Power efficient scheduling in multihop networks

Dinesh Rajan

Department of Electrical Engineering,
Southern Methodist University,
Dallas, Texas 75205.
E-mail: rajand@engr.smu.edu

ABSTRACT

In this paper, we present schedulers that minimize the total transmit power in a multihop wireless network. The focus is on guaranteeing an end-to-end delay bound for a single variable bit rate flow on a multihop fading channel. We first compute an analytical approximation for the transmit power required to send a variable bit rate source over a finite state fading channel. We then use this approximation to derive schedulers that have low complexity and near optimal performance over multihop networks where 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.

Keywords: Queuing, Scheduling, Multihop, Wireless network, Delay guarantees

1. INTRODUCTION

The capacity of wireless ad-hoc networks has been an area of significant research interest beginning with some significant results in.[1] Capacity issues have been considered under different constraints on interference, maximum node range, node mobility, hybrid ad-hoc-infrastructure networks[2–8]; these results typically study the scaling of throughput with number of nodes. Scheduling and capacity in multihop networks with interference constraints has been considered by many recent researchers (see[9–12] and references therein).

The importance of incorporating traffic models in the design of physical layer of wireless systems has been well recognized[13–15] and cross layer optimization has been an area of significant recent research.[16–21] Numerous scheduling mechanisms have been proposed that explore the intricate relationship between packet delay, fairness, throughput and power. Many prior works on scheduling[22–25] operate in the power limited regime and use 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[26]) for power consumption with increasing number of packets transmitted has been used in some analysis.[27–31] 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. There are also other information theoretic studies of effect of delay on physical layer performace. For example, in fading channels the concept of delay limited capacity is defined[32] as the minimum rate that can be transmitted over all channel states; for many channels like Rayleigh fading, the delay-limited capacity turns out to be zero.

The main problem that we explore in this paper is the optimal allocation of total end-to-end delay at the different nodes to optimize a desired function (for example, summation) of the transmit powers at each of the nodes in an ad-hoc wireless network. We do not consider the problem of routing or tradeoffs between single-hop versus multi-hop transmissions. Instead, we assume that such decisions have been made by the higher layers.[*] The importance of optimal delay allocation can be motivated by comparing the optimal multihop delay allocation scheme with a scheme in which the available delay is uniformly allocated to each of the hops: This comparison is given in Table 1 for a simple 2-hop network. The sum of the transmit powers across the 2-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.

The main contributions of this paper can be succinctly summarized as below:

[*] Although in cross layer optimization, routing decisions could depend on lower layer functionality, we omit such possibilities for now. Joint scheduling and routing has been considered in.[33]

- We construct an analytical approximation for the power 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 (non-negative) combinations of the transmit power at each of the nodes. The optimal strategy allocates a delay of 1 time-slot to all but the first node; the remaining delay is given to the first node.
- For arbitrary ad-hoc networks, where each hop is modeled as a finite state block fading channel, we provide a low complexity scheduling mechanism that has near optimal performance. The complexity of this scheduler scales well with increasing number of nodes.

Although there are many factors that affect delay in a network, in this paper, we only consider the queuing delay. For simplicity, the analytical results are derived for simple *i.i.d.* traffic models; representative results for MPEG and Ethernet traffic sources are also given.

The rest of the paper is organized as follows. In Section 2 we describe the system under investigation and summarize selected prior results on single hop scheduling (Section 2.1). In Section 2.2, we analyze single hop fading channels and formulate the multihop problem in Section 2.3. In Section 3, we explore scheduling in multihop networks; specifically, Section 3.1 focuses on multihop Gaussian networks and Section 3.2 focuses on multihop fading networks. Finally, we summarize and provide some conclusions in Section 4.

2. PRELIMINARIES AND PROBLEM FORMULATION

Consider an ad-hoc wireless network with nodes $N_1, N_2, \ldots, N_{m+1}$, which carries a single flow in the network. Let nodes N_1 and N_{m+1} be the source and destination, respectively, for the traffic (or flow). Further, assume that the packet is routed via nodes $N_2, N_3, \ldots, N_m$. In this paper, we assume that each node N_i is capable of simultaneously transmitting data to node N_{i+1} and recieving data from node N_{i-1}. Further, we do not consider the interference effect of one transmission on other transmissions in the network.

The traffic generated at source node N_1 is assumed to be bursty (time-varying) and is stored in a buffer of size B_1. We consider a time-slotted system and the source produces packets at an average rate of λ packets/time-slot. In time-slot n, the source produces a_n packets, each of 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. We assume that the buffers are large enough that buffer overflows do not occur. Analysis of packet outages or losses resulting for buffer overflows are given in.[34, 35] 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 is varied based on the selected number of packets, $u_{n,j}$.

