Distributed Multiuser Scheduling for Improving Throughput of Wireless LAN

Suhua Tang, *Member, IEEE*

***Abstract*—In wireless LANs, the performance of CSMA/CA might be degraded by several problems: (i) severe collisions in the uplink, (ii) head-of-line problem caused by fading in the downlink, and (iii) serious unfairness between uplink and downlink. In this paper, a distributed multiuser scheduling (DMUS) scheme is proposed to simultaneously address these problems. In DMUS, a node (i) computes its normalized SNR (signal to noise ratio) as the ratio of its instantaneous SNR to its average SNR, and (ii) contends via a contention window (CW) for the channel to initiate its uplink or downlink transmission when its normalized SNR is greater than a threshold. The contribution is threefold: (i) All three problems are solved in a unified framework by applying multiuser diversity in both uplink and downlink. Fresh SNR is exploited for distributed scheduling meanwhile airtime fairness is retained. (ii) SNR threshold and CW are jointly optimized to maximize throughput, taking into account time-variant link quality, collision probability and protocol overhead. (iii) Network performance is theoretically analyzed. Extensive simulations confirm that DMUS greatly improves total throughput under almost all scenarios compared with both the contention-based CSMA/CA scheme and the contention-free PCF scheme.**

***Index Terms*—CSMA/CA, multiuser diversity, multipath fading, airtime fairness, distributed scheduling, cross-layer design.**

I. Introduction

DATA traffic explodes on cellular networks with the fast growth of the smartphone market. Recently, the concept of mobile data offloading [1] was suggested to solve this problem. Specifically, most non-quality of service traffic can be offloaded from cellular networks to wireless LANs (WLANs), and its effect depends on the performance of WLANs.

Communications in WLANs take place between nodes and their associated access points (APs). There are three traffic patterns with different problems, as follows:

(i) **Uplink and its collision problem**. In the uplink, all nodes share a common wireless channel by the carrier sense multiple access (CSMA) scheme [2], which is further enhanced with collision avoidance (CA). Lack of a timely acknowledgement (ACK) after transmitting a DATA frame is regarded as a collision. Collision probability is alleviated by binary exponential backoff (BEB) and virtual carrier sense. CSMA works fairly well in times of light traffic. However, it is not well scalable with respect to the number of nodes. Frequent collisions caused by many contending nodes unavoidably degrade the performance of the whole network. Conventional schemes for reducing collision probability include exploiting multiple channels, directional antenna, transmit power control, adaptive backoff, carrier sense and transmit scheduling [3], [4]. A different trend is collision detection [5], which is exploited to diagnose packet losses in rate adaptation schemes [6]. Based on the new function of full-duplex wireless communications [7], collision notification [8] is also explored to quickly resolve a collision. Despite these efforts, controlling collision probability is still necessary since each collision inevitably leads to a waste of airtime.

(ii) **Downlink and its head of line (HOL) problem**. In the downlink, an AP forwards packets received from the Internet to its nodes. There is no contention because the AP is the only transmitter. But transmissions to a node in deep fading may fail. During the backoff and retransmission for a packet, packets destined to other nodes are blocked in the queue, resulting in the HOL problem [9]. Although rate adaptation [10] can partially alleviate this problem, transmissions at a low rate degrade spectrum efficiency.

(iii) **Hybrid mode and its unfairness problem**. In the hybrid mode, an AP has to contend for the channel for the downlink transmissions when other nodes compete to send in the uplink. Purely relying on CSMA/CA for channel access, the downlink transmissions might be starved by the uplink ones, which results in the unfairness problem. One solution is to allocate different resources for uplink and downlink by adjusting TXOP (transmit opportunity) [11].

Conventional methods separately address these problems and have limited effects. A new trend is to exploit cross layer design, and its usage in exploiting multiuser diversity (MUD) ([12], Chapter 6) has been proven to be effective in solving the HOL problem in cellular networks. There, a base station keeps monitoring signal to noise ratio (SNR) of each node and performs a centralized MUD scheduling. In decentralized networks without instantaneous feedback of SNR, MUD is usually realized by SNR-based distributed contention. There are some theoretical analyses of MUD for ALOHA-based uplink [13], [14] and protocol design for CSMA networks [15], [16], [17]. But these designs either are based on the splitting algorithm or require using multiple SNR thresholds for a node, which make a protocol complex. Moreover, fairness remains an issue when exploiting SNR for distributed scheduling. MUD is also applied in the downlink [18] to solve the HOL problem of WLANs. However, the rough link quality estimated from the statistics of packet loss constrains its performance. In addition, there is still no complete solution for all three problems and the protocol overhead is seldom considered.

In this paper, we propose a distributed multiuser scheduling (DMUS) scheme to simultaneously address the aforemen-

Manuscript received April 23, 2013; revised September 14 and December 16, 2013; accepted February 9, 2014. The associate editor coordinating the review of this paper and approving it for publication was A. Vosoughi.

S. Tang is with Adaptive Communications Research Laboratories, ATR, Japan (e-mail: suhua_tang@ieee.org).

Digital Object Identifier 10.1109/TWC.2013.040214.130707

1536-1276/14$31.00 © 2014 IEEE

tioned problems of WLANs. We have investigated the first problem in our previous work [19], minimizing collision probability by setting an SNR threshold. In this work, we further solve the three problems in a unified framework to realize a proportional fair scheduling [20] in a distributed way, so as to both improve total throughput and ensure airtime fairness. Generally, each node contends for the channel when its link quality is near its own peak. Specifically, each node computes its *normalized SNR*—the ratio of its instantaneous SNR to its average SNR, and contends for the channel via slotted contention when its normalized SNR is above the specified threshold. Then, the winning node initiates a data transmission, either in the uplink to the AP or in the downlink from the AP. Multiple packets can be transmitted in a burst at a high rate so as to reduce protocol overhead.

DMUS solves the three problems of WLANs as follows. (a) The performance of the multiple access channel is optimized via two stages: (i) reducing the number of nodes involved in a contention and improving transmit rate, both by setting a normalized SNR threshold, and (ii) using a contention window (CW) to mitigate collisions among the few nodes whose normalized SNR is above the threshold. (b) In the downlink, though data packets flow from an AP to nodes, actual transmissions are initiated in a distributed way by nodes instead of an AP. A transmission in the downlink is the same as in the uplink except two differences: (i) A node must first detect the presence of its downlink traffic via the notification from the AP. (ii) A node winning the channel contention first sends an invitation message to notify the AP that the node is ready to receive its downlink packets. By letting nodes, whose normalized SNR is above the threshold, contend to initiate their downlink transmissions, MUD is realized in the downlink and the HOL problem is avoided. (c) In the hybrid mode, nodes with downlink traffic directly contend with nodes with uplink traffic instead of relying on the AP to perform downlink scheduling. The contention among nodes in terms of CSMA/CA ensures fair chance and removes the unfairness between uplink and downlink.

Main contributions of this paper are threefold, as follows:

- A unified framework is suggested for applying MUD in both uplink and downlink of WLANs, exploiting fresh SNR for distributed scheduling. Besides controlling collision probability and avoiding fading, airtime fairness is also achieved.
- SNR threshold and CW size are jointly optimized to maximize the total throughput, which take into account time-variant link quality (transmit rate), collision probability (minislot contention) and protocol overhead (burst transmission).
- Network performance, under the optimal parameters, is theoretically analyzed and extensively evaluated.

Although nodes use optimized parameters provided by the AP, channel access remains distributed and contention-based. DMUS retains the simplicity of CSMA by using a single threshold, different from previous methods using multiple thresholds [17] or the complex splitting algorithm [15], [16]. In addition, DMUS exploits fresh SNR for distributed scheduling in the downlink transmission, compared with the centralized scheduling relying on statistical channel estimate [18]. Moreover, by using normalized SNR and burst transmission in DMUS, nearly perfect airtime fairness is achieved in most cases, even though nodes may have different average SNR. Simulation results confirm that DMUS greatly improves saturation throughput compared with both the contention-free PCF scheme and the contention-based CSMA/CA scheme. The main finding of this paper is: *for channel access in WLANs, contention is necessary to avoid fading and mitigate unfairness, but it should also be controlled*. By improving the performance of WLANs, DMUS can help to increase the gain of mobile data offloading [1].

The rest of this paper is organized as follows. Previous efforts on improving the performance of WLANs are reviewed in Sec. II, which include collision resolution, opportunistic and concurrent transmissions. Then, we propose the DMUS protocol in Sec. III, addressing how to optimize the multiple access channel via two stages, and how to conduct the channel access. In Sec. IV, we analyze the performance of DMUS and find the optimal parameters. Simulation results of the uplink, downlink and hybrid mode are presented and analyzed in Sec. V. We further discuss potential extensions of DMUS in Sec. VI. Finally, we conclude this paper with Sec. VII.

II. Related Work

WLANs exploit CSMA/CA, whose performance is affected by several factors. Some factors like fading and half-duplex transmission are common to wireless techniques while other factors such as carrier sense and collisions are specific to CSMA/CA. Many efforts have been devoted to solving these problems. We give a brief review of related work and make a short comparison in this section.

