Proceedings of the
46th IEEE Conference on Decision and Control
New Orleans, LA, USA, Dec. 12-14, 2007

Minimum Energy Transmission over a Wireless Fading Channel with Packet Deadlines

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

***Abstract*—We address transmission of data with deadline constraints over a wireless fading channel. Specifically, the system model consists of a wireless transmitter with data packets arriving to its queue having strict deadline constraints. The transmitter can control the transmission rate over time and the expended power depends on both the chosen rate and the present channel state. The objective is to obtain a rate-control policy that serves the data to meet the deadline constraints while also minimizing the transmission energy expenditure. Using a novel approach based on cumulative curves methodology and continuous-time stochastic control formulation, we obtain the optimal transmission policy under various specific scenarios. Utilizing these results, we then present an energy-efficient policy for the case of arbitrary packet arrivals and deadline constraints, and also give simulation results comparing its performance with a non-adaptive scheme.**

***Index Terms*—Energy efficiency, Delay constraints, Wireless fading channel, Rate control, Quality of Service.**

I. Introduction

Delay constraints and energy efficiency are important concerns in the design of modern wireless systems. Strict delay requirements frequently arise in many situations, for example, in video and real-time multimedia streaming in wireless data networks, in time-critical sensing applications in sensor networks and in real-time data communication in ad-hoc networks. Also, for wireless systems, battery energy is clearly a critical resource and minimizing the energy expenditure has numerous advantages in efficient battery utilization of mobile devices, increased lifetime of sensor and ad-hoc networks, and better utilization of energy sources in satellites. The work in this paper primarily addresses the above two concerns; specifically, the focus is to utilize dynamic rate-control to minimize the transmission energy cost subject to packet deadline constraints.

Modern wireless devices are equipped with channel-gain measurement and rate adaptation capabilities [1]–[3]. Channel measurement allows the transmitter-receiver pair to measure the fade state using a pre-determined pilot signal while rate control capability allows the transmitter to adjust the transmission rate over time. Such a control can be achieved in various ways that include adjusting the power level, symbol rate, coding scheme, constellation size or any combination of these approaches; further, in some technologies the receiver can detect these changes directly from the received data without the need for an explicit rate-change control information [3]. Since rate-adaptation can be done very rapidly over millisecond duration time-slots, this capability thus provides ample opportunity to optimize system performance.

The power-rate function defines the relationship that governs the amount of transmission power required to reliably transmit at a certain rate. Two fundamental aspects of this function, which are exhibited by most encoding/communication schemes, and hence are common assumptions in the literature are as follows [7]–[11], [15]. First, for a fixed bit error probability and channel state, the required transmission power is a convex function of the communication rate as shown in Figure 1(a). This implies, from Jensen's inequality, that transmitting data at low rates over a longer duration is more energy efficient as compared to high rate transmissions. Second, the wireless channel is time-varying which shifts the convex power-rate curves as a function of the channel state as shown in Figure 1(b). As good channel conditions require less transmission power, exploiting the channel variability over time can lead to reduced energy cost. Thus, it is evident that adapting the transmission rate intelligently over time, one can minimize the energy cost while ensuring also that the delay constraints are met.

In this paper, we consider a wireless transmitter with packets arriving to the queue having strict deadline constraints. The channel state varies stochastically over time and is assumed to be a Markov process. The objective is to obtain a rate-control policy that serves the packets within the deadline constraints and also minimizes the transmission energy expenditure. Towards this end, we first assume that the packet arrival information is known in advance and represent the arrivals, departures and deadline constraints in terms of cumulative curves. Using a continuous-time stochastic control formulation we obtain the optimal policy under various restricted setups. Then, based on the intuition developed from the optimal policy, we present a heuristic policy for the case of arbitrary (stochastic) packet arrivals and deadline constraints, and compare its performance with a non-adaptive scheme using simulation results.

Transmission power/rate control is an active area of research in communication networks in various different contexts. Adaptive network control and scheduling has been studied in the context of network stability [11], [13], average throughput [12], [14], average delay [7], [15] and packet drop probability [16]. This literature considers "average metrics" that are mea-

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 NASA Space Communication Project grant number NAG3-2835 and by ONR grant number N000140610064.

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

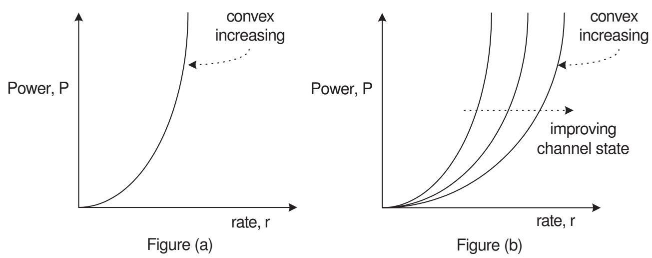

Fig. 1. Transmission power as a function of the rate and the channel state; (a) fixed channel state, (b) variable channel state.

sured over an infinite time horizon and hence do not directly apply for delay constrained/real-time data. Furthermore, for strict deadline constraints, rate-adaptation simply based on steady state distributions does not suffice and one needs to take into account the system dynamics over time; thus, introducing new challenges and complexity into the problem. Recent work in this direction includes [4], [8]–[10]. The work in [8] studied offline formulations assuming complete knowledge of the future and then devised heuristic online policies using the offline solution. The authors in [9] studied several data transmission problems using Dynamic Programming (DP), however, the specific problem that we consider in this work becomes intractable using this methodology. This is due to the large state space in the DP-formulation or the well-known "curse of dimensionality". The work in [10] studied energy-efficient data transmission over a static channel without fading. In our earlier work in [4], we utilized a cumulative curves approach to study data transmission with QoS constraints over a non-fading channel, while in [5], we considered a fading channel but a restricted setting with only a single deadline constraint.

II. System Model

We consider a continuous-time model of the system where the rate can be varied continuously in time. Clearly, such a model is an approximation but the assumption is justified, since in practice rate-adaptation can be done over time-slots of duration on the order of 1 msec [1], [2], whereas, the packet delay requirements are on the order of 100's of msec, thus justifying a continuous-time view of the system. A significant advantage of such a model is that it makes the problem mathematically tractable and yields simple solutions. The alternative discrete-time dynamic programming setup is intractable and would only yield numerical solutions without much insights. The results obtained using the continuous-time model can then be applied to the discrete-time system in a very straightforward manner by simply evaluating the solution at discrete times as done for the simulations in Section V-B.