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

$$Y_{n,j+1} = A_{n,j} X_{n,j} + \varepsilon_{n,j}, \tag{1}$$

where $\varepsilon_{n,j}$ is the complex circularly symmetric additive white Gaussian noise with zero mean and variance σ^2. For simplicity, we set $\sigma^2 = 1$. The channel gain $A_{n,j}$ 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. We use two models for the channel between node N_j and N_{j+1}; an AWGN channel and a block fading channel. In the AWGN channel model, we set $A_{n,j} = A_j \forall n$. In the block fading channel model, $A_{n,j} \in \{A_{j:i}, i = 1, \ldots, n_{ch-states}\}$. Further, the channel state at each hop is assumed to form a Markov chain with transition probabilities given by $q_{j:il}$, where $q_{j:il} = Pr\{A_{n+1,j} = A_{jl}/A_{n,j} = A_{ji}\}$. The invariant distribution of the Markov chain, denoted p_{ji}, can be easily calculated from this transition probability. Recognize that p_{ji} also represent the probability that the channel gain in the j^{th} hop equals A_{ji}.

Delay bound	Power of optimal allocation	Power based on on low complexity allocation	Power with equal delay division
2.5	3.39	3.45	4.2
7.5	0.31	0.39	1.41

Table 1. Total transmit power required in the network for three different delay allocation strategies. The power (in dB) normalized with respect to the power required at infinite delay.

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

$$x_{n+1,j} = x_{n,j} + u_{n,j-1} - u_n, \forall j = 1, \dots, m, \tag{2}$$

where, for notational simplicity we let $u_{n,0} = a_n$. At each node, the average packet delay, D_j, is related to the average buffer length via Little's theorem[36] as,

$$D_j = \frac{1}{\lambda} \mathbb{E}\left\{x_{n,j}\right\}, \tag{3}$$

where, $\lambda = \mathbb{E}\left\{a_n\right\}$ is the average packet arrival rate. We assume that all packets that arrive in time-slot n can be transmitted only in time-slot $(n+1)$ or later. A natural constraint on u_n is that $0 \leq u_n \leq x_n$. The smallest average delay in the system is achieved when all buffered packets are transmitted in each time-slot, *i.e.*, $u_{n,j} = x_{n,j}$, which implies that $D_j = \frac{1}{\lambda}\mathbb{E}\left\{u_{n,j}\right\} = 1$. The average end-to-end delay is given by $D_{ete} = \sum_{j=1}^{m+1} D_j$. Using our convention, the minimum end-to-end delay, D_{ete}, equals m, the number of hops.

The transmit power at node N_j during time-slot n depends on the number of packets transmitted, the coding and modulation scheme used and the desired performance in terms of bit error rate (BER) or frame error rate (FER). 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: a closed form expression can be used for required power and it also provides a lower bound on the power required by any practical transmission strategy. Moreover, advanced codes like LDPC and Turbo codes achieve performance very close to capacity with finite block lengths. Thus, we use the famous Gaussian capacity formulation[37] to derive the required power as $P_{n,j:i} = \frac{1}{|A_{ji}|^2}(e^{u_{n,j:i}} - 1)$. The total network power is then given by

$$\mathbb{E}\left[\sum_{i,j} p_{ji} P_{n,j:i}\right] = \sum_j \mathbb{E}\left[\sum_i p_{ji} P_{n,j:i}\right], \tag{4}$$

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

2.1. Single flow, single hop scheduling

We now summarize prior results on scheduling over a single hop wireless channel that will form the basis for the analysis in this paper. It is well known that on a fading channel, the optimal transmission scheme could exploit knowledge of instantaneous channel conditions to improve performance.[38] For instance, one could delay the transmission of packets on *bad* fading states and transmits more packets in the *good* channel states. Such a transmission policy results in reduced transmission power at the cost of increased delay 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, for example, in the INFOSTATION architecture.[39] Further, using such a transmission scheme the output traffic 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.[28,30] 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 lesser variation than the input traffic.

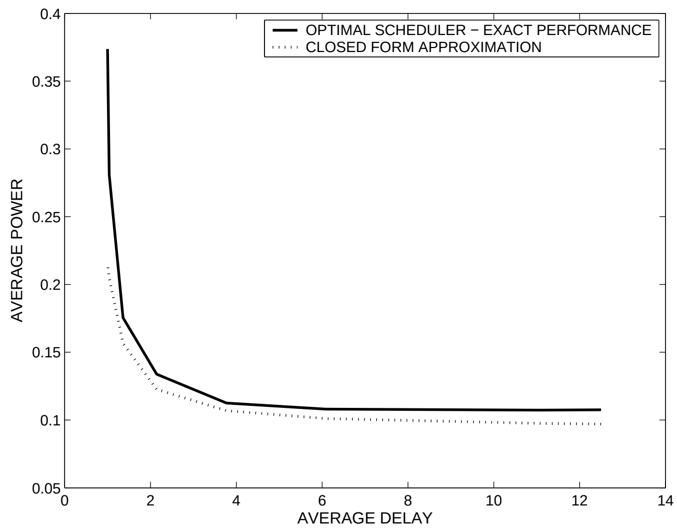

Figure 1. Plot to demonstrate the accuracy of the proposed closed form approximation for power required versus delay in a single hop fading channel.

