Performance Improvement for 802.11 Based Wireless Local Area Networks

Liang Zhang and Yantai Shu
Department of Computer Science
Tianjin University
Tianjin 300072, China
td_zl@hotmail.com

Oliver W.W. Yang
School of Information Technology and Engineering
University of Ottawa
Ottawa, Ontario, Canada, K1N 6N5
yang@site.uottawa.ca

Abstract - In typical installation of an 802.11 based wireless local area network (WLAN), mobile hosts would access the network through Access Points, even when two mobile stations in the same WLAN communicate. That is, all the packets in a WLAN are required to forward through the AP. Since the AP has the same priority as the other mobile stations to access the channel, according to the Medium Access Control Layer (MAC) protocol, the AP usually becomes a bottleneck in WLANs and the network performance degrades significantly. In this paper, we propose a new MAC layer protocol for WLANs in order to improve the throughput performance. Theoretical analysis and simulation results show that our new protocol works much better in WLAN than the standard DCF.

Keywords - Wireless Local Area Network, Access Point, DCF

I. INTRODUCTION

In recent years, wireless networks based on the IEEE 802.11 standard are becoming increasingly prevalent. IEEE 802.11 Wireless Local Area Network (WLAN) [1] in particular has gained a dominant share in the market among various emerging technologies for broadband wireless access. The focus is now turning to deploying these networks over hot spots such as airports, hotels, cafes, and other areas from which people can have public access to the Internet.

The IEEE 802.11 standard [1] provides a detailed MAC (Medium Access Control) and PHY (PHYSical) layer specification for WLANs. The IEEE 802.11 MAC provides an access to the shared wireless medium through two different access mechanisms: a mandatory random access protocol called DCF (the Distributed Coordination Function), and an optional polling-based protocol called the Point Coordination Function (PCF). DCF is the fundamental mechanism to access the shared medium. It is a random access method based on the carrier sense multiple access with collision avoidance (CSMA/CA) with a binary slotted exponential backoff mechanism. We focus the DCF in this paper.

Wireless networks can be used either to replace wired networks, or as an extension of the wired network infrastructure. There are two basic wireless network topologies: infrastructure WLANs and ad hoc network.

In ad hoc network, mobile stations communicate directly, if they can hear each other. This type of network is often formed on a temporary basis, and is commonly referred to as an Independent Basic Service Set (IBSS).

On the contrary, the basic building block of an IEEE 802.11 WLAN deployment is a BSS, which is composed of an access point (AP) and multiple stations associated with the AP. A Basic Service Set (BSS) consists of two or more wireless nodes, or stations that have recognized and established communications with each other. The main function of an AP is to form a bridge between wireless and wired LANs. The AP is analogous to a base station used in cellular phone networks. Stations do not communicate directly because all communications between stations, or between a station and a wired network node, have to go through the AP. The AP is not mobile, and form part of the wired network infrastructure.

In recent years, there are many interesting investigations on the performance evaluation of IEEE 802.11 and similar protocols [2-6]. Most of these studies are largely confined to Ad hoc network. In this paper we shall focus on WLANs where every packet needs the AP to forward even when the sender and receiver are all mobile stations in the same WLAN and close enough to hear with each other. Since the AP just has the same priority with other mobile stations to access the channel, it becomes a bottleneck and degrades the network throughput performance significantly. Furthermore, compared with a wired network, the limited bandwidth resource in a WLAN is already a big problem, and the AP forwarding simply makes it worse. In this paper, we propose an extension to the standard DCF protocol in WLANs by giving the AP a higher priority to access the channel. We expect our new protocol can improve the throughput performance and the channel utilization greatly, and we shall demonstrate this by our performance evaluation via theoretical analysis and simulation.

The rest of this paper is organized as follows. Section II presents the overview of the DCF of IEEE 802.11 MAC protocols. Section III describes our new protocol in detail. Section IV presents the performance evaluation. The simulation results and a conclusion are presented in Sections V and VI respectively.

For the remainder of the paper, the following symbols and notations are as follow:

- $n$: The number of current transmitting stations.
- $CW_{\text{min}}, CW_{\text{max}}$: The minimum and maximum contention window size as defined in IEEE 802.11 standard.
- $W$: The contention window size of $i^{\text{th}}$ backoff stage, it equals to $2^i W$, where $W = (CW_{\text{min}} + 1)$.
- $m$: The maximum backoff stage.
The probability that a station transmits in a randomly chosen slot time. After each unsuccessful transmission, a DIFS. The station transmits when the backoff time reaches a maximum value $2^{\tau}$. If the source is one of the mobile stations, it will increase the backoff time counter by one (except when the backoff time counter is already at its maximum) and then reactivate the decreasing process. The station transmits when the backoff time reaches zero.

