A Statistical Hybrid Traffic Control for IEEE 802.11e WLANs

Paolo Giacomazzi, Luigi Musumeci, Gabriella Saddemi, Giacomo Verticale
Politecnico di Milano, Dept. of Electronics and Information, Piazza Leonardo Da Vinci 32, 20133 Milano, ITALY
Tel.: +39-02-2399.3657 – Fax: +39-02-2399.3413 – E-mail: {giacomaz, musumeci, saddemi, vertical}@elet.polimi.it

Abstract — **The IEEE 802.11e WLAN supports prioritized Quality of Services (QoSs), but it needs additional mechanisms to support also the strict QoS required by real-time services. It has been shown that best performance is achieved when the WLAN is unsaturated. Therefore, a call admission control scheme is required to maintain the WLAN working in the unsaturated case. In this paper, we propose an analytical model for the capacity assignment necessary to satisfy the delay requirements for a given number of real-time flows belonging to two different traffic classes. At the same time, our analysis includes a rate control algorithm, as proposed in literature, that allows the best-effort traffic to fully use the residual bandwidth left by the real-time traffic, thereby achieving high channel utilization.**

I. Introduction

In a recent work [1], Chen et al. have shown that the IEEE 802.11 Distributed Coordination Function (DCF) can achieve maximum throughput and short delay only in the unsaturated case because of the low collision probability. In the saturated case, it suffers from a large collision probability, leading to low throughput and long delay. Effective tuning of the network in the unsaturated case is not easy to achieve, given that the 802.11 Enhanced Distributed Channel Access (EDCA) [2] [3] is in nature contention-based and distributed. To meet this goal, a call admission and rate control scheme is proposed and discussed in [1] [4] [5]: specifically, call admission control (CAC) is used for real-time and streaming traffic, and rate control (RC) for best-effort data traffic. The so called CARC (Call Admission and Rate Control) scheme utilizes a new measure of the network status, the channel business ratio, r_b, to exercise traffic regulation; r_b is the ratio of the time that the channel is busy to the total time of observation. Both successful transmissions and collisions contribute to r_b. The channel utilization, *cu*, which is the ratio of successful transmission periods to the total time, is almost the same as r_b, when the collision probability is very small [1].

If r_b is less then a proper threshold value, as the input traffic increases, the throughput keeps increasing linearly with r_b and the delay and delay variation do not change much and are small enough to support the real-time traffic. In this case, very few collisions are registered and $r_b \sim cu$. When r_b grows beyond a given threshold, empirically found at about 0.95, many collisions are registered, the throughput drops quickly, and the delay and delay variation increase drastically.

In the CARC scheme, call admission control checks different constraints to accommodate a new real-time flow. The adopted tests are based on the aggregate average data rate, aggregate peak rate, and delay bounds required by the specific applications.

In [1] and [4], simulation results have confirmed good CARC performance in terms of throughput, average delay, and delay variation. Nevertheless, this approach presents two drawbacks. First, if the CAC algorithm takes into account both the peak and the mean data rate for the real-time traffic, it is likely to reject many real-time flows in order to guarantee the negotiated QoS. Secondly, models to estimate the average delay are neither simple nor practical. To overcome these problems, in [5] it is proposed to remove the delay test from the admission algorithm and to make admission decisions based only on the mean rate. This simple approach cannot provide quantitative delay guarantees.

In this paper, we propose an alternative CAC algorithm, based on a two-moment description of traffic [6] [7], for real-time traffic over the IEEE 802.11e EDCA [8]. This scheme makes it possible to overcome the drawbacks concerning with the test based on the mean rate, the peak rate, and the delay estimate. We assume that a token bucket (TB) regulator guarantees each traffic flow to conform to a given QoS specification. In this case, a two moment description of traffic can be easily obtained and it can be used to calculate the link capacity necessary to satisfy the delay bound for a given number of traffic flows. As in the original proposal, the best-effort traffic is controlled according to the rate control principles proposed in CARC [4] [5] to keep low the collision probability.

This paper is organized as follows. Section II briefly introduces the EDCA mechanism and the CARC scheme. Section III presents our proposed admission control scheme with a brief description of the two-moment statistical traffic envelopes. Section IV discusses the analytical and simulation results. Finally, Section V concludes this paper.

