Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Computer Networks 50 (2006) 1749–1768
Computer Networks
www.elsevier.com/locate/comnet

Throughput analysis and optimal configuration of 802.11e EDCA

Albert Banchs [a,*], Luca Vollero [b]

[a] Universidad Carlos III de Madrid, Departamento de Ingeniería Telemática, Av. Universidad 30, 28911 Leganés (Madrid), Spain
[b] Consorzio Nazionale CINI, Via Diocleziano 328, 80124 Napoli, Italy

Received 11 February 2005; received in revised form 21 June 2005; accepted 7 July 2005
Available online 24 August 2005

Responsible Editor: L. Lenzini

Abstract

One of the main challenges with 802.11e EDCA is the configuration of the open parameters of this mechanism, namely *AIFS*, CW_{min}, CW_{max} and *TXOP_limit*. In this paper, we address the issue of finding the optimal configuration of these parameters in order to maximize the throughput performance of the WLAN. We first present a model to analyze the throughput performance of an EDCA WLAN as a function of these parameters. Then, based on this model, we search for the optimal EDCA configuration. We find out that, surprisingly, one of the EDCA parameters, the *AIFS*, is not used in the optimal configuration. We provide closed formulae, based on some approximations, for the configuration of the parameters CW_{min} and CW_{max}. The configuration of the *TXOP_limit* parameter is discussed based on a delay analysis provided elsewhere.

Keywords: 802.11e; EDCA; Wireless LAN; Throughput analysis; Optimal configuration; Weighted max–min fairness

1. Introduction

In recent years, much interest has been devoted to the design of wireless local area networks (WLAN's) with Quality of Service (QoS) support. The Enhancements Task Group (TGe) was formed under the IEEE 802.11 project to recommend an international WLAN standard with QoS support. This standard is called 802.11e and is being built as an extension of the basic WLAN 802.11 standard [1]. While the standardization process of 802.11e is still an ongoing effort, the main features of the upcoming standard have already been

* Corresponding author. Tel.: +34 91 624 8802; fax: +34 91 624 8849.
E-mail addresses: banchs@it.uc3m.es (A. Banchs), vollero@ieee.org (L. Vollero).

1389-1286/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2005.07.008

agreed upon and are unlikely to change. Our work here is based on the 802.11e standard draft [2].

The standard draft defines two different access mechanisms: the *Enhanced Distributed Channel Access* (EDCA) and the *HCF Controlled Channel Access* (HCCA). This paper focuses on the former. As EDCA is based on several open configurable parameters (namely CW_{min}, CW_{max}, *AIFS* and *TXOP_limit*), the challenge with this mechanism lies in its configuration. While there are some configuration recommendations for EDCA in the standard draft, these are not sustained analytically and do not guarantee optimized performance.

In this paper, we address the issue of optimally configuring EDCA. We start with the assumption that real-time traffic, if present, would better be served by HCCA, and EDCA is only used by delay insensitive applications whose performance depends mainly on throughput.[1] This assumption is supported by many papers in the literature (see, e.g., [4–7]), which argue that HCCA is better suited than EDCA for controlling the delay with which frames are delivered. However, which of EDCA and HCCA is better suited for serving real-time traffic is a debatable point. Here we do not aim at contributing to this debate but to look at the implications of using EDCA for delay insensitive applications only.

If one starts with the above assumption, then the throughput obtained by each of the stations sending through the WLAN (to which we refer hereafter as the *throughput distribution* of the WLAN) is the most relevant issue in the configuration of EDCA. In this paper, we first propose a model to analyze the throughput distribution resulting from a generic configuration of EDCA under saturation conditions.[2] Then, based on this model, we study which is the configuration that provides the best throughput distribution according to the *weighted max–min fairness* criterion for throughput allocation. In line with the 802.11e standard draft, our configuration is computed by a central point and then distributed to the stations.

To this date, there has been a remarkable amount of work to evaluate the throughput performance of EDCA via simulation (see, e.g., [8–11]). We argue, however, that an analytic model is a necessary step towards finding the optimal configuration. EDCA has been studied analytically in a number of publications [12–15]. The main novelty in the analysis presented here with respect to those previous works lies in the notion of the *k-slot time*,[3] which is the basis of our analysis. This concept is necessary for the study of the optimal configuration presented in the second part of the paper, which is our main contribution here and would not have been possible with any of the previous analyses.

The paper is outlined as follows. In Section 2 we briefly review the EDCA mechanism of the 802.11e standard draft. In Section 3 we present an analytical model to compute the throughput distribution under saturation conditions resulting from a given EDCA configuration; the model is thoroughly evaluated via simulation. In Section 4 we study which is the optimal configuration of EDCA according to the weighted max–min fairness criterion for throughput allocation; our optimal configuration is validated by means of analytical results. Section 5 contains some architectural considerations, and concluding remarks are given in Section 6.

2. 802.11e EDCA

This section briefly summarizes the EDCA mechanism as defined in the 802.11e standard draft [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 four CAF's, and each of the frames generated by the station is

[1] This type of applications are referred to as elastic applications in [3].

[2] With saturation conditions we mean that all the stations in the WLAN always have packets to transmit.

[3] The concept of *k-slot time* was defined and used in our previous work of [16] to analyze the delay performance of EDCA. In contrast to [16], the focus of this paper is on throughput performance.

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 parameter ($AIFS_i$), the CAF transmits. Otherwise, if the channel is sensed busy (either immediately or during the $AIFS_i$ period), the CAF starts a backoff process. The arbitration interframe space $AIFS_i$ takes a value of the form $DIFS + n\sigma$, where $DIFS$ (the DCF interframe space) 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 the contention window, and depends on the number of transmissions failed for the frame. At the first transmission attempt, CW_i is set equal to the minimum contention window parameter ($CW_i^{\min}$).

As long as the channel is sensed idle the backoff time counter is decremented once every time interval σ. When a transmission is detected on the channel, the backoff time counter is "frozen", and reactivated again after the channel is sensed idle for a certain period. This period is equal to $AIFS_i$ if the transmission is received with a correct CRC, and $EIFS - DIFS + AIFS_i$ otherwise, where $EIFS$ (the extended interframe space) is a physical layer constant.

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 equal to the physical layer constant SIFS (the short interframe space).

If the Ack is not received within a timeout given by the *Ack_Timeout* physical layer constant, the CAF assumes that the frame was not received successfully and reschedules the transmission by reentering the backoff process. The CAF then doubles CW_i (up to a maximum value given by the $CW_i^{\max}$ parameter), computes a new backoff time and starts decrementing the backoff time counter at an $AIFS_i$ time following the timeout expiry. 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 a SIFS period after the completion of the preceding transmission. The period of time a CAF is allowed to retain the right to access the channel is limited by the transmission opportunity limit parameter ($TXOP\_limit_i$).

All the frames transmitted in EDCA include a duration field that indicates the period of time that the transmission and corresponding Ack frame will occupy the channel. Upon receiving a frame, a station sets its NAV (Network Allocation Vector) timer to the value given by the frame's duration field. While this timer does not expire, the station behaves as if the medium was busy. Note that this mechanism protects the transmission from any station that receives correctly the duration field part of the transmission.

The RTS/CTS mechanism is defined as optional for EDCA. With this mechanism, a CAF that has a frame to transmit follows the same backoff procedure as described above, and then, instead of the frame, preliminarily transmits a special short frame called Request To Send (RTS). When the receiving station detects an RTS frame, it responds, after a SIFS time, with a Clear To Send (CTS) frame. The CAF is allowed to transmit its frame only if the CTS frame is correctly received; in this case, the frame transmission proceeds after a SIFS time, and it is followed by an Ack. Since both the RTS and CTS frames include a duration field, the RTS/CTS mechanism ensures that a frame transmission that uses this mechanism will not be disrupted by any station that receives correctly either the RTS or the CTS frames.

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* by 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_i and restarting the backoff process.

As it can be seen from the description of EDCA given in this section, the behavior of a CAF depends on a number of parameters, namely $CW_i^{\min}$, $CW_i^{\max}$, $AIFS_i$ and $TXOP\_limit_i$. These are configurable parameters that can be set to different values for different CAF's. The standard draft groups CAF's by Access Categories (AC's), all the CAF's of an AC having the same configuration, and limits the maximum number of AC's in the WLAN to 4.

The rest of the paper is devoted to the analysis of the throughput distribution of the WLAN as a function of the above EDCA parameters, and to the search for the optimal configuration of these parameters.

3. Throughput analysis

In this section we present an analysis of the throughput distribution of EDCA under saturation conditions and validate it via simulation. We initially assume that each station has only one CAF, and then extend the analysis to the multiple CAF's per station case. When there is only one CAF per station, the terms CAF and station are used indistinctly.

3.1. Model

Consider a WLAN with a fixed number of EDCA stations, each station with one CAF. We refer to a station with a CAF of AC i simply as a station of AC i. Let n be the number of AC's in the WLAN, n_i the number of stations of AC i and $\{CW_i^{\min}, CW_i^{\max}, AIFS_i, TXOP\_limit_i\}$ the EDCA parameters of AC i. Let us define m_i such that $CW_i^{\max} = 2^{m_i} CW_i^{\min}$ and A_i such that $AIFS_i = DIFS + A_i\sigma$. Let us further define S_k as the set of AC's with $A_i \leqslant k$ and N as the largest A_i in the WLAN.

Our analysis is built around two variables: (1) τ_i, the probability that a station of AC i transmits upon decrementing its backoff time counter and (2) $p(c_i)$, the probability that a transmission attempt of a station of AC i collides. To compute these two variables for each AC, our analysis is based on the following two approximations:

- To express τ_i as a function of $p(c_i)$, we use the approximation of [17,18] that each transmission attempt of a station collides with a constant and independent probability, i.e., we consider that $p(c_i)$ can be taken as a constant value.
- To express $p(c_i)$ as a function of the τ_i's, we use the approximation of [19] that backoff times follow a geometric distribution of parameter τ_i, i.e., we consider that stations transmit upon a backoff counter decrement with a constant and independent probability and therefore the τ_i's can be taken as constant values.

We note that the above two approximations are complementary. In fact, assuming constant τ_i's does not yield constant $p(c_i)$'s, and viceversa. We further note that it is intuitive that both approximations result more accurate as long as the n_i and $CW_i^{\min}$ values get larger.