A. Collision Avoidance and Detection

Soft reservation is exploited in [3] to reduce collision probability and idle channel, where nodes maintain a precedence relation among one another. This is extended in [4] by passing an implicit token between nodes. These schemes purely focus on collision avoidance and do not take link quality into account. In splitting algorithms [21], nodes involved in a contention are recursively divided into non-overlapping subsets so as to reduce the collision probability. In [15], this splitting algorithm is realized based on link quality. Although this procedure selects the optimal node, it makes the protocol complex and difficult to implement.

When the network is limited to WLANs, centralized schemes have been studied for contention free channel access, among which is PCF [2]. Although PCF removes collisions, channel efficiency may still be degraded by fading.

Rate adaption schemes run on top of CSMA to overcome the fading effect and decide the transmit rate for each outgoing packet. Practical rate adaptation schemes rely on statistics of packet loss as a trigger to drop the data rate. Recently, distinguishing a collision from channel fading [5] is used to diagnose packet losses so that backoff is taken after a collision while data rate is dropped at a fading. A collision may be detected, statistically based on channel congestion degree [22] or by exploiting the likelihood information of decoded bits [6]. By conveying collision information to a sender with collision notification [8], stopping a collided transmission early can reduce the channel waste. Although this is a promising technique, how to accurately and quickly detect a collision and how to reliably feed it back to the sender remain challenges.

B. Opportunistic Transmission

Knopp and Humblet [23] showed that in times of fading, the optimal power control scheme in a multiuser cell is to allocate all power to the node with the highest link quality. This is the origin of MUD. Many MUD schemes require a centralized controller to collect SNR and perform a scheduling.

In a distributed CSMA/CA network, it is difficult to obtain SNR of all nodes without causing much overhead. Some researchers suggested collecting SNR from potential communication peers via exchanging RTS/CTS [24] and on this basis transmit scheduling is performed. In such cases, the benefit of MUD depends on the number of active communication pairs for which packets can be scheduled.

Data packets are transmitted at the cost of protocol overhead and usually rate adaptation is only applied to the payload part. In consequence, transmissions between nodes with low link qualities take much time and may starve nodes with high qualities [25]. In contrast, at the high rate, the protocol overhead throttles throughput. To improve channel efficiency, multiple back-to-back data packets should be transmitted in a burst whenever the channel quality is good [26]. Opportunistic transmissions require accurate estimation of rates, which can be realized by exploiting physical layer information based on cross-layer design [6], [27].

In the uplink, each node usually only learns its own SNR. Exploiting this information in a distributed way, opportunistic slotted ALOHA [13], [14] is suggested under a collision model. MUD is extended to CSMA networks, using the channel gain-based splitting algorithm to resolve collisions [15], [16]. Slotted contention based on multiple thresholds (one threshold per slot) is used in [17] to prioritize nodes with high SNR, but a subsequent randomization procedure is still necessary in order to avoid consecutive collisions when multiple nodes happen to have the same SNR. These designs increase system complexity. Opportunistic transmission is also studied for general ad hoc networks under the collision model [28], and is further extended to the physical interference model [29]. However, these schemes lack practical protocol design. In the downlink, a joint rate control and packet scheduling scheme was suggested in [18] to solve the HOL problem. Because the AP does not have fresh SNR of each node, rough link quality is estimated based on statistics of packet loss, which limits the performance.

C. Concurrent Transmission

Carrier sense is used in CSMA/CA to reduce collision probability. However, the carrier sense range is much wider than the communication range, which leads to the exposed terminal problem. This is solved in [30] by using a centralized scheduling to enable concurrent transmissions. But the necessities of building a conflict graph and using a centralized control in channel access increase system complexity. Full duplex wireless communication recently attracts research interests again and is explored in [7], by exploiting multiple radios and advanced interference cancellation techniques.

D. A Short Comparison

The proposed DMUS scheme distinguishes itself from previous work in the following aspects.

- **Simplicity, effectiveness and fairness**. Instead of using a complex splitting algorithm [15], [16] or multiple SNR thresholds [17], DMUS realizes MUD by combining one normalized SNR threshold with minislot-based contention of CSMA/CA for the simplicity. Its scheduling is optimal with a high probability, though not always. In addition, DMUS achieves airtime fairness both among different nodes and between uplink and downlink.
- **Fresh SNR for downlink scheduling**. In DMUS, the downlink is converted to a multiple access channel by letting nodes initiate their downlink transmissions, and instantaneous SNR is used for distributed scheduling. In comparison, centralized MUD scheduling for the downlink relies on statistical estimation of link quality in [18].
- **Distributed scheduling.** An AP in DMUS uses periodical Notification frames for SNR detection, traffic indication, and signaling of SNR threshold and CW size. However, the node that will communicate with the AP is not specified by the AP, but determined based on normalized SNR at all nodes and the chosen random backoff timers at the nodes whose normalized SNR is above the threshold. This is quite different from centralized scheduling schemes like PCF and CENTAUR [30].
- **Unified framework**. Most previous efforts separately address collisions in the uplink [13], [14], HOL in the downlink [18], and the fairness problem [11]. In comparison, DMUS solves all three problems in a unified framework. In addition to realizing MUD in both uplink and downlink, the unfairness between them is also removed.

III. The DMUS Protocol

The basic procedure of DMUS is given below: An AP, based on average SNR and SNR distribution of all nodes, calculates a normalized SNR threshold and a CW size for all nodes. The AP includes the SNR threshold, the CW size and a traffic indication map (TIM[1]) [2] in its periodical Notification frames. Channel contention is divided into two stages. (i) All nodes detect instantaneous SNR of the Notification frame and only the nodes with normalized SNR greater than the specified threshold contend to access the channel. (ii) From these active contenders, one node is selected via minislot contention using the specified CW. The node initiates a data exchange with the AP, either transmitting to the AP in the uplink or receiving from the AP in the downlink. It occupies the channel for a fixed period during which multiple packets can be exchanged in a burst. For a downlink transmission, the first frame that a node transmits is an invitation message, notifying the AP that the node is ready to receive its buffered packets.

In the following, we first present the distributed contention model in Sec. III-A, explaining how to control the number of contenders and how to perform a minislot contention. In Sec. III-B, we talk about the channel access method for uplink, downlink and the hybrid mode. Then, we further discuss how

[1] Each bit in the TIM is associated with a node and a bit ‘1’ indicates that there are packets destined for the associated node.

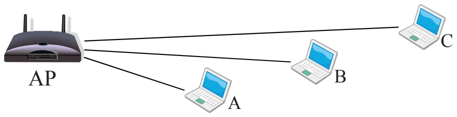


Fig. 1. A WLAN consisting of an AP and its associated nodes.

to recover the DMUS scheme from failure in Sec. III-C and how to reduce protocol overhead in Sec. III-D.

A. Distributed Contention Model

We first consider a basic DMUS model for a WLAN consisting of an AP and M nodes, as shown in Fig. 1. All nodes and the AP are within the same carrier sense range. Wireless transmissions between nodes and their AP take place in both uplink and downlink. The AP monitors the number of nodes associated with it and learns average SNR of each node by monitoring signal strength[2] of frames received from the node. Instantaneous SNR of a node is time-variant and follows the block Rayleigh fading model with its own parameter (average SNR). Long-time fading does not occur thanks to the movements of nodes or background objects. Extension of DMUS to other scenarios is discussed in Sec. VI.

An AP chooses proper parameters for channel access according to channel congestion degree [22]. When the channel congestion degree is less than a threshold, e.g., 40%, the channel is not congested, the AP sets the SNR threshold to $-\infty$ so that any node can transmit at its need without extra latency. When the channel congestion degree gets greater than this threshold, collisions should be controlled. Then, the SNR threshold and CW size are adjusted. How to find the optimal parameters is discussed in Sec. IV.

1) Control the number of contenders: Assume the i^{th} node has an average SNR $\bar{\gamma}_i'$ $(i = 1, 2, \cdots, M)$. Its *normalized SNR*, γ_i, is defined as the ratio of its instantaneous SNR γ_i' to its own average value $\bar{\gamma}_i'$, or $\gamma_i = \gamma_i'/\bar{\gamma}_i'$. γ_i has a probability density function $f_{\gamma_i}(\gamma) = \exp(-\gamma)$ and a cumulative distribution function $F_{\gamma_i}(\gamma) = 1 - \exp(-\gamma)$ under the Rayleigh fading model ([31], Chapter 3), the same for all nodes. $f_{\gamma_i}(\gamma)$ and $F_{\gamma_i}(\gamma)$ are written as $f(\gamma)$ and $F(\gamma)$ hereafter.

Normalization of SNR removes the effect of average SNR from the distribution function. In this way, $P(\gamma_i \geq \gamma_0) = 1 - F(\gamma_0)$, the probability with which a node has a normalized SNR above the threshold γ_0 and gets a chance to contend for the channel, is the same for all nodes. Comparing γ_i against γ_0 is equivalent to comparing γ_i' against $\gamma_0 \cdot \bar{\gamma}_i'$. In other words, each node contends for the channel when its link quality is near its own peak. But using a single normalized SNR threshold simplifies system design.