II. The IEEE 802.11e EDCA and the CARC scheme

The IEEE 802.11e EDCA has been standardized in order to support prioritized services in the IEEE 802.11 DCF, which only provides best-effort services in its current form. In EDCA, there are four access categories (ACs) to implement prioritized services. Each AC transmits packets with an independent channel access function, characterized by different values of the contention window and the backoff timer. Specifically, for AC[*i*] (*i* = 0, 1, 2, 3), the initial backoff window size is CW_{min} [i], the maximum backoff window size is CW_{max} [i], and the Arbitration Inter-Frame Space is AIFS[i]. For $0 \le i < j \le 3$, $CW_{min}[i] \ge CW_{min}[j]$, $CW_{max}[i] \ge CW_{max}[j]$, and $AIFS[i] \ge AIFS[j]$. Thus, we see that an AC with a higher priority, has a higher probability to gain channel access. When an application is admitted, it will be assigned to an AC.

To keep the network operating in the unsaturated case CARC scheme combines a call admission control (CAC) for

WIOPT 2008, 1st – 3rd Apr 2008, Berlin, Germany.

DOI 10.4108/ICST.WIOPT2008.3001

real-time traffic with a rate control (RC) for best-effort traffic [5]. In the following, we briefly present how the CARC scheme enhances the 802.11e EDCA in supporting QoS for real-time traffic. This paper presents the results referring to the system parameters in Table I. The idea is easily applicable to other versions of IEEE 802.11.

TABLE I. IEEE 802.11 SYSTEM PARAMETERS.

Bit rate for DATA packets	2 Mbit/s
Bit rate for RTS/CTS/ACK	1 Mbit/s
PLC Data rate	1 Mbit/s
Backoff Slot Time	20 μs
SIFS	10 μs
DIFS	50 μs
Phy header	192 bits
MAC header	224 bits
IP header	160 bits
DATA packet	8000 bits + Phy header + MAC header
RTS	160 bits + Phy header
CTS, ACK	112 bits + Phy header

A. Call Admission Control

As specified in the IEEE 802.11e standard, the call admission control is performed at the QoS Access Point (QAP), when the infrastructure mode is used. Even if this discussion can be extended to the ad hoc mode, only the infrastructure mode is considered in the following.

In the CAC scheme, three parameters, (R_{mean}, R_{peak}, L), are used to characterize a real-time flow, where R_{mean} is the average data rate and R_{peak} the peak data rate in bit/s, and L is the average packet length in bits. For CBR traffic, $R_{mean} = R_{peak}$. For VBR traffic, $R_{mean} < R_{peak}$. When RTS/CTS mechanism is used, the time associated with a successful transmission, denoted by T_{suc}, is obtained by:

$$T_{suc} = T_{RTS} + T_{CTS} + T_{data} + T_{ACK} + 3SIFS + AIFS \tag{1}$$

where T_{data} is the average packet transmission time for the packet of length L (included overheads), T_{RTS}, T_{CTS}, and T_{ACK} are calculated with the values given in Table I. The SIFS value is also given in Table I. The value for AIFS depends on the AC of the traffic flow and are given in Section IV.

Therefore, the channel utilization, cu, corresponding to the bandwidth requirement of a flow can be calculated as

$$cu = U(R) = (R/L)T_{suc} \tag{2}$$

where $U(R)$ [5] is the mapping function from the traffic rate, R, to the channel utilization. Thus, the bandwidth requirements of a flow can be translated into (cu_{mean}, cu_{peak}), where cu_{mean}= $U(R_{mean})$ and cu_{peak}= $U(R_{peak})$.

Before accepting a real-time flow of priority i, the QAP associates the flow with the appropriate AC_i and obtains $cu_{i,mean}$ and $cu_{i,peak}$ according to Equation (2).

