Understanding 802.11e Contention-Based Prioritization Mechanisms and Their Coexistence with Legacy 802.11 Stations

Giuseppe Bianchi, Universita degli Studi di Roma — Tor Vergata
Ilenia Tinnirello and Luca Scalia, Universita degli Studi di Palermo

Abstract

The IEEE 802.11e task group has reached a stable consensus on two basic contention-based priority mechanisms to promote for standardization: usage of different arbitration interframe spaces and usage of different minimum/maximum contention windows. The goal of this article is to provide a thorough understanding of the principles behind their operation. To this purpose, rather than limit our investigation to high-level (e.g. throughput and delay) performance figures, we take a closer look at their detailed operation, also in terms of low-level performance metrics (e.g., the probability of accessing specific channel slots). Our investigation on one hand confirms that AIFS differentiation provides superior and more robust operation than contention window differentiation. On the other hand, it highlights performance issues related to the coexistence between 802.11e contention-based stations with legacy 802.11 stations, and provides guidelines for the 802.11e parameter settings when such a coexistence is the goal.

The IEEE 802.11 technology [1] is experiencing impressive market success. Cheap and easy-to-install components, unlicensed spectrum, broadband capabilities, interoperability granted by standards and certifications (e.g., WiFi): these are a few of the key factors driving the evolution of WLAN from niche technology to public access means. The present challenge of WLANs is to offer a large portfolio of wireless mobile services to highly heterogeneous users with widely different requirements. This goal can be accomplished only by introducing suitable forms of service differentiation support at the various levels of a complex wireless network architecture.

A basic building block for service differentiation is the introduction of layer 2 prioritized delivery mechanisms for different traffic classes (and users), and support of quality of service (QoS) objectives. In switched Ethernet networks, service differentiation is managed within switches, through IEEE 802.1p priorities/virtual LAN IEEE 802.1Q tags. In IEEE 802.11 WLANs, all stations share the access to the same radio channel, and no switching operation is possible. Thus, the service differentiation mechanisms must be compulsorily introduced as medium access control (MAC) layer extensions.

For this reason, the 802.11e task group was established in July 1999, chartered to introduce QoS support at the MAC layer. The current 802.11e draft standard (v. 10.0 [2]) defines two mechanisms, enhanced distributed channel access (EDCA) and hybrid ccoordination function (HCF) controlled channel access (HCCA), both backward compatible with the legacy distributed coordination function (DCF) access mechanism defined by the 1999 standard [1]. A thorough presentation of the main features of 802.11e is provided in companion articles within this special issue [3, 4].

In this article we focus on the performance effectiveness of the priority mechanisms defined in the EDCA specification. EDCA considers two basic priority mechanisms for accessing the channel: different per-class setting of the contention window (CW) backoff parameters (CW_{min} and CW_{max}), and different per-class setting of the idle time after which a transmission may occur (arbitration interframe space, AIFS). Once a station accesses the channel, EDCA also provides the ability to differentiate the time interval for which a station is authorized to hold the channel (transmission opportunity, TXOP). Since a thorough understanding of the impact of different TXOP settings on service differentiation is somewhat immediate, and the TXOP feature is not mandatory and can be disabled, in the following we do not analyze this supplementary mechanism.

Performance evaluation of 802.11e/EDCA has been thoroughly carried out in recent literature. Different aspects have been investigated, including the coexistence of data, voice, and video applications [5, 6]; the need to integrate MAC-level service differentiation mechanisms with admission control policies [7, 8]; system capacity evaluation and the impact of different MAC parameter settings [9, 10]; and so on. All these

This work has been partially supported by the Italian Ministry of Education, University and Research (MIUR) in the frame of the Italian project PRIN-TWELVE.

Access category	CW_{min}	CW_{max}	AIFSN
AC_BK	*aCWmin*	*aCWmax*	7
AC_BE	*aCWmin*	*aCWmax*	3
AC_VI	*aCWmin/2*	*aCWmin*	2
AC_VO	*aCWmin/4*	*aCWmin/2*	2

■ Table 1. *EDCA default settings.*

works, as well as many other 802.11e/EDCA performance studies in the literature, not mentioned here for reasons of space, derive performance figures such as (per-class) throughput, delay statistics and fairness indexes. These performance indicators have a fundamental impact in terms of system dimensioning and parameters engineering, since they quantify the performance experienced by the customers of an 802.11e network.

This article takes a completely different point of view. Rather than providing system engineering insights, our goal is to give the reader thorough understanding of the principles and physical reasons behind the operation of the service differentiation mechanisms proposed in EDCA (AIFS and CW_{min} differentiation). To this purpose, we mainly focus our study on low-level performance figures that are "internal" to system operation (e.g., probability that specific slot times are occupied by MAC frames of given traffic classes), rather than limiting our investigation to high-level performance figures. Our proposed low-level view results are helpful not only to provide better understanding of the effectiveness of various EDCA mechanisms, but also to better assess more detailed technical issues, such as the important issue of coexistence between legacy DCF stations and EDCA terminals.

