The Book

The Man

May 14, 2006

Chapter 1

Opportunism in Wireless Networks: Principles and Techniques

Ahmad Khoshnevis and Ashutosh Sabharwal[1]

1.1 Opportunism: Avenues and Basic Principles

In this chapter, we will study how the variation in the sources and channels can be exploited to improve the power and spectral efficiency of wireless networks. A typical multi-hop wireless network is displayed in Figure 1.1, with an inside look into a typical node. At any node, there can be multiple sources with multiple queues, either based on the data generated at the node (a user web-surfing) or generated at other nodes (forwarding traffic in a multi-hop network like ad hoc, sensor or mesh network). Most sources produce information in time-varying bursts. This is clearly evident in multimedia sources like voice, music and video, which when compressed yield a time-varying amount of information. Voice conversations have periods of silence and speech interspersed, clearly de-marking periods of high informa-

[1]Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005, USA. Email: {farbod,ashu}@rice.edu. Phone: (713) 348-5057. Fax: (713) 348-6196.

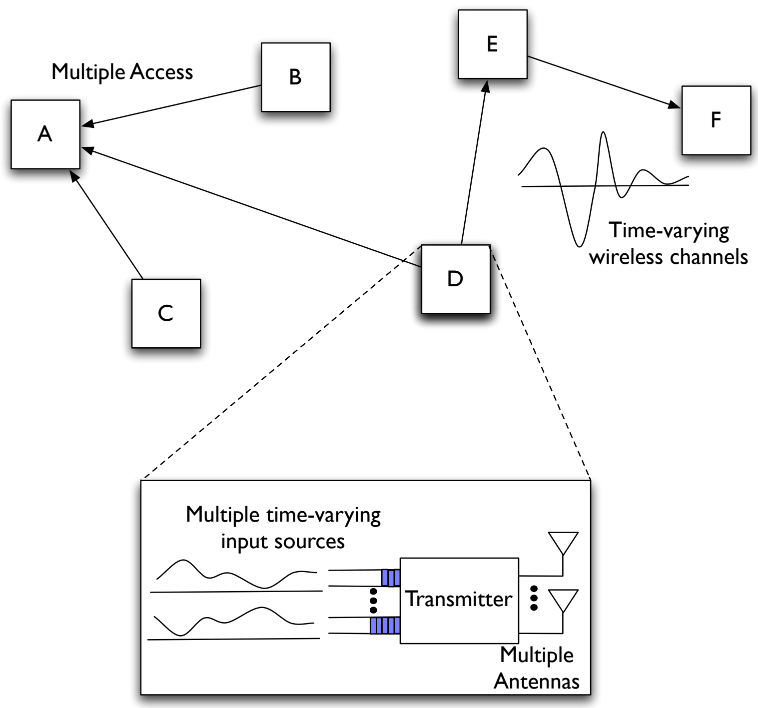

Figure 1.1: A typical multi-hop wireless network with multiple nodes. The figure also shows an expanded view of a typical node which could have multiple sources to be sent using a physical layer transmission over multiple antennas.

tion followed by low-information segments. The same time-varying burstiness is evident in a multi-tude of data sources ranging from web browsing to sensing natural scenery in sensor networks. In fact, the traffic in many networks is bursty at multiple time-scales, evident from abundance of self-similar traffic models studied in wireline [40, 6] and wireless networks [22]. Time-varying data rates of source implies a time-varying demand.

Analogous to time-varying sources is the well-known property of wireless channels, whose quality vary over time. The time-varying nature of the wireless channels can be attributed to mobility of users and their surrounding environment. Electromagnetic signals transmitted by base-station or mobile users reach the intended receiver via several paths; the multiple paths are caused by reflections from man-made and natural objects. Since the length of the each path may be different, the resultant received signal shows a wide fluctuations in its power profile, thereby complicating the design of spectrally efficient systems. If the transmitter and/or receiver is equipped with multiple antennas, then between any two nodes, there are a multiple wireless channels which may vary independent of each other depending on antenna configuration. This time-variability implies that channels have a data capacity which varies with time.

> *Thus, a typical wireless network consists of time-varying demand (sources) and time-varying capacity (channels).*

In this chapter, our goal will be to study techniques which exploit the inherent time-variability in networks to *improve* performance. Grouped together under the broad class of called *opportunistic* techniques, the aim is to conserve resource under poor channel condition (or high source demand) and judiciously use them in favorable conditions.

1.1.1 Degrees of Freedom

Performance of communication over wireless links can be characterized by the number of independent degrees of freedom along which a radio signal can be transmitted. The statistical degrees of freedom, available in spectral, temporal and spatial domain describe the dimensions available to send a wireless signal.

Temporal and Spectral Dimensions: Consider any signal, $X(t)$, which is transmitted over a bandwidth of W Hz for time duration of T seconds. It is well-known from [34] that any such signal has $\sim 2WT$ signaling dimensions. That is, only $2WT$ independent parameters are available at the transmitter to control the type of the signal which can be transmitted over the channel. Note that these dimensions are *not* statistical in form, i.e, they are available for any transmission as long as it is band- and time-limited. After transmission through the wireless channel, the signal undergoes distortions due to multi-path propagation between the transmitter and the receiver. The number of dimensions which determine the communication system performance are the ones observable at the receiver. Thus, if some of the dimensions are lost during the transmission due to poor channel conditions, the effective number of dimensions available for communication are less than $2WT$. Since the channel variations are best characterized statistically due to random fluctuations of the channel gain and phases, the receive degrees of freedom are also best characterized statistically.

Spatial Dimensions: The $2WT$ formula gives the number of spectro-temporal dimensions for single antenna transmitter and receiver pair. In wireless systems, spatial variations can also be exploited by the use of multiple transmit and receive antennas. For rich scattering environments, like those encountered in indoor channels or when then transmitter and receiver are located at low heights (potentially in dense deployments of sensor networks), each additional transmit-receive pair adds a new dimension. Thus, in a multiple antenna system with n transmit and n receive antennas, we have $2nWT$ dimensions for signaling [35, 37].

Spatial degrees of freedom are also available in multiple user systems. Due to differences in topological environment, the propagation environment between any transmitter and receiver is different between any other transmitter-receiver pair. Consider node A in Figure 1.1 which is communicating with three nodes (B, C and D). The three channels, A–B, A–C and A–D vary independent of each other and thus have the same characteristics as a multiple antenna system.

Source "Degrees of Freedom": In general, sources are not characterized in the same framework as the channels. However, there is certain "duality" between sources and channels, which allows a conceptual interpretation of many source opportunism based techniques as duals of channel opportunism based methods. For example, any arrival process $A(t)$can be characterized by its temporal properties and spectrum, since it is simply a time-series like a channel $H(t)$ or the transmitted signal $X(t)$. Similarly, there can be multiple independent sources at any node (as shown in Figure 1.1), which are analogous to multiple independent channels like in multiple antenna links between two nodes.

However, there are some fundamental differences between sources and channels depending on the network usage. While the channel between any two nodes (as long they are reasonably far away from each other) varies independent of the channel between any two other nodes, the sources at each node can be strongly correlated across nodes. For example in Figure 1.1, the data at node D travels to F via node E, which implies that there is a potential for correlation between the arrival process at nodes D, E and F. The only reason these arrival processes may not be perfectly correlated is due to random departures from each node due to channel time-variations. Thus, the random nature of transmissions can "de-correlate" the arrival process across multiple nodes [30].

1.1.2 Roadmap

> The general principle behind all opportunistic methods is to *adapt* transmitter actions based on the *current* conditions of the source and the channel.

For example, source burstiness has been conventionally exploited in communication networks via statistical multiplexing them over the same link using random access methods [1]; statistical multiplexing automatically provides more bandwidth to the user which has high current data need. To understand the basic principles, we will study simple bare-bones formulations in depth and then apply the lessons learnt to real networks. We will follow a rigorous approach to the derivation, all of them inspired by the information-theoretic method. The outline for the rest of the chapter and major milestones are as follows.

Source Opportunism: Focusing on a single point-to-point link, we will first study (Section 1.2) how scheduling packets from a bursty source can be used to reduce the power consumption of any node. The opportunism can be attributed to allowing more delay in the system to avoid sending large bursts of packets in any single transmission slot.

Temporal-spatial opportunism over a single link: Our next focus will multiple antenna links and how temporal and spatial degrees of freedom can be exploited in opportunistic manner (Section 1.3). The key concepts that if the transmitter can receive information (even if a few quantized bits [18]) about the channel conditions, then it can adapt its physical layer parameters like modulation, code rate and power to optimally use the channel conditions. Physical adaptation is commonly used in many current wireless networks (cellular networks like IS-95, 3G) and is proposed in upcoming standards (like IEEE 802.11n).

Temporal-spatial opportunism in ad hoc networks: The adaptive physical layer concepts are next applied to ad hoc networks, to allow opportunistic tuning of transmission rate (Section 1.4). Using the framework of IEEE 802.11 4-way handshake (RTS-CTS-DATA-ACK), we discuss Opportunistic Auto-Rate MAC protocol (OAR) that performs per packet rate adaptation to significantly improve the throughput compared to current systems.

