Two-phase Collision Avoidance to Improve Scalability in Wireless LANs

Seongil Han*, Yongsub Nam*, Yongho Seok†, Taekyoung Kwon* and Yanghee Choi*
*School of Computer Science and Engineering
Seoul National University, Seoul, Korea
Email: {sihan, ysnam, tk, yhchoi}@mmlab.snu.ac.kr
†Project - PLANETE
INRIA Sophia Antipolis, France
Email: Yongho.Seok@sophia.inria.fr

***Abstract*— IEEE 802.11 DCF exhibits poor scalability due to the large contention overhead. Therefore, the more the number of stations, the less the aggregate throughput. We propose a two-phase collision avoidance scheme to reduce the collision probability and to enhance the throughput performance. In our proposed scheme, contention among stations is resolved in two phases: SuperSlots and SubSlots. Also, our truncated backoff mechanism increases the throughput by reducing the idle time slots. We analyze the performance of our proposed scheme based on the previous analysis of 802.11 DCF. Both the analysis and simulation results exhibit that our proposed scheme achieves higher throughput than the current IEEE 802.11 backoff mechanism, and this differential increases with the number of stations in the network.**

***Index Terms*— IEEE 802.11 Distributed Coordination Function (DCF); Two-phase collision avoidance; SuperSlot; SubSlot; Truncated backoff;**

I. Introduction

Currently, the IEEE 802.11 protocol [1] is the most popular wireless LAN technology on the market. Based on the Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) mechanism, the IEEE 802.11 Distributed Coordination Function (DCF) is deemed scalable to the number of stations because of the binary exponential backoff mechanism. The theoretical maximum throughput of IEEE 802.11 is calculated in [3], which is only about 85% of the link bandwidth, due to the control and the contention overhead. And the actual throughput is substantially reduced as the number of stations increases.

In the CSMA/CA scheme, a contending station freezes the backoff timer when it senses that the medium is being used by another station. Then, if the station senses that the medium is idle after DCF Interframe Space (DIFS), it resumes decreasing the backoff counter. However, if two stations happen to have the same backoff counter, both counters will become zero at the same time, which will bring about a collision. Hence, the probability of such collision is the main obstacle to increasing throughput as the number of stations grows.

RTS/CTS exchange can reduce the collision probability by setting the Network Allocation Vector (NAV) for a specified duration, which is referred to as the virtual carrier sensing. However, RTS/CTS frames are transmitted in one of the basic rate set, which degrades the throughput performance, although it reduces the collision probability.

We propose a two-phase collision avoidance scheme that reduces the collision probability, and hence increases the throughput. In this paper, we first introduce the related work in Section II, and we detail our proposed scheme in Section III. Then, the analysis model of our proposed scheme is shown in Section IV, followed by the numerical results in Section V. Finally concluding remarks are given in Section VI.

II. Related Work

A number of research efforts have been made to reduce the collision probability of IEEE 802.11 so far. [4], [5] consider a separate control channel to schedule the channel access. Although these protocols can coordinate the transmission order a priori, they require additional hardware cost and complexity. Furthermore, the coverage of the busy tone signal may not be the same as that of the data transmission, which can bring about a hidden terminal problem.

Fast Collision Resolution (FCR) [6] is also proposed to reduce the collision probability. With FCR, when a new busy medium is detected, all stations performing the backoff process increase their own contention window exponentially and decide a new backoff counter. Since FCR reduces the collision probability by making one of the contending stations occupy the channel for a certain period of time, the unfairness issue arises. To overcome this problem, FCR additionally adopts the distributed SCFQ algorithm to dynamically adjust the maximum transmission limit of each station. However, FCR is expected to show worse performance as the network becomes saturated.

GDCF [7] is tailored to find the optimal contention window size to reduce the collision probability. Under GDCF, each station does not set the contention window size into the minimum size when the transmission is successful. Instead, GDCF halves the current contention window size after c consecutive successful transmissions. By this gentle approach, GDCF decreases the collision probability. However, since the

This work was supported by the Brain Korea 21 project of the Ministry of Education, 2006.

1-4244-0357-X/06/$20.00 ©2006 IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2006 proceedings.

c value cannot adapt to the network condition, it should be well-determined a priori.

