Dynamic Data Compression for Wireless Transmission over a Fading Channel

Michael J. Neely
University of Southern California
http://www-rcf.usc.edu/~mjneely

Abstract—We consider a wireless node that randomly receives data from different sensor units. The arriving data must be compressed, stored, and transmitted over a wireless link, where both the compression and transmission operations consume power. Specifically, the controller must choose from one of multiple compression options every timeslot. Each option requires a different amount of power and has different compression ratio properties. Further, the wireless link has potentially time-varying channels, and transmission rates depend on current channel states and transmission power allocations. We design a dynamic algorithm for joint compression and transmission, and prove that it comes arbitrarily close to minimizing average power expenditure, with an explicit tradeoff in average delay. The algorithm is simple to implement and does not require knowledge of probability distributions for packet arrivals or channel states.
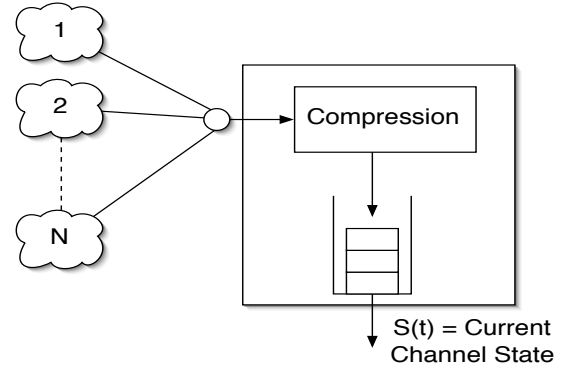

Fig. 1. Multiple sensors sending data to a single wireless link.

I. INTRODUCTION

We consider the problem of energy-aware data compression and transmission for a wireless link that receives data from N different sensor units (Fig. 1). Time is slotted with normalized slot durations $t \in \{0, 1, 2, \ldots\}$, and every timeslot the link receives a packet from a random number of the sensors. We assume that packets arriving on the same timeslot contain correlated data, and that this data can be compressed using one of multiple compression options. However, the signal processing required for compression consumes a significant amount of energy, and more sophisticated compression algorithms are also more energy expensive. Further, the data must be transmitted over a wireless channel with potentially varying channel conditions, where the transmission rates available on the current timeslot depend on the current channel condition and the current transmission power allocation. The goal is to design a joint compression and transmission scheduling policy that minimizes time average power expenditure.

This problem is important for modern sensor networks where correlated (and compressible) data flows over power limited nodes. Compressing the data can save power by reducing the amount of bits that need to be transmitted, provided that the transmission power saved is more than the power expended in the compression operation. It is important to understand the optimal balance between compression power and transmission power. Work in [1] considers this question for a wireless link with fixed transmission costs, and describes practical compression issues and reports communication-to-computation energy ratios for popular algorithms. Work in

[2] considers a similar static situation where the wireless channel condition is the same for all time. There, it is shown experimentally that compression can lead to a significant power savings when data is transmitted over multiple hops. The proposed algorithm of [2] uses a fixed data compression scheme, an adaptation of the Lempel-Ziv-Welch (LZW) compression algorithm for sensor networks. Techniques for distributed compression using Slepian-Wolf coding theory are considered in [3] [4]. Models of spatial correlation between data of different sensors are proposed in [5] and used to construct and evaluate energy-efficient routing algorithms that compress data at each stage.

The above prior work has concentrated on static environments where transmission power is directly proportional to the number of bits transmitted and/or traffic rates are fixed and known, so that compression and transmission strategies can be designed in advance. Here, we focus attention on a single link, but consider a stochastic environment where the amount of data received every slot is random, as is the current channel condition for wireless transmission. Further, the transmission rate is an arbitrary (possibly non-linear) function of transmission power. Optimal policies in this stochastic context are more complex, and more care is required to ensure transmissions are energy-efficient.

In this paper, we design a dynamic compression and transmission scheduling algorithm and prove that the algorithm pushes total time average power arbitrarily close to optimal, with a corresponding tradeoff in average delay. We assume the algorithm has a table of expected compression ratios for each compression option, and that, if channels are time-varying, current channel state information is available. Our algorithm bases decisions purely on this information and does

not require a-priori knowledge of the packet arrival or channel state probabilities. The algorithm is simple to implement and is robust to situations where these probabilities can change. This work is important as it demonstrates a principled method of making on-line compression decisions in a stochastic system with correlated data. Our solution applies the techniques of Lyapunov optimization developed in our previous work [6] [7], and is perhaps the first application of these techniques to the dynamic compression problem.

In the next section we describe the system model, and in Section III we characterize the minimum average power in terms of an optimization problem based on channel and packet arrival probabilities. In Section IV we develop an on-line algorithm that makes simple decisions based only on current information. We prove the algorithm achieves average power that can be pushed arbitrarily close to optimum via a simple control parameter that also affects an average delay tradeoff.