2) Channel contention: Minislot contention is performed among the nodes with normalized SNR greater than γ_0. DIFS (DCF inter-frame space) after the Notification frame ends, each contender selects a uniformly distributed random integer using the CW specified by the AP, to set up its backoff timer. The backoff timer counts down per idle minislot. The contender, whose backoff timer first reaches 0, grabs the channel and initiates its data exchange with the AP. Other contenders, detecting that the channel gets busy, cancel their timers and transmissions.

In CSMA/CA, the states of backoff timers and CW values are maintained at each node. CW is doubled at a collision and backoff timers are frozen when the channel is sensed as being busy and resumed when the channel is sensed as being idle again. This is so designed because *it is the same set of nodes that contend for the channel all the time in CSMA/CA*. As a comparison, in DMUS, a node decreases its non-zero backoff timer via the slotted contention only when its normalized SNR is greater than γ_0. When its normalized SNR gets less than γ_0, this node is not allowed to join the channel contention even though the channel may be sensed as being idle. When its normalized SNR gets greater than γ_0 again, the set of contending nodes has already changed. For this reason, the backoff timer is reset at the beginning of each contention in DMUS. The expected number of contenders under a given SNR threshold is small (refer to Table II), which can be handled by a fixed CW adapted to the number of nodes. Therefore, in DMUS, a fixed CW is used instead of doubling CW in times of a collision.

A transmission, either uplink or downlink, is always initiated by the node winning the channel contention. When a packet fails the transmission, its retransmission count is increased until the retransmission limit is reached where the packet is dropped. The retransmission will be performed when the node grabs the channel again.

B. Channel Access Sequence

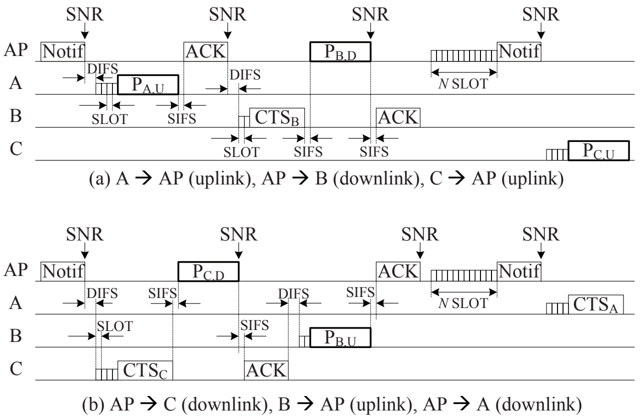
A channel access sequence starts with the AP broadcasting a Notification frame carrying the threshold γ_0, the CW size, and a TIM indicating the downlink traffic, as shown in Fig. 2. Each node measures its SNR on receiving this Notification frame, and contends to access the channel if its normalized SNR is greater than γ_0. The node winning the contention takes either action as follows: (i) transmitting a super frame to the AP if the node has uplink packets, or (ii) transmitting a CTS frame to initiate a downlink reception of a super frame from the AP in case the TIM indicates that there are packets for this node. In the normal case, only a single node communicates with the AP, and the transmit rate is determined by the instantaneous SNR. For example, in Fig. 2(a) node A wins the first contention and sends a super frame $P_{A,U}$ to the AP. The AP correctly receives $P_{A,U}$ and replies an ACK frame SIFS (short inter-frame space) later. Next, node B wins the contention and initiates the transmission of $P_{B,D}$ in the downlink. In Fig. 2(b), node C first initiates the downlink transmission of $P_{C,D}$ followed by an uplink transmission of $P_{B,U}$.

Each super frame in Fig. 2 has a preamble carrying the transmit rate of the payload. The payload may consist of several packets. It is designed that each super frame takes almost a fixed duration and by default each packet has a fixed length in bytes. The number of packets, n, in a super frame depends on the actual rate. More packets can be transmitted in a burst at a higher rate.

C. Recovery from Failure

In the normal case, one and only one node is selected via the minislot contention and the transmission will be

[2] WLAN device drivers provide signal strength of each received frame, which can be used to estimate the average SNR of a node without extra communication overhead.



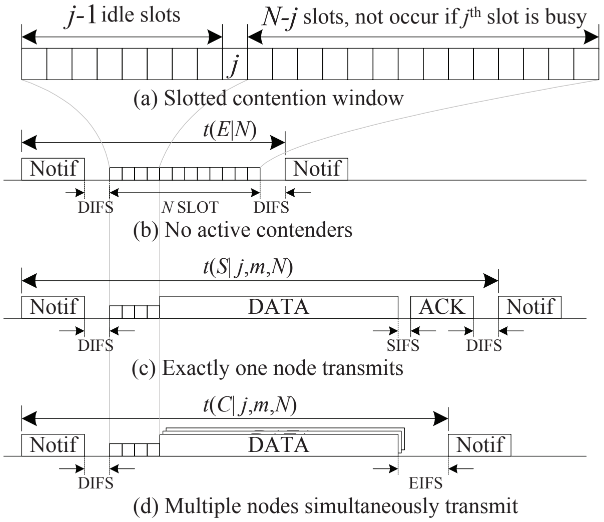
(a) A → AP (uplink), AP → B (downlink), C → AP (uplink)

(b) AP → C (downlink), B → AP (uplink), AP → A (downlink)

Fig. 2. Channel access in DMUS: (a) A transmits $P_{A,U}$ to AP, B initiates the reception of $P_{B,D}$ from AP, and then C transmits $P_{C,U}$ to AP. (b) C initiates the reception of $P_{C,D}$ from AP, and then B transmits $P_{B,U}$ to AP.

TABLE I
MAIN NOTATIONS USED IN THE PERFORMANCE ANALYSIS OF DMUS.

t_{DIFS}	Duration of a DIFS
t_{SIFS}	Duration of an SIFS
t_{EIFS}	Duration of an EIFS
t_{NOTIF}	Transmission time of a Notification frame
t_{BURST}	Transmission time of a DATA frame
t_{ACK}	Transmission time of an ACK frame
T_S	Duration of a minislot
N	Number of minislots in a contention window
m	Number of active contenders
M	Total number of nodes
C_m^M	Number of m-combinations out of M elements



(a) Slotted contention window

(b) No active contenders

(c) Exactly one node transmits

(d) Multiple nodes simultaneously transmit

Fig. 3. Channel access in DMUS: minislot contention and three cases.

successful if channel errors do not occur. There are also three abnormal cases in the channel access which require recovery: (i) It happens that all nodes experience fading and no node initiates the data exchange; (ii) More than two contending nodes choose the same value for their backoff timers and a collision occurs; (iii) Exactly one node talks with the AP but bit errors occur in the CTS/DATA/ACK. In case (i), the channel will remain idle for a continuous period of CW after the Notification frame ends. In case (ii) and (iii), a node or the AP fails to decode a frame. When CTS/DATA/ACK is erroneous, the channel remains idle for a duration of EIFS (Extended inter frame space). After this period, the AP, failing to detect a timely response, broadcasts a new Notification frame, which re-synchronizes the channel access sequence.

D. Reducing Protocol Overhead

The Notification frame broadcast by an AP is not always necessary. It can be omitted if the previous data transmission is successful. In the uplink channel access, each node overhears the ACK frame from the AP. In the downlink channel access, each node overhears the data frame from the AP. At the end of the overheard frame each node measures its SNR. After the current transmission is finished, each node decides to join the next channel access contention according to two factors: (i) its normalized SNR, and, (ii) whether it has an uplink packet or its downlink packet has not been received yet. This process, however, automatically stops if the channel remains idle for a continuous period of CW after the last transmission ends. Then, the recovery procedure works and the AP broadcasts a new Notification frame to start new data exchanges.

IV. Optimal Parameters and Analysis

The performance of DMUS depends on three parameters: the normalized SNR threshold γ_0, the CW size, and the packet size. In this section we discuss how to set these parameters to optimize the saturation performance. We first describe the effect of γ_0 on selecting candidate contenders and analyze the minislot contention in detail in Sec. IV-A. Then, channel efficiency and normalized throughput are defined in Sec. IV-B, based on which optimal γ_0 and CW size are obtained. The probability of selecting the best node via DMUS, using the optimal parameters, is analyzed in Sec. IV-C. Finally, SNR gain of DMUS is compared with that of an ideal MUD scheduling in Sec. IV-D.

In the analysis, we consider communications in the saturation situation where each node always has enough packets to send in a burst to or receive from its associated AP. Altogether there are M nodes, and a CW consists of N minislots with the period of each minislot being T_S. For the simplicity of analysis we focus on the uplink (downlink transmission only differs in that it has an extra CTS frame). The main notations are listed in Table I.

A. Minislot Contention

In DMUS, an AP determines γ_0 so that m out of M nodes will contend for the channel. The probability that m nodes have a normalized SNR greater than γ_0 is

$$P(m|\gamma_0) = C_m^M \cdot F(\gamma_0)^{M-m} \cdot [1 - F(\gamma_0)]^m. \qquad (1)$$

Collision probability due to the contention among these m nodes is reduced by minislot contention. These m contending nodes each start a timer, with a uniformly random integer value taken from $[1, N]$. Each node senses the channel per minislot, decreasing its timer if the channel is sensed as being idle and otherwise cancelling its timer. Fig. 3(a) shows the minislot contention, which may have three cases.