A. Transmission Model

Let h_t denote the channel gain, $P(t)$ the transmitted signal power and $P^{rcd}(t)$ the received signal power at time t. We make the common assumption [7]–[11], [15] that the required received signal power for reliable communication, with a certain low bit-error probability, is convex in the rate; i.e. $P^{rcd}(t) = g(r(t))$, where $g(r)$ is a non-negative, convex, increasing function for $r \geq 0$. Since the received signal power is given as $P^{rcd}(t) = |h_t|^2 P(t)$, the required transmission power to achieve rate $r(t)$ is given by,

$$P(t) = \frac{g(r(t))}{c(t)} \tag{1}$$

where $c(t) \triangleq |h_t|^2$. The quantity $c(t)$ is referred to as the *channel state* at time t. Its present value at time t is assumed known either through prediction or direct channel measurement but evolves stochastically in the future. In this paper, we take $g(r)$ to belong to the class of *Monomial* functions, namely, $g(r) = kr^n,\ n > 1, k > 0\ (n, k \in \mathbb{R})$. While this assumption helps mathematically as it leads to simple closed-form optimal solutions, more importantly, for most practical transmission schemes the function $g(\cdot)$ is described numerically and its exact analytical form is unknown. In such situations, one can obtain the best (least-square) approximation of that function to the form kr^n and then apply the results thus obtained. Without loss of generality, throughout the paper we take $k = 1$, since any other value of k simply scales the energy cost without affecting the optimal policy results.

B. Channel Model

We assume a general first-order, continuous-time, discrete state space Markov model for the channel state process. Such models have been widely used in the literature for wireless channel fading, and first-order Markov models have been proposed for common fading scenarios of Rayleigh, Rician etc. [17]–[19].

Denote the channel state process as $C(t)$ and the state space as $\mathcal{C}$. Let $c \in \mathcal{C}$ denote a particular channel state and $\{c(t), t \geq 0\}$ denote a sample path. Starting from state c, let $\mathcal{J}_c$ be the set of all states ($\neq c$) to which the channel can transition when the state changes. Let $\lambda_{c\tilde{c}}$ denote the channel transition rate from state c to $\tilde{c}$, then the sum transition rate at which the channel jumps out of state c is, $\lambda_c = \sum_{\tilde{c} \in \mathcal{J}_c} \lambda_{c\tilde{c}}$. The expected time that $C(t)$ spends in state c is $\frac{1}{\lambda_c}$, and one can view $\frac{1}{\lambda_c}$ as the channel coherence time in state c.

Now, define $\lambda \triangleq \max_c \lambda_c$ and a random variable, $Z(c)$, as,

$$Z(c) \triangleq \begin{cases} \tilde{c}/c, & \text{with probability } \lambda_{c\tilde{c}}/\lambda,\ \tilde{c} \in \mathcal{J}_c \\ 1, & \text{with probability } 1 - \lambda_c/\lambda \end{cases} \tag{2}$$

With this definition, we obtain a compact and simple description of the process evolution as follows. *Given a channel state c, there is an Exponentially distributed time duration with rate λ after which the channel state changes. The new state is a random variable which is given as $C = Z(c)c$.* Clearly, from (2) the transition rate to state $\tilde{c} \in \mathcal{J}_c$ is unchanged at $\lambda_{c\tilde{c}}$, whereas with rate $\lambda - \lambda_c$ there are indistinguishable self-transitions. This is a standard Uniformization technique and there is no process generality lost with the new description as it yields a stochastically identical scenario.

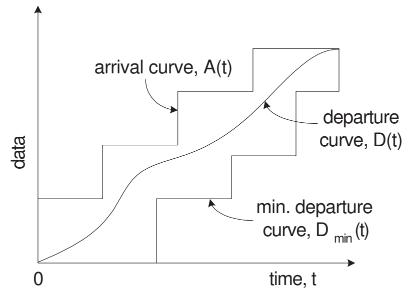

Fig. 2. Schematic diagram of $A(t)$, $D(t)$ and $D_{min}(t)$ curves.

C. Data Model

To describe the flow of data in the queue we utilize a cumulative curves approach [4], [20]. As will be evident later, such an approach provides an appealing visualization of the energy minimization problem. Let $A(t)$ and $D(t)$ be the arrival and the departure curve respectively. To model the packet deadlines (and other quality-of-service) constraints we introduce a new notion of a *minimum departure curve* denoted as $D_{min}(t)$. These curves 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* $[0,t]$.

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 (QoS) requirements.*

Thus, in simple terms, $A(t)$ denotes how data arrives to the queue, $D(t)$ denotes how data leaves the queue and $D_{min}(t)$ denotes the minimum amount of data that must depart to satisfy the QoS constraints. Clearly, we require that $D(t) \leq A(t),\ \forall t$ (causality constraints), i.e. only past data that has already arrived can be served and not future data. We also require $D(t) \geq D_{min}(t),\ \forall t$ (QoS constraints). Thus, we see that in a compact way the QoS and the causality constraints can be expressed as, $D_{min}(t) \leq D(t) \leq A(t),\ \forall t$. Figure 2 gives a schematic illustration of these curves for a packet arrival model. We, next, present a few illustrative examples of how $D_{min}(t)$ can model various common QoS constraints.

Delay Constraint: Consider an arrival curve $A(t)$ and a deadline constraint d on the data. In this case, $D_{min}(t) = 0, t \in [0,d)$ and $D_{min}(t) = A(t-d), t \geq d$; thus, $D_{min}(t)$ is simply a time-shifted version of $A(t)$. One can also model variable deadline constraints as presented in Section IV-B.

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 $D(t) \geq \max[A(t) - B, 0]$. This gives the minimum departure curve as $D_{min}(t) = \max[A(t) - B, 0]$. It is easy to incorporate a time-varying buffer constraint $B(t)$ as well.

