Distributed Multi-User Scheduling for Improving Throughput of Wireless LAN

Suhua TANG, Ryu MIURA, Sadao OBANA
ATR Adaptive Communications Research Laboratories, Japan
Email: {shtang, miurar, obana}@atr.jp

Abstract—Carrier Sense Multi-Access (CSMA) is a typical method to share the common channel in a Wireless LAN (WLAN). It works fairly well in times of light traffic. However, as the number of nodes in a WLAN increases quickly, severe collision greatly degrades network performance. In this paper we propose a Distributed Multi-User Scheduling (DMUS) scheme to solve this problem, taking time-variant link quality and rate adaptation into account. Instead of all nodes, only nodes with high instantaneous link quality are allowed to contend for the channel. By setting a suitable SNR threshold, at any instance only a small percentage of nodes join the contention. As a result, collision is mitigated, fairness is retained by independent fading, and the total throughput is increased since transmissions are finished at higher rates. Simulation results show when there are 40 nodes in a WLAN, the DMUS scheme improves total throughput by up to 49.6% compared with the Contention-Free scheme, and by up to 194.6% compared with the CSMA/CA scheme.

I. INTRODUCTION

Wireless LANs (WLAN) provide quick access to the Internet and are widely deployed in hotspot areas. Typically all nodes in a WLAN access the Internet by the associated Access Point (AP) and share the common wireless channel with the well-known Carrier Sense Multi-Access (CSMA) scheme [1].

CSMA has achieved great success in the Ethernet where collision can be detected quickly. In wireless networks, however, due to the half-duplex nature, each node cannot simultaneously transmit and receive with the same radio. Instead collision avoidance is used by the virtual carrier sense mechanism [2], [3]. Lack of a timely ACK after the transmission of a data packet is regarded as a collision. Then exponential back-off is started to alleviate collision. CSMA works fairly well in times of light traffic. However, as the number of nodes in a WLAN increases quickly, severe collision unavoidably degrades performance of the whole network. Since quick detection of collision is infeasible in a WLAN, the only choice to reduce collision is to control the number of nodes attending the contention. Some conventional schemes include channel diversity [4], directional antenna, transmit power control, adaptive backoff [5] and carrier sense [6]. Their common idea is to reduce the number of nodes in a collision domain.

Multi-path fading degrades performance of a WLAN with random signal fluctuation. It is usually regarded as being harmful and different diversity schemes, such as receive diversity, transmit diversity and time diversity, have been proposed to mitigate fading. Retransmission is also an effective solution.

Though either fading or increase in node density affects system performance, a joint consideration of the two factors may even improve system performance by exploiting multi-user scheduling ([7], Chapter 6). The basic idea is to exploit the independence of random fading. In cellular networks an AP can monitor Channel State Information (CSI) all the time and perform centralized scheduling. Multi-user diversity has already become a part of 3G wireless communication systems. In a WLAN, however, there is no feedback channel and an AP cannot detect CSI if a node keeps silence.

In this paper we propose a Distributed Multi-User Scheduling (DMUS) scheme, aiming to improve scalability and throughput of WLANs. Time-variant link quality under multi-path fading is considered and nodes are allowed to transmit only when their link quality is high enough. By setting a suitable SNR threshold, at any instance only a small percentage of nodes contend to access the common wireless channel. This SNR threshold is set by the AP, according to the number of nodes in the cell. In addition, it is a normalized value, which enables fair access of the channel by all nodes. In conventional CSMA a node has to double its contention window and perform exponential back-off in times of collision. In the proposed DMUS scheme a fixed contention window is adopted and the collision probability is controlled by the AP. The total throughput of a WLAN is improved since transmissions are finished at higher rate by avoiding deep fading. Simulation results indicate that DMUS can improve throughput by up to 194.6% compared with CSMA/CA in Rayleigh fading environment when there are 40 nodes in a WLAN.

The rest of the paper is organized as follows. Section II reviews previous efforts on improving throughput of WLANs. Then research motivation is addressed in Section III. In Section IV we propose our system model and discuss protocol design. In Section V by analysis we show how to find optimal parameters for minimizing collision and achieving high throughput. Simulation results and analysis are presented in Section VI. Finally we conclude the paper with Section VII.