(i) No node has a normalized SNR greater than γ_0 and during the whole contention period the channel is idle (denoted as E), as shown in Fig. 3(b). This probability is

$$P(E|\gamma_0) = P(0|\gamma_0), \qquad (2)$$

and the overhead time is

$$t(E|N) = t_{NOTIF} + t_{DIFS} + T_S \cdot N + t_{DIFS}, \quad (3)$$

$$t(E|\gamma_0, N) = t(E|N) \cdot P(E|\gamma_0). \quad (4)$$

Then, the AP needs to re-broadcast the Notification frame to initiate the next contention period.

(ii) The contention is successful without a collision (denoted as S), as shown in Fig. 3(c). This occurs when one contending node has the least timer value j. This probability is

$$P(S|j,m,N) = \begin{cases} \frac{1}{N}, & 1 \le j \le N, m = 1, \\ C_1^m \cdot \frac{(N-j)^{m-1}}{N^m}, & 1 \le j \le N-1, m \ge 2, \\ 0, & j = N, m \ge 2, \end{cases} \quad (5)$$

and a successful transmission takes the time

$$\begin{aligned} t(S|j,m,N) &= t_{NOTIF} + t_{DIFS} + j \cdot T_S + t_{BURST} \\ &+ t_{SIFS} + t_{ACK} + t_{DIFS}. \end{aligned} \quad (6)$$

The random waiting time (WS: waiting under a successful contention) $j \cdot T_S$ has an expected value

$$\begin{aligned} t\ (WS|m,N) &= \sum_{j=1}^{N} j \cdot T_S \cdot P(S|j,m,N) \quad (7) \\ &= \begin{cases} T_S \cdot N/2, & m = 1, \\ C_1^m \cdot T_S \cdot \sum\limits_{j=1}^{N-1} \left((\frac{j}{N})^{m-1} (1 - \frac{j}{N}) \right), & m \ge 2. \end{cases} \end{aligned}$$

The conditional success probability under a given m is

$$P(S|m,N) = \sum_{j=1}^{N} P(S|j,m,N), \quad (8)$$

the sum of $P(S|j,m,N)$ over all minislots. Each successful transmission on average takes the time

$$\begin{aligned} t(S|\gamma_0, N) &= \sum_{m=1}^{M} t(S|m,N) P(S|m,N) P(m|\gamma_0), \quad (9) \\ t(S|m,N) &= t_{NOTIF} + t_{DIFS} + t(WS|m,N) \\ &+ t_{BURST} + t_{SIFS} + t_{ACK} + t_{DIFS}. \end{aligned}$$

The total success probability is

$$P(S|\gamma_0, N) = \sum_{m=1}^{M} P(S|m,N) P(m|\gamma_0). \quad (10)$$

(iii) A collision occurs (denoted as C), as shown in Fig. 3(d). In the minislot contention, a collision occurs if more than two contending nodes have the same least timer value j. This probability is

$$P(C|j,m,N) = \begin{cases} \sum\limits_{k=2}^{m} C_k^m \cdot \frac{(N-j)^{m-k}}{N^m}, & 1 \le j \le N-1, m \ge 2, \\ \frac{1}{N^m}, & j = N, m \ge 2. \end{cases} \quad (11)$$

Due to the collision, the channel is wasted for a time

$$\begin{aligned} t(C|j,m,N) &= t_{NOTIF} + t_{DIFS} + j \cdot T_S \\ &+ t_{BURST} + t_{EIFS}. \end{aligned} \quad (12)$$

The random waiting time (WC: waiting under a collision) $j \cdot T_S$ has an expected value

$$\begin{aligned} t\ (WC|m,N) &= \sum_{j=1}^{N} j \cdot T_S \cdot P(C|j,m,N) \quad (13) \\ &= \frac{T_S}{N^{m-1}} + T_S \cdot \sum_{j=1}^{N-1} \sum_{k=2}^{m} C_k^m \cdot \frac{1}{N^{k-1}} \left(\frac{j}{N} \right)^{m-k} \left(1 - \frac{j}{N} \right), m \ge 2. \end{aligned}$$

The conditional collision probability under a given m is

$$P(C|m,N) = \sum_{j=1}^{N} P(C|j,m,N), \quad (14)$$

the sum of $P(C|j,m,N)$ over all minislots. Each collision on average takes the time

$$\begin{aligned} t(C|\gamma_0, N) &= \sum_{m=2}^{M} t(C|m,N) P(C|m,N) P(m|\gamma_0), \quad (15) \\ t(C|m,N) &= t_{NOTIF} + t_{DIFS} + t(WC|m,N) + t_{BURST} + t_{EIFS}. \end{aligned}$$

B. SNR Threshold, CW Size and Burst Size

Packets are transmitted with protocol overhead. In the above analysis only $t(S|\gamma_0, N)$ is used for successful transmissions and only t_{BURST} is used for actual data transmission. Therefore, we define channel efficiency as

$$\eta(\gamma_0, N) = \frac{t_{BURST} \cdot P(S|\gamma_0, N)}{t(E|\gamma_0, N) + t(S|\gamma_0, N) + t(C|\gamma_0, N)}, \quad (16)$$

where the numerator is the actual time used for data transmission and the denominator is the average time for each burst.

In DMUS, nodes are allowed to transmit only when their normalized SNR is greater than γ_0. The i^{th} node with an average SNR $\overline{\gamma}_i'$ and a normalized SNR γ_i can transmit $r_i = \log_2(1 + \gamma_i \cdot \overline{\gamma}_i')$ bits per second per Hz according to Shannon theory ([31], Chapter 4). Its average rate, under the condition $\gamma_i \ge \gamma_0$, can be computed as a conditional expectation

$$E(r_i|\gamma_i \ge \gamma_0) = \int_{\gamma_0}^{\infty} \log_2(1 + \gamma \cdot \overline{\gamma}_i') \cdot f_{\gamma_i}(\gamma) d\gamma \Big/ \int_{\gamma_0}^{\infty} f_{\gamma_i}(\gamma) d\gamma. \quad (17)$$

Since each node has the same chance to access the channel, the average transmission rate within the network is

$$E(r|\gamma \ge \gamma_0) = \frac{1}{M} \sum_{i=1}^{M} E(r_i|\gamma_i \ge \gamma_0). \quad (18)$$

Then, we define the normalized throughput (without considering bandwidth) as

$$\Gamma(\gamma_0, N) = \eta(\gamma_0, N) \cdot E(r|\gamma \ge \gamma_0). \quad (19)$$

$\Gamma(\gamma_0, N)$, under different thresholds γ_0 and CW sizes N, is shown in Fig. 4. The number of nodes is M=30, each node has the same average SNR 20dB and the packet size is 1000 bytes. Usually a larger γ_0 leads to more idle channel and a smaller γ_0 leads to more collisions. Fig. 4 shows that throughput can be maximized at a proper γ_0. By adjusting N, the global optimal pair (γ_0, N) can be found. The maximal throughput achieved under different N is almost the same, which indicates that γ_0 *plays a major role in maximizing the throughput.* Throughput curves in Fig. 4 are relatively flat near the maximal points,

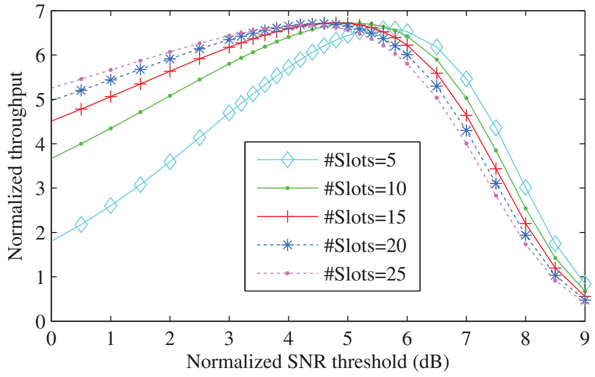


Fig. 4. Normalized throughput of a WLAN under different SNR thresholds and CW sizes (M=30 nodes, average SNR=20dB, packet size=1000bytes).

which indicate that DMUS is not sensitive to small errors in the estimation of γ_0.

γ_0 and N that an AP selects should maximize the normalized throughput as follows:

$$(\gamma_0, N) = \arg\max_{\widehat{\gamma}_0, \widehat{N}} \Gamma(\widehat{\gamma}_0, \widehat{N}). \tag{20}$$

Table II shows the optimal parameters found by grid search under different M. It is interesting to see that both γ_0 and N increase with M. A larger M leads to a higher collision probability, which is to be lowered by increasing γ_0. At a chosen γ_0, the number of contenders, m, changes randomly. The expected number of contenders is calculated as

$$\overline{m}(\gamma_0) = \sum_{m=1}^{M} m \cdot P(m|\gamma_0). \tag{21}$$

As shown in the fourth row of Table II, $\overline{m}(\gamma_0)$ increases with M, i.e., more contenders after comparing normalized SNR against γ_0 at a larger M. Accordingly, a larger N is necessary in order to mitigate collisions among these contenders. Compared with conventional CSMA/CA where all M nodes contend for the channel, the expected number of contenders, $\overline{m}(\gamma_0)$, is very small in DMUS, which means that most collisions are removed by setting the threshold γ_0. Therefore, contention among these few contenders can be handled by a fixed, refined CW, and the effect of a CW is less important in DMUS than in other contention-based schemes.

