

ANALYSIS AND CONFIGURATION OF IEEE 802.11E

ALBERT BANCHS AND PABLO SERRANO*

Abstract. The basic Medium Access Control (MAC) algorithm of 802.11 Wireless LANs has been defined for a long time, and considerable analytical and simulation work has been devoted to its evaluation since then. With the recently approved 802.11e standard, which leaves the MAC parameters of this algorithm open for configuration, a deep and complete understanding of the protocol becomes essential in order to determine the optimal configuration of the parameters that best satisfy the requirements of the applications using the WLAN. In this chapter we will provide a comprehensible description and analysis of this protocol and show by means of an example how this analysis can be applied to optimally configure 802.11e.

1. Introduction. Nowadays Wireless LANs (WLANs) have become very popular and are widely used to access the Internet, not only in home environments but also in office settings and hot-spots. The Medium Access Control (MAC) used by most of the WLANs today is based on the Distributed Coordination Function (DCF) mechanism of IEEE 802.11 standard [1]. The 802.11 standard also includes another mechanism, the Point Coordination Function (PCF) which, in contrast to DCF, is centralized. However, for a number of reasons PCF has seen a very limited deployment in today's WLAN cards, most of which implement only the DCF function. The reader is referred to Chapter 4 for an overview of the basic 802.11 protocol.

Recently, the IEEE 802.11 Working Group has approved a new standard, called 802.11e [2], that extends the basic 802.11 protocol with Quality of Service (QoS) capabilities. 802.11e includes two mechanisms, EDCA (Enhanced Distributed Channel Access) and HCCA (HCF controlled channel access). EDCA has been designed as an extension of DCF which (like DCF) is distributed, while HCCA is a centralized mechanism that extends PCF. The focus of this chapter is on the EDCA mechanism of 802.11e.

As compared to the traditional 802.11 DCF protocol, the main novelty introduced by 802.11e EDCA is to leave the MAC parameters of the algorithm open for configuration. This raises the issue of finding the optimal configuration of those parameters. Although the standard includes some pre-defined recommended configurations for the parameters, these configurations are not obtained analytically and do not guarantee an optimized performance. We argue that the development of an analytical model that can predict the performance resulting from a given parameter configuration is a necessary step toward finding the optimal EDCA configuration.

The 802.11 DCF protocol has been defined for a long time, and considerable analytical and simulation work has been devoted to its evaluation since its definition (see chapters 9 and 10 for an overview of this work). However, the 802.11e EDCA mechanism defines some extensions to DCF that require of new analyses in order to understand the performance of this new protocol. Indeed, with the recently approved 802.11e standard, a deep and complete understanding of the protocol becomes essential in order to determine the optimal configuration of the parameters that best satisfy the requirements of the applications using the WLAN.

In this chapter we provide a comprehensible description of the EDCA protocol and analyze its performance. To date, there has been a remarkable amount of work to evaluate the performance of 802.11e EDCA analytically. However, most of the existing analyses [3, 4, 5, 6, 7, 8, 9] are based on the unrealistic assumption that

*The authors are with the University Carlos III of Madrid.

all stations always have packets ready for transmission. This is commonly referred to as *saturation conditions*. Unlike these analyses, the model presented here does not assume saturation conditions but works for finite loads. Although some other previous analyses of non saturated WLANs are available in the literature [10, 11, 12, 13, 14], these are typically valid only for Poisson arrivals, in contrast to ours which does not make any assumption on the arrival process.

The rest of this chapter is structured as follows. We first present a brief description of the EDCA protocol as defined by the 802.11e standard. Next, we provide an analysis of the point of operation of an EDCA WLAN. Based on this analysis, we then present a model for the performance of EDCA. After this, we provide an example of how our analysis can be applied to finding the optimal EDCA configuration in order to meet a given set of service requirements. Finally, the chapter closes with some conclusions.

2. IEEE 802.11e EDCA. This section briefly summarizes the EDCA mechanism as defined in the 802.11e standard [2]. EDCA regulates the access to the wireless channel on the basis of the *channel access functions* (CAF's). A station may run up to 4 CAF's, and each of the frames generated by the station is mapped to one of these CAF's. Then, each CAF executes an independent backoff process to transmit its frames.

A CAF i with a new frame to transmit monitors the channel activity. If the channel is idle for a period of time equal to the arbitration interframe space ($AIFS_i$), the CAF transmits. Otherwise, if the channel is sensed busy (either immediately or during the $AIFS_i$), the CAF continues to monitor the channel until it is measured idle for an $AIFS_i$, and, at this point, the backoff process starts. The arbitration interframe space $AIFS_i$ takes a value of the form $DIFS + n\sigma$, where $DIFS$ and σ are constants dependent on the physical layer and n is a nonnegative integer.

Upon starting the backoff process, the CAF computes a random value uniformly distributed in the range $(0, CW_i - 1)$, and initializes its backoff time counter with this value. The CW_i value is called contention window, and depends on the number of transmissions failed for the frame. At the first transmission attempt, CW_i is set equal to a value CW_i^{min} , called minimum contention window.

The backoff time counter is decremented once every time interval σ as long as the channel is sensed idle, "frozen" when a transmission is detected on the channel, and reactivated when the channel is sensed idle again for an $AIFS_i$. As soon as the backoff time counter reaches zero, the CAF transmits its frame. A collision occurs when two or more CAF's start transmission simultaneously. An acknowledgement (Ack) frame is used to notify the transmitting CAF that the frame has been successfully received. The Ack is immediately transmitted at the end of the frame, after a period of time called short interframe space (SIFS).

If the Ack is not received within a specified Ack Timeout, the CAF assumes that the transmitted frame was not received successfully and schedules a retransmission reentering the backoff process. After each unsuccessful transmission, CW_i is doubled, up to a maximum value CW_i^{max} . If the number of failed attempts reaches a predetermined retry limit R , the frame is discarded.

