

Power Efficient Delay Allocation in Multihop Wireless Networks

Dinesh Rajan, *Member, IEEE*

Abstract—In this paper, we present delay allocation strategies that minimize the total transmit power in a multihop wireless network. The focus is on guaranteeing an end-to-end average delay bound for a single variable-bit-rate flow on a multihop fading channel. We first compute an analytical approximation for the transmit power that is required to send a variable-bit-rate source over a finite-state fading channel. We then use this approximation to derive a low-complexity and near-optimal delay allocation method for multihop networks when the fading processes on the multiple hops are independent. Properties of the optimal delay allocation are also studied; in the special case of a Gaussian network, the optimal delay allocation strategy is completely characterized. The tradeoff between single-hop transmission and multihop transmission is studied under an end-to-end delay constraint.

Index Terms—Delay guarantees, multihop, queuing, scheduling, wireless network.

I. INTRODUCTION

THE CAPACITY of wireless *ad hoc* networks has been an area of significant recent research and has been studied under different constraints on interference, maximum node range, node mobility, and hybrid *ad hoc*-infrastructure networks [1]–[8]; these results typically study the scaling of throughput with the number of nodes. Scheduling and capacity in multihop networks with interference constraints have been considered by many recent researchers (see [9]–[12] and references therein).

The importance of incorporating traffic models in wireless physical layer design has been well recognized [13]–[15], and cross-layer optimization has been widely investigated [16]–[22]. Numerous scheduling mechanisms that explore the intricate relationship between packet delay, fairness, throughput, and power have been proposed. Scheduling in the power-limited regime is considered in [23]–[26] using a linear model for the power consumption with increasing rate. A consequence of this linear model is that power savings (or equivalently throughput increases) with increasing delays are only possible in fading channels. On the other hand, a bandwidth-limited regime and an exponential model (as suggested by Shannon's capacity formulation or error rate analysis [27]) for power consumption with an increasing number of packets transmitted

Manuscript received December 30, 2005; revised June 10, 2006 and July 27, 2006. This work was supported by the National Science Foundation under Grant CCF-0546519. Parts of this paper were presented at the SPIE Conference on Fluctuations and Noise, Austin, TX, May 2005. The review of this paper was coordinated by Dr. S. Vishwanath.

The author is with the Department of Electrical Engineering, Southern Methodist University, Dallas, TX 75275-0338 USA.

Digital Object Identifier 10.1109/TVT.2007.897188

TABLE I
TOTAL TRANSMIT POWER REQUIRED IN A TWO-HOP FADING NETWORK FOR THREE DIFFERENT DELAY ALLOCATION STRATEGIES. THE POWER (IN DECIBELS) IS NORMALIZED WITH RESPECT TO THE POWER REQUIRED AT INFINITE DELAY

Delay bound	Power of optimal allocation	Power of low complexity allocation	Power with equal delay division
3.5	3.39	3.45	4.2
11.5	0.31	0.39	1.41

have been used in [28]–[32]. A consequence of this exponential relationship is that power savings with increasing delays are possible both in fading channels and in Gaussian channels with bursty traffic sources.

The problem that we explore in this paper is the optimal allocation of the total end-to-end delay at the different nodes of a multihop wireless network to optimize a desired function (e.g., summation) of the transmit powers at each of the nodes. A bandwidth-limited system is studied using a standard distance-2 interference model. The proposed framework is also used to compare single-hop versus multihop transmissions under end-to-end delay constraints. One of the main distinguishing characteristics of this work over other related works [33] is the joint consideration of bursty source, fading channels, and end-to-end delay constraints.

The importance of the delay allocation problem is motivated by the following simple example. Consider the transmission of a time-varying source over a single-hop fading channel. Intuitively, it is clear that increasing the queuing delay at the transmitter and waiting for *good* channel conditions results in improved performance (lower transmit power/higher throughput). Now, consider the same traffic being transmitted over a two-hop network with an end-to-end delay constraint. Clearly, increasing the delay at the first hop reduces the power at that hop but also reduces the available delay at the second hop, thereby leading to increased power requirements at the second hop. On the other hand, increasing the delay at the second hop reduces the power requirement at the second hop at the cost of increased power at the first hop. Thus, there is a tradeoff between the delay that is allocated to the two hops and the power that is required at the transmit nodes; exploring this tradeoff is the main focus of this paper.

The importance of *optimal* delay allocation is also motivated by comparing the optimal multihop delay allocation scheme (defined in Section IV) with a scheme in which the available delay is uniformly allocated to each of the hops: This comparison is given in Table I for a simple two-hop network. The sum of the transmit powers across the two hops is given for two different

values of end-to-end delay bound. It can be seen that uniformly allocating the available delay to each hop results in over 1-dB loss in performance compared to the optimal allocation. We also show examples in Section IV, where optimal allocation provides over 3-dB performance benefit.

The main contributions of this paper can be succinctly summarized.

- We construct an analytical approximation for the power that is required to transmit a bursty source over a finite-state block fading channel under an average delay bound. This approximation, which is accurate for medium to high delays, clearly shows the dependence of the transmit power on the source burstiness, channel conditions, and delay.
- For a multihop Gaussian network, we characterize the optimal end-to-end delay allocation strategy that minimizes any linear nonnegative combination of the transmit power at each of the nodes. The optimal strategy allocates a delay of one time slot to all but the first node; the remaining delay is given to the first node.
- For arbitrary fading *ad hoc* networks, where each hop is modeled as a finite-state block fading channel, we provide a low-complexity delay allocation mechanism that has significantly better performance than a simple equal delay allocation strategy. This low-complexity delay allocation is obtained by solving a convex optimization problem, and the complexity of computing the optimal solution scales with the number of nodes.
- We study the tradeoff between single-hop and multihop transmissions under finite delay constraints. We consider a simple straight line topology and characterize the achievable power-delay performance for various hops between sender and receiver. Similar linear network topologies have been considered in [33]–[37]. The optimal number of hops between transmitter and receiver depends on the source distribution, channel-fading statistics, and end-to-end delay constraint and is characterized in some special cases.

The results in this paper can be used for network capacity provisioning/planning, admission control, or for designing rate-throttling mechanisms. Furthermore, the proposed optimization framework can be easily modified to include upper bounds on transmit power at each node and, hence, are directly applicable in practical systems.

The focus of this paper is the energy expenditure in the transmission power amplifier, which is a significant fraction ($\approx 40\%$) of the total power consumption in a wireless communication system [38], [39]. Furthermore, to stay within focus, only the queuing delay is considered. Although there exists more sophisticated traffic models, we found that the resulting delay allocation problem is nontrivial, even with simplified independent identically distributed traffic models. Hence, analytical results are given for memoryless sources; representative numerical results for MPEG and Ethernet traffic sources are also given. The proposed algorithm requires knowledge of the channel statistics in each of the hops; this information can be gathered using an on-demand routing protocol such as *ad hoc* on-demand distance

vector (AODV). At each of the subsequent hops, instantaneous (1- or 2-bit information) channel knowledge at the transmitter is also required. Although considering routing protocols is beyond the scope of this paper, we briefly discuss the implementation of the proposed strategy in conjunction with standard *ad hoc* network routing algorithms.

The rest of this paper is organized as follows: In Section II, we describe the system under investigation. Section III focuses on single-hop scheduling, and selected prior results are summarized in Section III-A. In Section III-B, we analyze single-hop fading channels and derive an analytical approximation between transmit power and delay. In Section IV, we explore scheduling in multihop networks. In Section IV-A, the multihop problem is formulated, and Sections IV-B and C focus on multihop Gaussian and fading networks, respectively. Finally, we summarize in Section V.

II. PRELIMINARIES AND PROBLEM FORMULATION

Consider an *ad hoc* wireless network with nodes N_1, N_2, \dots, N_{m+1} , which carries a single flow. Let nodes N_1 and N_{m+1} be the source and destination, respectively, for the traffic (or flow). Furthermore, assume that packets are routed via nodes N_2, N_3, \dots, N_m . We assume that node N_i is not capable of simultaneously transmitting data to node N_{i+1} and receiving data from node N_{i-1} .

The traffic that is generated at source node N_1 is assumed to be bursty (time varying) and is stored in a buffer of size B_1 . Consider a time-slotted system, and let the source produce packets at an average rate of λ packets/time slot. Between time $n-1$ and n , the source produces a_n packets, each of fixed size S bits, where a_n has distribution $r(a_n)$. At intermediate node N_j , the incoming packets are stored in a buffer of size B_j . In this paper, we assume that the buffers are large enough that buffer overflows do not occur.¹ The number of packets in buffer B_j at the beginning of the n th time slot is denoted by $x_{n,j}$. The scheduler at node N_j chooses $u_{n,j}$ packets for transmission at the beginning of the time slot and uses power $P_{n,j}$ for transmission. Since the length of the time slot is fixed, the rate of transmission depends on the number of packets $u_{n,j}$ that is selected for transmission. In this paper, we consider only distributed scheduling, by which we mean that, at each node N_j , the scheduler cannot choose $u_{n,j}$, $P_{n,j}$ based on instantaneous buffer/channel state at the other nodes.

A. Channel Model

The received signal $Y_{n,j+1}$ at node N_{j+1} depends on the transmitted signal $X_{n,j}$ at node N_j and is given by

$$Y_{n,j+1} = \frac{A_{n,j}}{d_{j,j+1}^{2\beta}} X_{n,j} + \varepsilon_{n,j} \quad (1)$$

where $d_{j,j+1}$ is the distance between nodes N_j and N_{j+1} , β is the propagation loss coefficient (typical power loss factor with

¹Analysis of packet outages or losses resulting in buffer overflows is given in [40] and [41].

distance is between 2 and 4, and hence, $\beta \in [1, 2]$, and $\varepsilon_{n,j}$ is the complex circularly symmetric additive white Gaussian noise (AWGN) with zero mean and variance σ^2 . For simplicity, we set $\sigma^2 = 1$. We use two models for the channel between nodes N_j and N_{j+1} : 1) an AWGN channel and 2) a block fading channel. In the AWGN channel model, we set $A_{n,j} = A_j \forall n$. In the block-fading-channel model, the channel gain $A_{n,j} \in \{A_{ji}, i = 1, \dots, n_{\text{ch-states}}\}$ is assumed to be constant over the period of a time slot, i.e., the coherence interval of the channel is the same as the length of the time slot. Furthermore, the channel state at each hop is assumed to form a Markov chain with transition probabilities given by $q_{j:i\ell}$, where $q_{j:i\ell} = \Pr\{A_{n+1,j} = A_{j\ell}/A_{n,j} = A_{ji}\}$. The invariant distribution of the Markov chain, which is denoted by p_{ji} , can be easily calculated from the transition probabilities $q_{j:i\ell}$. Recognize that p_{ji} also represents the probability that the channel gain in the j th hop equals A_{ji} . We typically consider $n_{\text{ch-states}} = 2$, which corresponds to 1 bit of channel state information at the transmitter during every time slot.

