

Exploiting Multiuser Diversity for Medium Access Control in Wireless Networks

Xiangping Qin and Randall Berry

Dept. of Electrical and Computer Engineering

Northwestern University

2145 Sheridan Rd., Evanston IL 60208

Email: sandra@ece.nwu.edu, rberry@ece.nwu.edu

Abstract—Multiuser diversity refers to a type of diversity present across different users in a fading environment. This diversity can be exploited by scheduling transmissions so that users transmit when their channel conditions are favorable. Using such an approach leads to a system capacity that increases with the number of users. However, such scheduling requires centralized control. In this paper, we consider a decentralized medium access control (MAC) protocol, where each user only has knowledge of its own channel gain. We consider a variation of the ALOHA protocol, *channel-aware ALOHA*; using this protocol we show that users can still exploit multi-user diversity gains. First we consider a backlogged model, where each user always has packets to send. In this case we show that the total system throughput increases at the same rate as in a system with a centralized scheduler. Asymptotically, the fraction of throughput lost due to the random access protocol is shown to be $1/e$. We also consider a splitting algorithm, where the splitting sequence depends on the users' channel gains; this algorithm is shown to approach the throughput of an optimal centralized scheme. Next we consider a system with an infinite user population and random arrivals. In this case, it is proved that a variation of channel-aware ALOHA is stable for any total arrival rate in a memoryless channel, given that users can estimate the backlog. Extensions for channels with memory are also discussed.

I. INTRODUCTION

A fundamental trait of wireless channels is that they exhibit fading effects, due in part to mobility and other user interference. As a result of this time-variation, a user's channel suffers periods of severe decay, but also periods when the channel gain is stronger than average. When many users are present, different users will experience peaks in their channel quality at different times. This effect has been called multi-user diversity [7]. It can be exploited by scheduling transmissions when a user has favorable channel conditions. The more users that are present, the more likely it is that one user has a very good channel at any time; hence, the total throughput of such a system tends to increase with the number of users. Multiuser diversity has its roots in the work of Knopp and Humblet [7], where they presented a power control scheme for maximizing the information theoretic capacity of the uplink of a single cell with time-varying channels. Given the channel gain of each user, it is shown that capacity is maximized by allowing only the user with the best channel to transmit at any

time. Multiuser diversity underlies much of the recent work on "opportunistic" downlink scheduling [10], as in Qualcomm's High Data Rate (HDR) system (1xEV-DO) [1]. It has also been studied in the context of ad-hoc networks [5] and in multi-antenna systems [17].

In this paper, we consider exploiting multiuser diversity gains in a distributed way. As in [7], we consider an uplink model where a group of users are all communicating to a single receiver, such as an access point in a wireless LAN or a base station in a cellular setting. With a centralized approach, the scheduler must know each user's fading level; this could be gained, for example, by having the users estimate their channel gain and then transmit this information to the scheduler. In a large network with many users this type of approach will not scale well and the delay in conveying this information to the scheduler will limit performance. Instead of such an approach, we focus on the case where no centralized controller is available to schedule user transmissions, and we assume each user only has knowledge of its own fading level, but no knowledge of the fading levels of the other users in the cell. The estimation of a user's fading level may be based on a periodic pilot signal broadcast by the base-station.¹ Without a central controller, an approach as in [7] is precluded. This is similar in some ways to distributed power control problems, such as those studied in [14]. For this case, we show that multiuser diversity gains (i.e., a capacity that increases with the number of users) can still be achieved when the users access the uplink using a simple variation of the slotted ALOHA random access protocol [2], which we call channel-aware ALOHA. With this protocol, users randomly transmit with a transmission probability that is based on their channel gain.

The approach in this paper jointly addresses both physical layer and medium access control issues for a wireless network. Such cross-layer approaches in wireless networks have received much attention recently. Other examples that address both MAC and physical layer issues in wireless channels includes [19] and [16]; the focus of these papers is on various techniques for exploiting multi-user reception. Other work in this area includes work on power capture, such as [11], [8], and approaches that use channel coding to recover from collisions

This research was supported in part by the Motorola-Northwestern Center for Telecommunications.

¹This pilot signal should be broadcast in the same coherence bandwidth as the uplink channel, as, for example, in a time-division duplex system.

[15],[12].

We initially consider a backlogged system with n users, where each user always has a packet to send. In this setting, the throughput of the distributed MAC scheme is shown to increase with the number of users at the same rate as the optimal centralized scheme. Asymptotically, the ratio of the throughput of the channel-aware ALOHA to the throughput of a centralized scheduler is shown to be $1/e$, the same as the well-known ratio achieved by a standard slotted ALOHA system in an unfaded channel. This can be interpreted as saying that the only loss due to distributed channel knowledge is the loss due to contention for the channel. For a finite number of users, it is shown that the loss in throughput due to contention when fading is present is less than the loss in a channel without fading. In other words, lack of centralized control is less harmful in a fading environment. We also introduce a splitting algorithm; when the channel changes at a slow enough rate, this is shown by simulation to approach the optimal throughput. Next we consider a variation of the ALOHA protocol for random arrivals. For an infinite user, Rayleigh fading model, it is shown that the channel-aware ALOHA is stable for any total arrival rate. This stability is achieved by leveraging the increasing multiuser diversity as the number of backlogged users increases. Finally, we consider a variation of the model for a channel with memory, in this case the total throughput is shown to be the same as in the memoryless case, but the average delay increases.

The remainder of the paper is organized as follows: first we describe the model of the channel-aware ALOHA system and analyze this model for the case of a memoryless channel. In Sect. III, we compare this distributed MAC scheme with a centralized scheme and introduce a splitting algorithm. In Sect. IV, the random arrival case is addressed. Finally an extension of these results to a simple model of a channel with memory is discussed in Sect. V.