After a (successful or unsuccessful) frame transmission, before transmitting the next frame, the CAF must execute a new backoff process. As an exception to this rule, the protocol allows the continuation of an EDCA transmission opportunity (TXOP). A continuation of an EDCA TXOP occurs when a CAF retains the right to access the channel following the completion of a transmission. In this case, the CAF transmits a

new frame after a SIFS period following the completion of the transmission. When the station gains access to the channel, it is allowed to retain the right to access it for a duration equal to the transmission opportunity limit parameter ($TXOP_{Limit_i}$). If this parameter is set to zero, a station is allowed to transmit only one packet upon accessing the channel. In the rest of this paper, we assume this setting for the $TXOP_{Limit_i}$ parameter, and concentrate on the analysis of the other three parameters (CW_i^{min} , CW_i^{max} and $AIFS_i$)¹.

In the case of a single station running more than one CAF, if the backoff time counters of two or more CAF's of the station reach zero at the same time, a scheduler inside the station avoids the *internal collision*, granting the access to the channel to the highest priority CAF. The other CAF's of the station involved in the internal collision react as if there had been a collision on the channel, doubling their CW and restarting the backoff process.

The rest of the chapter is devoted to the analysis of the performance of a WLAN as a function of the EDCA parameters $\{CW_i^{min}, CW_i^{max}$ and $AIFS_i\}$ and to the search for their optimal configuration. For simplicity, in the rest of this chapter we assume that stations only execute one CAF and use indistinctly the terms station and CAF.

3. Point of Operation with a single AC. We next analyze the point of operation of an EDCA WLAN in the simplified case of a single AC, i.e. all stations have the same configuration. Since all the stations have the same $AIFS_i$, without loss of generality we take $AIFS_i = DIFS$. The input parameters to the analysis are the number of stations in the WLAN (N), their configuration $\{CW_i^{min}, CW_i^{max}, AIFS_i\}$ and the average arrival rate at the stations (R_i). The contents of this section have been published in [15].

The key variable upon which we base our analysis is τ_i , defined as the probability that a station transmits in a randomly chosen slot time. Based on this variable, the throughput r_i experienced by a station is computed as follows:

$$r_i = \frac{p(s_i)l}{p(s)T_s + p(c)T_c + p(e)T_e} \quad (3.1)$$

where $p(s_i)$ is the probability that a randomly chosen slot time contains a successful transmission of the given station, l is the average packet length, $p(s)$, $p(c)$ and $p(e)$ are the probabilities that a slot time contains a successful transmission, a collision or is empty, respectively, and T_s , T_c and T_e are the slot time durations in each case.

The above probabilities are computed as a function of τ_i as follows:

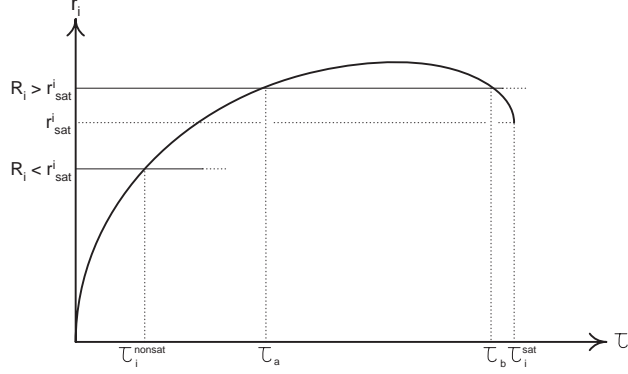
$$p(s_i) = \tau_i(1 - \tau_i)^{N-1} \quad (3.2)$$

$$p(s) = N\tau_i(1 - \tau_i)^{N-1} \quad (3.3)$$

$$p(e) = (1 - \tau_i)^N \quad (3.4)$$

$$p(c) = 1 - p(e) - p(s) \quad (3.5)$$

¹Note that the impact of the $TXOP_{Limit_i}$ parameter is typically small in realistic scenarios. In fact, for real-time traffic parameters are usually set such that the queue never grows to more than one packet, and therefore this parameter is not used, while for data traffic this parameter is set such that only one packet is transmitted upon accessing the channel, in order to avoid degrading the delay performance of real-time traffic.

FIG. 3.1. Throughput as a function of τ_i .

and T_s and T_c are calculated as

$$T_s = T_{PLCP} + \frac{H+l}{C} + SIFS + T_{PLCP} + \frac{ACK}{C} + DIFS \quad (3.6)$$

$$T_c = T_{PLCP} + \frac{H+l}{C} + EIFS \quad (3.7)$$

From the above, we have a formula to compute the throughput as a function of τ_i , $r_i(\tau_i)$. Figure 3.1 illustrates this function in the interval $[0, \tau_i^{sat}]$, τ_i^{sat} being the τ_i resulting from saturation conditions [16]. Note that τ_i can never take a value larger than τ_i^{sat} , since τ_i^{sat} corresponds to the case when all stations always have packets available for transmission. The value of τ_i^{sat} can be calculated from the configuration parameters as follows [16]:

$$\tau_i^{sat} = \frac{2(1-2p)(1-p)^{R+1}}{CW_i^{min}(1-(2p_i)^{m_i+1})(1-p_i) + (1-2p_i)(1-p)^{R+1} + CW_i^{min}2^{m_i}p_i^{m_i+1}(1-2p_i)(1-p_i^{R-m_i})} \quad (3.8)$$

where $p_i = (1 - \tau_i)^{N-1}$ is the probability that a transmission attempt collides and m_i is defined such that $CW_i^{max} = 2^{m_i}CW_i^{min}$. Hereafter we refer with saturation throughput (denoted as r_i^{sat}) to the throughput resulting from applying the above τ_i^{sat} value to Eq. (3.1).

Note that, in stable conditions, either the throughput is equal to the arrival rate, or the throughput is smaller than the arrival rate and the transmission queue grows indefinitely, which results in the station being saturated. Following this, it can be seen from Fig. 3.1 that the values of τ_i at which the WLAN stations can possibly operate as a function of the arrival rate R_i are the following ones, respectively:

- $R_i < r_i^{sat}$. In this case, the only value of τ_i for which the throughput is equal to the arrival rate is the value marked as $\tau_i^{nonstat}$ in the figure. Therefore, this is the τ_i of operation.
- $R_i > r_i^{sat}$. In this case, the throughput under saturation is smaller than the arrival rate, which yields three possible solutions. The first corresponds to operation under saturation (τ_i^{sat}): under these conditions, the queue grows indefinitely and as a result the stations always has a packet available for transmission. In addition, there are two other possible τ_i values of operation for which the throughput is equal to the arrival rate; those are marked in the figure as τ_a and τ_b , respectively.