In saturation conditions, a station always has a packet available for transmission, and needs to wait for a random backoff time before transmitting it. Let b represent the backoff time counter, and s the number of retransmissions suffered, after a backoff time counter decrement of the station. With our approximation that each transmission attempt collides with a constant and independent probability, it is possible to model the process $\{b, s\}$ with the same Markov chain as Fig. 5 of [18]. Then, τ_i can be computed as a function of $p(c_i)$ from Eq. (1), as given by [18].

$$\tau_i = \begin{cases} \dfrac{2(1-2p(c_i))(1-p(c_i)^{R+1})}{CW_i^{\min}(1-(2p(c_i))^{R+1})(1-p(c_i))+(1-2p(c_i))(1-p(c_i)^{R+1})}, & R \leqslant m_i, \\ \dfrac{2(1-2p(c_i))(1-p(c_i)^{R+1})}{CW_i^{\min}(1-(2p(c_i))^{m_i+1})(1-p(c_i))+(1-2p(c_i))(1-p(c_i)^{R+1})+CW_i^{\min}2^{m_i}p(c_i)^{m_i+1}(1-2p(c_i))(1-p(c_i)^{R-m_i})}, & R > m_i. \end{cases} \quad (1)$$

Let us define a *slot time* as the time interval between two consecutive backoff time counter decrements of a station with minimal $AIFS_i$, i.e., $AIFS_i = DIFS$. We say that the slot time is empty when there is no transmission ongoing on the channel during this interval. Let us further define a *k-slot time* as a slot time that is preceded by k or more empty slot times, and let $p(e_k)$ denote the probability that a randomly chosen *k-slot time* is empty.

As a station with $A_i = k$ starts decrementing its backoff time counter only after k empty slot times following a nonempty slot time, the backoff time 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 Fig. 1).

To compute $p(c_i)$, we use the approximation that backoff times follow a geometric distribution of parameter τ_i, with which a station of AC i transmits in each *k-slot time* with an independent probability τ_i. With this approximation, the probability $p(c_i)$ that a transmission of a station of AC i collides is equal to the probability that some other station transmits in a *k-slot time*, and the probability that a *k-slot time* is empty can be computed as the probability that a given station of AC i does not transmit multiplied by the probability that no other station transmits:

$$p(e_k) = (1-\tau_i)(1-p(c_i)), \quad (2)$$

which yields

$$p(c_i) = 1 - \frac{p(e_k)}{1-\tau_i}. \quad (3)$$

If the previous *k-slot time* before a given *k-slot time* is not empty, in this *k-slot time* only the AC's with $A_i \leqslant k$ (i.e., the AC's that belong to the set S_k according to our previous definition) may transmit, and, with our approximation, they transmit with an independent probability τ_i. If the previous *k-slot time* is 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 Fig. 2), $p(e_k)$ can be written as

$$p(e_k) = (1-p(e_k))\prod_{j\in S_k}(1-\tau_j)^{n_j} + p(e_k)p(e_{k+1}). \quad (4)$$

As (with our definition of N) in an *N-slot time* the stations of all AC's may transmit, the following equation holds:

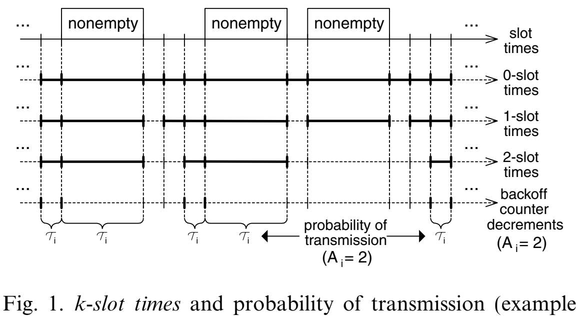


Fig. 1. *k-slot times* and probability of transmission (example with $k = 2$).

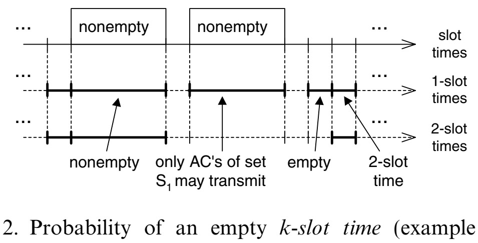


Fig. 2. Probability of an empty *k-slot time* (example with $k = 1$).

$$p(e_N) = \prod_{j \in S_N} (1 - \tau_j)^{n_j}. \quad (5)$$

Starting from τ_i $\forall i$, with Eq. (5) we can compute $p(e_N)$. Then, with Eq. (4) for $k = N - 1$, we can compute $p(e_{N-1})$. Applying this recursively, we can compute $p(e_k)$ $\forall k$. Then, from Eq. (3) we can compute $p(c_i)$ from which we can compute τ_i with Eq. (1). We conclude that with the above equations we can express each τ_i as a function of τ_i $\forall i$, with $i \in \{1, \ldots, n\}$. We therefore have a system of n non-linear equations on the τ_i's that can be resolved using numerical techniques. Uniqueness of the solution is given by Theorem 1,[4] Note that, as n is limited to 4 by the standard draft, our system of equations consists of only 4 equations at most and therefore can be solved at a reasonably low computational cost.

Once the values τ_i $\forall i$ have been obtained, the throughput experienced by a station of AC i, r_i, is computed as the average payload information transmitted by the station in a slot time divided by the average duration of a slot time [18]:

$$r_i = \frac{p(s_i) l}{p(s)T_s + p(c)T_c + p(e)\sigma}, \quad (6)$$

where $p(s_i)$ is the probability that a randomly chosen slot time contains a successful transmission of a given station of AC i, l is the average payload size of a transmission, $p(s)$, $p(c)$ 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 σ are the average slot time durations in each case.

To compute T_s and T_c, we assume here that the RTS/CTS option is not used and that all stations transmit a single frame of constant payload size when they gain access to the channel.[5] Note that, using the formulae given in [17] for T_s and T_c in the RTS/CTS and the variable payload size cases, the model can easily be extended to these cases. With the given assumptions:

$$T_s = T_{\mathrm{PLCP}} + \frac{H + l}{C} + SIFS + ACK + DIFS, \quad (7)$$

$$T_c = T_{\mathrm{PLCP}} + \frac{H + l}{C} + EIFS, \quad (8)$$

where T_{PLCP} is the PLCP (Physical Layer Convergence Protocol) preamble and header transmission time, H is the MAC overhead (header and CRC), ACK is the duration of an acknowledgement frame transmission and C is the channel bit rate.

The probability that a slot time is empty is by definition

$$p(e) = p(e_0). \quad (9)$$

Let us define p_k as the probability that a randomly chosen 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 is empty (see Fig. 2), this probability can be expressed as

$$p_k = p_{k-1} p(e_{k-1}). \quad (10)$$

Starting from $p_0 = 1$ (which holds by definition) and applying the above recursively it follows:

$$p_k = \prod_{j=0}^{k-1} p(e_j). \quad (11)$$

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

$$p(s_i) = \sum_{k=A_i}^{N} p(S_k) p(s_i / S_k), \quad (12)$$

where $p(S_k)$ is the probability that a randomly chosen slot time is allowed for transmission to only the AC's of set S_k and $p(s_i/S_k)$ is the probability that a slot time in which only the AC's of set S_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 S_k, for $k < N$, if the slot time is a *k-slot time* but not a $(k+1)$-*slot time*.[6] For $k = N$, we have that the AC's of set S_N are allowed to transmit in any *N-slot time*. Thus,

[4] The theorems and their proofs can be found at Appendix A.

[5] Note that, with this assumption, the *TXOP_limit*$_i$ parameter has no impact on the throughput distribution. This parameter is no further considered in our throughput analysis.

[6] Note 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 $A_i \leqslant k$ (i.e., the AC's of set S_k) may transmit in such a slot time.

$$p(S_k) = \begin{cases} p_k - p_{k+1}, & k \leqslant N-1, \\ p_N, & k = N. \end{cases} \quad (13)$$

The probability that a slot time in which only the AC's of set S_k may transmit contains a success of a given station of AC i corresponds to the probability that this station transmits and no other station of set S_k transmits:

$$p(s_i/S_k) = \tau_i(1-\tau_i)^{n_i-1} \prod_{j \in S_k \setminus i} (1-\tau_j)^{n_j}. \quad (14)$$

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

$$p(s) = \sum_{i \in S_N} n_i p(s_i), \quad (15)$$

and the probability that it contains a collision can be obtained from

$$p(c) = 1 - p(e) - p(s), \quad (16)$$

which terminates the throughput analysis.

3.2. Multiple CAF's per station

We now present an extension of the above model to the case of multiple CAF's per station. The key approximation of the extension is to assume that the τ_i's obtained in the case of one CAF per station also hold for this case. Indeed, as only a few of the collisions are avoided, the $p(c_i)$ values, and as a result the τ_i values, will be very similar in both cases.

With the above approximation, the only difference with respect to the analysis of the previous section is that, in case of multiple CAF's per station, internal collisions are resolved in favor of the highest priority CAF. Specifically, a transmission of a CAF i is successful as long as no CAF other than the lower priority CAF's of the same station transmits in the same slot time.

According to the above, we rewrite Eq. (14) for a CAF i belonging to AC j as follows:

$$p(s_i/S_k) = \tau_j(1-\tau_j)^{n_j-1-n_j^i} \prod_{l \in S_k \setminus j} (1-\tau_l)^{n_l-n_l^i}, \quad (17)$$

where n_l^i is the number of CAF's of the same station as CAF i that belong to AC l and have a lower priority than CAF i.

3.3. Model validation

To validate the model, in this section we compare its results with those obtained via simulation with a simulator based on the EDCA extension of [20] for *ns-2* [21]. For all tests, stations are located at a distance from each other such that all the frames that do not collide are received without errors. We take a fixed frame payload size of 1500 bytes, m_i equal to 6 (i.e., $CW_i^{\max} = 2^6 CW_i^{\min}$), R equal to 8 and the system parameters of Table 1. Unless otherwise stated, all AC's are configured with $A_i = 0$ and there is only one CAF per station. For the simulation results, average and 95% confidence interval values are given (note that in many cases confidence intervals are too small to be appreciated in the graphs).