Temporal-spectral-spatial opportunism in ad hoc networks: Finally, we show (Section 1.5) how all three dimensions can be fruit-fully exploited to further increase the capacity of OAR protocol by a multi-channel adaptation labeled Multichannel OAR (MOAR).

1.2 Source Opportunism

1.2.1 Problem Formulation

Consider a time-slotted system, each T_s seconds long, with an input buffer to queue arriving packets. In time slot n, there are a_n arriving packets, each of which is assumed to have same number of information bits. The simplest source model assumes independent and identically distributed (i.i.d.) arrival from one time-slot to another, according to the distribution $p(a_n)$. In more complex and realistic models, the arriving packets a_n can be correlated over time, such as signals generated by measurements of a thermometer. Though correlated models are important for practical systems, the key issue which affects the performance of the system is whether the scheduler at the transmitter has knowledge of the distribution of the arrival process or not. Thus, our problem formulations and the subsequent development in this section will explore how scheduler structure differs with and without the knowledge of source statistics.

Every time slot d_n packets are appropriately modulated as X_n and transmitted over an additive white Gaussian noise wireless channel, given by

$$Y_n = \sqrt{P_n} X_n + \epsilon_n. \tag{1.1}$$

Thus, each transmission packet X_n is unit power and power P_n is the power at which the packet is transmitted in slot n. Further the additive white Gaussian noise component, ϵ_n is assumed to be zero mean and unit variance. With the above setup, the queue evolves according to the following equation,

$$x_{n+1} = \max(x_n + a_n - d_n, 0). \tag{1.2}$$

The transmission $\sqrt{P_n} X_n$ has to be designed such that the receiver can decode the packets *reliably*. We will use an information-theoretic notion of reliability, motivated by [28]. To invoke mutual information based reliability, we require that the time slot duration, T_s, be long enough to encode d_n packets close to channel capacity. More precisely, rate at which d_n packets can be reliably decoded when sent in a fixed length T_s time-slot is given by

$$d_n = \log\left(1 + P_n\right), \tag{1.3}$$

where, for simplicity, we have assumed that the system bandwidth is 1 Hz. Equivalently, the power required to ensure reliable reception is given by

$$P_n = e^{d_n} - 1. \tag{1.4}$$

The relation between the number of scheduled packets d_n and the power required for their transmission shows that a linear increase in number of scheduled packets requires exponential increases in transmit power. The above convex relation between d_n and P_n will also be the reason why power control is important even *without* any channel variations.

Our objective is to minimize the average transmit power consumption of the transmitter such that the packet delay does not exceed an average or a maximum delay bound,

$$P^*(D_0) = \min_{\Theta(D_0)} \lim_{n \to \infty} \mathbb{E}\left\{P_n\right\} \tag{1.5}$$

The set of schedulers $\Theta(D_0)$ consists of all those schedulers which (a) do not drop any packets and (b) ensure that the queueing delay is no more than D_0, which may be measured on average or as the maximum packet delay. We will consider two important classes of schedulers. In the first class, the scheduler completely knows the packet arrival distribution and hence formulates its scheduling decisions based on source statistics. In contrast, the second class of schedulers work without any knowledge of the source statistics and hence aim to be perform only based on current arrivals without utilizing arrival distributions.

1.2.2 Fully Informed Schedulers

In this section, we will study the tradeoff between the average transmit power and the resulting average queuing delay of optimal schedulers. Our approach will be to first (a) characterize the nature of schedulers which achieve the minimum in optimization criterion (1.5), (b) study the nature of optimal schedulers to derive sub-optimal, computationally simple schedulers and lastly (c) derive an intuitive analytical approximation for the relation between average power and average queuing delay of optimal schedulers.

A scheduler is a mapping from the queue state (x_n) to the number of packet departures (d_n). Given the number of scheduled departures, the packet encoding and corresponding power P_n is decided by the physical using the Equation (1.4). We will consider both the set of randomized schedulers, $\mathcal{S} = \{\alpha | \alpha : x_n \mapsto d_n\}$, where a general scheduler mapping is defined by probabilities $\alpha_{j,i} = \mathrm{Prob}(d_n = j | x_n = i)$. Since $\alpha_{j,i}$ are probabilities, $\sum_{j=0}^{i} \alpha_{j,i} = 1$. The set of valid schedulers is thus all those probabilities which satisfy the above properties. A scheduler is called a deterministic if $\alpha_{j,i} \in \{0, 1\}$.

The two key metrics of interest are average queuing delay and average power of the scheduler, which we will denote by $D_{avg}(\alpha)$ and $P_{avg}(\alpha)$. The work in [29] characterizes the subset of schedulers $\Theta \subset \mathcal{S}$ which do not lose any packet (no packet dropped at transmitter and enough power used for receiver to ensure reliable decoding). Furthermore, it is shown that the delay and power of any zero-outage randomized scheduler can be expressed as a convex combination of delay and power of deterministic zero-outage schedulers, as described by the following result.

Theorem 1 (Characterization of delay and power) *Consider a queue with finite buffer size L and an input process with no more than M packet arrivals in one slot. Define $\mathbb{W}$ as the set of deterministic schedulers. Then, for any randomized scheduler $\phi \in \Theta$, there exists*

$\eta_i^{'} \in [0,1], i = 1, 2, \ldots, F' = |\mathbb{W} \cap \Theta|$, with $\sum_{i=1}^{F'} \eta_i^{'} = 1$, such that

$$D_{avg}(\phi) = \sum_{i=1}^{F'} \eta_i^{'} D_{avg}(\gamma_i) \tag{1.6}$$

$$P_{avg}(\phi) = \sum_{i=1}^{F'} \eta_i^{'} P_{avg}(\gamma_i), \tag{1.7}$$

where γ_i is a scheduler and $\gamma_i \in \mathbb{W} \cap \Theta$ for all i. Finally, the boundary of the achievable region in the delay-power plane is piecewise linear with the vertices achieved by deterministic schedulers.

Thus the average packet delay achieved by any zero outage randomized scheduler is given by a convex combination of the average packet delays achieved by all possible zero outage deterministic schedulers. In addition, the same convex combination of the average powers of the zero outage deterministic schedulers gives the average power of zero outage randomized policy.

To compute the optimal scheduler (randomized) for a desired delay bound D_{avg}, dynamic programming based technique commonly known as value iteration algorithm (VIA) [32] can be employed. For large values of buffer sizes, L, the number of states in the VIA increase exponentially and hence computing the optimal scheduler is computationally intensive. For the cases where the arrival distribution is measured in real-time and the optimal scheduler is adapted over time, the implementation of VIA can lead to an intractable design. There are two possible techniques to reduce the computational complexity of scheduler design.

A simple method for complexity reduction is to reduce the number of states in the VIA. The following state reduction is most useful for moderate to large delay scenarios. As delay increases beyond 1 time-slot, the scheduler tends to take the same action in several consecutive queue states. Thus, to reduce the state diagram size, multiple states can be morphed into one state. For instance, for all states $x_n = j, \ldots, j + K$, the scheduler can be constrained to take the same action u_n, thereby reducing the state space. It is clear that the constrained system will be suboptimal, but the loss due to additional constraints can be minimized by appropriate choice of the number and lengths of constraint intervals.

The second suboptimal scheduler is labeled the *log-linear* scheduler, described as follows. For a queuing delay of one time-slot the optimal scheduler flushes the buffer at all time-slots, *i.e.*, $d_n = x_n$ and the corresponding power required in each time-slot is proportional to $e^{d_n(x_n)}$. As the delay increases, we observe that the optimal scheduler tends to choose $d_n(x_n)$ so that the power in each time-slot is linearly proportional to x_n, thereby "equalizing" the power penalty in large buffer states x_n. For equalizing the power, the scheduler picks packets $d_n(x_n) \approx \log(x_n)$. Combined with the natural constraint that we cannot transmit more packets than available ($u_n \leq x_n$), we propose the following log-linear scheduler,

$$u_n = \min\left(x_n, \lfloor \log(\kappa x_n) \rfloor\right). \tag{1.8}$$

The parameter κ of the log-linear scheduler is chosen to meet the delay bound. For buffer states greater than $L-M$, the log-linear scheduler transmits at least $x_n-(L-M)$ packets to prevent buffer overflows. The log-linear scheduler greatly simplifies scheduler design since it requires only one parameter κ. The value of κ to achieve a certain average delay depends on the arrival distribution of the source. The following straightforward adaptive algorithm can be used to compute the value of κ dynamically to achieve any given average delay constraint based only on the knowledge of the mean arrival rate. In the adaptive scheduler, κ_n is a function of time and is updated in every time-slot as given below.

$$\kappa_n = \kappa_{n-1} + \Delta_\kappa \left(\hat{D}_n - D_0\right), \tag{1.9}$$

where $\hat{D}_n$ is the sample average delay given by the ratio of the sample average buffer length and arrival rate. The performance of the adaptive scheduler is shown to numerically converge to the performance of a log-linear scheduler that has full knowledge of arrival statistics [29]. Figure 1.2 compares the performance of the log-linear and the adaptive log-linear scheduler with that of the optimal scheduler, and shows that the suboptimal methods perform very well.