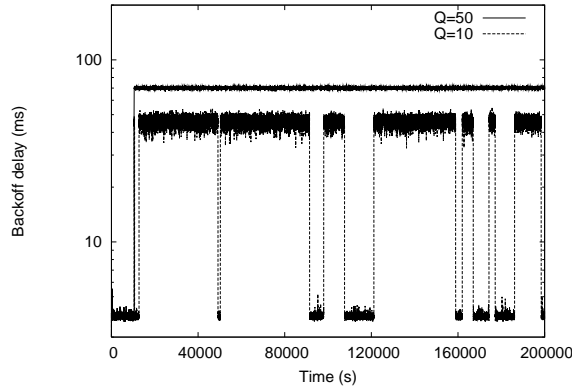
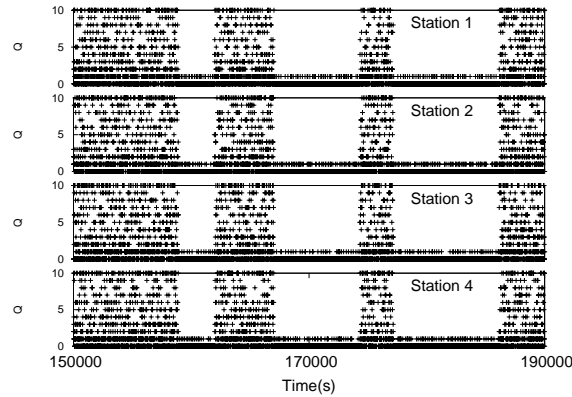
The pending challenge is to identify which of the three possible solutions for the case $R > r_i^{sat}$ corresponds to the actual τ_i of operation. From the above arguments, it follows that there are three possible τ_i values at which the WLAN can operate. However, when operating at these values, the WLAN does not remain fixed at the given τ_i but it suffers some perturbations around the operation point. These perturbations are caused by the probabilistic nature of the access mechanism. Indeed, a number of random consecutive random collisions leads to an increase in the number of transmissions which yields a higher τ_i temporarily, and similarly a random absence of collisions during a certain period yields a smaller τ_i . In the following we analyze the impact of these perturbations on the three equilibrium points identified:

- $\tau_i = \tau_a$. At this τ_i value, a temporary increase of τ_i yields a throughput larger than the arrival rate. If the throughput is larger than the arrival rate, the number of packets available for transmission and as a result the transmission probability τ_i tends to decrease. Similarly, it can be seen that, in case of a temporary decrease of τ_i , the system tends to increase τ_i . This shows that τ_a represents a stable point of equilibrium, since upon suffering perturbations the τ_i tends to go back to the original value.
- $\tau_i = \tau_b$. At this τ_i , a temporary increase of τ_i yields a throughput smaller than the arrival rate. With this throughput, not all the incoming packets are served and as a result the number of packets available for transmission increases, which yields an increase in the transmission probability τ_i . Under these conditions, we have that τ_i tends to increase continuously until reaching τ_i^{sat} . Similarly, it can be seen that a temporary decrease of τ_i leads to further decreasing τ_i until reaching τ_a . We conclude that τ_b represents an unstable equilibrium point and the system never operates at this τ_i except for very short time periods.
- $\tau_i = \tau_i^{sat}$. When operating at τ_i^{sat} , the throughput is smaller than the arrival rate and therefore the queues become full. Perturbations around this τ_i do not move the system away from this point unless they empty the queues of all the stations.

From the above, one of the three possible operation points has been discarded but there still remain two possible alternatives. In order to gain insight into this, we performed the following simulation experiments. We set a WLAN with 40 stations, each with a queue of 50 packets in our first test and of 10 packets in our second test. Traffic was generated at a rate of 110% the saturation throughput. We measured the average delay required to complete the backoff process; the delays resulting from these two tests, averaged over 10 second intervals, are given in Figure 3.2.

The results of Figure 3.2 show severe instability. At the beginning of the simulation, the average delay remains stable around a given value. At some point, it suddenly changes sharply to a new value, showing that the point of operation has changed. For the 50 packet queue length test, it then remains stable at the new value for the rest of the simulation run. For the 10 packet queue length test, the average delay keeps switching back and forth from one value to the other, although it stays most of the time around the high delay value.

In order to gain insight into the level of synchronization of the different stations in these tests, we looked at the evolution of the queue occupancy of 4 randomly chosen stations (station 1, 2, 3 and 4) over time. In Figure 3.3 we show the queue occupancy of the stations for a given time interval of the 10 packet length queue test. In the figure it can be clearly seen that all stations are synchronized: over the periods with

FIG. 3.2. *Unstable behavior.*FIG. 3.3. *Queue occupancy.*

long average delays, all the stations have their queue full, while in the periods with short average delays, the queues of all the stations are empty.

The above behavior is explained as follows. Upon starting the simulation, τ_i increases until reaching τ_a and it remains stable around this value for which the service rate is high enough to serve all incoming packets. However, as argued above, while operating at τ_a there are some perturbations around this value, and, at some point (when many collisions occur) these perturbations can bring τ_i to a value larger than τ_b . At this τ_i , the throughput is smaller than the arrival rate and therefore the queues start to fill up. This further increases τ_i until reaching saturation, when all queues are nonempty.

Since the arrival rate at saturation is larger than the service rate, τ_i remains then stable at the saturation value, although with some perturbations as well. In the case of the 10 packet queue length, these perturbations can, with some probability, empty the queues of all the stations and bring the system back to τ_a . This is however unlikely and, in fact, the WLAN spends most of the time saturated. For the 50 packet queue length case, there are too many packets in the queues, which makes it much harder for the perturbations to empty them. In fact, this does not occur for the entire simulation run.

