Voice Capacity Analysis of WLANs with Channel Access Prioritizing Mechanisms

Xinhua Ling, University of Waterloo

Yu Cheng, Illinois Institute of Technology

Xuemin (Sherman) Shen and Jon W. Mark, University of Waterloo

ABSTRACT

Voice over IP applications in wireless networks have gained increasing popularity in recent years. As a delay-sensitive real-time application, a VoIP flow is usually given higher priority in accessing the shared wireless channel than delay-insensitive non-real-time flows. In contention-based wireless networks two widely used prioritizing MAC mechanisms are class-dependent arbitration interframe space and class-dependent contention window. In this article we propose an analytical model to evaluate the effect of the two mechanisms on voice capacity (the maximum number of two-way voice flow pairs supportable) of ad hoc mode and infrastructure mode wireless LANs. We show that the AIFS mechanism has a relatively strong effect on WLAN voice capacity in the ad hoc mode, but not in the infrastructure mode; and the CW mechanism, when properly configured, has a mild effect on voice capacity in both modes.

INTRODUCTION

Voice over IP (VoIP) applications in wireless networks have gained increasing popularity in recent years. Meanwhile, wireless local area networks (WLANs) have been widely deployed for wireless Internet access. The overlap between these two technologies has naturally given rise to the VoIP over WLAN (VoWLAN) application [1, 2]. In the widely deployed IEEE 802.11-based WLANs, distributed coordination function (DCF) is the dominant medium access control (MAC) protocol, which guarantees equal long-term channel access probability to all stations. It has been reported that the DCF MAC is inefficient in protecting quality-of-service (QoS)-critical applications (e.g., VoIP sessions) from QoS-resilient applications (e.g., file transfer) [1]. To provide satisfactory QoS support to the delay-sensitive real-time application, a VoIP flow is usually given higher priority in accessing the shared wireless channel than delay-insensitive non-real-time flows. In contention-based wireless networks, two widely used prioritizing MAC mechanisms are:

- Class-dependent arbitration interframe space (AIFS)
- Class-dependent contention window (CW)

which have been included in the IEEE 802.11e amendment [3] to the legacy IEEE 802.11 standard.

Performance analysis of the AIFS and CW mechanisms has been a very hot topic (e.g., [4–6]). However, most of the studies are for saturated stations that always have MAC frames in or waiting for service, which are not appropriate for the analysis of VoIP flows usually characterized as bursty traffic [2]. A model that can analyze unsaturated stations, and is thus suitable for VoIP flows, is proposed in [7]. It has been successfully applied to analyze the access point (AP) multiplexing gain in serving variable bit rate (VBR) VoIP flows in an infrastructure mode WLAN using the CW mechanism only. The effect of AIFS on the performance of VoWLAN remains unexplored to a large extent.

In this article we propose a performance analysis model for studying the AIFS and CW mechanisms in a WLAN with contention-based channel access. We then apply it to study the effectiveness of these two mechanisms on improving the voice capacity, defined as the maximum number of two-way VoIP sessions supportable, of ad hoc and infrastructure mode WLANs. Both analytical and simulation results are presented to demonstrate that the AIFS and CW mechanisms have saliently different effects on improving WLAN voice capacity, depending on the WLAN operation mode and traffic load conditions.

The rest of the article is organized as follows. The main channel access prioritizing mechanisms are reviewed. The analytical model for performance study of the mechanisms is proposed. Detailed effects of the AIFS and CW mechanisms on voice capacity of WLANs are presented and discussed. Finally, concluding remarks are given.

CHANNEL ACCESS PRIORITIZING MECHANISMS

In the IEEE 802.11e standard, enhanced distributed channel access (EDCA) supports service differentiation mainly by distributed prioritized channel access among different access categories (ACs) with three AC-dependent parameters: AIFS, CW, and transmission opportunity (TXOP). The principles of these mechanisms are as follows.

With EDCA, user traffic is first classified into multiple ACs, such as voice, video, best effort, and background. Each station regulates its frame transmission using the contention parameters associated with each AC. When a station has a frame at the MAC sublayer buffer, it first senses the channel. If the channel is busy, it performs the backoff procedure by first setting the backoff counter (BC) to an integer sampled from the minimum CW size. Therefore, one differentiation mechanism is to assign a higher-priority AC a smaller value of minimum CW size such that higher-priority ACs statistically spend less time on backoff. After the channel becomes idle for the AC-dependent AIFS,[1] the station can count down the BC at the beginning of each idle slot and also the first slot of a channel busy period. Hence, the BC value represents the total number of generic slots[2] for which a station has to wait before it can transmit. Since higher-priority ACs are assigned smaller-value AIFS, they have greater chances to access the channel than low-priority ACs. Figure 1 shows an example of four ACs, where AC_1 has the highest priority. To illustrate the effect of different AIFS lengths, the time between two busy periods, except $AIFS_1$, is divided into four contention zones, Z_i, each with a length of M_i slots, $i = 1, 2, 3, 4$. In zone Z_1, only AC_1 stations are allowed to contend for channel access, while in zone Z_2 the contentions are between AC_1 and AC_2, i.e., contentions in zone Z_i involve AC_j, $j \leq i$. Consequently, each AC encounters different contentions in its allowable contention zones. After one station succeeds in contending for channel access, it can transmit for a duration up to the TXOP. Different TXOP durations can be assigned to different ACs to further differentiate the service.