CW_{min} and AIFS Differentiation

The EDCA proposal of the IEEE 802.11e Task Group is devised to differentiate the channel access probability among different traffic sources. As explained in greater detail in [3, 4], packets arriving at the MAC (MAC service data units, MSDUs) are mapped into four different access categories (ACs), which represent four different levels of service in contention for the shared medium. Each AC contends for the medium with the same rules as the standard DCF (i.e, wait until the channel is idle for a given amount of interframe space, IFS, and then access/retry following exponential backoff rules). The access probability differentiation is provided by using i) different AIFSs instead of the constant distributed IFS (DIFS) used in DCF, and ii) different values for the minimum/maximum CWs to be used for the backoff time extraction. Then, each AC is specified by the values AIFS[AC], CW_{min}[AC], and CW_{max}[AC]. The AIFS[AC] values each differ for an integer number of backoff slots. In particular, AIFS[AC] = AIFSN[AC]· aSlotTime + aSIFSTime, where AIFSN[AC] is an integer greater than 1 for normal stations and greater than 0 for APs.

Table 1 shows the default values of the channel access parameters defined in EDCA for the four ACs (BK = background, BE = best effort, VI = video, VO = voice). Note that these parameters are not fixed: in each beacon frame, the AP broadcasts the values chosen for each AC. Indeed, these values may also be dynamically adapted according to network conditions. Obviously, the smaller the AIFSN[AC] and CW_{min}[AC], the higher the probability of winning the contention with the other ACs. Separate queues are maintained in each station for different ACs, and each behaves as a single enhanced DCF contending entity. When more than one AC of the same station expires its backoff counter, a virtual collision occurs, and the highest-priority packet among the colliding ones is selected for actual transmission on the radio channel.

In the following, after discussing the fairness property of standard DCF, we introduce and separately analyze into details both CW_{min} and AIFS differentiation. These two prioritization mechanisms are individually evaluated considering as a reference the legacy DCF access, and specifically comparing the performance obtained when EDCA stations compete with standard DCF ones. This allows not only to understand the effect of the service differentiation parameters, but also to tackle the somewhat tricky issue of coexistence with legacy DCF stations. For sake of simplicity, we will assume that each EDCA station supports a single access category, and thus no virtual collision may occur inside the EDCA station.

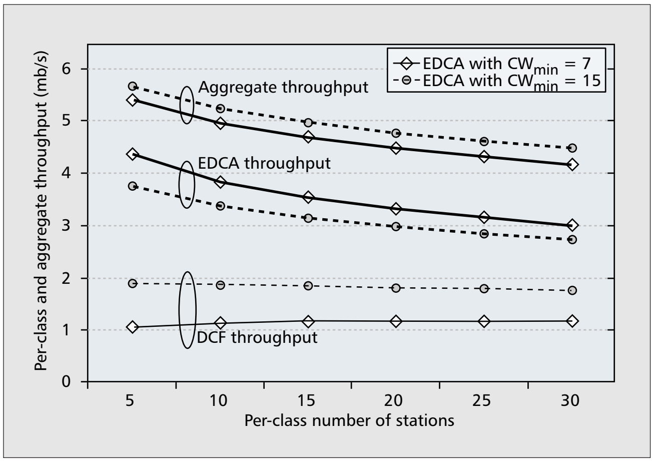
Fairness Issues in Standard DCF

According to the IEEE 802.11 DCF rules [1], a station with a new frame (MPDU) to transmit monitors the channel activity. If the channel is sensed idle for a period of time equal to a DIFS, the station transmits. Otherwise, if the channel is sensed busy (either immediately or during the DIFS), the station continues to monitor the channel until it is measured idle for a DIFS. At this point, the station generates a random backoff interval before transmitting to minimize the probability of collision with packets being transmitted by other stations. In addition, to avoid channel capture, a station must wait a random backoff time between two consecutive packet transmissions, even if the medium is sensed idle in the DIFS time.

DCF adopts an exponential backoff scheme, and employs a discrete-time backoff scale. The time immediately following an idle DIFS is slotted, and a station is allowed to transmit only at the beginning of each *Slot Time*. At each MPDU transmission, the backoff time is uniformly chosen in the range $(0, W-1)$. At the first transmission attempt for a given MPDU, the value $W-1$ is set equal to a parameter CW_{min}. CW_{min} is equal for all stations, and set to the value 31 in the case of 802.11b. After each unsuccessful transmission, W is doubled, until $W-1$ reaches a maximum value CW_{max}, equal to 1023 in the 802.11b case.

Assuming that all stations receive the same channel quality (or, more restrictively, that ideal channel conditions occur), this operation has been shown to be long-term fair [11] in terms of access probability. This means that on average, the same number of *successful* channel accesses is granted to each contending station or, equivalently, that over a long time interval stations that always have a frame ready for transmission will deliver the same amount of MPDUs.

It is interesting to note that long-term fairness can be derived as a corollary of the fact that in the absence of transmission errors, all stations experience the same probability of collision [12]. In order to understand this statement, we recall from [12] that under the assumption of greedy (saturated) traffic sources, DCF channel accesses can be considered to be composed of slot times of uneven size. A channel slot may be either empty, and thus lasts a slot time as specified by the standard (20 μs in the case of 802.11b), or busy, and thus lasts the amount of time necessary to complete a frame transmission (or a collision among two or more frames), plus the extra DIFS time necessary to permit other stations to access the channel again. Clearly, the time elapsing between two consecutive successful transmissions is related to the backoff param-



■ Figure 1. *DCF vs. EDCA throughput with $CW_{\min}$ differentiation.*

eters used by the station (CW_{min} and CW_{max}) and the probability that a transmitted frame collides.[1] Thus, as long as all stations encounter the same collision probability and employ the same backoff parameters, their time-averaged throughput performance is the same.