From the above we conclude that, whenever saturation conditions represent a

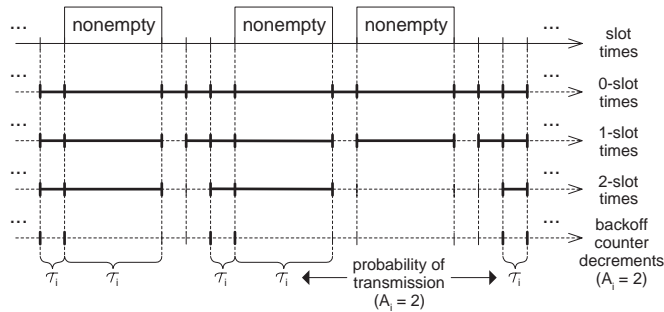


FIG. 4.1. k -slot times and probability of transmission (example with $k = 2$).

possible point of operation, under realistic queue sizes and after some initial transient, the WLAN will operate under saturation (τ_i^{sat}), and it will only operate at a different τ_i when the saturation throughput is larger than the arrival rate. This conclusion will be used in the following section in order to analyze the point of operation with multiple AC's. Note that phenomena similar to the above have been observed in a different context.

4. Point of Operation with multiple AC's. We now extend the analysis of the previous section to the case of multiple AC's. This part has been published in [17]. The input parameters to the analysis presented in this section are the number of stations of each AC (we denote by n_i the number of stations of AC i), the average arrival rate of the stations of each AC (denoted by R_i) and the configuration $\{CW_i^{min}, m_i, A_i\}$ of each AC. Note, in contrast to the previous section in which the $AIFS_i$ was not considered, here we have to consider this parameter since we have multiple AC's.

In the following we present a model to obtain the τ_i of operation of each AC i in the WLAN. We first analyze, in a scenario with multiple AC's, the τ_i of a saturated AC and the τ_i of a non saturated AC, respectively. Then, we combine both analyses in order to compute the τ_i values of all the AC's in the WLAN.

Let us start with the case of a **saturated AC** [9]. With the assumption of [18] that each transmission attempt collides with a constant and independent probability, we can model the behavior of this AC with the same Markov chain as figure 5 of [18]. Then, the probability that a station of a saturated AC transmits upon a backoff counter decrement can be computed from Eq. (3.8), which is a function of the probability p_i that a transmission attempt of a station of AC i collides.

To compute p_i , we proceed as follows. We start by defining a *slot time* as the time interval between two consecutive backoff counter decrements of a station with minimal $AIFS_i$ (i.e., $DIFS$). We say that a slot time is nonempty when it contains a collision or a successful transmission and that it is empty otherwise.

We further define a *k-slot time* as a slot time that is preceded by k or more empty slot times. Note that, since a station with $A_i = k$ starts decrementing its backoff counter only after k empty slot times following a nonempty slot time, we have that the backoff counter decrements of this station coincide with the boundaries of the *k-slot times*. Therefore, a station of AC i , with $A_i = k$, transmits in a *k-slot time* with probability τ_i , and does not transmit in any other slot time (see figure 4.1).

Based on the above definitions, we compute p_i as a function of the probability of

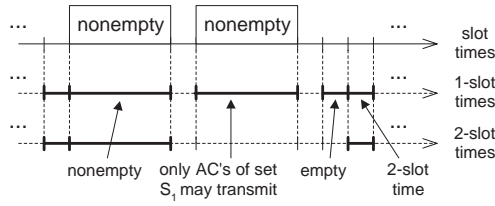


FIG. 4.2. Probability of an empty k -slot time (example with $k = 1$).

an empty k -slot time (denoted by $p(e_k)$) as follows. A k -slot time is empty as long as *i*) the considered station does not transmit, and *ii*) no other station transmits. The latter can be expressed as a function of p_i by noting that the probability of a collision corresponds to the case when some other station transmits. Thus,

$$p(e_k) = (1 - \tau_i)(1 - p_i) \quad (4.1)$$

which yields

$$p_i = 1 - \frac{p(e_k)}{1 - \tau_i} \quad (4.2)$$

Now let us focus on the probability that a given k -slot time is empty. If the previous k -slot time was nonempty, in this k -slot time only the AC's with $A_i \leq k$ may transmit. If the previous k -slot time was empty, the given k -slot time is preceded by $k + 1$ or more empty slot times, which is exactly the definition of $(k+1)$ -slot time, and therefore such a k -slot time is empty with probability $p(e_{k+1})$. Applying this reasoning (see figure 4.2), $p(e_k)$ can be written as

$$p(e_k) = (1 - p(e_k)) \prod_{j \in AC_k} (1 - \tau_j)^{n_j} + p(e_k)p(e_{k+1}) \quad (4.3)$$

where AC_k is the set of AC's with $A_i \leq k$.

Let Δ be the largest A_i in the WLAN. As (with this definition of Δ) in a Δ -slot time all stations may transmit, the following equation holds

$$p(e_\Delta) = \prod_{j \in AC_\Delta} (1 - \tau_j)^{n_j} \quad (4.4)$$

Starting from $\tau_i \forall i$, with equation (4.4) we can compute $p(e_\Delta)$. Then, with equation (4.3) for $k = \Delta - 1$, we can compute $p(e_{\Delta-1})$. Applying this recursively, we can compute $p(e_k) \forall k$. Then, p_i can be computed via equation (4.2) and, finally, τ_i^{sat} can be obtained from equation (3.8). As result, we can express the τ_i of a saturated AC, τ_i^{sat} , as a function of all the τ_i 's. This terminates the analysis for this case.

We next focus on the analysis of a **non saturated AC**. By definition, a station of a non saturated AC sees all the traffic it sends served; hence, the following equation holds,

$$R_i = r_i \quad (4.5)$$

The throughput r_i is computed from equation (3.1), which is a function of $p(s_i)$, $p(e)$, $p(s)$ and $p(c)$. The probability $p(e)$ is, by definition, $p(e_0)$, as all slot times are 0-slot times. This has already been computed for the saturated AC case.

Let us define p_k as the probability that a slot time is a k -slot time. Since a slot time is a k -slot time if and only if the previous slot time is a $(k-1)$ -slot time and it is empty, this probability can be expressed as

$$p_k = p_{k-1}p(e_{k-1}) \quad (4.6)$$

Starting from $p_0 = 1$ (which holds by definition), it follows

$$p_k = \prod_{j=0}^{k-1} p(e_j) \quad (4.7)$$

The probability that a random slot time contains a success of a given station of AC i can be computed as

$$p(s_i) = \sum_{k=A_i}^{\Delta} p(AC_k)p(s_i|AC_k), \quad (4.8)$$

where $p(AC_k)$ is the probability that a randomly chosen slot time is allowed for transmission to only the AC's of set AC_k , and $p(s_i|AC_k)$ is the probability that a slot time in which only the AC's of set AC_k may transmit contains a success of a given station of AC i .