In CARC, the CAC algorithm takes into account the peak rate, the mean rate, and the average delay requested by the application. If the ratio R_{peak}/R_{mean} is large for some applications, the admission test based on the peak rate parameter becomes very conservative with the rejection of many real-time flows. This drawback can be overcome by using only the mean rate in the admission control algorithm [4]. In addition, recognizing the difficulty to derive analytically the delay upper bound for a new flow, the delay test is removed from the admission control scheme, taking into account that, as long as the network is kept in the unsaturated case and the best-effort traffic is well controlled to isolate the effect on the real-time traffic, the delay for the real-time traffic should be small enough to meet the QoS requirements. Therefore, a simplified CAC algorithm is also supported in [4], based only on the mean rate parameter. In this case, a new real-time flow can be admitted only if the following test is satisfied:

$$cu_{A,mean} + cu_{i,mean} < CU_{rt} \tag{3}$$

where the parameter $cu_{A,\,mean}$ is the aggregate $cu_{i,\,mean}$, and CU_{rt} is the quota of the channel utilization that is available for the real-time traffic.

In the following, CU_{rt} is equal to 80% of the maximum channel utilization, CU_{max}. This number, chosen in [4], could be changed depending on the traffic composition in real networks. Therefore, the best-effort traffic is at least entitled to 20% of the channel utilization. This amount of channel utilization dedicated to the best-effort traffic can be also used to accommodate sizable fluctuations caused by the VBR real-time traffic.

It is clear that choosing an appropriate maximum channel utilization, CU_{max}, is critical in making both the CAC and RC work. In [1], it is shown that the maximum throughput and short delay can be obtained with CU_{max} in the range of 0.9 to 0.95, where the network works in the unsaturated case and the collision level is negligible. In the following, these parameters are set as in Table II.

TABLE II. CHANNEL UTILIZATION VALUES.

CU_{max}	0.93
$CU_{rt} = 0.80\ CU_{max}$	0.744

B. Rate Control

The transmission rate of the best-effort traffic is controlled based on two criteria. First, the best effort traffic should not affect the QoS level of the admitted real-time traffic. Second, the best-effort traffic should be able to promptly access the residual bandwidth left by the real-time traffic in order to efficiently utilize the channel. Clearly, to meet these criteria, each node needs to accurately estimate the total instantaneous rate of ongoing real-time traffic.

In the distributed rate control scheme proposed in [4], each node monitors the channel busyness ratio, r_b, during a period of T_{rb}. This approach is consistent with the original 802.11e protocol. Let us denote by r_{br} the contribution from the real-time traffic to r_b, and denote by R_{be} the data rate of the best-effort traffic at the node under consideration, with the initial value of R_{be} being conservatively set, say one packet per second. The node adjusts R_{be} after each T_{rb} according to the following:

$$R_{be,\,new} = R_{be,\,old} \times \frac{CU_{\max} - r_{br}}{r_b - r_{br}} \tag{4}$$

where $R_{be,\ new}$, and $R_{be,\ old}$ are the value of R_{be} after and before the adjustment. The node increases the rate of the best-effort traffic if $r_b < CU_{max}$ and decreases the rate otherwise.

If all the nodes adjust the rate of its own best-effort traffic according to Equation (4), r_b- r_{br} is the contribution from the total best-effort traffic to r_b. Thus, after one control interval T_{rb}, the channel utilization will be approximately CU_{max} [4].

III. Proposed Admission Control Scheme

We adopt the two-moment traffic and service description proposed in [6] [7] to implement a call admission test for a new real-time flow based on the delay bound violation probability, defined as the probability, p_i, that the delay of the flow of class i exceeds a given delay threshold, d_i. This approach is appropriate to handle real-time flows also when strict delay guarantees are requested. In order to match the QoS expectation from each flow, a traffic regulator is needed so that each traffic flow can be deterministically bounded by a function, called rate envelope. We propose to use a token bucket (TB) as a regulator, characterized by the r and b parameters, where r is the average token rate (bit/s) and b is the bucket size (bit). In addition, we make reference to the theory of the EDF scheduler to develop the statistical analysis applied to the statistical service envelope [6].

The EDF scheduling algorithm requires that incoming packets are marked with timestamp and then associated with a deadline depending on their traffic class. The EDF scheduler serves the packets in order of increasing deadline and is effective in guaranteeing QoS. In general, this mechanism is not easy to implement in a distributed system as a LAN environment. However, in [9] it is shown that the distribution of the channel access time in unsaturated EDCA can be approximated with the waiting time of the EDF scheduler if, for each couple (AC[i], AC[j]), with AC[i] having higher priority, the following conditions hold:

$$\frac{CW_{\min}[i]}{CW_{\min}[j]} = \frac{d_i}{d_j}$$

where d_i, d_j are the deadlines for the EDF scheduler.