In [29], a approximate relation between average power and average queuing delay is also

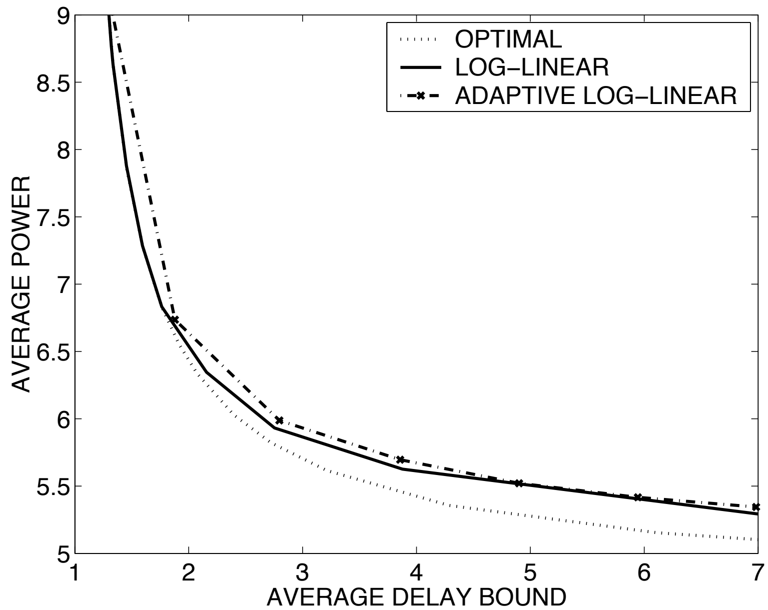

Figure 1.2: Comparison of log-linear and adaptive log-linear schedulers with corresponding optimal scheduler (Buffer length $L = 50$, maximum arrival $M = 6$, uniform arrival distribution).

derived for the optimal scheduler as

$$\log\left(1 + \frac{P_{av}}{\sigma^2}\right) \approx \lambda + \frac{\sigma_a^2}{4\lambda D_{avg}} \tag{1.10}$$

where P_{av} is the average transmit power and D_{avg} is the average queuing delay. Further λ is the average arrival rate and σ_a^2 is the variance of the arrival process $\{a_n\}$. Thus, we can see that as the delay increases, the average transmit power decreases. Furthermore, if the traffic is more bursty, σ_a^2, the gains from larger delay are also larger.

> Additional queuing delay allows the scheduler to smoothen the incoming traffic, thereby removing large swings in power requirements needed if very bursty traffic is transmitted with low delays.

Thus, a scheduler acts as"filter" on the incoming packet arrivals, where the bandwidth of the filter is determined by the ratio of burstiness of the traffic to average delay. In the next section, we will further strengthen this filtering intuition and show that it is in fact provably true that power-minimizing schedulers are in fact always low-pass filters.

1.2.3 Uninformed Robust Schedulers

In this section, we consider the problem of minimizing average transmit power with a constraint on maximum delay which any packet can tolerate. Let $D_{\max}$ denote the maximum delay constraint, then the following result can be proven.

Theorem 2 (Filter Characterization[16]) *A scheduler which guarantees the maximum delay $D_{\max}$ for any arrived packet is a linear time variant filter of size $D_{\max}$ denoted by*

$$u_n = \beta_0^n x_n + \beta_1^n x_{n-1} + \cdots + \beta_{D_{\max}-1}^n x_{t-D_{\max}+1}, \tag{1.11}$$

where the filter coefficients satisfy the following

$$\sum_{i=1}^{D_{\max}-1} \beta_i^{n+i} = 1, \forall n, \tag{1.12}$$

$$0 \leq \beta_i^n \leq 1, \forall n, i. \tag{1.13}$$

Furthermore, any time variant filter of size $D_{\max}$ which satisfies the above constraints is a valid scheduler which guarantees the maximum delay of $D_{\max}$ for any packet.

Theorem 2 turns the design of the guaranteed maximum delay scheduler into the problem of filter design with a linear structure. Therefore, we foresee the vast literature on linear filtering theory as a fundamental tool in designing power-efficient schedulers [9, 26]. Following result follows directly from basic filtering theory.

Corollary 3 *Every feasible time-invariant scheduler is a low-pass filter.*

For power-efficiency, additional delay helps by smoothening the input arrival process via queuing [29]. It is clear that by increasing the number of filter taps (equivalently increasing maximum possible scheduling delay), one can design better low-pass filters, leading to transmit power reduction.

> The intuition behind the fact that the optimal average power scheduler would have smoother output sequence than the input sequence comes from the convexity of the objective function of the optimization problem in (1.5).

We first present upper and lower bounds on the performance of optimal schedulers, which will serve as benchmarks for the performance of the following scheduler designs.

Theorem 4 *The average power of the optimal schedulers can be bounded as*

$$\mathbb{E}^{l+1}\left[2^{\frac{a}{l+D_{\max}}}\right] \leq P_{avg} \leq \mathbb{E}^{D_{\max}}\left[2^{\frac{a}{D_{\max}}}\right] \tag{1.14}$$

where the input arrival rates are i.i.d. random variables with distribution $f_A(a)$ *and* l *is any positive number. The lower bound is tight when* $D_{\max} = 1$ *and the two bounds converge as* $D_{\max} \to \infty$.

The filter design interpretation immediately allows us to design schedulers using filtering principles. The first such concept is that linear time-invariant filters, which leads to the question of designing optimal time-invariant scheduler (recall optimal schedulers are potentially time-varying filters, as shown in Theorem 2). It turns out that the optimal time-invariant scheduler is independent of input arrival distribution and in fact, a simple moving average filter.

Theorem 5 (Optimal time-invariant scheduler [16]) *The optimal time-invariant scheduler has the following form,*

$$u_n = \frac{x_n + x_{n-1} + \cdot + x_{n-D_{\max}+1}}{D_{\max}} \tag{1.15}$$

The performance of the optimal time-invariant scheduler in Theorem 5 is given by the upper bound in Theorem 4. In general, time-vaying schedulers perform much better. The next result derives a time-varying scheduler, adapts the coefficients β_i^n based on the current queue conditions to achieve improved performance.

Theorem 6 *At any given time slot n, let $(S_0^n, S_1^n, \ldots, S_{D_{\max}-1}^n)$ denote the vector of the scheduled output service rates out of the queue from the previous time slot $n-1$, and let a_n be the new arrival which needs to be scheduled. The optimal output service rate for the current time n, d_n and the new vector of the scheduled service rate for time $n+1$, $(S_0^{n+1}, S_1^{n+1}, \ldots, S_{D_{\max}-1}^{n+1})$, are determined by the water-filling solution as*

$$d_n = (\mu - S_0^n)^+ \tag{1.16}$$

$$S_i^{n+1} = (\mu - S_{i+1}^n)^+ \forall 0 \leq i \leq D_{\max} - 2 \tag{1.17}$$

$$S_{D_{\max}-1}^{n+1} = 0 \tag{1.18}$$

$$d_n + \sum_{i=0}^{D_{\max}-1} = x_n \tag{1.19}$$

Therefore the coefficients of the optimal robust time-varying scheduler are given by

$$\beta_i^n = \frac{(\mu - S_{i+1}^n)}{x_n}, \forall 0 \leq i \leq D_{\max} - 1 \tag{1.20}$$

The above discussion provides an extremely simple solution to the problem of robust scheduling for delay constrained inputs to a queue. There are some important observations which come out of above solution. First, to find the optimal robust scheduler in Theorem 6, even though we did not use the filtering property of the scheduler, the optimal solution turns of to be exactly a linear time-variant filter of size $D_{\max}$, which is not surprising due to the Theorem 2. Second, the optimal values of filter coefficient are exactly function of the past $D_{\max} - 2$ values of the input arrivals and the queue backlog, which is expected as we discussed earlier. Third, as we mentioned earlier, the optimal scheduler intuitively should try to make the output service rate as smooth as possible, which corresponds to the low-pass property of the scheduler. Indeed, the water filling solution is the best verification of this intuition. Fourth, the water-filling solution (shown in Figure 1.3) reveal that the optimal

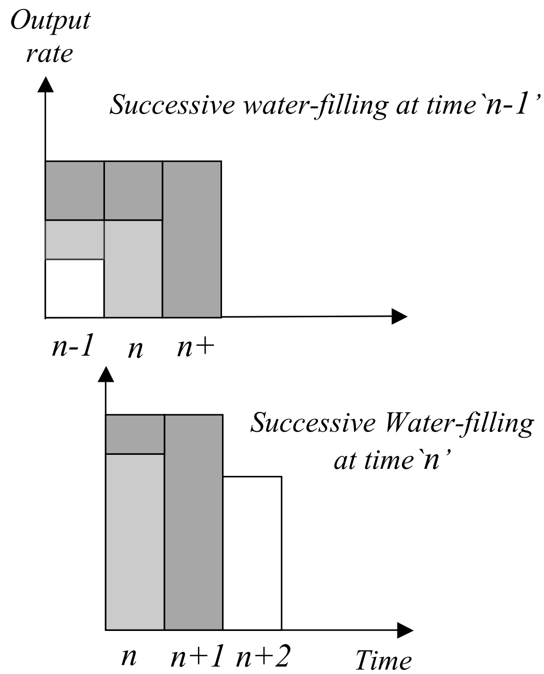

Figure 1.3: Water-filling solution for the time slot $n-1$ and subsequent time slot n.

solution always tries to push the scheduled time of the packets as far as possible and near the maximum value of the tolerable delay. Finally, Figure 1.4 shows the performance of the water-filling scheduler as compared to the upper and lower bounds.