A slot time is allowed for transmission to only the AC's of set AC_k (with $k < \Delta$) if the slot time is a k -slot time but not a $(k+1)$ -slot time². For $k = \Delta$, we have that in a Δ -slot time all AC's are allowed to transmit. Thus,

$$p(AC_k) = \begin{cases} p_k - p_{k+1} & k < \Delta, \\ p_{\Delta} & k = \Delta. \end{cases} \quad (4.9)$$

The probability $p(s_i|AC_k)$ corresponds to the case when the considered station transmits and no other station of set AC_k does:

$$p(s_i|AC_k) = \tau_i(1 - \tau_i)^{n_i-1} \prod_{j \in AC_k \setminus i} (1 - \tau_j)^{n_j} \quad (4.10)$$

The probability that a slot time contains a success can be computed as the sum of the individual success probabilities:

$$p(s) = \sum_{i \in AC_{\Delta}} n_i p(s_i) \quad (4.11)$$

from which $p(c)$ can be computed from $p(c) = 1 - p(e) - p(s)$.

Finally, by combining equations (4.5)–(4.11), we can express the τ_i of a non saturated AC as a function of all the τ_i 's as follows:

$$\tau_i^{nonsat} = \frac{\rho_i(1-p_i^{R+1})(p(s)T_s+p(e)T_e+p(c)T_c)}{l_i(1-\tau_i)^{n_i-1} \sum_{k=A_i}^{\Delta} p(AC_k) \prod_{j \in AC_k \setminus i} (1-\tau_j)^{n_j}} \quad (4.12)$$

which terminates the analysis of a non saturated AC.

²Note that a slot time that is a k -slot time but not a $(k+1)$ -slot time is preceded by exactly k empty slot times, and therefore only the AC's with $A_i \leq k$ (i.e. the AC's of set AC_k) may transmit in such a slot time.

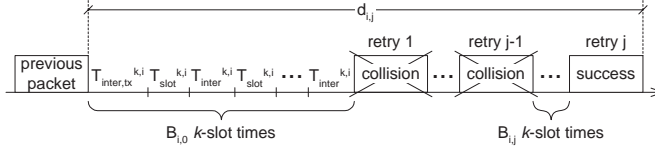
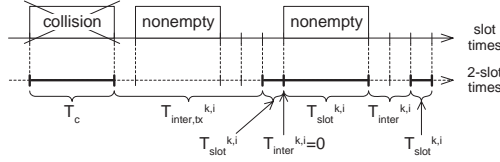
We now combine the above two analyses in order to obtain all the τ_i 's in the WLAN under stationary conditions. From the above analysis we have a method to compute the τ_i of a saturated and of a non saturated AC, respectively; the remaining challenge lies in determining which AC's are saturated and which are not. For this, we proceed step by step as follows in order to classify all the AC's into two sets, one set with the saturated AC's and the other with the non saturated ones:

- In the first step, we consider that all AC's are saturated (i.e., they are all in the set of saturated AC's) and compute their saturation throughputs. Note that, from the above, we can express each τ_i of a saturated AC as a function of all the τ_i 's. Therefore, we have a system of non-linear equations on the τ_i 's that can be resolved using numerical techniques. Once the τ_i values have been derived, we compute the throughput of all AC's by using the above analysis.
- We next compare the throughputs resulting from the first step against the arrival rates. If the throughput of an AC is larger than its arrival rate, we consider from this step on that this AC is not saturated, and move it to the set of non saturated AC's.
- In the second step, we take the new sets of saturated and non saturated AC's resulting from the first step and repeat the throughput computation. Note that, from the above analysis, we can express the τ_i of a saturated and of a non saturated AC, respectively, as a function of all the τ_i 's. Therefore, we have a new system of non-linear equations from which we can obtain the τ_i 's and the corresponding throughputs.
- In the next step, we compare again the throughputs obtained in the previous step for the saturated AC's against their arrival rates, and move those AC's whose throughput is larger than their arrival rate to the set of non saturated AC's. After this reorganization of the sets, we repeat the throughput computation.
- The above is done iteratively until the resulting throughputs of all the saturated AC's are smaller than their arrival rates. This last scenario represents a stable solution, and therefore the throughput values resulting from this step give us the throughput that each AC will obtain in the WLAN under stationary conditions.

Note that the above algorithm follows the conclusion drawn in the previous section. Indeed, we initially assume that all AC's are saturated and only move a given AC from the saturated set to the non-saturated one when there exists no possible solution with this AC saturated. With this method we end up, from all the possible solutions to the system of equations, with the one that has as many saturated AC's as possible. According to the arguments given in the previous section, this is the point of operation at which the WLAN will stabilize after some transient, and it is therefore the solution that we are interested in.

5. Delay Analysis. Next we present a model for the average delay in the point of operation obtained in the previous section. By average delay we understand the time elapsed between the start of the backoff process of a packet and its successful transmission. This is one of the main components of the end-to-end delay in a WLAN. This analysis has been published in [9].

The average average delay experienced by a non-dropped packet of a station of

FIG. 5.1. Average delay in case of j retries.FIG. 5.2. Components of the delay (example with $k = 2$).

AC i , d_i , is computed as

$$d_i = \frac{1}{P_{tx,i}} \sum_{j=0}^R P_{tx,i,j} d_{i,j} \quad (5.1)$$

where $P_{tx,i}$ and $P_{tx,i,j}$ are the probabilities that a packet of a station of AC i is not dropped and that it is successfully transmitted with j retries, respectively,

$$P_{tx,i} = \sum_{j=0}^R (1 - p_i) p_i^j, \quad P_{tx,i,j} = (1 - p_i) p_i^j \quad (5.2)$$

and $d_{i,j}$ is the average delay in case of j retries (see Figure 5.1),

$$d_{i,j} = \sum_{l=0}^j \left(T_{inter,tx}^{k,i} + B_{i,l} (T_{slot}^{k,i} + T_{inter}^{k,i}) \right) + jT_c + T_s \quad (5.3)$$

where $B_{i,l}$ is the average backoff time before retry l , $T_{slot}^{k,i}$ is the average duration of a k -slot time in which the considered station of AC i does not transmit, and $T_{inter,tx}^{k,i}$ and $T_{inter}^{k,i}$ are the average duration of the time between two k -slot times when the considered station transmits and does not transmit in the first one, respectively. Figure 5.2 illustrates these delay components under a given sequence of slot times following a collision of the considered station.