$CW_{\min}$ Differentiation

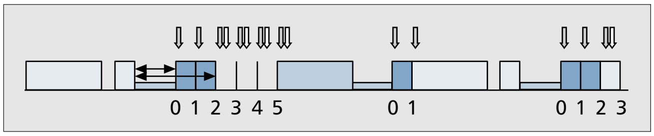
The idea behind the CW_{min} differentiation employed in EDCA is to change the amount of TXOPs provided to each traffic class. A station with a lower value of CW will reduce the average time needed to successfully deliver a packet, and thus experience improved performance in comparison to stations with higher CW values. The average value of the CW can be tuned through differentiated setting of the backoff parameters, and specifically of CW_{min} and CW_{max}.[2] In practice, in low network congestion situations, changes in the CW_{max} parameter have limited effects on throughput differentiation. For example, if we assume that the probability of collision is negligible, a station, regardless of its CW_{max}, will successfully transmit on average once every $CW_{min}/2 + 1$ slots, which corresponds to the average number $CW_{min}/2$ of backoff slots plus the slot used for transmission. Clearly, a station employing a double CW_{min} value will receive (if the collision probability is small) about half of theTXOPs other stations receive.

Figure 1 shows throughput results in a scenario in which N legacy 802.11b DCF stations share the channel with the same number N of EDCA stations. In order to be granted priority over the DCF stations, EDCA must be configured with CW_{min} values smaller than the legacy DCF value $CW_{min} = 31$. In the figure we have thus chosen $CW_{min} = 7$ and 15. EDCA stations have been configured with an AIFSN value equal to 3, which approximates (for reasons to become clear in the following sections) the DIFS setting for legacy DCF stations. In both DCF and EDCA cases, $CW_{max} = 1023$. The packet size has been fixed to 1500 bytes (Ethernet maximum transmission unit, MTU), and the retransmission limit is set to 7 for all the stations. Control frames are transmitted at a basic rate equal to 2 Mb/s, while the MPDU is transmitted at 11 Mb/s. We measured performance in saturation conditions. Although this assumption is not realistic for real-time application, it represents a very interesting case study to derive the limit performance, i.e., the maximum amount of bandwidth that high priority ACs can obtain sharing the channel with best effort DCF stations.

The figure shows that, as expected, for low values of N, the sharing of resources between EDCA and DCF stations is inversely proportional to the employed CW_{min} value. For example, in the case of $N = 5$, we see that the throughput performance of EDCA when $CW_{min} = 7$ is about four times the corresponding throughput performance of DCF (which uses $CW_{min} = 31$); similarly, when $CW_{min} = 15$, it is about double the DCF throughput.

We note that as the number of competing stations grows, the EDCA throughput significantly reduces, while the DCF one decreases only slightly. This phenomenon is not desirable, since performance degradations due to network congestion should be attributed first to best effort stations.

Finally, the figure shows that smaller CW_{min} values lead to smaller aggregate throughput. This is an obvious drawback of CW_{min} differentiation: the performance differentiation is paid for in terms of aggregate performance. This phenomenon is easily explained by considering that the reduction of the CW_{min} value may significantly increase the probability of collision on the channel, thus reducing the overall effectiveness of the random access mechanism.



■ Figure 2. *Slotted channel accesses and protected slots example.*

AIFS Differentiation

AIFS differentiation is motivated by a completely different (and somewhat more complex) physical rationale. Rather than differentiating the performance by changing the backoff structure (through different settings of the CW_{min} and CW_{max} parameters), the idea is to reserve channel slots for the access of higher-priority stations.

This is accomplished by using different AIFS values for different traffic classes. The AIFS is the amount of time a station defers access to the channel following a busy channel period. Once an AIFS has elapsed, the station access is managed by the normal backoff rules. Figure 2 graphically illustrates the AIFS differentiation mechanism for the case of two traffic classes. After every busy channel period, each station waits for a time equal to its AIFS value. If, as in the figure, the AIFS values are different, there is a period of time in which the stations with shorter AIFS values (the higher-priority stations) may access the channel, while the stations with longer AIFS values (lower-priority stations) are prevented from accessing the channel.

The EDCA specification imposes that AIFS values differ

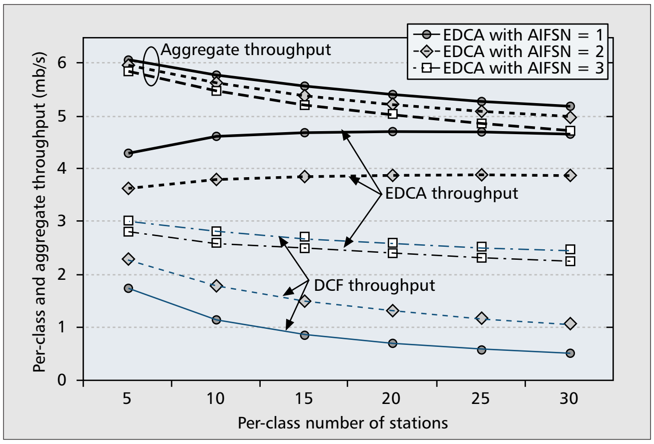
---

[1] *More formally, if* p *is the probability of collision, and neglecting the effect of the retransmission limit, the average time (measured in slots as defined above) between two consecutive successful transmissions is readily obtained as*

$$\frac{W_0}{2}+1+p\left(\frac{W_1}{2}+1+p\left(\frac{W_2}{2}+1+\ldots\right)\right),$$

where $W_i = \min 2^i(CW_{\min} + 1) - 1, CW_{\max})$ *is the backoff window used for the* i*th retransmission, and* $W_i/2 + 1$ *is the average duration, in slots, of the* i*th backoff period including the subsequent transmission slot. For small* p*, this expression can be approximated as* $CW_{\min}/2 + 1$.