As the TXOP durations need to be calculated by the AP, this mechanism is usually used only in the infrastructure mode WLAN where an AP exists. In contrast, the AIFS and CW mechanisms need no help from a central controller, so they are more widely used in distributed channel access protocols (e.g., the prioritized channel access protocol [8] specified by the Multi-Band OFDM Alliance for ultra-wideband wireless personal area networks). Therefore, our focus is on the study of these two mechanisms.

A PERFORMANCE ANALYSIS MODEL FOR THE PRIORITIZING MECHANISMS

In this section we give an overview of a performance analysis model [9] by which one can obtain the voice capacity of a WLAN with either the AIFS or CW mechanism, or both, in a unified manner.

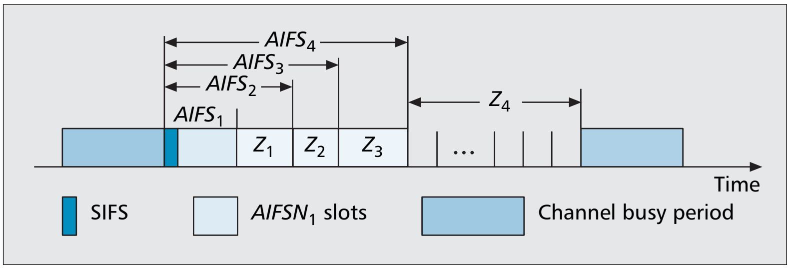


Figure 1. *An illustration of prioritized channel access for different station classes.*

THE SYSTEM MODEL

Consider a single-hop multiservice network consisting of K classes of stations, with N_k stations in each class. Specifically, all the stations are within the transmission range of one another, so there are no hidden terminals in the network, which is common in WLANs [1]. The time axis is slotted, and all the stations are synchronized so that all stations start their transmissions only at the beginning of a slot. In addition, all the stations can correctly sense the channel status. An ideal wireless channel without transmission error is assumed so that all transmitted frames may be lost only due to collisions caused by simultaneous transmissions from multiple stations, which is reasonable in a typical office/laboratory WLAN as demonstrated by the field measurement in [1]. MAC frame lengths or the physical layer data rates used by each class may be different. For simplicity, we assume that one station carries only one traffic flow. For stations in the same class, the incoming traffic is the same, and they receive the same type of service from the network. The above system model can represent both an ad hoc mode WLAN with peer-to-peer traffic and an infrastructure mode WLAN with either symmetric or asymmetric to-and-from AP traffic.

THE PERFORMANCE ANALYSIS MODEL

Due to the different AIFS values assigned, stations of different classes are eligible to contend for channel access in different zones, as shown in Fig. 1. More specifically, a station of class k is eligible to decrease its BC and transmit, if its BC has been decremented to zero, in zones Z_k to Z_K, for $k = 1, \ldots, K$. As a result, stations face different contention situations and experience different frame collision probabilities in different zones. On the other hand, each class k station performs its backoff procedure only depending on the transmission results of its frames transmitted in the eligible zones (i.e., it enters the next backoff stage) only when its frame experiences a collision, which occurs with probability of β_k, the mean collision probability of frames transmitted in any of the eligible zones.

With a given β_k, the number of *transmission trials* R_k of a class k frame is a random variable with geometric distribution. Since the channel access policy is the same for each frame of the same flow, the frame service pro-

[1] *In the standard,* $AIFS_i$ *equals the aggregate duration of an SIFS and* $SIFSN_i$ *slots, where* i *is the class index.*

[2] *A generic slot may refer to an idle time slot, a successful transmission, or a collision with respective probabilities.*

Channel rate	54 Mb/s	Slot time	9 μs
SIFS	16 μs	DIFS	34 μs
CW_{min}	16	CW_{max}	1024
Retry limit	7	PLCP and preamble	24 μs
Upper layer overhead	40 bytes	MAC overhead	34 bytes
Voice payload	80 bytes	ACK frame	14 bytes

Table 1. *Parameters of VOIP over WLAN.*

cess at the MAC sublayer of a station can be deemed as a renewal process with the service period for each frame as a renewal cycle. Therefore, R_k can also be deemed as a reward associated to the renewal cycle. By the renewal reward theorem, the probability γ_k of a class k station to transmit at the beginning of a randomly chosen generic slot in any of its eligible zones is given by

$$\gamma_k = \frac{\text{average number of transmission trial}}{\text{average total number of generic slots in a cycle}}, \tag{1}$$

which is a function of β_k. On the other hand, the collision probability $\beta_{k,z}$ of a class k frame in its eligible zone Z_z can be obtained as a simple function of γ_k, considering that a collision occurs only when two or more eligible stations transmit simultaneously. Hence, the mean value β_k can be obtained also as a function of γ_k. The resultant function and the one given by Eq. 1 can be jointly solved to obtain β_k and γ_k.