The minimal power delay bounded scheduler for transmitting bursty source over a fading channel has been derived[27, 28, 30] using dynamic programming methods. The optimal scheduler action is a combination of the action in the two cases above; smoothing of bursty arrivals and introducing rate variations to transmit more packets in good channel states and few packets in bad channel states. The variation of the transmit power with delay is shown in Figure 1; it can be seen that substantial savings in power are possible for small increases in delay.

Preliminary investigations have revealed that the scheduler 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.[40] 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 both the input is bursty and the channel is fading, the optimal scheduler exhibits behavior that is a combination of low-pass and high-pass; its actually a band-reject filter. These characteristics are depicted in Figure 2. The low pass, high pass and band reject nature of the output traffic spectrum is clearly evident from the figure.

2.2. Single hop fading channels: Analytical Performance evaluation

As will become evident from the next section, finding an exact solution to the delay allocation problem in the general multihop network is a problem of high computational complexity. In Section 3.2, we present a heuristic low-complexity method to obtain near optimal solution. The basis for our method is the derivation of a simple closed form analytical expression for the required transmit power at each hop as a function of the traffic, fading channel and delay bound. 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,

$$\bar{P}_j = \sum_{i=1}^{n_{ch-states}} \frac{p_{ji}}{|A_{ji}|^2} \left(e^{m_{ji}(D_j) + \frac{\sigma^2_{u_{j:i}}(D_j)}{2}} - 1 \right), \tag{5}$$

where, D_j is the delay, $m_{ji}(D_j)$ is the mean transmission rate, $\sigma^2_{u_{j:i}}(D_j)$ 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 derive an approximate value for $m_{ji}(D_j)$ and $\sigma^2_{u_{j:i}}$ in terms of the mean arrival rate, variance of the arrival rate and the fading channel statistics. The accuracy of the approximation is then numerically

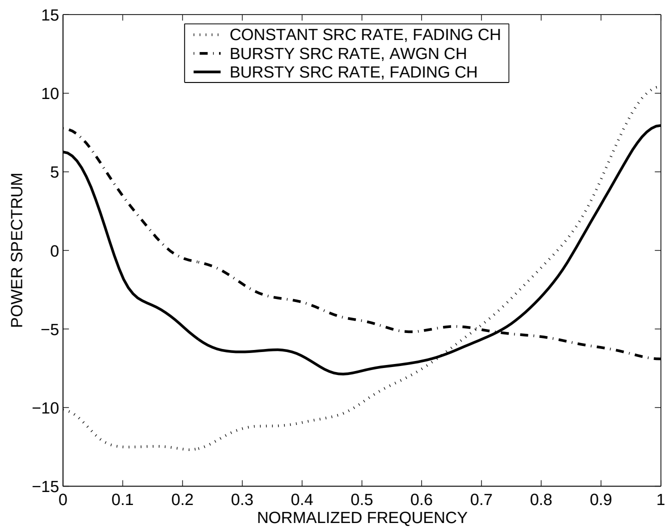

Figure 2. Frequency response of the output traffic at a node in three different scenarios.

studied and compared with the power of the optimal scheduler in Figure 1. 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-1} node forms the input traffic to node N_j. The variance of the output traffic at the node N_j can be derived (Appendix A) in terms of the variance of the input traffic to that node, as

$$\sigma_{u_j}^2 = \frac{\sigma_{u_{j-1}}^2}{2\lambda D_j - 1} + \frac{2\lambda D_j}{2\lambda D_j - 1}\frac{(D_j-1)^2}{D_j^2}\left[\sum_{i=1}^{n_{ch-states}} p_{ji} m_{ji}^2(\infty) - \lambda^2\right]. \tag{6}$$

An interesting point to note is that the variance of the output traffic could be lesser or greater than the variance of the input traffic. The output variance as shown in (6) depends on two factors: the first term due to the burstiness of the input source and the second term due to the burstiness introduced by matching the transmission rate to the channel state. For the transmission of a constant rate source, the first term equals 0 and the output variance depends only on channel fading characteristics. Note that the term $\sum_{i=1}^{n_{ch-states}} p_{ji} m_{ji}^2(\infty) - \lambda^2$ is a variance like quantity that shows how different the optimal transmission rates (equvalently gains) in the fading states are. In the case of an AWGN channel, the second term in (6) reduces to zero and the output variance only depends on input traffic variance and the delay.

The following examples qualitatively illustrate this relationship.

*E*xample 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 $Pr(A_{n,j} = 1) = 0.75, Pr(A_{n,j} = 100) = 0.25$. In this case the optimal transmission rates in the two fading states at infinite delay can be calculated using (33) as $m_{j1}(\infty) = 2.74$, $m_{j2}(\infty) = 11.76$. 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 2 for three different delay values. It is clearly evident that the output variance could be higher or lower than the input variance depending on the delay constraint. For this channel,

*E*xample 2: In this example, we let $A_j \in \{1, 100\}$ and set $Pr(A_{n,j} = 1) = 0.25, Pr(A_{n,j} = 100) = 0.75$. The optimal transmission rates in the two fading states at infinite delay are given by $m_{j1}(\infty) = 0$, $m_{j2}(\infty) = 6.67$. For same traffic arrivals as in Example 1, the ratio of output variance to input variance is given in Table 2. In this case, the output variance is strictly lesser than the input traffic variance.