B. Buffer and Delay Model

Recall that the output traffic at node N_{j-1} during time slot n is the input traffic to node N_j .² Thus, the buffer update is given by

$$x_{n+1,j} = x_{n,j} + a_{n+1,j} - u_{n,j} = x_{n,j} + u_{n,j-1} - u_{n,j} \quad \forall j \quad (2)$$

where $a_{n,j}$ represents the input traffic at node N_j during time slot n , and for notational simplicity, we let $u_{n,0} = a_n$. In this paper, the physical layer coding scheme is chosen to ensure that the selected packets are transmitted completely within one time slot.³ At each node, the average queuing delay D_j is related to the average buffer length via Little's theorem [42] and is given by

$$D_j = \frac{1}{\lambda} \mathbb{E}[x_{n,j}] \quad (3)$$

where $\lambda = \mathbb{E}[a_n]$ is the average packet arrival rate. A natural constraint on $u_{n,j}$ is that $0 \leq u_{n,j} \leq x_{n,j}$. We assume that all packets that arrive at the beginning of time slot n can be transmitted in time slot n or later. At each hop, the smallest average delay of one time slot is achieved when all arriving packets are transmitted in the same time slot, i.e., $u_{n,j} = a_{n,j}$, which implies that $D_j = (1/\lambda)\mathbb{E}[x_{n,j}] = 1$, assuming that the buffer is initially empty. The average end-to-end delay is given by $D_{\text{ete}} = \sum_{j=1}^m D_j$. Using our convention, the minimum end-to-end delay \bar{D}_{ete} equals m , which is the number of hops.

²The careful reader will rightly point out that node N_j might generate traffic on its own, which should be accounted for. However, as mentioned earlier, for simplicity, we only consider a single flow.

³The propagation delay in the wireless medium is assumed to be a constant and is not considered in this paper.

C. Interference Model

We consider a standard *distance-2* interference model in the network, i.e., while nodes N_1 and N_2 are in communication, nodes N_3 and N_4 cannot be in communication but nodes N_4 and N_5 can also be in simultaneous communication.⁴ The incorporation of distance-2 interference ensures that the minimum end-to-end delay is actually greater than m . For instance, consider a simple two-hop network with preassigned channel allocation as follows: Node N_1 transmits during every even time slot, and node N_2 transmits during every odd time slot. It can be easily shown that all packets arriving at node N_1 during even time slots experience a minimum end-to-end delay of 2, whereas all packets that arrive at node N_1 during odd time slots experience minimum end-to-end delay of 3. Thus, the minimum average end-to-end delay equals 2.5 time slots. Clearly, such a preassigned channel allocation scheme only provides a lower bound on the channel contention delay.

D. Scaled Down Equivalent Discrete Time Model at Each Node

In this work, we consider a preassigned channel assignment that is similar to that described for the two-hop network. In general, due to the distance-2 constraint, a node will acquire a channel for transmission once every k time slots, where $k = \min(m, 3)$, in a periodic or round-robin manner. Consequently, we will consider an equivalent discrete-time queuing model at each node, in which the time index is scaled down by factor k . The j th time slot in the equivalent queuing model at node N_1 corresponds to time slot kj in the original model. Similarly, the j th time slot in the equivalent model at node N_i corresponds to time slot $kj + i - 1$ in the original model. This equivalence is depicted pictorially in Fig. 1 for a simple two-hop network.

Let $\tilde{x}_{i,j}$ denote the number of packets in the buffer at node N_j during time slot i . We use the notation that quantities with $\tilde{\cdot}$ are the equivalents in the time-scaled model of the corresponding quantities in the original model.

The equivalent arrival process at node N_1 is now the sum of the original arrivals over k consecutive time slots, i.e., $\tilde{a}_{i,1} = \sum_{j=0}^{k-1} a_{ki+j}$. The distribution of the effective arrival process $\{\tilde{a}\}$ is given by the k -fold convolution of the distribution $r(a_n)$, and the average arrival rate equals $k\lambda$. The number of packets transmitted in time slot i is given by $\tilde{u}_{i,1} = u_{ki,1}$. Effective delay D_1 can be easily computed as

$$D_1 = k(\tilde{D}_1 - 1) + 1 + (k - 1)/2 \quad (4)$$

where \tilde{D}_1 is the queuing delay at node N_1 given by $\tilde{D}_1 = \mathbb{E}[\tilde{x}_{i,1}]/k\lambda$. The factor of k in (4) accounts for the scaling in time, and the additional term of $(k - 1)/2$ is due to the fact that packet arrivals are possible even in the time slots when node N_1 does not have access to the channel.

For all the subsequent nodes, the effective packet arrivals are given by $\tilde{a}_{i,j} = \tilde{u}_{i,j-1}$. The effective delay $D_i = k(\tilde{D}_i - 1) + 1$,

⁴Such an interference model is commonly used to model 802.11-based networks, which use a handshaking mechanism prior to data transfer.

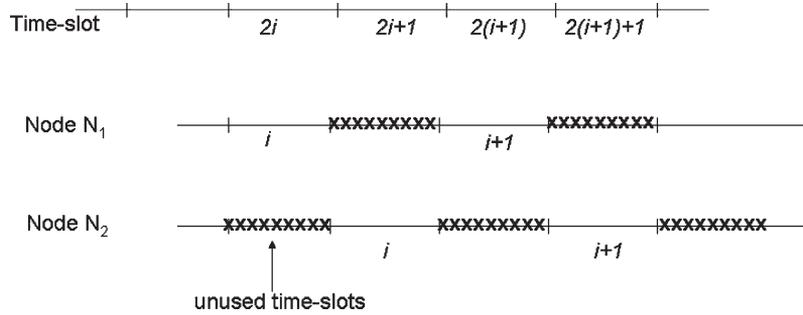


Fig. 1. Equivalent scaled-down discrete-time model.

where $\tilde{D}_i = \mathbb{E}[\tilde{x}_{j,i}]/k\lambda$. Recognize that, for nodes N_i , $i \neq 1$ packets arrive only once every k time slots; hence, there is no additive factor [like in (4)] in calculating D_i . The total end-to-end delay is thus given by

$$D_{ete} = \sum_{i=1}^m D_i = k \sum_{i=1}^m \tilde{D}_i - (k-1)m + \frac{k-1}{2}. \quad (5)$$

With the new equivalent queuing model at each node, the effective fading channel transition probabilities are easily computed using the Chapman–Kolmogorov equations [42, p. 259]. Specifically

$$\begin{aligned} \tilde{q}_{j:i} &= \Pr\{\tilde{A}_{k+1,j} = A_{j,l} / \tilde{A}_{k,j} = A_{j,i}\} \\ &= \sum_{c=0}^{\infty} q_{j:ic}^r q_{j:cl}^{k-r}, \quad k \geq 0. \end{aligned} \quad (6)$$

The corresponding invariant probability distribution \tilde{p}_{ji} can be easily calculated using these transition probabilities.

At node N_i , the scheduler now computes $\tilde{u}_{j,i}$ based on $\tilde{x}_{j,i}$ and $\tilde{A}_{j,i}$, as explained in [31]. The transmit power at node N_i during time slot n depends on the number of packets that were transmitted, the coding and modulation scheme that is used, and the desired performance in terms of bit error rate or frame error rate. In this paper, we assume reliable packet transmissions in the Shannon theoretic sense. Although such an assumption of reliability is only feasible asymptotically, it provides two benefits: 1) A closed-form expression can be used for required power and 2) it also provides a lower bound on the power that is required by any practical transmission strategy. Moreover, advanced codes such as LDPC and Turbo codes achieve performance that is very close to capacity with finite block lengths. Thus, we use the famous Gaussian capacity formulation [43] to derive the required power $P_{n,j:i}$ at node N_j during time slot n and channel state i as

$$\tilde{P}_{n,j:i} = \frac{d_{j,j+1}^{2\beta}}{|\tilde{A}_{j,i}|^2} (e^{S\tilde{u}_{n,j:i}/T_c} - 1) \quad (7)$$

where T_c is the length of the time slot. For simplicity, we set $S/T_c = 1$ in the rest of this paper. The average transmit power at node N_j is given by $P_j = (1/k)\mathbb{E}[\tilde{p}_{ji}\tilde{P}_{n,j:i}]$ since the

channel is used by each transmitter only once every k time slots. The total network power is then given by

$$\frac{1}{k} \sum_{j=1}^m \mathbb{E} \left[\sum_{i=1}^{n_{ch-states}} \tilde{p}_{ji} \tilde{P}_{n,j:i} \right] \quad (8)$$

where the expectation is over the distribution of $u_{n,j:i}$.

III. SINGLE-FLOW ONE-HOP CHANNEL

In this section, we focus on scheduling over a single-hop wireless channel that forms the basis for the multihop analysis in the rest of this paper. We first summarize prior results on scheduling in Section III-A and then derive a new closed-form approximate formula for transmit power in Section III-B. Numerical results are discussed in Section III-C.

A. Review: Delay Bounded Scheduling

It is well known that, on a fading channel, the optimal transmission scheme exploits knowledge of instantaneous channel conditions at the transmitter to improve performance [44]. For instance, one could delay the transmission of packets in *bad* fading states and transmit more packets in the *good* channel states. Such a transmission policy results in reduced transmission power at the cost of increased delay that is experienced by the incoming traffic. A similar concept of delaying packet transmission until one is closer to a base station, thereby ensuring high signal strength, has been used, in the INFOSTATION architecture [45]. Furthermore, using such a transmission scheme, the output traffic at a node could have larger variation than the input traffic at that node.