III. Problem Formulation

Consider a time interval $[0,T]$ and let $A(t)$ be the arrival curve over this period; we assume that $A(t)$ is known in advance. Based on the specific QoS requirements, the minimum departure curve $D_{min}(t)$ is obtained as outlined in Section II-C. Now, given $A(t)$ and $D_{min}(t)$ curves, the objective is to dynamically construct the departure curve $D(t)$ such that the expected energy cost is minimized. Note that while $A(t)$ and $D_{min}(t)$ are assumed known, the actual departure curve $D(t)$ followed would depend on the underlying channel sample path. Later in Section V, we utilize the optimal solution obtained under the above setup and present an energy-efficient policy for arbitrary packet arrivals (unknown in advance).

Optimal Control Formulation: We consider a continuous-time stochastic control formulation with the control being the chosen transmission rate. Let the system state be denoted as (D,c,t), where the notation means that at the present time t, the cumulative amount of data that has been transmitted is $D(t) = D$, and the channel state is $c(t) = c$. Let $r(D,c,t)$ denote a transmission policy and since the underlying channel process is Markov, it is sufficient to restrict attention to policies that depend only on the present system state [23].

Given a policy $r(D,c,t)$ the system evolves in time as a Piecewise-Deterministic-Process (PDP) as follows. It starts with the initial state $D(0) = 0$ and $c(0) = c_0$ ($c_0 \in \mathcal{C}$). Until τ_1, where τ_1 is the first time instant after $t = 0$ at which the channel changes, data is transmitted at the rate $r(D(t), c_0, t)$. Hence, over $t \in [0, \tau_1)$, $D(t)$ satisfies the differential equation,

$$\frac{dD(t)}{dt} = r(D(t), c_0, t) \tag{3}$$

Equivalently, $D(t) = D(0) + \int_0^t r(D(s), c_0, s)ds,\ t \in [0, \tau_1]$. Then, starting from the new state $(D(\tau_1), c(\tau_1), \tau_1)$ until the next channel transition we have, $\frac{dD(t)}{dt} = r(D(t), c(\tau_1), t)$, $t \in [\tau_1, \tau_2)$; and this procedure repeats in time.

A transmission policy, $r(D,c,t)$, is *admissible*, if it satisfies the following:

(a) $0 \leq r(D,c,t) < \infty$, (non-negativity of rate), and,

(b) $D_{min}(t) \leq D(t) \leq A(t), t \in [0,T]$, almost surely (a.s.), (deadline and causality constraints)

Now, define a *cost-to-go* function, $J_r(D,c,t)$, as the expected energy expenditure going forward in time starting from the state (D,c,t). Then,

$$J_r(D,c,t) = E\left[\int_t^T \frac{1}{c(s)} g(r(D(s), c(s), s))ds\right] \tag{4}$$

where the above expectation is taken over $\{c(s), s \in (t,T]\}$ and conditioned on the starting state $D(t) = D$, $c(t) = c$. Define a *minimum cost function*, $J(D,c,t)$, as the infimum of $J_r(D,c,t)$ over the set of all admissible transmission policies.

$$J(D,c,t) = \inf_{r(D,c,t)} J_r(D,c,t), \quad r(D,c,t) \text{ admissible} \tag{5}$$

The optimization problem, stated concisely, is to compute the optimal policy $r^*(D, c, t)$ that achieves the minimum cost $J(D, c, t)$.

Using stochastic control theory, it can be shown that $J(D, c, t)$ satisfies the following *Optimality Equation* [6], [21]–[23],

$$\min_{r\in[0,\infty)} \Big\{ \frac{g(r)}{c} + \frac{\partial J(D,c,t)}{\partial t} + r\frac{\partial J(D,c,t)}{\partial D} + \lambda(E_z[J(D,Z(c)c,t)] - J(D,c,t)) \Big\} = 0 \quad (6)$$

where $E_z[\cdot]$ above is expectation with respect to the $Z(c)$ variable which was defined in (2). The above partial differential equation is also referred to as the Hamilton-Jacobi-Bellman equation. The optimal rate $r^*(D, c, t)$ for a system state (D, c, t) is the value of r that minimizes (6) above.

IV. Optimal Transmission Policy

In the last section, we presented the optimality equation for a general setup. We now proceed to present analytical results for the optimal policy under various specific scenarios. To present the results, however, we require a few additional notations regarding the channel process. Let there be total m channel states in the Markov model and denote the various states $c \in \mathcal{C}$ as $c^1, c^2, \ldots, c^m$. Given a channel state c^i, the values taken by the random variable $Z(c^i)$ (defined in (2)) are denoted as $\{z_{ij}\}$, where $z_{ij} = c^j/c^i$. The probability that $Z(c^i) = z_{ij}$ is denoted as p_{ij}. Clearly, if there is no transition from state c^i to c^j, $p_{ij} = 0$. Also, as pointed out earlier, without loss of generality we take the multiplicative constant $k = 1$ in the power-rate function.

A. *BT-problem*

We consider first the following setup – the transmitter has B bits of data in the queue, there are no new arrivals and there is a single deadline T by which this data must be transmitted. We refer to this as the "*BT-problem*" and it was analyzed in detail in [5]. In terms of the cumulative curves, we have $A(t) = B,\ t \in [0, T]$ since the queue has B bits to begin with at time 0 and no more data is added. We have $D_{min}(t) = 0,\ t \in [0, T); D_{min}(T) = B$ since until the deadline $t < T$ there is no minimum data transmission requirement while at T the entire B bits must have been transmitted. A schematic diagram of this is given in Figure 3(a).

Theorem *I: (**BT-problem**) Consider the BT-problem with $g(r) = r^n$, $n > 1, n \in \mathbb{R}$ and a Markov channel model. The optimal policy, $r^*(D, c, t)$, and the function, $J(D, c, t)$, are,*

$$r^*(D, c^i, t) = \frac{B - D}{f_i(T - t)},\ i = 1, \ldots, m \quad (7)$$

$$J(D, c^i, t) = \frac{(B - D)^n}{c^i(f_i(T - t))^{n-1}},\ i = 1, \ldots, m \quad (8)$$

The functions $\{f_i(s)\}_{i=1}^m$ are the solution of the following ODE with the boundary conditions $f_i(0) = 0, f_i'(0) = 1, \forall i$[1]*,*

[1]For numerical evaluation of the ODE solution, take a small $\epsilon > 0$, set $f_i(\epsilon) = \epsilon, \forall i$ and use an initial-value ODE solver to obtain $\{f_i(s)\},\ s \geq \epsilon$.

