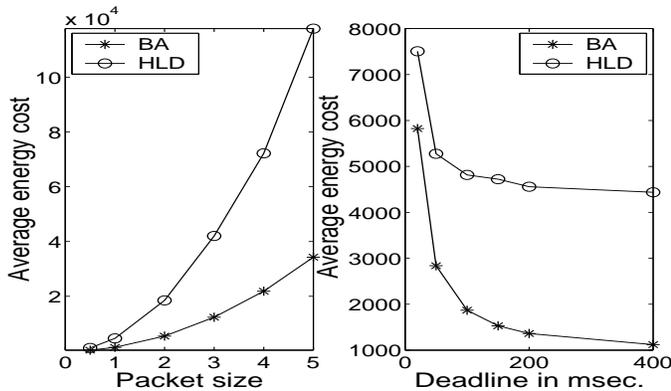


Fig. 13. (a) Average energy cost versus packet size and (b) Average energy cost versus packet deadline.

IV. TIME-VARYING POWER-RATE FUNCTION

In previous sections, we considered the time-invariant power-rate function case and utilized a cumulative curves methodology to obtain the optimal solution. The framework provided a graphical visualization of the problem and the optimal solution. In this section, we generalize those results and consider a time-varying power-rate function setup. Thus, now the function $P(t)$ has a time-varying dependence and is given as $P(t) = g(r(t), t)$. For a fixed time t_0 , the amount of power required to transmit at a certain rate r is governed by the convex function $g(\cdot, t_0)$, but now, this convex function could be different at different times.

A. Problem Formulation

The problem formulation remains the same as given in Section III-A with the data flows being described using cumulative curves and the objective is to obtain the minimum energy departure curve. The optimization problem is given as,

$$\begin{aligned} \min_{D(t) \in \Gamma} \quad & \mathcal{E}(D(t)) = \int_0^T g(D'(t), t) dt \\ \text{subject to} \quad & D_{\min}(t) \leq D(t) \leq A(t), \quad t \in [0, T] \end{aligned} \quad (17)$$

In the above formulation, we assume that $g(r, t)$ as a function of r is a strictly convex, increasing and continuously differentiable function for all t . We also assume that $g(r, t)$ is a deterministic function of time $t \in [0, T]$ and piecewise continuous in t .

The above formulation provides a general framework to model various scenarios involving time-variability in the system. It generalizes the problem in Section III-A to include time-dependent parameters in transmission arising due to phenomena such as beam-forming, antenna patterns etc. Since it models a more general power-rate cost function, one can also introduce an artificial cost for control purposes; for example, by imposing a high cost over certain intervals one can control the times over which data should be transmitted. Finally, it also models scenarios where we have a time-varying channel and the channel gain is predictable or known over time.

B. Optimality Properties

We proceed as in Section III by first considering the BT -problem and then extending the results to general $A(t)$ and $D_{\min}(t)$ curves. As in the time-invariant case, the BT -problem provides useful insights into the problem and also plays an important role as a building block.

1) *BT-problem*: Consider the BT -problem where the transmitter has B units of data in the queue and a deadline T by which this data must be transmitted using minimum energy. The following lemma gives the optimal solution for this problem; its proof is based on results from the theory of *Calculus of Variations* [29].

Lemma 6: *The optimal transmission rate $r^{opt}(t)$ for the BT -problem is given as,*

$$r^{opt}(t) = \max(0, r^*(t)) \quad (18)$$

where $r^*(t)$ is a unique positive value that satisfies $\frac{\partial}{\partial r} g(r, t)|_{r=r^*(t)} = k$ and k is a positive constant such that $\int_0^T r^{opt}(t) dt = B$.

Proof: See Appendix E. ■

Thus, we see that the optimal rate is such that the partial derivative of $g(r, t)$ with respect to r at the positive value $r^*(t)$ equals a constant k . The value of this constant is chosen such that the deadline constraint at T is met. We refer to the constant k as the “marginal cost” for the BT -problem. At any time t , if there exists a positive rate $r^*(t)$ for which the marginal cost is k , that rate is chosen as the transmission rate; otherwise, the transmission rate is 0 and no data is transmitted.

For positive transmission rate, since the marginal cost (or the first-derivative of $g(r, t)$ with respect to r) is the same for all t , it implies that infinitesimal changes in the rate would not change the total energy cost. This observation is intuitive, since otherwise, we could decrease the rate over the intervals when the marginal cost is high and correspondingly increase the rate over the intervals when the marginal cost is low; thereby, reducing the total energy cost and violating the optimality claim. Now, for all t such that $r^{opt}(t) = 0$ we must have $\frac{\partial}{\partial r} g(r, t)|_{r>0} > k$. This means that at all such times, the marginal cost is high and it is relatively costly to transmit the data, hence, the optimal policy chooses a zero rate.