1.3 Spatio-Temporal Opportunism Over a Single Link

Much like source variations can be exploited to improve the performance of the wireless systems, channel time-variations can also be productively used. Time-variations occurs in all degrees of freedom due to physical mobility of users and their surroundings, resulting in channel conditions changing with time, frequency and space. In this section, we will explore the fundamental ideas in exploiting channel variations for improved performance.

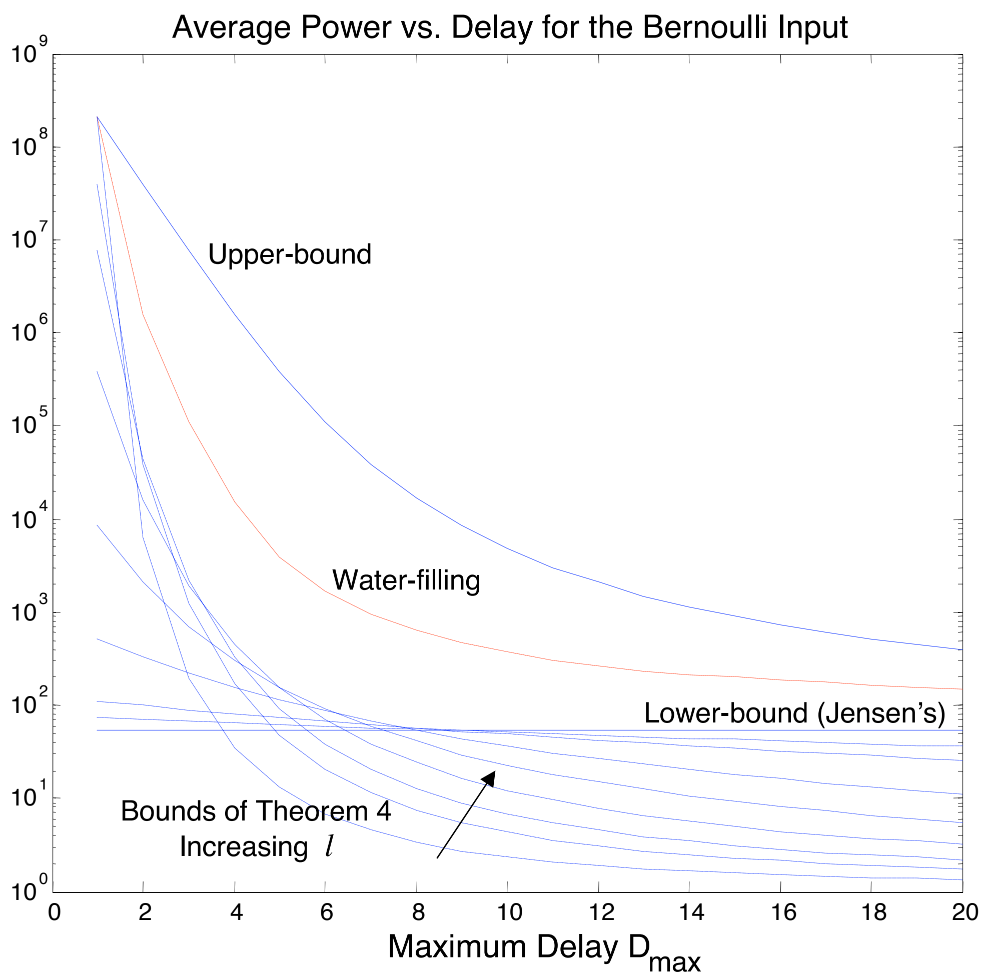

Figure 1.4: Performance of the optimal time-varying robust scheduler and the bounds on performance. Note that performance of the optimal time-invariant robust scheduler matches the upper bound.

1.3.1 Channel Models and Problem Formulation

In a traditional system representation, as depicted in Figure 1.5, a communication channel can be described by its transfer function $H(n)$. If the timescale of channel variation is much

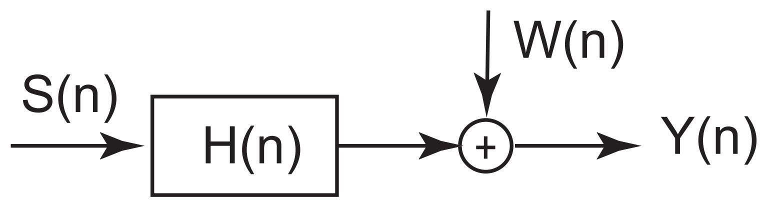

Figure 1.5: System representation of a communication channel.

larger than the duration needed for transmission of a packet, then $H(n)$ can be modeled simply by a complex-valued multiplicative factor known as channel coefficient. Such a channel is referred to as flat fading channel. Wireless systems with high data rate (in order of mega bits per second) and low mobility, such as indoor wireless LAN, wired line, and line-of-sight wireless links are examples of flat fading channel. The input-output relation of such system can be expressed by

$$Y(n) = H(n)S(n) + W(n), \tag{1.21}$$

where $H(n)$ is modeled by a random process. Distribution of $H(n)$ highly depends on the physical characteristic of the transmission path between the transmitter and the receiver and objects surrounding the antennas at both ends. In an environment with large number of scatterer (*e.g.* trees or furniture), $H(n)$ can be well approximated by a zero-mean complex Gaussian random process. Note that if we use an $N \times M$ matrix instead of a scalar for $H(n)$ in (1.21) and modify the input signal ,$S(n)$, and additive noise, $W(n)$ to $M \times 1$ vectors then (1.21) represents a multiple antenna system with M transmit and N receive antennas.

The capacity, or equivalently the maximum achievable data rate with arbitrarily small reliability, is the time average of the maximum achievable data rate at each time. The maximum achievable data rate at each time instance, n, is given by the maximum mutual information expression [35]

$$I(S(n); Y(n)) = \log\det\left(I + H(n)Q(n)H^{\dagger}(n)\right), \tag{1.22}$$

where $Q(n) = \mathbb{E}[S(n)S^{\dagger}(n)]$, and $\dagger$ is the complex conjugate transpose operation. Note that for a single antenna system where $S(n)$ is a scalar, $Q(n)$ is the total power of the input signal at time n. Similarly for a multiple antenna system the diagonal elements of $Q(n)$ are the average power on each antenna. Therefore, the total power spent at time n can be expressed by the sum of the diagonal elements (power on each antenna) of $Q(n)$, or more precisely

$$P(n) = \mathrm{tr}\{Q(n)\},$$

and the average power is given by simply a time average of the power dissipated at each time slot. That is

$$\begin{aligned} P_{av} &= \lim_{T\to\infty} \frac{1}{T} \sum_{n=1}^{T} P(n) \\ &\stackrel{(a)}{=} \mathbb{E}[P(H)], \end{aligned} \tag{1.23}$$

where (a) is the consequence of the law of large numbers, and expectation in (a) is taken over the distribution of H. In general the input signal can be decomposed into a power control factor and a fixed codebook, *i.e.*,

$$S(n) = P(n)x(n),$$

where $P(n)$ is the temporal adaptation scheme which controls the power and phase of the input signal, and the input $x(n)$ is independent of channel condition with a unit average power $\mathbb{E}[xx^{\dagger}] = 1$. Note that in (1.23), $P(n)$ is replaced by $P(H)$, which indicates the power allocated for each realization of channel and does not concern any specific time slot, whereas $P(n)$ indicates the power allocated at time n. Hence, both $P(n)$ and $P(H)$ represent the same quantity from different view.

When transmitter does not have any information about the channel, H, the best unbiased strategy is to divide the power equally among all antennas, and allocate the same amount of power at any given time instance. That is to say for any given time slot n, the allocated power is P_{av}, and the power allocated to each transmit antenna is P_{av}/M (M is the number of transmit antennas). But when channel state information is available at the transmitter, the allocated power can be adjusted such that it matches the channel condition. Therefore, at any time slot, n, the total power allocated will be a function of the channel realization at that time slot, $H(n)$, and the way the power is divided among transmit antennas is determined by power control algorithm. This will be discussed in more detail in next section.

The mutual information expression in (1.22) is the maximum achievable data rate at each time instance and because of $H(n)$ it is a time varying quantity. To find the capacity of the communication link it is enough to find the time average of expression (1.22), *i.e.*,

$$\begin{aligned} C(P_{av}) &= \lim_{T\to\infty} \frac{1}{T} \sum_{n=1}^{T} I(S(n);Y(n)) \\ &= \lim_{T\to\infty} \frac{1}{T} \sum_{n=1}^{T} \log\det\left(I + H(n)Q(n)H^{\dagger}(n)\right) \\ &= \mathbb{E}\left[\log\det\left(I + HQH^{\dagger}\right)\right]. \end{aligned} \tag{1.24}$$

The capacity of a communication system is solution of an optimization problem with a set of constraints. The solution as expected depends on the constraint set. What is described above is the capacity of a communication link with average power constraint. Interested readers can look at [5] for a complete treatment of the subject.