$B_{i,l}$ is computed as

$$B_{i,l} = \frac{CW_i^{min} 2^{\min(m_i, l)} - 1}{2} \quad (5.4)$$

$T_{slot}^{k,i}$ is computed as the sum of probability of success, empty and collision multiplied by the average slot time duration in each case,

$$T_{slot}^{k,i} = p(s_{k,i})T_s + p(e_{k,i})\sigma + (1 - p(s_{k,i}) - p(e_{k,i}))T_c \quad (5.5)$$

where $p(e_{k,i})$ and $p(s_{k,i})$ are the probabilities that a k -slot time in which the considered station does not transmit³ is empty and contains a success, respectively. We compute

³The condition that the considered station does not transmit holds in the computation of all probabilities until (5.11).

$p(e_{k,i})$ by applying a similar reasoning to (4.1),

$$p(e_{k,i}) = \frac{p(e_k)}{1 - \tau_i} \quad (5.6)$$

and $p(s_{k,i})$ by applying the theorem of total probability on the number of empty k -slot times preceding the k -slot time,

$$p(s_{k,i}) = \sum_{j=0}^{N-1-k} p(e_{k,i,j})p(s_{k,i,j}) + p(e_{k,i,N-k})'p(s_{k,i,N-k})' \quad (5.7)$$

where $p(e_{k,i,j})$ is the probability that a k -slot time is preceded by exactly j empty k -slot times, $p(s_{k,i,j})$ is the success probability of such a k -slot time, $p(e_{k,i,N-k})'$ is the probability that a k -slot time is preceded by $N - k$ or more empty k -slot times and $p(s_{k,i,N-k})'$ the success probability of such a slot time.

$p(e_{k,i,j})$ corresponds to having a generic k -slot time not empty followed by j empty k -slot times in which, respectively, AC's of sets $AC_k, AC_{k+1}, \dots, AC_{j+k-1}$ may transmit,

$$p(e_{k,i,j}) = (1 - p(e_{k,i})) \prod_{l=k}^{k+j-1} \prod_{m \in AC_l} (1 - \tau_m)^{n_{m,i}} \quad (5.8)$$

where $n_{m,i} = n_m - \delta_{im}$ (the Kronecker function δ_{im} accounts for the fact that the considered station does not transmit).

To compute $p(s_{k,i,j})$, we note that, after exactly j empty k -slot times, only the AC's of set AC_{k+j} may transmit,

$$p(s_{k,i,j}) = \sum_{l \in AC_{k+j}} n_{l,i} \tau_l (1 - \tau_l)^{n_{l,i}-1} \prod_{m \in AC_{k+j} \setminus l} (1 - \tau_m)^{n_{m,i}} \quad (5.9)$$

$p(e_{k,i,N-k})'$ and $p(s_{k,i,N-k})'$ are computed as follows,

$$p(e_{k,i,N-k})' = 1 - \sum_{j=0}^{N-k-1} p(e_{k,i,j}) \quad (5.10)$$

$$p(s_{k,i,N-k})' = \sum_{l \in AC_{\Delta}} n_{l,i} \tau_l (1 - \tau_l)^{n_{l,i}-1} \prod_{m \in AC_{\Delta} \setminus l} (1 - \tau_m)^{n_{m,i}} \quad (5.11)$$

Finally, $T_{inter}^{k,i}$ and $T_{inter,tx}^{k,i}$ are computed as

$$T_{inter}^{k,i} = (1 - p(e_{k,i})) \sum_{j=0}^k p(e_{k,i,j}) * T_{inter}^{k,i,j} \quad (5.12)$$

$$T_{inter,tx}^{k,i} = \sum_{j=0}^k p(e_{k,i,j}) * T_{inter}^{k,i,j} \quad (5.13)$$

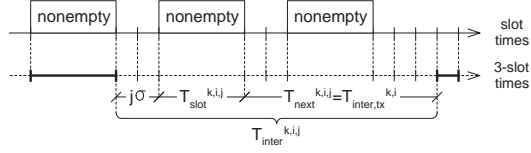


FIG. 5.3. Components of $T_{inter}^{k,i,j}$ (example with $k = 3$ and $j = 2$).

where $p(e_{k,i,j})^*$ is the probability that the interval between a nonempty k -slot time and the next k -slot time starts with exactly j empty slot times, and $T_{inter}^{k,i,j}$ is the average duration between the two k -slot times in this case. For $j = k$,

$$p(e_{k,i,k})^* = \prod_{l=0}^{k-1} \prod_{m \in AC_l} (1 - \tau_m)^{n_m}, \quad T_{inter}^{k,i,k} = k\sigma \quad (5.14)$$

and for $j < k$ (see Figure 5.3),

$$p(e_{k,i,j})^* = \left(1 - \prod_{m \in AC_j} (1 - \tau_m)^{n_m}\right) \prod_{l=0}^{j-1} \prod_{m \in AC_l} (1 - \tau_m)^{n_m} \quad (5.15)$$

$$T_{inter}^{k,i,j} = j\sigma + T_{slot}^{k,i,j} + T_{next}^{k,i,j} \quad (5.16)$$

where $T_{slot}^{k,i,j}$ is the average duration of a nonempty slot time preceded by a nonempty k -slot time followed by j empty slot times, and $T_{next}^{k,i,j}$ is the average duration between the end of this slot time and the next k -slot time,

$$T_{slot}^{k,i,j} = \left(1 - \frac{\sum_{m \in S_j} n_m \tau_m (1 - \tau_m)^{n_m - 1} \prod_{p \in S_p \setminus m} (1 - \tau_p)^{n_p}}{1 - \prod_{m \in S_j} (1 - \tau_m)^{n_m}}\right) \cdot T_c + \frac{\sum_{m \in S_j} n_m \tau_m (1 - \tau_m)^{n_m - 1} \prod_{p \in S_p \setminus m} (1 - \tau_p)^{n_p}}{1 - \prod_{m \in S_j} (1 - \tau_m)^{n_m}} T_s \quad (5.17)$$

$$T_{next}^{k,i,j} = T_{inter,tx}^{k,i} \quad (5.18)$$

Equations from (5.13) to (5.18) form a first order equation on $T_{inter,tx}^{k,i}$, from which we can isolate this term and then derive $T_{inter}^{k,i}$, which terminates the delay analysis.