According to the statistical approach [6] [7][10], the cumulative traffic $X(t)$ generated by a flow in a time interval of duration t, is described by its statistical traffic envelope, $B(t)$, which is defined as a random process such that the traffic generated by the flow in a time interval of duration t satisfies the condition:

$$\forall z, t: P\{X(t) > z\} \le P\{B(t) > z\}.$$

For a generic traffic flow, a suitable Gaussian statistical traffic envelope can be easily defined on the basis of its deterministic traffic envelope, $b(t)$ [6] [10] [11]:

$$E[B(t)] = rt$$
$$Var[B(t)] = rb(t)t - (rt)^2,$$

where

$$r = \lim_{t\to\infty} \frac{b(t)}{t} \cdot$$

A traffic flow, regulated by a token bucket with token rate r [bit/s] and burst size b [bit], admits the deterministic traffic envelope $b(t)=b+rt$. In turn, its Gaussian statistical traffic envelope is defined by:

$$E[B(t)] = rt$$
$$Var[B(t)] = rbt.$$

The *statistical service envelope* is defined as a probabilistic description of the service that can be offered to a traffic flow. The statistical use of the capacity assigned to each traffic flow enables the exploitation of statistical multiplexing. In this way, it is possible to overbook transmission resources in a controlled manner. A commonly used statistical performance target is the maximum probability, p, of exceeding the delay bound, d:

$$\Pr\{D > d\} \le p$$

where D is the actual delay experienced by traffic in the scheduler. The probability of violating a given delay bound, d, has the following statistical upper bound [10]:

$$\Pr\{D > d\} \le \Pr\left\{\max_{t\ge 0}\{B(t) - S(t+d)\} > 0\right\}, \tag{5}$$

where $S(t)$ is the statistical service envelope associated to the reference traffic flow. In general, the calculation of (5) is quite complex. The Maximum–Variance Approximation (MVA) is commonly adopted, under the assumption that both *B(t)* and *S(t)* are Gaussian [10]. Therefore, let:

$$\sigma^2(t) = Var[B(t) - S(t+d)], \tag{6}$$

$$\alpha(t) = \frac{0 - E[B(t) - S(t+d)]}{\sigma(t)}, \tag{7}$$

the MVA approximation provides:

$$\Pr\{D > d\} \le \Pr\left\{\max_{t\ge 0}\{B(t) - S(t+d)\} > 0\right\} \le e^{\frac{-\alpha_{\min}^2}{2}} \tag{8}$$

with

$$\alpha_{\min} = \min_{t\ge 0} \alpha(t).$$

Then, according to the MVA approximation, the statistical QoS target on the delay bound violation given by $\Pr\{D > d\} \le p$ can be satisfied if:

$$\exp\left(-\frac{\alpha_{\min}^2}{2}\right) \le p \cdot \tag{9}$$

By using the statistical approach, the resource allocation and admission control problems are significantly simplified.

In the following, we consider two traffic classes, namely AC[3] and AC[2], and extend the results of the statistical approach combined with an EDF scheduler. We suppose that N_3 traffic sources are offered over AC[3] and have the same token bucket parameters (r_3, b_3), while the N_2 sources over AC[2] have parameters (r_2, b_2). The QoS requirements expressed in terms of delay are d_3 and d_2, and delay violation probability is p, which must be the same for all the classes.

The Gaussian statistical traffic envelope of the traffic aggregate offered to AC[i] has the expected value:

$$E\{B_i(t)\} = N_i r_i t \ , \qquad (10)$$

and variance equal to:

$$VAR\{B_i(t)\} = N_i r_i b_i t \qquad (11)$$

On the other hand, the net capacity available, C_{av}, is a fraction of the link capacity, C, because of collisions and MAC protocol overhead.

The assumption of unsaturated case makes it possible to avoid calculating the impact of collisions by simply assuming that the channel utilization is below or equal to CU_{max}. The MAC protocol overhead is taken into account by weighting the channel capacity with the ratio

$$\eta = \frac{N_3 \frac{r_3}{L_3}}{N_3 \frac{r_3}{L_3} + N_2 \frac{r_2}{L_2}} \frac{T_{data,3}}{T_{suc,3}} + \frac{N_2 \frac{r_2}{L_2}}{N_3 \frac{r_3}{L_3} + N_2 \frac{r_2}{L_2}} \frac{T_{data,2}}{T_{suc,2}}$$

where $T_{suc,i}$, $T_{data,i}$ and L_i are the average time for successful delivery of a packet of AC[i], calculated as in Equation (1), the transmission time of a packet of AC[i], and the length of a packet of AC[i], respectively.