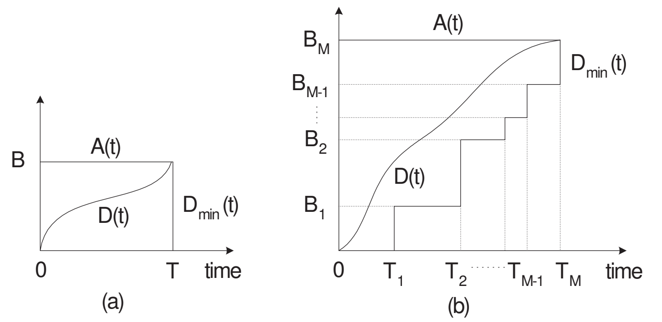

Fig. 3. Cumulative curves for (a) BT-problem, (b) Variable deadlines case.

$$f_1'(s) = 1 + \frac{\lambda f_1(s)}{n-1} - \frac{\lambda}{n-1}\sum_{k=1}^{m} \frac{p_{1k}}{z_{1k}} \frac{(f_1(s))^n}{(f_k(s))^{n-1}} \quad (9)$$

$$\vdots$$

$$f_m'(s) = 1 + \frac{\lambda f_m(s)}{n-1} - \frac{\lambda}{n-1}\sum_{k=1}^{m} \frac{p_{mk}}{z_{mk}} \frac{(f_m(s))^n}{(f_k(s))^{n-1}} \quad (10)$$

A complete proof of the above theorem can be found in [6], however, as a direct check it can be verified that the solution in (7)-(8) satisfies the optimality equation in (6).

The ODE system for $\{f_i(s)\}_{i=1}^m$ above can be easily solved numerically using standard techniques (eg. ODE solvers in MATLAB). Also, this computation needs to be done only once before the system starts operating. In fact, $\{f_i(s)\}$ can be pre-determined and stored in a table in the transmitter memory. Once the functions $\{f_i(s)\}$ are known, the closed form structure of the optimal policy in (7) requires no further computation. At time t, the transmitter simply looks at the amount of data D that has been transmitted, the channel state c^i, and then using the appropriate $f_i(\cdot)$ function it computes the transmission rate as $\frac{B-D}{f_i(T-t)}$.

From (7), we see that the optimal rate is linear in $(B - D)$ (amount of data left) with slope $\frac{1}{f_i(T-t)}$. Viewing the quantity $\frac{1}{f_i(T-t)}$ as the "urgency" of transmission at time t for channel state c^i, we obtain an appealing view of (7) as follows,

*optimal rate = amount of data left * urgency of transmission*

As $t \to T$, $f_i(T - t)$ goes to zero; thus, as expected, the urgency of transmission increases as t approaches the deadline. Another observation is that setting $\lambda = 0$ (no channel variations) gives $f_i(T - t) = T - t, \forall i$, and, $r^*(D, c, t) = \frac{B-D}{T-t}$. Thus, with no channel variations the optimal policy is to transmit at a rate that just empties the buffer by the deadline. This observation is consistent with the earlier results in the literature for non-fading/time-invariant channels [4], [8], [10].

Constant Drift Channel: In Theorem I, the optimal policy was presented under full generality on the Markov channel model. If we now impose a special structure on the channel model which we refer to as the *constant drift* channel, the ODE can be solved analytically. In this channel model, we make the assumption that the expected value of the random variable $1/Z(c)$ is independent of the channel state; i.e. $E[1/Z(c)] = \beta$, a constant. Thus, starting in state c, if $\tilde{c}$ denotes the next transition state, we have $E\left[\frac{1}{\tilde{c}}\right] = E\left[\frac{1}{Z(c)}\right]\frac{1}{c} = \frac{\beta}{c}$. This means that if we look at the process $1/c(t)$, the expected value of the next state is a constant multiple of the present state. We refer to β as the "*drift*" parameter of the channel process. If $\beta > 1$ the process $1/c(t)$ has an upward drift; if $\beta = 1$ there is no drift and if $\beta < 1$ the drift is downwards.

There are various situations where the above model is applicable at least over the time scale of the deadline. For example, when a mobile device is moving in the direction of the base station, the channel has an expected drift towards improving conditions and vice-versa. Similarly, in case of satellite channels, changing weather conditions such as cloud cover makes the channel drift towards worsening conditions and vice-versa. In many cases, the time scale of these drift changes is longer than the packet deadlines in which case a constant drift channel serves as an appropriate model.

Theorem *II: Consider the BT-problem with $g(r) = r^n$, $n > 1, n \in \mathbb{R}$ and a constant drift channel with drift β. The optimal policy, $r^*(D,c,t)$, and the function, $J(D,c,t)$, are,*

$$r^*(D,c,t) = \frac{B-D}{f(T-t)} \tag{11}$$

$$J(D,c,t) = \frac{(B-D)^n}{c(f(T-t))^{n-1}} \tag{12}$$

where $f(s) = \frac{(n-1)}{\lambda(\beta-1)}(1-\exp(-\frac{\lambda(\beta-1)}{n-1}s))$.

Proof: Omitted for brevity but can be found in [6]. ■

Note that for this channel model, the function $f_i(s)$ is the same for all channel states and is denoted as $f(s)$ above.

B. Variable Deadlines Setup

Consider now the second setup where the queue has M packets that are arranged and served in the earliest-deadline-first order. Let b_j be the number of bits in the j^{th} packet and T_j be the deadline for this packet; assume $0 < T_1 < T_2 < \ldots < T_M$. There are no new arrivals and the objective is to serve this data with minimum energy while meeting the deadline constraints. In terms of the cumulative curves, the setup is as shown in Figure 3(b). Let $B_j = \sum_{l=1}^{j} b_l$, where B_j is the cumulative amount of data in the first j packets. Then, $A(t) = B_M, \forall t$, since B_M bits are in the queue at time 0 and no more data is added. The curve $D_{min}(t)$ is piecewise-constant with jumps at times T_j; i.e. at time T_j, $D_{min}(T_j) = B_j$, since the first B_j bits must be transmitted by time T_j.