For given values of β_k and γ_k, the average frame service time of class k frames can be obtained with the help of a *virtual backoff event* method as follows. Consider a randomly chosen frame transmission over the channel; it occurs in different contention zones with different probabilities. Depending on the zone containing this transmission, different values of BC deduction are applied to different station classes — a station may not count down its BC if it is ineligible to contend in this zone. On average, however, a randomly chosen transmission will cause a class-dependent BC deduction $E[O_k]$ to an individual station, but the same delay $E[D]$ to all stations. Define such a transmission as a virtual backoff event to any station. Notice that the average total number of backoff slots $E[B_k]$ and transmission trials $E[R_k]$ can be obtained from β_k and γ_k. Then, considering the number of virtual backoff events that occur in $E[R_k] + E[B_k]$ generic slots and the associated delay they cause, the average frame service time ζ_k of a class k station is simply given by

$$\zeta_k = \frac{E[R_k]+E[B_k]}{E[O_k]} E[D]. \tag{2}$$

Comparing ζ_k with the average frame interarrival time, we can easily determine whether the station is saturated or not.

Performance Evaluation

In this section we apply the above analytical model to study the effects of the AIFS and CW mechanisms on improving the voice capacity of WLANs in both the ad hoc and infrastructure modes. Each voice flow generates a packet of 80 bytes every 10 ms, following an option in the G.711 codec for VoIP applications. Since a voice conversation is usually bidirectional, a new pair of voice flows are deemed admissible to the network if the stations carrying existing flows and the new ones will not become saturated. Therefore, the voice capacity (denoted N_v) refers to the maximum number of *voice flow pairs* admissible to the WLAN (i.e., the number of voice flows is equal to $2N_v$). We choose IEEE 802.11a as the physical (PHY) layer, and the highest possible PHY data rate of 54 Mb/s is used. Other parameters used in this study are given in Table 1 unless otherwise stated. We use Maple [10] to obtain the numerical results and compare them to simulation results from our event-driven simulator developed in C language [7]. We have found that in all the cases studied both results are very close to each other.

As our main focus is on the effects of channel access prioritizing mechanisms on voice capacity, an ideal channel is assumed in all the scenarios studied, which means transmitted frames are only lost due to collisions caused by simultaneous transmissions from multiple stations. Some previous studies (e.g., [11]) have shown that transmission errors caused by poor wireless channel conditions can be deemed independent of frame collisions. Meanwhile, such errors will reduce the probability of successful frame transmission even when there is no collision. Therefore, the voice capacity achievable in a practical WLAN may be lower than the results presented here.

Ad Hoc Mode WLAN

In the ad hoc WLAN considered, there is no AP, and all the nodes follow an EDCA-like MAC protocol. Each node is assumed to carry only one type of traffic flow, either voice or data. As background traffic competing for channel access with the voice flows, each data flow has a saturated source with MAC frame size 1500 bytes. A saturated source means that the MAC sublayer buffer is always nonempty, with frames being serviced or waiting for service. Although the data flows seem greedy by always competing with voice flows, they are delay-insensitive and thus elastic in their bandwidth requirement.

Effects of AIFS — The CW parameters of both voice and data flows are set to the [CW_{min}, CW_{max}] = [16, 1024] as in the 802.11a standard. Figure 2a shows the prioritizing effect of the AIFS mechanism. When this mechanism is not used (i.e., $M_1 = 0$), the voice capacity is N_v = 18, 11, 5 when there are N_d = 5, 10, 15 data flows, respectively. When the AIFS mechanism is in place, the voice capacity increases with M_1 until it reaches the maximum N_v^* of 26, which is common for all three data flow configurations. As a reference, the voice capacity with no background data flows in the studied network is $\hat{N}_v^*$