We validated the accuracy of the model by comparing analytical values with those obtained via simulation. The simulations were performed for a WLAN with the system parameters of the IEEE 802.11b physical layer. We considered two AC's: "voice" and "data". In the first AC, packets were generated at a constant rate of 64 Kbps modeling the behavior of a PCM voice codec. In the second AC, stations always had packets ready for transmission, modeling the behavior of a data transfer. In order to protect voice performance from data stations, the first AC was assigned smaller A_i , CW_i^{min} and m_i values. Two scenarios were studied, whose configurations are given in Table 5.1. Simulations were performed for a varying number of stations (both AC's had n_i stations each).

Figs. 5.4 plots the average delay and throughput values obtained analytically (lines) and via simulation (points) for voice and data AC's in scenarios 1 and 2.

Subplots are given for better observation of the low values. Simulation results are plotted with 95% confidence interval bars. We observe that scenario 1, by using a configuration with a larger level of differentiation, achieves a better preservation of voice performance. For all cases, analytical results closely match simulations, which confirms the accuracy of our model.

TABLE 5.1
EDCA Configurations

Scenario	A_i		CW_i^{min}		m_i	
	Voice	Data	Voice	Data	Voice	Data
1	0	5	16	128	1	5
2	0	1	32	64	2	4

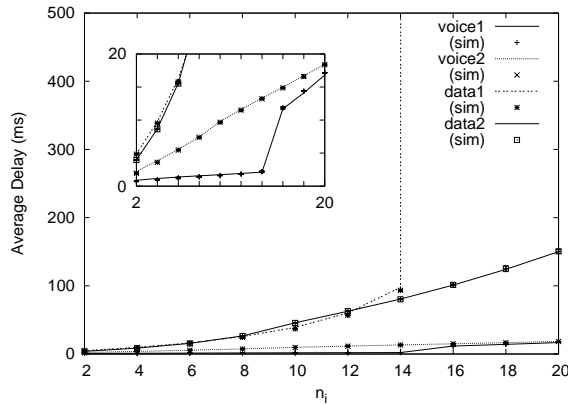
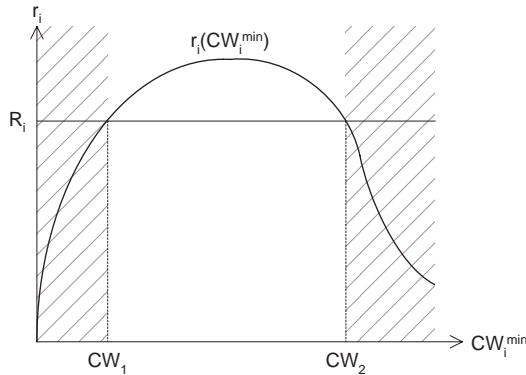
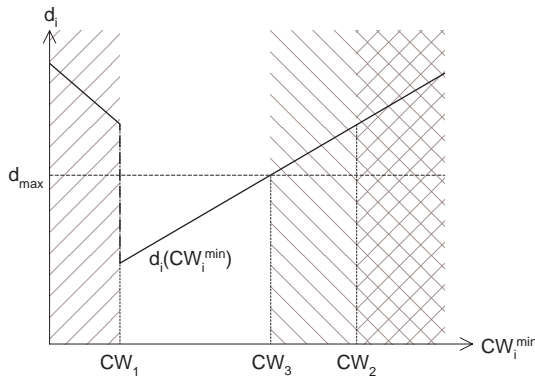


FIG. 5.4. Average delay: simulation versus analysis.

6. Configuration Example. In this section we present an algorithm that, given some desired performance, finds the optimal configuration that satisfies this quality criterion. Specifically, our algorithm takes as input the desired upper bound value for the average delay of all stations (D_{max}) and provides the configuration that satisfies the given requirements.

Since in the given scenario all stations have the same requirements and consequently the same configuration, we take $AIFS_i = DIFS$ for all stations. When the number of stations in the channel is unknown, CW_i^{max} is typically set larger than CW_i^{min} , so that after a collision the CW increases and thus the probability of a new collision is reduced. However, this is not necessary in our case, as the number of stations is known and therefore their CW_i^{min} can be directly set so that the resulting collision probability corresponds to optimal operation. In addition, if we set CW_i^{max} larger than CW_i^{min} , the delay of the packets that suffer one or more collision drastically grows, which harms jitter performance. Based on these arguments, we set $CW_i^{min} = CW_i^{max}$.

The above fixes the parameters $AIFS_i$ and CW_i^{max} , which leaves CW_i^{min} as the only parameter to configure. In the following, we first obtain a lower and an upper bound for CW_i^{min} and then, based on these bounds, we propose an algorithm to calculate the optimal CW_i^{min} .


 FIG. 6.1. CW_i^{min} bounds for throughput.

 FIG. 6.2. CW_i^{min} bound for average delay.

We start by analyzing the CW_i^{min} range that provides good throughput performance. According to Section 3, the WLAN will not be saturated as long as CW_i^{min} is set such that the following condition holds: $r_i(\tau_i^{sat}) > R_i$, where τ_i^{sat} is a function of CW_i^{min} as given by Eq. (3.8).

For any CW_i^{min} that does not meet the above condition, the outgoing rate will be smaller than the incoming one and as a result throughput performance will be degraded. As it can be observed from Figure 6.1, this imposes a lower and an upper bound on CW_i^{min} . Hereafter, we refer to these bounds as CW_1 and CW_2 , respectively.

We now analyze the CW_i^{min} range to meet the given delay performance requirements. According to the average delay analysis of Section 3.8, as long as the WLAN is not saturated (which is given by the above bounds) average delay is an increasing function of CW_i^{min} . As a result, the requirement that average delay cannot exceed a given D_{max} value imposes an additional upper limit on CW_i^{min} , which we refer to with CW_3 . Indeed, as it can be seen from Figure 6.2, for any CW_i^{min} value larger than CW_3 the average delay will not meet the given criterion.

