How Does Multiple-Packet Reception Capability Scale the Performance of Wireless Local Area Networks?

Ying Jun (Angela) Zhang, *Member*, *IEEE*, Peng Xuan Zheng, *Student Member*, *IEEE*, and Soung Chang Liew, *Senior Member*, *IEEE*

Abstract—Due to its simplicity and cost efficiency, wireless local area network (WLAN) enjoys unique advantages in providing high-speed and low-cost wireless services in hot spots and indoor environments. Traditional WLAN medium-access-control (MAC) protocols assume that only one station can transmit at a time: simultaneous transmissions of more than one station cause the destruction of all packets involved. By exploiting recent advances in PHY-layer multiuser detection (MUD) techniques, it is possible for a receiver to receive multiple packets simultaneously. This paper argues that such multipacket reception (MPR) capability can greatly enhance the capacity of future WLANs. In addition, the paper provides the MAC-layer and PHY-layer designs needed to achieve the improved capacity. First, to demonstrate MPR as a powerful capacity-enhancement technique, we prove a "superlinearity" result, which states that the system throughput per unit cost increases as the MPR capability increases. Second, we show that the commonly deployed binary exponential backoff (BEB) algorithm in today's WLAN MAC may not be optimal in an MPR system, and the optimal backoff factor increases with the MPR capability, the number of packets that can be received simultaneously. Third, based on the above insights, we design a joint MAC-PHY layer protocol for an IEEE 802.11-like WLAN that incorporates advanced PHY-layer signal processing techniques to implement MPR.

Index Terms—Wireless local area network, exponential backoff, multipacket reception.

1 Introduction

1.1 Motivation

The last decade has witnessed a surge of interest in wireless local area networks (WLANs), where mobile stations share a common wireless medium through contention-based medium access control (MAC). In WLANs, collision of packets occurs when more than one station transmits at the same time, causing a waste of bandwidth. Recent advances in multiuser detection (MUD) techniques [1] open up new opportunities for resolving collisions in the physical (PHY) layer. For example, in CDMA [2] or multiple-antenna [3] systems, multiple packets can be received simultaneously using MUD techniques without collisions. It is expected that with improved multipacket reception (MPR) capability from the PHY layer, the MAC layer will behave differently from what is commonly believed. In particular, to fully utilize the MPR capability for capacity enhancement in WLAN, it is essential to understand the fundamental impact of MPR on the MAC-layer design. As such, this paper is an attempt to study the MAC-layer throughput performance and the collision resolution schemes for WLANs with MPR.

- *Y.J. Zhang and S.C. Liew are with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong. E-mail: {yjzhang, scliew}@ie.cuhk.edu.hk.*
- *P.X. Zheng is with the Department of Computer Science, Purdue University, West Lafayette, IN 47907-2107. E-mail: pzheng@cs.purdue.edu.*

Manuscript received 5 Nov. 2007; revised 14 May 2008; accepted 9 Dec. 2008; published online 18 Dec. 2008.
For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-2007-11-0326. Digital Object Identifier no. 10.1109/TMC.2008.169.

1.2 Key Contributions

The key contributions of this paper are as follows:

- To demonstrate MPR as a powerful capacity-enhancement technique at the system level, we analyze the MAC-layer throughput of WLANs with MPR capability under both finite-node and infinite-node assumptions. Our model is sufficiently general to cover both carrier-sensing and noncarrier-sensing networks. We prove that in random-access WLANs, network throughput increases superlinearly with the MPR capability of the channel. That is, throughput divided by M increases as M increases, where M is the number of packets that can be resolved simultaneously. The superlinear throughput scaling implies that the achievable throughput per unit cost increases with MPR capability of the channel. This provides a strong incentive to deploy MPR in next generation wireless networks.
- We study the effect of MPR on the MAC-layer collision resolution scheme, namely exponential backoff (EB). When packets collide in WLANs, an EB scheme is used to schedule the retransmissions, in which the waiting time of the next retransmission will get multiplicatively longer for each collision incurred. In the commonly adopted binary exponential backoff (BEB) scheme (e.g., used in Ethernet [15], WiFi [16], etc.), the multiplicative (a backoff factor) is equal to 2. We show in this paper that the widely used BEB does not necessarily yield the close-to-optimal network throughput with the improved MPR capability from the PHY layer. As a matter of fact, BEB is far from

1536-1233/09/$25.00 © 2009 IEEE Published by the IEEE CS, CASS, ComSoc, IES, & SPS

optimum for both noncarrier-sensing networks and carrier-sensing networks operated in basic access mode. The optimal backoff factor increases with the MPR capability. Meanwhile, BEB is close to optimum for carrier-sensing networks when RTS/CTS access mode is adopted.

- Built on the theoretical underpinnings established above, we propose a practical protocol to fully exploit the MPR capability in IEEE 802.11-like WLANs. In contrast to [7], [8], we consider not only the MAC-layer protocol design, but also the PHY-layer signal processing to enable MPR in distributed random-access WLANs. As a result, the proposed protocol can be implemented in a fully distributed manner with marginal modification of current IEEE 802.11 MAC.

1.3 Related Work on MPR and Collision Resolution Schemes

The first attempt to model a general MPR channel in random-access wireless networks was made by Ghez et al. [4], [5] in 1988 and 1989, respectively, in which stability properties of conventional slotted ALOHA with MPR were studied under a simple infinite-user and single-buffer assumption. No collision resolution scheme (such as EB) was considered therein. This work was extended to CSMA systems by Chan et al. in [26] and to finite-user ALOHA systems by Naware et al. in [6]. It has been shown in [4], [5], [6] that MPR improves the stable throughput of ALOHA only when the MPR capability is comparable to the number of users in the system. In practical networks where the MPR capability is much smaller than the number of users, the stable throughput of conventional ALOHA is equal to 0, same as the case without MPR. To date, little work has been done to investigate the throughput enhancing capability of MPR in practical WLANs with collision resolution schemes. Our paper here is an attempt along this direction.

Protocols that exploit the MPR capability of networks have been studied by Zhao and Tong [7], [8]. In [7], a multiqueue service room (MQSR) MAC protocol was proposed for networks with heterogeneous users. The drawback of the MQSR protocol is its high computational cost due to updates of the joint distribution of all users' states. To reduce complexity, a suboptimal dynamic queue protocol was proposed in [8]. In both protocols, access to the common wireless channel is controlled by a central controller, which grants access to the channel to an appropriate subset of users at the beginning of each slot. In [27], Chan et al. proposed adding a MUD layer to facilitate MPR in IEEE 802.11 WLAN. To implement the MUD techniques mentioned in [27], the AP is assumed to have knowledge of the number of concurrent transmissions, the identities of the transmitting stations, and the channel coefficients. This information, while easy to get in a network with centralized scheduling (e.g., cellular systems), is unkown to the AP a priori in random access networks. Moreover, the preambles of concurrent packets overlap and, hence, it is difficult for the AP to have a good estimation of the channel coefficients with the current protocol. By contrast, this paper provides a solution to this issue by incorporating blind signal processing in the proposed protocol.

EB as a collision resolution technique has been extensively studied in different contexts [10], [11], [12], [13]. Stability upper bound of BEB has been given by Goodman et al. under a finite-node model in [10] and recently improved by Al-Ammal et al. [11]. The throughput and delay characteristics of a slightly modified EB scheme have been studied in [12] in the context of slotted ALOHA. The characteristics of EB in steady state is further investigated in [13] in time-slotted wireless networks with equal slot length. All the existing work on EB has assumed that the wireless channel can only accommodate one ongoing transmission at a time. This paper is a first attempt to look at EB for an MPR system.

The remainder of this paper is organized as follows: In Section 2, we describe the system model and introduce the background knowledge on MUD and EB. In Section 3, we prove that the maximum achievable throughput of MPR WLAN scales superlinearly with the MPR capability of the channel. In Section 4, the effect of MPR on EB is investigated. We show that the widely used BEB scheme is no longer close-to-optimal in MPR networks. To realize MPR in IEEE 802.11 WLANs, an MAC-PHY protocol is presented in Section 5. In Section 6, we discuss some practical issues related to MPR. Finally, Section 7 concludes this paper.

2 Preliminary and System Model

2.1 System Description

We consider a fully connected infrastructure WLAN where N infinitely backlogged mobile stations communicate with an access point (AP). We assume that the time axis is divided into slots and packet transmissions start only at the beginning of a slot. In addition, after each transmission, the transmitting stations have a means to discover the result of the transmission, i.e., success or failure. If the transmission fails due to collision, the colliding stations will schedule retransmissions according to a collision resolution scheme (e.g., EB). We assume that the channel has the capability to accommodate up to M simultaneous transmissions. In other words, packets can be received correctly whenever the number of simultaneous transmissions is no larger than M. When more than M stations contend for the channel at the same time, collision occurs and no packet can be decoded. We refer to M as MPR capability.