II. BACKGROUND AND RELATED WORK

CSMA in a WLAN depends on distributed contention and does not scale well to node density. As a result an increase in the number of nodes in a WLAN leads to higher collision probability and degrades network throughput. Previous efforts have suggested quick collision resolution or adaptive carrier sense. Others even suggest benefiting from the increase of nodes by exploiting multi-user diversity.
A. Collision Alleviation and Adaptive Carrier Sensing

In splitting algorithms [4] nodes involved in a collision are divided into several subsets so as to reduce the collision probability. An alternative approach is to exploit adaptive backoff, as suggested in [5].

Carrier sensing in CSMA effectively mitigates collisions, but due to the broadcast nature it also results in exposed terminals problem and reduces spatial reuse. This is one of the reasons why CSMA does not scale well. Yu et al. suggested adopting less sensitive carrier sensing to promote more spatial reuse of the channel, and alleviates the collision problem by adaptively adjusting the communication distance via “packet salvaging” at the MAC layer [6]. A similar scheme was introduced in [8], where a combination of receiving-sensitivity adaptation reduces stronger-last collision and a clear-channel-assessment adaptation balances the hidden/exposed terminal problems.

B. Multi-User MAC

Knopp and Humblet [9] showed that in times of fading, the optimal power control scheme in a multi-user cell is to allocate all power to the link with the highest quality based on the feedback of CSI. This is known as multi-user diversity. Many existing schemes require a centralized controller to collect CSI.

In a CSMA/CA network, some researchers suggested to collect the CSI from potential communication peers [10], [11] and on this basis performs scheduling. In such cases the benefit of multi-user diversity depends on the number of active communication pairs for which packets can be scheduled. In the up-link where all nodes contend to transmit to the same AP, these schemes hardly work.

In multi-access up-link channel usually each node only learns its own CSI to the AP. Opportunistic slotted Aloha [12], [13] was suggested under a collision model, with the basic idea being that in every slot each user transmits with a probability based on its own channel condition. How to reduce collision is a key issue.

C. Our Contribution

Previous efforts tend to separately consider scalability and multi-user scheduling. In comparison, we jointly consider both factors in WLANs. Compared with [5], [6], in the proposed DMUS scheme, the carrier sense mechanism remains the same as IEEE 802.11 [3] but contention window is fixed. The number of contenders is reduced and the contenders change over time according to their instantaneous CSI. Though existing multi-user diversity schemes [10], [11], [12] can mitigate fading, they still suffer from severe collision when there are many active senders. In the proposed scheme the number of contenders is determined by the AP and is made adaptive to the number of nodes in a WLAN.

III. Motivation and Overview

Distributed Coordination Function (DCF) [3] is the main MAC scheme in WLANs. It is based on CSMA/CA, which further relies on distributed contention. As the number of nodes increase, the collision probability inevitably increases and the total performance of the network is degraded.

Increase in the number of nodes is not always a minus effect. As shown in [7] (Chapter 6), in times of fading the system performance can be improved by multi-user scheduling. In cellular networks, feedback of CSI to the AP and centralized scheduling are always available. In a WLAN, however, there is no feedback channel. The common channel is shared in a distributed way. Although there have been some efforts on Point Coordination Function (PCF) where an AP performs contention free transmission, this optional function is almost not implemented at all. To effectively exploit multi-user diversity in a WLAN, two factors are taken into account in this work.

(i). CSI detection. We exploit the broadcast nature of wireless medium. In a WLAN, transmission always takes place between an AP and one of its nodes. Other nodes can overhear the signal (either data or ACK) from its AP and detect the CSI. (It was mentioned in [14] that at moderate speeds channel coherent time is long enough and the CSI keeps unchanged within the transmission of several packets.) Then according to their CSI, nodes contend for the wireless channel and only the winner talks with the AP.