The two-dimensional grid search of optimal parameters in Eq. (20) can be simplified. Fig. 4 shows that $\Gamma(\gamma_0, N)$ is a concave (and continuous) function of γ_0. At a fixed N, $\Gamma(\gamma_0, N)$ reaches its maximum at a certain value of γ_0. Hence, the optimal γ_0 can be computed from N via the partial derivative $\partial\Gamma(\gamma_0, N)/\partial\gamma_0 = 0$. Accordingly, the optimal parameters (γ_0, N) can be found by one-dimensional grid search in terms of N. In addition, $\Gamma(\gamma_0, N)$ will increase to the maximum when the parameter N changes from the old value (N') to the new optimal value. Then, the grid search can be further optimized by the following procedure. (1) Use N' as the initial seed of N, find its pairing γ_0' via the partial derivative and compute a reference throughput Γ'. (2) Adjust N to the next grid point, find its pairing γ_0 from the partial derivative and compute the throughput Γ. Repeat this process in both sides of N' until Γ gets less than Γ'. (3) Among the searched grid points, the one with the maximal throughput determines the optimal parameters N, γ_0. Because the change in the number of nodes or their average SNR in a real system is small within a short period, N is adjacent to N'. Only a small number of grid points are searched, which helps to greatly reduce the computation cost.

TABLE II
OPTIMAL PARAMETERS AND THE EXPECTED NUMBER OF CONTENDERS UNDER DIFFERENT NUMBERS OF NODES (AVERAGE SNR=20DB).

#nodes (M)	5	10	15	20	25	30	35	40
γ_0 (dB)	2.5	3.5	4.1	4.4	4.7	4.9	5.1	5.2
CW (N)	8	10	11	12	13	13	13	14
$\overline{m}(\gamma_0)$	0.84	1.06	1.14	1.27	1.31	1.36	1.38	1.46

The burst length, t_{BURST}, is an important parameter. As shown in [32], burst transmission effectively improves channel efficiency, especially when packet size is small or data rate is high. But as t_{BURST} increases, the throughput approaches a constant. Meanwhile, the delay and the cost of a collision also increase. Further considering the channel coherent time [26], t_{BURST} is empirically chosen to be 1.7 ms, corresponding to transmitting 1000 bytes at 6Mbps. It should be noted that this value can be further optimized according to application requirements.

C. Probability of Optimal Scheduling

The optimal scheduling is achieved when the node with the highest, normalized SNR is selected (it is equivalent to selecting the node with the highest SNR when all nodes have the same average SNR) by the minislot contention. But running in a distributed way, DMUS does not always lead to an optimal scheduling.

Let normalized SNR, $\gamma_m, m = 1, 2, \cdots, M$, be in the decreasing order and Z_m be the node corresponding to γ_m. The probability that k nodes have normalized SNR greater than γ_0 is $P(k|\gamma_0)$. The subsequent slotted contention among the k nodes is successful with a conditional probability $P(S|k, N)$. In this process, each of the k nodes is selected with the same chance $1/k$, so is the node Z_m when $m \leq k$. Accordingly,

$$P_{Z_m} = \sum_{k=m}^{M} \frac{1}{k} \cdot P(S|k, N) \cdot P(k|\gamma_0) \tag{22}$$

gives the total probability with which DMUS selects the node Z_m. Its numerical result is shown in Fig. 5, where the horizontal axis is m, the rank of nodes. It is clear that the node with the largest, normalized SNR is always selected with the highest probability, around 50%, and the top three nodes are selected with a probability around 70%. This trend is the same under different numbers of nodes.

D. Average SNR Gain

Next we analyze the SNR gain under the case where all nodes have the same average SNR. The SNR gain, achieved by an ideal selection combining (SC) [31], is $\sum_{i=1}^{M} 1/i$.

According to order statistics [33], γ_m, the m^{th} largest, normalized SNR, follows the distribution

$$f_{\gamma_m}(\gamma) = \frac{M!}{(M-m)!(m-1)!}[F(\gamma)]^{M-m} \cdot f(\gamma) \cdot [1-F(\gamma)]^{m-1}. \tag{23}$$

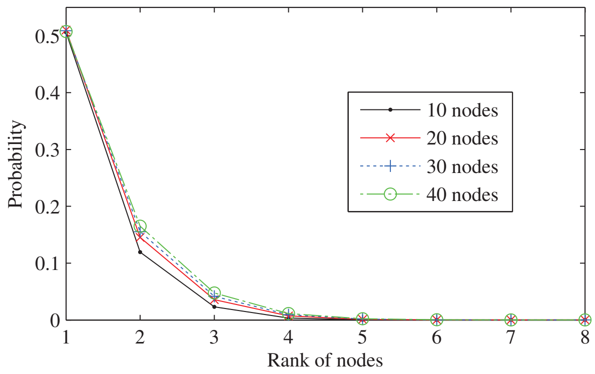


Fig. 5. Probability of selecting a node with m^{th} highest, normalized SNR in a WLAN via DMUS.

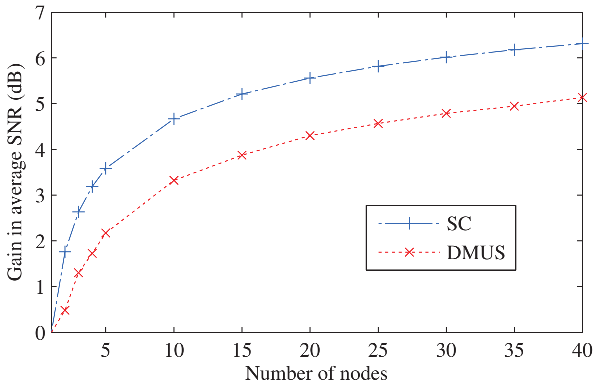


Fig. 6. SNR gain achieved by multiuser diversity in a WLAN: practical DMUS vs. ideal SC (all nodes have the same average SNR).

Under the condition that $\gamma_m \geq \gamma_0$, the average value of γ_m is

$$E(\gamma_m|\gamma_m \geq \gamma_0) = \int_{\gamma_0}^{\infty} \gamma \cdot f_{\gamma_m}(\gamma) d\gamma / \int_{\gamma_0}^{\infty} f_{\gamma_m}(\gamma) d\gamma. \qquad (24)$$

With the optimal γ_0 determined in Sec. IV-B, the above integration can be calculated numerically.

On the other hand, the SNR γ_m is meaningful only when the node Z_m is chosen and its transmission is successful, which has the probability P_{Z_m}. Therefore, the average of normalized SNR in all transmissions is

$$\sum_{m=1}^{M} E(\gamma_m|\gamma_m \geq \gamma_0) \cdot P_{Z_m}, \qquad (25)$$

which actually is the SNR gain of DMUS. Fig. 6 compares the SNR gain achieved by ideal SC and DMUS respectively. The difference between SNR gains is about 1dB under all cases. This small loss of SNR gain in DMUS is acceptable if we take the simplicity of DMUS into account.

V. Simulation Evaluation

We evaluate the performance of DMUS by packet level simulation, using the network simulator Qualnet [34]. DMUS is compared with CSMA/CA and PCF [2]. We also evaluate PCF+SC, where the AP learns the SNR of each node via out-of-band signaling and always selects the node with the highest, *normalized SNR* to communicate with. The overhead of SNR feedback is not involved in PCF+SC. Accordingly, PCF+SC determines the upper-bound of the achievable throughput of DMUS. RTS/CTS is often used in CSMA/CA for reducing collisions and negotiating transmit rates. CSMA/CA with RTS/CTS (CSMA/CA+RC) and PCF with RTS/CTS (PCF+RC) are also evaluated.

TABLE III
Burst factor n for IEEE 802.11a.

Rate (Mbps)	6	9	12	18	24	36	48	54
Burst factor n	1	1,2	2	3	4	6	8	9

The same rate adaptation scheme is adopted in CSMA/CA, PCF and DMUS, and works as follows: (i) Rate adjustment based on SNR. Each node measures SNR when overhearing frames from the AP and infers the rate under this SNR (based on an empirical SNR-rate table), and, (ii) Rate drop based on transmission failure, the same as ARF [35]. In DMUS, rate information for the downlink is carried in CTS and sent to the AP for next transmission. Rate adaptation in PCF+SC is always based on the actual SNR. In CSMA/CA+RC and PCF+RC, the rate is determined via RTS/CTS [10]. Burst transmission is applied in all schemes. IEEE 802.11a physical layer is used and the number of packets in a burst is described in Table III, calculated as r/r_{min} where r is the actual rate and r_{min} is the lowest rate 6Mbps. Except 9Mbps, all rates are multiples of 6Mbps. As for 9Mbps, the sender alternatively transmits 1 or 2 packets with equal probability.