In our model, the length of a time slot is not necessarily fixed and may vary under different contexts [9]. We refer to this variable-length slot as backoff slot hereafter. In WLANs, the length of a backoff slot depends on the contention outcome (hereafter, referred to as channel status). Let T_i denote the length of an idle time slot when nobody transmits, T_c denote the length of a collision time slot when more than M stations contend for the channel, and T_s denote the length of a time slot due to successful transmission when the number of transmitting stations is anywhere from 1 to M. The durations of T_i, T_c, and T_s depend on the underlying WLAN configuration. For noncarrier-sensing networks such as slotted ALOHA, the stations are not aware of the channel status, and the duration of all backoff slots is equal to the transmission time of a packet. That is,

$$T_{slot} = T_i = T_c = T_s = L/R, \tag{1}$$

where L is the packet size and R is the data transmission rate of a station. On the other hand, for carrier-sensing networks, stations can distinguish between various types of channel status and the durations of different types of slots may not be the same. For example, in IEEE 802.11 DCF basic-access mode

$$\begin{aligned} T_i &= \sigma, \\ T_s &= H + L/R + SIFS + \delta + ACK + DIFS + \delta, \\ T_c &= H + L/R + DIFS + \delta, \end{aligned} \tag{2}$$

where σ is the time needed for a station to detect the packet transmission from any other station and is typically much smaller than T_c and T_s; H is the transmission time of both PHY header and MAC header; *ACK* is the transmission time of an ACK packet; δ is the propagation delay; and *SIFS* and *DIFS* are the interframe space durations [16]. Similarly, in IEEE 802.11 DCF request-to-send/clear-to-send (RTS/CTS) access scheme, the slot durations are given by

$$\begin{aligned} T_i &= \sigma, \\ T_s &= RTS + SIFS + \delta + CTS + SIFS + \delta + H + L/R \\ &\quad + SIFS + \delta + ACK + DIFS + \delta, \\ T_c &= RTS + DIFS + \delta, \end{aligned} \tag{3}$$

where RTS and CTS denote the transmission time of RTS and CTS packets, respectively. By allowing the durations of T_i, T_c, and T_s to vary according to the underlying system, the analysis of this paper applies to a wide spectrum of various WLANs, including both noncarrier-sensing and carrier-sensing networks.

2.2 Multiuser Detection

This section briefly introduces the PHY-layer MUD techniques used to decode multiple packets at the receiver. Let $x_k(n)$ denote the data symbol transmitted by user k in symbol duration n. If there are K stations transmitting together, then the received signal at a receiver is given by

$$\begin{aligned} \mathbf{y}(n) &= \sum_{k=1}^{K} \mathbf{h}_k(n) x_k(n) + \mathbf{w}(n) \\ &= \mathbf{H}(n)\mathbf{x}(n) + \mathbf{w}(n), \end{aligned} \tag{4}$$

where $\mathbf{w}(n)$ denotes the additive noise, $\mathbf{H}(n) = [\mathbf{h}_1(n), \mathbf{h}_2(n), \ldots, \mathbf{h}_K(n)]$, and $\mathbf{x}(n) = [x_1(n), \ldots, x_K(n)]^T$. In multiple antenna systems, $\mathbf{h}_k$ is the channel vector, with the mth element being the channel coefficient from user k to the mth receive antenna.[1] In CDMA systems, vector $\mathbf{h}_k$ is the multiplication of the spreading sequence of user k and the channel coefficient from user k to the AP.

The receiver attempts to obtain an estimate of the transmitted symbols $\mathbf{x}(n)$ from the received vector $\mathbf{y}(n)$. To this end, various MUD techniques have been proposed in the literature. For example, the zero-forcing (ZF) receiver is one of the most popular linear detectors. It multiplies the received vector by the pseudoinverse of matrix $\mathbf{H}(n)$, denoted by $\mathbf{H}^{+}(n)$, and the decision statistics becomes

$$\begin{aligned} \mathbf{r}^{ZF}(n) &= \mathbf{H}^{+}(n)\mathbf{y}(n) \\ &= \mathbf{x}(n) + \mathbf{H}^{+}(n)\mathbf{w}(n). \end{aligned} \tag{5}$$

The minimum-mean-square-error (MMSE) receiver is the optimal linear detector in the sense of maximizing the signal-to-interference-and-noise ratio (SINR). The decision statistics is calculated as

$$\mathbf{r}^{MMSE}(n) = (\mathbf{H}(n)\mathbf{H}^{H}(n) + \eta\mathbf{I})^{-1}\mathbf{H}^{H}(n)\mathbf{y}(n), \tag{6}$$

1. In this paper, we assume that each station only transmits one data stream at a time.

where $\mathbf{I}$ is the identity matrix and η is the variance of the additive noise. Given the decision statistics, an estimate of $x_k(n)$ can be obtained by feeding the kth element of $\mathbf{r}^{ZF}(n)$ or $\mathbf{r}^{MMSE}(n)$ into a quantizer.

Other MUD techniques include maximum-likelihood (ML), parallel interference cancellation (PIC), successive interference cancellation (SIC), etc. Interested readers are referred to [1] for more details.

2.3 Exponential Backoff

EB adaptively tunes the transmission probability of a station according to the traffic intensity of the network. It works as follows: A backlogged station sets its backoff timer by randomly choosing an integer within the range $[0, W-1]$, where W denotes the size of the contention window. The backoff timer is decreased by one following each backoff slot. The station transmits a packet in its queue once the backoff timer reaches zero. At the first transmission attempt of a packet, $W = W_0$, referred to as the minimum contention window. Each time the transmission is unsuccessful, W is multiplied by a backoff factor r. That is, the contention window size $W_i = r^i W_0$ after i successive transmission failures.

3 Superlinear Throughput Scaling in WLANs with MPR

This section investigates the impact of MPR on the throughput of random-access WLANs. In particular, we prove that the maximum achievable throughput scales superlinearly with the MPR capability M. In practical systems, M is directly related to the cost (e.g., bandwidth in CDMA systems or antenna in multiantenna systems). Superlinear scaling of throughput implies that the achievable throughput *per unit cost* increases with M. This provides a strong incentive to consider MPR in next-generation wireless networks. As mentioned earlier, the transmission of stations is dictated by the underlying EB scheme. To capture the fundamentally achievable throughput of the system, the following analysis assumes that each station transmits with probability p_t in an arbitrary slot, without caring how p_t is achieved. The assumption will be made more rigorous in Section 4, which relates p_t to EB parameters such as r and W_0.

3.1 Throughput of WLANs with MPR

Define throughput to be the average number of information bits transmitted successfully per second. Let $S_N(M, p_t)$ denote the throughput of a WLAN with N stations when each station transmits at probability p_t and the MPR capability is M. Then, $S_N(M, p_t)$ can be calculated as the ratio between the average payload information bits transmitted per backoff slot to the average length of a backoff slot as follows:

$$S_N(M, p_t) = \frac{\sum_{k=1}^{M} k \Pr\{X = k\} L}{P_{idle} T_i + P_{coll} T_c + P_{succ} T_s}. \tag{7}$$

In the above, X is a random variable denoting the number of attempts in a slot

$$\Pr\{X = k\} = \binom{N}{k} p_t^k (1 - p_t)^{N-k}. \tag{8}$$

Let

$$P_{idle} = (1 - p_t)^N \tag{9}$$

be the probability that a backoff slot is idle,

$$P_{succ} = \sum_{k=1}^{M} \Pr\{X = k\} = \sum_{k=1}^{M} \binom{N}{k} p_t^k (1 - p_t)^{N-k} \tag{10}$$

be the probability that a backoff slot is busy due to successful packet transmissions, and

$$P_{coll} = \sum_{k=M+1}^{N} \Pr\{X = k\} = \sum_{k=M+1}^{N} \binom{N}{k} p_t^k (1 - p_t)^{N-k} \tag{11}$$

be the probability that a backoff slot is busy due to collision of packets.

The throughput of noncarrier-sensing networks, such as slotted ALOHA, can be obtained by substituting (1) into (7), which leads to following expression:

$$\begin{aligned} S_N(M, p_t) &= \frac{\sum_{k=1}^{M} k \Pr\{X = k\} L}{T_{slot}} \\ &= R \sum_{k=1}^{M} k \binom{N}{k} p_t^k (1 - p_t)^{N-k}. \end{aligned} \tag{12}$$

Similarly, the throughput of carrier-sensing networks, such as IEEE 802.11 DCF basic-access mode and RTS/CTS access mode, can be obtained by substituting (2) and (3) into (7), respectively.

We now derive the asymptotic throughput when the population size N approaches infinity. In this case, we assume that 1) the system has a nonzero asymptotic throughput and 2) the number of attempts in a backoff slot is approximated by a Poisson distribution with an average attempt rate $\lambda = Np_t$ [24, p. 258]. Both of these assumptions are valid under an appropriate EB scheme, which will be elaborated in Section 4. Let $S_\infty(M, \lambda)$ be the asymptotic throughput when MPR capability is M and average attempt rate is λ. Then, we derive from (7) that

$$\begin{aligned} S_\infty(M, \lambda) &= \lim_{N \to \infty} S_N \\ &= \frac{L \sum_{k=1}^{M} k \Pr\{X = k\}}{P_{idle} T_i + P_{coll} T_c + P_{succ} T_s} \\ &= \frac{L \sum_{k=1}^{M} k \frac{\lambda^k}{k!} e^{-\lambda}}{P_{idle} T_i + P_{coll} T_c + P_{succ} T_s} \\ &= \frac{L\lambda \sum_{k=0}^{M-1} \frac{\lambda^k}{k!} e^{-\lambda}}{P_{idle} T_i + P_{coll} T_c + P_{succ} T_s} \\ &= \frac{L\lambda \Pr\{X \le M - 1\}}{P_{idle} T_i + P_{coll} T_c + P_{succ} T_s}, \end{aligned} \tag{13}$$

where the third equality is due to the Poisson approximation. In particular, when $T_{slot} = T_i = T_c = T_s = L/R$,

$$S_\infty(M, \lambda) = R \sum_{k=0}^{M-1} \frac{\lambda^{k+1}}{k!} e^{-\lambda} = R\lambda \Pr\{X \le M - 1\}. \tag{14}$$

3.2 Superlinear Throughput Scaling

Having derived the throughput expressions for both finite-population and infinite-population models, we now address the question: how does throughput scale as M increases. In particular, we are interested in the behavior of the maximum throughput when the channel has a MPR capability of M. This directly relates to the channel-access efficiency that is achievable in MPR networks.