The up-link is a multi-access channel. Each node can detect its CSI and contend to transmit to the AP in a distributed way. The down-link is a broadcast channel, where the AP transmits to its associated nodes. Nodes need to provide some feedback to the AP so that AP knows to whom it can transmit efficiently. In the proposed scheme the AP indicates the down-link traffic in polling messages, and the actual transmission is initiated by an invitation from nodes (a reader interested in invitation-based scheme is referred to [15]). This effectively converts the broadcast channel to a multi-access channel.

(ii). Distributed MAC (scheduling and collision control). Both fading and collision may result in transmission failure, after which different procedures should be followed. The former triggers rate adaptation while the latter initiates backoff. However usually it is difficult to tell one from the other (it was mentioned in [16] that distinguishing failure reason is possible by complex operation). Then in a WLAN lack of an ACK always leads to backoff. This, however, is not efficient when a node in its peak rate has to wait for the channel (it is possible that another node with a much lower rate may occupy the channel first because it has a smaller random backoff value). Our policy is to let part of nodes contend for the channel. Specifically, only nodes whose links to the AP are near their peak rate are allowed to take part in the contention. Thanks to independent fading, the set of contenders changes dynamically according to CSI. This retains fair access.

When CSI has a high accuracy, with rate adaptation most transmission failure can be attributed to collision. The collision probability is further related with the number of nodes in the dynamic contention set. By adjusting this number, the collision probability can be controlled by the AP.
IV. NETWORK MODEL AND PROTOCOL

A. Link Quality Aware Distributed Contention Model

We consider a WLAN consisting of a single AP and \( M \) nodes, as shown in Fig. 1. Nodes exchange up/down link packets with their associated AP. All nodes and the AP are within the same carrier sense range. In other words, it is assumed that the hidden terminal problem does not occur at the AP. (Otherwise a busy tone can be used at the AP to indicate receiving status). Each link independently experiences block Rayleigh fading.

First consider the case where each link has the same average SNR \( \bar{\gamma}_i = \bar{\gamma} \). SNR of a link follows i.i.d. exponential distribution \( f(\gamma_i) = \exp(-\gamma_i/\bar{\gamma})/\bar{\gamma} \). Each node contends to access the channel. Analogous to selection diversity, average SNR of all transmissions will be \( \bar{\gamma}_i = \bar{\gamma}/M + 1/i \) with ideal multi-user scheduling.

Multi-user diversity effectively increases average SNR, however, in a non-linear way. The biggest gain is obtained by going from no diversity to two-branch diversity and in general increasing the number of links yields diminishing returns in terms of SNR gain. But the increase of nodes inevitably results in collision. Therefore it is necessary to reduce the number of nodes in the contention in order to reduce the collision probability. In the proposed scheme, an SNR threshold \( \gamma_0 \) is set by the AP according to the number of nodes in its cell. Only nodes detecting an instantaneous SNR no less than \( \gamma_0 \) on links to the AP will contend to access the channel.

According to the randomness of fading, a fixed SNR threshold can be set. When each link has the same average SNR, each node still has the same chance to access the channel. However, in the real environment, all links have different propagation paths and different path loss, and their average SNR also differs. With a fixed SNR threshold, nodes near the AP will have more chance to occupy the channel while nodes far away can hardly.

In the following we show that setting of the SNR threshold should be based on the normalized SNR in order to retain fairness. Assume the \( i^{th} \) link has an average SNR \( \gamma_i = \gamma_0 \). Let \( \gamma_i = \gamma_i/\bar{\gamma}_i \) be the SNR normalized with respect to its average value. The probability density function of \( \gamma_i \) is \( f(\gamma_i) = \exp(-\gamma_i/\bar{\gamma}_i) \). It is obvious that the normalized SNR of each link has the same density function. With the common SNR threshold set to \( \gamma_0 \), the distribution function of each normalized SNR is the same, \( P(\gamma_i \leq \gamma_0) = 1 - \exp(-\gamma_0) \). This restores fairness of channel access. The optimal value of \( \gamma_0 \) is discussed in Section V.

B. Detailed Channel Access Method

Up-link access and down-link access corresponding to the WLAN in Fig. 1 are shown in Figs. 2-3 respectively.