Therefore, C_{av} can be expressed as

$$C_{av} = 0.8 \cdot CU_{\max} \cdot \eta \cdot C$$

where the coefficient 0.8 is used to reserve 20% of the bandwidth to best-effort traffic.

Using the service curves for the EDF scheduler in [12], the average value of the statistical service envelopes available for service classes 3 and 2 can be calculated as a function of the available capacity:

$$\begin{aligned} \mathrm{E}\left\{S_3\left(t+d_3\right)\right\} &= \max\left\{0, C_{av}\left(t+d_3\right) - N_2 r_2\left(t+d_3-d_2\right)\right\} \\ \mathrm{E}\left\{S_2\left(t+d_2\right)\right\} &= \max\left\{0, C_{av}\left(t+d_2\right) - N_3 r_3\left(t+d_2-d_3\right)\right\} \end{aligned} \qquad (12)$$

with variances equal to:

$$\begin{aligned} \mathrm{VAR}\left\{S_3\left(t+d_3\right)\right\} &= N_2 r_2 b_2\left(t+d_3-d_2\right) \\ \mathrm{VAR}\left\{S_2\left(t+d_2\right)\right\} &= N_3 r_3 b_3\left(t+d_2-d_3\right) \end{aligned} \qquad (13)$$

The max operator in Equation (12) can be eliminated if one makes the conservative assumption that a service could be negative. By substituting (10), (11), (12), and (13) in (7), we obtain:

$$\begin{aligned} \alpha_3(t) &= \frac{t(C_{av} - A) + (C_{av} d_3 - E_3)}{\sqrt{tB + E_3 b_2}} \\ \alpha_2(t) &= \frac{t(C_{av} - A) + (C_{av} d_2 - E_2)}{\sqrt{tB + E_2 b_3}} \end{aligned}$$

where the symbolic parameters are defined as follows:

$$\begin{aligned} &A = N_3 r_3 + N_2 r_2 \\ &B = N_3 r_3 b_3 + N_2 r_2 b_2 \\ &D = N_3 r_3 b_3 d_3 + N_2 r_2 b_2 d_2 \\ &E_3 = N_2 r_2\left(d_3 - d_2\right) \\ &E_2 = N_3 r_3\left(d_2 - d_3\right) \end{aligned} .$$

The absolute minima of $\alpha_3(t)$ and $\alpha_2(t)$ are:

$$\alpha_{\min,3} = \alpha_{\min,2} = \frac{2}{B}\sqrt{\left(C_{av} - A\right)\left[C_{av} D + N_3 r_3 E_3\left(b_2 - b_3\right)\right]}. \qquad (14)$$

Finally, from Equation (8), we obtain the bound on the distribution of the access delay for AC[i]:

$$\Pr\{D_i > d_i\} \le \exp\left(-\frac{1}{2}\alpha_{\min,i}^2\right) \qquad (15)$$

Then, the Call Admission Control procedure is implemented by calculating Equation (15) for every new traffic flow: if the result is lower than the delay violation probability, p, the flow is accepted. Otherwise it is rejected.

IV. Performance Evaluation

To evaluate the performance in terms of throughput and delay, we have implemented the proposed CARC scheme in the ns-2 simulator. Our simulation scenario has been derived from [4] in order to have a good reference for the discussion of our simulation results.

In particular, we consider an 802.11e based WLAN with N mobile nodes. All nodes are within the transmission range of one another. In all simulations, channel rate is 2 Mbit/s and the RTS/CTS mechanism is used. The simulation time is 250 s. The other IEEE 802.11e system parameters are in Table I.