Given M, the maximum throughput can be achieved by optimizing the transmission probability p_t (or, equivalently, λ in the infinite-population model). The optimal transmission probability can, in turn, be obtained by adjusting the backoff factor r in practical WLANs, as will be discussed in Section 4. Let $S_N^*(M) = S_N(M, p_t^*(M))$ and $S_\infty^*(M) = S_\infty(M, \lambda^*(M))$ denote the maximum achievable throughputs, where $p_t^*(M)$ and $\lambda^*(M)$ denote the optimal p_t and λ when the MPR capability is M, respectively. In Theorem 1, we prove that the throughput scales superlinearly with M in noncarrier-sensing network with infinite population. In other words, $S_\infty^*(M)/M$ is an increasing function of M. In Theorem 2, we further prove that $S_\infty^*(M)/MR$ approaches 1 when $M \to \infty$. This implies that the throughput penalty due to distributed random access diminishes when M is very large. In Theorem 3 in Appendix A, we prove that the same superlinearity holds for WLANs with finite population.

Theorem 1 (Superlinearity). *$S_\infty^*(M)/M$ is an increasing function of M.*

It is obvious that, at the optimal $\lambda^*(M)$,

$$\begin{aligned} \left.\frac{\partial S_\infty(M, \lambda)}{\partial \lambda}\right|_{\lambda = \lambda^*(M)} &= R \sum_{k=0}^{M-1} \frac{(k+1)(\lambda^*(M))^k}{k!} e^{-\lambda^*(M)} \\ &\quad - R \sum_{k=0}^{M-1} \frac{(\lambda^*(M))^{(k+1)}}{k!} e^{-\lambda^*(M)} \\ &= 0. \end{aligned} \tag{15}$$

Consequently,

$$\sum_{k=0}^{M-1} \frac{(\lambda^*(M))^k}{k!} e^{-\lambda^*(M)} = \frac{(\lambda^*(M))^M}{(M-1)!} e^{-\lambda^*(M)} \tag{16}$$

or

$$\Pr\{X \le M - 1\}|_{\lambda = \lambda^*(M)} = M \Pr\{X = M\}|_{\lambda = \lambda^*(M)}. \tag{17}$$

To prove Theorem 1, we show that $S_\infty^*(M+1)/(M+1) \ge S_\infty^*(M)/M$ for all M in the following:

$$\begin{aligned} S_\infty^*(M+1) &= S_\infty(M+1, \lambda^*(M+1)) \\ &\ge S_\infty(M+1, \lambda^*(M)) \\ &= R \sum_{k=0}^{M-1} \frac{\lambda^*(M)^{k+1}}{k!} e^{-\lambda^*(M)} \\ &\quad + R \frac{\lambda^*(M)^{M+1}}{M!} e^{-\lambda^*(M)} \\ &= S_\infty(M, \lambda^*(M)) + R\lambda^*(M) \Pr\{X = M\}|_{\lambda = \lambda^*(M)} \\ &= \frac{M+1}{M} S_\infty^*(M), \end{aligned} \tag{18}$$

where the last equality is due to (14) and (17). Therefore, we have

$$\frac{S_\infty(M+1)}{M+1} \ge \frac{S_\infty(M)}{M} \ \forall M.$$

□

It is obvious that in a WLAN with MPR capability of M, the maximum possible throughput is MR when there exists a perfect scheduling. In practical random-access WLANs, the actual throughput is always smaller than MR, due to the throughput penalty resulting from packet collisions and idle slots. For example, the maximum throughput is well known to be Re^{-1} when $M = 1$. Theorem 2 proves that the throughput penalty diminishes as M becomes large. That is, the maximum throughput approaches MR even though the channel access is based on random contentions.

Theorem 2 (Asymptotic channel-access efficiency). $\lim_{M\to\infty} S^*_\infty(M)/MR = 1$.

Before proving Theorem 2, we present the following two lemmas:

Lemma 1.

a. $\lim_{M\to\infty} S_\infty(M)/\lambda R = 1$ *for any attempt rate* $\lambda < M$.
b. $\lim_{M\to\infty} S_\infty(M)/\lambda R = 0$ *for any attempt rate* $\lambda > M$.
c. $\lim_{M\to\infty} S_\infty(M)/\lambda R = 0.5$ *for attempt rate* $\lambda = M$.

Proof of Lemma 1a.

$$\begin{aligned} S_\infty(M,\lambda) &= R\lambda \Pr\{X \le M-1\} \\ &= R\lambda\left(1 - \sum_{k=M}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}\right) \\ &\ge R\lambda\left(1 - z^{-M} \sum_{k=M}^{\infty} \frac{(\lambda z)^k}{k!} e^{-\lambda}\right) \\ &\ge R\lambda\left(1 - z^{-M} e^{\lambda(z-1)}\right) \ \forall z > 1. \end{aligned} \tag{19}$$

Let $f(z) = R\lambda\left(1 - z^{-M}e^{\lambda(z-1)}\right)$ be the lower bound of $S_\infty(M)$. By solving

$$\frac{\partial f(x)}{\partial z} = R\lambda\left(Mz^{-M-1}e^{\lambda(z-1)} - \lambda z^{-M}e^{\lambda(z-1)}\right) = 0, \tag{20}$$

it can be easily found that $z^* = M/\lambda$ maximizes $f(z)$ and

$$\frac{f(z^*)}{\lambda R} = 1 - \left(\frac{\lambda}{M}\right)^M e^{M(1-\frac{\lambda}{M})}. \tag{21}$$

Since $z^* > 1$, $\lambda < M$. Let $\lambda = cM$, where $c < 1$. Equation (21) can be written as

$$\frac{f(z^*)}{\lambda R} = 1 - \left(ce^{1-c}\right)^M. \tag{22}$$

It is obvious that

$$ce^{1-c} < 1 \ \forall c \ne 1. \tag{23}$$

Therefore,

$$\begin{aligned} \lim_{M\to\infty} \frac{S_\infty(M,\lambda)}{\lambda R} &\ge \lim_{M\to\infty} \frac{f^*(z)}{\lambda R} \\ &= \lim_{M\to\infty} \left(1 - \left(ce^{1-c}\right)^M\right) \\ &= 1. \end{aligned} \tag{24}$$

On the other hand, the first equality of (19) implies

$$\frac{S_\infty(M,\lambda)}{\lambda R} \le 1. \tag{25}$$

Combining (24) and (25), we have

$$\lim_{M\to\infty} \frac{S_\infty(M,\lambda)}{\lambda R} = 1 \ \forall \lambda < M, \tag{26}$$

and Lemma 1a follows. □

Proof of Lemma 1b.

$$\begin{aligned} S_\infty(M) &= R\lambda \Pr\{X \le M-1\} = R\lambda \sum_{k=0}^{M-1} \frac{\lambda^k}{k!} e^{-\lambda} \\ &\le R\lambda z^{-M} \sum_{k=0}^{M-1} \frac{(\lambda z)^k}{k!} e^{-\lambda} \\ &\le R\lambda z^{-M} \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} e^{-\lambda} \\ &= R\lambda z^{-M} e^{\lambda(z-1)} \ \forall z < 1. \end{aligned} \tag{27}$$

Let $g(z) = R\lambda z^{-M}e^{\lambda(z-1)}$ be the upper bound of $S_\infty(M)$. By solving

$$\frac{\partial g(z)}{\partial z} = R\lambda\left(-Mz^{-M-1}e^{\lambda(z-1)} + \lambda z^{-M}e^{\lambda(z-1)}\right) = 0, \tag{28}$$

it can be easily found that $z^* = M/\lambda$ minimizes $g(z)$ and

$$\frac{g(z^*)}{R\lambda} = \left(\frac{\lambda}{M}\right)^M e^{M(1-\frac{\lambda}{M})}. \tag{29}$$

Since $z^* < 1$, $\lambda > M$. Let $\lambda = cM$, where $c > 1$. Equation (29) can be written as

$$\frac{g(z^*)}{R\lambda} = \left(ce^{1-c}\right)^M. \tag{30}$$

Due to (23),

$$\begin{aligned} \lim_{M\to\infty} \frac{S_\infty(M)}{R\lambda} &\le \lim_{M\to\infty} \frac{g^*(z)}{R\lambda} \\ &= \lim_{M\to\infty} \left(ce^{1-c}\right)^M = 0. \end{aligned} \tag{31}$$

On the other hand, it is obvious that

$$\frac{S_\infty(M)}{R\lambda} \ge 0. \tag{32}$$

Combining (31) and (32), we have

$$\lim_{M\to\infty} \frac{S_\infty(M)}{R\lambda} = 0 \ \forall \lambda > M, \tag{33}$$

and Lemma 1b follows.

Proof of Lemma 1c. To prove Lemma 1c, we note that the median of Poisson distribution is bounded as follows [20], [21]:

$$\lambda - \log 2 \le median \le \lambda + 1/3. \tag{34}$$

When $\lambda = M$ and $M \to \infty$, the median approaches M. According to the first equality of (14),

$$\begin{aligned} \lim_{M\to\infty} \frac{S_\infty(M)}{R\lambda} &= \lim_{M\to\infty} \Pr\{X \le M-1\} \\ &\approx \lim_{M\to\infty} \Pr\{X \le M\} = 0.5. \end{aligned} \tag{35}$$

□

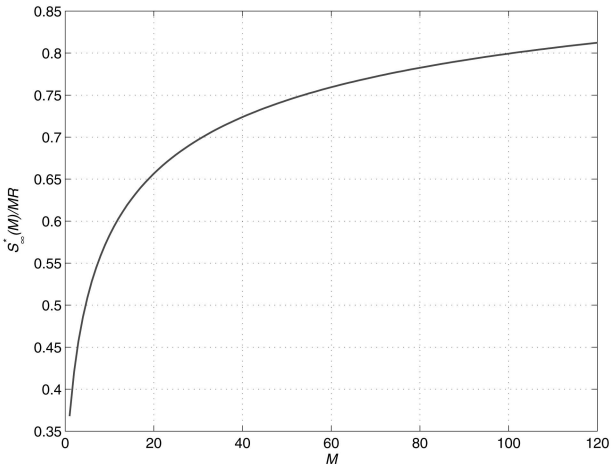


Fig. 1. Superlinear scalability of the throughput of noncarrier-sensing slotted ALOHA networks.