II. MODEL AND ANALYSIS - MEMORYLESS CHANNEL

A. Basic Model:

We consider a model of the uplink in a wireless network with n users all transmitting to a common receiver. The channel between each user and the receiver is modeled as a time-slotted, block-fading channel; if only the i th user transmits in a given time-slot, the received signal, $y_i(t)$ is given by

$$y_i(t) = \sqrt{H_i}x_i(t) + z(t),$$

where $x_i(t)$ is the transmitted signal, H_i is the time-varying channel gain, and $z(t)$ is additive white Gaussian noise. If user i uses transmission power P_i during this time-slot, the received power level for the user is given by $P_r = H_i P_i$. The channel gain is assumed to be fixed during each time slot and to randomly vary between time-slots. In this section, we assume that the channel gains of each user in each time-slot are i.i.d. random variables with probability density $f_H(h)$. For example, to model a Rayleigh fading channel, we have $f_H(h) = e^{-\frac{h}{h_0}}/h_0$, where h_0 is the average fading level. We

assume that at the start of each time-slot, each user knows their own channel gain during the slot, but not the gain of any other users. However, the distribution of each user's channel gain is known.

Given this distributed channel knowledge, we consider a variation of a slotted ALOHA protocol, where each user bases their transmission probability on their channel gain. As in a standard ALOHA model, we assume that if two or more users transmit packets in the same slot, a collision occurs and no data gets through.² After each slot, the users receive instantaneous $(0, 1, e)$ feedback indicating whether a slot was idle, contained a successful transmission or contained a collision.³ Initially, we focus on a model of a saturated system where all n users always have packets to send. In Sect. IV, random arrivals are considered.

In a standard ALOHA system, each backlogged user independently sends a packet in every slot with probability p . In this case, we assume a user only transmits when its channel gain is above a threshold H_0 . The threshold, H_0 , can be chosen to achieve a desired transmission probability. Let $F_H(h) = \int_h^\infty f_H(h) dh$ denote the channel gain's complimentary distribution function, and assume this is strictly decreasing. For a transmission probability p , the desired threshold is $H_0 = F_H^{-1}(p)$.

Let $R(P_r)$ be a function which denotes the rate at which a user can reliably transmit as a function of the received power. For example, if a user can transmit at rates approaching the Shannon capacity of the channel in each slot, then we have $R(P_r) = W \log(1 + \frac{P_r}{N_o W})$, where W is the bandwidth of channel. Initially, assume in each slot that when users transmit, they do so with a fixed rate; this will require a constant received power of P_r . When users transmit, they will simply invert the channel and transmit with power $P_t = \frac{P_r}{h}$.⁴ Suppose each user has a long-term average power constraint of \bar{P} , i.e.

$$\lim_{N \rightarrow \infty} \frac{1}{n} \sum_{n=1}^N P_n \leq \bar{P},$$

where P_n is the power used in the n th slot. For a given transmission probability p , this is equivalent to

$$\int_{F_H^{-1}(p)}^\infty f_H(h) \frac{P_r}{h} dh \leq \bar{P}. \quad (1)$$

To satisfy this constraint, we must have

$$P_r \leq \frac{\bar{P}}{\int_{F_H^{-1}(p)}^\infty f_H(h) \frac{1}{h} dh}.$$

²In particular, we do not consider any capture effects or multi-user reception; our main reason is to simplify the following discussion; however, many of these ideas could be extended to such settings.

³In most cases, additional feedback will be available and more elaborate MAC protocols can be used. Again, our reason for focusing on ALOHA is to illustrate the basic ideas in the simplest setting.

⁴Some improvement can be gained by allowing users to vary both their rate and power, but, as we will show below, this improvement is minor.

The total throughput of the system, $s(p, n)$ in bits/second is then given by

$$s(p, n) = (np(1-p)^{n-1}) R \left(\frac{\bar{P}}{\int_{F_H^{-1}(p)}^{\infty} f_H(h) \frac{1}{h} dh} \right). \quad (2)$$

This expression is the product of two terms. The first term represents the probability of a successful transmission in a slot; the second term gives the transmission rate of each success. The transmission probability p can be chosen to maximize this expression. Initially, we consider the sub-optimal choice of $p = \frac{1}{n}$; this choice maximizes the first term in (2) and simplifies the following analysis; subsequently, we will compare this choice to the optimal p which maximizes $s(p, n)$.

B. Throughput scaling

Given $p = \frac{1}{n}$, we denote the throughput in (2) by $s(n)$. We consider how $s(n)$ scales as n increases. Notice that the first term in (2) is decreasing with n and approaches the well-known asymptote of $\frac{1}{e}$. The threshold $H_0 = F_H^{-1}(\frac{1}{n})$ increases as n increases. Hence, with an average power constraint of \bar{P} , the maximum received power per slot, P_r , will also increase with n . This in turn will enable users to transmit at a higher rate. Taking both of these factors into account, the total throughput is increasing with n ; the rate of increase is given by the following proposition.⁵

Proposition 1: Assume $R(x)$ is a monotonically increasing and concave function of x , and $\frac{F_H(h)}{h} = O(f_H(h))$, then as $n \rightarrow \infty$,

$$s(n) = \Theta \left(R \left(\bar{P} n F_H^{-1} \left(\frac{1}{n} \right) \right) \right).$$

Proof: See Appendix I.

The p.d.f.'s for most widely used fading models have exponential tails, in which case the condition that $\frac{F_H(h)}{h} = O(f_H(h))$ is satisfied. As an example, consider a Rayleigh fading channel, i.e. $f_H(h) = \frac{1}{h_0} e^{-\frac{h}{h_0}}$, where $h_0 = \mathbb{E}H$, and assume $R(P_r) = W \log(1 + \frac{P_r}{N_0 W})$. From Prop. 1, the throughput $s(n)$ is then increasing at a rate of $\Theta(\log(n) + \log(\log(n)))$.

C. Optimal power/rate allocation

Now instead of channel inversion, we consider an optimal power allocation scheme. In other words, when a user transmits, it can vary both its transmission power and rate in order to maximize the total throughput. For simplicity, we focus on the case of a Rayleigh fading channel with $R(P) = W \log(1 + \frac{HP}{N_0 W})$; similar results hold for other fading distributions that satisfy the condition in Prop. 1. In this case we must specify a power allocation $P(h)$, which indicates the transmission power used for all states $h > H_0$, subject to an average power constraint. Thus for a given number of users,

⁵We use the notation $a_n = O(b_n)$ if $\lim_{n \rightarrow \infty} \frac{|a_n|}{|b_n|} < \infty$ and $a_n = \Theta(b_n)$ if both $\lim_{n \rightarrow \infty} \frac{|a_n|}{|b_n|} < \infty$ and $\lim_{n \rightarrow \infty} \frac{|b_n|}{|a_n|} < \infty$.