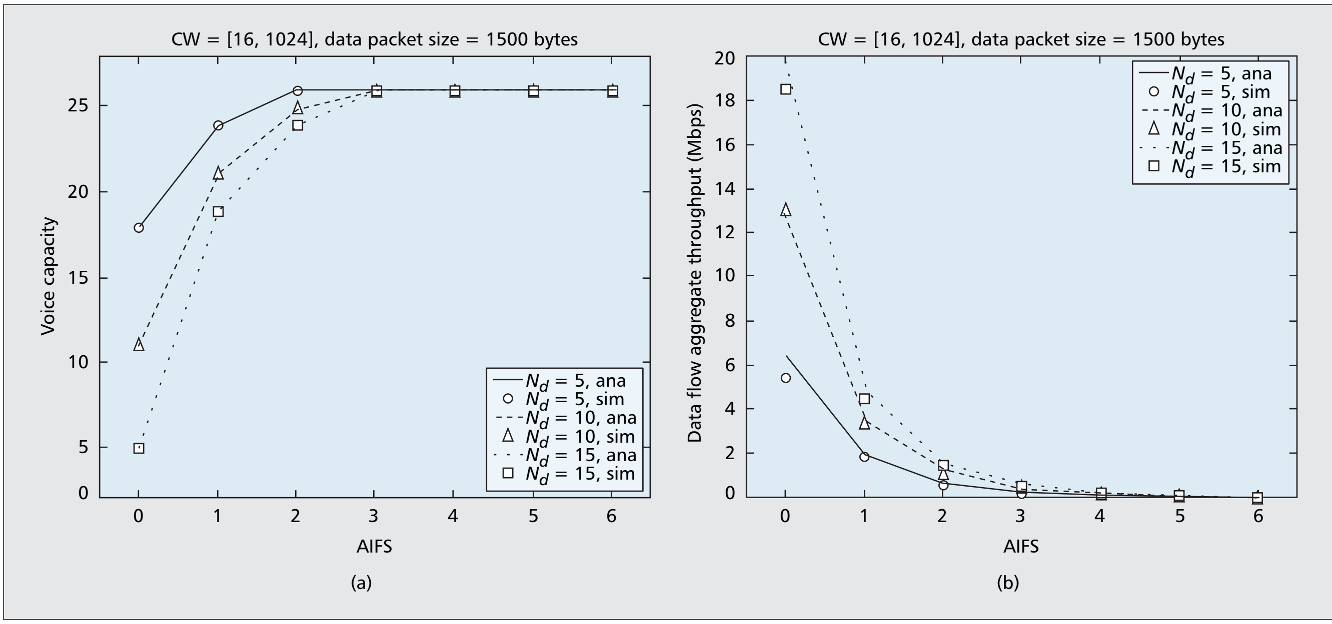


Figure 2. *Parameters of VoIP over WLAN.*

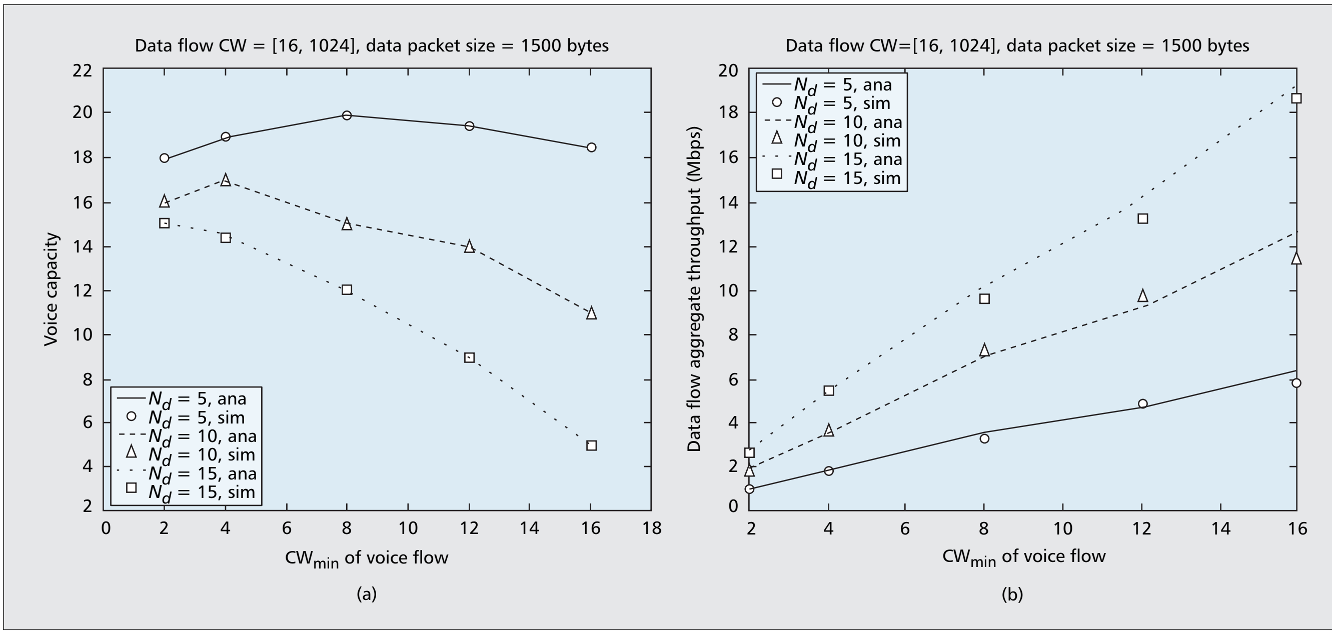
= 27, which is determined purely by the contentions among the voice flows. The marginal gap between the N_v^* reached by using the AIFS mechanism with $M_1 \geq 3$ and the $\hat{N}_v^*$ demonstrates that the AIFS mechanism is quite effective (with just a small value of M_1) in prioritizing voice flows over data flows.

The gain in voice capacity, however, is obtained at the cost of throughput loss of the data flows, as shown clearly in Fig. 2b. With the increase of M_1, the aggregate throughput of data flows keeps decreasing, even though the gain of voice capacity vanishes when M_1 becomes larger than three. This indicates that the AIFS mechanism will overly prioritize the voice flows when M_1 is larger than a certain threshold. Notice that this threshold depends on many factors (e.g., the characteristics of the voice and data flows and the CW parameters).