In the simulation, we consider a scenario where M nodes are placed around an AP. All nodes and the AP are in the same carrier sense range. Unless otherwise stated, $\lfloor M/2 \rfloor$ ($\lfloor x \rfloor$ is the maximal integer no more than x) nodes have a same low average SNR and the other nodes have a same high average SNR. Each node experiences independent block Rayleigh fading. We first evaluated the performance of all schemes under light traffic. In this case, the normalized SNR threshold in DMUS is set to $-\infty$ and each node transmits its packet without extra latency. Hence, DMUS has similar delay performance as other schemes. In the following, we focus on the performance of all schemes under saturation situations. In such scenarios, DMUS reduces average delay by decreasing the number of retransmissions via controlling collisions. We mainly show two results: (i) Total throughput of a WLAN, and, (ii) Airtime fairness computed using Jain's fairness index [36]. Simulation results are averaged over 50 runs.

A. *Performance of the Uplink*

Total throughput, achieved by different schemes in the uplink, is shown in Fig. 7. When the number of nodes increases, throughput of CSMA/CA decreases greatly due to severe collisions. Using RTS/CTS brings throughput of CSMA/CA near to that of PCF. In PCF, in times of fading, packets are transmitted at low rates. Therefore, PCF has a relatively low throughput, which does not change with the number of nodes. Using RTS/CTS hardly improves throughput of PCF. In DMUS, transmissions only take place among nodes with normalized SNR greater than the threshold and packets usually are transmitted at higher rates than in PCF. Therefore, DMUS achieves a much higher throughput than PCF.

Due to co-existence of nodes with high and low average SNR, total throughput achieved by all schemes does not change smoothly with the number of nodes. For DMUS,

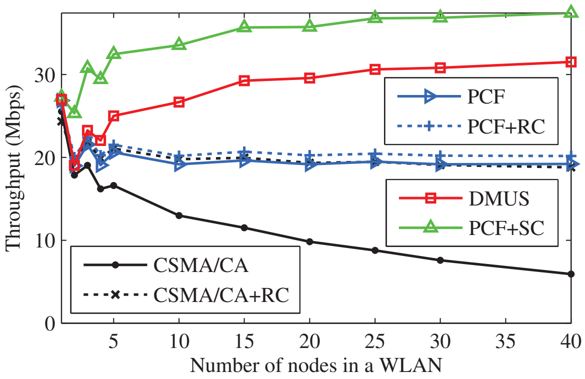


Fig. 7. Total throughput of the uplink under different numbers of nodes ($\lfloor M/2 \rfloor$ nodes: average SNR=14dB, other nodes: average SNR=22dB).

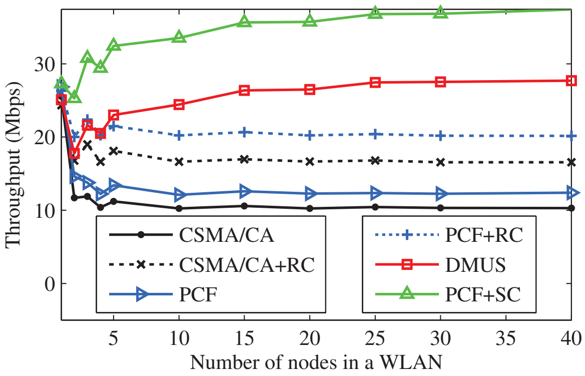


Fig. 8. Total throughput of the downlink under different numbers of nodes ($\lfloor M/2 \rfloor$ nodes: average SNR=14dB, other nodes: average SNR=22dB).

when $M = 1$, there is only a single node with high average SNR (22dB) and its throughput is 27.0Mbps. When $M = 2$, total throughput decreases to 19.0Mbps because half of the channel is shared by another poor node (average SNR=14dB). When M gets greater than 5, total throughput achieved by DMUS and PCF+SC increases with the number of nodes, benefiting from MUD scheduling. But diversity gain due to new nodes gradually diminishes and throughput approaches a constant value. When the number of nodes increases to 40, DMUS improves total throughput by 432.1% compared with CSMA/CA, by 67.5% compared with CSMA/CA+RC, and by 64.2% compared with PCF. Its throughput is 84.2% of the one achieved by the ideal scheme, PCF+SC.

B. Performance of the Downlink

Total throughput, achieved by different schemes in the downlink, is shown in Fig. 8. As for DMUS and PCF+SC, this figure reflects a similar trend as Fig. 7. There is no collision in the downlink for CSMA/CA. But throughput of CSMA/CA is still the lowest due to the HOL problem caused by channel fading. Because the AP has no knowledge of fresh SNR of each node, PCF also has a low steady throughput. Using RTS/CTS for rate adaptation, throughput of CSMA/CA and PCF can be effectively improved. But it is still much less than that of DMUS. In DMUS, by letting each node contend for the channel and initiate transmissions, the downlink is actually converted to a multiple access channel and the HOL problem is removed. Therefore, MUD applies to the downlink as well and DMUS achieves much higher throughput than other schemes.

According to Figs. 7, 8, a conclusion can be drawn as follows: *although too much contention (e.g., CSMA/CA in the uplink) is harmful, controlled contention is necessary in order to avoid fading and better exploit the channel in a distributed way*. DMUS adopts the optimal SNR threshold and CW size to exploit the controlled contention. In this way, it achieves much higher throughput than both CSMA/CA and PCF. *Even though half of the channel is shared by nodes with low average SNR, with sufficient nodes in the network, total throughput achieved by DMUS might get greater than the one achieved by a single node which has high average SNR and monopolizes the channel.* This is clearly reflected in Figs. 7, 8.

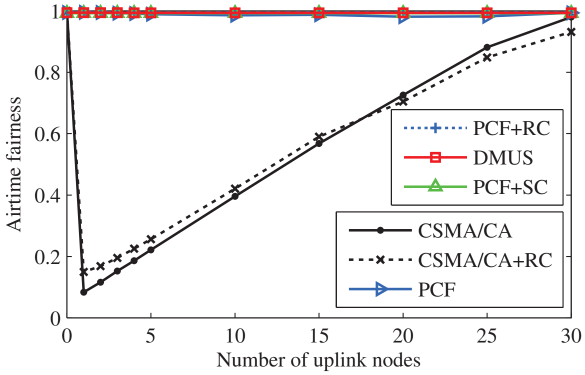


Fig. 9. Airtime fairness of the hybrid access mode under different numbers of uplink nodes ($\lfloor M/2 \rfloor$ nodes: average SNR=14dB, other nodes: average SNR=22dB).

C. Performance of the Hybrid Mode

In the hybrid mode, the number of nodes is fixed to 30. Each node either sends in the uplink or receives in the downlink, but not both. The number of total uplink nodes is adjusted.

Figure 9 shows Jain's fairness index computed from per-node airtime achieved by different schemes in the hybrid access mode. CSMA/CA has a quite different fairness curve from other schemes. This can be explained as follows: When there are m uplink nodes, the number of downlink nodes is $M - m$. Then, the AP and the m uplink nodes each occupy the channel with a share of $\frac{1}{m+1}$. When m is less than M, the AP is responsible for sending packets to the $M - m$ downlink nodes. The actual share of the channel is $\frac{1}{m+1} \cdot \frac{1}{M-m}$ for a downlink node and $\frac{1}{m+1}$ for an uplink node. For CSMA/CA, fairness index of airtime calculated according to these shares matches the result in Fig. 9. In comparison, nearly perfect airtime fairness is achieved in DMUS even though nodes have different average SNR and there are both uplink and downlink traffic. This is due to two factors: (i) It is not absolute SNR but normalized SNR that is used for channel contention in DMUS, which removes the effect of average SNR that may be different among nodes. (ii) By letting nodes initiate their own downlink transmission, nodes with downlink traffic directly contend with nodes with uplink traffic instead of relying on the AP to perform downlink scheduling. As a result, each node, either with uplink or downlink traffic, has the same channel share $1/M$. Therefore, the unfairness between uplink and downlink is removed.

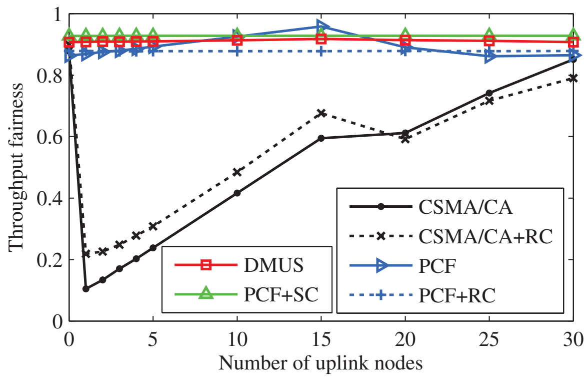


Fig. 10. Throughput fairness of the hybrid access mode under different numbers of uplink nodes ($\lfloor M/2 \rfloor$ nodes: average SNR=14dB, other nodes: average SNR=22dB).