In design of a communication system two common schemes have been adopted by the system engineers. One scheme is based on achieving the highest possible net throughput. That is to get as close as possible to the upper limit defined by (1.24). The performance metric used to evaluate different architectures and methods in this category is referred to as the *ergodic capacity* or average capacity given by expression (1.24).

Second scheme, considers a fixed transmission rate, R. However, the capacity of the channel (expression (1.22) is time varying and it may not be able to support transmission rate R at all times. The duration of time that the instantaneous capacity of the channel is smaller than the transmission rate, reliable communication is not possible, and transmission is said to be in outage. So the design issue is to pick the best rate R such that the net throughput of the system is maximized, or for any arbitrary rate R, find the adaptation schemes that minimizes the outage. We will not discuss this scheme, however, interested readers are referred to [3, 18]. Before delving into opportunistic resource allocation for maximizing throughput, we shall analyze the multiple antenna system in order to extract the important characteristics of channel matrix $H(n)$.

1.3.2 Multiple Antenna Representation

To this point we have not discussed the differences of a multiple antenna system with a single antenna. Consider the system model given in (1.21). Assume a multiple antenna system in which $H(n)$ is an N by M matrix. Applying singular value decomposition to $H(n)$ we can replace $H(n)$ by its SVD equivalent to get

$$\begin{aligned} Y(n) &= U(n)D(n)V^{\dagger}(n)S(n) + W(n) \\ \tilde{Y}(n) &= D(n)V^{\dagger}(n)S(n) + \tilde{W}(n), \end{aligned} \tag{1.25}$$

where U and V are unitary matrices, and D is a diagonal $m \times m$ matrix with $m = \min(N, M)$. The second equality in (1.25) is obtained by multiplying both sides of the first equation with $U^{\dagger}$. This operation is possible when receiver has a perfect knowledge of H, which is the case by the channel estimate unit at the receiver and training signals in the form of preambles in most of the current systems. The diagonal elements of D are denoted by $\lambda_i^{1/2}$, where λ_i's are the eigenvalues of $HH^{\dagger}$ when $N < M$ or $H^{\dagger}H$ when $N \leq M$. Expanding (1.25) we get

$$\begin{aligned} \tilde{y}_1(n) &= \lambda_1^{1/2}(n)\tilde{s}_1(n) + \tilde{w}_1(n) \\ \tilde{y}_1(n) &= \lambda_1^{1/2}(n)\tilde{s}_1(n) + \tilde{w}_1(n) \\ &\vdots \\ \tilde{y}_m(n) &= \lambda_m^{1/2}(n)\tilde{s}_m(n) + \tilde{w}_m(n), \end{aligned}$$

where $\tilde{s}_i(n)$ is the ith element of $V^{\dagger}(n)S(n)$. The multiple antenna system with M transmit and N receive antennas resemble m ($m = \min(M, N)$) independent single antenna channel.

These m *virtual* independent channels add new degrees of freedom (in addition to time and frequency) to the wireless system and are are labeled as *spatial degrees of freedom*.

Note that the m channels vary independent of each other, thus increasing the probability that one or several of them are in good condition (compared to a single-antenna system), thereby creating more opportunities for transmitter adaptation. For example, the m channels can be used independently to increase the net throughput. In this case if the net throughput of a single antenna system is denoted by $C_1(P_{av})$, the net throughput of the multiple antenna system can be at most as high as $mC_1(P_{av})$. Alternately, the m channels can be used to add redundancy and improve the error performance of the system. In the next section we explain in more detail different ways that the spatial degrees of freedom can be used and how its application may affect the net throughput.

1.3.3 Throughput Maximizing Resource Allocation

The problem of optimum power control for single and multiple antenna system is posed rigorously and solved in [10, 14]. We state the problem and its solution briefly and explain the physical interpretation of the result. The optimum signaling that maximizes the throughput of a multiple antenna communication system is the solution to the following optimization problem

$$\min_{(P(n),X(n))} I(S(n),Y(n)|H(n))$$
$$\text{Subject to: } \mathbb{E}[P(H)] \leq P_0.$$

The solution of the above optimization problem is known as matrix-water-filling given by

$$\frac{P_{i,i}(H)}{P_0/m} = \begin{cases} \frac{1}{\gamma_0} - \frac{1}{\bar{\gamma}\lambda_i} & \text{if } \lambda_i > \gamma_0 \\ 0, & \text{otherwise} \end{cases} \tag{1.26}$$

where $\bar{\gamma} = P_0/mN_0$, N_0 is the variance of the noise $W(n)$ and the γ_0 is the solution to power constraint

$$\mathbb{E}[P_{i,i}(H)] = \frac{P_0}{m}, \tag{1.27}$$

that is the average power allocated to any given eigenvalue over the long run should be the same and equal to $1/m$ of the total power P_0. For example a single antenna system the power allocation is depicted in figure 1.6. The transmitter and receiver have perfect information

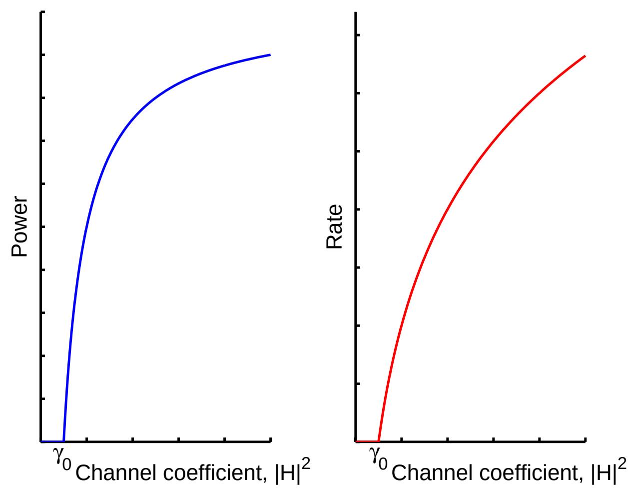

Figure 1.6: Power and rate adaptation based on channel condition for maximizing throughput.

about the channel state H. Transmitter adjust the transmission power and rate according to the power and rate expressions given in (1.26) and (1.22) respectively. The only practical issue of such system is due to the variable transmission rate. It requires a large storage capacity at the transmitter to hold different codebook for adjusting the rate and a complex receiver capable of decoding codes of variable rate.

> The channel dependent power allocation across the m spatial degrees of freedom, $P_{i,i}(H)$ is an example of spatio-temporal opportunistic technique . The power allocation puts more power on those spatial channels which are in good condition and puts less in poor condition. Thus, the optimization is performed in both the dimension of antennas and time.

The information at the transmitter is normally provided by the receiver through a feedback channel. To feedback the channel state information perfectly requires infinite amount of information to be fedback (a real number or matrix is represented with infinite number of bits). A more realistic scenario is to consider only a few amount of bits is available to feedback information to the transmitter. In this case the receiver quantizes the channel state and send back the quantized value using finite number of bits. The transmitter then uses a power and rate which is a quantized version of power and rate as shown in figure 1.6. A more complete treatment of the quantized power and rate allocation can be found in [23].

Spatial opportunism in some multiuser systems (multipoint-to-point or point-to-multipoint, like those in cellular systems) have also been studied [20, 38] and implemented in some practical systems like Qualcomm HDR [13]. The strategy there is similar to the above methods with m independent basestation to mobile user links replacing the m virtual channels in multiple antenna systems. The optimal method turns out to choose a single node which has the highest channel gain, SNR_i to allow sending at the highest possible data rate, $\log(1 + SNR_i)$. The resulting use of multiple independent channels is also labeled *multi-user diversity* to emphasize that the multiple spatial channels are due to multiple nodes in the system. Mutli-user diversity systems are only proportionally fair, implying that their short-term fairness properties can be very poor leading to large delays in the systems.

> Thus, performance improvement in channel condition based adaptation inherently introduce more delay in the transmission , much like packet scheduling to reduce transmission power.

1.4 Spatio-Temporal Opportunism in Ad hoc Networks

In this section we introduce a modification, Opportunistic Auto Rate (OAR), to the existing IEEE 802.11x protocol. The key idea of OAR is to opportunistically exploit high quality channels when they occur via transmission of multiple back-to-back packets. Consequently, nodes transmit more packets under high quality channels than under low quality channels. In order to be able to explain the OAR mechanism, we shall briefly describe the underlying IEEE 802.11 protocol.

1.4.1 Review of IEEE 802.11

As described in [27], IEEE 802.11 media access is based on the RTS/CTS mechanism [2]. In particular, a transmitting node must first sense an idle channel for a time period of Distributed InterFrame Spacing (DIFS) after which it generates a random backoff timer chosen uniformly from the range $[0, w-1]$, where w is referred to as the contention window. At the first transmission attempt, w is set to CW_{min} (minimum contention window). After the backoff timer reaches 0, the node transmits a short request-to-send (RTS) message. If successfully received, the receiving node responds with a clear-to-send (CTS) message. Any other node which hears either the RTS or CTS packet uses the data packet length information to update its Network Allocation Vector (NAV) containing the information of the period for which the channel will remain busy. Thus, all nodes including hidden node can defer transmission suitably to avoid collision. Finally, a binary exponential backoff scheme is used such that after each unsuccessful transmission, the value of w is doubled, up to the maximum value $2^m CW_{\min}$, where m is the number of unsuccessful transmission attempts.