II. SYSTEM MODEL

Consider the wireless link of Fig. 1 that operates in slotted time and receives packets from N different sensor units. If an individual sensor sends data during a timeslot, this data is in the form of a fixed length packet of size b bits, containing sensed information. Let $A(t)$ represent the number of sensors that send packets during slot t, so that $A(t) \in \{0, 1, \ldots, N\}$. The data from these $A(t)$ packets may be correlated, and hence it may be possible to compress the information within the $A(t)$ packets (consisting of $A(t)b$ bits) into a smaller data unit for transmission over the wireless link. This is done via a *compression function* $\Psi(a, k)$ defined as follows. There are $K + 1$ compression options, comprising a set $\mathcal{K} = \{0, 1, \ldots, K\}$. Option 0 represents no attempted compression, and options $\{1, 2, \ldots, K\}$ represent various alternative methods to compress the data. The function $\Psi(a, k)$ takes input $a \in \{0, 1, \ldots, N\}$ (representing the number of newly arriving packets) and compression option $k \in \mathcal{K}$, and generates a *random variable* output R, representing the total size of the data after compression.

Every timeslot the link controller observes the random number of new packet arrivals $A(t)$ and chooses a compression option $k(t) \in \mathcal{K}$, yielding the random compressed output $R(t) = \Psi(A(t), k(t))$. Let $P_{comp}(t)$ represent the power expended by this compression operation, and assume this is also a random function of the number of packets compressed and the compression option. The average compressed output $m(a, k)$ and the average power expenditure $\phi(a, k)$ associated with $A(t) = a, k(t) = k$ are defined:

$$m(a, k) = \mathbb{E}\{\Psi(A(t), k(t)) \mid A(t) = a, k(t) = k\} \quad (1)$$
$$\phi(a, k) = \mathbb{E}\{P_{comp}(t) \mid A(t) = a, k(t) = k\} \quad (2)$$

We assume the values of $m(a, k)$ and $\phi(a, k)$ are known so that a table such as given in Fig. 2 can be constructed. We assume $\Psi(a, 0) = ab$ and $\phi(a, 0) = 0$, as the compression option $k = 0$ does not compress any data and also does not expend any power. We assume that $m(a, k) \leq ab$ for all $a \in \{0, 1, \ldots, N\}$ and all $k \in \mathcal{K}$, so that compression is not expected to expand the data.

k	$\Psi(a, k)$	$\mathbb{E}\{\Psi(a, k)\}$	$\mathbb{E}\{P_{comp} \mid a, k\}$
0	ab	ab	$\phi(a, 0) = 0$
1	Random	$m(a, 1)$	$\phi(a, 1)$
2	Random	$m(a, 2)$	$\phi(a, 2)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
K	Random	$m(a, K)$	$\phi(a, K)$

Fig. 2.   An example table of expected compression outcomes.

A. Data Transmission and Queueing

The compressed data $R(t) = \Psi(A(t), k(t))$ is delivered to a queueing buffer for transmission over the wireless link (see Fig. 1). Let $U(t)$ represent the current number of bits (or *unfinished work*) in the queue. The queue backlog evolution is given by:

$$U(t + 1) = \max[U(t) - \mu(t), 0] + R(t) \quad (3)$$

where $\mu(t)$ is the transmission rate offered by the link on slot t. This rate is determined by the current channel condition and the current transmission power allocation decision, as in [7]. Specifically, the channel is assumed to be constant over the duration of a slot, but can potentially change from slot to slot. Let $S(t)$ represent the current channel state, which is assumed to take values in some finite set $\mathcal{S}$. We assume the channel state $S(t)$ is known at the beginning of each slot t, so that the link can make an opportunistic transmission power allocation decision $P_{tran}(t)$, yielding a transmission rate $\mu(t)$ given by:

$$\mu(t) = C(P_{tran}(t), S(t))$$

where $C(P, s)$ is the *rate-power* curve associated with the modulation and coding schemes used for transmission over the channel. We assume $C(P, s)$ is continuous in power P for each channel state $s \in \mathcal{S}$. Transmission power allocations $P(t)$ are restricted to some compact set $\mathcal{P}$ for all slots t, where $\mathcal{P}$ contains a maximum transmission power P_{max}. For example, the set $\mathcal{P}$ can contain a discrete set of power levels, such as the two element set $\mathcal{P} = \{0, P_{max}\}$. Alternatively, $\mathcal{P}$ can be a continuous interval, such as $\mathcal{P} = \{P \mid 0 \leq P \leq P_{max}\}$. We assume throughout that $0 \in \mathcal{P}$ and that $C(0, s) = 0$ for all channel states $s \in \mathcal{S}$, so that zero transmission power yields a zero transmission rate. Further, we assume that $C(P_{max}, s) \geq C(P, s)$ for all $s \in \mathcal{S}$ and all $P \in \mathcal{P}$, so that allocating maximum power yields the largest transmission rate that is possible under the given channel state.