An analogous effect has been shown for the transmission of bursty traffic over AWGN channels [29], [31]. By smoothing the variations in traffic arrival and transmitting at nearly constant rate, the transmission power is reduced at the cost of increased delays. Note that, in this case, the output traffic has less variation than the input traffic.

The minimal power delay bounded scheduler (referred to as optimal scheduler) for transmitting bursty source over a fading channel has been derived [28], [29], [31] using dynamic programming methods. The optimal scheduler action is a combination of the action in the two aforementioned cases: 1) smoothing of bursty arrivals and 2) introducing rate variations to transmit more packets in good channel states and a

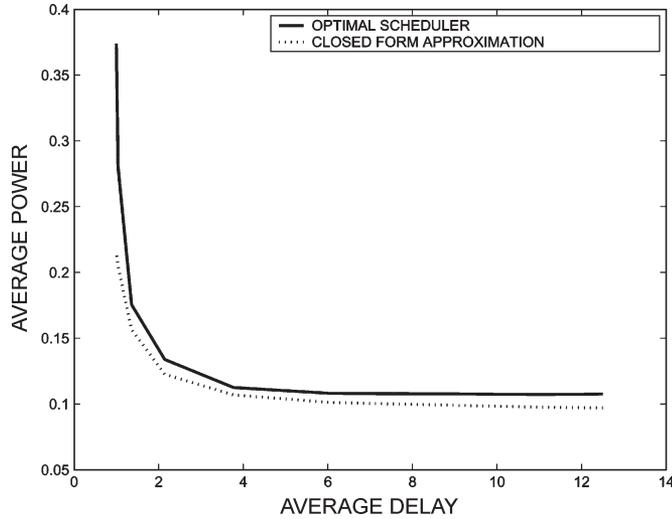


Fig. 2. Plot to demonstrate the accuracy of the proposed closed-form approximation for power required versus delay in a single-hop fading channel.

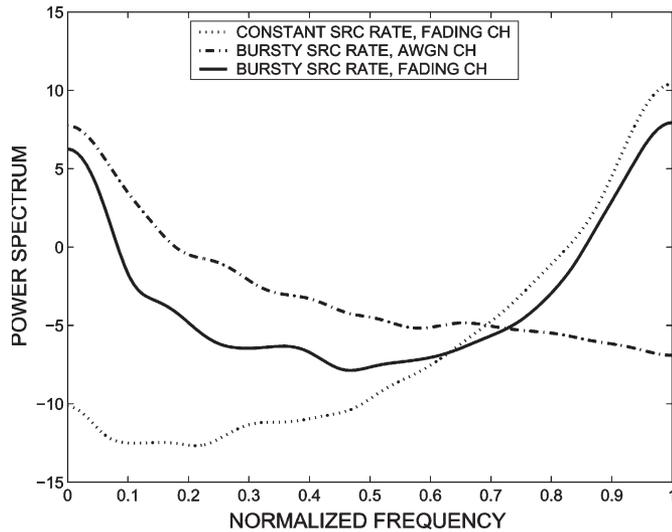


Fig. 3. Frequency response of the output traffic at a node in three different scenarios.

few packets in bad channel states. The variation of the transmit power with delay using the optimal scheduler is shown in Fig. 2; it can be seen that substantial savings in power are possible for small increases in delay.

Preliminary investigations have revealed that the scheduler that is designed for transmitting bursty sources over AWGN channels behaves like a low-pass filter. We also showed that the bandwidth of the filter depends inversely on the delay bound [46]. The filtering equivalence can be extended to scheduling of constant-rate sources over fading channels: In this case, it turns out that the scheduler behaves like a high-pass filter. When the input is bursty and the channel is fading, the optimal scheduler exhibits behavior that is a combination of low-pass and high-pass; the scheduler actually behaves like a band-reject filter. These characteristics are depicted in Fig. 3. The low-pass, high-pass, and band-reject nature of the output traffic spectrum is clearly evident from the figure.

B. Single-Hop Fading Channels: Analytical Performance Evaluation

In this section, we derive a simple closed-form analytical approximation for the required transmit power at each hop as a function of the traffic, fading channel, and delay bound. This approximation forms the basis for the proposed (Section IV) low-complexity delay allocation strategy for multihop networks. Since all results in this section are derived for a single-hop network, we use symbols from the original model rather than the equivalent model. The analytical approximation, which is derived in Appendix A based on a multivariate Gaussian distribution assumption on the traffic output of a scheduler, is given by

$$\begin{aligned} \bar{P}_j &= \mathbb{E} \left[\sum_{i=1}^{n_{\text{ch-states}}} p_{ji} P_{n,j;i} \right] \\ &\approx \sum_{i=1}^{n_{\text{ch-states}}} \frac{d_{j,j+1}^{2\beta} p_{ji}}{|A_{ji}|^2} \left(e^{m_{ji}(D_j) + \frac{\sigma_{u_{j,i}}^2(D_j)}{2}} - 1 \right) \end{aligned} \quad (9)$$

where D_j is the delay, $m_{ji}(D_j)$ is the mean transmission rate, $\sigma_{u_{j,i}}^2(D_j)$ is the variance of the transmission rates, $|A_{ji}|^2$ represents the effective channel gain, and p_{ji} is the probability of channel-fading state i in the j th hop. In Appendix A, we also derive an approximate value for $m_{ji}(D_j)$ and $\sigma_{u_{j,i}}^2(D_j)$ in terms of the mean arrival rate, variance of the arrival rate, and the fading channel statistics. The accuracy of the approximation is numerically studied and compared with the power of the optimal scheduler in Fig. 2. Recall that the optimal scheduler is computed using dynamic programming methods in [28] and [31]. It is evident from the figure that the approximation is accurate for moderate to large delay values.

As noted before, the output traffic at node N_j is the input traffic to node N_{j+1} . We define the *effective channel variance*, which is denoted by σ_{ch}^2 , as

$$\sigma_{\text{ch}}^2 = \left[\sum_{i=1}^{n_{\text{ch-states}}} p_{ji} m_{ji}^2(\infty) - \lambda^2 \right]. \quad (10)$$

This channel variance indicates how different the optimal transmission rates (equivalently gains) in the fading states are. The variance of the output traffic at node N_j can be derived (Appendix A) in terms of the variance of the input traffic to that node as

$$\sigma_{u_j}^2 = \frac{\sigma_{u_{j-1}}^2}{2\lambda D_j - 1} + \frac{2\lambda D_j}{2\lambda D_j - 1} \frac{(D_j - 1)^2}{D_j^2} \sigma_{\text{ch}}^2. \quad (11)$$

Thus, the variance of the output traffic is a convex combination of the variance of the input traffic and the effective channel variance that is scaled by $(D_j - 1)^2/D_j^2$. Furthermore, the variance of the output traffic could be less than or greater than the variance of the input traffic. For the transmission of a constant-rate source, the first term in (11) equals zero, and the output variance

TABLE II
RATIO OF OUTPUT TRAFFIC VARIANCE TO INPUT TRAFFIC VARIANCE FOR
DIFFERENT CHANNELS AND FOUR DIFFERENT DELAYS

Delay in time-slots	2	5	10	20	∞
Example 1	0.45	1.01	1.26	1.39	1.52
Example 2	0.27	0.56	0.69	0.76	0.83
Example 3	0.05	0.02	0.01	0.005	0

depends only on channel-fading characteristics. In the case of an AWGN channel, the second term in (11) reduces to zero, and the output variance only depends on input traffic variance and the delay. The following examples quantitatively illustrate the relationship between input and output traffic variances and delay.

C. Numerical Results: Input Versus Output Traffic Variances at a Node

Example 1: Consider a two-state fading channel with gains $A_j \in \{1, 100\}$. Assume that the probabilities of being in each individual state is given by $\Pr(A_{n,j} = 1) = 0.75$ and $\Pr(A_{n,j} = 100) = 0.25$. In this case, the optimal transmission rates in the two fading states at infinite delay can be calculated using (36), as $m_{j1}(\infty) = 2.74$ and $m_{j2}(\infty) = 11.76$. For a sample traffic arrival with mean arrival rate $\lambda = 5$ and variance $\sigma_a^2 = 10$, the ratio of the output variance to the input variance is given in Table II for various delays. Clearly, the output traffic variance could be higher or lower than the input variance, depending on the delay constraint. Asymptotically, as $D \rightarrow \infty$, the output traffic variance equals σ_{ch}^2 ; thus, the ratio of output to input traffic variance equals $\sigma_{ch}^2/\sigma_a^2 = 1.52$.

Example 2: In this example, we let $A_j \in \{1, 100\}$, and set $\Pr(A_{n,j} = 1) = 0.25$ and $\Pr(A_{n,j} = 100) = 0.75$. The optimal transmission rates in the two fading states at infinite delay are given by $m_{j1}(\infty) = 0$ and $m_{j2}(\infty) = 6.67$. For the same traffic arrivals as in Example 1, the ratio of output variance to input variance is given in Table II. In this case, the output variance is strictly less than the input traffic variance, but the output traffic variance increases with delay. Asymptotically, as $D \rightarrow \infty$, the ratio of output to input traffic variance equals $\sigma_{ch}^2/\sigma_a^2 = 0.834$.

Example 3: In this example, we let $A_j = 1$, i.e., an AWGN channel, and assume the same traffic statistics as in the earlier examples. In this case too, the output variance is strictly less than the input traffic variance, as illustrated in Table II. However, unlike Example 2, the output traffic variance decreases with delay and asymptotically reaches 0 as $D \rightarrow \infty$.

In summary, Examples 1 and 2 show that the ratio of output to input variance increases with delay; in the first case, the supremum of the ratio is greater than one, while in the second case, the supremum is less than one. Example 3 shows that this ratio of output to input variance decreases with delay.

IV. MULTIHOP NETWORKS

In this section, we investigate the delay allocation problem in a multihop network.