Early Backoff Announcement (EBA) [8] protocol is also designed to reduce the collision probability, by determining the transmission order a priori. Under EBA, a station announces its future backoff information, such as the number of backoff slots, within its MAC frame header. All other stations receiving the frame can avoid collisions by excluding the same backoff counter when choosing their future backoff counter. However, because the schedule information is transmitted in the frame header, the channel error or the hidden terminal problem can severely degrade the network performance.

III. Proposed Scheme

Our scheme comprises two mechanisms: two-phase collision avoidance and truncated backoff. The former reduces the collision probability by handling the collision with two-phase hierarchy, and the latter reduces the idle time slots due to the backoff process.

A. *Two-phase Collision Avoidance*

In our two-phase collision avoidance scheme, the slots are structured into a two-level hierarchy: *SubSlot*s and *SuperSlot*s. A SubSlot is identical in length to an IEEE 802.11 slot, and several SubSlots constitute a SuperSlot. The SuperSlot is the basic unit in the backoff process, which means that the size of the contention window is a multiple of the size of a SuperSlot. This is shown in Fig. 1, where a SubSlot is $20\mu s$ long, and a SuperSlot is $80\mu s$ long; a SuperSlot is composed of 4 SubSlots in this case.

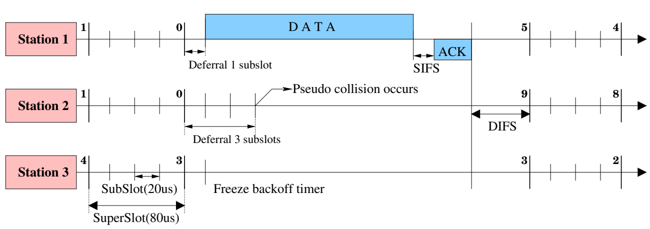


Fig. 1. Slot structure for two-phase collision avoidance

If a station has a frame to send, it randomly chooses a backoff counter among its contention window range and starts the backoff process. This backoff process is the same as that of DCF, except the slot size and the contention window size. In our scheme, the contention window size depends on the size of the SuperSlot. Let D denote the number of SubSlots in a SuperSlot, i.e., D is 4 in the above case. Then, the contention window size is calculated as follows, where CW_{DCF} denotes the contention window size of DCF in terms of an IEEE 802.11 slot.

$$\mathrm{CW} = \frac{\mathrm{CW_{DCF}} + 1}{\mathrm{D}} - 1.$$

The backoff counter is randomly selected in the range of zero and the contention window size, as in the equation below, where $\sigma_{SuperSlot}$ denotes the time duration of the SuperSlot. For example, if the current CW_{DCF} is 31, then an arbitrary number of SuperSlots between 0 and 7 will be chosen.

$$\mathrm{BackoffTime} = \sigma_{\mathrm{SuperSlot}} \times \mathrm{random}[0, \mathrm{CW}].$$

When the backoff timer reaches zero, the station selects a random number of SubSlots within the chosen SuperSlot (i.e., in the range 0∼3 in Fig. 1). The deferring time is calculated as follows, where $\sigma_{SubSlot}$ stands for the time duration of the SubSlot.

$$\mathrm{DeferringTime} = \sigma_{\mathrm{SubSlot}} \times \mathrm{random}[0, \mathrm{D} - 1].$$

The station now waits for the chosen deferring time, while continuing to sense the channel; this is referred to as a *deferral process*. When the deferral timer expires, the station transmits the frame. During the deferral process, no freezing mechanism is employed, which means that, if a station senses that the channel becomes busy before its timer reaches zero, it does not freeze the deferral timer but regards this situation as a collision, even though no actual collision occurs. We call this a *pseudo collision*. This pseudo collision does not affect another station which has already completed the deferral process and now transmits a frame. Meanwhile, the station that detects a pseudo collision doubles its contention window size and restarts the backoff process, as in the case of the actual collision.

In Fig. 1, the backoff timers of Station 1 and Station 2 reach zero at the same time, and Station 1 chooses a single SubSlot for the deferral process, while Station 2 chooses 3 SubSlots. Station 1 waits for a single SubSlot interval and then transmits a frame, while Station 2 discovers that the medium has become busy while sensing the channel (i.e., a pseudo collision occurs), doubles its contention window size, and restarts the backoff process. Meanwhile, Station 1 is unaware of the pseudo collision, and starts transmitting a frame. Overall, contention among stations is thus handled in two phases: the backoff process in SuperSlots and the deferral process in SubSlots.