B. Stochastic Assumptions and the Control Objective

For simplicity, we assume the packet arrival process $A(t)$ is i.i.d. over slots with a general probability distribution $p_A(a) = Pr[A(t) = a]$. Likewise, the channel state process $S(t)$ is i.i.d. over slots with a general distribution $\pi_s = Pr[S(t) = s]$.[1] The distributions $p_A(a)$ and π_s are not necessarily known to the link controller. Every slot the link controller observes the

[1]Using the T-slot Lyapunov drift techniques described in [7], our analysis can be generalized to show that the *same* algorithms we derive under the i.i.d. assumption yield similar performance for arbitrary ergodic arrival and channel processes, with delay bounds that increase by a constant factor related to the mixing times of the processes.

number of new packets $A(t)$, the current queue backlog $U(t)$, and the current channel state $S(t)$, and makes a compression decision $k(t) \in \mathcal{K}$ (expending power $P_{comp}(t)$) and a transmission power allocation $P_{tran}(t) \in \mathcal{P}$. The total time average power expenditure is given by:

$$\lim_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} [P_{comp}(\tau) + P_{tran}(\tau)]$$

The goal is to make compression and transmission decisions to minimize time average power while ensuring the queue $U(t)$ is stable, so that average backlog is finite.

Define r_{min} and r_{max} as follows:

$$r_{min} \triangleq \mathbb{E}\left\{\min_{k \in \mathcal{K}} m(A(t), k)\right\} \tag{4}$$

$$r_{max} \triangleq \mathbb{E}\{C(P_{max}, S(t))\} \tag{5}$$

where the expectations are taken over the randomness of $A(t)$ and $S(t)$ via the distributions $p_A(a)$ and π_s. Thus, r_{min} is the minimum average bit rate delivered to the queueing system (in units of bits/slot), assuming the compression option that results in the largest expected bit reduction is used every slot. The value r_{max} represents the maximum possible average transmission rate over the wireless link. We assume throughout that $r_{min} < r_{max}$, so that it is possible to stabilize the system.

Thus, there are two reasons to compress data: (i) In order to stabilize the queue, we may need to compress (particularly if $\mathbb{E}\{A(t)\} b > r_{max}$). (ii) We may actually save power if the power used to compress is less than the extra amount of power that would be used transmitting the extra data if it were not compressed.

C. Discussion of the System Model

This simple model captures a wide class of systems where data compression is important. The N sensor scenario of Fig. 1 captures the possibility of randomly arriving data that is *spatially correlated*. An example is when there are multiple sensors in an environment and only a random subset of them detect a particular event. The data provided by these sensors is thus correlated but not necessarily identical, as each observation can offer new information. The case of compression because of *time correlated* data can also be treated in this model by re-defining N to represent the time over which a frame of data samples are gathered, and interpreting $A(t)$ as the random number of packets arriving over the N mini-slots. This of course assumes compression is contained to data arriving within the same frame.

Our time-varying channel model is useful for systems with mobility, environmental changes, or restrictions that create time-varying transmission opportunities. This allows for opportunistic scheduling which can help to further reduce power expenditure. We do not consider the additional power required to measure the channel conditions here. Extensions that treat this issue can likely be obtained using the techniques for optimizing measurement decisions developed in [8]. A special case of the time-varying channel model is the *static channel* assumption, where $S(t)$ is the same for all timeslots t. This special case is similar to the static assumption in [1] [2].

However, this static channel scenario still creates an interesting problem that is much different from [1] [2]. Indeed, the random packet arrivals (with raw data rate that is possibly larger than link capacity) and the potentially non-linear rate-power curve necessitate a dynamic compression strategy that is not obvious, that depends on the packet arrival distribution, and that does not necessarily use the same compression option on every slot.

Here we assume that the compression options available within the set $\mathcal{K}$ are sufficient to ensure that the resulting data transmitted over the link has an acceptable fidelity. An example is *lossless* data compression, such as Huffman or Lempel-Ziv source coding, where all original data packets can be reconstructed at the destination. Alternatively, we might have some compression options $k \in \mathcal{K}$ representing *lossy* compression, provided that the distortion that may be introduced is acceptable.

III. MINIMUM AVERAGE POWER

Here we characterize the minimum time average power required for queue stability. We first define separate functions $h^*(r)$ and $g^*(r)$ that describe the minimum average power for compression and transmission, respectively, over a restricted class of stationary randomized algorithms. These functions depend on the steady state arrival and channel distributions $p_A(a)$ and π_s. We then show that these functions can be used to define system optimality over the class of all possible decision strategies, including strategies that do not necessarily make stationary and randomized decisions.