### III. DCF WITH AP HIGH PRIORITY

In WLANs, the main function of an AP is to form a bridge between wireless and wired LANs. Stations do not communicate directly because all communications between stations, or between a station and a wired network node, have to go through the AP. Therefore, every packet needs the AP to forward in WLANs even when the sender and receiver are all mobile stations in the same WLAN and close enough to hear with each other. However, since the AP just has the same priority with other mobile stations to access the channel, it has now become a bottleneck and would degrade the network throughput performance significantly.

For example, we assume there are $n$ pairs of mobile stations in the WLAN. Each pair consists of one mobile receiver (MR) and one mobile sender (MS). That is to say, there are $(n+1)$ active stations ($n$ MS plus the AP) in the WLAN where an active station is a station that has data to send.

Assuming all the active stations are in saturation mode, i.e., they always have data to send, the AP can achieve at most $1/(n+1)$ of the channel rate due to its same priority as other stations. Also each MR can obtain on the average $1/n$ of the AP’s throughput and $1/n(n+1)$ of the channel rate.

So the bandwidth achieved by every MR is in inverse proportion to $n$ approximately. That is why the throughput performance will be degraded seriously, when the number of active stations increases.

In our opinion, as every packet needs the AP to forward in WLANs, the AP should take at least half of the total bandwidth. In order to achieve this, we propose a new MAC layer protocol for 802.11 based WLANs, which can give the AP higher priority to access the channel so as to improve the throughput performance and the channel utilization.

In our new protocol, we can make a simple change to the decreasing rule of the backoff time counter in order to give the AP a higher priority. We know that when a mobile station gets the channel, it must be transmitting packet to the AP first. So when the AP receives a data packet correctly, it will transmit an ACK to indicate the successful reception after a SIFS interval. Both the mobile stations and the AP then take the following actions.

Action for mobile stations

After a DIFS medium idle time, the mobile station shall generate a random backoff period for an additional deferral time before transmitting. The backoff time counter is decremented while the medium is sensed idle, and frozen when a transmission is detected on the channel, and reactivated when the channel is sensed idle again for more than a DIFS. The station transmits when the backoff time reaches zero. After each unsuccessful transmission, $w$ is doubled, up to a maximum value $2^{m'}W = (CW_{\text{min}}+1)$, where $W$ equals to $(CW_{\text{max}}+1)$.

When the destination receives a data frame correctly, it will transmit an ACK to indicate the successful reception after the Short Inter-Frame Space (SIFS) interval. If the source station does not receive the ACK within a specified ACK_Timeout, the data is assumed to be lost and the source station schedules the retransmission.
Action for the AP

After a DIFS medium idle time, the AP also generates a random backoff period for an additional deferral time before transmitting. The backoff time counter is decremented while the medium is sensed idle. If the AP’s backoff counter reaches zero, it should send the data. After a successful transmission, the AP would begin the next contention process as already defined in the standard. If others obtain the channel during the AP’s backoff time, the AP should be the receiver, since it must be involved in any communication. When the transmission finished, the AP will set its backoff time counter be zero and reactivates the backoff time counter when the channel is sensed idle again for more than a DIFS. At this time, since its backoff time counter is zero, the AP should obtain the channel immediately and send the data. As all other mobile stations has increased their backoff time counter by one after the communication, there should not be any collision occur. As a result, the AP can now send one packet, after any mobile stations sending a packet. In this way, the AP with higher priority can obtain at least half of the bandwidth in the WLAN and will not be the bottleneck anymore.

IV. PERFORMANCE ANALYSIS

In this section, we propose a two-dimensional Markov chain model to evaluate the performance of our scheme. First, we will describe our model, which we use to get the stationary probability \( \tau \) that the station transmit a packet. In addition, we assume a fixed number \( n \) of mobile stations, and each always having a packet available for transmission.

A. Markov Chain Model

Let \( b(t) \) be the stochastic process representing the backoff-time counter for a given station at slot time \( t \). At the beginning of each slot time, if the medium is determined to be busy, the backoff time counter of each station is frozen; otherwise, the counter is decremented. Let \( s(t) \) be the stochastic process representing the backoff stage of the station at time \( t \). Also let \( q(t) \) be the probability that at least one of the \( (n - 1) \) remaining mobile stations would transmit. As an important approximation, if we now assume \( q \) is constant and independent each other. Then the two-dimensional process \( \{s(t), b(t)\} \) forms a discrete-time Markov chain shown in Fig.1.