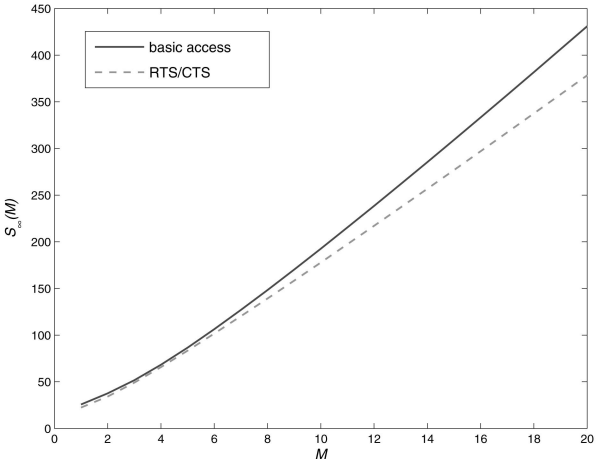


Fig. 2. Optimal throughput of carrier-sensing networks.

Lemma 2. *The optimal attempt rate* $\lambda^*(M) < M$ *and* $\lim_{M\to\infty} \lambda^*(M)/M = 1$.

Proof of Lemma 2. The mode of Poisson distribution is equal to $\lfloor\lambda\rfloor$, where $\lfloor\cdot\rfloor$ denotes the largest integer that is smaller than or equal to the argument. When $\lambda \geq M$,

$$\Pr\{X = M\} > \Pr\{X = i\}\ \forall 0 \leq i \leq M-1, \tag{36}$$

which conflicts with (17). Therefore, the optimal attempt rate is

$$\lambda^*(M) < M. \tag{37}$$

Combining (14), (17), (37), and Lemma 1, we have

$$\lim_{M\to\infty} M \Pr\{X = M\}\big|_{\lambda=\lambda^*(M)} = 1. \tag{38}$$

Let $\lambda^* = cM$, where $c < 1$. Thus, (38) can be written as

$$\lim_{M\to\infty} \frac{(cM)^M}{(M-1)!} e^{-cM} = 1 \tag{39}$$

and

$$\begin{aligned} c &= \lim_{M\to\infty} \frac{\left((M-1)!\right)^{1/M}}{M} e^c \\ &\approx \lim_{M\to\infty} \frac{(M!)^{1/M}}{M} e^c \\ &= e^{-(1-c)}, \end{aligned} \tag{40}$$

where the last equality is due to the Stirling's formula [14]. Solving (40), we have

$$\lim_{M\to\infty} \frac{\lambda^*}{M} = \lim_{M\to\infty} c = 1. \tag{41}$$

□

Proof of Theorem 2. From Lemmas 1 and 2, it is obvious that $\lim_{M\to\infty} S^*_\infty(M)/MR = 1$. □

The above results are illustrated in Fig. 1, where $S^*_\infty(M)/MR$ is plotted as a function of M in noncarrier-sensing slotted ALOHA systems.

Theorem 3 (Superlinearity with finite population). $S^*_N(M+1)/\ M+1 \geq S^*_N(M)/M$ *for all* $M < N$.

Proof of Theorem 3. See Appendix A.

In Theorems 1, 2, and 3, superlinearity is proved assuming that the network is noncarrier-sensing. In Figs. 2 and 3, the optimal throughputs $S^*_\infty(M)$ and $S^*_\infty(M)/M$ are plotted for carrier-sensing networks, respectively, with system parameters listed in Table 1. The figures show that system throughput is greatly enhanced due to the MPR enhancement in the PHY layer. Moreover, the superlinear throughput scaling holds for carrier-sensing networks when M is relatively large.

4 Impact of MPR on EB in WLAN MAC

In this section, we study the characteristic behaviors of WLAN MAC and EB when the channel has MPR capability. We first establish the relationship between transmission probability p_t (or λ) and EB parameters including backoff factor r and minimum contention window W_0. Based on the

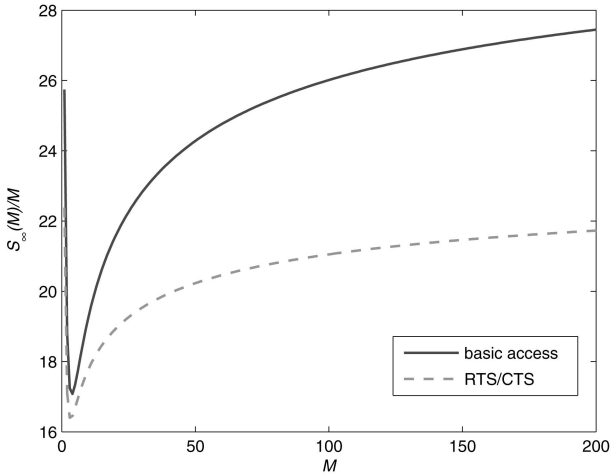


Fig. 3. Superlinear scalability of the throughput of carrier-sensing networks.

TABLE 1
System Parameters Used in Carrier-Sensing Networks
(Adopted from IEEE 802.11g)

Packet payload	8184 bits
MAC header	272 bits
PHY overhead	26 μs
ACK	112 bits + PHY overhead
RTS	160 bits + PHY overhead
CTS	112 bits + PHY overhead
Basic rate	6 Mbps
Data rate	54 Mbps
Slot time σ	9 μs
SIFS	10 μs

analysis, we will then study how the optimal backoff strategy changes with the MPR capability M.

4.1 Transmission Probability

We use an infinite-state Markov chain, as shown in Fig. 4, to model of operation of EB with no retry limit. The reason for the lack of a retry limit is that it is theoretically more interesting to look at the limiting case when the retry limit is infinitely large. Having said this, we note that the analysis in our paper can be easily extended to the case where there is a retry limit. The state in the Markov chain in Fig. 4 is the backoff stage, which is also equal to the number of retransmissions experienced by the station. As mentioned in Section 2, the contention window size is $W_i = r^i W_0$ when a station is in state i. In the figure, p_c denotes the conditional collision probability, which is the probability of a collision seen by a packet being transmitted on the channel. Note that p_c depends on the transmission probabilities of stations other than the transmitting one. In our model, p_c is assumed to be independent of the backoff stage of the transmitting station. In our numerical results, we show that the analytical results obtained under this assumption are very accurate when N is reasonably large.

With EB, transmission probability p_t is equal to the probability that the backoff timer of a station reaches zero in a slot. Note that the Markov process of MPR networks is similar to the ones in [9], [13], except that the conditional collision probability p_c is different for $M > 1$. Therefore, (42) can be derived in a similar way as [9], [13]

$$p_t = \frac{2(1 - rp_c)}{W_0(1 - p_c) + 1 - rp_c}, \tag{42}$$

where $rp_c < 1$ is a necessary condition for the steady state to be reachable. The detailed derivation of (42) is omitted due to page limit. Interested readers are referred to [9], [13].

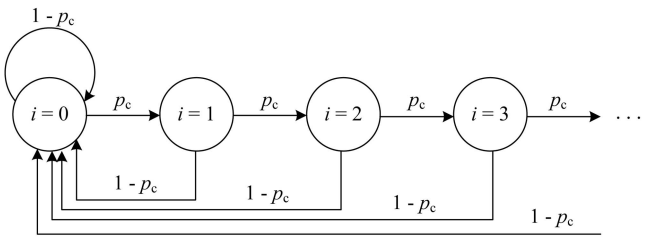


Fig. 4. Markov chain model for the backoff stage.

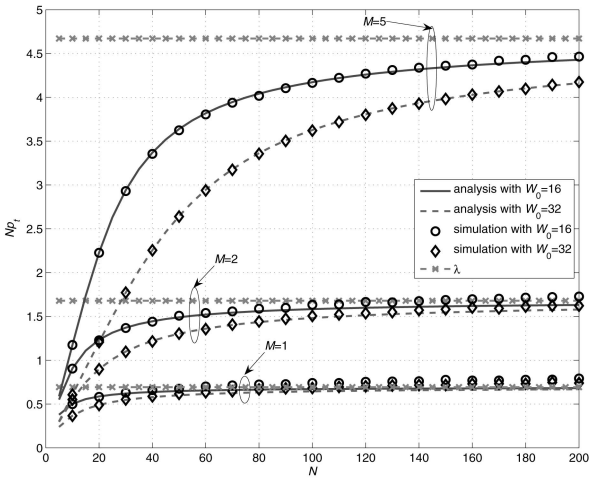


Fig. 5. Plots of Np_t versus N when $r = 2$; lines are analytical results calculated from (2) and (3), markers are simulation results.

Likewise, the conditional collision probability p_c is equal to the probability that there are M or more stations out of the remaining $N - 1$ stations contending for the channel. We thus have the following relationship:

$$p_c = 1 - \sum_{k=0}^{M-1} \binom{N-1}{k} p_t^k (1 - p_t)^{N-k-1}. \tag{43}$$

It can be easily shown that p_t is a decreasing function of p_c for any $r > 1$ in (42). Meanwhile, p_c is an increasing function of p_t in (43). Therefore, the curves determined by (42) and (43) have a unique intersection corresponding to the root of the nonlinear system. By solving the nonlinear system (42) and (43) numerically for different N, we plot the analytical results of Np_t in Fig. 5. In the figures, BEB is adopted. That is, $r = 2$. The minimum contention window size $W_0 = 16$ or 32. To validate the analysis, the simulation results are plotted as markers in the figures. In the simulation, the data are collected by running 5,000,000 rounds after 1,000,000 rounds of warm up. From the figures, we can see that the analytical results match the simulations very well. Moreover, it shows that Np_t converges to a constant quantity when N becomes large. This is a basic assumption in the previous section when we calculated the asymptotic throughput. The constant quantity that Np_t converges to can be calculated as follows.