As for the multi-access up-link, the AP first broadcasts a short Poll frame. This frame notifies nodes to start channel contention. At the end of the Poll frame each node measures the instantaneous SNR. Each node with at least one upstream packet starts a timer with a random value if its normalized SNR is greater than the threshold \( \gamma_0 \) specified in the Poll frame. The timers start to count down DIFS (DCF Inter-Frame Space) after the Poll frame ends. The node whose timer expires earliest sends its packet. Other nodes detecting the channel to become busy cancel their timers. In Fig. 2 node \( A \) wins the contention and sends packet \( PA_U \) to the AP. If the AP correctly receives the packet it replies an ACK SIFS (Short Inter-Frame Space) later.

There are three abnormal cases: (i) it happens that all links experience fading and no node sends a packet; (ii) more than two nodes send packets simultaneously and a collision occurs; (iii) the AP fails to decode the packet due to bit errors. In any case, the AP broadcasts a new Poll frame. Then the transmission procedure is resumed.

The down-link is essentially a broadcast channel. To enable distributed multi-user scheduling, in the proposed scheme it is converted into a multi-access channel, as shown in Fig. 3. In addition to \( \gamma_0 \), the Poll frame that the AP broadcasts also carries a bit map indicating down-link traffic. When a node receives the Poll frame and finds that the AP has buffered a packet destined to it itself, it sends an invitation to the AP at a suitable timing. Specifically, if its normalized SNR is greater than \( \gamma_0 \), the node contends to send a CTS frame by starting a random waiting timer DIFS after the Poll frame ends. This gratuitous CTS is to tell the AP the willingness of receiving a packet, and its target address is set to the node’s MAC address, from which the AP knows the correct receiver. If only a single node transmits a CTS, there is no collision and the AP transmits the buffered data packet on request. In Fig. 3, the AP sends packet \( PA_D \) to \( A \) after the first CTS ends.

In the real network, both up-link and down-link traffic exists. Therefore the procedures in Figs. 2-3 are executed
The transmission is initiated by the AP with a Poll frame, which carries the SNR threshold $\gamma_0$ and a down-streaming traffic bit map. On receiving this Poll frame, each node measures its SNR and takes action if its normalized SNR is greater than the specified threshold. This action may be either transmission of a data packet to the AP or transmission of a CTS frame to initiate reception of a packet from the AP. In the normal case, only a single node talks with the AP.

Although the Poll frame initiates packet exchanges, it is not always necessary and it can be omitted if previous transmission is successful. In the up-link channel access in Fig. 2 each node overhears the ACK from the AP. In the down-link channel access in Fig. 3 each node overhears data packet from the AP. At the end of the heard frame each node measures its SNR. After the current transmission is finished, each node decides whether to join the next channel access contention according to two factors: (i) its normalized SNR, and, (ii) whether it has an up-link packet or its down-link packet has not been received yet. If the transmission is not successful or the channel remains idle because all links experience fading simultaneously, the AP broadcasts a new Poll frame after the contention slots. This Poll frame re-synchronizes the channel access sequence.

V. Optimal Parameter and Analysis

Performance of the proposed DMUS scheme depends on parameter $\gamma_0$. In this section we discuss how to set the value to optimize system performance. Consider the communication in the saturation situation where each node always has a packet to send to or receive from its associated AP. When the normalized SNR $\gamma_0$ is no less than the specified threshold $\gamma_0$, a node is allowed to access the channel in terms of slotted contention. Assume altogether there are $M$ nodes, and the maximal contention consists of $N$ slots with the period of each slot being $T_S$. For the simplicity of analysis we focus on the up-link (down-link transmission differs only in that it has an extra CTS).

To reduce the collision probability, $\gamma_0$ should be changed dynamically so that at any time only a single link has an instantaneous SNR greater than $\gamma_0$. This ideal case is actually impossible in a WLAN due to lack of CSI feedback. In the proposed scheme, the AP determines $\gamma_0$ so that $m$ out of $M$ nodes will have the chance to access channel. Collision probability due to contention among these $m$ nodes is reduced by random slotted waiting. The probability that $m$ links have the normalized SNR greater than $\gamma_0$ is

$$P_m(\gamma_0) = \begin{cases} 1, & m = 1, \\ \sum_{i=1}^{N-1} (N - i)^{m-1}/N^m, & m \geq 2. \end{cases}$$  (2)

There are three possible cases after this contention period.