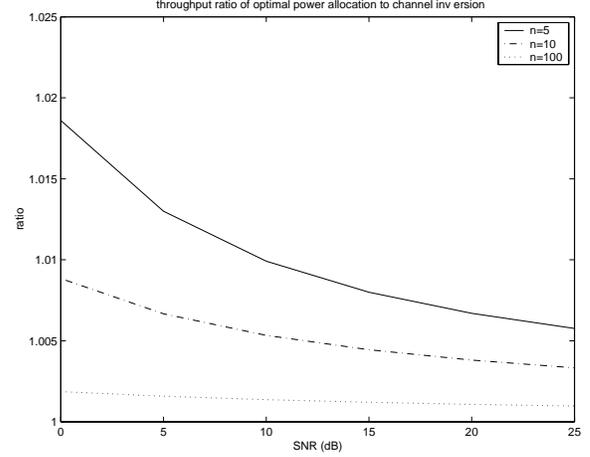


Fig. 1. Ratio of the throughput with optimal power allocation to that with channel inversion versus the expected received SNR for $n = 5, 10, 100$ users.

n , and still assuming $p = 1/n$, the optimal power allocation is the solution to the following optimization problem:

$$\begin{aligned} & \underset{P(h)}{\text{maximize}} && W \int_{h_0 \ln(n)}^{\infty} \frac{1}{h_0} e^{-\frac{h}{h_0}} \log \left(1 + \frac{P(h)h}{N_0 W} \right) dh \\ & \text{s.t.} && \int_{h_0 \ln(n)}^{\infty} \frac{1}{h_0} e^{-\frac{h}{h_0}} P(h) dh = \bar{P}. \end{aligned}$$

The solution of this optimization problem will be a well-known ‘‘water-filling’’ power allocation [4] over those channel states, $h > h_0 \ln(n)$. This can be written as,

$$P(h) = \begin{cases} N_0 W \left(\frac{1}{h'} - \frac{1}{h} \right), & h > \max\{h', h_0 \ln(n)\} \\ 0, & h < \max\{h', h_0 \ln(n)\}, \end{cases}$$

where h' is chosen to satisfy the average power constraint. For n large enough, it can be shown that $h' < h_0 \ln(n)$, in which case we have

$$h' = \frac{N_0 W}{n \bar{P} + n N_0 W \int_{h_0 \ln(n)}^{\infty} \frac{1}{h_0 h} e^{-\frac{h}{h_0}} dh}, \quad (3)$$

and the throughput becomes

$$\begin{aligned} s_{op}(n) &= W n \left(1 - \frac{1}{n} \right)^{n-1} \\ & \quad * \int_{h_0 \ln(n)}^{\infty} \frac{1}{h_0} e^{-\frac{h}{h_0}} \log \left(\frac{h}{h'} \right) dh. \end{aligned}$$

The next proposition states that asymptotically there is no advantage to optimally allocating the power.

Proposition 2: As $n \rightarrow \infty$, $\frac{s_{op}(n)}{s(n)} \rightarrow 1$, where $s(n)$ is given in (2).

The proof of this is similar to Prop. 1 and is omitted.

In Fig. 1, the throughput ratio of a system using the optimal power allocation to a system using channel inversion is shown as the function of the expected received SNR. As can be seen, the ratio is close to 1 even for a small number of users and a small SNR. The ratio decreases as the number of users increases and as the SNR increases.

D. Short-term Power Constraint

In addition to a long-term average power constraint, a user in wireless network may also have short-term power constraints, for example, limiting the power used for any transmission. We model this as a constraint on the transmission power a user can use in any one time-slot. For the model in Sect. II.B., let $P_m(n)$ be the maximum power used by any user in a system with n users. Under a long-term average power constraint \bar{P} ,

$$P_m(n) = \frac{P_r}{H_0} = \frac{\bar{P}}{H_0 \int_{F_H^{-1}(\frac{1}{n})}^{\infty} f_H(h) \frac{1}{h} dh}.$$

As the number of users increases, the fraction of time any one user transmits will decrease. Hence, under a long term power constraint, the power per transmission will increase with the number of users. This quantity grows linearly with the number of users, as summarized by the following proposition.

Proposition 3: Assume $R(x)$ is a monotonically increasing and concave function of x , and $\frac{F_H(h)}{h} = O(f_H(h))$, then as $n \rightarrow \infty$, $P_m(n) = \Theta(n)$.

The proof of this is similar to Prop. 1 and is omitted. It follows that a short term power constraint will eventually limit the throughput growth in Prop. 1. Next, we consider how the throughput scales with such a constraint. Specifically, we assume that each user's transmission power is limited to be less than P_m in each slot. As in Sect. II.B we assume the transmission rate is fixed. In this case, the maximum rate at which a user can transmit at for all channel gains $H > H_0$, is given by $R(H_0 P_m)$. If the threshold H_0 is still chosen so that the transmission probability is $1/n$, then the total throughput is given by

$$s_m(n) = \left(1 - \frac{1}{n}\right)^{n-1} R\left(P_m F_H^{-1}\left(\frac{1}{n}\right)\right). \quad (4)$$

This quantity can be shown to be increasing at rate $\Theta(R(P_m F_H^{-1}(\frac{1}{n})))$. Notice that this is a slower rate of growth than under an average power constraint. This can be explained as follows, with an average power constraint the total power available to the system is increasing with the number of users. This increase in power is responsible for part of the growth rate given in Prop. 1. Indeed, if the average power per user is normalized by the number of users, the growth rate under the average power constraint will be the same as with a short-term power constraint. For our example of a Rayleigh fading channel with $R(P_r) = W \log(1 + \frac{P_r}{N_o W})$, $s_m(n)$ in (4) can be shown to increase at rate $\Theta(\log(\log(n)))$.