Traffic is based on three different classes:

Voice Traffic (*VBR*): The VBR voice traffic is modeled as an *on*/*off* source with exponentially distributed *on* and *off* periods, T_{on} and T_{off}, of 300 ms average each. Traffic is generated during the *on* periods at a peak rate, R_p of 32 kbit/s with a packet size, L, of 160 bytes, thus the inter-packet time is 40 ms. An approximate set of Token Bucket parameters for this type of voice source has been obtained by applying the procedure outlined in [13], according to the following Equations:

$$r = R_p \frac{T_{on}}{T_{on} + T_{off}} = 16\ kbit/s$$

$$b = 2T_{on}\left(\frac{T_{off}}{T_{on} + T_{off}}\right)^2 R_p + L = 760\ bytes$$

By considering also the Phy, MAC, and IP overheads, shown in Table I, the real Token Bucket parameters become r_{voice} = 23.2 kbit/s and b_{voice} = 1102 bytes. This class is served with the highest priority level (class 3).

Video Traffic (CBR): The video traffic is modeled as 64 kb/s CBR traffic with a packet size of 1000 bytes, thus the inter-packet time is 125 ms. For a CBR source the Token Bucket parameters (r, b) are simply the data rate, 64 kb/s, and the packet size, 1000 bytes, respectively. By including the Phy, MAC, and IP overheads shown in Table I, the token bucket parameters become r_{video} = 68.6 kbit/s and b_{video} = 1072 bytes. The video traffic is served with a lower priority level (class 2).

Data Traffic Model: We use the greedy best-effort TCP Reno traffic as the background data traffic with a packet size of 1000 bytes. The access category is AC[0].

TABLE III. QoS PARAMETERS FOR THE VOICE AND VIDEO TRAFFIC.

Access Category	AIFS[*i*]	CW$_{min}$[*i*]	CW$_{max}$[*i*]
AC[3]	50 µs	16	31
AC[2]	60 µs	32	63
AC[0]	80 µs	128	255

The AIFS and CW parameters are set as in Table III. In this setting, it is clear that the voice traffic has the highest priority and the TCP traffic has the lowest priority in terms of channel access. Further, as said in Section III, the window ranges of different classes do not overlap and the ratio between CW$_{min}$[3] and CW$_{min}$[2] imposes that the ratio between d_3 and d_2 will be 1/2. These constraints allow us to use Equation (15).

In the simulation, the traffic load is gradually increased. Specifically, until the admission control scheme allows to accept the flows, a new voice flow is added at the time instants of $6 \times i$ seconds ($0 \leq i \leq 10$). Likewise, a video flow is added two seconds later. Only 1 TCP flow is added 4 seconds later the start of the simulation.

Furthermore, to simulate the real scenario where the starting instants of real-time flows are randomly spread over time, the start of a voice flow is delayed a random period uniformly distributed in [0 ms, 40 ms], and that of a video flow delayed a random period uniformly distributed in [0 ms, 125 ms]. Note that in the simulation period between [T_{stop}, 250 s], where T_{stop} is the acceptance time of the last flow according to CAC scheme, we stop injecting more flows into the network in order to observe how well the scheme performs in a steady state.

For the proposed CAC scheme, the QoS requirements expressed in terms of bound delay and delay violation probability of the voice and video traffic are shown in Table IV. A flow is accepted if these QoS requirements can be satisfied without jeopardizing already accepted flows.

TABLE IV. QoS PARAMETERS FOR THE VOICE AND VIDEO TRAFFIC.

Traffic	Access Category	delay bound d_i (ms)	delay violation probability p_i
Voice	AC [3]	100	10^{-3}
Video	AC [2]	200	

The schedulable region, i.e. the set of pairs (N_{voice}, N_{video}) that satisfy the QoS requirement of Table IV is shown in Figure 1 and is obtained by solving Equation (15) for all the pairs.

The CAC procedure proposed in the CARC scheme, requires knowledge both of the peak rate and the mean rate parameters for the real-time traffic; in addition, it requires an estimate of the average delay. Even neglecting the difficulty of evaluating these parameters, a Call Admission based on the mean values is not capable of providing strict guarantees. This aspect is addressed by the new scheme through the use of Equation (15), which guarantees the maximum probability of exceeding a pre-assigned delay bound. The knowledge of the schedulable region makes it possible the choice of different operation points all capable of satisfying the traffic requirements. In particular, we consider in more detail the case with a maximum number of 10 voice flows and 10 video flows. In addition, our scenario includes a TCP session allowed to use any available channel capacity left by real-time traffic.