(i) No links have normalized SNR greater than $\gamma_0$ and during the whole contention period the channel is idle. The probability is

$$P_E(\gamma_0) = P_0^T(\gamma_0),$$  (3)

and the overhead time is

$$t_E = t_{DIFS} + T_S \cdot N + t_{Poll},$$  (4)

where $t_{Poll}$ is the transmission time of the Poll frame and $t_{DIFS}$ is the duration of DIFS. Then the AP needs to re-broadcast the Poll frame for the purpose of SNR detection and initiating next contention period.

(ii) Exactly single link has normalized SNR greater than $\gamma_0$. There is no collision and the contention is successful. The probability is

$$P_S(\gamma_0) = \sum_{m \geq 1} P_T^T(\gamma_0)P_m(\gamma_0),$$  (5)

and the transmission time is

$$t_S = t_{DIFS} + t_{rand} + L/r(\gamma) + t_{SIFS} + t_{ACK},$$  (6)

where $t_{rand}$ is the random slotted waiting time before transmission, $L$ is the packet length and $r(\gamma)$ is the actual rate. $t_{SIFS}$ is the duration of SIFS and $t_{ACK}$ is the transmission time of the ACK frame.

(iii) More than two links have normalized SNR greater than $\gamma_0$ and a collision occurs. The probability is

$$P_C(\gamma_0) = \sum_{m=2}^{M} P_T^T(\gamma_0)(1 - P_m(\gamma_0)) = 1 - P_E(\gamma_0) - P_S(\gamma_0),$$  (7)

and the consumed time is

$$t_C = t_{DIFS} + t_{rand} + \max_{i \in S_c}[L_i/r(\gamma_i)],$$  (8)

where $S_c$ contains the packets that encountered collision and the channel is wasted for a period equaling the longest transmission time of these colliding packets.

Channel utility is defined as follows:

$$\eta(\gamma_0) = \frac{P_S(\gamma_0) t_S}{P_E(\gamma_0) t_E + P_S(\gamma_0) t_S + P_C(\gamma_0) t_C}.$$  (9)

The $\gamma_0$ that an AP selects should maximize channel utility $\eta(\gamma_0)$. It can be calculated by the following equation.

$$\gamma_0 = \arg \max_{\gamma_0} \eta(\gamma_0).$$  (10)

In the simplest case where $t_E \approx t_S \approx t_C$, $\eta(\gamma_0) \approx P_S(\gamma_0)$ and $\gamma_0$ can be selected by

$$\gamma_0 = \arg \max_{\gamma_0} P_S(\gamma_0).$$  (11)
γ₀ calculated by Eq.(11) only depends on M, the number of nodes associated with the AP, and N, the number of contention slots. For the simplicity of protocol design, we adopt Eq.(11) instead of Eq.(10). It is worth pointing out again that γ₀ is a normalized parameter and it does not rely on the absolute value of SNR. The AP advertises γ₀ in its periodic Poll frame.

Figures 4-5 respectively show the effect of N and M on \( P_S(\gamma_0) \) under different SNR thresholds \( \gamma_0 \). In Fig. 4, five curves correspond to \( N=1, 5, 10, 20 \) and 50 respectively. \( \gamma_0 \) where \( P_S(\gamma_0) \) reaches the peak, is large when \( N \) is small. This is to reduce collision. As \( N \) increases, maximal \( P_S(\gamma_0) \) also increases. However, the increase of \( P_S(\gamma_0) \) is gradually diminishing. In the following, \( N \) is chosen to be 20. In Fig. 5, as the number of nodes increases, the maximal \( P_S(\gamma_0) \) almost approaches a steady value. Meanwhile the \( \gamma_0 \), where \( P_S(\gamma_0) \) reaches the peak, gradually increases, corresponding to the effect of selection diversity.