B. *Truncated Backoff Scheme*

Since the unit interval used by the backoff timer is the length of a SuperSlot, that length is significant. Suppose a station has a frame to send. If that station senses that the channel has become busy while waiting to decrement the backoff timer, it freezes the timer and only resumes the backoff process when the medium becomes idle again.

When the station resumes with the frozen timer value, the value will be decremented by one after each SuperSlot so long as there is no transmission in the SuperSlot. If a busy medium is detected in the SuperSlot, the waiting time will be wasted because it cannot decrement the backoff timer. This wasted time will not significantly affect the performance in the case of 802.11 DCF, because the slot time is very short. But the length of a SuperSlot is relatively large, so that the performance of our scheme may be degraded by this wasted time, particularly when a small number of stations are in contention.

1-4244-0357-X/06/$20.00 ©2006 IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2006 proceedings.

We therefore decide that the backoff timer should be truncated by one SuperSlot immediately after it senses that the channel is idle and waits for a DIFS period, as shown in Fig. 2. With this aggressive approach, we can achieve higher throughput, particularly when the number of stations is small.

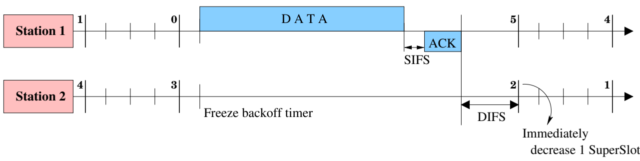


Fig. 2. Example of truncated backoff scheme

At the moment, when a station senses that the medium has become idle and is waiting for DIFS, the backoff timer of every station is greater than or equal to one. (Any station whose backoff timer was zero would either transmit a frame, or detect a pseudo collision and invoke the backoff process.) Thus, we can decrement the backoff timer of each station by one without increasing the collision probability, which will reduce the wasted time due to idle slots. One possible exception is that a station receives a frame from the above layer which has completed the post-backoff process. In this case, the station will transmit the frame right after DIFS, which can increase the collision probability. However, it is negligible considering the gain from the truncated backoff.

EDCA in [2] also defines the similar backoff process to our truncated backoff. Under EDCA, each station decrements the backoff counter before carrier sensing, but initiates the frame transmission after an additional idle slot when the backoff counter reaches zero. The main difference is that EDCA will show better throughput performance than IEEE 802.11 DCF when the network is heavily loaded, while our truncated backoff is tailored to achieve better performance even when the network is lightly loaded. When the network is heavily loaded, however, our scheme will show better performance than EDCA, owing to the two-phase collision avoidance mechanism.

IV. Throughput Analysis

The stochastic model of our scheme is based on the Markov chain model in [3]. If we do not take the deferral process into account, our model is very similar to that in [3]. The only difference is that, our Markov chain has states in units of one SuperSlot time, while in [3], in units of one slot time. However, the deferral process requires additional states. In DCF, when the backoff counter of a station reaches zero, the station transmits a frame. If the transmission is successful (with probability of $1-p$), the state transits to to the initial backoff stage; here, p is the conditional collision probability. Otherwise, the state transits to the next backoff stage. In our scheme, on the other hand, when the backoff counter of a station is decremented to zero, the station starts the deferral process. During the deferral process, the state transfers from the uniformly selected deferral counter value to zero. When the deferral counter reaches zero, a frame is transmitted and the state is changed to the initial backoff stage. On the other hand, if a collision (either pseudo or actual) occurs during the deferral process, the state transfers to the next backoff stage.

We modify the equations in [3] to model the deferral process. In the following equation, τ is the probability that a station begins the deferral process. Note that τ is same as the transmission probability of [3]; unless otherwise mentioned, we borrow the notation from [3] for convenience.

$$\tau = \frac{2(1-2p)}{(1-2p)(W+1)+Wp(1-(2p)^m)}.$$

In the above equation, W and m are the contention window size in SuperSlots and the maximum backoff stage, respectively. But, p is modified to represent the conditional probability of both the actual and the pseudo collisions. This is because the pseudo collision is regarded same as the actual collision in our scheme.