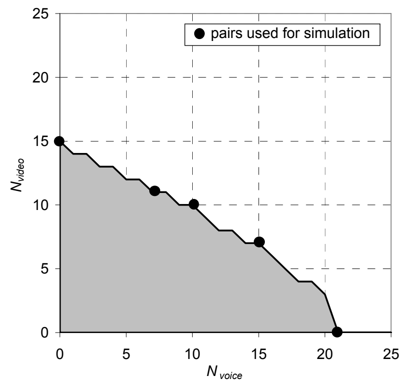

Fig. 1: Schedulable region of the proposed CAC scheme.

Figure 2 shows the throughput for the three traffic classes throughout the simulation. At the beginning, the TCP traffic has high throughput. During the simulation, neither real-time nor best-effort packets are lost. Then, as more real-time flows are admitted, the TCP throughput gradually drops as a result of the rate control as well as the high priority of real-time traffic. As desired, TCP traffic does not completely starve, because real-time traffic is upper bounded by the CU_{rt} value. In addition, TCP traffic is allowed to use any available channel capacity left by the real-time traffic. Therefore, the total channel utilisation due to different types of traffic is high. Moreover, in the unsaturated case, the collision probability is very small and, then, the channel utilisation coincides with the channel busyness ratio.

Simulation results confirm that delay requirements are satisfied for all pairs of the schedulable curve. For this purpose, simulation results have been obtained with the some pairs of voice flows and video flows laying on the curve. and marked with a circle in Figure 1.

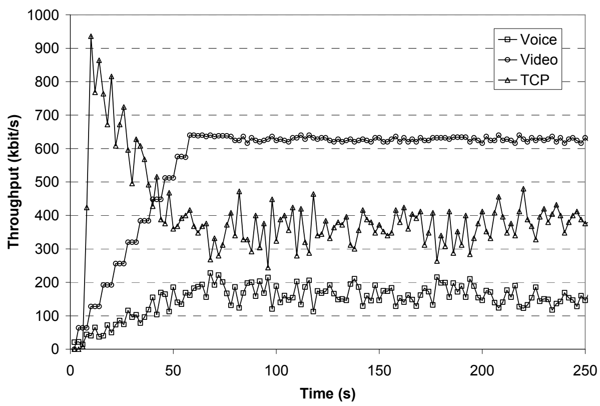

Fig. 2: Aggregate throughput of the proposed CAC scheme.

Delay parameters, shown in Table V, are indicated in terms of the mean, the standard deviation (SD), the 97 percentile, 99 percentile, and 99.9 percentile delays. These results demonstrate that the delay requirements (Table IV) of the real-time traffic can be fully met along the schedulable curve. In all cases, the values of the mean delay are very low. In addition, also the values of the 99.9 percentile are below the target delay bounds.

TABLE V. THE MEAN, STANDARD DEVIATION (SD), AND 97'TH, 99'TH, 99.9'TH PERCENTILE DELAYS (MS) FOR VOICE AND VIDEO WHEN S-CARC SCHEME IS USED WITH THREE CLASSES OF SERVICES.

Simulation		**N**	**Mean (ms)**	**SD (ms)**	**Percentile (ms)**		
					97	***99***	***99.9***
1	Video	15	9.91	7.87	26	35	60
2	Voice	7	7.48	6.82	23	33	52
	Video	11	11.97	10.86	38	50	83
3	Voice	10	7.61	7.01	24	34	53
	Video	10	12.65	11.51	40	47	85
4	Voice	15	8.14	8.08	27	40	67
	Video	7	12.74	14.14	43	67	142
5	Voice	21	5.78	5.19	17	24	46

TABLE VI. THE MEAN, STANDARD DEVIATION (SD), AND 97'TH, 99'TH, 99.9'TH PERCENTILE DELAYS (MS) FOR VOICE AND VIDEO WHEN CARC SCHEME IS USED WITH THREE CLASSES OF SERVICES [4].

	N	**Mean (ms)**	**SD (ms)**	**Percentile (ms)**		
				97	***99***	***99.9***
Voice	10	6.5	5.1	18.5	24.6	41.1
Video	10	12.3	7.4	29.2	37.1	70.8

TABLE VII. THE MEAN, STANDARD DEVIATION (SD), AND 97'TH, 99'TH, 99.9'TH PERCENTILE DELAYS (MS) FOR VOICE AND VIDEO WHEN S-CARC SCHEME IS USED WITHOUT TCP FLOWS.