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

Table 2. Ratio of output traffic variance to the input traffic variance for different channels and four different delays.

*E*xample 3: In this example, we let $A_j = 1$, *i.e.*, its an AWGN channel. In this case too, the output variance is strictly lesser than the input traffic variance as illustrated in Table 2.

Using this approximation for the power required at a node, the summation of the powers required in the entire network is given by,

$$\bar{P}_{net} = \sum_{j=1}^{m} \bar{P}_j = \sum_{j=1}^{m} \sum_{i=1}^{n_{ch-states}} \frac{p_{ji}}{|A_{ji}|^2} \left(e^{m_{ji}(D_j) + \frac{\sigma^2_{u_{ji}}(D_j)}{2}} - 1 \right). \tag{7}$$

2.3. Problem formulation

The optimization problem os interest is posed as follows:

$$\min_{\sum_{j=1}^{m} D_j \leq \bar{D},\ D_j \geq 1} \quad \sum_j \mathbb{E}\left[\sum_i p_{ji} P_{n,j:i} \right], \tag{8}$$

where the variables of optimization are $u_{n,j:i}, \forall i, j, n$. Recognize that in its general form, this optimization problem is extremely complex to solve due to the dependency of input traffic at one node on the output traffic at the previous node. The optimal results given in Table 1 were computed in the simple 2-hop case based on an exhaustive search method and serves as a lower bound on the total transmit power required to support that traffic under delay constraint. Instead, we use a statistical characterization of the output traffic at each node to simplify the problem. Specifically, we characterize 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.

The optimization problem is then rewritten as follows:

$$\min_{\sum_{j=1}^{m} D_j \leq \bar{D},\ D_j \geq 1} \quad \bar{P}_{net} \tag{9}$$

Now, the variables of optimization are just the delays allocated to the various nodes; conditioned on a particular delay being allocated, the optimal scheduler for that hop is calculated using dynamic programming techniques.[27,30] It can be easily shown that the optimizing function in (9) is convex and the constraints are linear. Hence, this problem is easily solved (even for large number of hops) using a numerical optimizer (e.g., the *fmincon* constrained optimization function in MATLAB). In some simple cases, analytical solutions to (9) are also possible. The performance of the end to end scheduler based on this approximation is given in Table 1; it is clear that the low complexity solution is near optimal.

3. MULTIHOP NETWORKS

In this section, we investigate the delay allocation problem in a multihop network. Our strategy is to completely understand the effects in isolation of the bursty source and fading channel on the optimization problem and then combine the two effects.

3.1. Gaussian Channels, Bursty Source

In the specific case of a Gaussian network, *i.e.*, channel between each 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 generated at node N_1 is bursty. This traffic has to be transmitted to destination node N_{m+1} via nodes $N_2, N_3, \ldots, N_m$. There is an end-to-end delay constraint D_{ete} for flow. Any linear non-negative combination of the transmit powers at nodes $N_1, N_2, \ldots, N_m$ is minimized by allocating delay of $D_{ete} - 1$ at the first link and delay of 1 at all the other links.

Recall that in our notation, the minimum delay at each node equals 1 time-slot and thus the minimum delay over the entire network equals the number of hops.

Proof: The proof is by contradiction. The proof also uses the following property of scheduling over single hop Gaussian networks; the larger the delay allocated to a link, the smoother is the output traffic at that node and lower is the required transmit power.[27, 30]

In the "optimal" scheduler let the delay at two successive hops be D_i and D_{i+1}, where $D_{i+1} \neq 1$ and let the corresponding transmit powers equal P_i and P_{i+1}. We now construct a different scheduler, labeled ALT, at nodes N_i and N_{i+1} with delays D'_i and D'_{i+1} such that $D'_i + D'_{i+1} = D_i + D_{i+1}$. Further, we will show that the total power of scheduler ALT, is lower than the "optimal" scheduler.

Consider the two different buffers B_i and B_{i+1} at nodes N_i and N_{i+1}. Now, in scheduler ALT, assume that both the buffers are at node i with one preceding the other and the scheduler at node N_{i+1} has delay of 1 time-slot. 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 modification; we assume that packets that arrive into 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 D'_i at node i equals $D_i + D_{i+1} - 1$ and delay $D'_{i+1} = 1$. Further, the output traffic at node N_{i+1} using the new scheduler is the same as the "optimal" scheduler and hence the transmit power at node N_{i+1} is unchanged. However, the transmit power at node N_i is reduced since it is allocated a larger delay. Hence, by contradiction the power is minimized by allocating a delay of 1 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 an 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, over an AWGN channel, the variance of the output traffic strictly decreases with increasing delay constraint. Thus, the total power is minimized by allocating all the delay to the first node.