1. *Homogeneous configuration:* In the simplified case of only one AC (i.e., same configuration for all stations), our model becomes equivalent to

Table 1
WLAN system parameters

T_{PLCP}	192 μs
H	34 bytes
C	11 Mbps
σ	20 μs
SIFS	10 μs
PIFS	30 μs
DIFS	50 μs
ACK	304 μs
EIFS	SIFS + ACK + DIFS
Ack_Timeout	SIFS + ACK

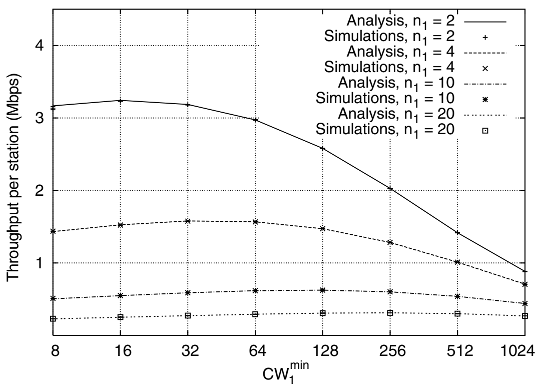


Fig. 3. Homogeneous configuration.

the one of [18] for DCF. Fig. 3 shows the analytical and simulation results obtained for this homogeneous scenario, for varying n_1 and CW_1^{min} values.

2. *Different* CW_i^{min}*:* We next study the throughput distribution when we have two AC's (AC 1 and AC 2) with a different CW_i^{min} each (CW_1^{min} and CW_2^{min}). Figs. 4 and 5 show the throughputs obtained by AC 1 and AC 2 stations when CW_2^{min}/CW_1^{min} is equal to 2 and 10, respectively. Results are given as a function of CW_1^{min} (in the x axis) and for different numbers of stations (namely, 2 and 10 stations per AC, i.e., $n_1 = n_2 = 2$ and $n_1 = n_2 = 10$).

3. *Different* $AIFS_i$*:* Figs. 6 and 7 show the throughput distribution resulting from two AC's sharing the channel with a different $AIFS_i$ each;

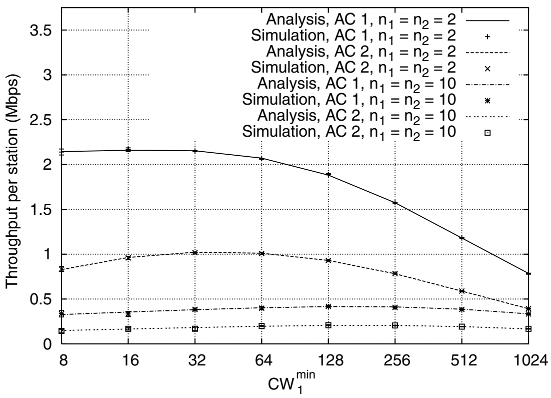


Fig. 4. Different CW_i^{min}; $CW_2^{min}/CW_1^{min} = 2$.

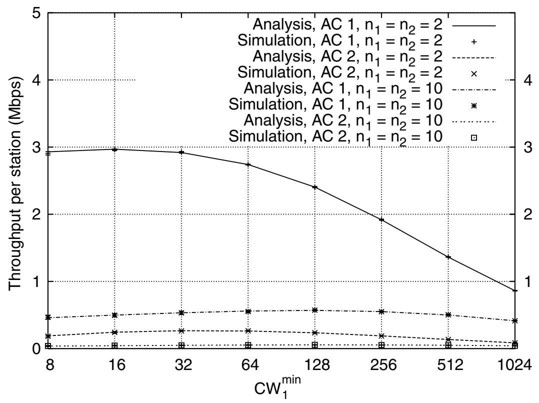


Fig. 5. Different CW_i^{min}; $CW_2^{min}/CW_1^{min} = 10$.

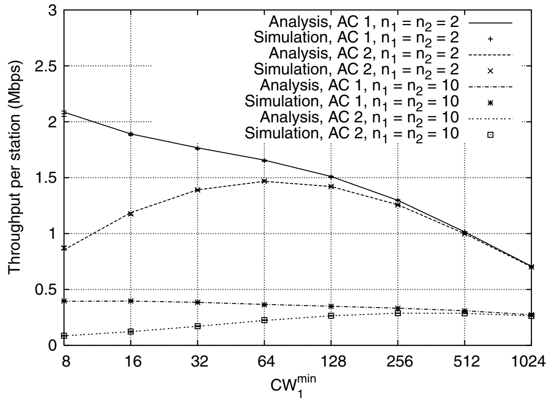


Fig. 6. Different $AIFS_i$; $A_2 = 1$.

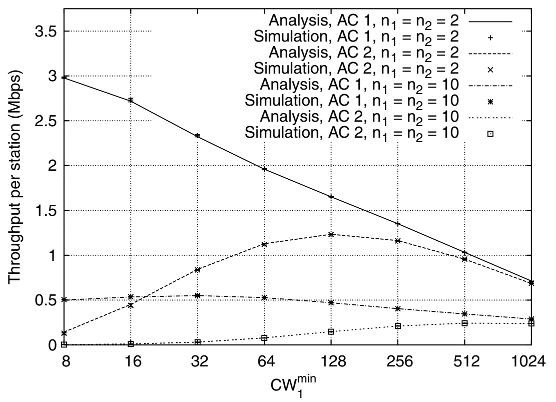


Fig. 7. Different $AIFS_i$; $A_2 = 5$.

A_1 is equal to 0 in both figures, while A_2 takes the values 1 and 5, respectively. Results are given for different CW_i^{min} values (the same for both AC's) and numbers of stations.

4. *Heterogeneous configuration:* Our next scenario corresponds to 4 AC's, $i \in \{1, \ldots, 4\}$, sharing the channel with a different CW_i^{min} and $AIFS_i$ each. Specifically, we take $CW_i^{min} = 2^{i-1}CW_1^{min}$ for $i \in \{2, \ldots, 4\}$ and $A_i = i - 1$ for $i \in \{1, \ldots, 4\}$. Results, in Figs. 8 and 9, are given for different CW_1^{min} values as well as different numbers of stations per AC (2 and 10, respectively).

5. *Multiple CAF's per station:* To validate the model for the multiple CAF's per station case, we study a scenario analogous to the experiment of Fig. 8: two stations share the wireless channel

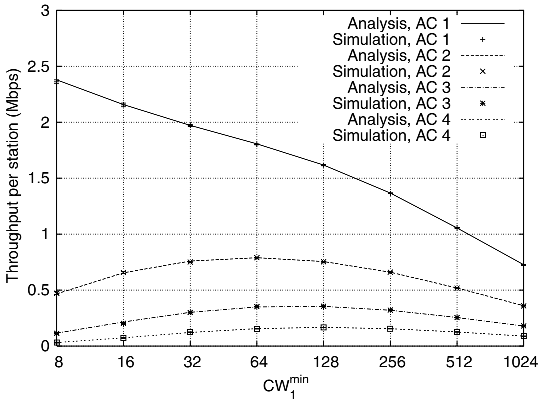


Fig. 8. Heterogeneous configuration; $n_i = 2$.

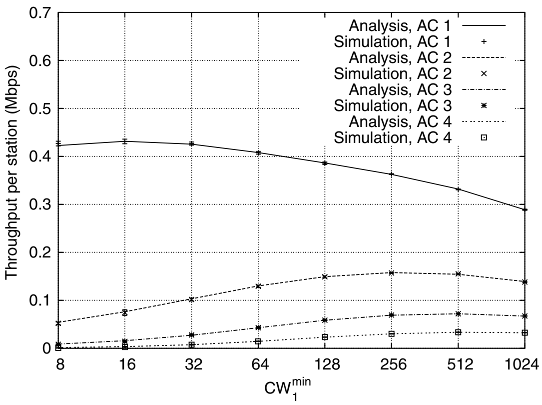


Fig. 9. Heterogeneous configuration; $n_i = 10$.

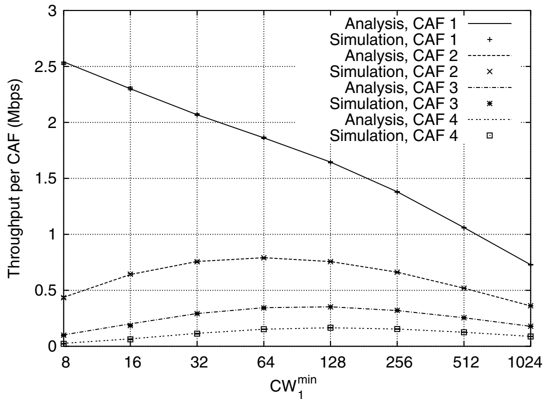


Fig. 10. Multiple CAF's per station.

with four CAF's each, $i \in \{1,\ldots,4\}$, with CAF i belonging to AC i (all AC's configured in the same way as the analogous experiment) and with a higher priority than the CAF's j of the station with $j < i$. The results are given in Fig. 10. Compared to Fig. 8, we observe a (slightly) higher level of differentiation, caused by the additional prioritization introduced.

6. *Discussion:* The simulations performed validate our model, as simulation results match accurately the analytical ones: analytical results (lines) follow closely simulations (points) in all experiments. As predicted by the analysis, accuracy increases with n_i and $CW_i^{\min}$; however, simulations show that also for a small number of stations (as small as 2 per AC) and low $CW_i^{\min}$ values (as low as 8) results are accurate.

4. Optimal configuration for throughput allocation

In the previous section we have analyzed the throughput distribution with 802.11e EDCA as a function of a number of open parameters, the configuration of which is an important unresolved issue. In this section we address, based on this analysis, the issue of finding the optimal configuration of EDCA for throughput allocation.

4.1. Throughput allocation criterion

While there are many different criteria proposed in the literature for throughput allocation, *weighted max–min fairness* [22–24] is a widely accepted one.[7] The weighted max–min fair allocation is the one that maximizes the minimum r_i/w_i in the system, r_i being the throughput allocated to entity i and w_i the entity's weight.