[2] *In early versions of the 802.11e draft specifications, a further parameter conisdered for differentiation was the persistence factor (i.e., the multiplicative factor for the CW increment after a collision), which is set to 2 in the binary exponential backoff used by DCF. But it was soon understood that its effect is similar to, though less effective than, change in just the* $CW_{\min}$ *and* $CW_{\max}$ *values. Hence, it was abandoned in later standardization stages.*



■ Figure 3. *DCF vs. EDCA throughput with AIFS differentiation.*

for an integer number of slot times. This implies that the channel access can be still considered slotted, and stations may access the channel only at the discrete time instants indicated in Fig. 2 by arrows. We note that there are some discrete instants of time, hereafter referred to as *protected slots*, where only high-priority stations may access the channel. In the figure, the protected slots are shaded and indicated by a single arrow. A low-priority station can access the channel only if no high-priority station has accessed the channel in one of the previous protected slots (in the example considered, the difference between AIFS values has been set to 2, and the protected slots are those indexed as 0 and 1, while slots numbered from 2 may be accessed by both classes).

A fundamental issue of AIFS differentiation is that protected slots occur after every busy channel period. This implies that the percentage of protected slots significantly increases as long as network congestion increases. In fact, a greater number of competing stations implies that the average number of slots between consecutive busy channel periods reduces, and thus the fraction of protected slots over the total number of idle slots gets larger.

Figure 3 shows throughput results in a scenario in which N legacy DCF stations share the channel with the same number N of EDCA stations. EDCA stations have been configured with the standard DCF backoff parameters (CW_{min} = 31 and CW_{max} = 1023). All other simulation settings are the same as in Fig. 1.

First, Fig. 3 shows that EDCA stations configured with an AIFSN value equal to 3 achieve performance close to that of legacy DCF stations. This counterintuitive result (we recall that a DIFS is equal to an AIFS with AIFSN = 2) requires detailed analysis of the backoff counter decrement mechanism used in EDCA, and is justified in the second part of this article.

Moreover, Fig. 3 shows that performance depends dramatically on the AIFSN setting. It is remarkable to note that by lowering the EDCA AIFSN setting of just one slot, AIFS differentiation shows impressive effectiveness in protecting the EDCA priority traffic from the legacy DCF traffic, especially in the presence of network congestion (e.g., 30 + 30 stations). Indeed, AIFS differentiation correctly reacts to network congestion by penalizing DCF stations, while EDCA stations do not experience an aggregate throughput reduction (which, on the contrary, actually shows a slight increase).[3]

The analysis of the aggregate AIFS throughput curves also leads to a very interesting conclusion: contrary to CW_{min} differentiation, *the AIFS mechanism is beneficial in terms of aggregate throughput performance*. This is a direct consequence of the fact that AIFS differentiation introduces protected slots in which a lower number of stations compete, thus increasing the effectiveness of the overall random access mechanism.

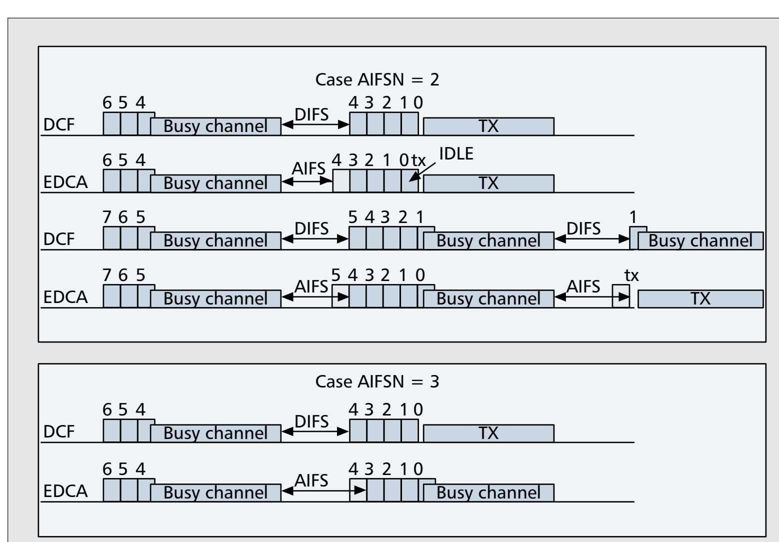
A Closer Look

The previous discussion highlights the different operation and behavior of AIFS and CW_{min} differentiation, and has shown how these translate into high-level performance figures (specifically, throughput performance has been used as a benchmark). In order to understand how these mechanisms operate in scenarios where EDCA stations compete with legacy DCF terminals, it is necessary to take a closer look at some further technical details and provide additional performance insights in terms of low-level performance metrics.

Backoff Counter Decrement Rules

EDCA slightly differs from DCF in terms of how the backoff counter is managed (decremented, frozen, resumed). However, such an apparently minor difference (which might perhaps appear as a technicality) indeed has some important consequences in terms of performance of EDCA access categories, especially when they compete with legacy DCF stations.