1.4.2 Multi-rate IEEE 802.11

The physical-layers in IEEE 802.11a/b/g/n protocols are designed to have *multi-rate* transmission capabilities. So in good channel condition the transmitter can send data at a rate greater than the base rate (*i.e.*, 2 Mbps in IEEE 802.11b, which is the protocol we will focus in our discussion). Figure 1.7 shows a sample channel variation with time for a mobile speed of 2.5 m/s. The received power shows wide fluctuations such that the supported data rate varies between 5.5 and 2 Mbps with almost equal probability. In practice, depending on the existence and strength of a line-of-sight path and the distance from the transmitter, the rates that channel can support may vary within the entire range of the lowest to highest possible data rate (the highest available rate in IEEE 802.11a is 54 Mbps and 11 Mbps for IEEE 802.11b). Auto Rate Fallback (ARF) [15] was the first commercial implementation that exploited multi-rate capability in IEEE 802.11. In ARF, senders use the error rates of previous transmissions to adaptively select future (attempted) transmission rates. That is, after a number of consecutive successful transmissions, the sender modifies its modulation

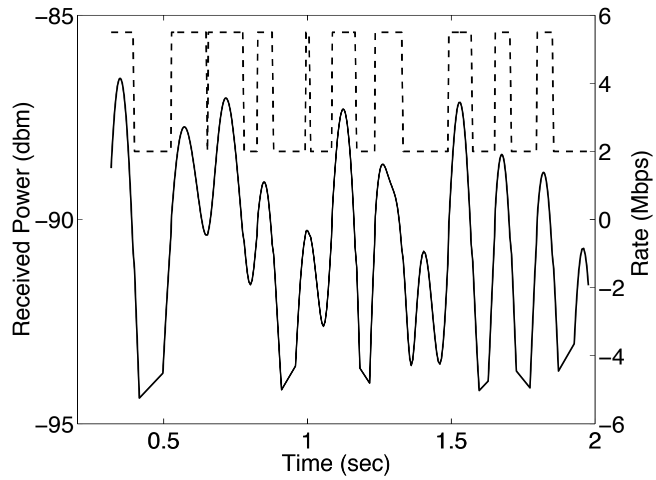

Figure 1.7: Illustration of channel condition variation

scheme to increase transmission rate. Similarly after consecutive losses it reduces the data rate. Consequently, if a mobile user has (for example) a perpetually high-quality channel, the user will eventually transmit at higher data rates. Receiver Based Auto Rate (RBAR) was

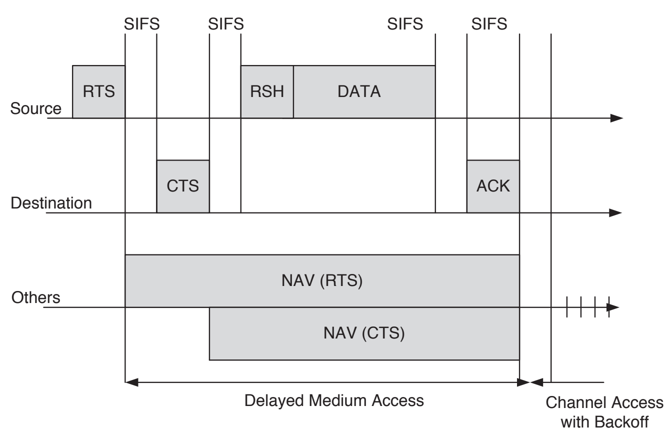

Figure 1.8: IEEE 802.11 with RBAR enhancement

proposed in [12], in which the key idea is for receivers to control the transmission rate. To guarantee that all stations receive the messages (RTS,CTS and ACK) error free, all messages in IEEE 802.11 must be sent at the base rate. Using the received RTS message, receiver determines the maximum possible transmission rate for a given acceptable bit error rate. The receiver sends back the calculated rate to the tranmitter by adding it into a special field of the CTS message. Note that all other nodes overhearing this message are also informed of the modified transmission rate. This message is termed reservation-sub-header (RSH) and is inserted preceding data transfer as illustrated in Figure 1.8. With the RSH message, overhearing nodes can modify their NAV values to the new potentially decreased transmission time. The explicit messaging in RBAR causes a quick adaptation to channel variations and extracts significant throughput gains as compared to ARF.

Although ARF and RBAR increase the data rate, at each access of channel they tend to send a single packet. Since the packet length is fixed, a channel access with lower transmission rate keeps the channel for a longer time in compare with higher data rate packet transmission. The key OAR mechanism is to allow a flow to keep the channel for multiple packet transmission (instead of for a single packet) when channel quality is good and higher data rate is feasible. In OAR, the same amount of *time* is granted to a sender as if the sender is transmitting at the base rate. For example, if the base rate is 2 Mbps and the channel condition is measured such that transmission at 11 Mbps is feasible, the sender is granted a channel access time to send $\lfloor 11/2 \rfloor = 5$ packets.

Note that all three extensions of IEEE 802.11 protocol are examples of channel adaptation developed in Section 1.3.3. With the difference that optimum power allocation is not included and transmitted power is kept constant and only adaptive rate allocation is considered. Next section gives insight on how OAR can be implemented and added to IEEE 802.11 [31].

1.4.3 OAR Protocol

In order to allow the sender to hold the channel for multiple packet transmission, the *fragmentation mechanism* of IEEE 802.11 can be exploited. The fragmentation mechanism defined as part of the IEEE 802.11 standard and mandated for implementation provides a simple and practical way for nodes to hold the channel for multiple packets when high data rates are measured. In particular, as in RBAR, the receiver indicates the available physical-layer rate via the RTS/CTS messages. If the data rate is above the base rate, the *more fragments* flag in *frame control* field of MAC header is set by the sender until $\lfloor$transmission_rate / base_rate$\rfloor$ packets are transmitted. The duration field carried by the data and the subsequent ACK is also updated to indicate that the medium will be busy until the end of next data packet.

Thus, as described above, each data packet and ACK serve as a virtual RTS/CTS so that no additional RTS/CTS frames need to be generated after the initial RTS/CTS handshake. The sender must also set the *fragment number* subfield in the *sequence control* field of the MAC header to be 0. This prevents the receiver from treating the data packet as a part of an actual fragmented packet.

To illustrate the protocol time-line consider the following example in which node 1 has a good channel (11 Mbps) and node 2 has a poor channel (5.5 Mbps). As depicted in Figure 1.9, in one case node 1 uses the OAR modification and in another case uses RBAR. Using OAR, node 1 retains channel access and sends four more packets without any channel contention. Whereas, RBAR goes into contention immediately after the transmission of node 1 is completed. The backoff counter for node 2 was frozen while node 1 was transmitting, and node 1 picks a new backoff counter after finishing its transmission. Thus, with high probability, node 2 gets access to the channel and sends its data at 5.5 Mbps. Thus, RBAR loses in throughput due to extra contention after every packet in addition to not capitalizing fully when good channels are encountered.

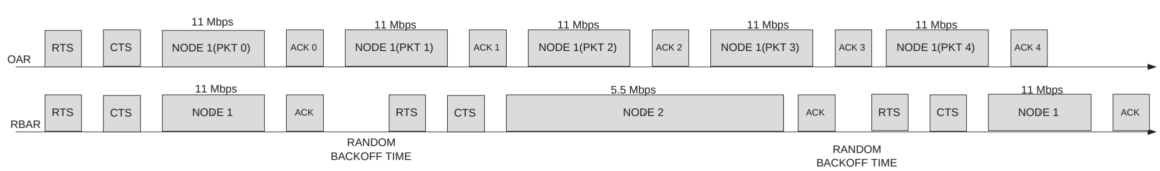

Figure 1.9: Illustration of OAR and RBAR Time-lines for a two node system. Node 1 is in a better channel state than node 2 (11 versus 5.5 Mbps).

Figure 1.10 shows the throughput gain of OAR over RBAR for different number of flows. Observe that in all cases, OAR results in significant throughput gains as compared to RBAR in the range of 42% to 56%. OAR extracts this gain by holding the channel when it is good to the longest extent possible subject to maintaining the base-rate time shares. Moreover, observe that as the number of flows increases, the throughput gain of OAR as compared to RBAR also increases. This increase is due to two factors. First, with a higher number of contending flows, the fraction of time that a flow with a good channel is accessing the medium is higher, resulting in increasing gains. That is, since flows send a single packet at the base rate, but up to 5 consecutive packets at higher rates, additional users provide more

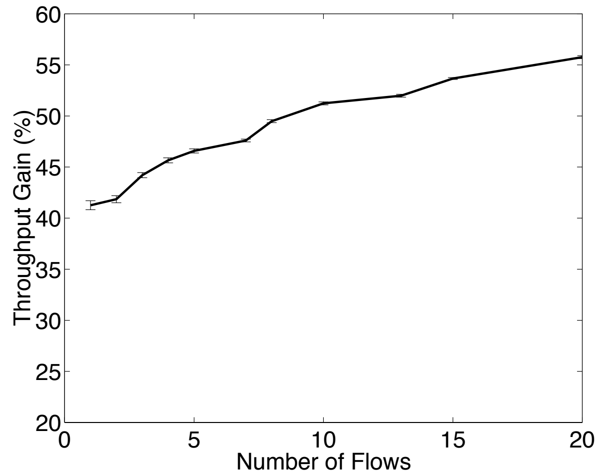

Figure 1.10: OAR throughput gain as a function of the number of flows.

opportunities for a flow to be in a good-channel state thereby extracting further gains from opportunistic scheduling. Second, OAR reduces the contention overhead since nodes with a good channel exploit the channel for transmission of additional packets. Since IEEE 802.11 as well as RBAR have increasing contention times for an increasing number of nodes, OAR extracts increased throughput gains with a larger number of nodes by decreasing contention time.