In this section we set our objective to find the configuration that provides weighted max–min fairness in the WLAN, the WLAN stations being

[7] Weighted max–min fairness is, e.g., the criterion provided by Weighted Fair Queueing, which is the most widely implemented mechanism for throughput allocation in wired links. Many works in the literature have aimed at providing weighted max–min fairness in WLAN (see, e.g., [25–28]).

our *entities*,[8] and the saturation throughput of a WLAN station its *allocated throughput*. Note that the saturation throughput in a WLAN corresponds to the notion of *allocated throughput* in weighted max–min fairness: the former assumes that all stations always have packets to transmit, while the latter assumes that all entities are using all the throughput to which they are entitled.

In the rest of the paper we refer to a configuration providing a better throughput performance than another when it provides a greater $\min(r_i/w_i)$.

4.2. $AIFS_i$ configuration

According to the above, our goal is to find the configuration that maximizes the minimum r_i/w_i in the WLAN, r_i being the saturation throughput of a station of AC i and w_i the weight assigned to the AC. From the throughput analysis of Section 3, we have seen that the throughput distribution depends on the $\{\tau_i, A_i\}$ configuration. Our goal can therefore be reformulated as to find the $\{\tau_i, A_i\}$ configuration that leads to $\max(\min(r_i/w_i))$. We initially assume that we have the freedom to fix the τ_i's to any value.

Following Theorem 2, we can restrict our search to the configurations with $A_i = 0\ \forall i$, since, according to the theorem, for any configuration with $A_i \neq 0$ for some i, there exists a configuration with $A_i = 0\ \forall i$ that provides equal or better throughput performance. We note that the implications of this theorem, i.e., that the *AIFS* mechanism performs worse than the *CW* mechanism for throughput allocation, are rather surprising and one of the main findings of this paper. Indeed, as the *AIFS* mechanism avoids collisions between high and low priority stations in some of the slot times (specifically, those following a transmission), it would (wrongly) seem that it is more efficient than the *CW* mechanism, for which collisions between high and low priority stations may occur in any slot time.

Note that, with $A_i = 0\ \forall i$, the throughput analysis of Section 3.1 is simplified to Eq. (6) with $p(s)$ and $p(c)$ computed according to Eqs. (15) and (16), and $p(s_i)$ and $p(e)$ computed as follows:

$$p(s_i) = \tau_i(1-\tau_i)^{n_i-1}\prod_{j\in S\setminus i}(1-\tau_j)^{n_j}, \tag{18}$$

$$p(e) = \prod_{i\in S}(1-\tau_i)^{n_i}, \tag{19}$$

where S is the set of all the AC's in the WLAN.

4.3. $CW_i^{\min}$ and m_i Configuration

Following Theorem 3,[9] we can further restrict our search to the solutions that satisfy

$$\frac{r_i}{r_j} = \frac{w_i}{w_j} \quad \forall i,j, \tag{20}$$

since, according to the theorem, for any configuration that does not satisfy the above condition, there exists a configuration that satisfies the condition and provides equal or better throughput performance. This reduces our problem to finding the τ_i values that maximize the $\min(r_i/w_i)$ under the constraint of Eq. (20).

From Eq. (20) we have

$$r_j = \frac{w_j}{\sum_{i\in S} n_i w_i} r, \tag{21}$$

where r is the total throughput in the WLAN.

With the above, the problem of finding the optimal configuration can be reformulated as to find the τ_i values that maximize r subject to the condition of Eq. (20), as any other set that satisfies this condition will lead to smaller $r_i\ \forall i$, and therefore smaller $\min(r_i/w_i)$.

The total throughput r can be expressed as

$$r = \frac{p(s)l}{p(s)T_s + p(c)T_c + p(e)\sigma} = \frac{l}{T_s - T_c + \frac{p(e)(\sigma - T_c) + T_c}{p(s)}}. \tag{22}$$

[8] We initially search for the optimal configuration when there is only one CAF per station, and use indistinctly the terms CAF and station. In Section 4.5 we extend the search to the multiple CAF's per station case and, in this case, our *entities* are the CAF's.

[9] We note that Theorem 3 and the *CW* configuration analysis presented here partially reuses some of our previous results of [29].

As l, T_s, and T_c are constant, maximizing the following expression will result in the maximization of r:

$$\hat{r} = \frac{p(s)}{p(e)(\sigma - T_c) + T_c}. \tag{23}$$

With the simplified throughput analysis given in Section 4.2, the condition of Eq. (20) can be rewritten as

$$\frac{\tau_i(1-\tau_j)}{\tau_j(1-\tau_i)} = \frac{w_i}{w_j}. \tag{24}$$

Under the assumption of $\tau_i \ll 1\ \forall i$—which is reasonable in optimal operation, as large τ_i values would lead to a high collision probability—Eq. (24) is approximately equivalent to

$$\frac{\tau_i}{\tau_j} \approx \frac{w_i}{w_j} \tag{25}$$

which yields

$$\tau_1^{\text{opt}} = w_1 \frac{\sqrt{(b\sigma)^2 + abc\sigma} - b\sigma}{bc}. \tag{31}$$

Finally, applying Eq. (24) to τ_1^{opt}, we obtain our approximation to the optimal τ_i values:

$$\tau_i^{\text{opt}} = \frac{w_i \tau_1^{\text{opt}}}{w_1(1-\tau_1^{\text{opt}}) + w_i \tau_1^{\text{opt}}}. \tag{32}$$

The remaining challenge is to find the $CW_i^{\min}$ and m_i configuration that leads to the optimal τ_i values obtained above. From Eq. (1), we have that τ_i can be adjusted as a function of two parameters, $CW_i^{\min}$ and m_i. As a consequence, we have one level of freedom to adjust these parameters in order to obtain the desired τ_i. If we fix m_i, then the $CW_i^{\min}$ value that leads to τ_i^{opt} can be computed from Eq. (1) as given by Eq. (33), where $p(c_i)$ is computed from the τ_i^{opt} values according to Section 3.

$$CW_i^{\min} = \frac{(1-2p(c_i))(1-p(c_i)^{R+1})}{(1-(2p(c_i))^{m_i+1})(1-p(c_i)) + 2^{m_i} p(c_i)^{m_i+1}(1-2p(c_i))(1-p(c_i)^{R-m_i})} \left(\frac{2}{\tau_i^{\text{opt}}} - 1\right). \tag{33}$$

and $\hat{r}$ can be approximated by

$$\hat{r} \approx \frac{a(\tau_1/w_1) - b(\tau_1/w_1)^2}{c(\tau_1/w_1) + \sigma}, \tag{26}$$

where AC 1 is taken as reference, with

$$a = \sum_{i \in S} n_i w_i, \tag{27}$$

$$b = \sum_{i \in S} \sum_{j \in S \setminus \{1,\ldots,i\}} n_i n_j w_i w_j, \tag{28}$$

$$c = \sum_{i \in S} n_i w_i (T_c - \sigma). \tag{29}$$

The optimal value of τ_1, τ_1^{opt}, that maximizes $\hat{r}$ can then be obtained by

$$\left.\frac{\mathrm{d}\hat{r}}{\mathrm{d}\tau_1}\right|_{\tau_1=\tau_1^{\text{opt}}} = 0 \Rightarrow bc\left(\frac{\tau_1^{\text{opt}}}{w_1}\right)^2 + 2b\sigma\left(\frac{\tau_1^{\text{opt}}}{w_1}\right) - a\sigma = 0, \tag{30}$$

Note, however, that Eq. (33) does not necessarily yield an integer $CW_i^{\min}$ value; to meet the requirement that contention windows must take integer values, we round $CW_i^{\min}$ to the closest integer, i.e.,[10]

$$CW_i^{\min} = \text{round int}\left(CW_i^{\min}\ (\text{Eq. } (33))\right). \tag{34}$$

In the rest of this paper, unless otherwise specified, we set $m_i = 0\ \forall i$,[11] from which $CW_i^{\min} = CW_i^{\max} = CW_i$. With this setting, Eq. (33) is simplified to

[10] Note that, for $\tau_i^{\text{opt}} \ll 1$, the error resulting from the rounding operation is very small.

[11] Note that in the DCF standard, in which $CW_i^{\min}$ and $CW_i^{\max}$ are statically set, $CW_i^{\max}$ is chosen to be larger than $CW_i^{\min}$ so that after each unsuccessful transmission the contention window is increased, reducing thus the probability of a new collision. However, this is not necessary in our case, as we can directly adjust the contention window values so that the resulting collision probabilities correspond to optimal operation.

$$CW_i = \frac{2}{\tau_i^{opt}} - 1. \tag{35}$$

4.4. $TXOP\_limit_i$ Configuration

The remaining open issue is to find the optimal $TXOP\_limit_i$ configuration. Throughput performance increases with larger $TXOP\_limit_i$ values, since larger transmission times means lower overhead for each transmitted bit. However, $TXOP\_limit_i$ can not be set based only on throughput performance considerations, as throughput performance would be optimized with infinite payload size transmissions, but this would lead to infinite delays, which is clearly undesirable.

Based on the above, we propose to set the $TXOP\_limit_i$ of all the AC's to the maximum acceptable value according to delay and/or other considerations, while configuring the other three parameters (CW_i^{min}, CW_i^{max} and $AIFS_i$) following the algorithm given in this paper. For example, given a maximum allowed delay, if the other EDCA parameters are configured according to the formulae given here, then the value of the $TXOP\_limit_i$ parameter can be computed from our delay model of EDCA in [16].

In the rest of the paper, we assume that the configuration of the $TXOP\_limit_i$ parameter is set to a fixed value (corresponding to a certain payload size) and do not further consider this parameter.

4.5. Multiple CAF's per station

For the multiple CAF's per station case, Eq. (24) is rewritten as follows:

$$\frac{\tau_i}{\tau_j \prod_{k \in H_j} 1 - \tau_k} = \frac{w_i}{w_j}, \tag{36}$$

where CAF i is the highest priority CAF of a station, CAF j is another CAF of the same station and H_j is the set of CAF's of the station with higher priority than CAF j.

Our computation of the optimal configuration in this case is based on the following observation. Since collisions in an optimally configured WLAN are very infrequent, the impact of avoiding some of them due to the multiple CAF's feature is negligible. As a result, the optimal τ_i values will be very similar to the ones obtained in the previous section. In particular, if we keep one of the τ_i's (e.g., the highest priority CAF of each station) to the τ_i^{opt} value of the previous section, and calculate the others such that Eq. (36) is satisfied, then the resulting configuration will very accurately approximate the optimal configuration.