E. Optimal transmission probability

So far we have been assuming that the transmission probability for each user is $p = \frac{1}{n}$; as noted in Sect. II.A., this is generally a suboptimal choice for p . In this section we look at the difference in throughput when p is chosen optimally. We focus on the Rayleigh fading example with a short-term power constraint of P_m . In this case, given that each user

transmits with probability p , then the channel threshold is $H_0 = -h_0 \ln(p)$, and the total throughput is given by

$$s_m(p, n) = np(1-p)^{n-1} W \log\left(1 - \frac{P_m h_0 \ln(p)}{N_o W}\right) \quad (5)$$

The optimal transmission probability, $p^*(n)$ is the value of p that maximizes (5). The next proposition shows that $p(n) = 1/n$ is a good approximation of the optimal transmission probability and it approaches the optimum as n approaches infinity. A similar result holds under a long term power constraint.

Proposition 4: For any finite n , $p^*(n) = \frac{\alpha(n)}{n}$, where $0 < \alpha(n) < 1$, and $\alpha(n) \rightarrow 1$ as $n \rightarrow \infty$.

Proof: See Appendix II.

From this proposition, it follows that $s_m(p^*(n), n)$ also grows at rate $\Theta(\log(\log(n)))$, and $s_m(p^*(n), n)$ approaches $s_m(n)$ as n increases; therefore the preceding discussion applies equally well when the optimal transmission probability is used.

F. Numerical comparisons

In this section, we present a numerical study that compares the throughput of the channel-aware ALOHA protocol with three other cases under a short-term power constraint.

The first case we compare to is a slotted ALOHA system where there is no fading and the channel between each user has a constant channel gain of h_0 . In this case, given a short-term power constraint of P_m , the maximum rate a user can send in a slot is

$$R = W \log\left(1 + \frac{P_m h_0}{N_o W}\right)$$

independent of the number of users. Thus, in this case the total throughput is given by

$$s_{nf}(n) = W \left(1 - \frac{1}{n}\right)^{n-1} \log\left(1 + \frac{P_m h_0}{N_o W}\right). \quad (6)$$

As $n \rightarrow \infty$ this quantity approaches a constant value of $\frac{W}{e} \log\left(1 + \frac{P_m h_0}{N_o W}\right)$.

The second case is an ALOHA system with Rayleigh fading where users do not base their transmissions on the channel state. As in the channel aware ALOHA, users attempt to transmit at a constant rate R and still do "channel inversion". With Rayleigh fading, a user will not be able to transmit as the channel gain approaches to zero and satisfy the short-term power constraint. Therefore, there will still be a threshold h_{min} , and only above this threshold will users transmit, *i.e.* users will not transmit when the channel fading is severe. The difference in this case is that we assume that this threshold does not change with the number of users. We choose h_{min} subject to a short term power constraint to maximize the average throughput, *i.e.*

$$h_{min} = \arg \max_h \left\{ \log\left(1 + \frac{P_m h}{N_o W}\right) e^{-\frac{h}{h_0}} \right\}. \quad (7)$$

The throughput in this case is

$$s_a(n) = W \left(1 - \frac{1}{n}\right)^{n-1} \log \left(1 + \frac{P_m h_{min}}{N_o W}\right) e^{-\frac{h_{min}}{h_0}}. \quad (8)$$

The third case we consider is a TDM system. In a fading channel, assume each user is assigned a time slot in a TDM frame. During each time slot only one user can transmit. As in the second case, we assume users transmit at a constant transmission rate R , and only transmit when the channel gain is larger than h_{min} given in (7). The throughput in this case is given by

$$s_{TDM}(n) = \log \left(1 + \frac{P_m h_{min}}{N_o W}\right) e^{-\frac{h_{min}}{h_0}}. \quad (9)$$

Figure 2 shows a comparison of the total throughput as a function of the number of users in all four cases. The parameters are the same as before. Notice for the first case when fading is not present, the throughput approaches to a constant as expected. For the second case, the throughput is even lower than in the first case. The poor performance can be attributed to the fact that the MAC protocol ignores the channel variations leaving the physical layer to compensate for deep fades. For the third case, notice that for small values of n , this TDM approach has a higher throughput than the channel aware ALOHA system. As n grows, however, the ALOHA approach quickly achieves higher throughputs, despite the fact that collisions occur. This is interesting as in a wire-line channel, for a backlogged system, a TDM approach would always be preferable to any random access technique. However in the wireless setting, the channel-aware ALOHA system has a higher throughput when enough users are present to provide sufficient multiuser diversity. Fig. 3 shows the number of users required for channel-aware Aloha to have a higher throughput than TDM under both long-term power constraint and short-term power constraints as a function of a function of the SNR. Under a long-term power constraint, channel inversion is used in channel-aware Aloha, while the optimal power allocation is used in the TDM scheme. It can be seen that when the SNR is not too large, only small number of users are required to out perform TDM. Therefore, the channel-aware Aloha would perform better than the TDM approach in a fading environment when power is limited.

III. COMPARISONS WITH CENTRALIZED SCHEDULERS

A. Throughput scaling

In this section, we compare the throughput scaling of the channel aware ALOHA protocol with that achieved by a centralized scheduler. In this section, we are still considering a backlogged system. As in [7], to maximize the total throughput, the optimal centralized scheduler will schedule the user with the best channel gain in each slot. Assuming variable-rate transmissions and a short-term power constraint of P_m , the average throughput achieved, $s_{ct}(n)$, is then

$$s_{ct}(n) = \mathbb{E} \left(R(P_m, \max_{i=1, \dots, n} H_i) \right). \quad (10)$$

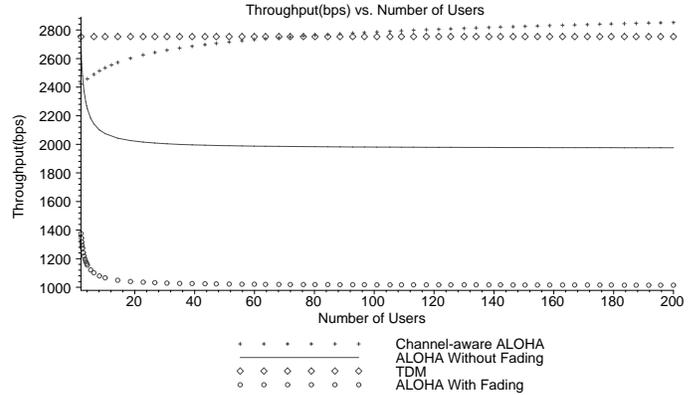


Fig. 2. Comparison of throughput versus the number of users under a short-term power constraint. The parameter used here are $W = 1\text{KHz}$, $P_m = 40\text{W}$, $N_o = 0.001$ and $h_0 = 1$.