Finally, note that for both RBAR and OAR, the throughput gains as compared to base-rate IEEE 802.11 are significant. For example, for 10 flows, RBAR obtains a gain of 230% throughput as compared to IEEE 802.11, and OAR obtains an additional throughput gain of 51% as compared to RBAR, or 398% above IEEE 802.11.

1.5 Spatio-Temporal-Spectral Opportunism in Ad hoc Networks

In many wireless systems, the available spectrum is more than what is used, thus the system has multiple channels (frequency bands) available for transmission. If the channel condition of all the frequency bands are known at the transmitter, then sender can choose the channel with best quality in order to maximize its throughput. However, knowledge of channel qualities is not free and there is a cost (requires time and power) associated to channel measurement.

The key idea of Multi-band Opportunistic Auto Rate (MOAR) is to allow nodes to find the band with the best channel quality. The search should be done in an optimal way such that the net throughput is maximized while the measurement overhead is kept as low as possible.

1.5.1 MOAR Protocol

In this section we describe how MOAR employs a band skipping technique within the IEEE 802.11 framework.[2] All nodes initially reside on a single common frequency band, known as the *home band*. DATA transmission is preceded by the sender transmitting an RTS packet to the receiver on the home band. On reception of the RTS frame, the receiver makes the decision to skip by comparing the measured SNR to a channel skip threshold.[3] If the measured SNR is low, the sender and receiver skip to a new channel in search of a better quality channel, whereas if the measured SNR is high, data is transferred on the current frequency channel as in the OAR protocol, in which nodes transfer multiple back-to-back packets in proportion to their channel quality.

On making the decision to skip, the receiver selects a band to skip to and piggy-backs this channel on the CTS packet. After transmitting the CTS frame, the receiver immediately skips to the new frequency channel and waits for another RTS from the receiver for a time equal to the CTS timeout value as mandated by the IEEE 802.11 standard. Since we assume that in a realistic setting channel conditions on other frequency channels are unknown, the band to which the receiver decides to skip is selected *randomly* among the available frequency bands. Yet, if information regarding channel conditions or interference on some other band is known (e.g, in a wireless LAN scenario where the Access Point (AP) may have information regarding interference on other bands), the receiver can take that into account to make a better decision about which band to skip to. However, for the purpose of this discussion we

[2] Although our discussion of MOAR is within the context of the RTS/CTS mechanism within the DCF mode of IEEE 802.11, the concepts are equally applicable to other RTS/CTS based protocols such as SRMA [36] and FAMA [8].

[3] A reasonably accurate estimate of the received SNR can be made from physical-layer analysis of the PHY layer preamble to each packet.

do not require the existence of such information.

If after skipping to a new band, the receiver does not receive another RTS from the sender within a CTS timeout period, the receiver node switches back to the home band and starts contending for channel access as mandated by the IEEE 802.11 standard.

Once the sender receives confirmation of the choice of band to skip to from the receiver (via a CTS frame), it immediately skips to that band. Note that the time elapsed for switching channels is $\sim 1\mu s$ [7] and has negligible overhead. After skipping to the selected channel, the transmitter and receiver renegotiate the data rate via another RTS/CTS exchange which also serves the dual purpose of measuring the channel. In case the channel quality on the new band is measured to be below the skip threshold, the sender-receiver pair can choose to skip again in search of a better quality channel.

Since RTS/CTS exchange prior to any channel skip is done at the base rate on the home band, all nodes within radio range of the receiver and the transmitter can also decode these packets. However, some nodes (including nodes within radio range of the sender but outside the radio range of the receiver) may be unable to hear the CTS packet and are unable to detect whether a decision to skip bands was made or not. Moreover, even though nodes within radio range of the receiver can correctly decode a CTS packet and infer that a decision to skip has been made, they are unable to set a correct defer time since it is not known *a priori* how many times the sender-receiver pair may skip in search of a better quality channel. This can lead to problems similar to the hidden terminal problem [2].

To solve the problem mentioned above, all MOAR nodes upon reception of an RTS/CTS packet defer (via the Network Allocation vector, NAV) for a fixed amount of time corresponding to a maximum time, D_{skip}, necessary for the transmitter and receiver to skip (multiple times, if required) to a better quality channel and finish the DATA/ACK transmission. D_{skip} is given by

$$D_{skip} = K \cdot T_{overhead} + T_D, \tag{1.28}$$

where, K is the maximum number of allowed channel skips, $T_{overhead}$ is the time taken for RTS/CTS exchange at base rate, including all the defer timers (EIFS, SIFS, DIFS etc) as mandated by the IEEE 802.11 standard. The time period T_D represents the time to send data packet at the base rate.

We refer to D_{skip} as a *temporary reservation*, to denote the fact that the reservation is not an actual reservation but represents a maximal amount of reservation time. A temporary reservation serves to inform the neighboring nodes that a reservation has been requested but the duration of the reservation is not known. Any node that receives the temporary reservation is required to treat it the same as an actual reservation with regard to later transmission requests; that is if a node overhears a temporary reservation it must update its NAV so that any later requests it receives that would conflict with the temporary reservation must be denied. Thus the temporary reservation serves as a placeholder until either a new reservation is received or is canceled. If the sender-receiver pair decide not to skip bands then they can proceed with the DATA/ACK exchange on the home band as dictated by OAR in which case other nodes can replace the temporary reservation with the exact reservation, as carried in the DATA/ACK packets.

Once the transmitter and the receiver conclude the DATA/ACK transmission by skipping to one or more bands, they return to the home band. The ACK for the final packet (recall in OAR, nodes can send multiple back-to-back packets) is sent in the home band at base rate by the sender, and the sender rebroadcasts the same ACK. The two-way ACK in the home-channel is necessary to ensure that all the nodes within the range of either the sender and/or the receiver can correctly infer the end of channel skipping and cancel the temporary reservation timer. In case a node is unable to hear either the updated reservation or the DATA/ACK transmission signalling the end of the temporary reservation, it would be able to contend for the channel again after the temporary reservation has expired.

1.5.2 Performance Study of MOAR

Here we consider random topologies representative of a wireless LAN and consider a scenario where the link distance is uniformly distributed in a circular area with diameter 250 m. We fix the Ricean fading parameter to K=4 and also set the size of the estimation window to

60 packets, as discussed in the previous section. Figure 1.11 shows the average throughput gain of MOAR over OAR (computed on per-flow basis and then averaged over flows) as well as the 95% confidence interval values of the percentage gain for each number of flows. The curve labeled "Look-ahead" represents the genie-aided protocol in which channel state information for all the 11 bands is known *a priori* and thus flows need to skip at a maximum of one time to the band with known higher rate than the present band. This serves as an upper bound to the gain that MOAR can extract over OAR. We also implement the optimal skipping rule and plot the throughput gains of MOAR with optimal skipping over OAR. The opportunistic gain that MOAR can extract is dependent upon the distance between

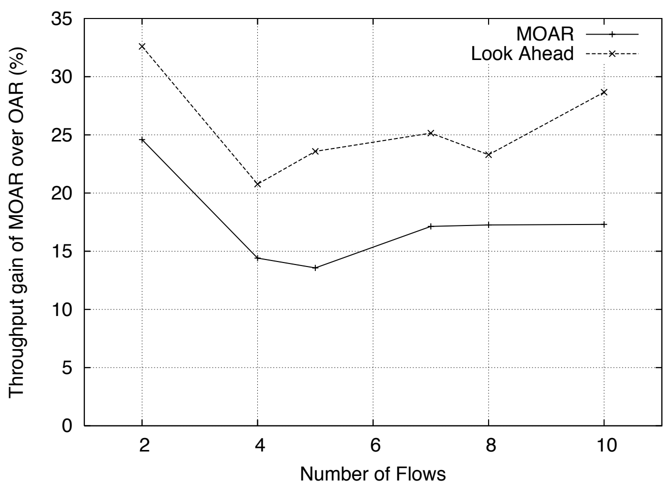

Figure 1.11: Throughput gain of MOAR for random fully connected topologies

the sender and receiver of a flow. For a given random topology, some of the flows are located in a region where the opportunistic gain obtained by skipping bands is not significant. These nodes, besides contributing little to the net overall gain that MOAR can obtain, actually reduce the opportunistic gain for better located nodes. The reason for this can be attributed to the random nature of the MAC. Whenever nodes with lower opportunistic gain access the medium, nodes which are better located to exploit the opportunistic gain through band skipping defer medium access. Thus, the net opportunistic gain that can be obtained by exploiting channel diversity is reduced. However, on average MOAR still outperforms OAR by 14-24%. Also note that the gain of MOAR with optimal skipping is very close to the maximum gain achievable if the channel condition on all the 11 bands is known *a priori*. Thus, in realistic systems where channel state information on other channels may be unavailable, the optimal skipping rule can still enable MOAR to capture most of the performance gains available via opportunistic skipping.