As compared to the time-invariant power-rate function case, clearly, the optimal rate now is not constant over time; however, interestingly, the marginal cost is constant. Thus, the constant slope property from before translates here into a constant marginal cost property. As a check, if we remove the time-dependence in $g(r, t)$, then $r^*(t)$ would be a constant. This gives $r^{opt}(t) = r^*$ and from $\int_0^T r^{opt}(t) dt = B$, we get $r^* = \frac{B}{T}$. Thus, the optimal solution is constant-rate transmission in conformity with the result in Section III-B.

As concrete examples for illustration, we now specialize (18) to two specific forms of $g(r, t)$, namely, the *Monomial* class and the *Exponential* class of functions.

Example 4: (Monomial Class) Let $g(r, t) = \frac{r^n}{c(t)}$, $n > 1$, $c(t) > 0$, be the class of positive monomial functions with $c(t)$ representing the channel gain or the time-dependent

parameter. For any positive constant k , $\frac{\partial}{\partial r} \frac{r^n}{c(t)}|_{r=r^*(t)} = k$, gives,

$$r^*(t) = \left(\frac{kc(t)}{n} \right)^{\frac{1}{n-1}} \quad (19)$$

Since k and $c(t)$ are positive, we have $r^*(t) > 0, \forall t$, and from (18) we get $r^{opt}(t) = \left(\frac{kc(t)}{n} \right)^{\frac{1}{n-1}}$. The value of k such that the deadline constraint is met is obtained from, $\int_0^T r^{opt}(t) dt = B$, which gives, $k^{\frac{1}{n-1}} = \frac{B}{\gamma}$, where $\gamma = \int_0^T (c(t)/n)^{\frac{1}{n-1}} dt$. Substituting back in $r^{opt}(t)$ finally gives,

$$r^{opt}(t) = \frac{B}{\gamma} \left(\frac{c(t)}{n} \right)^{\frac{1}{n-1}} \quad (20)$$

Example 5: (Exponential Class) Let $g(r, t) = \frac{\alpha^r - 1}{c(t)}$, $\alpha > 1$, $c(t) > 0$, be the class of exponential functions with $c(t)$ being the time-dependent parameter. Note that taking $\alpha = 2$ and $c(t) = |h(t)|^2$, gives the Shannon formula for the power per bit. For the exponential case, $\frac{\partial g(r^*, t)}{\partial r} = \frac{\alpha^{r^*(t)} \ln(\alpha)}{c(t)} = k$, gives,

$$r^{opt}(t) = \max \left(0, \frac{\ln(k) - \ln(\ln(\alpha)/c(t))}{\ln(\alpha)} \right) \quad (21)$$

The value of k such that the deadline constraint is met is obtained from $\int_0^T \max \left(0, \frac{\ln(k) - \ln(\ln(\alpha)/c(t))}{\ln(\alpha)} \right) dt = B$.

Returning back to the solution in (18), we next show an interesting monotonicity property with respect to the marginal cost k . This is presented in the lemma below.

Lemma 7: Let $r^{opt}(t)$ be given by (18) for some $k \geq 0$ and $D^{opt}(t) = \int_0^t r^{opt}(s) ds$. Then, $D^{opt}(t)$ is monotonically non-decreasing in k , unique for a given value of k and zero throughout for $k = 0$. Furthermore, for $D^{opt}(T) = B > 0$, there is a unique positive value of k that achieves it.

Proof: See Appendix F ■

From the above lemma, we see that given B and T , a binary search would be sufficient to obtain the value k numerically.

2) **General Case:** Thus far, we have presented results for the *BT*-problem; these can now be generalized to the setup with general $A(t)$ and $D_{min}(t)$ curves. Theorem V below gives the optimality criterion for this case and is a generalization of Theorem I presented earlier. It states that, if there exists any two points on an admissible departure curve that can be replaced with a constant marginal-cost solution without violating the admissibility constraints, the new departure curve obtained will have a lower energy cost. The notation, “constant marginal-cost curve over time-interval $[a, b]$ between data-points $[B_1, B_2]$ ” will refer to the departure curve, $L(t)$, obtained using the solution in (18) as follows: $L(a) = B_1$, $L(t) = L(a) + \int_a^t r(s) ds$, $t \in [a, b]$, where $r(s) = \max(0, r^*(s))$ and marginal-cost k chosen such that $L(b) = B_2$. From Lemma 7, this value of k and the corresponding $L(t)$ are unique.