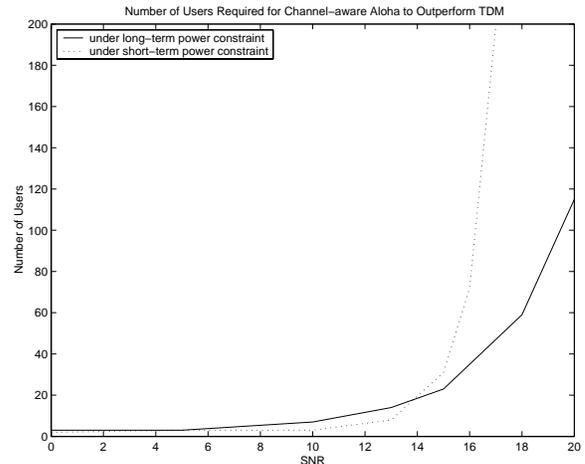


Fig. 3. Number of users required for Channel-aware ALOHA to outperform TDM under both short-term and long-term power constraints.

The rate of growth of this quantity depends on the growth of the maximum of n i.i.d. random variables. The following result from extreme order statistics [3] is helpful in characterizing this growth [17].

Lemma 1: Let z_1, \dots, z_n be i.i.d. random variables with a complimentary distribution function $F(\cdot)$ and p.d.f. $f(\cdot)$ satisfying $F(z) < 1$ for all z , $F(z)$ is twice differentiable for all z , and $\lim_{z \rightarrow \infty} \left[\frac{F(z)}{f(z)} \right] = c > 0$ for some constant c . Then $\max_{1 \leq i \leq n} z_i - l_n$ converges in distribution to a limiting random variable with c.d.f. $\exp(-e^{-\frac{x}{c}})$, where l_n is given by $F(l_n) = 1/n$.

Common fading distributions, such as Rayleigh and Ricean, satisfy the assumptions of this lemma. For the above system, this implies the maximum channel gain grows like l_n , where

$l_n = F_H^{-1}(\frac{1}{n})$. Therefore with an optimum scheduler, the throughput will increase at the rate $R(P_m F_H^{-1}(\frac{1}{n}))$; this is the same as the rate at which $s_m(n)$ is growing in the ALOHA system under a short-term power constraint (cf. in Sect. II.D). Also notice that in the ALOHA system, we restricted ourselves to using a fixed transmission rate; with this restriction we are still able to achieve the optimal growth rate in throughput. The following proposition shows the ratio of $s_m(n)$ to the transmission rate, $s_{ct}(n)$ converges to $1/e$ as n approaches to infinity.

Proposition 5: If $R(\cdot)$ is strictly increasing, then

$$\lim_{n \rightarrow \infty} \frac{s_m(n)}{s_{ct}(n)} = \lim_{n \rightarrow \infty} (1 - \frac{1}{n})^{n-1} = 1/e$$

Proof: Let $z_n = R(P_m h_n)$. Because $R(\cdot)$ is monotonically increasing, we have

$$\max z_n = \max R(P_m h_n) = R(P_m \max(h_n)).$$

The complimentary distribution function $F(z) = F_H(R^{-1}(z)/P_m)$, satisfies the conditions of Lemma 1 (assuming $R^{-1}(\cdot)$ is twice differentiable). Hence, $\max z_n - l_n$ converges in distribution to a limiting random variable with c.d.f. $\exp(-e^{-x/c})$, where $l_n = R(P_m F_H^{-1}(1/n))$.

Therefore, we have

$$\lim_{n \rightarrow \infty} E(\max z_n - l_n) = c_0,$$

where $c_0 = E(X)$ and X is a random variable with c.d.f. $\exp(-e^{-x/c})$, for some constant c . Recall,

$$s_m(n) = \left(1 - \frac{1}{n}\right)^{n-1} R(P_m F_H^{-1}(1/n)).$$

It follows that,

$$\lim_{n \rightarrow \infty} s_{ct}(n) - s_m(n) \left(1 - \frac{1}{n}\right)^{-(n-1)} = c_0.$$

Dividing by $s_m(n)$ and taking limits, we have the desired result. \blacksquare

Prop. 5 implies that that $\lim_{n \rightarrow \infty} \frac{R(P_m F_H^{-1}(1/n))}{s_{ct}(n)} = 1$; this means the transmission rate averaged over all non-collision slots in the channel-aware ALOHA ($R(P_m F_H^{-1}(1/n))$) approaches the average rate in the optimal centralized scheme $E(R(\max(h_n)))$, as n approaches to infinity. In other words, asymptotically, the only penalty incurred due to distributed channel knowledge is due to the contention inherent in the ALOHA protocol.

B. Finite user comparisons

In the previous section, we assumed that in the ALOHA system all users transmitted at a fixed rate, while in the centralized approach a user employed variable rate transmission. In this section, we assume that users in both the ALOHA system and the centralized approach use variable rate transmission with a fixed transmission power of P_m . For the ALOHA case, we still assume that each user only transmits when $H > H_0$, however

the rate at which the user transmits is given by $R(HP_m)$. In this case, the throughput for the ALOHA system is given by

$$\tilde{s}_m(n) = \left(1 - \frac{1}{n}\right)^{n-1} E(R(P_m H)|H > H_0),$$

where we are still assuming $H_0 = F_H^{-1}(1/n)$.