For large N, the number of attempts in a slot can be modeled as a Poisson process [24, p. 258]. That is,

$$\Pr\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \tag{44}$$

where

$$\lambda = \lim_{N \to \infty} Np_t. \tag{45}$$

The conditional collision probability in this limiting case is given by

$$\lim_{N \to \infty} p_c = \Pr\{X \geq M\} = 1 - \sum_{k=0}^{M-1} \frac{\lambda^k}{k!} e^{-\lambda}. \tag{46}$$

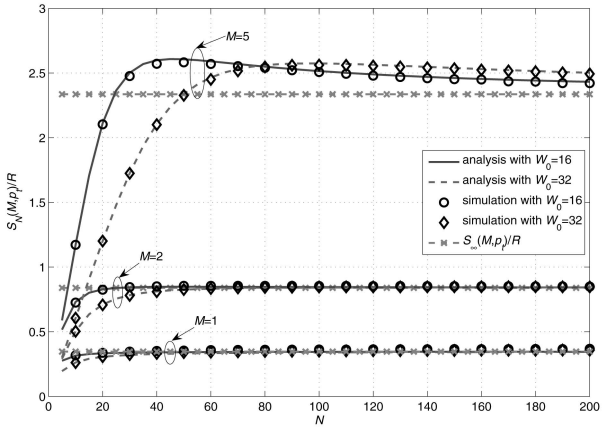


Fig. 6. Normalized throughput of noncarrier-sensing-slotted ALOHA networks when $r = 2$.

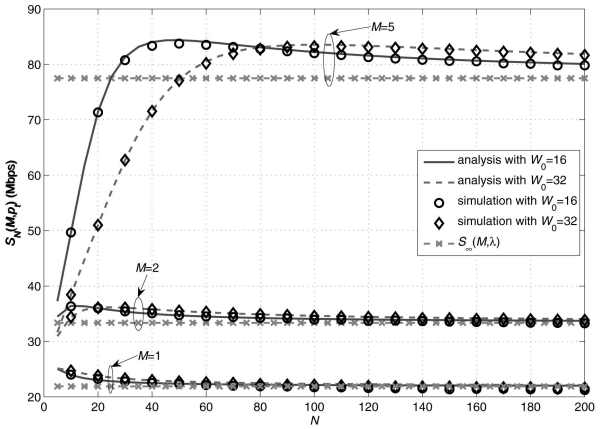


Fig. 7. Throughput of carrier-sensing basic-access networks when $r = 2$.

When the system is steady, the total attempt rate $\lambda = \lim_{N\to\infty} Np_t$ should be finite. Therefore,

$$\lim_{N\to\infty} p_t = \lim_{N\to\infty} \frac{2(1-rp_c)}{W_0(1-p_c)+1-rp_c} = 0, \tag{47}$$

which implies

$$\lim_{N\to\infty} p_c = \frac{1}{r}. \tag{48}$$

Combining (46) and (48), we get the following equation:

$$\sum_{k=0}^{M-1} \frac{\lambda^k}{k!} e^{-\lambda} = 1 - \frac{1}{r}, \tag{49}$$

where λ can be calculated numerically from (49) given M and r. Fig. 5 shows that Np_t calculated from (42) and (43) does converge to λ when N is large.

Note that the relationship between p_t, λ, and EB established above does not depend on the duration of the underlying backoff slots, and therefore can be applied in both noncarrier-sensing and carrier-sensing networks.

Before leaving this section, we validate another assumption adopted in Section 3. That is, EB guarantees a nonzero throughput when N approaches infinity. To this end, the throughput of slotted ALOHA is plotted as a function of N in Fig. 6 when BEB is adopted. It can be seen that the throughputs with the same M converge to the same constant as N increases, regardless of the minimum contention window W_0. Similar phenomenon can also be observed in carrier-sensing networks, as illustrated in Fig. 7, where the throughput of IEEE 802.11 WLAN with basic-access mode is plotted with detailed system parameters listed in Table 1. The asymptotic throughput when N is very large depends only on the MPR capability M and the backoff factor r.

4.2 Optimal Backoff Factor

In Section 3, we have investigated the maximum network throughput that is achieved by optimal transmission probability $p_t^*(M)$ and $\lambda^*(M)$. The previous section shows that transmission probability is a function of backoff factor r. Mathematically, the optimal r that maximizes throughput can be obtained by solving the equation $\partial S(M)/\partial r = 0$.

In this section, we investigate how the optimal backoff factor r changes with the MPR capability M. In Figs. 8 and 9, we plot the throughput as a function of r for both noncarrier sensing networks and carrier-sensing networks in basic-access mode. From the figure, we can see that the optimal r that maximizes throughput increases with M for moderate to large M. This observation can be intuitively explained for noncarrier-sensing networks by (14), (49), and Lemma 1 as follows. Equations (14) and (49) indicate that

$$\frac{S_\infty(M,\lambda)}{R\lambda} = \Pr\{X \le M-1\} = 1 - \frac{1}{r}. \tag{50}$$

As Lemma 1 indicates, when M is large, $S_\infty(M,\lambda)/R\lambda$ increases with M and eventually approaches 1. Consequently, r increases with M.

As the figures show, the throughput decreases sharply when r moves from the optimal r^* to 1. On the other hand, it is much less sensitive to r when r is larger than the r^*. Therefore, in order to avoid dramatic throughput degradation, it is not wise to operate r in the region between 1 and r^*. Note that when M is large, r^* is larger than 2. This implies that the widely used BEB might be far from optimal in MPR WLANs. To further see how well BEB works, we plot the ratio of the

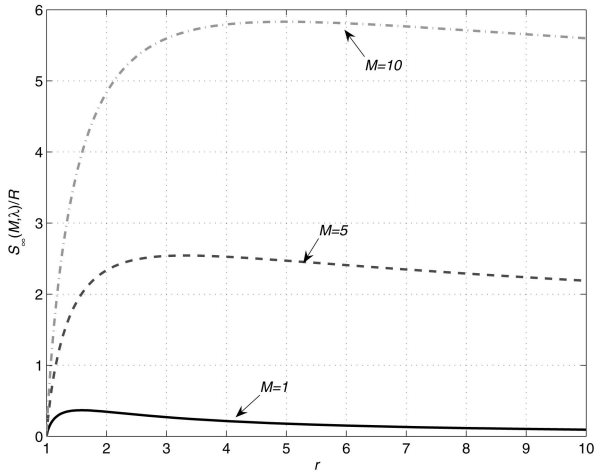


Fig. 8. Throughput versus r for noncarrier-sensing slotted ALOHA networks.

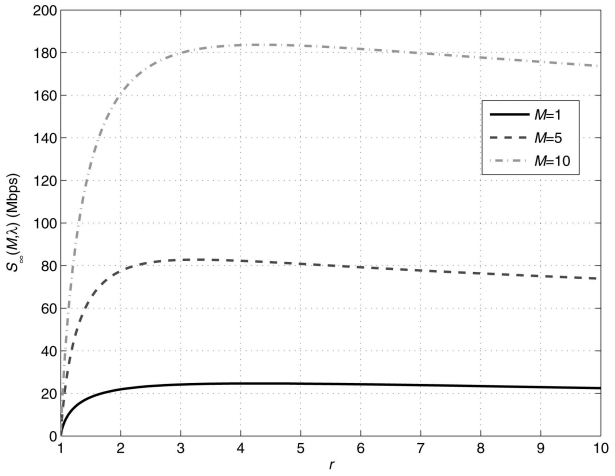


Fig. 9. Throughput versus r for carrier-sensing networks with basic-access mode.

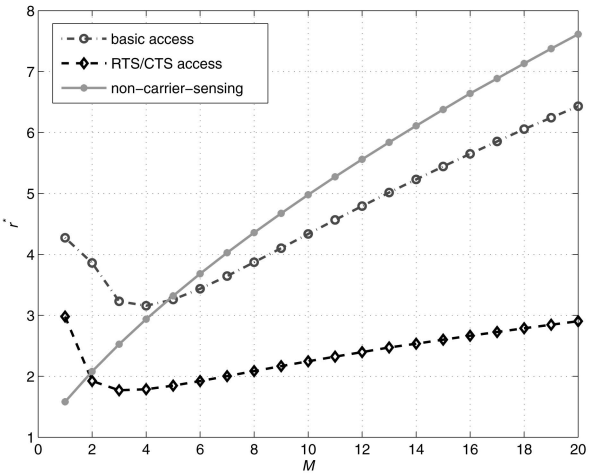


Fig. 11. Optimal r versus M.

throughput obtained by BEB to the maximum achievable throughput in Fig. 10. The optimal r that achieves the maximum throughput is plotted versus M in Fig. 11. In the figures, we can see that BEB only achieves a small fraction of the maximum achievable throughput when M is large in noncarrier-sensing and IEEE 802.11 basic-access mode. For example, when $M = 10$ BEB only achieves about 80 percent of the maximum throughput in noncarrier-sensing networks. In RTS/CTS mode, in contrast, the performance of BEB is close to optimal for a large range of M. Therefore, we argue from an engineering point of view that BEB (i.e., $r = 2$) is a good choice for RTS/CTS access scheme, while on the other hand, tuning r to the optimal is important for noncarrier-sensing and basic-access schemes.

Having demonstrated the significant capacity improvement that MPR brings to WLANs, we are highly motivated to present practical protocols to implement MPR in the widely used IEEE 802.11 WiFi. In particular, we will propose protocols that consist of both MAC-layer mechanisms and PHY-layer signal processing schemes in the next section.