Theorem V: (Optimality Criterion) Let $D(t)$ be an admissible departure curve and $L(t)$ be the constant marginal-cost curve over time-interval $[a, b]$ between data-points

$[D(a), D(b)]$, $0 \leq a < b \leq T$. If $L(t)$ is admissible, i.e. $D_{min}(t) \leq L(t) \leq A(t)$, and, $L(t) \neq D(t)$, the new departure curve $\tilde{D}(t)$ constructed as,

$$\begin{aligned} \tilde{D}(t) &= D(t), \quad t \in [0, a) \\ &= L(t), \quad t \in [a, b] \\ &= D(t), \quad t \in (b, T] \end{aligned}$$

satisfies $\mathcal{E}(\tilde{D}(t)) \leq \mathcal{E}(D(t))$, where $\mathcal{E}(\cdot)$ is as given in (17).

Proof: First note that since $L(t)$ is admissible, the new curve $\tilde{D}(t)$ is also admissible. Consider,

$$\mathcal{E}(\tilde{D}(t)) - \mathcal{E}(D(t)) = \mathcal{E}(L(t)) - \int_a^b g(D'(t), t) dt \quad (22)$$

From Lemmas 6 and 7, we know that $L(t)$ is the unique curve, that has the least energy cost among all departure curves that would transmit $(D(b) - D(a))$ units of data over time interval $[a, b]$. Thus, $\mathcal{E}(L(t)) \leq \int_a^b g(D'(t), t) dt$, which completes the proof. ■

From the above theorem, we see that whenever admissible, segments of the optimal departure curve follow the constant marginal cost curve. This property translated into constant rate (straight line) segments in the time-invariant power-rate function case, as outlined earlier in Theorem I. Thus, we see that the pictorial representation and the properties from the time-invariant case apply here in terms of constant marginal costs. Lastly, as illustrative examples for the time-varying case, we re-visit Examples 1 and 2 in Section III-C and obtain the departure curve that satisfies the optimality criterion. The algorithms presented below are obtained by translating the respective ones from the time-invariant case, where instead of constant-slope segments we will be seeking constant marginal-cost segments.

Example 6: Consider the setup in Example 1 where there is a stream of N packet arrivals and a deadline T by which all the data must depart. The curves $A(t)$ and $D_{min}(t)$ for this problem are depicted in Figure 7. To obtain the departure curve satisfying the optimality criterion proceed as follows. Start at time 0; let $\{k_i\}$, $i = 1, \dots, N$, be the marginal costs to meet each of (t^i, A^i) points individually, i.e. k_i is the marginal cost associated with optimally transmitting A^i bits over time $[0, t^i]$. Let k_{min} be the minimum among $\{k_i\}$ and i_{min} the corresponding index of the minimizing jump point. The first segment of $D^{opt}(t)$ is then the constant marginal cost solution between $[0, t^{i_{min}}]$ with marginal cost k_{min} . Now, starting at $(t^{i_{min}}, A^{i_{min}})$ repeat the algorithm by shifting the origin to this point and considering the jump points beyond $t^{i_{min}}$, i.e. considering all i such that $t^i > t^{i_{min}}$. Finally, the algorithm stops when $t^{i_{min}} = T$.

Example 7: Consider the setup in Example 2 where the queue has M data packets with the j^{th} packet having b_j bits and a deadline t_j , $j = 1, \dots, M$. For this problem, the curves $A(t)$ and $D_{min}(t)$ are shown in Figure 9(a). As in the previous example, the departure curve satisfying the optimality criterion is constructed as follows. At time 0, let $\{k_j\}$, $j = 1, \dots, M$ be the marginal costs to meet the (t_j, B_j) points, i.e. k_j is the marginal cost associated with optimally transmitting

B_j bits over time $[0, t_j]$. Let k_{max} be the maximum among $\{k_j\}$ and j_{max} be the corresponding index of the maximizing jump point. The first segment of $D^{opt}(t)$ is then the constant marginal-cost solution between $[0, t_{j_{max}}]$ with marginal cost k_{max} . Now, starting at $(t_{j_{max}}, B_{j_{max}})$ repeat the algorithm by shifting the origin to this point and considering the jump points beyond $t_{j_{max}}$. The algorithm finally stops at the step when $t_{j_{max}} = T$.

V. CONCLUSION

We considered the problem of transmitting quality-of-service constrained data over a finite-time horizon, using minimum transmission energy expenditure. The problem was formulated using a calculus approach and posed as a continuous-time optimization. This novel formulation provided a graphical visualization of the problem and an appealing “string visualization” for the optimal policy. Utilizing the general solution, we considered various examples and presented the optimal policy under these setups. An online policy was also developed and it was shown through simulation results to provide significant gains as compared to the head-of-line drain policy. The energy minimization formulation was then extended to incorporate a time-varying power-rate function.