The ratio $r_f(n) = \frac{\tilde{s}_m(n)}{s_{ct}(n)}$ can be viewed as a measure of the loss in throughput due to the medium access control protocol. In a channel without fading, the ratio of the throughput when using ALOHA compared to the throughput with a centralized scheduler is simply $r_{nf}(n) = (1 - \frac{1}{n})^{n-1}$. Both $r_f(n)$ and $r_{nf}(n)$ converge to $\frac{1}{e}$ as $n \rightarrow \infty$. Next, we compare these two quantities for finite n and show that $r_f(n) > r_{nf}(n)$ for all finite n . In other words, the penalty for lack of coordination is smaller in a fading channel. This is summarized in the following proposition.

Proposition 6: For all finite n , $r_f(n) > r_{nf}(n)$.

Proof: From the definitions of the various quantities this is equivalent to showing that $E(R(P_m H)|H > H_0) > E(R(P_m H_{max,n}))$, where $H_{max,n} = \max_{i=1, \dots, n} \{H_i\}$.

To show this we use the following stochastic ordering result [13]: for two random variables, X and Y , if $\Pr(X > a) \geq \Pr(Y > a)$ for all a , then $E[f(X)] \geq E[f(Y)]$ for all increasing functions f .

Let X be the channel gain H when a successful transmission occurs in the ALOHA systems, thus

$$\Pr(X > h) = \begin{cases} nF_H(h) & \text{for all } h > H_0 \\ 1 & \text{otherwise,} \end{cases}$$

where we have used that $\Pr(H > H_0) = 1/n$. Let $Y = H_{max,n}$, then

$$\begin{aligned} \Pr(Y > h) &= 1 - p(Y < h) \\ &= 1 - (1 - F_H(h))^n \end{aligned}$$

For $h < H_0$, $\Pr(X > h) = 1$, thus $\Pr(X > h) > \Pr(Y > h)$. For all $h > H_0$, let $z = F_H(h)$. Since $nz > 1 - (1 - z)^n$ for all $0 < z < 1$, we again have $\Pr(X > h) > \Pr(Y > h)$ for all $h > H_0$. By assumption, $R(x)$ is monotonically increasing, thus applying the above result we have $E(R(P_m H)|H > H_0) > E(R(P_m H_{max,n}))$ as desired. \blacksquare

C. Splitting Algorithm

If the time-scale over which the channel varies is larger than the round-trip time between each user and the receiver, then the throughput of the channel aware ALOHA approach can be improved by using a type of splitting algorithm to resolve collisions [2]. Furthermore, by basing the splitting decision on the users' channel gains, the splitting algorithm could be used to find the user with the best channel at the beginning of each slot in a distributed way. Assume that the round-trip time is β , where this is smaller than the time-slot used to send packets. At the beginning of each slot, consider using several mini-slots with length β to communicate with the base station and find the best user.

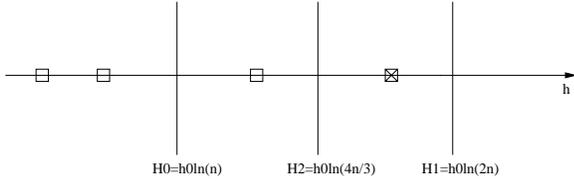


Fig. 4. One example of the splitting sequence

As before, we assume instantaneous $(0, 1, e)$ feedback after each mini-slot. Based on this feedback information, we set the threshold for the next slot to attempt to maximize the success probability in that slot. One example of such a splitting sequence is shown in Fig. 4. Initially we set the threshold $H_0 = h_0 \ln(n)$. Users whose channel gains are above H_0 transmit in the first mini-slot. If a collision occurs, we know there are two or more users with channel gains above H_0 . The probability of more than two users' channels being above H_0 is much less than the probability that exactly two users are above H_0 when n is large. Therefore we set the next threshold $H_1 = h_0 \ln(2n)$ so that the probability that one user's channel is above H_1 given that two users' channels are above H_0 is $1/2$. In the second mini-slot, users whose channel gains are above H_1 transmit. In the example, no users will transmit in this slot; thus we know there must be two or more users whose channel gains are within the range $H_0 < H < H_1$. The next threshold is set to be $H_2 = h_0 \ln(4n/3)$, again this is chosen so that the probability that one user's channel is in the range $H_2 < H < H_1$ given that two users' channels are in the range $H_0 < H < H_1$ is $1/2$. After the third mini-slot, a "1" feedback is received and the user with the best channel gain is found and will transmit in the rest of the slot.

Generally, at the beginning of each slot, we set the threshold $H_0 = h_0 \ln(n)$ initially. Then based on the feedback received in each mini-slot, there are the following three possibilities: (i) if the feedback is "1" (success), the requested user will continue to transmit the data packet; (ii) if feedback is "0" (idle), the threshold is lowered; (iii) if the feedback is e (collision), we know there are multiple users within the current range, so we increase the threshold. In each mini-slot, users whose channel gain is larger than the threshold attempt to transmit. This process continues until a successful transmission occurs or there are no more mini-slots in the time-slot. In the next time-slot a new search interval begins. Compared to the previous approaches, the extra overhead involved is the mini-slots we use to transmit the requests.

Fig. 5 presents simulation results showing the ratio of the throughput from the splitting algorithm compared to a centralized scheduling scheme for a Rayleigh fading channel. This ratio decreases slightly as the number of users increases; this is because with more users, more overhead is required for splitting. Fig. 6 shows the number of mini-slots required as the number of users in the system increases. Notice that for the range shown, the number of required mini-slots is small, which suggests that this approach is effective.

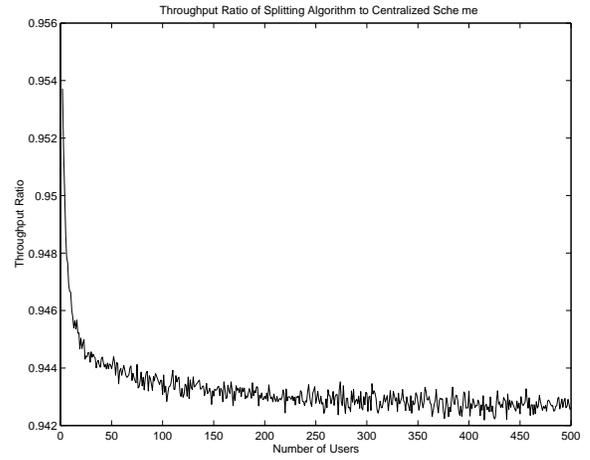


Fig. 5. Ratio of throughput of splitting algorithm to centralized scheduler for different numbers of users.