Let D be the number of SubSlots in a SuperSlot and i be the number of stations whose backoff counter reaches zero. If a station randomly chooses to start a transmission at the $j^{th}(0 \le j \le D-1)$ SubSlot in the chosen SuperSlot, an actual collision occurs when one or more stations among i stations choose j^{th} SubSlot to transmit a frame. If one or more stations select their SubSlots in the range of 0 to $j-1$, a pseudo collision occurs. Let p_i be the collision (either pseudo or actual) probability when there are i contending stations. Then, p_i is calculated as follows.

$$p_i = \sum_{j=0}^{D-1} \frac{1}{D}\left(1-\left(\frac{D-j-1}{D}\right)^i\right).$$

Now, the conditional collision probability, p, is expressed as follows, when there are n stations in the BSS. Note that the transmitting station is excluded from the summation.

$$p = \sum_{i=1}^{n-1} \binom{n-1}{i} \tau^i (1-\tau)^{n-1-i} \cdot p_i.$$

The probability that i stations among n stations choose the same SuperSlot for a transmission is $\binom{n}{i}\tau^i(1-\tau)^{n-i}$. Now, let $P_{tr}(j)$ be the probability that there is at least one transmission in the j^{th} SubSlot of the given SuperSlot. $P_{tr}(j)$ is obtained by the following equation, where all i stations select j^{th} to $D-1^{th}$ SubSlot and at least one station selects j^{th} SubSlot.

$$P_{tr}(j) = \sum_{i=1}^{n} \binom{n}{i}\tau^i(1-\tau)^{n-i} \cdot \left(\left(\frac{D-j}{D}\right)^i - \left(\frac{D-j-1}{D}\right)^i\right).$$

Also, let $(P_{tr}P_s)(j)$ be the probability that a successful transmission occurs in the j^{th} SubSlot of the given SuperSlot. Then, $(P_{tr}P_s)(j)$ is obtained by the similar manner. However, only one station among i stations selects j^{th} SubSlot in this case to have a successful transmission in j^{th} SubSlot.

$$(P_{tr}P_s)(j) = \sum_{i=1}^{n} \binom{n}{i}\tau^i(1-\tau)^{n-i} \cdot \binom{i}{1}\frac{1}{D}\left(\frac{D-j-1}{D}\right)^{i-1}.$$

Now, let P_{tr} and $P_{tr}P_s$ be the probability that there is at least one transmission at a given moment, and the probability

1-4244-0357-X/06/$20.00 ©2006 IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2006 proceedings.

TABLE I
MAC AND PHY PARAMETERS

SIFS	10 μs
DIFS	50 μs
EIFS	364 μs
σ	20 μs
Propagation delay	1 μs
BasicRate	2 Mbps
DataRate	11 Mbps
PLCP length	192 bits @ 1 Mbps
MAC header (ACK, Data)	(14, 28) bytes @ BasicRate
(CW_{min}, CW_{max})	(31, 1023)

that a single (successful) transmission is under way at a given moment, respectively. They are obtained by summing up the probabilities of each possible case.

$$P_{tr} = \sum_{j=0}^{D-1} P_{tr}(j)$$
$$P_{tr}P_s = \sum_{j=0}^{D-1} (P_{tr}P_s)(j).$$

The aggregate throughput ratio, S, is expressed as follows, where Π_i, Π_s, and Π_c denote the average idle slot time, the average time spent for the successful transmission, and the average time spent in the collision, in a slot time, respectively. Also, $E[P]$ is the average payload size of a MAC frame.

$$S = \frac{P_{tr}P_s \cdot E[P]}{\Pi_i + \Pi_s + \Pi_c}.$$

Π_i, Π_s, and Π_c are the same as those in [3] except that we add the average deferring time $(j \cdot \sigma)$ to T_s and T_c, which are the average time spent for the successful transmission of a frame, and the average time spent in the collision, respectively. Here, σ denotes the duration of a SubSlot. Recall that the duration of a SubSlot is equal to the slot time of IEEE 802.11 (e.g., in IEEE 802.11b, 20 μs).

$$\Pi_i = (1 - P_{tr}) \cdot D \cdot \sigma$$
$$\Pi_s = \sum_{j=0}^{D-1} (P_{tr}P_s)(j) \cdot (j \cdot \sigma + T_s)$$
$$\Pi_c = \sum_{j=0}^{D-1} \{P_{tr}(j) - (P_{tr}P_s)(j)\} \cdot (j \cdot \sigma + T_c).$$

T_s and T_c are the same as those in [3] except that we put EIFS for T_c instead of DIFS. In the equations below, H, $E[P*]$, and δ represent the header overhead (both PHY and MAC), the average transmission time for the longer frame involved in collision and the propagation delay, respectively.

$$T_s = H + E[P*] + SIFS + ACK + DIFS + 2\delta$$
$$T_c = H + E[P*] + EIFS + \delta.$$

We conduct our analysis using Mathematica 4. We use MAC and PHY parameters in Table 1. The aggregate throughput ratio, S, obtained by our analysis is compared with our simulation result in Fig. 3(c); the simulation scenario is given in the next section.