The continuous-time cumulative curves approach taken in this paper leads to a simple optimal solution and avoids many of the complications associated with discrete-time approaches and dynamic-programming. Hence, we believe that this approach provides promise for future research into issues related to delay and quality-of-service constraints in wireless networks. Promising extensions to this work include the consideration of multiple users and multi-hop traffic.

REFERENCES

- [1] A. K. Katsaggelos, Y. Eisenberg, F. Zhai, R. Berry and T. Pappas, “Advances in efficient resource allocation for packet-based real-time video transmission”, *Proceedings of the IEEE*, vol. 93, no. 1, Jan. 2005.
- [2] A. Ephremides, “Energy concerns in wireless networks”, *IEEE Wireless Communications*, vol. 9, issue 4, pp. 48-59, August 2002.
- [3] A. Jalali, R. Padovani, R. Pankaj, “Data throughput of CDMA-HDR a high efficiency high data rate personal communication wireless system”, *IEEE Vehicular Technology Conf.*, vol. 3, 2000.
- [4] S. Nanda, K. Balachandran and S. Kumar, “Adaptation techniques in wireless packet data services”, *IEEE Communications Magazine*, Jan. 2000.
- [5] Murtaza Zafer, “Dynamic rate-control and scheduling algorithms for quality-of-service in wireless networks”, PhD Thesis, Massachusetts Institute of Technology (MIT), 2007.
- [6] M. Zafer and E. Modiano, “A calculus approach to minimum energy transmission policies with quality of service guarantees”, *In Proc. of IEEE INFOCOM 2005*, vol. 1, pp. 548-559, March 2005.
- [7] Berry R., Gallager R. “Communication over fading channels with delay constraints”, *IEEE Tran. on Information Theory*, vol. 48, no. 5, May 2002.
- [8] B. Prabhakar, E. Biyikoglu, A. El Gamal, “Energy-efficient transmission over a wireless link via lazy packet scheduling”, *IEEE INFOCOM 2001*, vol. 1, pp 386 - 394, April 2001.
- [9] Fu A., Modiano E., Tsitsiklis J. “Optimal energy allocation for delay-constrained data transmission over a time-varying channel”, *INFOCOM 2003*, vol. 2, pp 1095 - 1105, April 2003.
- [10] Khojastepour M., Sabharwal A. “Delay-constrained scheduling: power efficiency, filter design and bounds”, *INFOCOM 2004*, March 2004.
- [11] Nuggehalli P., Srinivasan V., Rao R.R. “Delay constrained energy efficient transmission strategies for wireless devices”, *INFOCOM 2002*, vol. 3, pp. 1765 - 1772, June 2002.
- [12] E. Uysal-Biyikoglu and A. El Gamal, “On adaptive transmission for energy-efficiency in wireless data networks,” in *IEEE Tran. on Information Theory*, Dec. 2004.
- [13] Cruz R., “A Calculus for Network delay” (Parts I&II), *IEEE Transactions on Information Theory*, Jan., 1991.
- [14] Collins B. and Cruz R., “Transmission policies for time varying channels with average delay constraints”, *Proceedings, 1999 Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, 1999.
- [15] Ata B., “Dynamic power control in a wireless static channel subject to a quality of service constraint”, *To appear in Operations Research*.
- [16] L. Tassiulas and A. Ephremides, “Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks”, *IEEE Transactions on Automatic Control*, vol. 37, no. 12, Dec. 1992.
- [17] M. J. Neely, E. Modiano, and C. E Rohrs, “Dynamic power allocation and routing for time varying wireless networks”, *IEEE Journal on Selected Areas in Communications*, Jan. 2005.
- [18] A. L. Stolyar, “Maximizing queueing network utility subject to stability: greedy primal-dual algorithm”, *Queueing Systems* vol. 50, no.4, pp.401-457, 2005.
- [19] A. Eryilmaz, R. Srikant and J. Perkins, “Stable scheduling policies for fading wireless channels,” *IEEE/ACM Transactions on Networking*, vol. 13, no. 2, pp. 411-424, April 2005.
- [20] X. Liu, E. Chong, N. Shroff, “A framework for opportunistic scheduling in wireless networks ” *Computer Networks*, 41, pp. 451-474, 2003.
- [21] S. Borst, P. Whiting, “Dynamic rate control algorithms for HDR throughput optimization”, *IEEE INFOCOM*, Alaska, April 2001.
- [22] M. Zafer, E. Modiano, “Joint scheduling of rate-guaranteed and best-effort services over a wireless channel”, *IEEE CDC-ECC*, Spain, Dec. 2005.
- [23] A. Goldsmith, “The capacity of downlink fading channels with variable rate and power”, *IEEE Tran. Vehicular Technology*, vol. 45, pp. 1218-1230, Oct. 1997.
- [24] J. Le Boudec and P. Thiran, “Network Calculus”, *Springer Verlag*, LNCS 2050, 2001.
- [25] J. D. Salehi, Z. Zhang, J. Kurose, D. Towsley, “Supporting stored video: Reducing rate variability and end-to-end resource requirements through optimal smoothing”, *IEEE/ACM Transactions on Networking*, vol. 6, no. 4, Aug. 1998.
- [26] B. Gaujal, N. Navet and C. Walsh, “Shortest-path algorithms for real-time scheduling of FIFO tasks with minimal energy use”, *ACM Transactions on Embedded Computing Systems*, Vol. 4, No. 4, November 2005.
- [27] I.S. Gradshteyn and I.M. Ryzhik, “*Table of Integrals, Series and Products*”, Academic Press, San diego, 2000.
- [28] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, 3rd edition, Singapore, 1976.
- [29] E.R. Pinch, *Optimal Control and the Calculus of Variations*, Oxford University Press, 1993.
- [30] D. Luenberger, *Optimization by vector space methods*, John Wiley & sons, 1969.