In this Markov chain, let \( m \) be the maximum backoff stage (which also means the maximum retransmission count), and let \( m' \) be the maximum count of the exponential increase of \( CW \) from \( CW_{\text{min}} + 1 \) to \( CW_{\text{max}} + 1 \). Note that the value of \( m \) may be larger than that of \( m' \). As represented in Fig.1, once \( CW \) reaches \( CW_{\text{max}} + 1 \), the CW shall remain at the value until it is reset. Therefore, we have

\[
\begin{align*}
W_i &= 2^i W \quad i \leq m' \\
W_i &= 2^i W \quad i > m'
\end{align*}
\]

where \( W = (CW_{\text{min}} + 1) \), and \( 2^n W = (CW_{\text{max}} + 1) \).

The only non-null one-step transition probabilities of this Markov chain are\(^1\)

\[
\begin{align*}
\Pr\{i, k, i, k-1\} &= q_i \quad k \in [2, W_i + 1] \quad i \in [0, m] \\
\Pr\{i, W_i - 1, i, W_i - 1\} &= q_i \quad i \in [0, m] \\
\Pr\{i, k, i, k+1\} &= 1 - q_i \quad k \in [0, W_i - 2] \quad i \in [0, m] \\
\Pr\{0, k, i, 0\} &= (1 - q_i) / W_0 \quad k \in [0, W_0 - 1] \quad i \in [0, m] - 1 \quad (2) \\
\Pr\{i, k, i-1, 0\} &= q_i / W_i \quad k \in [0, W_i - 1] \quad i \in [1, m] \\
\Pr\{0, k, |m, 0\} &= 1 / W_0 \quad k \in [0, W_0 - 1]
\end{align*}
\]

These transition probabilities account for 1) the fact that the backoff time decrement is stopped and the backoff counter increases by one (except when it is already at its maximum), when the channel is sensed busy; 2) the fact that at the beginning of each slot time the backoff time is decremented; 3) the fact that a new packet following a successful packet transmission starts at backoff stage 0, and thus the backoff is initialized uniformly chosen in the range \([0, W_0]\); 4) the situation when an unsuccessful transmission occurs at backoff stage \( i-1 \), the backoff stage increases, and the new initial backoff value is uniformly chosen in the range \([0, W_i - 1]\); 5) the need to reset \( CW \) to \( CW_{\text{min}} \) at the maximum backoff stage, and to restart the backoff stage for a new frame after.

Let \( b_{i,k} = \lim_{t \to +\infty} \Pr\{s(t) = i, b(t) = k\} \) be the stationary distribution of Markov chain. First we note that

\[
b_{i-1,0} \cdot q = b_{i,0} \quad 0 < i \leq m \quad , \quad (3)
\]

from which we can show easily

\[
b_{i,0} = q \cdot b_{i,0} \quad 0 \leq i \leq m \
\]

(4)

According to the global balance equations and transitions in the chain, we have

\[
b_{0, W_i - 1} = \frac{1}{1 - q} \left( (1 - q) \sum_{j=0}^{m-1} b_{j,0} W_j + b_{m,0} W_0 + q \cdot b_{0, W_i - 2} \right) \
\]

\(\text{Figure 1. Markov chain model of backoff window scheme}\)

\(1\) \( P\{i, k, |i, k\} = P\{s(t+1) = i, b(t+1) = k\} \mid s(t) = i, b(t) = k\} \).
\[ b_{0,k} = \frac{(1-q)\sum_{j=0}^{m-1} b_{j,0} + b_{m,0}}{W_0} \quad (6) \]
\[ b_{0,k+1} (1-q) + b_{0,k-1} q \quad 1 < k < W_0 - 1 \]
\[ b_{0,1} = \frac{(1-q)\sum_{j=0}^{m-1} b_{j,0} + b_{m,0}}{W_0} + b_{0,2} (1-q) \quad (7) \]

So according (5), (6) and (7), we have
\[ b_{0,k} = \frac{(W_0 - j) (1-q)^{k-j}}{W_0 (1-q)^k} \quad (1 \leq k \leq W_0 - 1) \quad (8) \]

Similarly, we have
\[ \frac{1}{1-q} \left( h_{i,k} \cdot q + b_{i,k} \right) \quad 0 < i < m, k = W_i - 1 \]
\[ b_{i,k} = \frac{h_{i-1,k} \cdot q + b_{i-1,k} \cdot (1-q)}{W_i} \quad 0 < i \leq m, 1 \leq k < W_i - 1 \quad (9) \]