A direct solution of the variable deadlines problem by solving the PDE equation in (6) is fairly difficult, due to the complexity of the multiple deadline constraints involved. Therefore, we consider a natural decomposition of it in terms of multiple BT-problems. This provides a simple transmission policy which can be shown to be optimal under the constant drift channel model. A visual comparison of the two diagrams in Figure 3 suggests the following approach. First, we can visualize the deadline constraints in terms of the cumulative amounts as $\{B_jT_j\}_{j=1}^{M}$ constraints, that is, a total of B_j bits must be transmitted by deadline T_j ($j = 1,\ldots,M$). Clearly, each B_jT_j constraint is like a BT-problem except that now there are multiple such constraints that all need to be satisfied. For every time t and channel state c, we know the optimal transmission rate to meet each of the B_jT_j constraint individually (assuming only this constraint existed), thus, to meet all the constraints a natural solution is to simply choose the maximum rate among them.

More precisely, given a system state (D, c^i, t) and using (7), the rate function to satisfy an individual B_jT_j constraint is $\frac{B_j-D}{f_i(T_j-t)}$. Let $\tilde{r}(D,c,t)$ denote the transmission rate for our proposed policy, then, $\tilde{r}(\cdot)$ is the maximum value among the rates for all B_jT_j constraints for which $(B_j \geq D, T_j \geq t)$.

$$\tilde{r}(D,c^i,t) = \max_{j:(B_j\geq D, T_j\geq t)} \frac{B_j-D}{f_i(T_j-t)} \tag{13}$$

Clearly, by construction, all the B_jT_j constraints are satisfied since at all times we choose the maximum value among rates required to meet each of the remaining constraints. Hence, the policy in (13) is admissible. Furthermore, since the policy in (13) is based on the BT-solution, it inherits the useful properties of that solution. As before, the functions $\{f_i(s)\}_{i=1}^{m}$ can be obtained numerically using a standard ODE solver and this computation needs to be done only once before system operation. Having pre-computed $\{f_i(s)\}$, the online computation is minimal. At time t, the transmitter simply looks at the cumulative data transmitted D and the channel state c^i; then, using the appropriate $f_i(\cdot)$ function it computes the maximum among a set of values as in (13).

The transmission policy in (13) applies for a general Markov channel model, and more specifically as shown below, it is in fact the optimal policy for the constant drift channel model.

Theorem *III:* ***(Variable Deadlines Case)*** *Consider the variable deadlines problem with $g(r) = r^n$, $n > 1, n \in \mathbb{R}$ and a constant drift channel with drift β. The optimal rate, $r^*(D,c,t)$ for $D_{min}(t) \leq D \leq A(t)$, $t \in [0, T_M)$ is given as,*

$$r^*(D,c,t) = \max_{j:(B_j\geq D, T_j\geq t)} \frac{B_j-D}{f(T_j-t)} \tag{14}$$

where $f(s) = \frac{(n-1)}{\lambda(\beta-1)}(1-\exp(-\frac{\lambda(\beta-1)}{n-1}s))$.

Proof: See Appendix A. ■

C. Arrivals with a Single Deadline

In this section, we consider a setup with arrivals as follows. There are M packet arrivals to the queue with the first packet arrival at time $T_0 = 0$ and the rest arriving at times $\{T_j\}_{j=1}^{M-1}$, where $0 < T_1 < T_2 < \ldots < T_{M-1}$. Let b_j be the number of bits in the j^{th} packet arrival. The deadline constraint is that all the data must depart by time $T > T_{M-1}$. This problem has motivations in a sensor network scenario where the data collected at certain times must be transmitted to a central node within a particular time-interval.

In terms of the cumulative curves we have the following picture. Let $A_j = \sum_{l=1}^{j} b_l$; where A_j denotes the cumulative

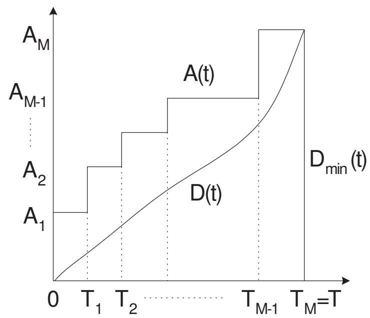

Fig. 4. Cumulative curves for the arrivals with a single deadline case.

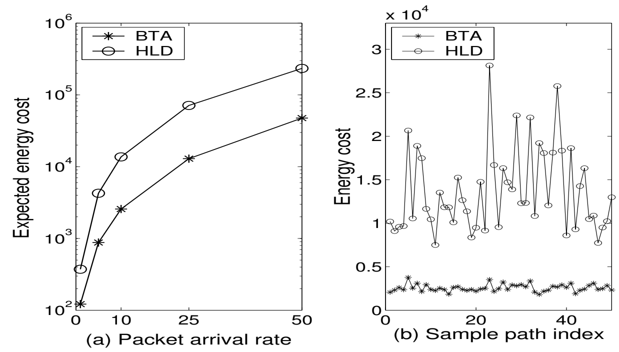

Fig. 5. Energy cost comparison for Poisson arrival process for (a) different arrival rate, (b) different sample paths.

amount of data arrived in the first j packets. Then, $A(t)$ is a piecewise-constant curve with jumps at times T_j as depicted in Figure 4; i.e. at time T_j, $A(T_j) = A_{j+1},\ j = 0, \ldots, M-1$ and $A(T) = A_M$. The minimum departure curve is $D_{min}(t) = 0,\ t \in [0, T); D_{min}(T) = A_M$, since for $t < T$ there is no minimum transmission requirement while at T the entire A_M bits must be transmitted. For convenience we define, $T_M \triangleq T$.