There are two interesting observations from Fig. 2. The first is that the prioritizing effect of the AIFS mechanism (in terms of voice capacity increase) is the strongest at the first step with M_1 changing from zero to one and weakens for further increments of M_1. Take the configuration of $N_d = 15$ as an example. At the first step of M_1 change, the voice capacity N_v increases by an astonishing 280 percent (from 5 to 19). In contrast, when M_1 changes from one to two, the resultant N_v increase is just about 26 percent (from 19 to 24). The gain of N_v is even smaller when M_1 increases from two to three, and there is no gain of N_v when M_1 further increases up to six. This phenomenon can be explained as follows. When M_1 changes from zero to one and with the original $N_v = 5$, the immediate effect is that the data flows have a lower chance to decrease their backoff counters; thus, there is more bandwidth to accommodate voice flows. As the number of voice flows increases, two contradicting effects on the increase of voice capacity appear. On one hand, it will be more difficult for data flows to decrease their backoff counters, and thus more bandwidth will be usable by voice flows, which tends to increase voice capacity. On the other hand, the competition among the increased voice flows will be fiercer, so the service time of voice frames will be longer, which hinders the increase of voice capacity. When these two effects become balanced, the increase of voice capacity stops at $N_v = 19$. Notice that the bandwidth consumed by data flows at this point ($M_1 = 1$) is much lower than that of $M_1 = 0$. Hence, when M_1 increases to two, the bandwidth further given up by data flows (to accommodate voice flows) is lower than that in the previous step, resulting in a smaller increase in voice capacity. A similar situation occurs when M_1 increases further. Finally, when the bandwidth given up by the data flows is not high enough to accommodate one more pair of voice flows, voice capacity remains at N_v^*. The above argument also explains the second observation: N_v^* is reached with $M_1 = 2$ for $N_d = 5$ data flows, but with $M_1 = 3$ for $N_d = 10$ and 15 data flows. This is because with fewer competing data flows, the N_v^* voice flows need weaker prioritization to grasp enough bandwidth from the data flows.

Effects of CW — The effects of CW parameters are investigated with the AIFS mechanism disabled ($M_1 = 0$). That is, the voice flows are prioritized over data flows purely by adjusting the CW parameters. Previous studies [5, 7] have shown that CW_{min} has a greater prioritizing effect than CW_{max} in most cases. Therefore, in this study only the minimum contention window of voice flows (CW_{min}^v) is adjusted, while CW_{max} is fixed as 1024. The CW parameters for the background data flows are kept at [16, 1024] as in the previous scenario.

Usually, a smaller CW_{min} means higher priority for channel access [5]. As an illustration, the aggregate throughput of data flows keeps on decreasing with CW_{min}^v for all three configurations, as shown in Fig. 3b. Therefore, one may



■ **Figure 3.** *Effect of AIFS in ad hoc mode.*

expect that voice capacity increases with the decrease of CW^v_{min}, similar to the case of increasing the value of M_1 in the AIFS mechanism. Figure 3a shows the voice capacity for different CW_{min} values of voice flows for three data traffic configurations. Interestingly, only the changing trend of voice capacity for $N_d = 15$ is as expected, while those of the other two configurations are not. For $N_d = 5$, the voice capacity slightly increases from 18 to 20 when CW^v_{min} shrinks from 16 to 8; but it starts dropping when CW^v_{min} further decreases. A similar trend is observed for $N_d = 10$, with the maximum voice capacity of 17 reached at $CW^v_{min} = 4$.

The reason behind this interesting phenomenon is as follows. With a given number of voice flows, the direct effects provided by using a smaller CW^v_{min} are twofold:

- It prioritizes the channel access of the voice flows, so the voice capacity has a potential to increase.
- It intensifies the contention among the flows (both voice and data), which lengthens the service time of voice frames and thus may obstruct the voice capacity increase.[3]

Therefore, there is a net increase of voice capacity when the former effect outweighs the latter, which corresponds to situations with moderate values of CW^v_{min}. When the CW^v_{min} further decreases to smaller values, the net change of voice capacity goes in the opposite direction due to the greatly deteriorated contention among all the flows. The turning point of the above process depends on the total carried traffic in the network, reflected indirectly by the numbers of voice and data flows here. In general, the network with fewer data flows can support more voice flows with a properly selected CW^v_{min}.[4]

Comparison of the Two Mechanisms — It can be seen from the above that the AIFS mechanism is more effective in squeezing the bandwidth consumed by the background data flows than the CW mechanism, which results in higher voice capacity of the ad hoc WLAN in this study. Moreover, the AIFS mechanism has a coarser granularity in channel access prioritization. As an example, increasing M_1 from zero to one (the minimum increase possible) gives a loss of aggregate throughput of data flows close to that caused by halving the CW^v_{min} from the original value. The obvious side effect is that the background data flows are relatively easier to starve with the AIFS mechanism.

INFRASTRUCTURE MODE WLAN

Unlike the traffic pattern in ad hoc WLANs, a VoIP conversation in an infrastructure WLAN usually consists of an uplink (from a station to the AP) flow and a downlink (from the AP to a station) one. A serious constraint on the voice capacity of this mode is the so-called *AP bottleneck problem* [2], which is caused by the highly unbalanced voice traffic between the AP and the stations. Specifically, when there are *n* conversations in the WLAN, the AP will carry a traffic load *n* times of that on a station. Therefore, if the AP and a station have the same priority in accessing the channel, the former becomes saturated long before the latter does; thus, the AP is the bottleneck. In principle the solution to this problem is to balance the maximum throughputs of uplink and downlink flows [2, 12] while keeping the AP and stations unsaturated, which usually means that the AP should have a certain level of priority in channel access.