V. NUMERICAL RESULTS

We evaluated our scheme in terms of the collision probability, the throughput, and the fairness by performing simulations using NS-2. These experiments show the effect of the number of stations and the SuperSlot size. The frame size is 1500 bytes and a saturated traffic model is used, in which each station always has a frame to send and contends for the channel access whenever the medium is idle. The SubSlot length is set to 20 μs, which is identical to the slot length of IEEE 802.11b.

Fig. 3(a) depicts the actual collision probability for each scheme as the number of stations increases. When there are many stations, the collision becomes dominant with IEEE 802.11b, while our scheme performs better and better as the length of the SuperSlot increases. In Fig. 3(a), DCF shows the collision probability of 0.3 when there are 100 stations in the BSS, while our scheme with 160 μs SuperSlot exhibits the collision probability below 0.15, which is about half of DCF. The decrease in the collision probability is directly related to the throughput performance, as shown in Fig. 3(b).

The curves in Fig. 3(b) show aggregate throughput of the proposed scheme and of IEEE 802.11b as the number of stations increases. As expected, throughput decreases as the number of stations increases. But the performance of the proposed scheme decreases more slowly when the SuperSlot is made larger. Also, the relative advantage of our scheme increases with the number of stations.

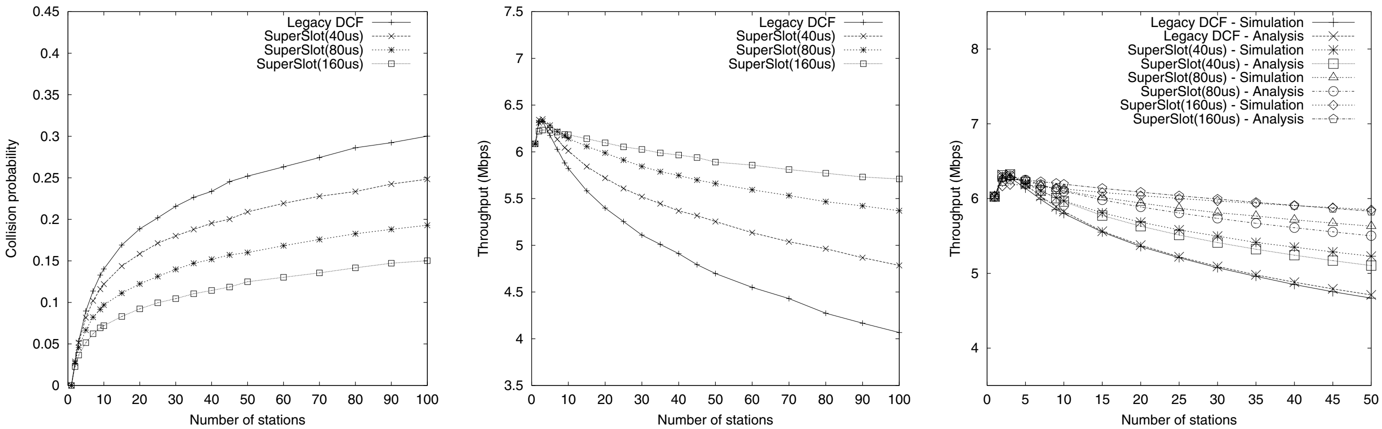
However, when the number of stations is vary small (fewer than five), the proposed scheme achieves slightly less throughput than IEEE 802.11b. This is because the length of the SuperSlot is larger than the slot time of the IEEE 802.11. In this case, the larger the size of the SuperSlot is, the less the aggregate throughput is. However, this difference is negligible. Also, when the number of stations is above five, the larger the SuperSlot size, the more the aggregate throughput.