In standard DCF, the backoff counter is decremented at each idle slot time, frozen during channel activity periods, and resumed after the medium is sensed idle again for a DIFS interval. This implies that a legacy DCF station, after a DIFS, resumes the backoff counter to the discrete value the station had at the instant of time the busy channel period started. An illustrative example is shown in Fig. 4. Here, a busy channel period (i.e., a transmission from one or more other stations) starts while the backoff counter of the considered DCF station is equal to 4. This value is frozen during the busy channel period and resumed, again to value 4, only a DIFS after the



■ Figure 4. *Backoff counter management in EDCA and DCF.*

[3] *However, the reader should be careful to note that the aggregate throughput is shared among the EDCA stations, and thus, as the number of EDCA stations grows, the per-station throughput ultimately reduces. If a minimum rate must be guaranteed to each station (as in the case of voice or video flows), solutions devised to enforce an upper bound on the number of competing stations (e.g., admission control mechanisms) are necessary.*

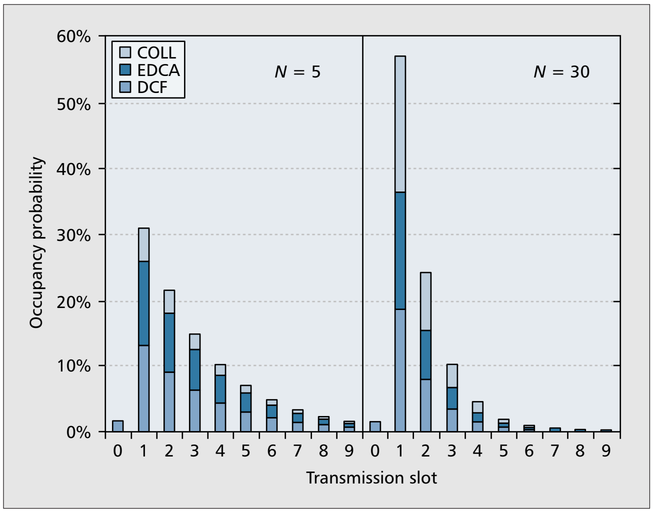


Figure 5. *Per-slot occupancy probability — AC_BE vs. DCF.*

end of the busy period. As a consequence, it is decremented to value 3 only a slot after the DIFS.

In EDCA, the backoff counter is also decremented at every idle slot time and frozen during channel activity periods. But it is resumed one slot time before the AIFS expiration. This means that when the AIFS timer elapses, the backoff counter will already be decremented by one unit. Moreover, since a single MAC operation per slot is permitted (backoff decrement or packet transmission [2, clause 9.9.1.3]), when the counter decrements to 0, the station cannot transmit immediately, but has to wait for a further backoff slot if the medium is idle, or a further AIFS expiration if the medium is busy.

In order to understand how these different rules affect the channel access probability, refer to the example shown in Fig. 4. Let us first focus on the case AIFSN = 2 (top figure), which corresponds to using an AIFS equal to the DCF DIFS. In the example, two stations encounter a busy channel period with the same backoff counter value. However, at the end of the channel activity, we see that the DCF station resumes its counter to a value equal to the frozen value (4 in the example), while the EDCA station resumes and decrements its counter. In the case of a single busy channel period encountered during the backoff decrement process, this difference is compensated for by the fact that the EDCA station has to wait an extra slot; unlike the DCF station, it transmits in the slot following the one in which the backoff counter is decremented to 0 (as illustrated in the two top diagrams of Fig. 4).

However, in the presence of several busy channel periods encountered during the backoff decrement process (which is very likely to happen in the presence of several competing stations), the EDCA station gains a backoff counter decrement advantage for every encountered busy period with respect to the DCF station. This implies that for an AIFSN equal to a DIFS, the EDCA station has an advantage over DCF. Indeed, Fig. 4 shows that there is also a second reason why, with the same AIFSN = 2, the EDCA gains priority over DCF stations. In fact, as shown in the figure, an EDCA station may actually transmit in the slot immediately following a busy channel period (it is sufficient that the busy channel period be encountered while the backoff counter was equal to 0 — last case in the figure). Conversely, a DCF station cannot defreeze a backoff counter value equal to zero. Thus, the only case in which it can access the slot immediately following a busy period is when it extracts a new backoff counter after a successful transmission, exactly equal to 0.

In order to synchronize the EDCA and DCF backoff decrements, it appears appropriate to set the AIFSN value equal to 3. In this case, as we can see in Fig. 4, although the EDCA station has a higher IFS, after each busy slot the backoff evolution of the two target stations is the same. However, since the EDCA station has to wait for a further channel slot after the counter expiration, the access probabilities of the two stations does not coincide, since for a given extraction, the EDCA station has always to wait for a slot more than the DCF station. However, this results in just a slightly higher access probability for the DCF station (loosely speaking, the EDCA station resembles the operation of a DCF station with a CW_{min} value increase of just one unit).

Coexistence of EDCA AC_BE and Legacy DCF Stations

The throughput results shown in Fig. 3 show that for the same CW parameters, EDCA throughput performance is similar to that of legacy stations when the AIFSN parameter is set to the value 3 (i.e., the EDCA AC_BE, Table 1), rather than to a legacy DIFS (i.e., AIFSN = 2). The discussion carried out in the previous section has provided a qualitative justification.

The goal of Fig. 5 is to back up the previous qualitative explanation with quantitative results. To this purpose, we have numbered slots according to the time elapsed after a busy channel period. The slot immediately following a DIFS is indexed as slot 0. Given the end of each busy period, the next transmission will occur after an integer number x of idle slots: we refer to x as the transmission slot. Under the assumption of ideal channel conditions, a successful transmission occurs if, in a transmission slot, only one station transmits; otherwise, a collision occurs. Figure 5 reports the probability distribution that a transmission occurs at a given slot for two different load scenarios: N = 5 (i.e., 5 EDCA stations competing with 5 DCF stations) and N = 30. Only the first 10 slots are plotted, since most transmissions are originated after very few idle backoff slots. In addition, the figure further details in different shades the probability that a transmission occurring at a given slot results in a collision, success for an EDCA station, or success for a DCF station.