A. The Functions $h^*(r)$ and $g^*(r)$

Definition 1: For any value r such that $r_{min} \leq r \leq b\mathbb{E}\{A(t)\}$, the *minimum-power compression function* $h^*(r)$ is defined as the *infimum value* h for which there exist probabilities $(\gamma_{a,k})$ for $a \in \{0, 1, \ldots, N\}$, $k \in \mathcal{K}$, such that the following constraints are satisfied:

$$\sum_{a=0}^{N} \sum_{k=1}^{K} p_A(a) \gamma_{a,k} \phi(a, k) = h \tag{6}$$

$$\sum_{a=0}^{N} \sum_{k=1}^{K} p_A(a) \gamma_{a,k} m(a, k) \leq r \tag{7}$$

$$\gamma_{a,k} \geq 0 \quad \text{for all } a, k \tag{8}$$

$$\sum_{k=1}^{K} \gamma_{a,k} = 1 \quad \text{for all } a \tag{9}$$

Intuitively, the $(\gamma_{a,k})$ values define a stationary randomized policy that observes the current arrivals $A(t)$ and uses compression option k with probability $\gamma_{a,k}$ whenever $A(t) = a$. The expression on the left hand side of (6) is the expected compression power $\mathbb{E}\{P_{comp}(t)\}$ for this policy. Likewise, the expression on the left hand side of (7) is the expected number of bits $\mathbb{E}\{R(t)\}$ at the output of the compressor for this policy. The value of $h^*(r)$ is thus the smallest possible average power due to compression, infimized over all such stationary randomized policies that yield $\mathbb{E}\{R(t)\} \leq r$. Note from (4) that it is possible to have a stationary randomized

policy that yields $\mathbb{E}\{R(t)\} = r_{min}$, and hence the function $h^*(r)$ is well defined for any $r \geq r_{min}$. Further, the following lemma shows that the infimum value $h^*(r)$ can be *achieved* by a particular stationary randomized algorithm.

Lemma 1: For any r such that $r_{min} \leq r \leq b\mathbb{E}\{A(t)\}$, there exists a particular stationary randomized policy that makes compression decisions $k^*(t)$ as a random function of the observed $A(t)$ value (and independent of queue backlog), such that:

$$\mathbb{E}\{\phi(A(t), k^*(t))\} = h^*(r) \qquad (10)$$
$$\mathbb{E}\{m(A(t), k^*(t))\} = r \qquad (11)$$

where the above expectations are taken with respect to the steady state packet arrival distribution $p_A(a)$ and the randomized compression decisions $k^*(t)$.

Proof: Omitted for brevity. ∎

Similar to the function $h^*(r)$, we define $g^*(r)$ as the smallest possible average transmission power required for a stationary randomized algorithm to support a transmission rate of at least r. The precise definition is given below.

Definition 2: For any value r such that $0 \leq r \leq r_{max}$, the *minimum-power transmission function* $g^*(r)$ is defined as the *infimum value* g for which there is a stationary randomized power allocation policy that chooses transmission power $P_{tran}(t)$ as a random function of the observed channel state $S(t)$ (and independent of current queue backlog), such that:

$$\mathbb{E}\{P_{tran}(t)\} = g \qquad (12)$$
$$\mathbb{E}\{C(P_{tran}(t), S(t))\} \geq r \qquad (13)$$

The function $g^*(r)$ is well defined whenever $r \leq r_{max}$ because it is possible to satisfy the constraint (13). Indeed, note by (5) that the policy $P_{tran}(t) = P_{max}$ for all t yields $\mathbb{E}\{C(P_{tran}(t), S(t))\} = r_{max}$. Furthermore, it is easy to show that the inequality constraint in (13) can be replaced by an equality constraint. Because the set $\mathcal{P}$ is compact and the function $C(P, s)$ is continuous in power P for all channel states $s \in \mathcal{S}$, it can be shown that the infimum average power $g^*(r)$ can be *achieved* by a particular stationary randomized policy. Specifically, for any r such that $0 \leq r \leq r_{max}$, there exists a stationary randomized algorithm that chooses transmission power $P^*_{tran}(t)$ that yields:

$$\mathbb{E}\{C(P^*_{tran}(t), S(t))\} = r \qquad (14)$$
$$\mathbb{E}\{P^*_{tran}(t)\} = g^*(r) \qquad (15)$$

B. Structural Properties of $h^*(r)$ and $g^*(r)$

It is not difficult to show that $h^*(r)$ is a non-increasing function of r (because less compression power is required if a larger compressor output rate is allowed), and that $g^*(r)$ is a non-decreasing function of r (because more transmission power is required to support a larger transmission rate). Further, both functions are convex. It is interesting to note that in the special case when there is no channel state variation so that $C(P, s) = C(P)$, and when the function $C(P)$ is strictly increasing and concave, then $g^*(r) = C^{-1}(r)$, i.e., it is given by the inverse of $C(P)$. Details on the structure of the $g^*(r)$ function in the general time-varying case are given in [6].