From Figure 4, we see that this cumulative curves picture can be viewed as a "dual" of the variable deadlines case. Earlier, we had constraints from $D_{min}(t)$ but now there are causality constraints from the arrival curve $A(t)$ and a final deadline constraint at time T. Thus, using similar reasoning as in the variable deadlines case, we can obtain a transmission policy as follows. First, note that a constraint $A_jT_j,\ j = 1, \ldots, M-1$ requires that no more than A_j bits must be transmitted before time T_j, while A_MT_M requires that the queue must be empty by time T_M. Starting from a system state (D, c^i, t) and without considering other constraints, emptying the buffer by time T_j (i.e. transmitting A_j bits by time T_j) is equivalent to a BT-problem with $B = A_j$ and $T = T_j$. From (7) the rate for this is given as $\frac{A_j - D}{f_i(T_j - t)}$. Now, to ensure that none of the A_jT_j constraints is violated, i.e. no more than A_j bits is transmitted by time T_j, a natural solution is to choose the minimum rate among them. More precisely, let $\tilde{r}(D, c, t)$ denote the proposed policy we then have,

$$\tilde{r}(D, c^i, t) = \min_{j:(A_j \geq D, T_j \geq t)} \frac{A_j - D}{f_i(T_j - t)} \tag{15}$$

By construction all the causality constraints are satisfied since at all times we choose the minimum rate among those needed to meet the A_jT_j points. Also, for $t > T_{M-1}$, $\tilde{r}(\cdot)$ reduces to choosing a rate that meets the A_MT_M constraint, hence, the deadline constraint is also satisfied. Thus, the policy in (15) is admissible and further, as in the variable deadlines case, it is also optimal for the constant drift channel model.

Theorem *IV: **(Arrivals with Single Deadline)** Consider the "arrivals with a single-deadline" problem with $g(r) = r^n$, $n > 1, n \in \mathbb{R}$ and a constant drift channel with drift β. The optimal rate, $r^*(D, c, t)$, for $D_{min}(t) \leq D \leq A(t),\ t \in [0, T_M)$ is given as,*

$$r^*(D, c, t) = \min_{j:(A_j \geq D, T_j \geq t)} \frac{A_j - D}{f(T_j - t)} \tag{16}$$

where $f(s) = \frac{(n-1)}{\lambda(\beta-1)}(1 - \exp(-\frac{\lambda(\beta-1)}{n-1}s))$.

Proof: The proof is similar to that of Theorem III and is omitted here for brevity. It can be found in [6]. ■

V. General Packet Arrivals with Deadlines

The understanding gained in the last section provides a useful tool for addressing the more general setup involving arbitrary packet arrivals to the queue. In this section, we treat such a setup and present a heuristic energy-efficient policy based on the variable-deadlines solution. We call it the "***BT-Adaptive***" (BTA) policy. We also present simulation results comparing its performance with a non-adaptive scheme.

A. BT-Adaptive (BTA) Policy

Consider an arbitrary stream of packet arrivals to the queue with each packet having a deadline by which it must depart. Assuming that the arrivals occur at discrete times, clearly, at the instant immediately following a packet 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_M of data with variable deadlines, identical to the case considered in Section IV-B. Not assuming any knowledge of the future arrivals and using (13), we have an energy-efficient policy to empty the transmitter buffer. As this policy is followed, at the next packet arrival instant the above procedure is then simply repeated. Summarizing, the BTA policy is as follows:

Transmit the data in the queue with the rate as given in (13); at every packet arrival instant re-arrange the data in the earliest-deadline-first order to obtain a new set of B_jT_j values by including the new packet and its deadline; re-initialize D to zero and follow (13) thereafter.

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

B. Simulation Results

In this section, we present simulation results to evaluate the performance of the BT-Adapative policy. For comparison purposes we consider the "Head-of-Line Drain" (HLD) policy. In HLD policy, the data in the queue is arranged and served in the earliest-deadline-first 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 left until its deadline. The transmission rate at t 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.

The setup is as follows. The queue has Poisson packet arrivals with each packet having 1 unit of data and a deadline of 200 msec. The channel model is a two state model with the two states denoted as the "good" (c^g) and the "bad" (c^b) state. Let λ be the transition rate among the states; we take the following parameter values, $c^g = 1$, $c^b = 0.2$, $\lambda = 50$. Thus, the average time spent in a state before the channel transitions is $1/50$ seconds, or 20 msec. On each simulation run, the total time over which the 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. We take $g(r) = r^2$, hence the energy cost per slot is $\frac{r^2 dt}{c}$. The expected energy cost is obtained as an average of the total cost over the sample runs.

Figure 5(a) is a plot of the expected energy cost, plotted on a log scale, versus the packet arrival rate. As is evident from the plot, the BTA policy has a much lower energy cost compared to the HLD policy and as the arrival rate increases the two costs are roughly an order of magnitude apart. 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 with the BTA policy also adapting the rate with the channel state. As the arrival rate increases and due to the bursty nature of the Poisson process, the queue tends to have more packets. The BTA policy then adapts based on the channel 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. BTA policy has energy gains not just in an average sense but even on individual sample paths. This is shown in Figure 5(b) for 50 sample paths for arrival rate 10 packets/sec.

VI. Conclusion

We considered transmission of delay-constrained data over time-varying channels with the objective of minimizing the total transmission energy expenditure. We adopted a novel approach based on cumulative curves and stochastic control theory to obtain specific optimal solutions for an otherwise difficult set of problems. We first obtained the optimal policy for transmitting B bits of data by deadline T. Using this solution and a graphical decomposition approach, we obtained the optimal solution for the "variable deadlines" and the "arrivals with a single deadline" case for the constant drift channel model. Finally, based on the intuition developed in the above, we devised an energy-efficient policy for arbitrary packet arrivals and deadline constraints, and evaluated its performance through simulations. Various extensions to be explored include a network model with multiple transmitter-receiver pairs and multi-hop transmissions with end-to-end delay constraints.