It should be noted that DMUS ensures airtime fairness instead of throughput fairness. When nodes almost have the same channel share in terms of airtime, nodes near to the AP will have higher throughput. This is consistent with practical systems. Throughput fairness, under the same scenario as in Fig. 9, is shown in Fig. 10. In this evaluation, nodes have different average SNR. Therefore, throughput fairness is degraded and is not as perfect as airtime fairness.

VI. Some Extensions

In this section, we investigate potential extensions of DMUS. We first examine how to use different normalized SNR thresholds in the network so that nodes may have different channel shares, a kind of service differentiation. Each node still uses only one normalized SNR threshold, which may be different from that of other nodes. In this case, we consider two service priorities, one low priority group with M_1 nodes using a normalized SNR threshold $\gamma_{0,1}$ and the other high priority group with M_2 nodes using a normalized SNR threshold $\gamma_{0,2}$, where $M_1 + M_2 = M$. For the simplicity, it is assumed that all nodes share the same CW. In this design, the ratio of airtimes obtained by a high priority node and by a low priority node is set to $\rho : 1$ ($\rho > 1$), which is realized by $1 - F(\gamma_{0,2}) = \rho \cdot [1 - F(\gamma_{0,1})]$. Then, we have $\gamma_{0,2} < \gamma_{0,1}$. When there are m_1 low priority nodes with normalized SNR above $\gamma_{0,1}$ and m_2 high priority nodes with normalized SNR above $\gamma_{0,2}$, $m = m_1 + m_2$ nodes will contend for the channel via slotted contention. Using the probability

$$P(m|\gamma_{0,1},\gamma_{0,2}) = \sum_{m_1+m_2=m} P(m_1|\gamma_{0,1},M_1)P(m_2|\gamma_{0,2},M_2), \qquad (26)$$

$$P(m_i|\gamma_{0,i},M_i) = C_{m_i}^{M_i} \cdot F(\gamma_{0,i})^{M_i-m_i} \cdot [1-F(\gamma_{0,i})]^{m_i}, i = 1, 2,$$

$$1 - F(\gamma_{0,2}) = \rho \cdot [1 - F(\gamma_{0,1})],$$

in place of $P(m|\gamma_0)$ in Eq. (1) and following the procedure in Sec. IV, $\gamma_{0,1}$, $\gamma_{0,2}$ and CW can be jointly optimized.

With $\rho = 2$ and $M = 30$, we evaluate the per-node airtime and throughput of DMUS in the uplink, where all nodes have the same average SNR=20dB. The results are shown in Fig. 11. A high priority node has nearly double airtime as a low priority node, consistent with the setting $\rho = 2$. The channel share of a high priority node equals $\rho/(M_1 \cdot 1 + M_2 \cdot \rho) = \rho/(\rho \cdot M - M_1 \cdot (\rho - 1))$ and increases with M_1, which is also consistent with the evaluation result. The transmit rate of a high priority node averaged over the range $[\gamma_{0,2}, \infty)$ using Eq. (17) is less than that of a low priority node averaged over the range $[\gamma_{0,1}, \infty)$. Therefore, a high priority node achieves a throughput less than the double of that of a low priority node, although it occupies double airtime.

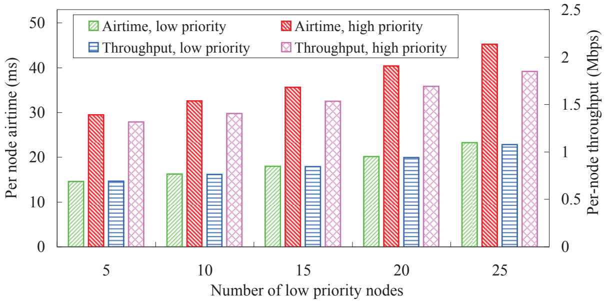


Fig. 11. Per-node airtime and throughput under different numbers of low-priority nodes (two priorities, the number of total nodes is 30, all nodes have the same average SNR=20dB).

TABLE IV
Optimal Normalized SNR Threshold γ_0 and CW Size N Under Different Average SNR (M=30 Nodes).

SNR (dB)	10	12	14	16	18	20	22	24	26	28	30
γ_0 (dB)	5.3	5.2	5.1	5.0	5.0	4.9	4.8	4.8	4.7	4.7	4.6
CW (N)	10	11	11	12	12	13	14	14	15	15	16

Next, we examine DMUS in a multi-cell scenario. APs and M nodes run on the same channel[3] and are in the same carrier sense range. Each node is randomly placed in a cell. This differs from the one-cell scenario as follows: (i) We assume that inter-AP coordination is possible via the wired backbone network connecting all APs. Each AP learns average SNR of all nodes in its carrier sense range via inter-AP cooperation and computes the SNR threshold and CW size for all nodes in that range. (ii) APs take turns to transmit their Notification frames so that each node can estimate fresh SNR. Then, each node contends for the channel based on its normalized SNR and communicates with its associated AP after winning the contention. We evaluate the performance of a two-cell network with the following settings: all nodes experience independent block Rayleigh fading and have the same average SNR; there are two APs and the number of nodes is fixed at $M = 30$.

Parameters used in this evaluation are shown in Tab. IV. As average SNR increases, the normalized SNR threshold is decreased so as to reduce the chance that the channel is wasted without any transmission. Meanwhile, CW is increased to control the collision probability.

Uplink throughput, under different average SNR, is shown in Fig. 12. In all schemes, total throughput increases with average SNR and approaches a steady value, which is limited by the maximal rate (54Mbps in 802.11a). DMUS has a satisfactory throughput compared with PCF+SC. The performance difference between DMUS and PCF+SC is partially due to the factor that the SNR detection overhead is not involved in PCF+SC. DMUS achieves much higher throughput than PCF in most cases. Only when average SNR is very high (≥30dB) will PCF achieve a similar throughput as DMUS. This is

[3] DMUS can simply run in parallel if adjacent cells work on different channels.

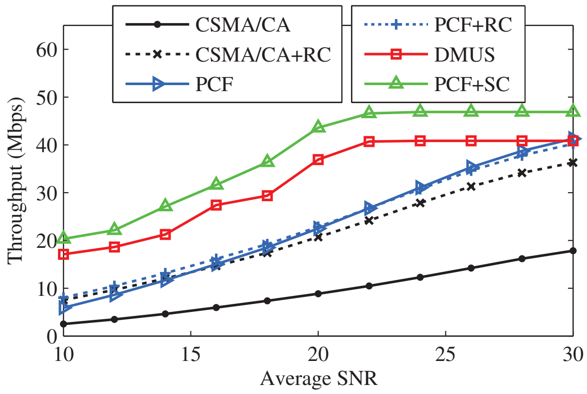


Fig. 12. Total throughput of the uplink in a two-cell network (All nodes have the same average SNR and the number of nodes is fixed to 30).

because when average SNR is above 30dB, the instantaneous SNR is high enough to support the highest rate of 802.11a. But with 802.11n which supports much higher rates (up to 600Mbps), it is expected that DMUS will be superior to PCF in a much wider SNR range.

DMUS can also be extended to support general channel models, which may be different for nodes or even unknown to the AP. In the latter case, the i^{th} node should collect its SNR statistics, compute an estimated density function $f_{\gamma_i'}(\gamma)$ of its instantaneous SNR γ_i', and report $f_{\gamma_i'}(\gamma)$ to the AP, using the mechanism defined by IEEE 802.11k. The AP further computes the cumulative distribution function $F_{\gamma_i'}(\gamma)$. The i^{th} node should have its own SNR threshold $\gamma_{0,i}$, and contend to access the channel only if γ_i' is greater than $\gamma_{0,i}$, which occurs with the probability $P(\gamma_i' \geq \gamma_{0,i}) = 1 - F_{\gamma_i'}(\gamma_{0,i})$. By (i) using $F_0 = F_{\gamma_i'}(\gamma_{0,i})$ instead of $F(\gamma_0)$ for all nodes in Eq. (1), and, (ii) using $f_{\gamma_i'}(\gamma)$ instead of $f(\gamma)$ in Eq. (17), F_0 and CW size can be found by using the procedure in Sec. IV. Then, $\gamma_{0,i}$ is computed according to $P(\gamma_i' \geq \gamma_{0,i}) = 1 - F_0$. Fairness is still retained since each node contends to access the channel with the same probability $1 - F_0$.

The general model can be simplified in some cases. (i) For Rayleigh fading, γ_i' of the i^{th} node can be expressed by the product of normalized SNR γ_i and average SNR $\overline{\gamma}_i$ as $\gamma_i' = \gamma_i \cdot \overline{\gamma}_i$. Let $\gamma_{0,i} = \gamma_0 \cdot \overline{\gamma}_i$. When $\overline{\gamma}_i$ is known, $F_0 = F_{\gamma_i'}(\gamma_0 \cdot \overline{\gamma}_i)$ is a function of the normalized SNR threshold γ_0. Therefore, γ_0 and CW size can be directly computed and node i contends for the channel if γ_i' is greater than $\gamma_0 \cdot \overline{\gamma}_i$, or $\gamma_i = \gamma_i'/\overline{\gamma}_i$ is greater than γ_0. (ii) For log-normal shadowing model, the log value of instantaneous SNR of the i^{th} node, $l_i' = \log \gamma_i'$, follows normal distribution $\mathcal{N}(\mu_i, \sigma_i)$, where μ_i and σ_i are mean and standard deviation of l_i'. l_i' can also be expressed by the normalized value l as $l_i' = \mu_i + \sigma_i \cdot l$. With μ_i and σ_i known, $F_0 = F_{l_i'}(\mu_i + \sigma_i \cdot l_0)$ is a function of l_0. l_0 and CW size can be directly found by using the procedure in Sec. IV. Then, node i contends to access the channel if $(l_i' - \mu_i)/\sigma_i$ is greater than l_0. By simulation evaluation, we confirmed that DMUS also works well for the log-normal shadowing channel, both improving throughput and retaining airtime fairness compared with CSMA/CA.

