A Delay Model for IEEE 802.11e EDCA

Albert Banchs, *Member, IEEE,* and Luca Vollero, *Member, IEEE*

***Abstract*— In this letter we present a model to analyze the delay behavior of the EDCA mechanism of IEEE 802.11e. Simulation results validate the accuracy of our analysis.**

***Index Terms*— WLAN, 802.11e, EDCA, delay analysis.**

I. Introduction

In this letter we present a model for the backoff delay under saturation conditions of the *Enhanced Distributed Channel Access* (EDCA) mechanism of the upcoming 802.11e standard [1]. By backoff 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. With saturation conditions we mean that all the stations in the WLAN always have packets to transmit. Note that this corresponds to the worst case and thus provides us with an upper bound of the delay.

With EDCA, transmissions are regulated by the following backoff algorithm. Upon starting the backoff process, a station i initializes its backoff time counter to a random value uniformly distributed in the range $(0, CW_i - 1)$, with CW_i initially set to CW_i^{min}. 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 a period of time equal to $AIFS_i$. The value of σ is a constant defined by the standard, and $AIFS_i$ takes a value of the form $DIFS + n\sigma$, where $DIFS$ is another constant defined by the standard and n is a nonnegative integer.

The station transmits when the backoff time counter reaches zero. A collision occurs when two or more stations start transmission simultaneously. After each unsuccessful transmission, CW_i is doubled, up to a maximum value CW_i^{max}, and the backoff process is restarted. If the number of failed retries reaches a predetermined retry limit R, the packet is discarded.

From the above explanation, it can be seen that the behavior of a station depends on a number of configurable parameters ($AIFS_i$, CW_i^{min} and CW_i^{max}) that can be set to different values for different Access Categories (AC's). The rest of this letter is devoted to the study of the delay as a function of the number of stations and configuration of each AC.

Manuscript received October 21, 2004. The associate editor coordinating the review of this letter and approving it for publication was Prof. C.-F. Chiasserini. This work was funded in part by the European Community's Sixth Framework Programme with the IST FP6 Integrated Project DAIDALOS. The work was performed during a visit of L. Vollero funded by the IST FP6 Network of Excellence E-NEXT. Apart from this, the European Commission has no responsibility for the content of this paper.

A. Banchs is with the Universidad Carlos III de Madrid, 28911 Madrid, Spain (email: banchs@it.uc3m.es).

L. Vollero is with the Università degli Studi di Napoli "Federico II", 80125 Naples, Italy (email: vollero@unina.it).

Digital Object Identifier 10.1109/LCOMM.2005.06015.

II. Delay Analysis

Consider a WLAN with a fixed number of EDCA stations, each belonging to some AC i, $i \in \{1, \ldots, n\}$. Let n_i be the number of stations of AC i and $\{CW_i^{min}, CW_i^{max}, AIFS_i\}$ the corresponding EDCA parameters. 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 $A_i \leq k$ and N as the largest A_i in the WLAN.

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 counter decrement of the station. With the assumption of [2] 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 [2]. Then, the probability that a station of AC i transmits upon a backoff counter decrement can be computed as [2],

$$\tau_i = \frac{2(1-2p_i)(1-p_i^{R+1})}{CW_i^{min}(1-(2p_i)^{m_i+1})(1-p_i)+(1-2p_i)(1-p_i^{R+1})+ +CW_i^{min}2^{m_i}p_i^{m_i+1}(1-2p_i)(1-p_i^{R-m_i})} \tag{1}$$

where p_i is the probability that a transmission attempt of a station of AC i collides.

Let us define a *slot time* as the time interval between two consecutive backoff 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 k*-slot time* is empty.

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

To compute p_i, we use the assumption of [3] 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 assumption, the probability p_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 the considered station does not transmit multiplied by the probability that no other station transmits,

$$p(e_k) = (1-\tau_i)(1-p_i) \tag{2}$$

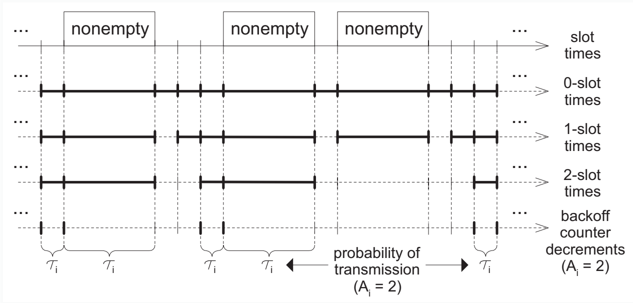

Fig. 1. *k-slot times* and probability of transmission (example with $k = 2$).