References

[1] A. Jalali, R. Padovani and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency high data rate personal communication wireless system", *IEEE Vehicular Technology Conf.*, vol. 3, 2000.
[2] S. Nanda, K. Balachandran and S. Kumar, "Adaptation techniques in wireless packet data services", *IEEE Comm. Magazine*, Jan. 2000.
[3] L. Tsaur and D. Lee, "Closed-loop architecture and protocols for rapid dynamic spreading gain adaptation in CDMA networks",*IEEE INFOCOM*, 2004.
[4] M. Zafer and E. Modiano, "A Calculus approach to minimum energy transmission policies with quality of service guarantees", *IEEE INFOCOM 2005*, March 2005.
[5] M. Zafer and E. Modiano, "Continuous-time optimal rate control for delay constrained data transmission", *43rd Allerton conf. on Communication, Control and Computing*, Monticello, Sept. 2005.
[6] M. Zafer, "Dynamic rate-control and scheduling algorithms for quality-of-service in wireless networks", *PhD. Thesis, Massachusetts Institute of Technology*, Sept. 2007. http://web.mit.edu/murtaza/www/
[7] R. Berry and R. Gallager, "Communication over fading channels with delay constraints", *IEEE Tran. on Information Theory*, May 2002.
[8] E. Uysal-Biyikoglu and A. El Gamal, "On Adaptive transmission for energy-efficiency in wireless data networks," in *IEEE Trans. Information Theory*, December 2004.
[9] A. Fu, E. Modiano and J. Tsitsiklis, "Optimal energy allocation for delay constrained data transmission over a time-varying channel", *IEEE INFOCOM 2003*, vol. 2, pp 1095 - 1105, April 2003.
[10] M. Khojastepour and A. Sabharwal, "Delay-constrained scheduling: power efficiency, filter design and bounds", *IEEE INFOCOM*, March 2004.
[11] M. Neely, E. Modiano and C. Rohrs, "Dynamic power allocation and routing for time varying wireless networks," *IEEE INFOCOM*, 2003.
[12] A. Goldsmith, "The capacity of downlink fading channels with variable rate and power", *IEEE Tran. Vehicular Technology*, vol. 45, Oct. 1997.
[13] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks", *IEEE Tran. on Automatic Control*, Dec. 1992.
[14] X. Liu, E. Chong and N. Shroff, "A framework for opportunistic scheduling in wireless networks " *Computer Networks*, vol. 41, 2003.
[15] B. Collins and R. Cruz, "Transmission policies for time varying channels with average delay constraints", *Proceedings, 1999 Allerton conf. on Communication, Control and Computing*, Monticello, IL, 1999.
[16] B. Ata, "Dynamic power control in a wireless static channel subject to a quality of service constraint", *Operations Research* 53, 2005.
[17] F. Babich and G. Lombardi, "A Markov model for the mobile propagation channel", *IEEE Tran. on Vehicular Technology*, vol. 49, no. 1, pp. 63-73, Jan. 2000.
[18] E. N. Gilbert, "Capacity of burst-noise channel", *Bell Syst. Tech. J.*,vol. 39, no. 9, pp. 1253-1265, Sept. 1960.
[19] A. Goldsmith and P. Varaiya, "Capacity, mutual Information and coding for finite state markov channels", *IEEE Trans. Information Theory*, 1996.
[20] J. Le Boudec, P. Thiran, *Network calculus*, Springer-Verlag, 2001.
[21] W. Fleming and H. Soner, *Controlled Markov Processes and Viscosity Solutions*, Springer-Verlag, 1993.
[22] M. Davis, *Markov Models and Optimization*, Chapman and Hall, 1993.
[23] B. Oksendal, *Stochastic Differential Equations*, Springer, 5^{th} edn., 2000.

Appendix A
Proof of Theorem III – *Variable Deadlines Setup*

The proof outline is as follows. We start with the functional form for $r^*(D, c, t)$ as given in (14), obtain the minimum cost function $J(D, c, t)$ and check that these satisfy the PDE equation in (6). The admissibility of $r^*(\cdot)$ has already been discussed in Section IV-B. For brevity and simplicity, we only treat the two-packet case $M = 2$; full details of the proof can be found in [6].

Two Packet Case: Consider the rate function in (14); if this is the optimal policy then for every admissible system state (D,c,t) we must have $r^*(D,c,t)$ as the minimizing value of r in (6). Using the first-order condition for the minimization, this gives, $\frac{\partial J(D,c,t)}{\partial D} = -\frac{g'(r^*(D,c,t))}{c}$, which by integrating with respect to D gives the $J(\cdot)$ function. After obtaining the $J(\cdot)$ function, we verify optimality by proving that $r^*(\cdot)$ and $J(\cdot)$ satisfy the PDE equation, i.e.,

$$\Big\{\frac{g(r^*(D,c,t))}{c} + \frac{\partial J(D,c,t)}{\partial t} + r^*(D,c,t)\frac{\partial J(D,c,t)}{\partial D} + \lambda(E_z[J(D,Z(c)c,t)] - J(D,c,t))\Big\} = 0 \quad (17)$$

To proceed, first consider the state space $(D,c,t) \in [B_1,B_2]\times\mathcal{C}\times[T_1,T_2)$ – that is, we are looking at time $t \geq T_1$ and all admissible D values over this time. Starting from (D,c,t) in this state space, clearly, the problem is identical to the BT-problem with $B = (B_2 - D)$ and $T = (T_2 - t)$. From (11), the optimal rate function must be $r^*(D,c,t) = \frac{B_2-D}{f(T_2-t)}$. In conformation, the rate function in (14) over this state space also reduces to the same form. Thus, over this state space, (14) is trivially the optimal policy.

Next consider the state space $(D,c,t) \in [0,B_2]\times\mathcal{C}\times[0,T_1)$; thus now we are considering $0 \leq t < T_1$ and all admissible D values over this time which are $[0,B_2]$. Fix a value of t and c, then, as a function of D the rate $r^*(\cdot)$ in (14) has the following two possibilities.

Case 1: Suppose $\frac{B_2}{f(T_2-t)} > \frac{B_1}{f(T_1-t)}$. For a fixed t, we see that both $\frac{B_1-D}{f(T_1-t)}$ and $\frac{B_2-D}{f(T_2-t)}$ are linear in D. Figure 6(a) gives a schematic picture of the two curves and from the figure it is clear that since $B_2 > B_1$, the two curves do not intersect over $D \in [0,B_1]$. Thus, in this case the maximizing function for all $D \in [0,B_2]$ is $\frac{B_2-D}{f(T_2-t)}$ and so, $r^*(D,c,t) = \frac{B_2-D}{f(T_2-t)}$.