APPENDIX A

PROOF OF THEOREM II - Uniqueness

Let us assume that the admissible departure curve satisfying the optimality criterion, $\hat{D}(t)$, is not unique. Let $D_1(t)$ and $D_2(t)$ be two such distinct curves. From the boundary conditions we have $D_1(0) = D_2(0) = 0$ and $D_1(T) = D_2(T) = D_{min}(T)$. Since $D_1(t) \neq D_2(t)$ the two curves must differ over some time interval in $[0, T]$. Let $t = a$ be the first instant at which the two curves differ and $t = b$ be the first time instant after $t = a$ at which they are equal again. Note that $b \leq T$ as at time T , $D_1(T) = D_2(T)$. Without loss of generality let $D_1(t) > D_2(t)$, $t \in (a, b)$. From the admissibility of the two curves we have,

$$D_{min}(t) \leq D_2(t) < D_1(t) \leq A(t), \quad t \in (a, b) \quad (23)$$

By assumption, since both curves $D_1(t)$ and $D_2(t)$ satisfy the optimality criterion, Lemmas 2-4 apply for points of slope changes. As $D_1(t)$ is strictly greater than $D_{min}(t)$ in $t \in (a, b)$ it follows from those lemmas that its slope cannot decrease in (a, b) . This implies that $D_1(t)$ is convex in (a, b) (it could be linear as well). Similarly as $D_2(t)$ is strictly less than $A(t)$

in $t \in (a, b)$, its slope cannot increase and hence it must be concave in (a, b) . It is clear that starting with $D_1(a) = D_2(a)$ and having $D_1(t)$ convex and $D_2(t)$ concave in $t \in (a, b)$, the two curves cannot be equal again at $t = b$ which leads to a contradiction. Finally, if both curves are linear in (a, b) with equality at $t = a$ and $t = b$, then this violates the assumption that $D_1(t) \neq D_2(t)$, $t \in (a, b)$.

To show that $\tilde{D}(t)$ minimizes the energy cost in (3), we proceed as follows. First, as defined in the problem statement in (3), we have $\tilde{D}(t) \in \Gamma$, where Γ is the set of all non-decreasing, continuous functions with bounded right-derivative for all $t \in [0, T]$. In addition, we also assume that $|D'(t)| < \mathcal{M}$, $\forall D(t) \in \Gamma$, $\forall t \in [0, T]$, where $\mathcal{M} > 0$ is chosen large enough so that all practical policies of interest (with finite-energy cost) are included in Γ . Also, the curves $A(t)$ and $D_{min}(t)$ are assumed to have a bounded right-derivative for all $t \in [0, T]$.