A. Problem Formulation

The optimization problem of interest is posed as follows:

$$\min_{\tilde{u}_{n,j:i} \forall i, j, n} \sum_{j=1}^m \frac{w_j}{k} \mathbb{E} \left[\sum_{i=1}^{n_{ch-states}} \tilde{p}_{ji} \tilde{P}_{n,j:i} \right] \quad (12)$$

where $\tilde{u}_{n,j:i} \forall i, j, n$ are the variables of optimization, \bar{D} is the bound on the end-to-end delay, and w_j is the weighting of the power at node N_j . A higher weight w_j can be used for instance when battery resources at node N_j are scarce. In this paper, for simplicity, we choose w_j to be constant for all j 's. In other words, we only consider the minimization of the sum of transmit powers at the different nodes. Recognize that the optimization problem (12) is extremely complex to solve due to the dependence of input traffic at one node on the output traffic at the previous node. The optimal results that are given in Table I are computed in the simple two-hop case based on an exhaustive search and serves as a lower bound on the total transmit power that is required to support that traffic under delay constraint. For each possible value of \bar{D}_1 , the optimal scheduler at node N_1 is computed; using the output traffic at node N_1 as the input to node N_2 , the optimal scheduler at node N_2 is computed with delay $\bar{D} - \bar{D}_1$. The resulting sum power is compared with the sum power using all other possible delay allocation values, and the minimum value is selected.

To make the problem tractable, we use a statistical characterization of the output traffic at each node. Specifically, we use the derived relationships [in (11)] for the mean and variance of output traffic at each node to determine the input traffic at the next node. Such a characterization allows us to decouple the optimizing function into tractable functions. Using (9), the summation of the powers that are required in the entire network can be approximated as

$$\begin{aligned} \bar{P}_{net} &= \sum_{j=1}^m \frac{w_j}{k} \bar{P}_j \\ &= \sum_{j=1}^m \sum_{i=1}^{n_{ch-states}} \frac{w_j d_{j,j+1}^{2\beta} p_{ji}}{k |\tilde{A}_{ji}|^2} \left(e^{m_{ji}(\bar{D}_j) + \frac{\sigma_{u_{ji}}^2(\bar{D}_j)}{2}} - 1 \right) \end{aligned} \quad (13)$$

where we have used the quantities in the equivalent scaled-down discrete model.

The optimization problem is then rewritten as follows:

$$\min_{\tilde{u}_{n,j:i} \forall i, j, n} \bar{P}_{net}. \quad (14)$$

Now, the variables of optimization are delays \bar{D}_j that are allocated to the various nodes; conditioned on delay \bar{D}_j being allocated to node N_j , the optimal scheduler that minimizes the power at node N_j is calculated using dynamic programming techniques [28], [31]. In Appendix C, we have shown that

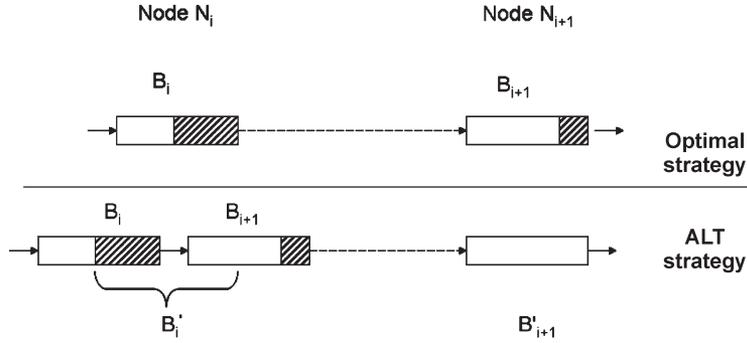


Fig. 4. Buffer actions in *optimal* and ALT strategies, as used in the proof of Theorem 1. The delay at node N_{i+1} using ALT strategy equals 1, i.e., all packets in the buffer are transmitted during every time slot.

the optimizing function in (14) is convex. Furthermore, the constraints in (14) are linear. Hence, optimization problem (14) is easily solved (even for a large number of hops) using a numerical optimizer (e.g., the *fmincon* constrained optimization function in Matlab). In some simple cases, analytical solutions to (14) are also possible and discussed in the next section. The performance of the end-to-end scheduler based on this approximation is given in Table I; clearly, the proposed low-complexity solution is near optimal.

It should be reiterated that the proposed allocation strategy requires knowledge of channel statistics at all hops and the input traffic distribution. One can conceive implementation strategies wherein such statistics can be gathered adaptively, for example, using an on-demand routing algorithm such as AODV [47]. For instance, each node can append its channel statistics, along with route request frames (RREQ), and the destination can then make the optimal delay allocation and notify the intermediate nodes of the delays at each hop using the route reply (RREP) packets.

To understand the relative impacts of the source and channel on delay allocation, we study in detail the following special cases: 1) Gaussian network and bursty source and 2) fading network and bursty source.

B. Delay Allocation: Gaussian Network and Bursty Source

It turns out that, in a Gaussian network, i.e., the channel between successive nodes N_i and N_{i+1} is a Gaussian channel, the optimal delay allocation strategy is quite simple and is characterized by the following theorem.

Theorem 1: Consider a multihop wireless network with m hops, and let each hop of the network be modeled as a time-invariant additive Gaussian noise channel. The incoming traffic or traffic that is generated at node N_1 is bursty. This traffic is transmitted to destination node N_{m+1} via nodes N_2, N_3, \dots, N_m . There is an end-to-end delay constraint D_{ete} for flow. Any linear nonnegative combination of the transmit powers at nodes N_1, N_2, \dots, N_m is minimized by allocating delay $\tilde{D}_i = 1$, $i \neq 1$ and the remaining delay to node N_1 .

Proof: The proof is by contradiction. The proof also uses the following property of scheduling over single-hop Gaussian channels; the larger the delay that is allocated to a link, the smoother the output traffic at that node, and the lower the required transmit power [28], [31].

Let the “optimal” delay allocation at two successive hops be \tilde{D}_i and \tilde{D}_{i+1} , where $\tilde{D}_{i+1} \neq 1 (\implies \tilde{D}_{i+1} > 1)$, and let the corresponding transmit powers equal \tilde{P}_i and \tilde{P}_{i+1} . We now construct a different delay allocation strategy, which is labeled ALT, at nodes N_i and N_{i+1} with delays \tilde{D}'_i and \tilde{D}'_{i+1} such that $\tilde{D}'_i + \tilde{D}'_{i+1} = \tilde{D}_i + \tilde{D}_{i+1}$. Furthermore, we set $\tilde{D}'_{i+1} = 1$ and show that the total power of strategy ALT is lower than the optimal delay allocation strategy.

Consider buffers B_i and B_{i+1} at nodes N_i and N_{i+1} , respectively. Now, in strategy ALT, assume that both the buffers are at node N_i with one preceding the other and the scheduler at node N_{i+1} has a delay of one time slot, as depicted in Fig. 4. The ALT scheduler action at node N_i includes the scheduler action of the “optimum” scheduler at nodes N_i and N_{i+1} with one minor difference; packets that arrive at the second buffer at node N_i during time slot n using scheduler ALT can be transmitted in time slot n or later. Clearly, with this new scheduling mechanism, the delay \tilde{D}'_i at node i equals $\tilde{D}_i + \tilde{D}_{i+1} - 1$ and delay $\tilde{D}'_{i+1} = 1$. Furthermore, the output traffic at node N_{i+1} using the new scheduler is the same as the optimal scheduler; hence, the transmit power at node N_{i+1} is unchanged, i.e., $\tilde{P}'_{i+1} = \tilde{P}_{i+1}$. However, the transmit power at node N_i is reduced since it is allocated a larger delay $\tilde{D}'_i = \tilde{D}_i + \tilde{D}_{i+1} - 1 > \tilde{D}_i$ since $\tilde{D}_{i+1} > 1$. Thus, the total power of strategy ALT is lower than the power using the optimal strategy, which is a contradiction. Hence, the power is minimized by allocating a delay of one time slot to node N_{i+1} . By simple extension, we can prove that the optimal allocation of delay at all the hops after the first hop equals 1. ■

We now provide a qualitative justification for the theorem. For a given delay constraint, the average transmit power at each node increases with increasing burstiness (or variability) of incoming traffic. Moreover, in an AWGN channel, the variance of the output traffic strictly decreases with increasing delay [31]. Thus, the total power is minimized by allocating as much delay as possible to the first node without violating the end-to-end delay bound.

To further understand and characterize the behavior of the optimal delay allocation strategy, we consider the following cases. For simplicity, we consider a straight-line topology with $m + 1$ uniformly spaced nodes and study the performance for various values of m . Let d_0 be the distance between nodes N_1 and N_{m+1} .

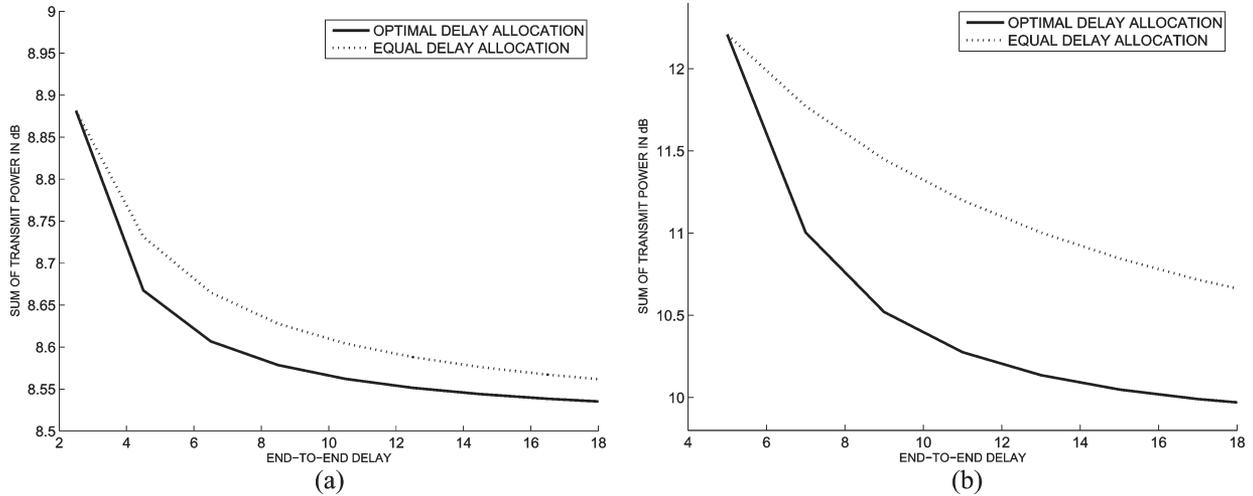


Fig. 5. Plot of the total transmit power versus end-to-end delay using optimal delay allocation and equal delay allocation in two different cases. (a) Gaussian network. (b) Fading network with one hop being in deep fade. Simple on-off traffic source is considered with zero- or ten-packet arrivals in each time slot.