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 the equal delay allocation strategy, it is assumed that each node schedules its packets optimally to minimize the local transmission power given delay constraints at that node. The ratio of the power required using equal delay allocation to all hops versus optimal allocation to all hops can be calculated using (7) as,

$$10 \log \left(\frac{P_{eq}}{P_{opt}} \right) = \frac{\sum_{k=1}^{m} e^{\frac{\sigma_a^2}{(2\lambda D_{ete}/m - 1)^k}}}{\sum_{k=1}^{m} e^{\frac{\sigma_a^2}{(2\lambda (D_{ete}-1)-1)(2\lambda-1)^{(k-1)}}}} \tag{10}$$

A plot of this ratio is given in Figure 3 for two different traffic arrival streams - one stream contains Ethernet traffic trace and the other stream is an MPEG trace. It is clear that substantial power savings result when the delay is allocated optimally across the multiple hops.

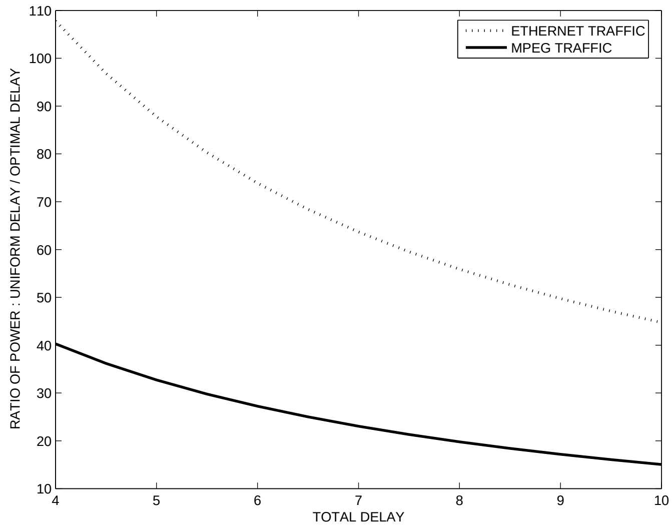

Figure 3. Plot of ratio of power required in dB with optimal delay allocation versus uniform delay allocation for two different traffic traces. A 4-hop wireless AWGN network is considered and hence smallest end-to-end delay equals 4.

3.2. Fading channel, constant rate source

Analogous to case 1, 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 than the input traffic at that node.[†] Qualitatively, larger the variation in $|A_{ij}|^2$ at a given hop, the larger is 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.

3.3. Bursty Source, Fading Channels

In this case, the optimal allocation of delay depends on the characteristics of the source and the fading channels at the different hops. Although no generalization like in case 1 (Theorem 1) can be made, we observed the following properties.

Empirically observed Property 1: Given an end to end delay D_{ete} across a n-hop wireless network, setting the delay at each of the hops as an affine function of the total delay D_{ete} results in near optimal performance. The slope and intercept of this affine function is (clearly) different for the different hops and depends on the input traffic arrival statistics, and the fading characteristics of the different hops. This property is illustrated in Figure 5 for a simple 4-hop network.

Empirically observed property 2: In our preliminary investigations, we found that in the optimal allocation of delay across the various nodes, the *spectrum* of the transmitted data rates across the various links were "matched" to one another. (*e.g.* Figure 4 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 approaches. The 'matching' of sources and channels at the bit/signal level has been considered[42] to optimize performance. Our results indicate that similar 'matchings' between traffic arrivals and fading process occurs at the packet level. Unlocking this connection could lead to scheduler designs similar to the impedance matching approach and should be considered in future work.

[†]A similar observation on output entropy of a queue is made in.[41]

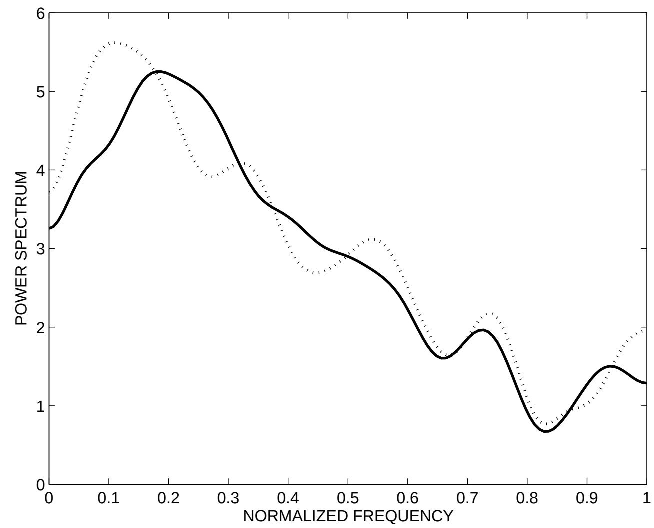

Figure 4. Plot of power spectrum of the output traffic in a simple 2-hop wireless network.