Let \mathcal{B} denote the space of continuous functions defined on $[0, T]$ with the supremum norm, $\|f\| = \sup_{t \in [0, T]} f(t)$; this space is then a Banach space [28]. Let Ω denote the set of all admissible departure curves, i.e. $\{D(t) : D(t) \in \Gamma \text{ and } D_{min}(t) \leq D(t) \leq A(t)\}$. We then have $\Omega \subset \mathcal{B}$. First, we claim that Ω is a convex set. To see this, consider $D_1(t), D_2(t) \in \Omega$ and let $D_3(t) = xD_1(t) + (1-x)D_2(t)$, $x \in [0, 1]$. Since $D_1(t), D_2(t)$ are continuous, non-decreasing and have bounded right-derivative, it is easy to see that $D_3(t)$ also has these properties. Further, we also have $x D_{min}(t) \leq x D_1(t) \leq x A(t)$ and $(1-x) D_{min}(t) \leq (1-x) D_2(t) \leq (1-x) A(t)$, which gives, $D_{min}(t) \leq D_3(t) \leq A(t)$; thus, the causality and the QoS constraints are also satisfied. Next, we show that Ω is compact. To see this, consider a sequence of admissible departure curves $\{D_n(t)\}_{n=1}^{\infty}$. Since $|D'(t)| \leq \mathcal{M}$, $\forall D(t) \in \Omega$, we have, $|D_n(t_2) - D_n(t_1)| \leq \mathcal{M}|t_2 - t_1|$, which makes the sequence of functions $\{D_n(t)\}$ form an equi-continuous family of functions. From [28] (Thm. 7.25, pg. 158), it then follows that there is a subsequence that converges in the supremum norm. Thus, this limit function is continuous and since $D_n(t)$ satisfies the causality and the QoS constraints for all n , it is satisfied by the limit function as well. Hence, the limit function lies in Ω and we see that Ω is compact. Now, consider the energy cost function $\mathcal{E}(D(t))$ as given in (3) with $g(\cdot)$ being strictly convex. We next show that $\mathcal{E}(D(t))$ is also strictly convex. Consider $D_1(t), D_2(t) \in \Omega$ and let $D_3(t) = xD_1(t) + (1-x)D_2(t)$, $x \in [0, 1]$. Then, $\mathcal{E}(D_3(t)) = \int_0^T g(xD_1'(t) + (1-x)D_2'(t)) dt < \int_0^T (xg(D_1'(t)) + (1-x)g(D_2'(t))) dt$. Thus, we see that, $\mathcal{E}(D_3(t)) < x\mathcal{E}(D_1(t)) + (1-x)\mathcal{E}(D_2(t))$. From above, we see that (3) involves an optimization of a strictly convex functional over a compact convex set. Thus, it has a unique minimizer in Ω [30]. From Theorem I, the necessary condition for any admissible departure curve to be the minimizer is that it must satisfy the optimality criterion and since such a curve is unique, it must be the optimal solution.

APPENDIX B

PROOF OF THEOREM III - Minimal Maximum Power

Consider an admissible departure curve $D(t)$ that is not optimal. Let $[a, b]$ be the interval over which the optimality

criterion is violated. Then, based on the construction in Theorem I we obtain a new curve $\tilde{D}(t)$ that is also admissible. The line segment $L(t)$ between $[a, b]$ in $\tilde{D}(t)$ always has a slope that is less than the maximum slope of $D(t)$ between $[a, b]$. As $\tilde{D}(t) = D(t)$, $t \notin (a, b)$, the overall maximum slope of $\tilde{D}(t)$ cannot exceed that of $D(t)$. Thus,

$$\max_{t \in [0, T]} \tilde{D}'(t) \leq \max_{t \in [0, T]} D'(t) \quad (24)$$

If $\tilde{D}(t) = D^{opt}(t)$ then we are done. If not proceed as follows.

From Theorem II, we know that $D^{opt}(t)$ is unique and minimizes the energy cost for any non-negative, convex increasing power-rate function $g(r)$. In particular, consider the sequence of functions $g_n(r) = r^n$. For any $D(t)$ we know that, $\lim_{n \rightarrow \infty} \left(\int_0^T (D'(t))^n dt \right)^{1/n} = \max_{t \in [0, T]} D'(t)$. Since $D^{opt}(t)$ minimizes the integral for all n , we obtain $\max_{t \in [0, T]} (D^{opt})'(t) \leq \max_{t \in [0, T]} D'(t)$ as required.

APPENDIX C

PROOF OF LEMMA 5

(a) Let \tilde{t} be the point at which $L_D(t)$ intersects $D_{min}(t)$ first. By definition, $L_D(t) < A(t)$, $\forall t \in (t_0, \tilde{t})$. The proof now follows in two parts. First, we show that any line in \mathcal{F} with slope less than L'_D must intersect $D_{min}(t)$ at or before \tilde{t} and second that this line does not intersect $A(t)$ in (t_0, \tilde{t}) . Consider $L(t) \in \mathcal{F}$ with slope less than L'_D , then, $L(t) < L_D(t)$, $\forall t > t_0$. Hence, at time \tilde{t} we have $L(\tilde{t}) < L_D(\tilde{t}) = D_{min}(\tilde{t})$. If instead, \tilde{t} is the discontinuity point for $D_{min}(t)$, then, $L_D(t) - D_{min}(t)$ changes sign at \tilde{t} and so $L(t) - D_{min}(t)$ must have changed sign earlier at $t \leq \tilde{t}$. Thus, we see that $L(t)$ must intersect $D_{min}(t)$ at or before \tilde{t} . Next, since $L(t) < L_D(t) < A(t)$ in $t \in (t_0, \tilde{t})$, the line $L(t)$ cannot intersect $A(t)$ first. This completes the proof of part (a) in the lemma. Along similar lines as above, part (b) follows.