1) *Equal Delay Versus Optimal Delay Allocation:* The sum power using optimal and equal delay allocation strategies is plotted in Fig. 5(a) for a simple on-off traffic source. The input distribution is $r(a_n) = 0.5\delta(0) + 0.5\delta(10)$, i.e., in each time slot, either zero or ten packets are received. Using both the strategies, the power decreases with increasing delay constraint, as expected. The significant savings in power using the optimal allocation scheme are clearly evident from the figure.

For the Gaussian network, we can also analytically approximate the gains of optimal delay allocation over uniform delay allocation strategy. It should be added that, with both the delay allocation strategies, each node schedules its packets optimally to minimize the local transmission power given the local delay constraint at that node. Let $\tilde{D}_{eq,i}$ represent the delay at each of the nodes using the equal delay allocation strategy; then, using (5), it is easy to show that $\tilde{D}_{eq,i} = 1 + (1/mk)(D_{ete} - m - (k-1)/2)\forall i$, where D_{ete} is the total end-to-end delay. In the optimal delay allocation case, $\tilde{D}_{opt,i} = 1, i > 1$; thus, using (5), $\tilde{D}_{opt,1} = 1 + (1/k)(D_{ete} - m - (k-1)/2)$. The ratio of the power that is required using equal delay allocation to all hops versus optimal allocation to all hops can be calculated using (13) as

$$\left(\frac{P_{eq}}{P_{opt}}\right) = \frac{\sum_{k=1}^m e^{\frac{\sigma_a^2}{(2\lambda\tilde{D}_{eq,1}-1)^k}}}{\sum_{k=1}^m e^{\frac{\sigma_a^2}{(2\lambda D_{opt,1}-1)(2\lambda-1)^{(k-1)}}}}. \quad (15)$$

A plot of this ratio is given in Fig. 6 for two different traffic arrival streams: One stream contains Ethernet traffic trace obtained from [48], and the other stream is an MPEG trace. Clearly, substantial power savings are possible when the delay is allocated optimally across the multiple hops. It is also interesting to note that maximum gains arise only at medium delays, which is of high practical value. When $\tilde{D}_j = 1$ or when $\tilde{D}_j \rightarrow \infty$, the optimal allocation strategy only provides marginal gains over the equal delay allocation strategy.

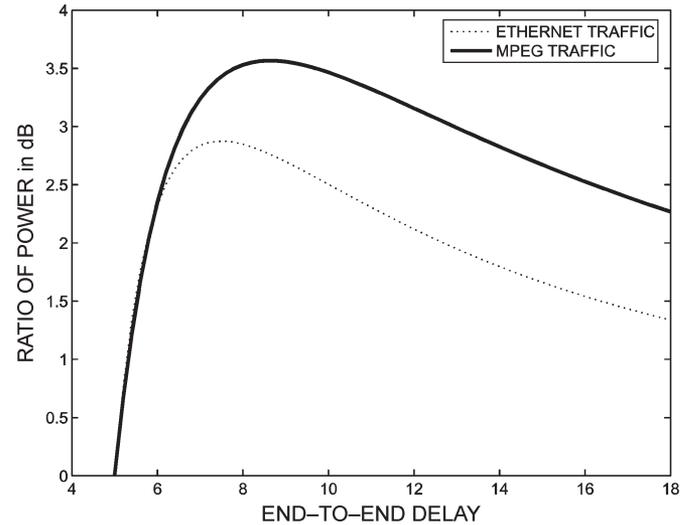


Fig. 6. Plot of ratio of power in decibels with equal delay allocation versus optimal delay allocation for two different traffic traces. A four-hop wireless AWGN network is considered; hence, the smallest end-to-end delay equals 5.

2) *Single-Hop Versus Multihop Transmission:* The proposed framework can also be used to study the performance of varying numbers of hops between transmitter and receiver nodes. For a fixed d_0 , we study the variation of \bar{P}_{net} versus m . The sum transmit power \bar{P}_{net} for various numbers of hops and delays is plotted in Fig. 7(a) and (b) for two different arrival rates λ . In Fig. 7(a) and (b), the traffic arrivals have uniform distribution over support sets $\{0, 1, 2\}$ and $\{0, \dots, 8\}$, respectively. As expected for fixed m , the power decreases with delay. However, for a fixed end-to-end delay bound, the power versus m exhibits interesting nonmonotonic behavior. We now provide a qualitative explanation for behavior in Fig. 7(a) and (b).

Recognize that the power at each hop depends on the rate and distance to the next hop in a nonlinear way. First, consider $m \leq 3, \implies k = m$. Now, increasing m from 1 to 3 has three consequences: 1) The average rate of transmission in each node increases by k ; thus, power increases as $e^{k\lambda}$. 2) The distance

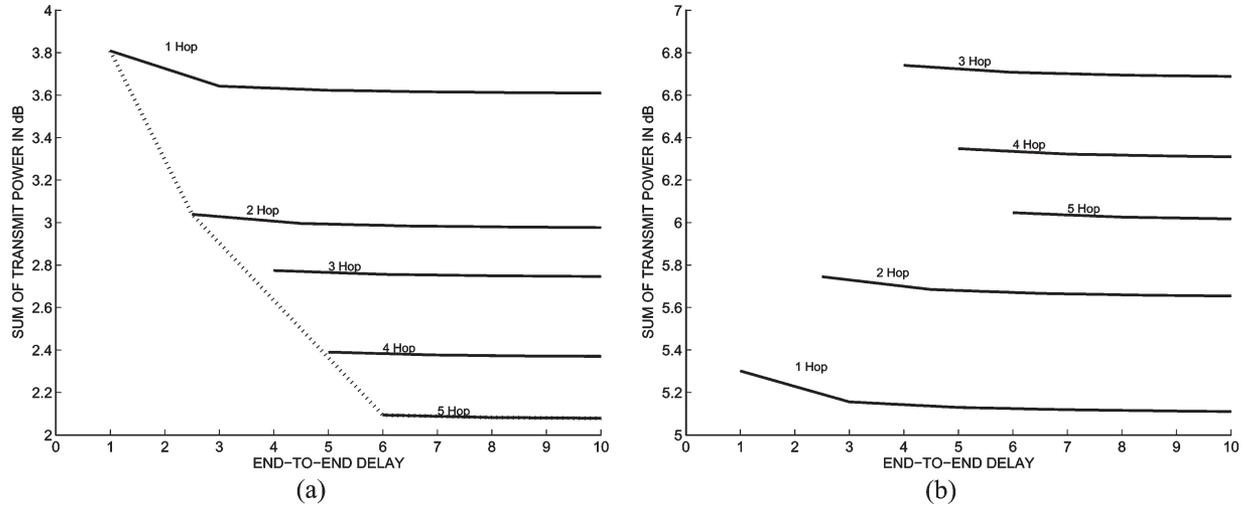


Fig. 7. Plot of the total transmit power versus end-to-end delay for a Gaussian network in two different cases. (a) Small average arrival rate $\lambda = 1$. (b) Large $\lambda = 4$.

between nodes decreases, and the power per hop behaves as $(d_0/m)^{2\beta}$. 3) The number of transmitters increases linearly with m , but at any given time only one transmitter is active. Thus, the third factor does not increase or decrease \bar{P}_{net} as m goes from 1 to 3. For small λ , the reduction due to factor 2 is larger than the increase due to factor 1; hence, the power decreases as m increases from 1 to 3, which is a result that is confirmed by Fig. 7(a). For large λ , the reverse is true, and the power decreases as m increases from 1 to 3; Fig. 7(b) confirms the result.

As the number of nodes continue to increase beyond 3, $k = 3$ stays constant. Again, consider the three factors mentioned previously: 1) The average rate of transmission does not increase since k is a constant, and this term does not affect \bar{P}_{net} . 2) The distance between nodes decreases as m increases, and the power decreases as $(d_0/m)^{2\beta}$. 3) The number of transmitters increases linearly with m , but each transmitter is used only once every k time slots. Hence, the power at a node has to be scaled by a factor m/k to obtain the total network power. The third issue causes a linear increase in power with m , whereas the second issue causes a decrease in power that is on the order of $m^{-2\beta}$. Since $\beta > 1$, the total power reduces with increasing number of hops when $m > 3$. Fig. 7(a) and (b) confirms this qualitative analysis. A similar conclusion is stated in [33], where, for large rates, single-hop performance is superior to multihop performance without using any interference cancellation. Asymptotically, as $m \rightarrow \infty$, the end-to-end delay $D_{\text{ete}} \rightarrow \infty$, and that case is studied next.

3) *Large-Delay Analysis:* Asymptotically, for large delays, the scheduler completely smoothes the burstiness of the source, and the output traffic depends only on the channel conditions. For a Gaussian network, each node transmits at constant rate $k\lambda$, and the transmit power at each hop is given by $(1/k)(1/|A_j|^2)(d_0/m)^{2\beta}(e^{k\lambda} - 1)$. The total power in the network is given by

$$\bar{P}_{\text{net}} = m \frac{1}{k} \frac{1}{|A_j|^2} \left(\frac{d_0}{m} \right)^{2\beta} (e^{k\lambda} - 1). \quad (16)$$

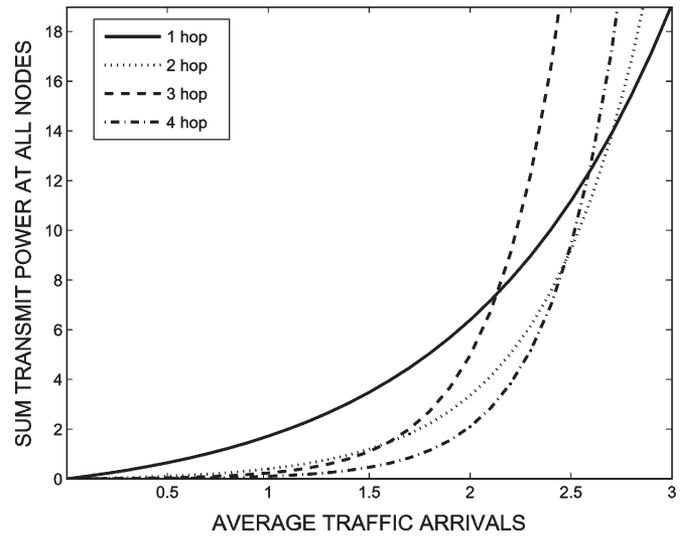


Fig. 8. Total power that is required in a Gaussian network as a function of the mean traffic arrival rate λ .