In this section we study the effectiveness of both the AIFS and CW mechanisms in solving the AP bottleneck problem in the context of contention-based channel access. Since the presence of any data traffic will only reduce the number of simultaneous calls [13], no background data flows are considered in the following study so that we can focus on balancing the uplink and downlink voice flows.[5] Notice that an

[3] *Notice that this negative effect does not exist in the AIFS mechanism when the number of voice flows is fixed. It only comes into the picture when more voice flows are admitted into the network, which brings fiercer contention among voice flows.*

[4] *The value of CW^v_{min} in this case is usually larger than that for fewer voice flows, as determined by the nature of distributed channel access contention among the flows.*

[5] *We can effectively reduce the effect of background data flows by setting a large* $AIFS_N$[DATA]*, as shown in the study of ad hoc mode WLANs.*

important factor here is that the low-priority uplink voice flows also have a QoS requirement (i.e., the received service rate is larger than the traffic arrival rate), which is not required by the low-priority data in the previous case.

We have chosen 10 different baseline cases, each with two pairs of CW parameters ([CW_{min}, CW_{max}]) for the AP and stations, respectively. Setting the AP to AC_1 and the stations to AC_2, the difference in their AIFSs is again denoted by M_1 slots. For each baseline case, the M_1 is further changed from zero to four to investigate the effect of the AIFS mechanism. The voice capacities of the infrastructure WLAN in all the above 50 different cases are summarized in Table 2. Since the analytical results are the same as the simulation ones for all cases studied, we present just one set of results in the table to avoid unnecessary duplication.

Case no.	AP CW [min, max]	Station CW [min, max]	M_1				
			0	1	2	3	4
1	[32, 1024]	[32, 1024]	24	24	24	24	23
2	[16, 1024]	[32, 1024]	28	28	27	27	26
3	[8, 1024]	[32, 1024]	30	29	28	(28)	(27)
4	[4, 1024]	[32, 1024]	30	28	(28)	(27)	(27)
5	[16, 1024]	[16, 1024]	28	28	27	27	26
6	[8, 1024]	[16, 1024]	30	29	27	(27)	(27)
7	[4, 1024]	[16, 1024]	30	28	(28)	(27)	(27)
8	[8, 1024]	[8, 1024]	29	29	27	(26)	(26)
9	[4, 1024]	[8, 1024]	30	27	(27)	(26)	(26)
10	[2, 1024]	[8, 1024]	28	(26)	(27)	(28)	(27)

Note: A number with () means the bottleneck changes from the AP to the stations.

■ **Table 2.** *Voice capacity of infrastructure mode WLAN.*

Effects of AIFS — The effect of using only the AIFS mechanism to prioritize the AP can be found by examining the voice capacity along each row for cases 1, 5, and 8 in Table 2. It can be seen that the use of AIFS does not increase voice capacity, but reduces it in most of the cases! This seems counterintuitive, especially when considering the strong prioritizing effect of the AIFS mechanism manifested in the ad hoc WLAN. The reason is as follows. According to the AIFS mechanism, after each channel busy period, the AP can have a guaranteed BC decrement of M_1 if its BC value at the end of the preceding busy period is no less than M_1, or it can transmit a frame without collision otherwise (since no station can start a transmission in zone Z_1, as shown in Fig. 1). Therefore, the AP can decrease its BC at a faster speed than the other stations, as expected from the prioritizing effect of the AIFS. However, these stations are at a disadvantage in channel access contention and experience longer frame service times. Consequently, there are more stations with nonempty MAC buffers at any time instant; thus, more stations contend for channel access with the AP in zone Z_2. This causes a longer frame service time for the AP, which greatly offsets the aforementioned advantage of the AP. As a result, with the increase of M_1, more and more stations contend with the AP, and the bandwidth usable by the AP decreases, which leads to decreased voice capacity. We may call this a *neither-side-gain* effect. Moreover, the AP remains the bottleneck when M_1 is relatively small, as the advantage in channel access for the AP is not strong enough to provide it extra bandwidth to handle its heavy load. However, when M_1 is set so large that the stations suffer too long frame service times, they become saturated even before the AP does — the bottleneck then moves to the stations, as shown in Table 2.

It is noteworthy that the offset effect mentioned above does not exist in the ad hoc WLAN studied earlier, because the number of low-priority flows (saturated data flows) competing with the high-priority voice flows is *fixed* in that case. Moreover, the number of high-priority flows in that case is larger than one, so there is a higher chance that a high-priority flow transmits in zone Z_1, leading to more frequent channel busy periods and thus more Z_1s, which "amplifies" the prioritizing effect of the AIFS mechanism. Hence, the voice flows as a group can keep on squeezing the usable bandwidth of the saturated data flows with the increase in the value of M_1.

Effects of CW — Compared with AIFS, the CW mechanism provides some positive results on voice capacity. The column with $M_1 = 0$ shows the net effect of using the CW mechanism only. Cases 1 to 4, 5 to 7, and 8 to 10 share the same uplink CW parameters, respectively. In each of the above groups the downlink CW_{min} decreases by half at each step, corresponding to the increase of the priority of the downlink voice flows. When the CW_{min} of uplink flows is relatively large (e.g., 32 or 16), the uplink flows do not contend aggressively. Therefore, voice capacity increases when the AP is given higher priority, until it finally reaches the maximum value of 30. In contrast, when the uplink CW_{min} is small (e.g., 8), the uplink flows contend more aggressively. Therefore, with 30 flow pairs, when the AP's CW_{min} is further decreased to 2 as in case 10, the channel access contention becomes too fierce, resulting in frequent collisions and thus longer frame service times for both uplink and downlink flows, which finally leads to a decrease in voice capacity. This is similar to the changing pattern of $N_d = 10$ in Fig. 3a, which is also due to the relatively large number of competing voice flows therein (compared to that of $N_d = 15$).