**VI. Performance Evaluation**

In this section we evaluate the performance of the proposed DMUS scheme and compare it with CSMA/CA [3] and Contention Free (in Contention Free each node takes turn to transmit in a way similar to TDMA, exempt from collision). The same rate adaptation scheme is adopted in three schemes but multi-user diversity is only exploited in DMUS.

In the physical layer IEEE 802.11b parameters are used. In CSMA/CA and DMUS a receiving busy tone is necessary to make sure that hidden terminal problem does not occur. Contention Free is not pure TDMA and each transmission takes different time due to rate variations. For the convenience of simulation we developed a discrete event simulator by matlab.

In the simulation the AP is located at the center of a circle while a number of nodes are evenly distributed on the circle. The average SNR between a node and the AP is 13dB. It is assumed that each link experiences independent block Rayleigh fading. According to SNR of overheard frames, each node/AP performs rate adaptation based on an empirical SNR-rate table. By varying the number of nodes in the WLAN, we check the different performance trends of three schemes.

In the following we evaluate performance of multi-access up-link under saturation situations. Each node always has data ready to send to the AP. Each packet has a fixed size equaling 512 bytes. We consider the following three metrics: (i) channel utility defined in Eq.(9), actually calculated as the ratio of channel occupation time of successful transmissions to the whole simulation time, (ii) Total throughput of a WLAN, and, (iii) Jain’s fairness index in terms of long-term throughput, as shown in the following equation where \( r_i \) is the throughput of the \( i^{th} \) node.

\[
f(r_1, r_2, r_3, ..., r_n) = \frac{\left(\sum_{i=1}^{n} r_i\right)^2}{n \sum_{i=1}^{n} r_i^2}.
\]

Figures 6-7 show channel utility and throughput respectively. Since Contention Free has no contention at all, during nearly 80% of the time the channel is utilized for valid transmission. However, in Contention Free, each node takes turn to transmit to the AP. In times of fading, packets are transmitted at low rates or even transmission failure occurs. Contention cannot be completely removed in DMUS and channel utility of DMUS is less than that of Contention Free. However, in DMUS, only transmissions over links with high quality are allowed. Since packets usually are transmitted at higher rates in DMUS than in Contention Free, DMUS has a much higher throughput, as reflected in Fig. 7.

Channel utility reaches its maximum in three schemes when there are merely two nodes. Throughput of Contention Free and CSMA/CA also reaches the peak at the same condition. As the number of nodes further increases, the multi-user diversity due to new users diminishes in DMUS and the throughput approaches a steady value. Channel utility of DMUS also approaches steadiness. In contrast, due to severe collision, channel utility and throughput of CSMA/CA decrease quickly. Though Contention Free has much higher throughput than CSMA/CA, its throughput is much less than that of DMUS, owing to the effect of fading. When the number of nodes increases to 40, DMUS improves the cell throughput by 194.6% compared with CSMA/CA, and by 49.6% compared with Contention-Free.

Figure 8 shows long-term fairness of three schemes. Fairness decreases a little as the number of nodes increases.
Though DMUS cannot achieve nearly perfect fairness as Contention Free, its fairness is much higher than that of CSMA/CA. Decrease in its fairness can be attributed to remaining collisions in the contention and imperfect randomness of channel state.

VII. CONCLUSION AND FUTURE WORK

CSMA/CA is the de facto standard of wireless channel access method. However its performance is limited by the number of associated nodes and multi-path fading. We have shown that though either factor may be harmful, a joint consideration may improve the system performance. In the proposed Distributed Multi-User Scheduling (DMUS) scheme, by only letting nodes with high instantaneous link quality contend for the channel, collision is mitigated, throughput is improved and long-term fairness is retained. The optimal parameter is also given by the analysis. Simulation results indicate that the proposed DMUS scheme can improve the throughput by up to 194.6% compared with CSMA/CA and by up to 49.6% compared with Contention-Free. In the future we will enhance DMUS with proportional fair scheduling and service differentiation, and apply it to WLAN and CSMA/CA based inter-vehicle communications.

ACKNOWLEDGMENT

This work was supported by National Institute of Information and Communications Technology (NICT) of Japan.

REFERENCES