C. Minimum Average Power for Stability

The following theorem establishes the minimum time average power required for queue stability in terms of the $h^*(r)$ and $g^*(r)$ functions. We consider all possible algorithms for making compression decisions $k(t) \in \mathcal{K}$ and transmission power decisions $P_{tran}(t) \in \mathcal{P}$ over time, including algorithms that are not necessarily in the class of stationary randomized policies.

Theorem 1: Let $A(t)$ and $S(t)$ be ergodic with steady state distributions $p_A(a)$ and π_s, respectively (such as processes that are i.i.d. over slots, or more general Markov modulated processes). Assume that $r_{min} < r_{max}$ (defined in (4), (5)). Then any joint compression and transmission rate scheduling algorithm that stabilizes the queue $U(t)$ yields a time average power expenditure that satisfies:

$$\limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\{P_{comp}(\tau) + P_{tran}(\tau)\} \geq P^*_{av}$$

where P^*_{av} is defined as the optimal solution to the following problem:

$$\text{Minimize:} \qquad h^*(r) + g^*(r) \qquad (16)$$
$$\text{Subject to:} \quad r_{min} \leq r \leq \min[r_{max}, b\mathbb{E}\{A(t)\}] \qquad (17)$$

Proof: Omitted for brevity. ∎

The above theorem shows that time average power must be greater than or equal to P^*_{av} for queue stability. The result can be understood intuitively by observing that if r is the rate of bits arriving to the queue from the compressor, then average transmission power can be minimized while maintaining stability by pushing the time average transmission rate down closer and closer to r. The optimization problem corresponding to this definition of P^*_{av} may be difficult to solve in practice, as it would require exact knowledge of the $h^*(r)$ and $g^*(r)$ functions, which in turn requires full a-priori knowledge of the distributions $p_A(a)$ and π_s. In the next section, we design a simple class of dynamic algorithms that stabilize the queue without this knowledge, and that push time average power arbitrarily close to P^*_{av}.

IV. THE DYNAMIC COMPRESSION ALGORITHM

Our dynamic algorithm is decoupled into separate policies for data compression and transmission rate scheduling. It is defined in terms of a control parameter $V > 0$ that affects an energy-delay tradeoff.

The Dynamic Compression and Transmission Algorithm:

Compression: Every slot t, observe the number of new packet arrivals $A(t)$ and the current queue backlog $U(t)$, and choose compression option $k(t) \in \mathcal{K}$ as follows:

$$k(t) = \arg\min_{k \in \mathcal{K}} [U(t)m(A(t), k) + V\phi(A(t), k)]$$

If there are multiple compression options $k \in \mathcal{K}$ that minimize $U(t)m(A(t), k) + V\phi(A(t), k)$, break ties arbitrarily.

Transmission: Every slot t, observe the current channel state $S(t)$ and the current queue backlog $U(t)$, and choose a transmission power $P_{tran}(t) \in \mathcal{P}$ as follows:

$$P_{tran}(t) = \arg\max_{P \in \mathcal{P}} [U(t)C(P, S(t)) - VP]$$

The compression policy involves a simple comparison of $K + 1$ values found by evaluating the $m(a, k)$ and $\phi(a, k)$ functions for all $k \in \mathcal{K}$, and can easily be accomplished in real time. The transmission policy is a special case of the Energy Efficient Control Algorithm (EECA) developed in [6], and typically can also be solved very simply in real time.

Theorem 2: (Algorithm Performance) Suppose packet arrivals $A(t)$ are i.i.d. over slots with distribution $p_A(a)$, and channel states $S(t)$ and are i.i.d. over slots with distribution π_s. For any control parameter $V > 0$, the dynamic compression and transmission scheduling algorithm yields power expenditure and queue backlog that satisfy the following:

$$
\begin{aligned}
\overline{P}_{tot} &\leq P_{av}^* + B/V & (18) \\
\overline{U} &\leq \frac{B + V(P_{max} + \phi_{max})}{(r_{max} - r_{min})} & (19)
\end{aligned}
$$

where $\overline{P}_{tot}$ and $\overline{U}$ are the time averages for power expenditure and queue backlog, defined:

$$
\begin{aligned}
\overline{P}_{tot} &\triangleq \limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\left\{P_{comp}(\tau) + P_{tran}(\tau)\right\} \\
\overline{U} &\triangleq \limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\left\{U(\tau)\right\}
\end{aligned}
$$

and where B and ϕ_{max} are constants given by:

$$
\begin{aligned}
B &\triangleq \frac{1}{2}\left[\sigma^2 + \mathbb{E}\left\{C(P_{max}, S(t))^2\right\}\right] & (20) \\
\phi_{max} &\triangleq \mathbb{E}\left\{\max_{k \in \mathcal{K}} \left[\phi(A(t), k)\right]\right\} & (21)
\end{aligned}
$$

where σ^2 is an upper bound on $\mathbb{E}\left\{R(t)^2\right\}$ for all slots t. For example, if no compression operation expands the data, then $R(t) = \Psi(A(t), k) \leq bA(t)$ for all t, and hence σ^2 is defined:

$$
\sigma^2 \triangleq b^2 \mathbb{E}\left\{A(t)^2\right\}
$$

We prove Theorem 2 in the next subsection. Note that the parameter $V > 0$ can be chosen to make B/V arbitrarily small, ensuring by (18) that time average power is arbitrarily close to the optimal value P_{av}^*. However, the resulting average queue backlog bound grows linearly with V. By Little's Theorem, the average queue backlog is proportional to average delay [9]. This establishes an explicit tradeoff between average power expenditure and average delay.

A. Lyapunov Performance Analysis for Theorem 2

Our proof relies on the performance optimal Lyapunov scheduling techniques from [7] [6]. First define the following quadratic Lyapunov function of queue backlog $U(t)$:

$$
L(U(t)) \triangleq \frac{1}{2} U(t)^2
$$

Define the one-step conditional Lyapunov drift $\Delta(U(t))$ as follows:

$$
\Delta(U(t)) \triangleq \mathbb{E}\left\{L(U(t+1)) - L(U(t)) \mid U(t)\right\} \quad (22)
$$

The following simple lemma from [7] shall be useful.

Lemma 2: (Lyapunov drift [7]) Let $L(U(t))$ be a non-negative function of $U(t)$ with Lyapunov drift $\Delta(U(t))$ defined in (22). If there are stochastic processes $\alpha(t)$ and $\beta(t)$ such that every slot t and for all possible values of $U(t)$, the conditional Lyapunov drift satisfies:

$$
\Delta(U(t)) \leq \mathbb{E}\left\{\beta(t) - \alpha(t) \mid U(t)\right\} \quad (23)
$$

then:

$$
\limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\left\{\alpha(\tau)\right\} \leq \limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}\left\{\beta(\tau)\right\}
$$

The proof involves taking expectations of (23), using iterated expectations, and summing the resulting telescoping series (see [7] for details).

The queue backlog $U(t)$ for our system satisfies the queue evolution equation (3). The Lyapunov drift is given by the following lemma.

Lemma 3: (Computing $\Delta(U(t))$) Under the queue evolution equation (3) and using the quadratic Lyapunov function $L(U(t)) = \frac{1}{2} U(t)^2$, the Lyapunov drift $\Delta(U(t))$ satisfies the following for all t and all $U(t)$:

$$
\Delta(U(t)) \leq B - U(t)\mathbb{E}\left\{\mu(t) - m(A(t), k(t)) \mid U(t)\right\} \quad (24)
$$

where $\mu(t) = C(P_{tran}(t), S(t))$, and B is given in (20). The expectation above is taken with respect to the random channels and arrivals $S(t)$ and $A(t)$, and the potentially random control actions $k(t)$ and $P_{tran}(t)$.

Proof: From (3) we have:

$$
\begin{aligned}
\frac{1}{2}U(t+1)^2 &= \frac{1}{2}\left(\max[U(t) - \mu(t), 0] + R(t)\right)^2 \\
&\leq \frac{1}{2}\left[U(t)^2 + \mu(t)^2 + R(t)^2\right] \\
&\quad - U(t)(\mu(t) - R(t))
\end{aligned}
$$

and hence (taking conditional expectations given $U(t)$):

$$
\begin{aligned}
\Delta(U(t)) &\leq \frac{1}{2}\mathbb{E}\left\{\mu(t)^2 + R(t)^2 \mid U(t)\right\} \\
&\quad - U(t)\mathbb{E}\left\{\mu(t) - R(t) \mid U(t)\right\}
\end{aligned}
$$

It is clear that the value $\frac{1}{2}\mathbb{E}\left\{\mu(t)^2 + R(t)^2 \mid U(t)\right\}$ is less than or equal to the constant B defined in (20), and hence:

$$
\Delta(U(t)) \leq B - U(t)\mathbb{E}\left\{\mu(t) - R(t) \mid U(t)\right\} \quad (25)
$$