Figure 5 shows that DCF stations are the only ones that can transmit in the slot immediately following the last busy period. Also, it confirms that a transmission in slot 0 is always successful (as it is originated by a station that has just terminated a successful transmission). Indeed, a trans-

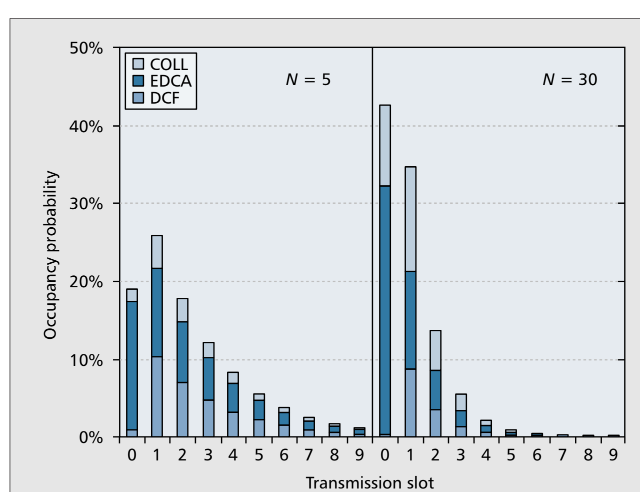


Figure 6. *Per-slot occupancy probability — AIFS differentiation.*

mission in the slot immediately following a busy period is a rare event, since it requires the station that has just experienced a successful transmission to extract a new backoff counter exactly equal to 0. The figure also shows that in the slots with index greater than 0, DCF and EDCA stations experience almost the same success probability, with a negligible advantage for DCF. For example, in the case $N = 5$ a DCF success occurs, almost constantly through the various slot indexes, in about 42.5 percent of the cases vs. 41 percent of EDCA, while for $N = 30$ these numbers reduce to about 32.5 percent and 31.3 percent, respectively, due to the increased probability of collision.

The fundamental conclusion is that *by using AIFSN = 3, an EDCA station can be set to operate approximately as a legacy DCF station*. With reference to the proposed EDCA parameter settings reported in Table 1, we thus conclude that an EDCA station belonging to AC_BE will experience similar performance to a legacy DCF station. The above quantitative analysis also justifies why DCF shows slightly superior throughput performance over EDCA AC_BE, as found in Fig. 3 under the case of AIFSN = 3.

AIFSN = 2 and Legacy DCF Stations

As shown in Table 1, AIFSN = 2 is the minimal setting allowed for an EDCA station. The rationale is that both AIFSN = 0 and AIFSN = 1 are already reserved in the 802.11 standard for the short IFS (SIFS) and the point coordination function (PCF) IFS (PIFS), respectively.

However, as discussed above, the different mechanism employed in EDCA for decrementing the backoff counter suggests that by using AIFSN = 2 (i.e., AIFS = DIFS), an EDCA station is nevertheless expected to gain priority over a legacy DCF station. This was indeed shown in Fig. 3, and is strikingly confirmed by Fig. 6, which, similar to Fig. 5, reports the probability distribution that a transmission occurs at a given slot for the scenario of N DCF stations competing with N EDCA stations configured with AIFSN = 2 and standard CW parameters ($CW_{min} = 31$ and $CW_{max} = 1023$).

Figure 6 shows, for two different load conditions ($N = 5$ and $N = 30$), how the channel slots are occupied by the contending stations. From the figure we see that slot 0 is almost protected for EDCA stations, since it is rarely accessed by DCF stations. Channel slots with index higher than 0 are instead accessed by both classes with comparable probability.

Figure 6 allows us to see a number of interesting considerations. First, the probability of collision in the protected slots (specifically, slot 0) is lower than in the other slots (e.g., for $N = 5$, a collision in slot 0 occurs only in about 8.5 percent of cases vs. an average of 17 percent in the remaining slots, and these numbers for $N = 30$ become 24.5 vs. 38.5 percent), due to the reduced number of competing stations. Second, and most interesting, as the network load increases, the probability of accessing low-indexed slots significantly increases. The reason is that the number of slots between two consecutive busy channel periods significantly reduces in high load. But this implies that a large amount of accesses occur in slots 0 (more than 40 percent in $N = 30$, Fig. 6), and thus are almost exclusively dedicated to EDCA stations, with a definite gain in terms of service differentiation effectiveness (Fig. 3).

As a conclusion, *the usage of AIFSN = 2 in EDCA (i.e, AIFS = DIFS) provides a significant priority for EDCA stations over legacy DCF stations*. This is an extremely important fact, as it allows AIFS differentiation to be effectively deployed even when DCF stations share the same channel; thus, apparently, there seems to be no room for AIFS levels between the IFSs reserved by the standard (SIFS and PIFS), and the legacy DIFS.