Figs. 4(a) and 4(b) show the results of the short term fairness and the long term fairness, respectively. We use the fairness index defined in [9]:

$$\text{Fairness Index} = \frac{(\sum_{\rm i} {\rm T_i}/\phi_{\rm i})^2}{{\rm n}\sum_{\rm i} ({\rm T_i}/\phi_{\rm i})^2},$$

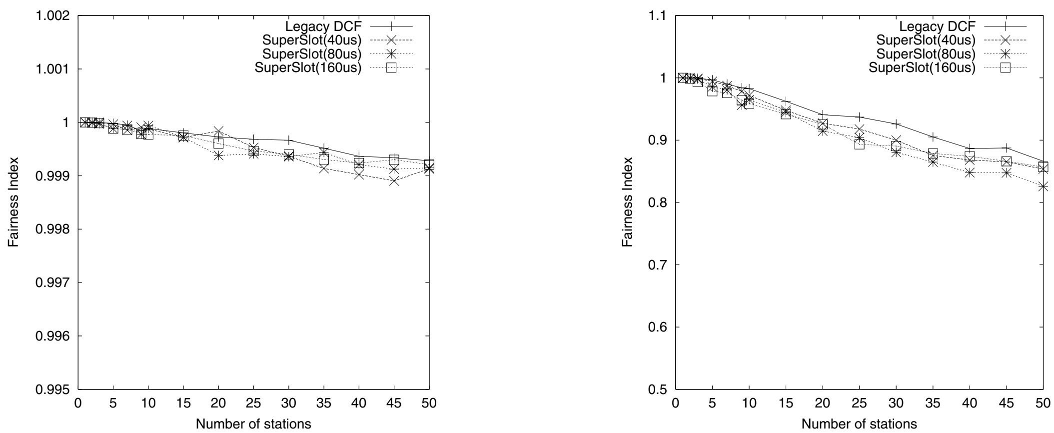
where n, T_i, and ϕ_i denote the number of stations, the throughput of the flow i, and the weight of the flow i, respectively. Here, we assume all the stations have the same weight. The fairness index ranges between 0 and 1, and 1 means that each station achieves exactly same throughput. We measure and average the short-term fairness every second; in case of the long-term fairness, we measure it every 100 seconds. In Fig. 4(a), the long-term fairness index values of DCF and the SuperSlots lie between 0.999 and 1, which means both protocols achieve almost fair throughput among stations. In terms of the short-term fairness, both DCF and our scheme show decreasing fairness performance with the number of stations.

1-4244-0357-X/06/$20.00 ©2006 IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2006 proceedings.



(a) Simulated collision probability (b) Simulated throughput (c) Throughput comparison

Fig. 3. Comparison of throughput and collision probability



(a) Fairness index (long-term) (b) Fairness index (short-term)

Fig. 4. Comparison of fairness index

VI. CONCLUSION

This paper has introduced a collision avoidance scheme which improves the scalability of IEEE 802.11. The proposed scheme consists of two mechanisms: two-phase collision avoidance and truncated backoff. The former reduces the collision probability by handling the collision with two-phase hierarchy. And the latter reduces the idle time due to the overhead of the SuperSlot. Both simulation and analysis results show that the proposed scheme increasingly outperforms IEEE 802.11 DCF as the number of stations grows, while maintaining comparable fairness.

REFERENCES

[1] IEEE Computer Society. 802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, June 1997.

[2] IEEE Computer Society. 802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements (802.11e), Nov. 2005.

[3] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," IEEE Journal of Selected Areas Communications, vol. 18, pp. 535-547, March 2000.

[4] X. Yang and N. H. Vaidya, "Explicit and Implicit Pipelining for Wireless Medium Access Control," Proc. IEEE Vehicular Technology Conference, Fall 2003.

[5] A. Tzamaloukas and J. J. Garcia-Luna-Aceves, "A Receiver-Initiated Collision-Avoidance Protocol for MultiChannel Networks," Proc. IEEE INFOCOM '01, 2001.

[6] Y. Kwon, Y. Fang and H. Latchman, "A Novel MAC Protocol with Fast Collision Resolution for Wireless LANs," Proc. IEEE INFOCOM '03, 2003.

[7] C. Wang, W. Tang, K. Sohraby and B. Li, "A Simple Mechanism on MAC Layer to Improve the Performance of IEEE 802.11 DCF," Proc. First International Conference on Broadband Networks '04, 2004.

[8] J. Choi, J. Yoo, S. Choi and C. Kim, "EBA: An Enhancement of the IEEE 802.11 DCF via Distributed Reservation," IEEE Transactions on Mobile Computing, Vol. 4, Issue 4, pp. 378-390, July-Aug. 2005.

[9] R. Jain, A. Durresi and G. Babic, "Throughput Fairness Index: An Explanation," ATM Forum/99-0045, Frb. 1999.

1-4244-0357-X/06/$20.00 ©2006 IEEE
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2006 proceedings.