Case 2: Suppose $\frac{B_2}{f(T_2-t)} \leq \frac{B_1}{f(T_1-t)}$. In this case, the two functions $\frac{B_1-D}{f(T_1-t)}$ and $\frac{B_2-D}{f(T_2-t)}$ are plotted in Figure 6(b). From the figure it is clear that since $B_1 < B_2$ the two curves must intersect at some $\tilde{B} \in [0,B_1]$ which satisfies $\frac{B_1-\tilde{B}}{f(T_1-t)} = \frac{B_2-\tilde{B}}{f(T_2-t)}$. Thus, in this case we get $r^*(D,c,t) = \frac{B_1-D}{f(T_1-t)}$ for $D \in [0,\tilde{B}]$ and $r^*(D,c,t) = \frac{B_2-D}{f(T_2-t)}$ for $D \in [\tilde{B},B_2]$. Define,

$$\tilde{B}(t) \triangleq \begin{cases} 0, & \text{if } \frac{B_1}{f(T_1-t)} < \frac{B_2}{f(T_2-t)} \\ \frac{\frac{B_1}{f(T_1-t)} - \frac{B_2}{f(T_2-t)}}{\frac{1}{f(T_1-t)} - \frac{1}{f(T_2-t)}}, & \text{otherwise} \end{cases} \quad (18)$$

Using (18), $r^*(\cdot)$ can be written in a more compact form as,

$$r^*(D,c,t) = \begin{cases} \frac{B_2-D}{f(T_2-t)}, & \tilde{B}(t) \leq D \leq B_2 \\ \frac{B_1-D}{f(T_1-t)}, & 0 \leq D < \tilde{B}(t) \end{cases} \quad (19)$$

Using $\frac{\partial J(D,c,t)}{\partial D} = -\frac{g'(r^*(D,c,t))}{c}$ (note $g(r) = r^n$) and integrating with respect to D with the boundary condition $J(B_2,c,t) = 0$, we obtain,

$$J(D,c,t) = \begin{cases} \frac{(B_2-D)^n}{c(f(T_2-t))^{n-1}}, & \tilde{B}(t) \leq D \leq B_2 \\ \frac{(B_1-D)^n}{c(f(T_1-t))^{n-1}} + \frac{(B_2-\tilde{B}(t))^n}{c(f(T_2-t))^{n-1}} \\ -\frac{(B_1-\tilde{B}(t))^n}{c(f(T_1-t))^{n-1}}, & 0 \leq D < \tilde{B}(t) \end{cases} \quad (20)$$

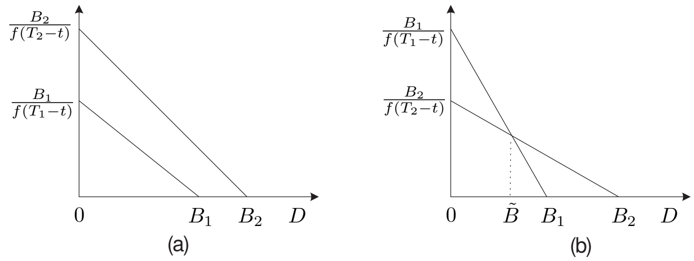

Fig. 6. Proof of Theorem III for the two packet case, (a) case $\frac{B_2}{f(T_2-t)} > \frac{B_1}{f(T_1-t)}$ and (b) case $\frac{B_2}{f(T_2-t)} \leq \frac{B_1}{f(T_1-t)}$.

We now show that (19) and (20) satisfy the optimality PDE equation in (17). Consider first $D \in [\tilde{B}(t),B_2]$, then, from (20) we have $J(D,c,t) = \frac{(B_2-D)^n}{c(f(T_2-t))^{n-1}}$ and from (19) we have $r^*(D,c,t) = \frac{B_2-D}{f(T_2-t)}$. Substituting in the left hand side (LHS) of (17) gives (let $s = T_2 - t$),

$$LHS = \frac{(n-1)(B_2-D)^n}{c(f(s))^n}\left(f'(s) - 1 + \frac{\lambda(\beta-1)}{n-1}f(s)\right) = 0, \quad (\text{since, } f'(s) = 1 - \tfrac{\lambda(\beta-1)}{n-1}f(s)) \quad (21)$$

Thus from above, (17) is satisfied over $D \in [\tilde{B}(t),B_2]$. If $\tilde{B}(t) = 0$, we are done. So, now suppose $\tilde{B}(t) > 0$.

Consider $D \in [0,\tilde{B}(t))$, then, from (19) we have $r^*(D,c,t) = \frac{B_1-D}{f(T_1-t)}$ and from (20) we have $J(D,c,t) = Q(c,t) + H(D,c,t)$, where for simplicity of exposition we define $Q(c,t) \triangleq \left(\frac{(B_2-\tilde{B}(t))^n}{c(f(T_2-t))^{n-1}} - \frac{(B_1-\tilde{B}(t))^n}{c(f(T_1-t))^{n-1}}\right)$ and $H(D,c,t) = \frac{(B_1-D)^n}{c(f(T_1-t))^{n-1}}$. Substituting in (17) gives,

$$LHS = \left(\frac{\partial Q(c,t)}{\partial t} + \lambda(E_z[Q(Z(c)c,t)] - Q(c,t))\right) + \Big\{\frac{g(r^*(D,c,t))}{c} + \frac{\partial H(D,c,t)}{\partial t} + r^*(D,c,t)\frac{\partial H(D,c,t)}{\partial D} + \lambda(E_z[H(D,Z(c)c,t)] - H(D,c,t))\Big\}$$

Using identical steps that lead to (21), it can be shown that the terms within the curly bracket above equal zero. Now consider the first-bracket terms. Let $Q(c,t) = Q_2(c,t) - Q_1(c,t)$, where $Q_2(c,t) = \frac{(B_2-\tilde{B}(t))^n}{c(f(T_2-t))^{n-1}}$ and $Q_1(c,t) = \frac{(B_1-\tilde{B}(t))^n}{c(f(T_1-t))^{n-1}}$. We have,

$$\frac{\partial Q_2(c,t)}{\partial t} + \lambda(E_z[Q_2(Z(c)c,t)] - Q_2(c,t)) = \frac{(n-1)(B_2-\tilde{B}(t))^n}{c(f(T_2-t))^n}\left(-\frac{\tilde{B}'(t)f(T_2-t)n}{(B_2-\tilde{B}(t))(n-1)} + 1\right)$$

A similar expression as above is obtained for the term $Q_1(c,t)$. Combining the two and using $\frac{B_1-\tilde{B}(t)}{f(T_1-t)} = \frac{B_2-\tilde{B}(t)}{f(T_2-t)}$, gives,

$$\frac{\partial Q(c,t)}{\partial t} + \lambda(E_z[Q(Z(c)c,t)] - Q(c,t)) = 0$$

This completes the verification that the functions in (19) and (20) satisfy the PDE equation in (17), and, thus proves that they form the optimal solution.