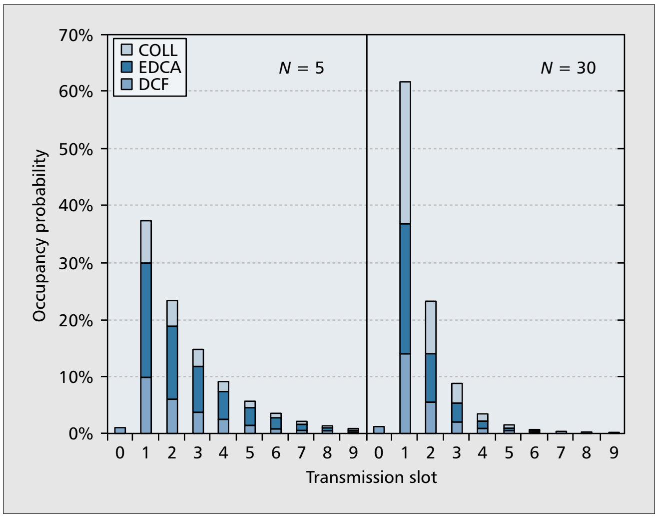


Figure 7. *Per-slot occupancy probability — CW_{min} differentiation.*

Further Remarks on AIFS vs. CW_{min} Differentiation

The capability to dynamically adapt to network congestion without significant performance impairments makes AIFS differentiation an extremely effective approach. This is not so when CW_{min} differentiation is employed: as the number of stations in the network grows, too small CW_{min} settings may lead to a dramatic increase in the probability of collision among frames. This is demonstrated in Fig. 7, which shows the probability that a transmission occurs at a given slot, and furthermore the probability that such a transmission results in success or collision. In Fig. 7 N DCF stations compete with N EDCA stations with parameters $CW_{min} = 15$, $CW_{max} = 1023$, and AIFSN = 3.

By comparing Fig. 7 with Fig. 6, we can immediately observe the different effects of CW_{min} and AIFS as differentiation parameters. In AIFS differentiation, the number of transmission opportunities granted to different classes mainly differ because of the protected slots, while they are similar in slots with index greater than 0. Conversely, in CW_{min} differentiation, the transmission opportunities are differentiated slot by slot. EDCA stations obtain more transmission grants because of the lower backoff expiration times. For example, on the left of Fig. 7, for $N = 5$ the EDCA successful accesses are almost double the DCF ones in all the transmission slots accessed by both classes. This happens because the DCF stations employ a CW_{min} value that is double the EDCA one, and the collision probability is small.

As network congestion increases, it is more and more likely that EDCA stations experience collisions and double the CW, so CW_{min} differentiation is less effective. In fact, on the right of the figure, we observe that for $N = 30$ the increment of the collisions results in a reduction of the difference between EDCA and DCF successful transmissions, which are no more related by an exact 2:1 ratio (e.g. 13 percent DCF successes and 22 percent EDCA successes in the slot number 1). These considerations justify the curves behavior in Fig. 1, where the main effect of the network load increment is the reduction of the EDCA throughput performance. Furthermore, the figure also shows that the probability to collide is significantly greater in the case of CW_{min} differentiation.

Collisions can become dramatic in high load conditions, in the case of reduction of both the CW_{min} and the CW_{max} parameter, as suggested by the default values of Table 1 for the access categories AC_VI and AC_VO. Figure 8 shows the slot occupancy distribution for the case of N legacy DCF

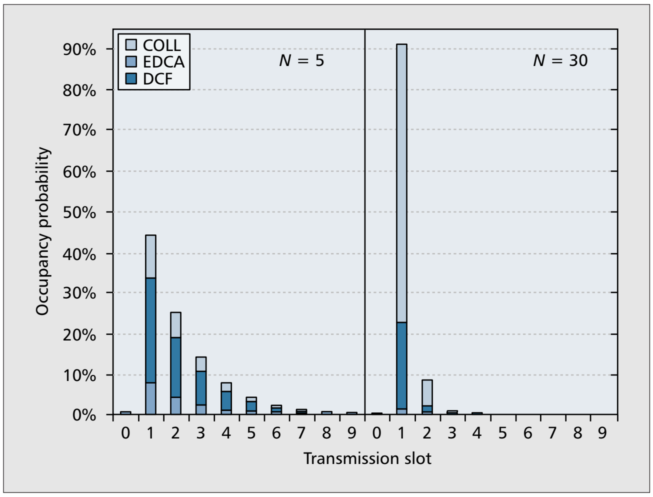


Figure 8. *Per-slot occupancy probability — $CW_{\min}$ and $CW_{\max}$ differentiation.*

stations competing with N EDCA stations that exploit the default CW parameters recommended for the AC_VI access category: $CW_{min} = 15$ and $CW_{max} = 31$. Different from Table 1, the AIFSN value has been set to 3 in order to specifically focus on just the effect of CW_{min} differentiation, not on the joint effect of both mechanisms. The figure shows that the priority advantage of EDCA over DCF is provided through a significant increase of the probability that an EDCA station transmits in a low indexed slot. However, the figure shows that a significant price to pay is a very high collision probability, resulting from the small adopted CW_{max}. This can be dramatic in high load conditions: as the number of stations grows, almost 80 percent of the transmission slots result in a collision. Note that with smaller CW_{min} and CW_{max} values (e.g., those recommended for AC_VO) the situation is much worse, and a significant collision probability may occur even with a small number of competing stations.

To conclude, the CW differentiation operation may lead to situations in which most of the channel slots are trashed by collisions. This can be considered as a fundamental inefficiency of the CW differentiation mechanism in high load conditions, which is unavoidable as long as small CW values have to be employed in order to allow prioritization over the legacy DCF stations.