Following the above reasoning, we compute the optimal τ_i of the highest priority CAF of a station using Eq. (32), and then calculate the configuration of the remaining CAF's of the station, in order of decreasing priority, applying Eq. (36) as follows:

$$\tau_j^{opt} = \frac{w_j}{w_i} \frac{\tau_i^{opt}}{\prod_{k \in H_j} 1 - \tau_k^{opt}}, \tag{37}$$

from which the CW_i values are then obtained using Eq. (35).

4.6. Optimal configuration validation

The optimal configuration proposed here is based on a number of approximations. To assess the validity of the configuration proposed, in this section we compare it with the result of performing an exhaustive search over the entire configuration space. Specifically, we evaluate (analytically or via simulation) the throughputs resulting from all possible configurations (or a range wide enough) and choose the one that leads to the maximum $\min(r_i/w_i)$, against which we compare our configuration.

In the following, we refer to the three methods mentioned above as "our approximation" to the optimal configuration, "analytical exhaustive search" and "simulation exhaustive search". Note that the analytical and simulation exhaustive search methods are unfeasible for practical use, as they require a large amount of time and computational resources to find the optimal configuration; our intent here is to use them as a benchmark to assess the accuracy of our approximation.

In experiments 1, 2, 3 and 4 we restrict the exhaustive search to the configurations with $A_i = 0$ and $m_i = 0$. Experiments 5 and 6 confirm that these restrictions do not affect the search for the optimal. Unless otherwise stated, the same parameters as in Section 3.3 are used.

1. *One AC:* We first study the simplest possible scenario in which there is only one AC present in the WLAN with weight $w_1 = 1$. In this case, our objective of maximizing the minimum r_i/w_i is equivalent to finding the CW_1 value that, configuring the AC with this value, maximizes the individual throughputs.

Fig. 11 shows the optimal configuration resulting from our approximation, analytical exhaustive search and simulation exhaustive search, for different numbers of stations. The resulting throughputs are given for each case; the throughputs obtained analytically are represented with lines, and the throughputs obtained via simulation with points and confidence intervals. For the simulation exhaustive search, the simulation throughputs have been obtained by rerunning the simulations for the selected configuration with different seed values.

We observe that the optimal CW_1 values given by our approximation are very close to the ones obtained with the exhaustive search methods, and the resulting throughputs are practically identical. As throughput is the only relevant metric for our objective of maximizing $\min(r_i/w_i)$, we conclude that these results validate our approach.

We also observe from the figure that simulation results match the analytical ones very accurately. In fact, simulations of Section 3.3 show that the accuracy of our model increases with CW_i, and, for non-small CW_i values, analytical results match simulations almost perfectly. The results obtained in the present experiment show that the optimal CW_i for two stations is already relatively large, and it increases with the number of stations. From this, it follows that the throughput results obtained analytically and via simulation will be practically identical for any scenario configured optimally.

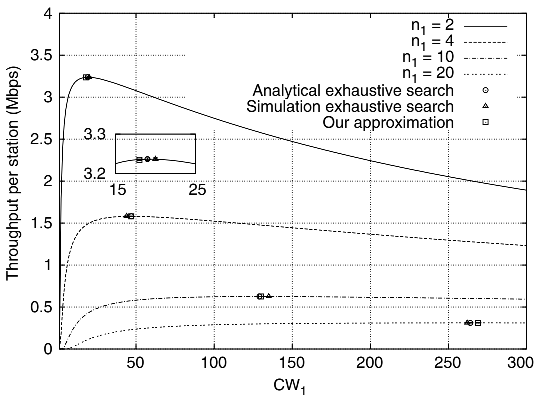


Fig. 11. One AC.

For a number of AC's greater than 1, the simulation exhaustive search method becomes practically unfeasible, as the number of simulations required to explore an area wide enough to be sure that it contains the optimal becomes too large.[12] As a consequence, in the remaining experiments of this section we restrict ourselves to the comparison of our approximation against the analytical exhaustive search method. As, according to above paragraph, analytical and simulation throughputs are practically identical for an optimally configured WLAN, we argue that the difference in terms of the provided $\min(r_i/w_i)$ between the analytical and simulation exhaustive search methods will be negligible.

2. *Two AC's:* In case of two different AC's with weights w_1 and w_2, the problem of maximizing the minimum r_i/w_i acquires two dimensions: the CW_i configuration of AC 1, CW_1, and of AC 2, CW_2. Our goal is therefore to find the CW_1 and CW_2 values that maximize the minimum r_i/w_i among the two AC's.

Fig. 12 gives the $\min(r_i/w_i)$, obtained analytically, with the configuration corresponding to our approximation (points) and with the optimal configuration resulting from performing an analytical exhaustive search in the $\{CW_1, CW_2\}$ space (lines). Results, given for different numbers of stations per AC ($n_1 = n_2 = 2$ and $n_1 = n_2 = 10$) and different weights ($w_1 = 1, w_2 \in [1, 10]$), confirm that our approximation provides a throughput performance very close (almost identical) to the optimal for the two AC's case.

3. *Multiple AC's:* To assess performance in case of multiple AC's, we study a scenario with four AC's, $i \in \{1, \ldots, 4\}$, each AC i with n_i stations ($n_i = 2$ and 10) and a weight $w_i = w_1 + (i-1)j$, with $w_1 = 1$ and $j \in [1, 10]$. Fig. 13 shows the min

[12] Note that the behavior of the throughput around the optimal may be quite flat (see Fig. 11 for $n_1 \geqslant 10$), which makes the area to explore very large.

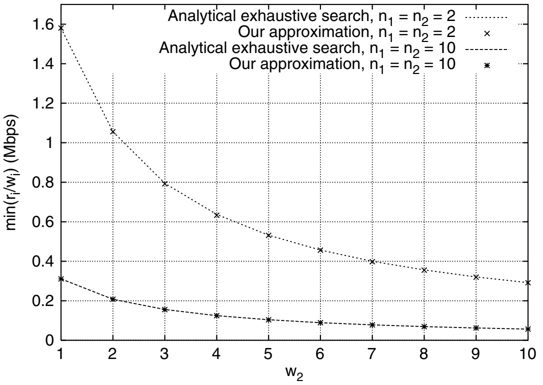


Fig. 12. Two AC's.

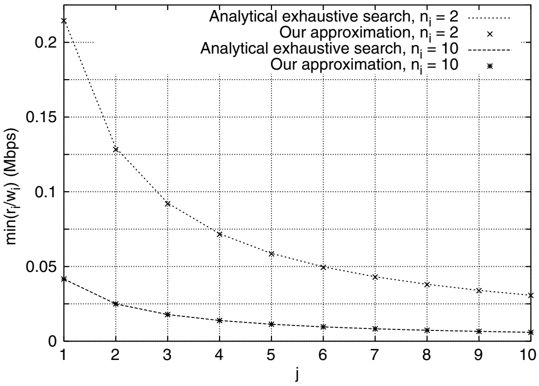


Fig. 14. Multiple CAF's per station.

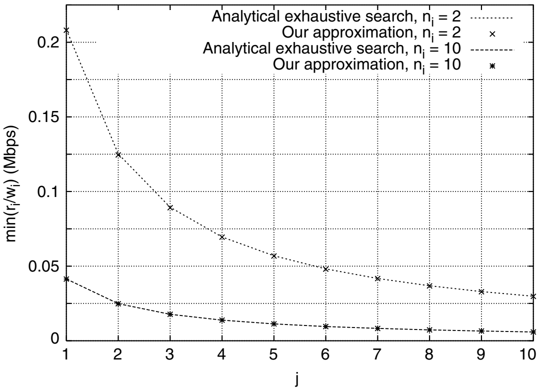


Fig. 13. Multiple AC's.

(r_i/w_i) values, obtained analytically, corresponding to the analytical exhaustive search method and our approximation to the optimal configuration. Results validate our approach also for this case, as the throughput performance given by our approximation is very close to the performance of the analytical exhaustive search method.

4. *Multiple CAF's per station:* Fig. 14 shows the result of running the experiment analogous to Fig. 13 with four CAF's (each with a different weight) per station. Results show that our configuration performs very close to the analytical exhaustive search method. As a result of avoiding some of the collisions, there is some improvement in throughput performance as compared to Fig. 13, although the improvement is so small that it can barely be appreciated in the graphs.

5. $A_i \neq 0$: To illustrate the result of Theorem 2, according to which the optimal configuration is met with $A_i = 0\ \forall i$, we compare the $\min(r_i/w_i)$ values, obtained analytically, from performing an analytical exhaustive search in the $\{CW_i\}$ space for different A_i settings. Table 2 gives the result of this experiment for two AC's, with $A_1 = 0$ and $A_2 \in [0, 10]$. As predicted by Theorem 2, the best performance is achieved when $A_2 = 0$, and then performance decreases as A_2 increases, even though with some irregularities caused by the limited granularity on the τ_i values, which the theorem does not account for. In any case, results validate our choice of $A_i = 0\ \forall i$, as this setting always provides the best performance.