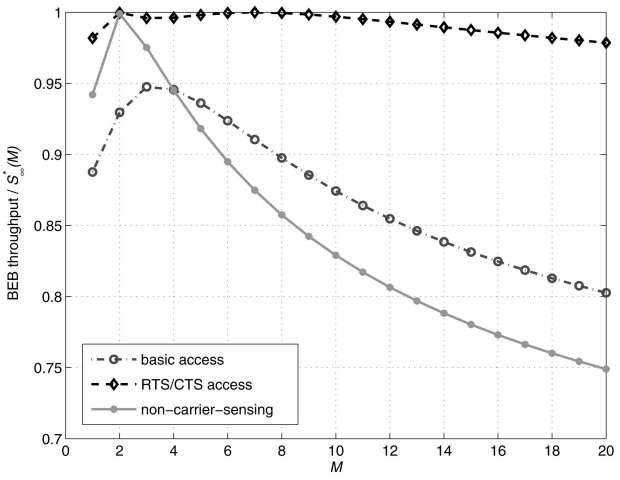


Fig. 10. Ratio of BEB throughput to the maximal throughput versus M.

5 MPR Protocol for IEEE 802.11 WLAN

In this section, we present an MPR protocol for IEEE 802.11 WLAN with RTS/CTS mechanism. The proposed protocol requires minimum amendment at mobile stations, and hence will be easy to implement in practical systems. Throughout this section, we assume that the MPR capability is brought by the multiple antennas mounted at the AP. This assumption complies with the hardware request of the latest MIMO-based WLAN standards. However, the proposed MAC-PHY protocol can be easily extended to CDMA networks, as the received signal structures in multiantenna and CDMA systems are almost the same (refer to Section 2.3).

5.1 MAC Protocol Design

The MAC protocol closely follows the IEEE 802.11 RTS/CTS access mechanism, as illustrated in Fig. 12. A station with a packet to transmit first sends an RTS frame to the AP. In our MPR-MAC model, when multiple stations transmit RTS frames at the same time, the AP can successfully detect all the RTS frames if and only if the number of RTSs is no larger than M. When the number of transmitting stations exceeds M, collisions occur and the AP cannot decode any of the RTSs. The stations will retransmit their RTS frames after a backoff time period according to the original IEEE 802.11 protocol. When the AP detects the RTSs successfully,

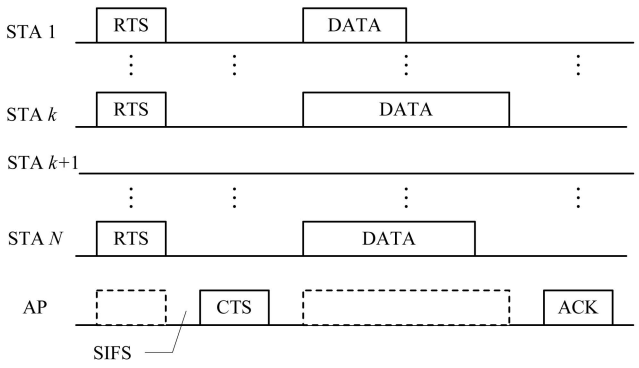


Fig. 12. Time line example for the MPR MAC.

it responds after an SIFS period, with a CTS frame that grants transmission permissions to all the requesting stations. Then, the transmitting stations will start transmitting DATA frames after an SIFS, and the AP will acknowledge the reception of the DATA frames by an ACK frame.

The formats of the RTS and Data frames are the same as those defined in 802.11, while the CTS and ACK frames have been modified to accommodate multiple transmitting stations for MPR. In particular, there are M receiver address fields in the CTS and ACK frames to identify up to M intended recipients.

As described above, our MPR MAC is very similar to the original IEEE 802.11 MAC. In fact, to maintain this similarity in the MAC layer, the challenge is pushed down to the physical layer. For example, in the proposed MPR MAC, the AP is responsible to decode all the RTSs transmitted simultaneously. However, due to the random-access nature of WLAN, the AP has no priori knowledge of who the senders are and the channel state information (CSI) on the corresponding links. This imposes a major challenge on the PHY layer, as the MUD techniques introduced in Section 2, such as ZF and MMSE cannot be directly applied. To tackle these problems, we introduce the physical layer techniques in the next section.

5.2 PHY-Layer Signal Processing Mechanism

In this section, we propose a PHY mechanism to implement MPR in IEEE 802.11. The basic idea is as follows: RTS packets are typically transmitted at a lower data rate than the data packets in IEEE 802.11. This setting is particularly suitable for blind detection schemes which can separate the RTS packets without knowing the prior knowledge of the senders' identities and CSI [17], [18]. Upon successfully decoding the RTS packets, the AP can then identify the senders of the packets. Training sequences, to be transmitted in the preamble of the data packets, are then allocated to these users to facilitate channel estimation during the data transmission phase. Since the multiple stations transmit their data packets at the same time, their training sequences should be mutually orthogonal. In our system, no more than M simultaneous transmissions are allowed. Therefore, a total of M orthogonal sequences are required to be predefined and made known to all stations. The sequence allocation decision is sent to the users via the CTS packet.

During the data transmission phase, CSI is estimated from the orthogonal training sequences that are transmitted in the preamble of the data packets. With the estimated CSI, various MUD techniques can be applied to separate the multiple data packets at the AP. Using coherent detection, data packets can be transmitted at a much higher rate than the RTS packets without involving excessive computational complexity.

As MUD techniques have been introduced in Section 2, we focus on the blind separation of RTS packets in this section. Assume that there are K stations transmitting RTS packets together. Then, the received signal in symbol duration n is given by (4), where the $(m,k)\text{th}$ element of $\mathbf{H}$ denotes the channel coefficient from user k to the $m\text{th}$ antenna at the AP. Assuming that the channel is constant over an RTS packet, which is composed of N symbol periods, we obtain the following block formulation of the data:

$$\mathbf{Y} = \mathbf{HX} + \mathbf{W}, \tag{51}$$

where $\mathbf{Y} = [\mathbf{y}(1), \mathbf{y}(2), \ldots, \mathbf{y}(N)]$, $\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(N)]$, and $\mathbf{W} = [\mathbf{w}(1), \mathbf{w}(2), \ldots, \mathbf{w}(N)]$. The problem to be addressed here is the estimation of the number of sources K, the channel matrix $\mathbf{H}$, and the symbol matrix $\mathbf{X}$, given the array output $\mathbf{Y}$.

5.2.1 Estimation of the Number of Sources K

For an easy start, we ignore the white noise for the moment and have $\mathbf{Y} = \mathbf{HX}$. The rank of $\mathbf{H}$ is equal to K if $K < M$. Likewise, $\mathbf{X}$ is full-row-rank when N is much larger than K. Consequently, we have $rank(\mathbf{Y}) = K$ and K is equal to the number of nonzero singular values of $\mathbf{Y}$. With white noise added to the data, K can be estimated from the number of singular values of $\mathbf{Y}$ that are significantly larger than zero.

5.2.2 Estimation of X and H

In this paper, we adopt the Finite Alphabet (FA)-based blind detection algorithm to estimate $\mathbf{X}$ and $\mathbf{H}$, assuming that K is known. The ML estimator yields the following separable least-squares minimization problem [17]:

$$\min_{\mathbf{H},\mathbf{X}\in\Omega} \|\mathbf{Y} - \mathbf{HX}\|_F^2, \tag{52}$$

where Ω is the finite alphabet to which the elements of $\mathbf{X}$ belong and $\|\cdot\|_F^2$ is the Frobenius norm. The minimization of (52) can be carried out in two steps. First, we minimize (52) with respect to $\mathbf{H}$ and obtain

$$\hat{\mathbf{H}} = \mathbf{YX}^+ = \mathbf{YX}^H(\mathbf{XX}^H)^{-1}, \tag{53}$$

where $(\cdot)^+$ is the pseudoinverse of a matrix. Substituting $\hat{\mathbf{H}}$ back into (52), we obtain a new criterion, which is a function of $\mathbf{X}$ only

$$\min_{\mathbf{X}\in\Omega} \|\mathbf{YP}_{\mathbf{X}^H}^{\perp}\|_F^2, \tag{54}$$

where $\mathbf{P}_{\mathbf{X}^H}^{\perp} = \mathbf{I} - \mathbf{X}^H(\mathbf{XX}^H)^{-1}\mathbf{X}$ and $\mathbf{I}$ is the identity matrix. The global minimum of (54) can be obtained by enumerating over all possible choices of $\mathbf{X}$. Reduced-complexity iterative algorithms that solve (54) iteratively, such as ILSP and ILSE, were introduced in [18]. Not being one of the foci of this paper, the details of ILSP and ILSE are not covered here. Interested readers are referred to [19] and the references therein.

Note that the scheme proposed in this section is only one way of implementing MPR in WLANs. It ensures that the orthogonal training sequences are transmitted in the preambles of data packets. This leads to highly reliable channel estimation that facilitates the user of MUD techniques. Moreover, the modification to the original protocol is mainly restrained within the AP. Minimum amendment is needed at mobile stations.

6 Discussions

6.1 Random Channel Error

In our analysis so far, we have assumed that packet error rate due to random fading effect is negligible when the number of simultaneous transmission is smaller than M and is close to 1 otherwise. This assumption is quite accurate when data packets are well protected by error

correction codes (e.g., convolutional codes in IEEE 802.11 protocol) and linear MUD is deployed at the receiver. The simplification allows us to focus on the effect of MPR on WLAN without the need to consider signal processing details such as coding and detection schemes.

In this section, we relax the assumption and investigate how random channel errors would affect our analysis. Fortunately, we can prove that superlinear throughput scaling still holds even when random channel error is taken into account, as detailed in the following. Denote by $P_M^{err}(k)$ the packet error rate due to wireless channel fading when k packets are transmitted at the same time in a network with MPR capability M. Then, $P_M(k) = 1 - P_M^{err}(k)$ is the packet success rate, which is the probability that a packet *survives* random channel fading [25]. Typically, $P_M(k) \geq P_M(k')$ for $k \leq k'$ and $P_M(k) \geq P_{M'}(k)$ for $M \geq M'$. Assuming linear detectors, we have $P_M(k) \approx 0$ if $k > M$ and $P_M(M) \approx P_{M'}(M')$ for $M \neq M'$ [22].