Conclusions

This article has tackled the issue of the performance evaluation of EDCA service differentiation mechanisms. In comparison with the existing literature, we have put greater attention in analyzing the basic operation of the service differentiation mechanisms, not only in terms of high-level performance figures (such as throughput and delay), but also in terms of per-slot access probability distributions. This has allowed us to gain some additional understanding of the detailed operation of the EDCA mechanisms and the physical reasons behind their operation.

Our conclusion is that AIFS differentiation is a superior mechanism to CW_{min} differentiation for a number of reasons. First, it does not trade off service differentiation with aggregate performance impairment. Second, it is natively adaptive to network congestion. Third, even a single slot difference among AIFS values may result in a substantial difference in terms of performance.

A further contribution of this article is analysis of the coexistence between EDCA and legacy DCF stations. We have shown that the different backoff counter decrement mechanisms used in EDCA allow gaining, in practice, one extra slot to be used for AIFS differentiation: by setting the EDCA AIFS equal to the DCF DIFS (we recall this is the minimum possible setting for the AIFS value, according to the present standard draft), EDCA traffic experiences substantially higher access priority. Our results show that AIFS differentiation is effectively deployable in an hybrid EDCA/DCF scenario even if there seems to be a problem of lack of “space” available.

References.

[1] IEEE 802.11 WG, IEEE Std 802.11, “International Standard for Information Technology. Telecommunications and Information Exchange Between Systems — Local and Metropolitan Area Networks. Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,” 1999.

[2] IEEE 802.11 WG, IEEE Std 802.11e/D10.0, “Draft Amendment to Standard [for] Information Technology — Telecommunications and Information Exchange Between Systems — LAN/MAN Specific Requirements — Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Quality of Service (QoS) Enhancements,” Sept. 2004.

[3] N. Ramos, D. Panigrahi, and S. Dey, “Quality of Service Provisioning in 802.11e Networks: Challenges, Approaches, and Future Directions,” *IEEE Network*, this issue.

[4] Q. Ni, “Performance Analysis and Enhancements for IEEE 802.11e Wireless Networks,” *IEEE Network*, this issue.

[5] S. Choi *et al.*, “IEEE 802.11e Contention-Based Channel Access (EDCF) Performance Evaluation,” *Proc. IEEE ICC ’03*, Anchorage, AK, May 2003.

[6] D. He and C. Q. Shen, “Simulation Study of IEEE 802.11e EDCF,” *Proc. IEEE VTC Spring 2003.*

[7] A. Lindgren, A. Almquist, and O. Schelen, “Quality of Service Schemes for IEEE 802.11 Wireless LANs — An Evaluation,” *ACM MONET*, vol. 8, 2003, pp. 223–35.

[8] Y. Xiao, H. Li, and S. Choi, “Protection and Guarantee for Voice and Video Traffic in IEEE 802.11e Wireless LANs,” *Proc. IEEE INFOCOM ’04.*

[9] J. Zhao, Z. Guo, Q. Zhang, and W. Zhu, ‘Performance Study of Service Differentiation in IEEE 802.11,” *Proc. IEEE GLOBECOM ’02.*

[10] Y. Xiao, “An Analysis for Differentiated Services in IEEE 802.11 and IEEE 802.11e Wireless LANs,” *Proc. IEEE ICDCS ’04*, Mar. 2004.

[11] M. Heusse *et al.*, “Performance Anomaly of 802.11b,” *Proc. IEEE INFOCOM ’03.*

[12] G. Bianchi, “Performance Analysis of the IEEE 802.11 Distributed Coordination Function,” *IEEE JSAC*, vol. 18, no. 3, Mar 2000, pp. 535–47.

Biographies

GIUSEPPE BIANCHI (giuseppe.bianchi@uniroma2.it) received a Laurea degree in electronic engineering from Politecnico di Milano, Italy, in 1990, and a specialization degree in information technology from CEFRIEL, Milano, in 1991. He was an assistant professor at Politecnico di Milano from 1993 to 1998, and an associate professor at the University of Palermo from 1998 to 2003. He is currently an associate professor of networking at the University of Roma Tor Vergata, Italy. He spent 1992 as a visiting researcher at Washington University, St. Louis, Missouri, and 1997 as a visiting professor at Columbia University, New York, New York. His research activities span several areas, including multiple access and mobility management in wireless local area networks, design and performance evaluation of broadband networking protocols, and QoS support in IP networks. He has participated in several European (IST, ITEA, ESA) and Italian (PRIN, FIRB) projects, frequently in managing and/or coordination roles. He was co-organizer of the first ACM workshop on Wireless Mobile Internet (ACM WMI ’01), the first and second ACM Workshops on Wireless Mobile Applications over WLAN Hot-Spot (ACM WMASH ’03 and ’04), and the third IEEE International Workshop on Multiservice IP Networks (IEEE QoS-IP ’05).

ILENIA TINNIRELLO (ilenia.tinnirello@tti.unipa.it) is an assistant professor at the University of Palermo since January 2005. She received a Laurea degree in electronic engineering and a Ph.D. in communications in April 2000 and February 2004, respectively. She spent 2004 as a visiting researcher at Seoul National University, Korea. Her research activity has been mainly focused on wireless networks and, in particular, on multiple access algorithms.

LUCA SCALIA (luca.scalia@tti.unipa.it) received a Laurea degree in electronic engineering from the University of Palermo, Italy, in November 2002. Since January 2002 he is involved as a research engineer in the Italian research project FIRB-Primo, where he deals with 802.11 MAC layer issues for QoS support as well as MAC/PHY cross-layer scheduling solutions for differentiated service access. Since January 2004 he is a Ph.D. student in communications at the University of Palermo.