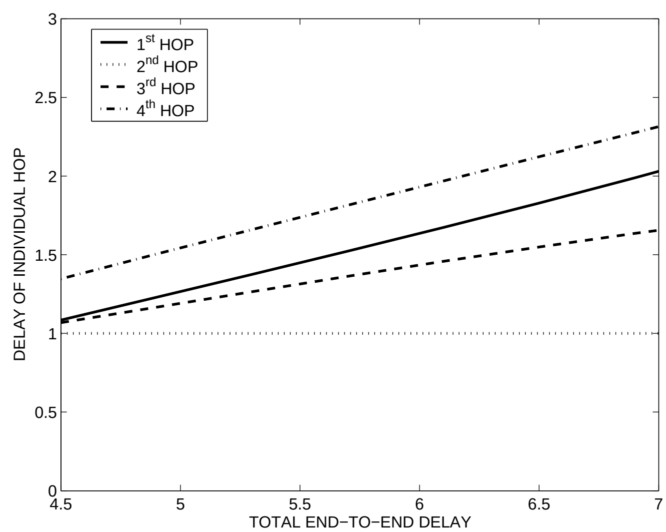

Figure 5. Near optimal allocation of delay at each of the hops as a function of end-to-end delay. The affine nature of the dependence of each individual hop delay on end-to-end delay is clearly evident.

4. CONCLUSIONS 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.

The proposed framework can also be used to making routing decisions. Specifically, to compare between two different routes, the proposed framework can be used to compute the total power 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. Such an algorithm is different from traditional power based routing algorithms, in that it provides a unified framework for considering the effect of source bustiness, channel variations and end-to-end delay bounds.

The proposed scheduling framework can also be extended to include physical constraints *e.g.* ensuring that each node cannot transmit and receive at the same time or introduce a distance-2 interference constraint that well models the 802.11 MAC.

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

Recall that the buffer update is given by

$$x_{n+1,j} = x_{n,j} + u_{n,j-1} - u_n, \tag{11}$$

assuming that there are no buffer overflows. Taking the expectation on both sides of the above equation, we obtain

$$\mathbb{E}[u_{n,j-1}] = \mathbb{E}[u_{n,j}] = \mathbb{E}[a_n] = \lambda. \tag{12}$$

Squaring (11) and taking the expectation on both sides, we get,

$$\begin{aligned} \mathbb{E}[x^2_{n+1,j}] &= \mathbb{E}[x^2_{n,j}] + \mathbb{E}[u^2_{n,j-1}] + \mathbb{E}[u^2_{n,j}] + 2\lambda\mathbb{E}[x_{n,j}] - 2\lambda^2 - 2\mathbb{E}[x_{n,j}u_{n,j}] && (13) \\ \implies \sigma^2_{u_{j-1}} + \sigma^2_{u_j} + 2\lambda^2 D_j &= 2\mathbb{E}[x_{n,j}u_{n,j}] \quad \text{using (3)} && (14) \end{aligned}$$

We now approximate the scheduler action to be linear function of the buffer occupancy; a similar approximation is used for scheduling over an AWGN channel.[30] Hence,

$$\begin{aligned} u_{n,j:i} &= \mu_{ji}x_{n,j} + \nu_{ji} && (15) \\ \implies \mathbb{E}[u_{n,j:i}] &= m_{ji}(D_j) = \mu_{ji}\lambda D_j + \nu_{ji}, && (16) \end{aligned}$$

where, $m_{ji}(D_j)$ represents the average number of packets transmitted in fading channel state A_{ji}. Squaring (15) and taking expectation, we obtain,

$$\begin{aligned} \mathbb{E}[u^2_{n,j:i}] &= \mu^2_{ji}\mathbb{E}[x^2_{n,j}] + \nu^2_{ji} + 2\mu_{ji}\nu_{ji}\lambda D_j && (17) \\ \implies \sigma^2_{u_{j:i}} &= \mu^2_{ji}\mathbb{E}[x^2_{n,j}] - \mu^2_{ji}\lambda^2 D^2_j = \mu^2_{ji}\sigma^2_{x_j} && (18) \end{aligned}$$

Thus,

$$\frac{\sigma_{u_{j:i}}}{\mu_{ji}} = \sigma_{x_j}, \tag{19}$$

which is a constant for all i. We now use a heuristic form for $m_{ji}(D_j)$ as

$$m_{ji}(D_j) = m_{ji}(\infty) + \frac{\lambda - m_{ji}(\infty)}{D_j} = m_{ji}(\infty)\frac{D_j - 1}{D_j} + \frac{\lambda}{D_j}. \tag{20}$$

This heuristic is basically a linear interpolation operation for average transmission rate in each fading state, between λ and $m_{ji}(\infty)$. The heuristic form is thus optimal at the two extreme delays of 1 and ∞. Substituting from (15) into (14),

$$\sigma^2_{u_{j-1}} + \sigma^2_{u_j} + 2\lambda^2 D_j = 2\left\{\sum_i p_{ji}\mu_{ji}\mathbb{E}[x^2_{n,j}] + \lambda D_j \sum_i p_{ji}\nu_{ji}\right\} \tag{21}$$

Now, we set $\mu_{ji} = \frac{1}{\lambda D_j}, \forall i$. Thus, $\nu_{ji} = m_{ji}(D_j) - 1, \forall i$. Using these values for μ_{ji} and ν_{ji}, along with (19), we obtain,

$$\begin{aligned}
\sigma^2_{u_{j-1}} + \sigma^2_{u_j} + 2\lambda^2 D_j &= 2\sum_i \left(\frac{p_{ji}}{\mu_{ji}} (\sigma^2_{u_{j:i}} + \mu^2_{ji}\lambda^2 D_j^2) + \lambda D_j p_{ji} m_{ji}(D_j) \right) - 2\lambda D_j && (22) \\
&= 2\lambda D_j(\sigma^2_{u_{j:i}} + 1) + 2\lambda^2 D_j - 2\lambda D_j \quad \text{for any } i && (23) \\
&= 2\lambda D_j[\sigma^2_{u_{j:i}} + \lambda] && (24)
\end{aligned}$$