The plot of sum power versus average arrival rate λ is given in Fig. 8 for various numbers of hops. It is clear from the plot that the number of hops that minimize the power depends on λ . The set of λ 's for which a one-hop route results in lower power than a two-hop route can be easily computed in closed form as $\lambda < \log(2^{2\beta-1} - 1)$. Similar bounds can be computed to compare a m -hop and m' -hop route. In addition, recognize that the total power monotonically decreases with m for $m > 3$ and asymptotically reaches 0. However, the total delay increases monotonically with m ; thus, the given data can be transmitted with arbitrarily small power $\epsilon \rightarrow 0$ only as $D \rightarrow \infty$.

C. Delay Allocation: Fading Network and Bursty Source

Analogous to the earlier case (Gaussian network, bursty source), even in this scenario, the transmit power decreases monotonically with increasing delay bound. However, the output traffic at a given node could have more or less burstiness

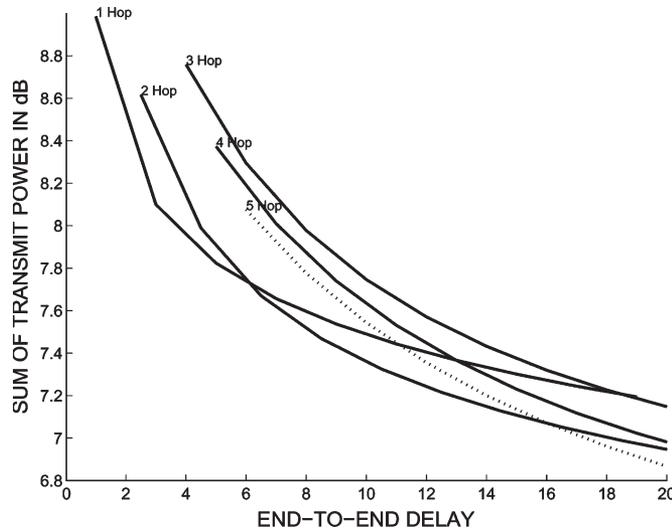


Fig. 9. Plot of power versus delay in a fading network for various numbers of hops. The figure clearly shows that the optimal hops between transmit and receive nodes depend on the delay bound. Input traffic has uniform distribution over $\{0, \dots, 4\}$ and channel gain $A_j \in \{1, 300\} \forall j$.

than the input traffic at that node.⁵ Qualitatively, the larger the variation in $|A_{ij}|^2$ at a given hop, the larger the variance of the output traffic, i.e., the output is more bursty. As a result of this increased burstiness, the optimal delay allocation can be substantially different from case 1. For example, using a simple two-state fading channel with gains 1 and 1000 and a two-hop network, the optimal delay allocation strategy gives nearly 80% of the total delay to the second hop and the remaining 20% to the first hop.

Similar to the Gaussian network, we now study the following special cases in detail.

1) *Equal Versus Optimal Delay Allocation*: The plot of the power that is required using optimal and equal delay allocation strategies is given in Fig. 5(b). Simple on-off traffic is considered with either ten or zero packets arriving each time slot. The average rate λ equals 5 packets/time slot, and a four-hop network is considered. The channel gains $A_j \in \{30\}$, $j > 1$ and $A_1 \in \{1, 300\}$, i.e., one of the hops is in deep fade and the other hops experience a Gaussian channel. The probability of being in deep fade equals 0.5. The figure clearly shows that additional power savings of about 1 dB are possible using the proposed delay allocation strategy as compared to the uniform delay allocation strategy. As in the Gaussian network, the maximum benefit due to delay allocation strategy is obtained for *intermediate* delay values; at the two extremal delays of $\bar{D}_j = 1$ and $\bar{D}_j \rightarrow \infty$, no additional gains are obtained beyond equal delay allocation.

2) *Single-Hop Versus Multihop Transmission*: The plot of power versus delay for various numbers of hops is given in Fig. 9. The average channel gain $A_j \in \{1, 300\}$, i.e., each of the hops experiences one deep fade state with a probability of 0.5. The traffic arrival distribution is considered to be uniform distribution over support set $\{0, \dots, 4\}$ with average rate $\lambda = 2$. As expected, for a fixed m , the power decreases monotonically

as delay increases. However, the variation of power with m is nonmonotonic for a fixed end-to-end delay bound. For large delays, data are only sent during the *good* fading state at each hop. In that case, using an analysis that is similar to the Gaussian case, one can see that larger m results in lower power. For smaller delays, the optimal number of hops depends on the actual D_{ete} and fading characteristics. No generalizations such as Theorem 1 on the optimal allocation are possible.

For large λ as before, increasing the number of hops from 1 to 3 results in increased power. Furthermore, an increase in m beyond 3 results in a decrease in power. Similarly, for very small λ , increasing m beyond 1 results in decreased power consumption. The plots in these cases are similar to the Gaussian network and are not shown. The qualification of *large* and *small* λ depends on fading characteristics.

3) *Large-Delay Analysis*: For a fading network, at asymptotically large delays, the output traffic at each node depends on channel state \tilde{A}_{ji} and equals $m_{ji}(\infty)$ (36); note, however, that λ in (36) should be replaced by $k\lambda$ since each hop is only used once every k time slots. The total power in the network is given by

$$\bar{P}_{net} = m \frac{1}{k} \left(\frac{d_0}{m} \right)^{2\beta} \sum_{i=1}^{n_{ch}-states} \frac{(e^{m_{ji}(\infty)} - 1)}{|\tilde{A}_{ji}|^2} \quad (17)$$

assuming identical channel statistics across all hops. The behavior of power versus λ is similar to the Gaussian case (Fig. 8), except for a different scaling factor $\sum_i (e^{m_{ji}(\infty)} - 1)/|A_{ji}|^2$, instead of $(e^{k\lambda} - 1)/|A_j|^2$ in (16).

4) *Empirical Properties*: As discussed in the previous cases, the optimal allocation of delay depends on the characteristics of the source and the fading channels at the different hops. Although no generalization such as Theorem 1 can be made, we observed the following properties.

Empirically observed property 1: Given an end-to-end delay D_{ete} across a m -hop wireless network, the variation of delay at each of the hops is an affine function of the total delay D_{ete} . The slope and intercept of this affine function depends on the input traffic arrival statistics and the fading characteristics of the different hops. This property is illustrated in Fig. 10 for a simple four-hop network. Since $m = 4$, the minimum $D_{ete} = 5$. The fading gains in the different hops are chosen at random and are given by $A_1 \in \{300\}$, $A_2 \in \{3, 30\}$, $A_3 \in \{8, 20\}$, and $A_4 \in \{5, 10\}$. Each of these states has equal probability. The input traffic has uniform distribution over support set $\{0, \dots, 10\}$. This property could be used to derive closed-form delay allocation strategies in future work.

Empirically observed property 2: We found that, using the optimal allocation of delay at the various nodes, the spectra of the transmitted data rates across the various links were “matched” to one another (e.g., Fig. 11 shows the spectrum in the general case of fading channels and bursty source in which delay is split between the two links). This critical observation leads us to conjecture that a deeper connection exists between scheduler design and spectral analysis. The “matching” of sources and channels at the bit/signal level has been considered [50] to optimize performance. Our results indicate that similar matchings between traffic arrivals and the fading process occur

⁵A similar observation on output entropy of a queue is made in [49].

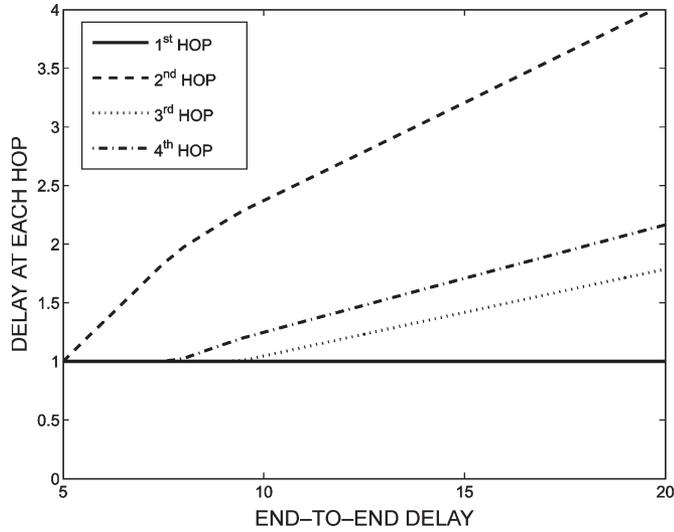


Fig. 10. Allocation of delay at each of the hops as a function of end-to-end delay. The affine nature of the dependence of delay at each hop for large end-to-end delay is clearly evident.