For simplicity, assume that $T_{slot} = T_s = T_i = T_c = L/R$. Then, asymptotic throughput is given by

$$\begin{aligned} S_\infty(M,\lambda) &= R\sum_{k=1}^{M} k\frac{\lambda^k}{k!}e^{-\lambda}P_M(k) \\ &= R\sum_{k=0}^{M-1}\frac{\lambda^{k+1}}{k!}e^{-\lambda}P_M(k+1). \end{aligned} \tag{55}$$

At the optimal $\lambda^*(M)$, $\frac{\partial S_\infty(M,\lambda)}{\partial \lambda} = 0$. Consequently,

$$\begin{aligned} &\sum_{k=0}^{M-1}\frac{(\lambda^*(M))^k}{k!}e^{-\lambda^*(M)}P_M(k+1) \\ &= \sum_{k=0}^{M-2}\frac{(\lambda^*(M))^{k+1}}{k!}e^{-\lambda^*(M)}(P_M(k+1) - P_M(k+2)) \\ &\quad + \frac{(\lambda^*(M))^M}{(M-1)!}e^{-\lambda^*(M)}P_M(M) \\ &\leq \frac{(\lambda^*(M))^M}{(M-1)!}e^{-\lambda^*(M)}P_M(M). \end{aligned} \tag{56}$$

We are now ready to prove superlinear throughput scaling $\frac{S_\infty^*(M+1)}{M+1} \geq \frac{S_\infty^*(M)}{M}$ in the following:

$$\begin{aligned} S_\infty^*(M+1) &\geq S_\infty(M+1, \lambda^*(M)) \\ &= R\sum_{k=0}^{M-1}\frac{\lambda^*(M)^{k+1}}{k!}e^{-\lambda^*(M)}P_{M+1}(k+1) \\ &\quad + R\frac{\lambda^*(M)^{M+1}}{M!}e^{-\lambda^*(M)}P_{M+1}(M+1) \\ &\geq S_\infty(M,\lambda^*(M)) + R\frac{\lambda^*(M)^{M+1}}{M!}e^{-\lambda^*(M)}P_M(M) \\ &\geq \frac{M+1}{M}S_\infty^*(M), \end{aligned} \tag{57}$$

where the last inequality is due to (56).

6.2 Near-Far Effect

One implicit assumption in our analysis is that each station transmits at the same data rate R. In practice, stations experience different channel attenuation to the AP due to their random locations. If stations transmit at the same power level, then the data rate sustainable on each link would differ. In this case, the airtime occupied by a busy period is dominated the lowest data rate involved. Hence, the effective throughput enjoyed by high-rate stations would degenerate to the level of the lowest rate. Such problem, known as "performance anomaly," is not unique to MPR. It exists in all multirate IEEE 802.11 networks. Fortunately, performance anomaly only causes the data rate R in our throughput expression to degrade to R_{min}, where R_{min} is the lowest possible data rate. Therefore, it will not affect the scaling law of throughput in MPR networks.

6.3 Comparison with Multiuser SIMO Systems

In this paper, we have demonstrated the drastic increase in spectrum efficiency brought by MPR. To implement MPR, modification is needed in both MAC and PHY layers, as discussed in Section 5. With the same hardware enhancement (e.g., having M antennas at the AP), an alternative is to let each link transmit at a higher data rate, but keep the single-packet reception restriction unchanged. This essentially becomes a traditional WLAN with single-input-multiple-output (SIMO) links.

The capacity of an SIMO link increases logarithmically with the number of antennas at the receiver [23], that is,

$$R_{SIMO} \approx R_{SISO} + \log(M), \tag{58}$$

where R_{SISO} is the data rate of a single-input-single-output (SISO) link. In contrast, the data rate R of each link in MPR WLAN is set to R_{SISO}, for antenna, diversity is used to separate multiple data streams instead of increasing the rate of one stream therein.

With (58), the throughput of WLAN with SIMO links is

$$S_N^{SIMO} = \frac{L\sum_{k=1}^{M} k\Pr\{X=k\}}{P_{idle}^{SIMO}T_i^{SIMO} + P_{coll}^{SIMO}T_c^{SIMO} + P_{succ}^{SIMO}T_s^{SIMO}}, \tag{59}$$

where the expressions for P_{idle}^{SIMO}, P_{coll}^{SIMO}, and P_{succ}^{SIMO} are the same as (9), (10), and (11) with $M = 1$, respectively. Likewise, T_i^{SIMO}, T_c^{SIMO}, and T_s^{SIMO} are the same as (1), (2), or (3) except that R is replaced by R_{SIMO}. Specifically, throughput in the ALOHA case becomes

$$S_N^{SIMO} = (R + \log(M))Np_t(1-p_t)^{N-1}, \tag{60}$$

and the optimal p_t that maximizes the throughput is equal to $1/N$. In particular, the maximum achievable throughput when N is large is

$$S_\infty^{SIMO*}(M) = (R + \log(M))e^{-1}. \tag{61}$$

It is obvious that the normalized throughput $\frac{S_\infty^{SIMO*}(M)}{M}$ decreases with M in SIMO networks. This, in contrast to the superlinear throughput scaling in MPR networks, suggests that multiple antennas at the AP should be used to resolve simultaneous transmissions instead of increasing per-link data rate in random access WLANs.

7 Conclusion

With the recent advances in PHY-layer MUD techniques, it is no longer a physical constraint for the WLAN channel to

accommodate only one packet transmission at one time. To fully utilize the MPR capability of the PHY channel, it is essential to understand the fundamental impact of MPR on the MAC layer. This paper has studied the characteristic behavior of random-access WLANs with MPR. Our analysis provides a theoretical foundation for the performance evaluation of WLANs with MPR, and it is useful for system design in terms of setting operating parameters of MAC protocols.

Our analytical framework is general and applies to various WLANs including non-carrier-sensing and carrier-sensing networks. In Theorems 1 and 3, we have proved that the throughput increases superlinearly with M for both finite and infinite-population cases. This is the case in noncarrier-sensing networks for all M and in carrier-sensing networks for moderate to large M. Moreover, Theorem 2 shows that the throughput penalty due to distributed random access diminishes when M approaches infinity. Such scalability provides strong incentives for further investigations on engineering and implementation details of MPR systems. Based on the analysis, we found that the commonly deployed BEB scheme is far from optimum in most systems except the carrier-sensing systems with RTS/CTS four-way handshake. In particular, the optimum backoff factor r increases with M for large M. We further note that the throughput degrades sharply when r is smaller than the optimum value, while it is much less sensitive to r when r exceeds the optimum.

Having understood the fundamental behavior of MPR, we propose practical protocols to exploit the advantage of MPR in IEEE 802.11-like WLANs. By incorporating advanced PHY-layer blind detection and MUD techniques, the protocol can implement MPR in a fully distributed manner with marginal modification of MAC layer.

APPENDIX A

SUPERLINEAR THROUGHPUT SCALING IN WLANS WITH FINITE POPULATION

Theorem 4 (Superlinearity with finite population). *$S_N^*(M+1)/(M+1) \geq S_N^*(M)/M$ for all $M < N$.*

From (12), we have

$$\begin{aligned} S_N(M,p_t) &= R\sum_{k=1}^{M} k\binom{N}{k}p_t^k(1-p_t)^{N-k} \\ &= R\frac{Np_t}{1-p_t}\sum_{k=0}^{M-1}\binom{N}{k}p_t^k(1-p_t)^{N-k} \\ &\quad - R\frac{p_t}{1-p_t}\sum_{k=0}^{M-1} k\binom{N}{k}p_t^k(1-p_t)^{N-k} \\ &= R\frac{Np_t}{1-p_t}\Pr\{X\leq M-1\} \\ &\quad -\frac{p_t}{1-p_t}S_N(M-1,p_t), \end{aligned} \tag{62}$$

and

$$S_N(M+1,p_t)=R\frac{Np_t}{1-p_t}\Pr\{X\leq M\}-\frac{p_t}{1-p_t}S_N(M,p_t). \tag{63}$$

Meanwhile,

$$\begin{aligned} S_N(M+1,p_t) &= R\sum_{k=1}^{M+1} k\binom{N}{k}p_t^k(1-p_t)^{N-k} \\ &= S_N(M,p_t)+R(M+1)\Pr\{X=M+1\}. \end{aligned} \tag{64}$$

Substituting (64) to (63), we get

$$\begin{aligned} S_N(M,p_t) &= RNp_t\Pr\{X\leq M\} \\ &\quad - R(1-p_t)(M+1)\Pr\{X=M+1\} \\ &\quad \forall M<N, p_t. \end{aligned} \tag{65}$$

At the optimal $p_t^*(M)$, the derivative $\partial S_N(M,p_t)/\partial p_t = 0$. Thus,

$$\begin{aligned} &\frac{\partial S_N(M,p_t)}{\partial p_t}\bigg|_{p_t=p_t^*(M)} = RN\Pr\{X\leq M\}\big|_{p_t=p_t^*(M)} \\ &+R(M+1)\left(1-\frac{M+1}{p_t}\right)\Pr\{X=M+1\}\big|_{p_t=p_t^*(M)}=0. \end{aligned} \tag{66}$$

Combining (65) and (66), we have

$$\begin{aligned} S_N(M,p_t^*(M)) &= p_t^*(M)\frac{\partial S_N(M,p_t)}{\partial p_t}\bigg|_{p_t=p_t^*(M)} \\ &\quad - R(M+1)\big(p_t^*(M)-(M+1)\big)\Pr\{X=M+1\}\big|_{p_t=p_t^*(M)} \\ &\quad - R(M+1)\big(1-p_t^*(M)\big)\Pr\{X=M+1\} \\ &= RM(M+1)\Pr\{X=M+1\}\big|_{p_t=p_t^*(M).} \end{aligned} \tag{67}$$

It is obvious that

$$\begin{aligned} &S_N(M+1,p_t^*(M+1))\geq S_N(M+1,p_t^*(M)) \\ &\quad = S_N(M,p_t^*(M))+R(M+1)\Pr\{X=M+1\}\big|_{p_t^*(M)}. \end{aligned} \tag{68}$$

Substituting (67) to (68), we have

$$\begin{aligned} &S_N(M+1,p_t^*(M+1)) \\ &\quad \geq S_N(M,p_t^*(M))+\frac{S_N(M,p_t^*(M))}{M} \\ &\quad = S_N(M,p_t^*(M))\frac{M+1}{M}. \end{aligned} \tag{69}$$

Hence, $S_N^*(M+1)/(M+1) \geq S_N^*(M)/M$ for all $M < N$. □

ACKNOWLEDGMENTS

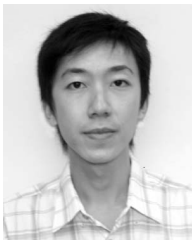
This work was supported in part by the Competitive Earmarked Research Grant (Project numbers 418506, 418707, and 414106) established under the University Grant Committee of Hong Kong. Peng Xuan Zheng was with the Department of Information Engineering, The Chinese University of Hong Kong, when this work was completed.