Combining to (8) and (9), we obtain
\[ b_{i,k} = \frac{(W_i - j)(1-q)^{k-j}}{W_i (1-q)^k} \quad (0 \leq i \leq m, 1 \leq k \leq W_i - 1) \quad (10) \]

Finally, according to (4) and (10), all the values \( b_{i,k} \) can be expressed as functions of the value \( b_{i,0} \) and the probability \( q \). Therefore, \( b_{i,0} \) can be determined by imposing the normalization condition for stationary distribution and can be expressed as functions of the probability \( q \).
\[ 1 = \sum_{i=0}^{m} \sum_{k=0}^{W_i-1} b_{i,k} \quad (11) \]

The probability \( \tau \) that a station transmits in a randomly chosen slot time can now be expressed as
\[ \tau = \sum_{i=0}^{m} \sum_{k=0}^{W_i-1} b_{i,k} = \frac{1 - q^{m+1}}{1 - q} - b_{0,0} \quad (12) \]

In the stationary state, a station transmits a packet with probability \( \tau \). So we have
\[ q = 1 - (1-\tau)^{n+1} \quad (13) \]

Equations (11), (12), and (13) represent a nonlinear system in the two unknowns \( \tau \) and \( q \), which can be solved by numerical techniques.

**B. Throughput Analysis**

Let \( P_s \) be the probability that there is at least one transmission in the considered slot time. And let \( P_r \) be the probability that a transmission is successful, given the probability \( P_s \). So we have
\[ P_r = 1 - (1-\tau)^n \quad (14) \]

Now we are able to express the normalized system throughput \( S \) as the ratio.
\[ S = \frac{E[\text{Payload Information in a slot time}]}{E[\text{Length of a slot time}]} \]
\[ = \frac{(1-P_r)\sigma + P_s P_r T_c + (1-P_r) P_s T_c}{P_s P_r E[P]} \quad (16) \]

Here, \( T_c \) and \( T_r \) are the average time the channel is sensed busy because of a successful transmission or a collision respectively. The \( E[P] \) is the average packet length and \( \sigma \) is the duration of an empty slot time.

Let \( H = \text{PHYhdr} + \text{MAChdr} \) be the packet header, and \( \delta \) be the propagation delay. Let \( T_c \) be the average time the channel is sensed busy by each station during a collision. According to the standard [1], after transmitting a DATA/RTS frame, the station shall wait for an ACKTimeout/CTSTimeout interval. If the response does not occur during the ACKTimeout/CTSTimeout interval, the station shall conclude that the transmission of the DATA/RTS has failed. According to [3], in the basic access case, we have
\[ \{ T_c = \text{DIFS} + H + E[P] + \delta + \text{SIFS} + \text{ACK} + \delta \} \]
\[ \{ T_r = \text{DIFS} + \text{RTS} + \delta + \text{SIFS} + \text{CTS} + \delta + \text{SIFS} + H + E[P] + \delta + \text{SIFS} + \text{ACK} + \delta \} \quad (17) \]

We defined the Goodput \( G \) of the WLAN in this paper as the sum of the end-to-end throughput in WLANs. That is, when two mobile stations communicate with each other through the AP, the goodput is the end-to-end transmission rate between these two mobile stations. When a mobile station communicates with a wired station through the AP, the system throughput is the end-to-end transmission rate between the mobile station and the AP.

As we assume above, that all mobile stations work in saturation mode. In our analysis, we assume the UDP protocol is used. As TCP is a self-adapting protocol, we cannot make it work in saturation mode. However, in Section V, we shall prove that our new protocol can also improve the network performance using TCP protocol by simulation.

In our scenario, mobile stations in the WLAN communicate with each other through the AP. We assume there are \( n \) \((n+1)\) pairs of mobile stations in the WLAN. Each pair contains one MS and one MR. The MS sends UDP data packets to the MR in a saturation mode through the AP. In such case, the goodput of the WLAN is just the forwarding rate of the AP. As we discussed above, the AP can only achieve on the average \( l/(n+1) \) of the system throughput. So we have
\[ G_{\text{Standard}} = \frac{S}{n+1} \frac{P_s P_r E[P]}{(1-P_r)\sigma + P_s P_r T_c + (1-P_r) P_s T_c} \quad (19) \]

As for our new protocol, the AP can send a frame
immediately after any mobile station sending a frame. It means that as long as a mobile station can obtain the wireless channel successfully, a frame will be forward to the MR in the duration \(2^*T_s\). So we have

\[
G_{\text{new}} = \frac{P_s P_r E[P]}{(1-P_s)\sigma + P_s P_r \cdot (2T_s) + (1-P_s)P_r T_r}.
\]  

(20)

From the comparison between (19) and (20), it is easy to conclude that our new protocol will improve the throughput performance of the WLAN significantly. The performance evaluation in Section V will elucidate this further.