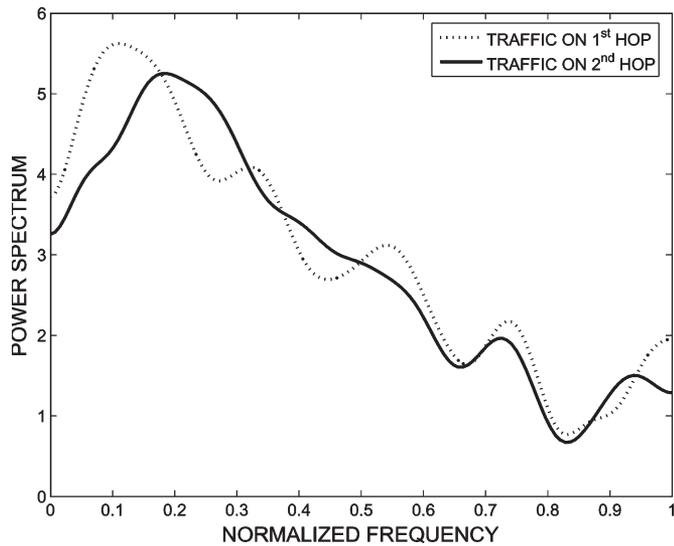


Fig. 11. Plot of power spectrum of the output traffic in a simple two-hop wireless network.

at the packet level. Unlocking this connection could lead to delay allocation strategies that are similar to the impedance matching approach and should be considered in future work.

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a framework for studying the delay allocation problem in an *ad hoc* wireless network. A closed-form expression for the total required power in a network is derived as a function of the delay allocation, and approximation is exploited to find near optimal schedulers.

This work can be extended in many different ways. For instance, the effect of multiuser interference on delay allocation problem needs to be considered in future work. The proposed framework can also be used to make routing decisions. Specifically, to compare between two different routes, the proposed framework can be used to compute the total power that is

required in the two routes for a given end-to-end delay constraint. The total power can then be used as the metric to make routing decisions. Furthermore, other channel allocation strategies beyond the temporal fair allocation that is used in this paper should be considered in future investigations.

APPENDIX A DERIVATION OF APPROXIMATE CLOSED-FORM EXPRESSION FOR POWER CONSUMPTION IN FADING CHANNELS WITH DELAY CONSTRAINTS

Recall that the buffer update is given by

$$x_{n+1,j} = x_{n,j} + u_{n,j-1} - u_{n,j}. \quad (18)$$

Taking expectation on both sides of (18), we obtain

$$\mathbb{E}[u_{n,j-1}] = \mathbb{E}[u_{n,j}] = \mathbb{E}[a_n] = \lambda. \quad (19)$$

Squaring (18) and taking expectation on both sides, we get

$$\begin{aligned} \mathbb{E}[x_{n+1,j}^2] &= \mathbb{E}[x_{n,j}^2] + \mathbb{E}[u_{n-1,j-1}^2] + \mathbb{E}[u_{n,j}^2] \\ &\quad + 2\lambda\mathbb{E}[x_{n,j}] - 2\lambda^2 - 2\mathbb{E}[x_{n,j}u_{n,j}] \\ &\implies \sigma_{u_{j-1}}^2 + \sigma_{u_j}^2 + 2\lambda^2 D_j = 2\mathbb{E}[x_{n,j}u_{n,j}] \end{aligned} \quad (20)$$

where the last step is obtained by using (3). We now approximate the scheduler action to be a linear function of the buffer occupancy; a similar approximation is used for scheduling over an AWGN channel [31]. Denote by $u_{n,j;i}$ the number of packets that are transmitted at time slot n if the fading channel state in the j th hop equals A_{ji} . Then

$$u_{n,j;i} = \mu_{ji}x_{n,j} + \nu_{ji} \quad (21)$$

$$\implies \mathbb{E}[u_{n,j;i}] = m_{ji}(D_j) = \mu_{ji}\lambda D_j + \nu_{ji} \quad (22)$$

where $m_{ji}(D_j)$ represents the average number of packets that are transmitted in fading channel state A_{ji} , and μ_{ji} and ν_{ji} are constants that determine the number of packets to be transmitted in fading state i at node N_j [31]. Squaring (21) and taking expectation, we obtain

$$\mathbb{E}[u_{n,j;i}^2] = \mu_{ji}^2\mathbb{E}[x_{n,j}^2] + \nu_{ji}^2 + 2\mu_{ji}\nu_{ji}\lambda D_j. \quad (23)$$

Squaring (22) and subtracting from (23), the variance of the output traffic is given by

$$\sigma_{u_{j;i}}^2 = \mu_{ji}^2\mathbb{E}[x_{n,j}^2] - \mu_{ji}^2\lambda^2 D_j^2 = \mu_{ji}^2\sigma_{x_j}^2. \quad (24)$$

Thus

$$\frac{\sigma_{u_{j;i}}}{\mu_{ji}} = \sigma_{x_j} \quad \forall i. \quad (25)$$

We now use the following heuristic functional form for selecting $m_{ji}(D_j)$ as a function of delay:

$$m_{ji}(D_j) = m_{ji}(\infty) + \frac{\lambda - m_{ji}(\infty)}{D_j} = m_{ji}(\infty) \frac{D_j - 1}{D_j} + \frac{\lambda}{D_j}. \quad (26)$$

This heuristic is basically a linear interpolation for the average transmission rate in each fading state, between λ and $m_{ji}(\infty)$. The heuristic form is thus exact at the two extreme delays of 1 and ∞ . The optimal transmission rate $m_{ji}(\infty)$ in the various fading states, at infinite delay, is calculated in Appendix B. Substituting from (21) into (20)

$$\begin{aligned} & \sigma_{u_{j-1}}^2 + \sigma_{u_j}^2 + 2\lambda^2 D_j \\ &= 2 \left\{ \sum_i p_{ji} \mu_{ji} \mathbb{E}[x_{n,j}^2] + \lambda D_j \sum_i p_{ji} \nu_{ji} \right\}. \end{aligned} \quad (27)$$

Now, we set $\mu_{ji} = 1/\lambda D_j \forall i$ and $\sigma_{u_{j,i}} = \sigma_{u_{j,k}} \forall i, k$. Thus, $\nu_{ji} = m_{ji}(D_j) - 1 \forall i$. Using these values for μ_{ji} and ν_{ji} , along with (25), we obtain

$$\begin{aligned} & \sigma_{u_{j-1}}^2 + \sigma_{u_j}^2 + 2\lambda^2 D_j \\ &= \dots 2 \sum_i \left(\frac{p_{ji}}{\mu_{ji}} \left(\sigma_{u_{j,i}}^2 + \mu_{ji}^2 \lambda^2 D_j^2 \right) \right. \\ & \quad \left. + \lambda D_j p_{ji} m_{ji}(D_j) \right) - 2\lambda D_j \\ &= 2\lambda D_j \left(\sigma_{u_{j,i}}^2 + 1 \right) + 2\lambda^2 D_j \\ & \quad - 2\lambda D_j \text{ (for any } i) \\ &= 2\lambda D_j \left[\sigma_{u_{j,i}}^2 + \lambda \right]. \end{aligned} \quad (28)$$

Characterizing the Output Traffic Variance

The variance of the traffic output at a node can be derived as

$$\mathbb{E}[u_{n,j}^2] = \sum_i p_{ji} \mathbb{E}[u_{n,j,i}^2] \quad (29)$$

$$\Rightarrow \sigma_{u_j}^2 = \sum_i p_{ji} \left\{ \sigma_{u_{j,i}}^2 + m_{ji}^2(D_j) \right\} - \lambda^2 \quad (30)$$

$$= \sigma_{u_{j,k}}^2 + \sum_i p_{ji} m_{ji}^2(D_j) - \lambda^2 \text{ (for any } k). \quad (31)$$

Substituting from (31) into (28), we obtain

$$\begin{aligned} \sigma_{u_{j,k}}^2 &= \frac{1}{2\lambda D_j - 1} \left[\sigma_{u_{j-1}}^2 - \lambda^2 + \sum_i p_{ji} m_{ji}(D_j)^2 \right] \\ &= \frac{1}{2\lambda D_j - 1} \left[\sigma_{u_{j-1}}^2 + \frac{(D_j - 1)^2}{D_j^2} \right. \\ & \quad \left. \times \left(\sum_i p_{ji} m_{ji}(\infty)^2 - \lambda^2 \right) \right]. \end{aligned} \quad (32)$$

Now, combining (31) and (32), we obtain the variance of the output traffic in terms of the variance of the input traffic as

$$\sigma_{u_j}^2 = \frac{\sigma_{u_{j-1}}^2}{2\lambda D_j - 1} + \frac{2\lambda D_j}{2\lambda D_j - 1} \frac{(D_j - 1)^2}{D_j^2} \sigma_{\text{ch}}^2 \quad (33)$$

where σ_{ch}^2 is defined in (10). The transmit power in each channel state is calculated by assuming that the traffic output

in each channel state has a Gaussian distribution with mean $m_{ji}(D_j)$ and variance $\sigma_{u_{n,j,i}}^2$. Thus

$$\begin{aligned} \mathbb{E}[P_{n,j}] &= \bar{P}_j = \sum_{i=1}^{n_{\text{ch-states}}} \frac{1}{\sqrt{2\pi\sigma_{u_j}^2}} \\ & \quad \dots \int_{-\infty}^{\infty} \frac{p_{ji} d_{j,j+1}^{2\beta}}{|A_{ji}|^2} (e^{u_{n,j,i}} - 1) e^{-\frac{(u_{n,j,i}-\lambda)^2}{2\sigma_{u_{j,i}}^2}} du_{n,j,i}. \\ &= \sum_{i=1}^{n_{\text{ch-states}}} \frac{p_{ji} d_{j,j+1}^{2\beta}}{|A_{ji}|^2} \left(e^{\lambda + \frac{\sigma_{u_{j,i}}^2}{2}} - 1 \right). \end{aligned} \quad (34)$$

APPENDIX B

OPTIMAL TRANSMISSION RATE IN EACH FADING STATE AT INFINITE DELAY

The average rate of transmission $m_{ji}(\infty)$ in each channel-fading state at infinite delay can be easily computed using standard water-filling techniques as follows: The total power over a fading channel is given by $\sum_i (p_{ji}/|A_{ji}|^2)(e^{m_{ji}(\infty)} - 1)$. The rates $m_{ji}(\infty)$ are computed to minimize the total power as follows:

$$m_{ji}(\infty) = \arg \min_{\substack{\sum_i p_{ji} m_{ji}(\infty) = \lambda, \\ \sum_i p_{ji} = 1, m_{ji}(\infty) > 0}} \sum_i \frac{p_{ji} d_{j,j+1}^{2\beta}}{|A_{ji}|^2} (e^{m_{ji}(\infty)} - 1). \quad (35)$$

Using standard Lagrangian methods, we can easily show that

$$m_{jk}(\infty) = \frac{1}{\sum_i p_{ji}} \left(\lambda + \sum_i 2p_{ji} \log \left(\frac{A_{jk}}{A_{ji}} \right) \right) \quad (36)$$

where i is the summation over states for which $0 \leq m_{ji}(\infty) \leq (\lambda/p_{ji})$.