Effects of Integrating AIFS and CW — The effects of jointly using the AIFS and CW mechanisms are reflected by the results with $M_1 > 0$ and unequal CW parameter pairs between downlink and uplink flows in Table 2. We can see that the effect of AIFS is almost unchanged from the

The results suggest that techniques for reducing the protocol overhead, such as compressing upper layer headers of the voice frame and aggregating the downlink frames for multicast multiplexing, should be used to further improve the voice capacity in addition to the two prioritizing mechanisms.

case when the CW mechanism is not used. An exception is in case 10, where the CW_{min} of the downlink flows is 2. When $M_1 = 1$, the aforementioned neither-side-gain effect occurs, so the voice capacity is less than that of $M_1 = 0$. However, when $M_1 = CW_{min} = 2$, the downlink flows have much greater advantage because most of the downlink frames can be transmitted successfully in zone Z_1 without contention from the uplink flows. Notice that there are still chances for the uplink flows to transmit at the beginning of zone Z_2, which may collide with a downlink frame whose initial backoff counter has a value of 2. Such collisions no longer occur when M_1 is set to 3, so the downlink frames are always transmitted successfully in zone Z_1 and collisions only occur among the uplink frames, which results in slightly increased voice capacity.[6]

Nevertheless, when M_1 is further increased to 4, the length of Z_1 is unnecessarily long because slots 3 and 4 in Z_1 are always idle and wasted when the AP does not have a frame to serve. Hence, compared to the case of $M_1 = 3$, less bandwidth is available for the uplink flows to share, which results in decreased voice capacity. From this case we can see that there is a complicated effect of jointly using the two mechanisms when M_1 is close to the CW_{min} of the high priority class, especially when their values are small.

Other Solutions to the AP-Bottleneck Problem — With centralized control, it is relatively easy to give the AP priority over the stations (e.g., [14]). However, the study of centralized control methods is beyond the scope of this article. Some other approaches to increase voice capacity in general wireless networks (e.g., upper-layer header reduction [14]) and/or specifically in infrastructure WLANs (e.g., multiplex-multicast for downlink VoIP traffic [1]) do not interfere with channel access contention, so they can be used in parallel with the two prioritizing mechanisms.

Conclusion

We have proposed an analytical model for performance analysis of two widely used channel access prioritizing mechanisms and successfully applied it to evaluate their effects on the voice capacity of WLANs in different situations. Comparing the resultant voice capacities of WLANs with and without the two mechanisms, we have the following observations. In the ad hoc WLAN, the AIFS mechanism is very effective in prioritizing the voice flows and suppressing the bandwidth used by data flows; thus, its deployment results in a relatively large increase of voice capacity. The CW mechanism, on the contrary, can only provide a mild increase in voice capacity due to its mild prioritizing effect. In the infrastructure mode with no background data traffic, the contention is between the voice flows carried by the AP (downlink traffic) and those carried by the stations (uplink traffic). Since the low-priority uplink flows also have a QoS requirement, the AIFS mechanism renders no improvement of voice capacity mainly due to the neither-side-gain effect caused by the strong prioritization of the downlink flows. In contrast, the CW mechanism can still provide a moderate increase in voice capacity. The results suggest that techniques for reducing the protocol overhead, such as compressing upper layer headers of the voice frame and aggregating the downlink frames for multicast multiplexing, should be used to further improve voice capacity in addition to the two prioritizing mechanisms.

[6] *The voice capacity achieved in this case is the same as that of* $M_1 = 0$, *but the bottleneck is on the stations' side.*

References

[1] W. Wang, S. C. Liew, and V. O. K. Li, "Solutions to Performance Problems in VoIP over a 802.11 Wireless LAN," *IEEE Trans. Vehic. Tech.*, vol. 54, no. 1, Jan. 2005, pp. 366–84.

[2] L. X. Cai *et al.*, "Voice Capacity Analysis of WLAN with Unbalanced Traffic," *IEEE Trans. Vehic. Tech.*, vol. 55, no. 3, May 2006, pp. 752–61.

[3] IEEE Std. 802.11e-2005, Amendment to IEEE Std. 802.11 1999, 2005.

[4] J. W. Robinson and T. S. Randhawa, "Saturation Throughput Analysis of IEEE 802.11e Enhanced Distributed Coordination Function," *IEEE JSAC*, vol. 22, no. 5, June 2004, pp. 917–28.

[5] Y. Xiao, "Performance Analysis of Priority Schemes for IEEE 802.11 and IEEE 802.11e Wireless LANs," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, July 2005, pp. 1506–15.

[6] X. Chen *et al.*, "Supporting QoS in IEEE 802.11e wireless LANs," IEEE Trans. Wireless Commun., vol. 5, no. 8, Aug. 2006, pp. 2217–27.

[7] Y. Cheng *et al.*, "A Cross-Layer Approach for WLAN Voice Capacity Planning," *IEEE JSAC*, vol. 25, no. 4, May 2007, pp. 678–88.