6. $m_i \neq 0$: The fact that contention windows must take integer values introduces some rounding

Table 2
$\min(r_i/w_i)$ (Kbps), $A_i \neq 0$

A_i	$n_1 = n_2 = 2$		$n_1 = n_2 = 10$	
	$w_1 = 1$ $w_2 = 2$	$w_1 = 1$ $w_2 = 10$	$w_1 = 1$ $w_2 = 2$	$w_1 = 1$ $w_2 = 10$
0	1056.11	291.73	207.46	56.70
1	1051.88	291.65	207.09	56.69
2	1052.56	291.58	206.85	56.69
3	1051.94	291.57	206.49	56.68
4	1049.83	291.54	206.05	56.67
5	1047.72	291.49	205.53	56.66
6	1044.32	291.21	204.94	56.64
7	1040.60	291.23	204.27	56.63
8	1037.20	290.88	203.56	56.61
9	1033.38	290.90	202.77	56.58
10	1029.17	290.65	201.92	56.55

Table 3
$\min(r_i/w_i)$ (Kbps), $m_i \neq 0$

m_i	$n_1 = n_2 = 2$		$n_1 = n_2 = 10$	
	$w_1 = 1$ $w_2 = 2$	$w_1 = 1$ $w_2 = 10$	$w_1 = 1$ $w_2 = 2$	$w_1 = 1$ $w_2 = 10$
0	1056.11 (1056.11)	291.73 (291.32)	207.46 (207.45)	56.70 (56.68)
1	1053.06 (1051.64)	291.70 (291.12)	207.37 (207.30)	56.70 (56.70)
2	1054.87 (1054.83)	291.69 (287.68)	207.42 (207.41)	56.70 (56.57)
5	1055.09 (1050.15)	291.68 (289.16)	207.40 (207.40)	56.70 (56.65)
10	1054.99 (1050.20)	291.71 (288.97)	207.40 (207.39)	56.70 (56.64)

errors with respect to the optimal τ_i values, as discussed in Section 4.3. To see whether lower rounding errors can be achieved by allowing $m_i \neq 0$, we study the scenario of the experiment of Fig. 12 (with $w_2 = 2$ and 10) for different m_i values. Results for the analytical exhaustive search and our approximation (the latter in parenthesis) are given in Table 3. They show that throughput performance is approximately the same for all m_i values. The reason is that rounding errors are always very small. We conclude that this validates the choice $m_i = 0$ of this paper, as performance is (with minor deviations) as good as with any other choice.

5. Architectural considerations

The optimal configuration we propose has to be computed by a centralized entity, which we call the *Configuration Server* (CS), that needs to keep track of the number of stations in each AC and the associated weights. The CS could be located, e.g., at the Access Point (AP).

One way of obtaining this data is by forcing admission control for all AC's (this is allowed by the standard draft). Then, all the stations issue a request before joining an AC, and, in this way, the CS can keep track of the number of stations in every AC. The weight associated to an AC needs to be provided to the CS via some external signaling or static configuration.

With the above information, the CS can compute the optimal configuration for each AC, and then, announce it to the WLAN stations by means of the beacon frames (as specified by the standard draft). This message exchange is illustrated in Fig. 15. The upper limit on the allowed number of AC's (4) ensures that the signaling load resulting from announcing the configuration is kept small.

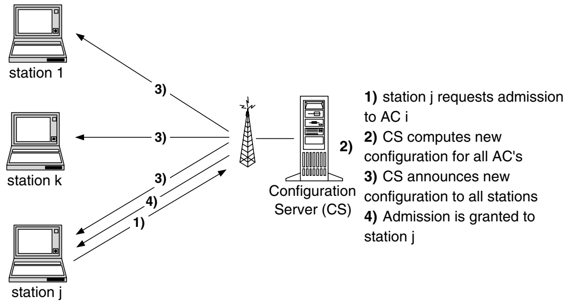


Fig. 15. Optimal configuration message exchange.

One crucial issue of the architecture described above is the computational efficiency of the operations performed at the CS. With the algorithm proposed in this paper, the computation of the optimal configuration is very efficient, as it can be performed with a few basic (addition, product, square root) operations. The computation of the resulting throughput distribution, if needed, is equally efficient, as the τ_i values are already given by the optimal configuration computation and do not need to be obtained from the non-linear system of equations of Section 3.

6. Conclusions

This paper consists of two main parts. In the first part we presented a model to analyze the throughput distribution of an 802.11e EDCA WLAN. Simulation results showed that the model closely approximates the 802.11e EDCA protocol. Moreover, they showed that the model provides particular accuracy in our region of interest, i.e., for a WLAN optimally configured.

The second part and main contribution of the paper was the search for the optimal configuration of EDCA with respect to the weighted max–min fairness criterion. This criterion allocates to each station a throughput proportional to its assigned weight. One possible application of the proposed configuration is, e.g., to offer services of different quality and (possibly) different price for data access in a WLAN hot spot.

Our optimal configuration was not validated via simulation, as finding the optimal configuration by means of simulations is unfeasible in terms of simulation time. However, we gave clear evidence that our optimal configuration provides a performance very close to the "real" optimal, as we validated it against the throughput analysis, which, in turn, was validated against simulations.

One of the main findings of this paper was that the $AIFS_i$ parameter is not used in the optimal configuration. Indeed, at a first glance it would wrongly seem that the *AIFS* mechanism provides a more efficient means for throughput allocation than the *CW* mechanism, as with *AIFS* collisions are avoided in some of the slot times. Another contribution of the paper was the computation of closed formulae for the configuration of the $CW_i^{\min}$ and $CW_i^{\max}$ parameters. These formulae were shown to accurately approximate the optimal configuration.

The $TXOP\_limit_i$ parameter was also discussed. We reasoned that throughput performance increases as the $TXOP\_limit_i$ gets larger, but counter-argued that, on account of delay considerations, this parameter cannot be set arbitrarily large. Given a maximum acceptable delay, and the configuration of the other three EDCA parameters given in this paper, the configuration of the $TXOP\_limit_i$ parameter can be set based on a delay analysis provided elsewhere [16].

The architecture described in this paper for the optimal configuration of EDCA WLANs is currently being implemented in the framework of the DAIDALOS project[13] and its architecture for QoS over heterogeneous wireless networks [30]. Initial experimental results on EDCA configurations in our implementation are presented in [31].

Acknowledgement

The work described in this paper is based on results of the IST FP6 Integrated Project DAIDALOS. It was performed during a visit of Luca Vollero at the University Carlos III of Madrid funded by the IST FP6 Network of Excellence E-NEXT. DAIDALOS and E-NEXT receive research funding from the European Community's Sixth Framework Programme. Apart from this, the European Commission has no responsibility for the content of this paper.

[13] See http://www.ist-daidalos.org/.

Appendix A

Theorem 1. *There exists one and only one solution to the non-linear system of equations of Section 3.*

Proof. Let us consider $p(e_0)$. From Eqs. (1) and (3), it can be seen that τ_i, $i \in S_0$, is a continuous and monotone increasing function of $p(e_0)$. Applying Eq. (4) and considering this relationship between τ_i and $p(e_0)$, it can be seen that $p(e_1)$ is a continuous and monotone increasing function from 0 to 1 of $p(e_0)$. Applying this recursively, it follows that τ_i $\forall i$ is a continuous and monotone increasing function of $p(e_0)$ and $p(e_k)$ $\forall k$ is a continuous and monotone increasing function from 0 to 1 of $p(e_0)$.

Applying the above to Eq. (5), we have that the left-hand side of the equation is a continuous and monotone increasing function of $p(e_0)$ that starts from 0 and grows up to 1, while the right-hand side is a continuous and monotone decreasing function of $p(e_0)$ that starts at $\prod_{k \in S_N} 1 - \tau_k(0)$ and reduces down to $\prod_{k \in S_N} 1 - \tau_k(1)$; from this, it follows that this equation has a unique solution. Taking the resulting $p(e_0)$ and undoing all the previous steps we have a solution to the system. Uniqueness of the solution is given by the fact that all relationships are bijective and any solution must satisfy Eq. (5), which (as we have shown) has only one solution. □

Theorem 2. *For any configuration* $\{\tau_i, A_i\}$ *with some* $A_i \neq 0$, *there exists an alternative configuration* $\{\tau_i', A_i'\}$ *with* $A_i' = 0$ $\forall i$ *for which* $\min(r_i'/w_i) \geqslant \min(r_i/w_i)$.

Proof. Let N be the maximum A_i value in the original configuration $\{\tau_i, A_i\}$. Note that, according to the theorem statement, we have $N \neq 0$. Let us define (following Section 3) S_N as the set of AC's with $A_i \leqslant N$, i.e., all the AC's in the WLAN. Let us further define T_N as the set of AC's with $A_i = N$, i.e., the AC's with maximum A_i in the original configuration.

We first consider the following alternative configuration to the original one, which we denote by $\{\tau_i^*, A_i^*\}$. In this alternative configuration, all the AC's that do not belong to set T_N are left with their original configuration, i.e., $A_i^* = A_i$ and $\tau_i^* = \tau_i$ for $i \notin T_N$. The AC's that belong to T_N (i.e., the ones with maximum A_i) are set with A_i^* and τ_i^* as follows. Their A_i^* is set equal to $N-1$. Note that with this, the maximum A_i^* of the alternative configuration is $N-1$ instead of N. Their τ_i^* is set equal to the only solution of the system formed by Eqs. (38) and (39):[14]

$$\prod_{k \in T_N} (1-\tau_k^*)^{n_k} = 1 - p(e_{N-1}) + p(e_{N-1}) \prod_{k \in T_N} (1-\tau_k)^{n_k}, \tag{38}$$

$$\frac{\tau_i^*(1-\tau_j^*)}{\tau_j^*(1-\tau_i^*)} = \frac{\tau_i(1-\tau_j)}{\tau_j(1-\tau_i)}, \quad \forall i,\ j \in T_N, \tag{39}$$

where $p(e_{N-1})$ is the probability that, with the original configuration, an $(N-1)$-*slot time* is empty.

We first proof that $\tau_i^* \leqslant \tau_i$ for $i \in T_N$. From Eq. (38) we have

$$1 - \prod_{i \in T_N} (1-\tau_i^*)^{n_i} = p(e_{N-1}) \left(1 - \prod_{i \in T_N} (1-\tau_i)^{n_i} \right) \leqslant 1 - \prod_{i \in T_N} (1-\tau_i)^{n_i}, \tag{40}$$

which yields

$$\prod_{i \in T_N} (1-\tau_i^*)^{n_i} \geqslant \prod_{i \in T_N} (1-\tau_i)^{n_i}. \tag{41}$$

From Eq. (39) we have

$$\tau_i^* > \tau_i \quad \text{for some } i \in T_N \Rightarrow \tau_j^* > \tau_j \forall j \in T_N. \tag{42}$$

Finally, combining Eqs. (41) and (42) leads to $\tau_i^* \leqslant \tau_i$ for $i \in T_N$.

Let $p(e)^*$ denote the probability of empty, and $p(s_i)^*$ the probability of a success of a station of AC i, with the alternative configuration $\{\tau_i^*, A_i^*\}$. Next, we show that the following relationship holds between these probabilities and the ones of the original configuration: $p(e)^* = p(e)$ and $p(s_i)^* \geqslant p(s_i)\ \forall i$.