Noting that $R(t) = \Psi(A(t), k(t))$ and using iterated expectations, we have:

$$
\begin{aligned}
&\mathbb{E}\left\{R(t) \mid U(t)\right\} \\
&= \mathbb{E}\left\{\Psi(A(t), k(t)) \mid U(t)\right\} \\
&= \mathbb{E}\left\{\mathbb{E}\left\{\Psi(A(t), k(t)) \mid U(t), A(t), k(t)\right\} \mid U(t)\right\} \\
&= \mathbb{E}\left\{m(A(t), k(t)) \mid U(t)\right\}
\end{aligned}
$$

where we have used the definition of $m(a, k)$ given in (1). Using this equality in (25) yields the result. ∎

Following the Lyapunov optimization framework of [7] [6], we add a weighted cost term to the drift expression. Specifically, from (24) we have:

$$
\begin{aligned}
\Delta(U(t)) &+ V\mathbb{E}\left\{P_{comp}(t) + P_{tran}(t) \mid U(t)\right\} \leq \\
&B - U(t)\mathbb{E}\left\{C(P_{tran}(t), S(t)) - m(A(t), k(t)) \mid U(t)\right\} \\
&+ V\mathbb{E}\left\{P_{comp}(t) + P_{tran}(t) \mid U(t)\right\} \quad (26)
\end{aligned}
$$

where we have just added an additional term to both sides of (24). Note that $\mathbb{E}\{P_{comp}(t)\,|\,U(t)\}$ can be expressed as follows (using iterated expectations):

$$
\begin{aligned}
\mathbb{E}\{P_{comp}(t)\,|\,U(t)\} \\
&= \mathbb{E}\{\mathbb{E}\{P_{comp}(t)\,|\,U(t),A(t),k(t)\}\,|\,U(t)\} \\
&= \mathbb{E}\{\phi(A(t),k(t))\,|\,U(t)\}
\end{aligned}
$$

Using this equality in the right hand side of (26) and re-arranging terms yields:

$$
\begin{aligned}
\Delta(U(t)) + V\mathbb{E}\{P_{comp}(t) + P_{tran}(t)\,|\,U(t)\} \leq \\
B - \mathbb{E}\{U(t)C(P_{tran}(t),S(t)) - VP_{tran}(t)\,|\,U(t)\} \\
+ \mathbb{E}\{U(t)m(A(t),k(t)) + V\phi(A(t),k(t))\,|\,U(t)\} \quad (27)
\end{aligned}
$$

Now note that we have not yet used the properties of the dynamic compression and transmission policy. Indeed, the above expression (27) is a bound that holds for any compression and transmission scheduling decisions $k(t) \in \mathcal{K}$, $P_{tran}(t) \in \mathcal{P}$ that are made on slot t, including randomized decisions. However, note that the dynamic compression and transmission strategy is designed specifically to minimize the right hand side of (27) over all alternative decisions that can be made on slot t. Indeed, the compression algorithm observes $A(t)$ and $U(t)$ and chooses $k(t) \in \mathcal{K}$ to minimize $U(t)m(A(t),k(t)) + V\phi(A(t),k(t))$, which thus minimizes the following term over all alternative decisions that can be made on slot t:

$$
\mathbb{E}\{U(t)m(A(t),k(t)) + V\phi(A(t),k(t))\,|\,U(t)\}
$$

Similarly, the algorithm chooses $P_{tran}(t) \in \mathcal{P}$ to minimize the right hand side of (27) over all transmission power decisions that can be made on slot t. It follows that the right hand side of (27) is less than or equal to the corresponding expression with $P_{tran}(t)$ and $k(t)$ replaced by $P_{tran}^*(t)$ and $k^*(t)$, where $P_{tran}^*(t)$ and $k^*(t)$ are any other (possibly randomized) policies that satisfy $P_{tran}^*(t) \in \mathcal{P}$ and $k^*(t) \in \mathcal{K}$:

$$
\begin{aligned}
\Delta(U(t)) + V\mathbb{E}\{P_{comp}(t) + P_{tran}(t)\,|\,U(t)\} \leq \\
B - \mathbb{E}\{U(t)C(P_{tran}^*(t),S(t)) - VP_{tran}^*(t)\,|\,U(t)\} \\
+ \mathbb{E}\{U(t)m(A(t),k^*(t)) + V\phi(A(t),k^*(t))\,|\,U(t)\} \quad (28)
\end{aligned}
$$

Now let r_1 be any particular value that satisfies $r_{min} \leq r_1 \leq b\mathbb{E}\{A(t)\}$, and let $k^*(t)$ be the stationary randomized policy that yields:

$$
\begin{aligned}
\mathbb{E}\{\phi(A(t),k^*(t))\} &= h^*(r_1) & (29) \\
\mathbb{E}\{m(A(t),k^*(t))\} &= r_1 & (30)
\end{aligned}
$$