APPENDIX C

CONVEXITY OF OPTIMIZING FUNCTION (14)

We know that the sum of convex functions is a convex function; hence, it is sufficient to prove that each of the terms inside the summation in (14) is convex. Now, consider the term $(e^{m_{ji}(D_j) + (\sigma_{u_{j,i}}^2(D_j)/2)} - 1)$. Since e^x is a convex nondecreasing function, it is sufficient to show that the term in the exponent is a convex function (see [51] for rules on convexity of composite functions). Using (26) and (32), the term in the exponent can be equivalently rewritten in the form $\sum_a D_j^a$, where $a < 0$. Each of these terms is a convex function of D_j (see, e.g., [51]). Hence, the function is convex in $D_j \forall j$.

REFERENCES

- [1] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.
- [2] N. Bansal and Z. Liu, "Capacity, delay and mobility in wireless ad-hoc networks," in *Proc. INFOCOM*, San Francisco, CA, Apr. 2003, pp. 1553–1563.

- [3] U. Kozat and L. Tassioulas, "Throughput capacity with random adhoc networks with infrastructure support," in *Proc. 9th Annu. ACM Int. Conf. Mobile Comput. Netw.*, Sep. 2003, pp. 55–65.
- [4] B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *Proc. INFOCOM*, San Francisco, CA, Apr. 2003, pp. 1543–1552.
- [5] R. Negi and A. Rajeswaran, "Capacity of power constrained ad-hoc networks," in *Proc. INFOCOM*, 2004, pp. 443–453.
- [6] M. Grossglauser and D. Tse, "Mobility increases capacity of adhoc wireless networks," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 477–486, Aug. 2002.
- [7] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case," in *Proc. INFOCOM*, New York, Jun. 2002, pp. 1577–1586.
- [8] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 765–779, Mar. 2005.
- [9] V. S. Anil Kumar, M. V. Marathe, S. Parthasarathy, and A. Srinivasan, "End-to-end packet scheduling in wireless ad-hoc networks," in *Proc. ACM-SIAM SODA*, 2004, pp. 1014–1023.
- [10] H. Balakrishnan, C. L. Barrett, V. S. Anil Kumar, M. V. Marathe, and S. Thite, "The distance-2 matching problem and its relationship to the MAC-layer capacity of ad-hoc wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1069–1079, Aug. 2004.
- [11] S. Ramanathan and E. Lloyd, "Scheduling algorithms for multihop radio networks," *IEEE/ACM Trans. Netw.*, vol. 1, no. 2, pp. 166–172, Jan. 1993.
- [12] C. Barrett, G. Istrate, M. Marathe, S. Thite, and V. S. Anil Kumar, *Link Scheduling Problems in Packet Radio Networks*, 2001. unpublished manuscript.
- [13] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 2, pp. 124–142, Mar. 1985.
- [14] A. Ephremides and B. Hajek, "Information theory and communication networks: An unconsummated union," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2416–2434, Oct. 1998.
- [15] A. Ephremides, "Some wireless networking problems with a theoretical conscience," in *Codes, Curves and Signals. Common Threads in Communications*. Norwell, MA: Kluwer, 1998, pp. 201–212.
- [16] T. ElBatt and A. Ephremides, "Joint scheduling and power control for wireless ad-hoc networks," in *Proc. INFOCOM*, 2002, pp. 976–984.
- [17] E. Yeh, "Delay optimal rate allocation in multiaccess communications: A cross layer view," in *Proc. Int. Workshop Multimedia Signal Process.*, St. Thomas, U.S. Virgin Islands, Dec. 2002, pp. 404–407.
- [18] L. Tong, V. Naware, and P. Venkatasubramanian, "Signal processing in random access: A cross-layer perspective," *IEEE Signal Process. Mag.*, vol. 21, no. 5, pp. 29–39, Sep. 2004.
- [19] R. A. Berry and E. M. Yeh, "Cross-layer wireless resource allocation," *IEEE Signal Process. Mag.*, vol. 21, no. 5, pp. 59–68, Sep. 2004.
- [20] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-layer design for wireless networks," *IEEE Commun. Mag.*, vol. 41, no. 10, pp. 74–80, Oct. 2003.
- [21] E. Shih, S. H. Cho, N. Ickes, R. Min, A. Sinha, A. Wang, and A. Chandrakasan, "Physical layer driven algorithm and protocol design for energy-efficient wireless sensor networks," in *Proc. MOBICOM*, Rome, Italy, Jul. 2001, pp. 272–287.
- [22] D. Wu and R. Negi, "Power control and scheduling for guaranteeing quality of service in cellular networks," in *Wirel. Commun. Mob. Comput., to be published*.
- [23] B. E. Collins and R. L. Cruz, "Transmission policies for time varying channels with average delay constraints," in *Proc. Allerton Int. Conf. Commun., Control Comput.*, Monticello, IL, 1999, pp. 709–717.
- [24] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "CDMA data QoS scheduling on the forward link with variable channel conditions," Bell Labs., Lucent Technol., Murray Hill, NJ, Apr. 2000. Tech. Rep.
- [25] N. Joshi, S. R. Kadaba, S. Patel, and G. S. Sundaram, "Downlink scheduling in CDMA data networks," in *Proc. MOBICOM*, 2000, pp. 179–190.
- [26] A. Bedekar, S. C. Borst, K. Ramanan, P. A. Whiting, and E. M. Yeh, "Downlink scheduling in CDMA data networks," *Probability, Networks and Algorithms (PNA)*, 1999.
- [27] R. G. Gallager, *Information Theory and Reliable Communication*. Hoboken, NJ: Wiley, 1968.
- [28] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1135–1149, May 2002.
- [29] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay and rate constrained transmission policies over wireless channels," in *Proc. GLOBECOM*, San Antonio, TX, Nov. 2001, pp. 806–810.
- [30] B. Prabhakar, E. U. Biyikoglu, and A. El Gamal, "Energy-efficient transmission over a wireless link via lazy packet scheduling," in *Proc. INFOCOM*, Anchorage, AK, Apr. 2001, pp. 386–394.
- [31] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay bounded packet scheduling of bursty traffic over wireless channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 125–144, Jan. 2004.
- [32] A. El Gamal, C. Nair, B. Prabhakar, E. Uysal-Biyikoglu, and S. Zahedi, "Energy-efficient scheduling of packet transmissions over wireless networks," in *Proc. INFOCOM*, 2002, pp. 1773–1782.
- [33] M. Sikora, J. N. Laneman, M. Haenggi, D. J. Costello, Jr., and T. E. Fuja, "Bandwidth- and power-efficient routing in linear wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2624–2633, Jun. 2005.
- [34] R. Mathar and J. Mattfeldt, "Analyzing routing strategy NFP in multihop packet radio networks on a line," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 977–988, Feb.–Apr. 1995.
- [35] J. A. Silvester and L. Kleinrock, "On the capacity of multihop slotted aloha networks with regular structure," *IEEE Trans. Commun.*, vol. COM-31, no. 8, pp. 974–982, Aug. 1983.
- [36] J. Jun and M. L. Sichitiu, "The nominal capacity of wireless mesh networks," *IEEE Wireless Commun.*, vol. 10, no. 5, pp. 8–14, Oct. 2003.
- [37] C. Florens, M. Sharif, and R. McEliece, "Delay issues in linear sensory networks," *IEEE Trans. Inf. Theory*, Aug. 2006. Abridged version appeared at IEEE International Symp. Inf. Theory, Jul. 2004.
- [38] B. Muller, *Filling Stations for Cell Phones*, 2002. [Online]. Available: http://w4.siemens.de/Ful/en/archiv/pof/heft1_02/artikel04/
- [39] C.-F. Chiasserini and R. R. Rao, "Energy efficient battery management," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 7, pp. 1235–1245, Jul. 2001.
- [40] D. Rajan, A. Sabharwal, and B. Aazhang, "Outage behaviour with delay and CSIT," in *Proc. ICC*, Paris, France, Jun. 2004, pp. 578–582.
- [41] A. Sabharwal, D. Rajan, and B. Aazhang, "Dual problems in power control," in *Proc. Allerton Conf. Commun., Control Comput.*, Monticello, IL, Oct. 2001.
- [42] D. P. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
- [43] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [44] E. Biglieri, J. G. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [45] D. J. Goodman, J. Borras, N. B. Mandayam, and R. Yates, "INFOSTATIONS: A new system model for data and messaging services," in *Proc. Veh. Technol. Conf.*, May, 1997, pp. 969–973.
- [46] D. Rajan, "Towards universal power efficient scheduling in wireless channels," in *Proc. ICC*, Paris, France, Jun. 2004, pp. 123–127.
- [47] E. M. Royer and C. K. Toh, "A review of current routing protocols for ad hoc mobile wireless networks," *IEEE Commun. Mag.*, vol. 6, no. 2, pp. 46–55, Apr. 1999.
- [48] V. Paxson, "Empirically-derived analytic models of wide-area TCP connections," *IEEE/ACM Trans. Netw.*, vol. 2, no. 4, pp. 316–336, Aug. 1994.
- [49] B. Prabhakar and R. G. Gallager, "Entropy and the timing capacity of discrete queues," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 357–370, Feb. 2003.
- [50] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
- [51] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, Mar. 2004.



Dinesh Rajan (S'99–M'02) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, India, in 1997 and the M.S. and Ph.D. degrees in electrical and computer engineering from Rice University, Houston, TX, in 1999 and 2002, respectively.

In 2002, he joined the Department of Electrical Engineering, Southern Methodist University, Dallas, TX, where he is currently an Assistant Professor. His current research interests include communications theory, wireless networks, and information theory.

Prof. Rajan was the recipient of a National Science Foundation CAREER award in 2006.