We next propose an algorithm to compute the optimal CW_i^{min} based on the lower bound (CW_1) and two upper bounds (CW_2 and CW_3) obtained above.

From the above, we have that any CW_i^{min} that falls within the bounds meets the given quality criterion. The remaining challenge is to choose one CW_i^{min} value

within this range. Based on the following argument, we choose the largest possible value. As it can be observed from Figure 6.2, in the given range delay performance improves as CW_i^{min} decreases. The problem, however, is that as CW_i^{min} approaches CW_1 , there is the risk of suffering a sharp performance decrease. In order to avoid this, we choose the CW_i^{min} value that, while meeting the given criterion, falls as far as possible from this critical point.

We next present our algorithm resulting from all the above considerations. Note that the algorithm is extremely efficient as each of the steps only involves the calculation of one equation of first or second order:

- In the first step, we compute CW_1 and CW_2 by solving $r_i(CW_i^{min}) = R_i$.
- Next, we compute CW_3 by solving $d = D_{max}$.
- As a final step, the algorithm compares the lower bound (CW_1) with the minimum of all upper bounds (CW_2, CW_3): if $CW_1 > \min(CW_2, CW_3)$, there exists no CW_i^{min} value that satisfies the desired quality criterion and the algorithm indicates that it is not possible to admit the request.
- Otherwise, the algorithm terminates by giving the following optimal configuration: $CW_i^{min} = \min(CW_2, CW_3)$.

The above illustrates how, in a specific environment with average delay requirements, our analysis can be used to determine the optimal configuration that satisfies these requirements.

7. Conclusion. The EDCA mechanism of the 802.11e standard leaves the parameters with which each station contends to access the wireless medium open for configuration. This raises the issue of determining the optimal configuration of those parameters in order to satisfy a given set of requirements of the applications using the WLAN. A first step toward finding this optimal configuration is the analysis of the WLAN behavior for a given configuration.

This chapter has been devoted to the analysis of the WLAN performance as a function of the EDCA parameters. We first have analyzed the single AC and seen that, when saturation is a possible point of operation, after some transient the WLAN stabilizes at this point. Based on this conclusion, we have presented a method to find the point of operation for the multiple AC's case that finds the solution with as many saturated AC's as possible. The performance of this point of operation in terms of average delay has also been analyzed.

In the last section of the chapter, we have shown by means of an example how the presented analysis can be used to find the WLAN's optimal configuration. Specifically, we have considered the single AC case with some requirements on the average delay and have proposed a method to find (if it exists) the EDCA configuration that meets these requirements.

REFERENCES

- [1] IEEE 802.11, *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications*, IEEE Standard, IEEE, 1999.
- [2] IEEE 802.11, *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS)*, Supplement to IEEE 802.11 Standard, 2005.
- [3] J. W. ROBINSON AND T. S. RANDHAWA *Saturation Throughput Analysis of IEEE 802.11e Enhanced Distributed Coordination Function*, IEEE Journal on Selected Areas in Communications, 22 (2004).
- [4] J. HUI AND M. DEVETSIKIOTIS, *A unified model for the performance analysis of IEEE 802.11e EDCA*, IEEE Transactions on Communications, 53 (2005), pp. 1498–1510.

- [5] Z. KONG, D. H. K. TSANG, B. BENSOU AND D. GAO, *Performance Analysis of IEEE 802.11e Contention-Based Channel Access*, IEEE Journal on Selected Areas in Communications, 22 (2004), pp. 2095–2106.
- [6] V. RAMAIYAN, A. KUMAR AND E. ALTMAN, *Fixed point analysis of single cell IEEE 802.11e WLANs: uniqueness, multistability and throughput differentiation*, Proc. of ACM SIGMETRICS, 2005.
- [7] H. ZHU AND I. CHLAMTAC, *Performance Analysis for IEEE 802.11e EDCA Service Differentiation*, IEEE Transactions on Wireless Communications, 4 (2005).
- [8] A. BANCHS AND L. VOLLERO, *Throughput Analysis and Optimal Configuration of 802.11e EDCA*, Computer Networks, 50 (2006).
- [9] A. BANCHS AND L. VOLLERO, *A Delay Model for 802.11e EDCA*, IEEE Communications Letters, 9 (2005).
- [10] P. E. ENGELSTAD AND O. N. OSTERBO, *Non-saturation and saturation analysis of IEEE 802.11e EDCA with starvation prediction*, Proc. of the 8th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems (MSWIM), 2005.
- [11] G.-R. CANTIENI, Q. NI, C. BARAKAT AND T. TURLETTI, *Performance Analysis of Finite Load Sources in 802.11b Multirate Environments*, Computer Communications, 28 (2005), pp. 1095–1109.
- [12] J. W. TANTRA, C. H. FOH, I. TINNIRELLO AND G. BIANCHI, *Analysis of the IEEE 802.11e EDCA under Statistical Traffic*, Proc. of ICC, 2006.
- [13] K. DUFFY, D. MALONE AND D. LEITH, *Modeling the 802.11 Distributed Coordination Function in non-saturated conditions*, IEEE Communications Letters, 9 (2005).
- [14] B. LI AND R. BATTITI, *Analysis of the 802.11 DCF with Service Differentiation Support in Non-saturation Conditions*, Proc. of the Fifth International Workshop on Quality of future Internet Services, 2004.
- [15] P. SERRANO, A. BANCHS, T. MELIA AND L. VOLLERO, *Performance Anomalies of nonoptimally configured WLANs*, Proc. of WCNC, 2006.
- [16] G. BIANCHI, *Performance Analysis of the IEEE 802.11 Distributed Coordination Function*, IEEE Journal on Selected Areas in Communications, 18 (2000).
- [17] P. SERRANO, A. BANCHS AND A. AZCORRA, *Analysis of IEEE 802.11e EDCA under non-saturation conditions*, Personal Wireless Communications, to appear.
- [18] H. WU, Y. PENG, K. LONG, S. CHENG AND J. MA, *Performance of reliable transport protocol over IEEE 802.11 wireless LAN: Analysis and enhancement*, Proc. of IEEE INFOCOM, 2002.