Characterizing the output traffic variance

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

$$\begin{aligned}
\mathbb{E}[u^2_{n,j}] &= \sum_i p_{ji}\mathbb{E}[u^2_{n,j:i}] && (25) \\
\implies \sigma^2_{u_j} &= \sum_i p_{ji} \left\{ \sigma^2_{u_{j:i}} + m^2_{ji}(D_j) \right\} - \lambda^2 && (26) \\
&= \sigma^2_{u_{j:k}} + \sum_i p_{ji} m^2_{ji}(D_j) - \lambda^2 \quad \text{for any } k && (27) \\
&= \frac{\sigma^2_{u_{j-1}}}{2\lambda D_j - 1} - \frac{2\lambda D_j}{2\lambda D_j - 1} \frac{(D_j-1)^2}{D_j^2} \left[\lambda^2 - \sum_i p_{ji} m^2_{ji}(\infty) \right] && (28)
\end{aligned}$$

Applying (27) into (24), we obtain,

$$\begin{aligned}
\sigma^2_{u_{j:k}} &= \frac{1}{2\lambda D_j - 1} \left[\sigma^2_{u_{j-1}} - \lambda^2 + \sum_i p_{ji} m_{ji}(D_j)^2 \right] && (29) \\
&= \frac{1}{2\lambda D_j - 1} \left[\sigma^2_{u_{j-1}} - \lambda^2 \frac{(D_j-1)^2}{D_j^2} + \frac{(D_j-1)^2}{D_j^2} \sum_i p_{ji} m_{ji}(\infty)^2 \right] && (30)
\end{aligned}$$

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 $\sigma^2_{u_{n,j:i}}$. Thus,

$$\begin{aligned}
\mathbb{E}\{P_{n,j}\} = \bar{P}_j &= \sum_{i=1}^{n_{ch-states}} \frac{1}{\sqrt{2\pi\sigma^2_{u_j}}} \int_{-\infty}^{\infty} \frac{p_{ji}}{|A_{ji}|^2} (e^{u_{n,j:i}} - 1) e^{-\frac{(u_{n,j:i}-\lambda)^2}{2\sigma^2_{u_{j:i}}}} du_{n,j:i}. && (31) \\
&= \sum_{i=1}^{n_{ch-states}} \frac{p_{ji}}{|A_{ji}|^2} \left(e^{\lambda + \frac{\sigma^2_{u_{j:i}}}{2}} - 1 \right) && (32)
\end{aligned}$$

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. For the problem at hand, this transmission rate is computed as follows. The total power over a fading channel is given by $\mathbb{E}[P] = \sum_i \frac{p_{ji}}{|A_{ji}|^2} \left(e^{m_{ji}(\infty)} - 1 \right)$, where $\sum_i p_{ji} = 1$, $\sum_i p_{ji} m_{ji}(\infty) = \lambda$, and $m_i(\infty) > 0$. Using standard Lagrangian methods, we can easily show that

$$m_{jk}(\infty) = \frac{1}{\sum_i p_{ji}} \left(\lambda + \sum_i 2p_{ji} \log \left(\frac{A_{jk}}{A_{ji}} \right) \right), \tag{33}$$

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

ACKNOWLEDGMENTS

This work has been supported in part by Nokia Inc.