In a real system, it is possible that a node may experience a long-time fading due to lack of movement. To address this problem, each node uses a new flag, LTF (long-time fading), to track its state. The LTF flag of a node is set if its normalized SNR is below the threshold for a continuous period, and cleared otherwise. The DMUS scheme is modified to work with normalized SNR, LTF and CW as follows: A node with LTF cleared contends for the channel using the CW when its normalized SNR is above the threshold. A node with LTF set neglects its normalized SNR, and contends for the channel with a same probability $1 - F_0$ (by mimicking a Bernoulli process) as other nodes not in long-time fading. This heuristic method ensures that nodes in long-time fading are not starved.

VII. Conclusion and Future Work

The performance of WLANs may be greatly degraded by the collision, HOL and unfairness problems. State-of-the-art methods usually focus on one problem by exploiting multiuser diversity, either using complex splitting algorithm to find the optimal node or using multiple SNR thresholds to prioritize nodes with high SNR. In contrast, we have suggested DMUS to solve all these problems in a unified framework. The proposed scheme is both simple (using only one normalized SNR threshold in combination with minislot-based contention of CSMA/CA) and effective (the scheduling is optimal with a high probability). In addition, fresh SNR is exploited for downlink scheduling, and airtime fairness is achieved not only among different nodes but also between uplink and downlink. Simulation results confirm that DMUS has a high throughput gain compared with contention-free PCF and a much higher throughput gain in comparison with contention-based CSMA/CA.

Service differentiation in terms of airtime is touched a little in this paper. It can be further extended from two aspects: (i) Exploiting multiple SNR thresholds inside a node so as to realize different access categories defined in IEEE 802.11, e.g., using a low SNR threshold for real-time traffic to reduce delay and using a high SNR threshold for background traffic to improve throughput. (ii) Exploiting hybrid channel access where high priority traffic is transmitted under the control of the AP, and low-priority traffic is transmitted via distributed scheduling. These are left as future work.

References

[1] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: how much can WiFi deliver?" *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 536–550, Apr. 2013.

[2] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 2012.

[3] Z. Zeng, Y. Gao, K. Tan, and P. R. Kumar, "CHAIN: introducing minimum controlled coordination into random access MAC," in *Proc. 2011 IEEE INFOCOM*, pp. 2669–2677.

[4] G. Hosseinabadi and N. H. Vaidya, "Token-DCF: an opportunistic MAC protocol for wireless networks," in *Proc. 2013 COMSNETS*, pp. 1–9.

[5] S. Rayanchu, A. Mishra, D. Agrawal, S. Saha, and S. Banerjee, "Diagnosing wireless packet losses in 802.11: separating collision from weak signal," in *Proc. 2008 IEEE INFOCOM*, pp. 735–743.

[6] M. Vutukuru, H. Balakrishnan, and K. Jamieson, "Cross-layer wireless bit rate adaptation," in *Proc. 2009 ACM SIGCOMM*, pp. 3–14.

[7] J. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proc. 2010 ACM MobiCom*, pp. 1–12.

[8] S. Sen, R. R. Choudhury, and S. Nelakuditi, "CSMA/CN: carrier sense multiple access with collision notification," in *Proc. 2010 ACM MobiCom*, pp. 25–36.

[9] P. Bhagwat, P. Bhattacharya, A. Krishna, and S. K. Tripathi, "Enhancing throughput over wireless LANs using channel state dependent packet scheduling," in *Proc. 1996 IEEE INFOCOM*, vol. 3, pp. 1133–1140.

[10] G. Holland, N. H. Vaidya, and P. Bahl, "A rate-adaptive MAC protocol for multi-hop wireless networks," in *Proc. 2001 MobiCom*, pp. 236–251.

[11] W.-S. Lim, D.-W. Kim, and Y.-J. Suh, "Achieving fairness between uplink and downlink flows in error-prone WLANs," *IEEE Commun. Lett.*, vol. 15, no. 8, pp. 822–824, Aug. 2011.

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

[13] X. Qin and R. Berry, "Distributed approaches for exploiting multiuser diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 392–413, Feb. 2006.

[14] S. Adireddy and L. Tong, "Exploiting decentralized channel state information for random access," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 537–561, Feb. 2005.

[15] X. P. Qin and R. Berry, "Opportunistic splitting algorithms for wireless networks," in *Proc. 2004 IEEE INFOCOM*, vol. 3, pp. 1662–1672.

[16] J. Wang, H. Zhai, Y. Fang, J. M. Shea, and D. Wu, "OMAR: utilizing multiuser diversity in wireless ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 5, no. 12, pp. 1764–1779, Dec. 2006.

[17] C.-S. Hwang and J. M. Cioffi, "Opportunistic CSMA/CA for achieving multi-user diversity in wireless LAN," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2972–2982, June 2009.

[18] Q. Xia, X. Jin, and M. Hamdi, "Cross layer design for the IEEE 802.11 WLANs: joint rate control and packet scheduling," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2732–2740, July 2007.

[19] S. Tang, R. Miura, and S. Obana, "Distributed multi-user scheduling for improving throughput of wireless LAN," in *Proc. 2009 IEEE ICC*.

[20] H. Kushner and P. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1250–1259, July 2004.

[21] R. Garces and J. J. Garcia-Luna-Aceves, "Collision avoidance and resolution multiple access for multichannel wireless network," in *Proc. 2000 IEEE INFOCOM*, vol. 2, pp. 595–602.

[22] P. A. K. Acharya, A. Sharma, E. M. Belding, K. C. Almeroth, and K. Papagiannaki, "Congestion-aware rate adaptation in wireless networks: a measurement-driven approach," in *Proc. 2008 IEEE SECON*, pp. 1–9.

[23] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. 1995 IEEE ICC*, vol. 1, pp. 331–335.

[24] Z. Ji, Y. Yang, J. Zhou, M. Takai, and R. Bagrodia, "Exploiting medium access diversity in rate adaptive wireless LANs," in *Proc. 2004 ACM MobiCom*, pp. 345–359.

[25] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *Proc. 2003 IEEE INFOCOM*, vol. 2, pp. 836–843.

[26] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, "Opportunistic media access for multirate ad hoc networks," in *Proc. 2002 ACM MobiCom*, pp. 24–35.

[27] S. Sen, N. Santhapuri, R. R. Choudhury, and S. Nelakuditi, "AccuRate: constellation based rate estimation in wireless networks," in *Proc. 2010 USENIX NSDI*, pp. 12–12.

[28] D. Zheng, W. Ge, and J. Zhang, "Distributed opportunistic scheduling for ad-hoc communications: an optimal stopping approach," in *Proc. 2007 ACM MobiHoc*, pp. 1–10.

[29] W. Ge, J. Zhang, J. E. Wieselthier, and X. Shen, "PHY-aware distributed scheduling for ad hoc communications with physical interference model," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2682–2693, May 2009.

[30] V. Shrivastava, N. Ahmed, S. Rayanchu, S. Banerjee, S. Keshav, K. Papagiannaki, and A. Mishra, "CENTAUR: realizing the full potential of centralized WLANs through a hybrid data path," in *Proc. 2009 ACM MobiCom*, pp. 297–308.

[31] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[32] Y. Xiao, "IEEE 802.11n: enhancements for higher throughput in wireless LANs," *IEEE Wireless Commun. Mag.*, vol. 12, no. 6, pp. 82–91, Dec. 2005.

[33] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. Wiley-Interscience, 2003.

[34] "Qualnet." Available: http://web.scalable-networks.com/content/qualnet

[35] A. Kamerman and L. Monteban, "WaveLAN-II: a high-performance wireless LAN for the unlicensed band," *Bell Labs Technical J.*, vol. 2, no. 3, pp. 118–133, Aug. 1997.

[36] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," DEC, Tech. Rep. TR-301, Sept. 1984.

Suhua Tang received the B.S. degree in Electronic Engineering in 1998 and the Ph.D. degree in Information and Communication Engineering in 2003, both from University of Science and Technology of China. From Oct. 2003 to Mar. 2014, he was with Adaptive Communications Research Laboratories, ATR, Japan. Since Apr. 2014, he is with Department of Communication Engineering and Informatics, Graduate School of Informatics and Engineering, The University of Electro-Communications, Japan, and is a guest researcher at ATR. His research interests include green communications, network coding, cross-layer design, mobile ad hoc networks and inter-vehicle communications. He is a member of IEICE.