[8] ECMA-368, High Rate Ultra Wideband PHY and MAC Standard, 1st ed. Dec. 2005, http://www.ecma-international.org/publications/standards/Ecma-368.htm

[9] X. Ling, "Performance Analysis of Distributed MAC Protocols for Wireless Networks," Ph.D. dissertation, Univ. of Waterloo, Ontario, Canada, 2007.

[10] A. Heck, *Introduction to Maple*, 3rd ed., Springer-Verlag, 2003.

[11] Y. Zheng *et al.*, "Performance Analysis of IEEE 802.11 DCF in Imperfect Channels," *IEEE Trans. Vehic. Tech.*, vol. 55, no. 5, Sept. 2006, pp. 1648–56.

[12] D. Gao, J. Cai, and C. Chen, "Capacity Analysis of Supporting VoIP in IEEE 802.11e EDCA WLANs," *Proc. IEEE GLOBECOM '06*, Nov.–Dec. 2006.

[13] S. Garg and M. Kappes, "Can I Add a VoIP Call?" *Proc. IEEE ICC '03*, May 2003, pp. 779–83.

[14] P. Wang, H. Jiang, and W. Zhuang, "IEEE 802.11e Enhancement for Voice Service," *IEEE Wireless Commun.*, vol. 13, no. 1, Feb. 2006, pp. 30–35.

Biographies

Xinhua Ling (x2ling@bbcr.uwaterloo.ca) received a B.Eng. (1993) degree in radio engineering from Southeast University, China, an M.Eng. (2001) degree in electrical engineering from the National University of Singapore, and a Ph.D. (2007) degree in electrical and computer engineering from the University of Waterloo, Canada. Since May 2007 he has been a post-doctoral fellow at the University of Waterloo. His research interests are in the area of cellular, WLAN, WPAN, ad hoc networks, and their internetworking, including MAC protocol design and performance analysis, and cross-layer design for QoS support.

Yu Cheng [S'01, M'04] (cheng@iit.edu) received B.E. and M.E. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1995 and 1998, respectively, and a Ph.D. degree in electrical and eomputer engineering from the University of Waterloo in 2003. From September 2004 to July 2006 he was a postdoctoral fellow in the Department of Electrical and Computer Engineering, University of Toronto, Canada. Since August 2006 he has been an assistant professor in the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago. His research interests include application-oriented networking, Internet performance analysis and resource allocation, network measurement, wireless networks, and wireless/wireline interworking. He received a Postdoctoral Fellowship Award from the Natural Sciences and Engineering Research Council of Canada in 2004.

Xuemin (Sherman) Shen [M'97, SM'02] (xshen@bbcr.uwaterloo.ca) received a B.Sc. (1982) degree from Dalian Mar-

itime University, China, and M.Sc. (1987) and Ph.D. degrees (1990) from Rutgers University, New Jersey, all in electrical engineering. He is a professor, university research chair, and associate chair for graduate studies, Department of Electrical and Computer Engineering, University of Waterloo. His research focuses on mobility and resource management in interconnected wireless/wired networks, UWB wireless communications systems, wireless security, and ad hoc and sensor networks. He is a co-author of three books, and has published more than 300 papers and book chapters in wireless communications and networks, control, and filtering. He served as Technical Program Committee Chair for IEEE GLOBECOM '07, General Co-Chair for Chinacom '07 and QShine '06, and was the Founding Chair for the IEEE Communications Society Technical Committee on P2P Communications and Networking. He also serves as a Founding Area Editor for *IEEE Transactions on Wireless Communications*; Editor-in-Chief for *Peer-to-Peer Networking and Application*; and an Associate Editor for *IEEE Transactions on Vehicular Technology*, *KICS/IEEE Journal of Communications and Networks*, *Computer Networks*, *ACM/Wireless Networks*, and *Wireless Communications and Mobile Computing* (Wiley). He has also served as Guest Editor for *IEEE JSAC*, *IEEE Wireless Communications*, and *IEEE Communications Magazine*. He received the Excellent Graduate Supervision Award in 2006, and the Outstanding Performance Award in 2004 from the University of Waterloo, the Premier's Research Excellence Award in 2003 from the Province of Ontario, Canada, and the Distinguished Performance Award in 2002 from the Faculty of Engineering, University of Waterloo. He is a registered Professional Engineer of Ontario, Canada.

JON W. MARK [M'62, SM'80, F'88, LF'03] (jwmark@bbcr.uwaterloo.ca) received a Ph.D. degree in electrical engineering from McMaster University, Canada, in 1970. Upon graduation, he joined the Department of Electrical Engineering (now Electrical and Computer Engineering) at the University of Waterloo, became a full professor in 1978, and served as department chairman from July 1984 to June 1990. In 1996 he established the Centre for Wireless Communications (CWC) at the University of Waterloo and has since been serving as the founding director. His current research interests are in wireless communications and wireless/wireline interworking, particularly in the areas of resource management, mobility management, and end-to-end information delivery with QoS provisioning. He is a co-author of the textbook *Wireless Communications and Networking* (Prentice Hall, 2003). He is a Life Fellow of the IEEE and has served as a member of a number of editorial boards, including *IEEE Transactions on Communications*, *ACM Wireless Networks*, and *Telecommunication Systems*. He was a member of the Inter-Society Steering Committee of *IEEE/ACM Transactions on Networking* from 1992 to 2003, and a member of the IEEE ComSoc Awards Committee during the period 1995–1998.