Such a policy exists by (10) and (11) of Lemma 1. Similarly, let r_2 be any value that satisfies $0 \leq r_2 \leq r_{max}$, and let $P_{tran}^*(t)$ be the stationary randomized power allocation policy that yields:

$$
\begin{aligned}
\mathbb{E}\{C(P_{tran}^*(t),S(t))\} &= r_2 & (31) \\
\mathbb{E}\{P_{tran}^*(t)\} &= g^*(r_2) & (32)
\end{aligned}
$$

Such a policy exists by (14) and (15). Further, the stationary randomized policies of (29)-(32) base decisions only on the current $A(t)$ and $S(t)$ states, which are i.i.d. over slots (and are

hence independent of the current queue backlog $U(t)$). Thus, the expectations of (29)-(32) are the same when conditioned on $U(t)$. Plugging (29)-(32) into the right hand side of (28) thus yields:

$$
\begin{aligned}
\Delta(U(t)) + V\mathbb{E}\{P_{comp}(t) + P_{tran}(t)\,|\,U(t)\} \leq \\
B - U(t)(r_2 - r_1) + V(h^*(r_1) + g^*(r_2)) \quad (33)
\end{aligned}
$$

The above inequality holds for all r_1 and r_2 that satisfy $r_{min} \leq r_1 \leq b\mathbb{E}\{A(t)\}$ and $0 \leq r_2 \leq r_{max}$. Let $r_1 = r_2 = r^*$, where r^* is the value of r that optimizes the problem in (16) and (17) of Theorem 1, so that $P_{av}^* = h^*(r^*) + g^*(r^*)$. Plugging into (33), we have:

$$
\Delta(U(t)) + V\mathbb{E}\{P_{comp}(t) + P_{tran}(t)\,|\,U(t)\} \leq B + VP_{av}^*
$$

Using the above drift inequality in the Lyapunov Drift Lemma (Lemma 2) and defining $\alpha(t) = VP_{comp}(t) + VP_{tran}(t)$ and $\beta(t) = B + VP_{av}^*$ yields $\overline{P}_{tot} \leq P_{av}^* + B/V$, proving equation (18) of Theorem 2.

Now choose $r_1 = r_{min}$ and $r_2 = r_{max}$. Plugging into (33) and noting that $P_{comp}(t) \geq 0$ and $P_{tran}(t) \geq 0$ gives:

$$
\begin{aligned}
\Delta(U(t)) \leq B - U(t)(r_{max} - r_{min}) \\
+ V(h^*(r_{min}) + g^*(r_{max}))
\end{aligned}
$$

Using the above drift inequality in the Lyapunov Drift Lemma (Lemma 2) and defining $\alpha(t) = U(t)(r_{max} - r_{min})$ and $\beta(t) = B + V(h^*(r_{min}) + g^*(r_{max}))$, we have:

$$
\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=0}^{t-1} \mathbb{E}\{U(\tau)\} \leq \frac{B + V(h^*(r_{min}) + g^*(r_{max}))}{(r_{max} - r_{min})}
$$

The result of (19) follows because $g^*(r_{max}) \leq P_{max}$ and $h^*(r_{min}) \leq \phi_{max}$. This completes the proof of Theorem 2.

REFERENCES

[1] K. Barr and K. Asanović. Energy aware lossless data compression. *Proc. of First International Conf. on Mobile Systems, Applications, and Services,* May 2003.

[2] C. Sadler and M. Martonosi. Data compression algorithms for energy-constrained devices in delay tolerant networks. *Proc. of ACM Conf. on Embedded Networked Sensor Systems (SenSys),* 2006.

[3] S. S. Pradhan, J. Kusuma, and K. Ramchandran. Distributed compression in a dense microsensor network. *IEEE Signal Processing Magazine,* vol. 19, no. 2, pp. 51-60, March 2002.

[4] R. Cristescu, B. Beferull-Lozano, and M. Vetterli. Networked Slepian-Wolf:theory, algorithms, and scaling laws. *IEEE Transactions on Information Theory,* vol. 51, no. 12, Dec. 2005.

[5] S. Pattem, B. Krishnamachari, and R. Govindan. The impact of spatial correlation on routing with compression in wireless sensor networks. *Proc. IPSN,* 2004.

[6] M. J. Neely. Energy optimal control for time varying wireless networks. *IEEE Transactions on Information Theory,* vol. 52, no. 7, July 2006.

[7] L. Georgiadis, M. J. Neely, and L. Tassiulas. Resource allocation and cross-layer control in wireless networks. *Foundations and Trends in Networking,* vol. 1, no. 1, pp. 1-149, 2006.

[8] C. Li and M. J. Neely. Energy-optimal scheduling with dynamic channel acquisition in wireless downlinks. *Proc. of 46th IEEE Conf. on Dec. and Control (CDC),* Dec. 2007.

[9] D. P. Bertsekas and R. Gallager. *Data Networks.* New Jersey: Prentice-Hall, Inc., 1992.