Since TCP is a bidirectional protocol, all stations (mobile stations and the AP) will send the data, when mobile stations communicate with each other or with wired stations using TCP protocol. In such case, the AP is a bottleneck too. Our protocol can give the AP more chances to get the channel, so as to improve the performance of WLANs. The simulation result can be found in Section V.

V. PERFORMANCE EVALUATION

We shall first validate our method and our analysis by ns-2 simulation [8]. Fig. 2 shows the WLAN under consideration. There is one Server (S) in the wired networks and several Wireless Stations (WS) in the WLAN. The channel rate is set to 11Mbps and the packet size is 1024 bytes.

![Network simulation scenario](image)

We first consider the scenario where mobile stations communicate with each other using UDP. We assume there are \(n\) pairs of mobile stations in the WLAN. Each pair contains one MS and one MR. As shown in Fig.3, the simulation results (points with “x”) of our new protocol agree very well with our analytical results (solid lines), thus demonstrating the analytical equation (20) of goodput for our new protocol based on Fig.1 is quite capable of capturing the real network scenarios. Likewise, we find such agreement for the analytical goodput equation (19) for the standard DCF [2].

As we can see, the goodput performance of our new protocol is much better (at least double) than that of the standard DCF in WLANs. The more mobile nodes in the WLAN is, the more obvious the advantage of our protocol. When the number of mobile station is 30, the goodput of our protocol is about 16 times larger than that of the standard DCF. This is because the AP is not a bottleneck in our protocol as discussed in Section IV.

![Goodput performance compare for UCP pair scenario](image)

![Fairness performance compare](image)

Fairness performance is also a very important property in wireless protocol. Figure 4 presents the fairness index of both the standard DCF and our modified DCF. We shall use the commonly accepted fairness index defined in [7] to measure the fair bandwidth allocation capability. That is,

\[
FI = \left( \sum_{i=1}^{n} x_i \right)^2 / \left( n \sum_{i=1}^{n} x_i^2 \right),
\]

where \(FI\) is fairness index, \(n\) is the number of contending stations, \(x_i\) is the bandwidth shared by the \(i\)th contending station. As shown in Fig. 4, our modified DCF maintains the same good fairness performance as the standard DCF (while improving the goodput performance of the WLAN greatly).

The second scenario is similar to the first scenario except that TCP is used instead of UDP. Whenever the MS sends data to the MR, the MR should return the ACK. As TCP is a self-adapting protocol, all stations should be in their unsaturated mode. If the MS/MR does not receive the ACK/DATA packet, it will remain silent and not contend for the channel. Such behavior gives the AP more chance to obtain the channel. Therefore, the goodput performance does not degrade as seriously as using UDP. As shown in Fig.5, our new protocol in such scenario also can obtain better performance than the standard DCF. In our protocol, the AP can send one packet without contending the channel, after any mobile stations sending a packet. Such process saves the contention time of the AP and reduces the probability of contention.
Finally, we consider the downloading scenario where mobile stations in the WLAN constantly download data from the wired server as shown in Fig. 2 using the TCP protocol. This downloading scenario also has its unsaturated mode. As we know the TCP is a bidirectional protocol, mobile stations downloading data from the wired station should contend the channel to return the ACK packet, which also benefit the AP in our new protocol. The analytical and simulation results in Fig.6 show the goodput performance of our new protocol is better than that of standard DCF.

VI. CONCLUSIONS

In this paper, we have proposed a new MAC layer protocol for 802.11 based WLANs to improve the network performance. Compared with standard DCF, our new protocol can give the AP more change to access the channel, so that the AP will not be the bottleneck any more. Our theoretical analysis proved that our protocol can improve the network performance. We have also done extensive simulations to evaluate the performance of our new protocol in several various scenarios. The results prove the correctness of our analysis. Both simulation results and theoretical analysis shows our protocol can enhance the performance of a WLAN largely.

ACKNOWLEDGMENTS

This research was supported in part by the National Natural Science Foundation of China (NSFC) under grant No. 60472078, by a grant from the Cisco University Research Program Fund at Community Foundation Silicon Valley, and by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Grant No. RGPIN-2003-42878.

REFERENCES