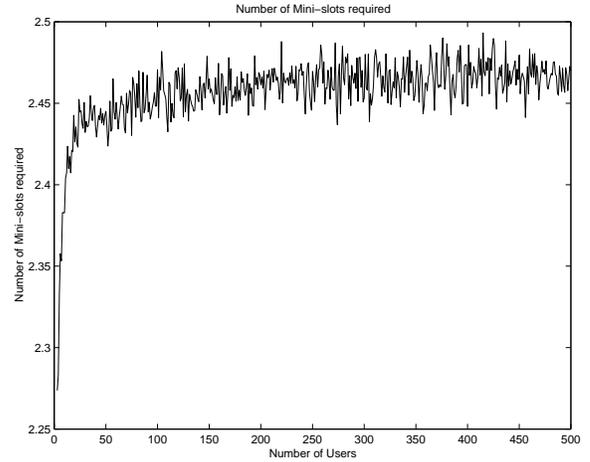


Fig. 6. The number of mini-slots required for a success by the splitting algorithm

The slot length used in Fig. 5 is 4ms, and β is 0.1 ms. If the ratio of β to the slot length decreases, throughput of the splitting algorithm will increase and approaches to the central scheduling scheme; if the ratio increases, the performance approaches the ALOHA algorithm. A more detailed analysis of this approach can be found in [20].

IV. RANDOM ARRIVALS

In previous sections, we assumed all nodes are backlogged and have an infinite reservoir of packets to send. In this section, we relax this assumption and assume packets randomly arrive with total arrival rate λ . First, we consider an infinite user model, where it is assumed that each new packet arrives to a new user [2]. Such a model is reasonable for a system with a large number of users, each with a small arrival rate. We assume that the number of backlogged users n in each slot is known. Practically, the backlog would have to be estimated using an algorithm such as the Pseudo-Bayesian algorithm [2].

We still consider an approach where users base their transmission on whether their channel gain is above a threshold H_0 ; however, now we allow this threshold to depend on the backlog. Specifically, we assume that

$$H_0(n) = \begin{cases} h_{min} & \text{for } F_H^{-1}(\frac{1}{n}) < h_{min} \\ F_H^{-1}(\frac{1}{n}) & \text{for } F_H^{-1}(\frac{1}{n}) \geq h_{min}. \end{cases} \quad (11)$$

Here h_{min} is the minimum threshold above which the user will transmit regardless of the backlog.

As in Sect. II.D., we consider a model with a short-term power constraint of P_m . Given that n users are backlogged, each user will transmit at rate $R(P_m H_0(n))$ if successful. As n increases, the transmission rate $R(P_m H_0(n))$ will also increase. If all packets have a fixed length of L bits, then the time needed to transmit a packet is $L/R(P_m H_0(n))$, which will decrease as n increases. We consider a slotted-time model, where the length of time-slots vary with the backlog according to this relationship. Packet arrivals are assumed to be independent in each time-slot with an expected arrival rate of $\lambda L/R(P_m H_0(n))$. In this section, we still assume that the channel variation is memoryless between slots.⁶

Given the above assumptions, we consider over what range of arrival rates, λ , the system is stable. The following proposition states that if $R(P_m H_0(n))$ is unbounded (as in the Rayleigh fading model), then the system will be stable for any total arrival rate.⁷

Proposition 7: Under memoryless fading, if $R(P_m H_0(n))$ is unbounded, then the infinite user, channel-aware ALOHA system is stable for any arrival rate λ .

Proof: Let for $t = 1, 2, \dots$, let $n(t)$ denote the backlog at the start of the t th time-slot. Given the memoryless assumption, $\{n(t)\}$ will be a Markov chain. To show that the system is stable, it is sufficient to show the following drift condition[2]: there exists some $D > 0$, $N > 0$ such that

$$E(n(t+1) - n(t) | n(t) = n) \leq -D \quad (12)$$

for all $n \geq N$.

Given that $n(t) = n$, each user will transmit with probability $1/n$ in each slot, therefore the departure rate in packets per time-slot would be $(1 - \frac{1}{n})^{n-1}$. The arrival rate in packets per time-slot would be $\lambda(n) = \lambda L/R(P_m H_0(n))$. Thus we have,

$$E(n(t+1) - n(t) | n(t) = n) = -(1 - \frac{1}{n})^{n-1} + \lambda(n) \quad (13)$$

As n approaches to infinity, $\lambda(n)$ decreases to 0, while $(1 - \frac{1}{n})^{n-1}$ approaches to $1/e$. Therefore for any $\delta < 1/e$, an N can be found such that (12) is satisfied, and so the system is stable. ■

Figure 7 illustrates the idea behind this stability result. This figure shows both the total arrival rate and departure rate normalized in units of packets per time-slot, as a function

⁶Since the slot sizes are variable, this may seem to be a questionable assumption. In Sect. V, we discuss removing this restriction.

⁷However, for high enough arrival rates this requires a prohibitively high diversity gain and the underlying physical model becomes unrealistic.