Applying Eq. (38) to $p(e_{N-1})^*$ we have

$$p(e_{N-1})^* = \prod_{k \in S_N} 1 - \tau_k^* = (1 - p(e_{N-1})) \prod_{k \in S_{N-1}} 1 - \tau_k + p(e_{N-1}) \prod_{k \in S_N} (1-\tau_k)^{n_i}. \tag{43}$$

From Eq. (4) it can be seen that the right-hand side of the above equation is equal to $p(e_{N-1})$, which yields $p(e_{N-1})^* = p(e_{N-1})$. Applying Eq. (4) recursively, it follows $p(e_k)^* = p(e_k)\ \ \forall k \leqslant N-1$, and, therefore, $p(e)^* = p(e)$.

Combining the above result with Eq. (11) leads to $p_k^* = p_k\ \forall k \leqslant N-1$ and $p(S_k)^* = p(S_k)$ $\forall k \leqslant N{-}2$. Applying Eqs. (10) and (13) it can be further seen that $p(S_{N-1}) = p(S_{N-1})^*(1-p(e_N))$ and $p(S_N) = p(S_{N-1})^*\ p(e_N)$.

To show $p(s_i)^* \geqslant p(s_i)\ \forall i$, we proceed in two steps: $i \notin T_N$ and $i \in T_N$. For $i \notin T_N$, this is given by

$$\begin{aligned} p(s_i)^* &= \sum_{k=A_i}^{N-1} p(S_k)^* \tau_i^* (1-\tau_i^*)^{n_i-1} \prod_{j \in S_k \setminus i} (1-\tau_j^*)^{n_j} \\ &= \sum_{k=A_i}^{N-2} p(S_k) \tau_i (1-\tau_i)^{n_i-1} \prod_{j \in S_k \setminus i} (1-\tau_j)^{n_j} \\ &\quad + p(S_{N-1})^* \tau_i (1-\tau_i)^{n_i-1} \prod_{j \in S_{N-1} \setminus i} (1-\tau_j)^{n_j} \\ &\quad \times \prod_{j \in T_N} (1-\tau_j^*)^{n_j} \\ &= \sum_{k=A_i}^{N-2} p(S_k) \tau_i (1-\tau_i)^{n_i-1} \prod_{j \in S_k \setminus i} (1-\tau_j)^{n_j} \\ &\quad + p(S_{N-1})^* \tau_i (1-\tau_i)^{n_i-1} \prod_{j \in S_{N-1} \setminus i} (1-\tau_j)^{n_j} \\ &\quad \times \left(1 - p(e_{N-1}) + p(e_{N-1}) \prod_{k \in T_N} (1-\tau_k)^{n_k} \right) \end{aligned}$$

[14] As, from Eq. (39), we can express any τ_k^* as an increasing function from 0 to 1 of some reference τ_i^*, we have that $\prod_{k \in T_N} (1-\tau_k^*)^{n_k}$ is a continuous and decreasing function from 1 to 0 in the range $\tau_i^* \in (0,1)$. Uniqueness is then given by $1 > 1 - p(e_N - 1) + p(e_N - 1) \prod_{k \in T_N} (1-\tau_k)^{n_k} > 0$.

$$= \sum_{k=A_i}^{N-2} p(S_k)\tau_i(1-\tau_i)^{n_i-1} \prod_{j\in S_k\setminus i}(1-\tau_j)^{n_j} + p(S_{N-1})\tau_i(1-\tau_i)^{n_i-1} \prod_{j\in S_{N-1}\setminus i}(1-\tau_j)^{n_j} + p(S_N)\tau_i(1-\tau_i)^{n_i-1} \prod_{j\in S_N\setminus i}(1-\tau_j)^{n_j} = p(s_i). \quad (44)$$

To show $p(s_i)^* \geqslant p(s_i)$ for $i \in T_N$, we define $p(s_{T_N})$ as the probability that a slot time contains a success of some station of an AC that belongs to T_N. From Eq. (39) it follows:

$$p(s_i)^* = \frac{p(s_{T_N})^*}{p(s_{T_N})} p(s_i), \quad (45)$$

from which, if we proof $p(s_{T_N})^* \geqslant p(s_{T_N})$, we have $p(s_i)^* \geqslant p(s_i)$. $p(s_{T_N})^*$ is given by Eq. (46), and $p(s_{T_N})$ by Eq. (47), K being the same value in both equations, and $f()$ the same function but evaluated at a different point.

$$\begin{aligned} p(s_{T_N})^* &= p(S_{N-1})^* \prod_{k\in S_{N-1}}(1-\tau_k^*)^{n_k} \times \left(\sum_{k\in T_N} n_k\tau_k^*(1-\tau_k^*)^{n_k-1} \prod_{j\in T_N\setminus k}(1-\tau_j^*)^{n_j}\right) \\ &= p(S_{N-1})^* \prod_{k\in S_{N-1}}(1-\tau_k^*)^{n_k} \left(1-\prod_{k\in T_N}(1-\tau_k^*)^{n_k} - \left(1-\prod_{k\in T_N}(1-\tau_k^*)^{n_k} - \sum_{k\in T_N} n_k\tau_k^*(1-\tau_k^*)^{n_k-1} \prod_{j\in T_N\setminus k}(1-\tau_j^*)^{n_j}\right)\right) \\ &= p(S_N) \prod_{k\in S_{N-1}}(1-\tau_k)^{n_k}\left(1-\prod_{k\in S'_N}(1-\tau_k)^{n_k}\right) \times \left(1-\frac{1-\prod_{k\in T_N}(1-\tau_k^*)^{n_k} - \sum_{k\in T_N} n_k\tau_k^*(1-\tau_k^*)^{n_k-1}\prod_{j\in T_N\setminus k}(1-\tau_j^*)^{n_j}}{1-\prod_{k\in T_N}(1-\tau_k^*)^{n_k}}\right) \\ &= K\left(1-\frac{1-\prod_{k\in T_N}(1-\tau_k^*)^{n_k} - \sum_{k\in T_N} n_k\tau_k^*(1-\tau_k^*)^{n_k-1}\prod_{j\in T_N\setminus k}(1-\tau_j^*)^{n_j}}{1-\prod_{k\in T_N}(1-\tau_k^*)^{n_k}}\right) \\ &= K\left(1-f(\tau_j^*, j\in T_N)\right). \end{aligned} \quad (46)$$

$$\begin{aligned} p(s_{T_N}) &= K\left(1-\frac{1-\prod_{k\in T_N}(1-\tau_k)^{n_k} - \sum_{k\in T_N} n_k\tau_k(1-\tau_k)^{n_k-1}\prod_{j\in T_N\setminus k}(1-\tau_j)^{n_j}}{1-\prod_{k\in T_N}(1-\tau_k)^{n_k}}\right) \\ &= K(1-f(\tau_j, j\in T_N)). \end{aligned} \quad (47)$$

Taking partial derivatives, it can be seen that $f()$ is an increasing function on all its arguments:[15]

$$\begin{aligned} \frac{\partial f}{\partial \tau_j} &= \frac{n_j(1-\tau_j)^{n_j-1}\prod_{k\in T_N\setminus j} 1-\tau_k^{n_k}}{(1-\prod_{k\in T_N} 1-\tau_k)^2} \times \left(-1+(n_j-1)\frac{\tau_j}{1-\tau_j} + \sum_{k\in T_N\setminus j} n_k \frac{\tau_k}{1-\tau_k} + \prod_{k\in T_N}(1-\tau_k)^{n_k} + \tau_j(1-\tau_j)^{n_j-1}\prod_{k\in T_N\setminus j}(1-\tau_k)^{n_k}\right) \\ &> \frac{n_j(1-\tau_j)^{n_j-1}\prod_{k\in T_N\setminus j} 1-\tau_k^{n_k}}{(1-\prod_{k\in T_N} 1-\tau_k)^2} \times \left(-1+(n_j-1)\tau_j + \sum_{k\in T_N\setminus j} n_k\tau_k + \prod_{k\in T_N}(1-\tau_k)^{n_k} + \tau_j(1-\tau_j)^{n_j-1}\prod_{k\in T_N\setminus j}(1-\tau_k)^{n_k}\right) \\ &> 0. \end{aligned} \quad (48)$$

Since, as we have shown, $\tau_j^* \leqslant \tau_j \ \forall j \in T_N$, we have from Eqs. (46) and (47) that $p(s_{T_N})^* \geqslant p(s_{T_N})$ and therefore $p(s_i)^* \geqslant p(s_i)$ for $i \in T_N$.

Applying the above procedure recursively to the configuration resulting from the previous step for $N-1, N-2, \ldots, 1$ it can be seen that there exists a configuration $\{\tau_i'', A_i''\}$ such that $A_i'' = 0 \ \forall i$, $p(e)'' = p(e)$ and $p(s_i)'' \geqslant p(s_i) \ \forall i$.