REFERENCES

[1] S. Verdu, *Multiuser Detection.* Cambridge Univ. Press, 1998.
[2] P.B. Rapajic and D.K. Borah, "Adaptive MMSE Maximum Likelihood CDMA Multiuser Detection," *IEEE J. Selected Areas Comm.*, vol. 17, no. 12, pp. 2110-2122, Dec. 1999.
[3] I.E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585-595, Nov. 1999.

[4] S. Ghez, S. Verdu, S. Schwartz, "Stability Properties of Slotted ALOHA with Multipacket Reception Capability," *IEEE Trans. Automatic Control,* vol. 33, no. 7, pp. 640-649, July 1988.
[5] S. Ghez, S. Verdu, S. Schwartz, "Optimal Decentralized Control in the Random Access Multipacket Channel," *IEEE Trans. Automatic Control,* vol. 34, no. 11, pp. 1153-1163, Nov. 1989.
[6] V. Naware, G. Mergen, and L. Tong, "Stability and Delay of Finite-User Slotted ALOHA with Multipacket Reception," *IEEE Trans. Information Theory,* vol. 51, no. 7, pp. 2636-2656, July 2005.
[7] Q. Zhao and L. Tong, "A Multiqueue Service Room MAC Protocol for Wireless Networks with Multipacket Reception," *IEEE/ACM Trans. Networking,* vol. 11, no. 1, pp. 125-137, Feb. 2003.
[8] Q. Zhao and L. Tong, "A Dynamic Queue Protocol for Multi-Access Wireless Networks with Multipacket Reception," *IEEE Trans. Wireless Comm.,* vol. 3, no. 6, pp. 2221-2231, Nov. 2004.
[9] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE J. Selected Areas Comm.,* vol. 18, no. 3, pp. 535-547, Mar. 2000.
[10] J. Goodman, A.G. Greenberg, N. Madras, and P. March, "Stability of Binary Exponential Backoff," *J. ACM,* vol. 35, no. 3, pp. 579-602, 1988.
[11] H. Al Ammal, L.A. Goldberg, and P. MacKenzie, "An Improved Stability Bound for Binary Exponential Backoff," *Theory of Computing Systems,* vol. 30, pp. 229-244, 2001.
[12] D.G. Jeong and W.S. Jeon, "Performance of an Exponential Backoff Scheme for Slotted-ALOHA Protocol in Local Wireless Environment," *IEEE Trans. Veichular Technology,* vol. 44, no. 3, pp. 470-479, Aug. 1995.
[13] B.J. Kwak, N.O. Song, and L.E. Miller, "Performance Analysis of Exponential Backoff," *IEEE/ACM Trans. Networking,* vol. 13, no. 2, pp. 343-355, Apr. 2005.
[14] W. Feller, "Stirling's Formula," *An Introduction to Probability Theory and Its Applications,* vol. 1, third ed., pp. 50-53, Willey, 1968.
[15] *ANSI/IEEE Std 802.3-1985, IEEE Standards for Local Area Networks: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications,* IEEE, 1985.
[16] *IEEE Std 802.11-1997, Local and Metropolitan Area Networks-Specific Requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,* IEEE, Nov. 1997.
[17] S. Talwar, M. Viberg, and A. Paulraj, "Blind Separation of Synchronous Co-Channel Digital Signals Using an Antenna Array, Part I: Algorithms," *IEEE Trans. Signal Processing,* vol. 44, no. 5, pp. 1184-1197, May 1996.
[18] C.B. Papadias and A.J. Paulraj, "A Constant Modulus Algorithm for Multiuser Signal Separation in Presence of Delay Spread Using Antenna Arrays," *IEEE Signal Processing Letters,* vol. 4, no. 6, pp. 178-181, June 1997.
[19] S. Talwar, M. Viberg, and A. Paulraj, "Blind Estimation of Multiple Co-Channel Digital Signals using an Antenna Array," *IEEE Signal Processing Letters,* vol. 2, no. 1, pp. 29-31, Feb. 1994.
[20] K.P. Choi, "On the Medians of Gamma Distributions and an Equation of Ramanujan," *Proc. Am. Math. Soc.,* vol. 121, no. 1, pp. 245-251, May 1994.
[21] J. Chen and H. Rubin, "Bounds for the Difference Between Median and Mean of Gamma and Poisson Distributions," *Statistics and Probability Letters,* vol. 4, pp. 281-283, 1986.
[22] J.H. Winters, J. Salz, and R.D. Gitlin, "The Impact of Antenna Diversity on the Capacity of Wireless Communication Systems," *IEEE Trans. Comm.,* vol. 42, nos. 2-4, pp. 1740-1751, Feb. 1994.
[23] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications.* Cambridge Univ. Press, 2005.
[24] M.H. DeGroot and M.J. Schervish, *Probability and Statistics,* third ed. Addison Wesley, 2002.
[25] J. Proakis, *Digital Communications,* fourth ed. McGraw-Hill, 2000.
[26] D.S. Chan, T. Berger, and L. Tong, "On the Stability and Optimal Decentralized Throughput of CSMA with Multipacket Reception Capability," *Proc. Allerton Conf. Comm., Control, and Computing,* Sept.-Oct. 2004.
[27] D.S. Chan, P. Suksompong, J. Chen, and T. Berger, "Improving IEEE 802.11 Performance with Cross-Layer Design and Multipacket Reception via Multiuser Iterative Decoding," IEEE 802.11-05/0946r0, Sept. 2005.

Ying Jun (Angela) Zhang received the BEng degree (with honors) in electronic engineering from Fudan University, Shanghai, China, and the PhD degree in electrical and electronic engineering from the Hong Kong University of Science and Technology. Since January 2005, she has been with the Department of Information Engineering, Chinese University of Hong Kong, where she is currently an assistant professor. She is on the editorial boards of the *IEEE Transactions on Wireless Communications* and *Security and Communication Networks*. She served as a technical program committee cochair of the Communication Theory Symposium at IEEE ICC 2009, a track chair of ICCCN 2007, and a publicity chair of IEEE MASS 2007. She has served as a technical program committee member for leading conferences including IEEE ICC, IEEE GLOBECOM, IEEE WCNC, IEEE ICCCAS, IWCMC, IEEE CCNC, IEEE ITW, IEEE MASS, MSN, ChinaCom, etc. She is an IEEE Technical Activity Board GOLD Representative, a 2008 IEEE GOLD Technical Conference Program Leader, an IEEE Communication Society GOLD Coordinator, and a member of the IEEE Communication Society Member Relations Council. Her research interests include wireless communications and mobile networks, adaptive resource allocation, optimization in wireless networks, wireless LAN/MAN, broadband OFDM and multicarrier techniques, and MIMO signal processing. As the only winner from engineering science, she won the Hong Kong Young Scientist Award in 2006, conferred by the Hong Kong Institution of Science. She is a member of the IEEE.



Peng Xuan Zheng received the BE degree in computer science and engineering from Zhejiang University, Hangzhou, China, in 2004, and the MPhil degree in information engineering from the Chinese University of Hong Kong in 2007. He is currently working toward the PhD degree at Purdue University. His research interests include mobile ad hoc networks and wireless sensor networks. He is a student member of the IEEE.

Soung Chang Liew received the SB, SM, EE, and PhD degrees from the Massachusetts Institute of Technology (MIT). From 1984 to 1988, he was at the MIT Laboratory for Information and Decision Systems, where he investigated fiber-optic communications networks. From March 1988 to July 1993, he was at Bellcore (now Telcordia), New Jersey, where he engaged in broadband network research. He is currently a professor and the chairman of the Department of Information Engineering, The Chinese University of Hong Kong. He is also an adjunct professor at Southeast University, China. His current research interests include wireless networks, Internet protocols, multimedia communications, and packet switch design. Both he and his student won best paper awards at the First IEEE International Conference on Mobile Ad Hoc and Sensor Systems (IEEE MASS 2004) and the Fourth IEEE International Workshop on Wireless Local Network (IEEE WLN 2004). Separately, TCP Veno, a version of TCP to improve its performance over wireless networks proposed by him and his student, has been incorporated into a recent release of Linux OS. In addition, he initiated and built the first inter-university ATM network tested in Hong Kong in 1993. In addition to academic activities, he is also active in the industry. He cofounded two technology start-ups in Internet software and has served as a consultant to many companies and industrial organizations. He is currently a consultant for the Hong Kong Applied Science and Technology Research Institute (ASTRI), providing technical advice as well as helping to formulate R&D directions and strategies in the areas of wireless internetworking, applications, and services. He is a holder of four US patents, a fellow of the IEE and HKIE, and a senior member of the IEEE. He is listed in *Marquis Who's Who in Science and Engineering*. He is the recipient of the first Vice-Chancellor Exemplary Teaching Award from the Chinese University of Hong Kong. His publications can be found at www.ie.cuhk.edu.hk/soung.