APPENDIX D

PROOF OF OPTIMALITY FOR THE $D^{opt}(t)$ ALGORITHM

From Theorem II we know that $D^{opt}(t)$ is unique. Hence it suffices to prove that the constructed curve satisfies the optimality criterion.

Let $D_c(t)$ denote the constructed curve. It is obvious from the construction that at all points where the slope changes Lemma 2 is satisfied. We next show that Lemmas 3 and 4 are also satisfied. Let t_0 be the starting instant at some iteration and suppose that L_o intersects $D_{min}(t)$ first, i.e. at t_1 (as in the algorithm) we have $L_o(t_1) = D_{min}(t_1)$. Also, suppose that $L_o(t_1) \neq A(t_1^-)$. From the chosen t_1 in step 2, it is clear that $L_o(t) < A(t)$ in $(t_0, t_1]$. Thus, if we pick a line $L_1 \in F_A$ with slope close to $L'_o (= \beta_o)$, then L_1 would intersect $A(t)$ beyond t_1 . More precisely, there exists an $\epsilon > 0$ such that any $L_1 \in F_A$ with slope $\beta_o < L'_1 < \beta_o + \epsilon$ intersects $A(t)$ first at $\tilde{t} > t_1$. Now, it follows that at the next iteration, starting from time t_1 , the new set \mathcal{F}_A must at least contain all lines with slopes in $(\beta_o, \beta_o + \epsilon)$, hence, the optimal line starting at time t_1 (at the new iteration) cannot have slope greater than β_o (β_o here refers to the optimal slope for the iteration at t_0). Thus, we see that Lemma 3 is satisfied at t_1 .

Similarly, if in step 2, we have $L_o(t_1) = A(t_1)$ (or $A(t_1^-)$), then, using a similar argument as above it can be seen that starting from time t_1 , the new set \mathcal{F}_{D_m} must at least contain all lines with slopes in $(\beta_o, \beta_o - \epsilon)$. Hence, the optimal line starting at time t_1 cannot have slope less than β_o , and this shows that Lemma 4 is satisfied at t_1 . Note that, if at t_1 we have $L_o(t_1) = D_{min}(t_1) = A(t_1^-)$, then, it does not matter how the slope changes beyond t_1 .

Thus, we see that starting at $(0, 0)$, at every iteration of the algorithm (every constructed segment of $D_c(t)$) Lemmas 2-4 are satisfied. This implies that around every point where the slope of $D_c(t)$ changes we cannot construct an admissible line segment (as outlined in Theorem I); hence, $D_c(t)$ satisfies the optimality criterion.

APPENDIX E PROOF OF LEMMA 6

As presented in Section III-B, the two curves $A(t)$ and $D_{min}(t)$ for the BT-problem are, $A(t) = B, \forall t \in [0, T]$, and $D_{min}(t) = 0, t \in [0, T], D_{min}(T) = B$. The admissibility criterion is $0 \leq D(t) \leq B$ and $D(T) = B$. Re-phrasing the BT-problem as a calculus of variations problem we get [29],

$$\begin{aligned} \min_{r(t)} \quad & \mathcal{E}(D(t)) = \int_0^T g(r(t), t) dt \\ \text{subject to} \quad & D'(t) = r(t), \quad D(T) = B \\ & r(t) \geq 0, \quad t \in [0, T] \end{aligned} \quad (25)$$

Using [29], the Hamiltonian for the above is, $H(D, r, t) = g(r, t) + \lambda(t)r$, and from Pontryagin's maximum principle (which is also a sufficient condition in our case due to convexity) the optimal value $r^{opt}(t)$ satisfies, $r^{opt}(t) = \arg \max_{r \geq 0} H(D^{opt}, r, t) = \arg \max_{r \geq 0} (g(r, t) + \lambda(t)r)$. We also have, $\dot{\lambda}(t) = -\frac{\partial H}{\partial D} = 0$, which implies $\lambda(t) = \text{constant}$. Taking $k = -\lambda(t)$ as the constant and substituting back in the $r^{opt}(t)$ equation, we get, $r^{opt}(t) = \arg \max_{r \geq 0} (g(r, t) - kr)$. The solution to this maximization is as given in (18). Since, $g(r, t)$ is strictly convex and increasing in r , we have that r^* is unique. Finally, to ensure that a total of B units of data is transmitted by the deadline T , the value of k must be chosen such that $\int_0^T r^{opt}(t) dt = B$.