REFERENCES

1. P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. on Information Theory* **46**, pp. 388–404, March 2000.
2. N. Bansal and Z. Liu, "Capacity, delay and mobility in wireless ad-hoc networks," in *Proc. of INFOCOM*, (San Francisco, CA), April 2003.
3. U. Kozat and L. Tassiulas, "Throughput capacity with random adhoc networks with infrastructure support," in *Proc. 9^{th} Annual ACM International Conference on Mobile Computing and networking*, September 2003.
4. B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *Proc. of INFOCOM*, (San Francisco, CA), April 2003.
5. R. Negi and A. Rajeswaran, "Capacity of power constrained ad-hoc networks," in *Proc. of INFOCOM*, 2004.
6. M. Grossglauser and D. Tse, "Mobility increases capacity of adhoc wireless networks," *IEEE/ACM Transactions on Networking* **10**, pp. 477–486, August 2002.
7. M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case," in *Proc. of INFOCOM*, (New York, NY), June 2002.
8. M. Gastpar and M. Vetterli, "On the capacity of large gaussian relay networks," *to appear in IEEE Transactions on Information Theory* .
9. V. A. Kumar, M. V. Marathe, S. Parthasarathy, and A. Srinivasan, "End-to-end packet scheduling in wireless ad-hoc networks," in *Proc. ACM-SIAM Symposium on Discrete Algorithms(SODA)*, pp. 1014–1023, 2004.
10. H. Balakrishnan, C. L. Barrett, V. S. A. 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 Journal on Select Areas in Communications* **22**, pp. 1069–1079, August 2004.
11. S. Ramanathan and E. Lloyd, "Scheduling algorithms for multihop radio networks," *IEEE/ACM Transactions on Networking* , pp. 166–172, January 1993.
12. C. Barrett, G. Istrate, M. Marathe, S. Thite, and V. S. A. Kumar, "Link scheduling problems in packet radio networks," in *unpublished manuscript*, 2001.
13. R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. on Information Theory* **31**, pp. 124–142, Mar. 1985.
14. A. Ephremides and B. Hajek, "Information theory and communication networks: An unconsummated union," *IEEE Trans. on Information Theory* **44**, 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*, pp. 201–212, Kluwer Academic publishers, 1998.
16. T. ElBatt and A. Ephremides, "Joint scheduling and power control for wireless ad-hoc networks," in *Proc. of INFOCOM*, 2002.
17. E. Yeh, "Delay optimal rate allocation in multiaccess communications: A cross layer view," in *Proceedings of 2002 International workshop on multimedia signal processing*, (St. Thomas, U.S. Virgin Islands), December 2002.
18. L. Tong, V. Naware, and P. Venkitasubramanian, "Signal processing in random access : A cross-layer perspective," *IEEE Signal Processing Magazine* **21**, September 2004.
19. R. Berry and E. Yeh, "Cross layer wireless resource allocation - fundamental performance limits," *IEEE Signal Processing Magazine* , July 2004.
20. S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-layer design for wireless networks," *IEEE Communications Magazine* , October 2003.
21. E. Shih, S. 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 *Proceedings of MOBICOM*, (Rome, Italy), July 2001.
22. B. Collins and R. L. Cruz, "Transmission policies for time varying channels with average delay constraints," in *Proc. Allerton Intnl. Conf. on Comm., Control and Computing*, pp. 709–717, (Monticello, IL), 1999.

23. 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," *Technical Report, Bell Laboratories, Lucent Technologies.* , 2000.
24. N. Joshi, S. R. Kadaba, S. Patel, and G. S. Sundaram, "Downlink scheduling in CDMA data networks," in *Proceedings of MOBICOM*, pp. 179–190, 2000.
25. 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.
26. R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley and Sons, Inc., 1968.
27. R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. on Information Theory* **48**, pp. 1135–1149, May 2002.
28. D. Rajan, A. Sabharwal, and B. Aazhang, "Delay and rate constrained transmission policies over wireless channels," in *Proc. of GLOBECOM*, (San Antonio, TX), November 2001.
29. B. Prabhakar, E. U. Biyikoglu, and A. E. Gamal, "Energy-efficient transmission over a wireless link via lazy packet scheduling," in *Proc. of INFOCOM*, (Anchorage, Alaska), April 2001.
30. D. Rajan, A. Sabharwal, and B. Aazhang, "Delay bounded packet scheduling of bursty traffic over wireless channels," *IEEE Trans. on Information Theory* **50**(1), pp. 125–144, 2004.
31. A. E. Gamal, C. Nair, B. Prabhakar, E. Uysal-Biyikoglu, and S. Zahedi, "Energy-efficient scheduling of packet transmissions over wireless networks," in *Proc. of INFOCOM*, 2002.
32. S. Hanly and D. Tse, "Multi-access fading channels: Part II: Delay-limited capacities," *IEEE Trans. on Information Theory* **44**, pp. 2816–2831, November 1998.
33. M. Kodialam and T. Nandagopal, "Characterizing achievable rates in multi-hop wireless networks: the joint routing and scheduling problems," in *Proc.* 9^{th} *Annual ACM International Conference on Mobile Computing and networking*, pp. 42–54, 2003.
34. D. Rajan, A. Sabharwal, and B. Aazhang, "Outage behaviour with delay and CSIT," in *Proceedings of International Conference on Communications (ICC)*, pp. 578–582, (Paris, France), June 2004.
35. A. Sabharwal, D. Rajan, and B. Aazhang, "Dual problems in power control," in *Proceedings of Allerton Conference on Communication, Control and Computing*, (Monticello, IL), October 2001.
36. D. P. Bertsekas and R. Gallager, *Data Networks*, Prentice Hall, 1992.
37. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.
38. E. Biglieri, J. G. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. on Information Theory* **44**, pp. 2619–2692, October 1998.
39. D. J. Goodman, J. Borras, N. B. Mandayam, and R. Yates, "INFOSTATIONS: A new system model for data and messaging services," *Proceedings of Vehicular Technology Conference* , pp. 969–973, May 1997.
40. D. Rajan, "Towards universal power efficient scheduling in wireless channels," in *Proceedings of International Conference on Communications (ICC)*, pp. 123–127, (Paris, France), June 2004.
41. B. Prabhakar and R. G. Gallager, "Entropy and the timing capacity of discrete queues," *IEEE Trans. on Information Theory* **49**, pp. 357–370, February 2003.
42. M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. on Information Theory* **49**, pp. 1147–1158, May 2003.