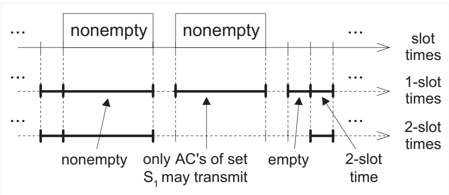

Fig. 2. Probability of an empty *k-slot time* (example with $k = 1$).

which yields

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

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 $A_i \leq k$ (i.e. the AC's that belong to the set S_k according to our previous definition) may transmit, and, with the above assumption, 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 a *N-slot time* the stations of all AC's may transmit, the following equation holds

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

Starting from $\tau_i \ \forall i$, with (5) we can compute $p(e_N)$. Then, with (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 (3) we can compute p_i from which we can compute τ_i with (1). We conclude that with the above equations we can express each τ_i as a function of $\tau_i \ \forall i$, 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.

Once the values $\tau_i \ \forall i$ have been derived, the average backoff delay experienced by a non-dropped packet of a station of AC i, d_i, is computed as

$$d_i = \frac{1}{P_{tx,i}} \sum_{j=0}^{R} P_{tx,i,j} \, d_{i,j} \quad (6)$$

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,

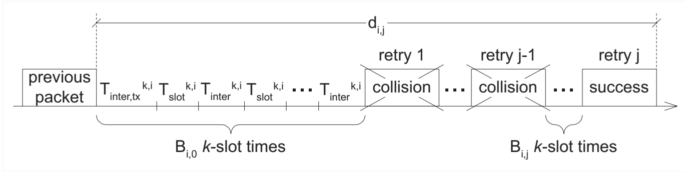

Fig. 3. Average delay in case of j retries.

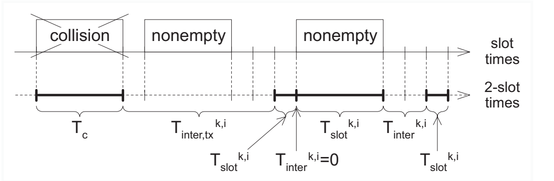

Fig. 4. Components of the delay (example with $k = 2$).

$$P_{tx,i} = \sum_{j=0}^{R} (1 - p_i) p_i^j \ , \quad P_{tx,i,j} = (1 - p_i) p_i^j \quad (7)$$

and $d_{i,j}$ is the average delay in case of j retries (see Fig. 3),

$$d_{i,j} = \sum_{l=0}^{j} \left(T_{inter,tx}^{k,i} + B_{i,l} (T_{slot}^{k,i} + T_{inter}^{k,i}) \right) + j T_c + T_s \quad (8)$$

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. T_s and T_c are the average duration of a slot time that contains a success and a collision. Fig. 4 illustrates these delay components under a given sequence of slot times following a collision of the considered station.

Expressions to compute T_c and T_s, both for RTS/CTS and no RTS/CTS, are given in [2]. $B_{i,l}$ is computed as

$$B_{i,l} = \frac{CW_i^{min} 2^{min(m_i, l)} - 1}{2} \quad (9)$$

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

$$T_{slot}^{k,i} = p(s_{k,i}) T_s + p(e_{k,i}) \sigma + (1 - p(s_{k,i}) - p(e_{k,i})) T_c \quad (10)$$

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 transmit[1] is empty and contains a success, respectively. We compute $p(e_{k,i})$ by applying a similar reasoning to (2),

$$p(e_{k,i}) = \frac{p(e_k)}{1 - \tau_i} \quad (11)$$

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*,

$$p(s_{k,i}) = \sum_{j=0}^{N-1-k} p(e_{k,i,j}) p(s_{k,i,j}) + p(e_{k,i,N-k})' p(s_{k,i,N-k})' \quad (12)$$

[1]The condition that the considered station does not transmit holds in the computation of all probabilities until (16).

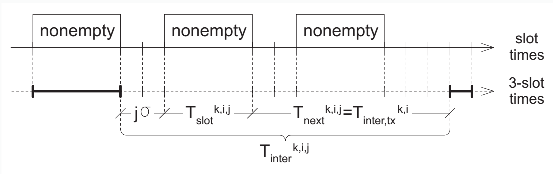

Fig. 5. Components of $T_{inter}^{k,i,j}$ (example with $k=3$ and $j=2$).