APPENDIX F PROOF OF LEMMA 7

Let k_1, k_2 be two positive values such that $0 < k_1 < k_2$. Let $r_{k_1}^{opt}(t), r_{k_2}^{opt}(t)$ be the corresponding optimal rate functions as given in (18). Suppose at time t , we have $r_{k_1}^{opt}(t) > 0$, then, due to strict convexity $\frac{\partial}{\partial r} g(r, t)$ is an increasing function of r and since $k_2 > k_1$ the unique r^* value for k_2 must be greater than for k_1 . This gives, $r_{k_2}^{opt}(t) > r_{k_1}^{opt}(t)$. If instead at time t , we have $r_{k_1}^{opt}(t) = 0$, then, $r_{k_2}^{opt}(t)$ can be either 0 or positive. Thus, we see that $r_{k_2}^{opt}(t) \geq r_{k_1}^{opt}(t), \forall t$, with equality only if both are zero. This shows that $D^{opt}(t)$ is non-decreasing in k . For a given k value, the uniqueness of $D^{opt}(t)$ follows since r^* is unique due to strict convexity.

Now suppose $k = 0$, then, since $g(r, t)$ is increasing in r we have $\frac{\partial}{\partial r} g(r, t) \geq 0, \forall r$ (a non-negative function of r). But, we

also know that $\frac{\partial}{\partial r} g(r, t)$ is an increasing function in r , thus, there cannot be any positive r^* such that $\frac{\partial}{\partial r} g(r, t)|_{r=r^*} = 0$ ($= k$ as taken). This gives $r^{opt}(t) = 0$ and $D^{opt}(t) = 0, \forall t$.

Lastly, suppose that $D^{opt}(T) = B > 0$ and let k_1, k_2 be two distinct k values such that $\int_0^T r_{k_1}^{opt}(s) ds = \int_0^T r_{k_2}^{opt}(s) ds = B$. Without loss of generality assume $k_2 > k_1$. From the earlier arguments we know that whenever $r_{k_1}^{opt}(t) > 0$, we have $r_{k_2}^{opt}(t) > r_{k_1}^{opt}(t)$. Since $B > 0$, an interval exists over which $r_{k_1}^{opt}(t) > 0$. Thus, we see that $\int_0^T r_{k_1}^{opt}(s) ds < \int_0^T r_{k_2}^{opt}(s) ds$, which leads to a contradiction, hence there is a unique k value that achieves $D^{opt}(T) = B$.



Murtaza Zafer received his B.Tech degree in Electrical Engineering from the Indian Institute of Technology (IIT) Madras, India, in 2001, his S.M. and Ph.D. degrees in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology (MIT), MA, USA, in 2003 and 2007 respectively. Currently, he is a Research Staff Member at IBM Thomas J. Watson Research Center, NY, USA, where his research is on communication and sensor networks. He spent the summer of 2004 at the Mathematical Sciences Research center, Bell

Laboratories Alcatel-Lucent Inc., and the summer of 2003 at Qualcomm, Inc.

His research interests are in queueing theory, information theory, control and optimization theory, algorithms and protocol design with applications to communication networks. He is the co-recipient of the best Student Paper award at WiOpt conference, 2005. He also received the Siemens and Philips award for academic excellence in 2001.



Eytan Modiano received his B.S. degree in Electrical Engineering and Computer Science from the University of Connecticut at Storrs in 1986 and his M.S. and PhD degrees, both in Electrical Engineering, from the University of Maryland, College Park, MD, in 1989 and 1992 respectively. He was a Naval Research Laboratory Fellow between 1987 and 1992 and a National Research Council Post Doctoral Fellow during 1992-1993. Between 1993 and 1999 he was with MIT Lincoln Laboratory where he was the project leader for MIT Lincoln Laboratory's Next Generation Internet (NGI) project. Since 1999 he has been on the faculty at MIT; where he is presently an Associate Professor. His research is on communication networks and protocols with emphasis on satellite, wireless, and optical networks.

He is currently an Associate Editor for IEEE Transactions on Information Theory, The International Journal of Satellite Communications, and for IEEE/ACM Transactions on Networking. He had served as a guest editor for IEEE JSAC special issue on WDM network architectures; the Computer Networks Journal special issue on Broadband Internet Access; the Journal of Communications and Networks special issue on Wireless Ad-Hoc Networks; and for IEEE Journal of Lightwave Technology special issue on Optical Networks. He was the Technical Program co-chair for WiOpt 2006, IEEE Infocom 2007, and ACM MobiHoc 2007.