1.6 Conclusions

In this chapter, we discussed the general ideas of *adaptation* to exploit the inherent time-variability of sources and channels in wireless networks. The significant gain in all cases makes opportunistic algorithms as default candidates for any new design of high-performance network elements.

It is necessary to say that the adaptive algorithms are not limited to what is mentioned in this chapter. For example in [39, 19] an adaptive beamforming scheme is investigated. Beamforming algorithms, matches the phase of the input signal to the phase of the channel such that the received signals (in a multiple antenna system) are combined constructively. Opportunism can be exploited in scheduling in many ways. In this chapter we talked about source scheduling of a single queue. Opportunistic scheduler can be extended to a multiuser scheduler [24, 21], in addition to the best user scheduling adding new spatial degrees of freedom even in single-antenna systems, often referred to as *multiuser diversity* [11, 41].

In a multiple antenna system another way of exploiting the spatial degrees of freedom is by antenna selection in which the transmitter only uses one or a subset of possible transmit antennas [25, 33]. On higher networks layers, opportunistic routing can be done in which the best route or a subset of routes among all possible routes is chosen [4, 17]. We believe that there are numerous many other ways that one can uses the information about source and channel, and adapt the transmission scheme in a beneficial way.

References

[1] Dimitri P. Bertsekas and Robert Gallager. *Data Networks*. Prentice Hall, 1992.

[2] V. Bharghavan, S. Demers, S. Shenker, and L. Zhang. MACAW: A media access protocol for wireless LANs. In *Proceedings of ACM SIGCOMM '94*, London, UK, 1994.

[3] E. Biglieri, G. Caire, and G. Taricco. Limiting performance of block-fading channels with multiple antennas. *IEEE Trans. Inform. Theory*, 47(4):1273–1289, May 2001.

[4] C. Cetinkaya and E. Knightly. Opportunistic traffic scheduling over multiple network paths. *IEEE INFOCOM'04*, 3:1928–1937, 2004.

[5] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons Inc., 1991.

[6] M. E. Crovella and A. Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. *IEEE/ACM Transacations on Networking*, 5(6):835–846, Dec. 1997.

[7] R. Garces and J. J. Garcia-Luna-Aceves. Collision avoidance and resolution multiple access for multichannel wireless networks. In *Proceedings of IEEE INFOCOM 2000*, pages 595–602, Tel Aviv, Israel, 2000.

[8] J. J. Garcia-Luna-Aceves and C. Fullmer. Floor acquisition multiple access (FAMA) in single-channel wireless networks. *ACM Mobile Networks and Application (MONET), Special issue on Ad Hoc Networks*, 4:157–174, 1999.

[9] J. C. Geromel, M. C. De Oliveira, and J. Bernussou. Robust filtering of discrete-time linear systems with parameter dependent lyapunov functions. *SIAM Journal on Control and Optimization*, 41(3):700–711, 2002.

[10] A. J. Goldsmith and P. P. Varaiya. Capacity of fading channels with channel side information. *IEEE Trans. Info. Th.*, 43(6):1986–92, Nov. 1997.

[11] M. Grossglauser and D. N. C. Tse. Mobility increases the capacity of adhoc wireless networks. *IEEE Transactions on Networking*, 10(4):477–486, Aug. 2002.

[12] G. Holland, N. Vaidya, and P. Bahl. A rate-adaptive MAC protocol for multi-hop wireless networks. In *Proceedings of ACM MOBICOM'01*, Rome, Italy, 2001.

[13] TIA/EIA IS-856. "CDMA 2000: High rate packet data air interface specification". *Std*, November 2000.

[14] S. K. Jayaweera and H. V. Poor. Capacity of multiple-antenna system with both receiver and transmitter channel state information. *IEEE Trans. Info. Theory*, 49(10):2697–2709, Oct. 2003.

[15] A. Kamerman and L. Monteban. WaveLAN II: A high-performance wireless LAN for the unlicensed band. *Bell Labs Technical Journal*, pages 118–133, Summer 1997.

[16] M. A. Khojastepour and A. Sabharwal. Delay-constrained scheduling: Power efficiency, filter design and bounds. *IEEE INFOCOM*, 2004.

[17] A. Khoshnevis and A. Sabharwal. Network channel estimation in cooperative wireless networks. *Canadian Workshop on Information Theory*, May 2003.

[18] A. Khoshnevis and A. Sabharwal. On the asymptoti performance of multiple antenna channels with fast channel feedback. *submitted to IEEE Trans. on Inform. Theory*, 2005.

[19] I-M. Kim, S-C Hong, S. S Ghassemzadeh, and V. Tarokh. Opportunistic beamforming based on multiple weighting vectors. *IEEE Transaction on Wireless Communications*, 4(6):2683–2687, Nov. 2005.

[20] R. Knopp and P. A. Humblet. Information capacity and power control in single-cell multiuser communications. In *Proc. ICC*, pages 331–335, June 1995.

[21] C. Li and X. Wang. Adaptive opportunistic fair scheduling over multiuser spatial channels. *IEEE Transaction on Cp,,imocatopms*, 53(10):1708–1717, Oct. 2005.

[22] Q. Liang. Ad hoc wireless network traffic: Self-similarty and forecasting. *IEEE Communications Letters*, 6(7):297–299, July 2002.

[23] L. Lin, R. D. Yates, and P. Spasojevic. Adaptive transmission with discrete code rates

and power levels. *IEEE Trans. on Communications*, 51(12):2115–2125, Dec. 2003.

[24] X. Liu, E. K. P. Chong, and N. B. shroff. Opportunistic transmission scheduling with resource-sharing constraints in wireless networks. *IEEE Journal on Selected Areas in Communicaitons*, 19(10):2053–2064, Oct. 2001.

[25] Y. Liu and E. Knightly. Opportunistic fair scheduling over multiple wireless channels. *IEEE INFOCOM'03*, 2:1106–1115, 30 March-3 April 2003.

[26] S. O. R. Moheimani, A. V. Savkin, and I. R. Petersen. Robust filtering , prediction, smoothing and observability of uncertain systems. *IEEE Transactions on Circuits and Systems*, 45(4):446–457, 1998.

[27] B. O'Hara and A. Petrick. *IEEE 802.11 Handbook, A Designer's Companion*. IEEE Press, 1999.

[28] L. H. Ozarow, S. Shamai, and A. D. Wyner. Information theoretic considerations for cellular mobile radio. *IEEE Trans. Info. Th.*, 43(2):359–378, May 1994.

[29] D. Rajan, A. Sabharwal, and B. Aazhang. Delay-bounded packet scheduling of bursty traffic over wireless channels. *IEEE Trans. Info. Th.*, 50(1):125–144, January 2004.

[30] A. Sabharwal, A. Khoshnevis, and E. W. Knightly. Opportunistic spectral usage: Bounds and a multip-band csma/ca protocol. *Submitted to IEEE/ACM Transacations on Networking*, 2005.

[31] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly. OAR: An opportunistic auto-rate media access protocol for ad hoc networks. *To appear in ACM/Baltzer Mobile Networks and Applications (MONET): Special Issue on Selected Papers from MobiCom 2002*, 2003.

[32] Linn I. Sennott. *Stochastic Dynamic Programming and the control of Queueing Systems*. John Wiley and Sons, Inc., 1999.

[33] N. Sharma and L. H. Ozarow. A study of opportunism for multiple-antenna systems. *IEEE Transaction on Information Theory*, 51(5):1804–1814, May 2005.

[34] D. Slepian. On bandwidth. *Proc. IEEE*, 64(3):292–300, March 1976.

[35] I. E. Telatar. Capacity of multi-antenna Gaussian channels. Technical report, AT&T Bell Labs, 1995. (Appeared in *Eur. Trans. Tel.*, vol. 10, no. 6, pp. 585–595, 1999).

[36] F. A. Tobagi and L. Kleinrock. Packet switching in radio channels: Part III - polling and (dynamic) split-channel reservation multiple access. *IEEE Transactions on Communications*, 24(8), 1976.

[37] D. Tse and P. Viswanath. *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.

[38] P. Vishwanath, D. Tse, and R. Laroia. Opportunistic beamforming using dumb antennas. *IEEE Transactions on Information Theory*, 48(6):1277–1294, June 2002.

[39] P. Viswanath, D. N. C. Tse, and R. Laroia. Opportunistic beamforming using dumb antennas. *IEEE Transaction on Information Theory*, 48(6):1277–1294, June 2002.

[40] W. Willinger, M. S. Taqqu, r. Sherman, and D. V. Wilson. Self-similarity through high-variability: statistical analysis of ethernet LAN traffic at the source level. *IEEE/ACM Transacations on Networking*, 5(1):71–86, Feb. 1997.

[41] D. Wu and R. Negi. Utilizing multiuser diversity for efficient support of quality of service over a fading channel. *IEEE Transaction on Vehicular Technology*, 55(3):1298–1206, May 2005.