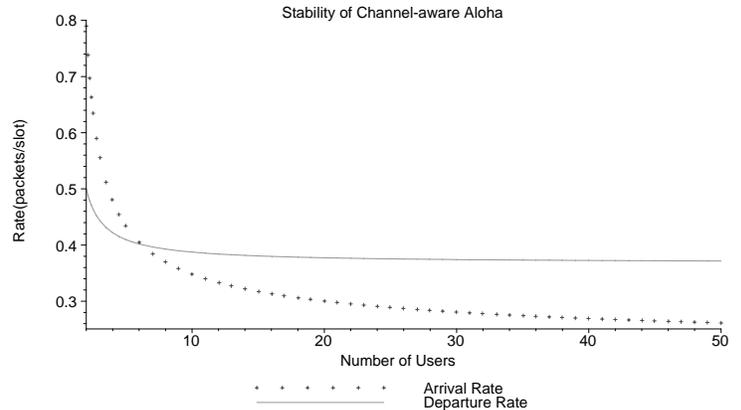


Fig. 7. Stability of channel-aware ALOHA.

of the backlog. The parameters used in the figure are $\lambda = 0.6$ packets/second, $L = 1000$ bits/packets, $W = 1$ kHz and $\frac{P_m h_0}{N_0 W} = 1$. For small backlogs the normalized arrival rate is larger than the departure rate, and so the backlog will tend to increase. Eventually, for high enough backlogs, the arrival rate will drop below the departure rate; the system will stabilize around the point where these curves cross. As the arrival rate increases, the system will stabilize around a larger backlog; this is because more users are needed to provide the multiuser diversity gain necessary to stabilize the system. The higher backlog results in a larger delay. This is illustrated in Fig. 8. This figure shows simulation results of the delay for a system with a finite number of users for various arrival rates, λ . In these simulations, arrivals are assumed to be from a Poisson process. For each curve the total arrival rate is fixed as the number of users varies. Notice that for a given arrival rate, the delay decreases as the number of users increase; this is due to the increased multiuser diversity.

We note that Prop. 7 does not imply that a system with a finite number of users is stable for any arrival rate. For example, consider a system with n users and symmetric traffic. If the arrival rate λ satisfies $\lambda * L/R(P_m H_0(n)) > 1/e$, the system will be unstable.

V. CHANNEL WITH MEMORY

We have been considering the case where the channel is memoryless from slot to slot. In this section, we consider a simple model that incorporates channel memory. Specifically, we still assume that each backlogged user only transmits when its channel is above a threshold value, H_0 . When the channel is memoryless, the transmission attempts of a user are an i.i.d. sequence of random variables. For a channel with memory, we assume that the sequence of transmission attempts by a backlogged user can be modeled as a two-state Markov chain

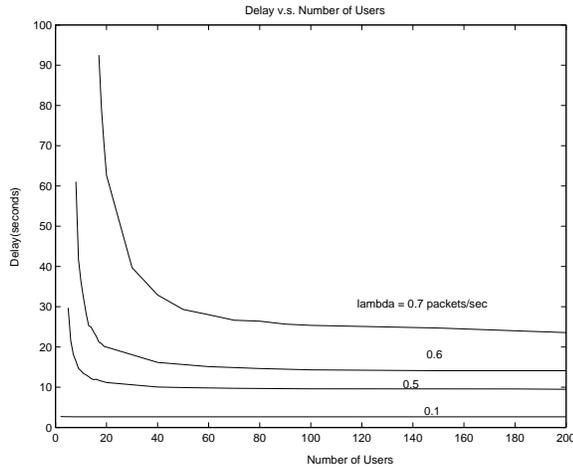


Fig. 8. Delay v.s. number of users

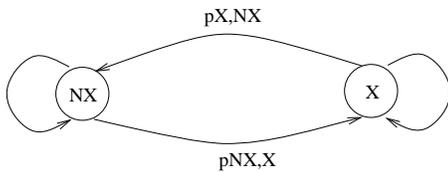


Fig. 9. Markov chain representing each user's transmission state for the channel model with memory.

as shown in Fig. 9. The state X corresponds to time-slots when the user will transmit (*i.e.*, the channel gain is greater than H_0), and the state NX corresponds to slots when the user does not transmit. The transition probability from state X to NX and from state NX to X is denoted by $p_{X,NX}$ and $p_{NX,X}$, respectively. If the steady-state probability of state X (*i.e.*, the probability that a user attempts a transmission) is p , then it can be shown that the transition probabilities must have the form $p_{X,NX} = r(1 - p)$ and $p_{NX,X} = rp$, for some $r > 0$. Notice that $r = 1$ corresponds to a memoryless channel; smaller values of r correspond to increased channel memory.

First, consider a backlogged system with n users; each user transmitting according to the Markov chain in Fig. 9. Because the users' channel gains are independent of each other, the steady-state probability of a successful transmission is still given by $np(1 - p)^{n-1}$, as in the memoryless case. The total throughput, will only depend on the steady-state distribution, hence the analysis in Sect. II can be seen to apply in this case as well. In other words, for the backlogged model, channel memory has no effect on the total throughput.

Next, we consider the random arrival case from Sect. IV. As in the memoryless case, we assume that the length of a time-slot needed to send a packet will vary depending on the backlog. When n users are backlogged, each user's transmission attempts can be modeled as in Fig. 9, where the transition probabilities of this Markov chain are parameterized by the quantity r . Only now this parameter will vary with the number of backlogged users; we denote this by $r(n)$. As the

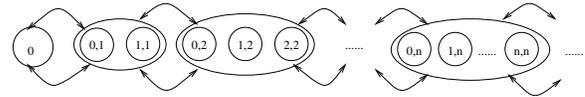


Fig. 10. Markov chain model for the whole process.

backlog increases, there are two factors that effect $r(n)$ – first for a fixed threshold, H_0 , shorter time-slots will lead to an increased correlation between time-slots and hence a smaller $r(n)$; on the other hand, for a fixed time-slot, as the threshold, H_0 , increases, the probability that the channel stays above the threshold will decrease, leading to larger $r(n)$. The exact behavior of $r(n)$ will depend on the specifics of the underlying channel model. We will look at one example of this at the end of this section.

Given a backlog of n users, assume that the channel threshold is still set so the transmission probability, p is $1/n$. In this case, the transition probabilities in Fig. 9 become $p_{X,NX} = r(n)(1 - 1/n)$ and $p_{NX,X} = r(n)/n$.

The entire system can also be modeled as a Markov chain, a part of which is shown in Fig. 10. Each state has two parameters, (k, n) , k stands for the number of backlogged users whose channel gains are above the threshold level and n stands for the number of backlogged users. Given that the system is in state (k, n) and no new arrivals occur, the probability of a successful transmission attempt in the next slot is given by

$$\begin{aligned}
 p_s(k, n) &= k(p_{X,NX})^{k-1} p_{X,X} (p_{NX,NX})^{n-k} \\
 &\quad + (n - k)p_{NX,X}
 \end{aligned}$$