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 $p(s_{k,i,N})'$ 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 $S_k, S_{k+1}, \ldots, S_{j+k-1}$ may transmit,

$$p(e_{k,i,j}) = (1-p(e_{k,i})) \prod_{l=k}^{k+j-1} \prod_{m \in S_l} (1-\tau_m)^{n_{m,i}} \tag{13}$$

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 S_{k+j} may transmit,

$$p(s_{k,i,j}) = \sum_{l \in S_{k+j}} n_{l,i}\tau_l(1-\tau_l)^{n_{l,i}-1} \prod_{m \in S_{k+j}\setminus l} (1-\tau_m)^{n_{m,i}} \tag{14}$$

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

$$p(e_{k,i,N-k})' = 1 - \sum_{j=0}^{N-k-1} p(e_{k,i,j}) \tag{15}$$

$$p(s_{k,i,N-k})' = \sum_{l \in S_N} n_{l,i}\tau_l(1-\tau_l)^{n_{l,i}-1} \prod_{m \in S_N\setminus l} (1-\tau_m)^{n_{m,i}} \tag{16}$$

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

$$T_{inter}^{k,i} = (1-p(e_{k,i})) \sum_{j=0}^{k} p(e_{k,i,j})^* T_{inter}^{k,i,j} \tag{17}$$

$$T_{inter,tx}^{k,i} = \sum_{j=0}^{k} p(e_{k,i,j})^* T_{inter}^{k,i,j} \tag{18}$$

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 $j=k$,

$$p(e_{k,i,k})^* = \prod_{l=0}^{k-1} \prod_{m \in S_l} (1-\tau_m)^{n_m} \ , \ T_{inter}^{k,i,k} = k\sigma \tag{19}$$

and for $j<k$ (see Fig. 5),

$$p(e_{k,i,j})^* = \left(1 - \prod_{m \in S_j} (1-\tau_m)^{n_m}\right) \prod_{l=0}^{j-1} \prod_{m \in S_l} (1-\tau_m)^{n_m} \tag{20}$$

$$T_{inter}^{k,i,j} = j\sigma + T_{slot}^{k,i,j} + T_{next}^{k,i,j} \tag{21}$$

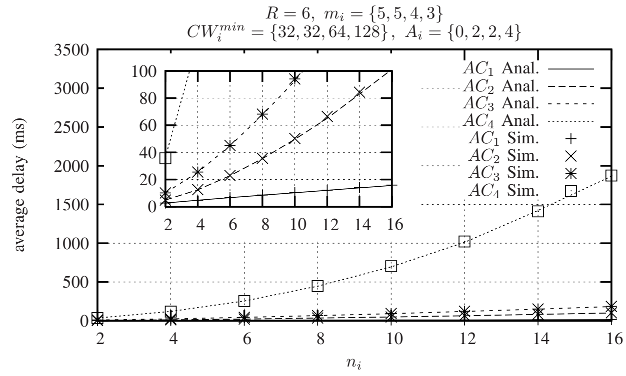

Fig. 6. Saturation backoff delay: simulation versus analysis.

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*,

$$T_{slot}^{k,i,j} = \left(1 - \frac{\sum_{m \in S_j} n_m\tau_m(1-\tau_m)^{n_m-1} \prod_{p \in S_p\setminus m}(1-\tau_p)^{n_p}}{1-\prod_{m \in S_j}(1-\tau_m)^{n_m}}\right) \cdot T_c + \frac{\sum_{m \in S_j} n_m\tau_m(1-\tau_m)^{n_m-1} \prod_{p \in S_p\setminus m}(1-\tau_p)^{n_p}}{1-\prod_{m \in S_j}(1-\tau_m)^{n_m}} T_s \tag{22}$$

$$T_{next}^{k,i,j} = T_{inter,tx}^{k,i} \tag{23}$$

Equations from (18) to (23) 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 analysis.

III. Simulation Results

We validated the accuracy of our analysis by comparing analytical results against simulations.

Fig. 6 gives the average delays, obtained analytically (lines) and via simulation (points), corresponding to 4 AC's. The subplot in the figure gives a zoom of the y-axis for a better observation of the low delays. Experiments are performed for a varying number of stations (n_i in the x axis) and $\{m_i, CW_i^{min}, A_i\}$ configurations (given on top of the figure). For all tests, a fixed packet length of 1500 bytes, the system parameters of the IEEE 802.11a physical layer and the no RTS/CTS option have been used.

We can observe that, for all experiments, analytical results coincide almost exactly with simulations, the error in all cases being well below 1%. We also assessed that the times required to compute the analytical results were very short (a few tenths of ms with a Pentium 4 PC of 2.66 GHz CPU speed). We conclude that the proposed analysis is accurate and computationally efficient.

References

[1] *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS)*, IEEE Draft Supplement to IEEE Std 802.11, Rev. 802.11e/D9.0, July 2004.

[2] H. Wu *et al.*, "Performance of reliable transport protocol over IEEE 802.11 wireless LAN: Analysis and enhancement," in *Proc. IEEE INFOCOM'02*, June 2002, pp. 599–607.

[3] F. Cali, M. Conti, and E. Gregori, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Trans. Networking*, vol. 8, pp. 785–799, Dec. 2000.