We further consider the configuration $\{\tau_i', A_i'\}$ where $A_i' = 0 \ \forall i, \tau_j' = \tau_j''$ for the AC for which r_j''/r_j is minimum, and τ_i' is the only solution to the following equation for any other AC:

$$\frac{\tau_i'(1-\tau_j')}{\tau_j'(1-\tau_i')} = \frac{r_i}{r_j}, \quad (49)$$

[15] Note that in a slot time either none of the stations of the AC's of set T_N transmits, a given station of AC j transmits and no other station of the AC's of set T_N transmits, or at least one of the other stations of the AC's of set T_N transmits. The probability of the latter event is smaller than the sum of the individual transmission probabilities, which yields the inequality of Eq. (48).

where r_i is the throughput of a station of AC i with the original configuration.[16]

From the fact that r''_j/r_j is minimum for AC j:

$$\frac{r''_j}{r_j}\min \Rightarrow \frac{r''_i}{r_i} \geqslant \frac{r''_j}{r_j} \Rightarrow \frac{r''_i}{r''_j} \geqslant \frac{r_i}{r_j}\ \forall i. \quad (50)$$

The combination of Eqs. (49), (50) and $\tau'_j = \tau''_j$ leads to $\tau'_i \leqslant \tau''_i$ for $i \neq j$. This implies

$$p(e)' = \prod_{k\in S_N}(1-\tau'_k)^{n_k} \geqslant p(e)'' = p(e) \quad (51)$$

and

$$\begin{aligned} p(s)' &= \frac{\sum_{k\in S_N} n_k r'_k}{r'_j} p(s_j)' \\ &= \frac{\sum_{k\in S_N} n_k r_k}{r_j} \tau'_j(1-\tau'_j)^{n_j-1} \prod_{k\in S_N\setminus j}(1-\tau'_k)^{n_k} \\ &\geqslant \frac{\sum_{k\in S_N} n_k r_k}{r_j} p(s_j)'' \geqslant \frac{\sum_{k\in S_N} n_k r_k}{r_j} p(s_j) = p(s). \end{aligned} \quad (52)$$

Since the configuration $\{\tau'_i, A'_i\}$ leads to the same throughput ratios as $\{\tau_i, A_i\}$, and greater empty and success probabilities, it also leads to a greater throughput for all the AC's:

$$\begin{aligned} r'_i &= \frac{r_i}{\sum_{k\in S_N} n_k r_k} \cdot \frac{p(s)' l}{p(s)' T_s + p(c)' T_c + p(e)' \sigma} \\ &\geqslant \frac{r_i}{\sum_{k\in S_N} n_k r_k} \cdot \frac{p(s) l}{p(s) T_s + p(c) T_c + p(e) \sigma} = r_i\ \forall i. \end{aligned} \quad (53)$$

Finally, since $r'_i \geqslant r_i\ \forall i, \min(r'_i/w_i)$ will be greater than or equal to $\min(r_i/w_i)$, which terminates the proof of the theorem. □

Theorem 3. *For any configuration* $\{\tau_i\}$ *such that* $r_i/r_j \neq w_i/w_j$ *for some* i, j, *there exists an alternative configuration* $\{\tau'_i\}$ *that satisfies* $r'_i/r'_j = w_i/w_j\ \forall i,j$ *for which* $\min(r'_i/w_i) \geqslant \min(r_i/w_i)$.

Proof. Let us consider the configuration $\{\tau'_i\}$ where $\tau'_j = \tau_j$, j being the AC for which r_j/w_j is minimum, and τ'_i is the only solution to

$$\frac{\tau'_i(1-\tau'_j)}{\tau'_j(1-\tau'_i)} = \frac{w_i}{w_j}, \quad (54)$$

for $i \neq j$.

From the above equation it follows that $r'_i/r'_j = w_i/w_j$, which proofs the first part of the theorem.

From $\tau'_j = \tau_j, r'_i/r'_j = w_i/w_j$ and $r_i/r_j \geqslant w_i/w_j$ (given by r_i/w_i being minimum for AC j), it follows that $\tau'_i \leqslant \tau_i\ \forall i \neq j$. This yields $p(e)' \geqslant p(e)$ and $p(s_j)' \geqslant p(s_j)$, which in turn yields $r'_j \geqslant r_j$.

Since $r'_i/r'_j = w_i/w_j$ we have that $\min(r'_i/w_i) = r'_j/w_j$, and since $r'_j \geqslant r_j$, we have

$$\min\left(\frac{r'_i}{w_i}\right) = \frac{r'_j}{w_j} \geqslant \frac{r_j}{w_j} = \min\left(\frac{r_i}{w_i}\right), \quad (55)$$

which terminates the proof of the theorem. □

[16] Note that Eq. (49) implies $r'_i/r'_j = r_i/r_j$.

References

[1] IEEE 802.11, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, Standard, IEEE, August 1999.

[2] IEEE 802.11e/D13.0, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS), Draft Supplement to IEEE 802.11 Standard, January 2005.

[3] S. Shenker, Fundamental design issues for the future internet, IEEE Journal on Selected Areas in Communications 13 (7) (1995) 1176–1188.

[4] T. Korakis, L. Tassiulas, Providing Quality of Service guarantees in wireless LANs compliant with 802.11e, Computer Networks 47 (2) (2005) 239–255.

[5] A. Grilo, M. Macedo, M. Nunes, A scheduling algorithm for QoS support in IEEE 802.11e networks, IEEE Wireless Communications 10 (3) (2004) 36–43.

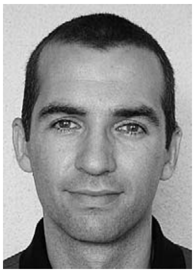
[6] P. Ansel, Q. Ni, T. Turletti, An efficient scheduling scheme for IEEE 802.11e, in: Proceedings of WiOpt'04, Cambridge, UK, March 2004.

[7] P. Garg, R. Doshi, R. Greene, M. Baker, M. Malek, X. Cheng, Using IEEE 802.11e MAC for QoS over wireless, in: Proceedings of IEEE IPCCC'03, Phoenix, AZ, April 2003.

[8] S. Mangold, S. Choi, P. May, O. Klein, G. Hiertz, L. Stibor, IEEE 802.11e wireless LAN for Quality of Service, in: Proceedings of European Wireless'02, Florence, Italy, February 2002.

[9] S. Choi, J. del Prado, S. Shankar, S. Mangold, IEEE 802.11e contention-based channel access (EDCF) performance evaluation, in: Proceedings of IEEE ICC'03, Anchorage, AK, May 2003.

[10] A. Grilo, M. Nunes, Performance evaluation of IEEE 802.11e, in: Proceedings of IEEE PIMRC'02, Lisbon, Portugal, September 2002.
[11] D. Gu, J. Zhang, QoS enhancement in IEEE 802.11 wireless local area networks, IEEE Communications Magazine 41 (6) (2003) 120–124.
[12] Y. Xiao, Enhanced DCF of IEEE 802.11e to support QoS, in: Proceedings of IEEE WCNC'03, New Orleans, LA, March 2003.
[13] G. Bianchi, I. Tinnirello, Analysis of priority mechanisms based on differentiated inter frame spacing in CSMA-CA, in: Proceedings of IEEE VTC'03 Fall, Orlando, FL, October 2003.
[14] J.W. Robinson, T.S. Randhawa, Saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function, IEEE Journal on Selected Areas in Communications 22 (5) (2004) 917–928.
[15] Z. Kong, D.H.K. Tsang, B. Bensaou, D. Gao, Performance analysis of IEEE 802.11e contention-based channel access, IEEE Journal on Selected Areas in Communications 22 (10) (2004) 2095–2106.
[16] A. Banchs, L. Vollero, A delay model for IEEE 802.11e EDCA, IEEE Communications Letters 9 (6) (2005) 508–510.
[17] G. Bianchi, Performance analysis of the IEEE 802.11 distributed coordination function, IEEE Journal on Selected Areas in Communications 18 (3) (2000) 535–547.
[18] H. Wu, Y. Peng, K. Long, S. Cheng, J. Ma, Performance of reliable transport protocol over IEEE 802.11 wireless LAN: analysis and enhancement, in: Proceedings of IEEE INFOCOM'02, New York, NY, June 2002.
[19] F. Cali, M. Conti, E. Gregori, Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit, IEEE/ACM Transactions on Networking 8 (6) (2000) 785–799.
[20] EDCF extension for *ns-2*, Available from: <http://www.tkn.tu-berlin.de/research/802.11e_ns2/>.
[21] The *ns-2* simulator, Available from: <http://www.isi.edu/nsnam/ns/>.
[22] D. Bertsekas, R.G. Gallager, Data Networks, Prentice-Hall, Englewood Cliffs, NJ, 1987.
[23] A.K. Parekh, R.G. Gallager, A generalized processor sharing approach to flow control in integrated services networks: the multiple node case, IEEE/ACM Transactions on Networking 2 (2) (1994) 137–150.
[24] L. Massoulie, J. Roberts, Bandwidth sharing: objectives and algorithms, IEEE/ACM Transactions on Networking 10 (3) (2002) 320–328.
[25] T. Nandagopal, S. Lu, V. Bharghavan, A unified architecture for the design and evaluation of wireless fair queuing algorithms, in: Proceedings of ACM MOBICOM'99, Seattle, WA, August 1999.
[26] S. Lu, V. Bharghavan, R. Srikant, Fair scheduling in wireless packet networks, in: Proceedings of ACM SIGCOMM'97, Cannes, France, August 1997.
[27] N.H. Vaidya, P. Bahl, S. Gupta, Distributed fair scheduling in wireless LAN, in: Proceedings of ACM MOBICOM'00, Boston, MA, August 2000.
[28] A. Banchs, X. Pérez-Costa, Distributed weighted fair queuing in 802.11 wireless LAN, in: Proceedings of IEEE ICC'02, New York, NY, April 2002.
[29] A. Banchs, X. Pérez-Costa, D. Qiao, Providing throughput guarantees in IEEE 802.11e wireless LANs, in: Proceedings of the 18th International Teletraffic Congress (ITC18), Berlin, Germany, September 2003.
[30] G. Carneiro, C. García, P. Neves, Z. Chen, M. Wetterwald, M. Ricardo, P. Serrano, S. Sargento, A. Banchs, The DAIDALOS architecture for QoS over heterogeneous wireless networks, in: Proceedings of IST Mobile & Wireless Communications Summit, Dresden, Germany, June 2005.
[31] A. Banchs, A. Azcorra, C. García, R. Cuevas, Applications and challenges of the 802.11e EDCA mechanism: an experimental study, IEEE Network Special Issue on Wireless Local Area Networking: QoS Provisioning & Resource Management 19 (4) (2005).



Albert Banchs received his M.Sc. and Ph.D. degrees in telecommunications from the Technical University of Catalonia in 1997 and 2002, respectively. His Ph.D. received the national award for best thesis on broadband networks granted by the Professional Association of Telecommunication Engineers. He worked for the International Computer Science Institute, Berkeley, in 1997, for Telefonica I+D, Madrid, in 1998, and for NEC Network Laboratories, Heidelberg, from 1998 to 2003. Since 2003 he is with the University Carlos III of Madrid. His current research interests include resource allocation, QoS, and performance evaluation of wireless and wired networks.

Luca Vollero received a M.Sc. degree in telecommunications from the University of Naples, Federico II in 2001, and a Ph.D. degree in computer science from the same university in 2005. Currently he is a researcher at C.I.N.I., the Italian University Consortium for Computer Science. His primary research interests include wireless networks, multimedia data processing and performance analysis of embedded systems. He is a member of IEEE, IEEE Computer Society and IEEE Communications Society.