Simulation		**N**	**Mean (ms)**	**SD (ms)**	**Percentile (ms)**		
					97	***99***	***99.9***
1	Video	15	9.89	7.78	26	35	58
2	Voice	7	7.47	6.76	23	32	50
	Video	11	11.96	10.80	37	50	79
3	Voice	10	6.90	6.35	24	33	47
	Video	10	11.84	10.45	39	47	80
4	Voice	15	8.12	7.98	27	39	64
	Video	7	12.74	15.0	43	66	131
5	Voice	21	5.76	5.08	17	24	44

The case with $N_{voice} = N_{video} = 10$ is also discussed in [4] with reference to the CARC scheme, whose CAC procedure is based on the peak rate, the mean rate, and the average delay equal to 100 ms and 200 ms for voice and video traffic, respectively. The delay parameters, as obtained by simulation in [4], are reported in Table VI in order to compare the two schemes. From Table V and Table VI, we register similar performance. The CARC scheme needs simulation to verify the real performance of the system, while the proposed CARC scheme is able to foresee a closer match between analytical and simulation results.

Finally, simulation results confirm that, as long as the network is kept working in the unsaturated case, the rate control algorithm allows the best-effort traffic not to influence the QoS requirements of real-time services. Performance parameters, shown in Table VII, have been obtained through simulation without TCP flows in the system. There are little differences between delay values given in Table V and in Table VII: therefore, the best-effort traffic does not affect the QoS level of the admitted real-time traffic. In conclusion, a distributed rate control scheme is able to promptly access the residual bandwidth left by the real-time traffic in order to efficiently utilize the channel capacity.

V. CONCLUSIONS

To support strict QoS for real-time traffic, the IEEE 802.11e WLAN is tuned to operate in the unsaturated case. This goal can be achieved by introducing a Call Admission and Rate Control (CARC) scheme, as proposed in literature. In this paper, we have proposed an alternative Call Admission scheme based on a two-moment statistical description of traffic. We have first introduced an analytical model to determine the number of users belonging to two classes of real-time traffic that can be served by a channel of capacity C, without violating the delay requirements. In our analytical model, we have used the token bucket as a traffic regulator and have exploited the approximate equivalence of the access time of the IEEE 802.11e EDCA and the queuing delay of the EDF scheduler. Simulation results have confirmed that our Admission Control scheme ensures QoS guarantees for the real-time traffic in terms of maximum probability of exceeding a given delay bound. In addition, simulation results have confirmed that the high channel utilization can be achieved by adding best-effort traffic, since the available bandwidth can be used without influencing the strict QoS requirements of real-time services.

REFERENCES

[1] X. Chen, H. Zhai, and Y. Fang, "How well can the IEEE 802.11 wireless LAN support quality of service," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 3084-3094, Nov. 2005.

[2] S. Choi, *et al.*, "IEEE 802.11e contention-based channel access (EDCF) performance evaluation," in *Proc. IEEE ICC'03*, May 2003.

[3] S, Mangold, *et al.*, "IEEE 802.11e wireless LAN for quality of service," in *Proc. European Wireless'02*, Florence, Italy, Feb. 2002.

[4] X. Chen, *et al.*, "Supporting QoS in IEEE 802.11e Wireless LANs," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, August 2006.

[5] X. Chen, H. Zhai, and Y. Fang, "A Call Admission and Rate Control Scheme for Multimedia Support over IEEE 802.11 Wireless LANs," *ACM Wireless Networks*, Vol.12, No.4, pp.451-463, May 2006.

[6] J. Qiu and E. Knightly, "Inter-Class Resource Sharing using Statistical Service Envelopes," *in Proceedings of IEEE INFOCOM '99*, New York, NY, March 1999.

[7] E. Knightly, "Resource allocation for multimedia traffic flows using rate variance envelopes", *ACM/Springer-Verlag Multimedia System Journal*, Vol. 7, No. 6, pages 477-485, November 1999.

[8] Draft Supplement to Part 11: Wireless Medium Access Control (MAC) Enhancements for Quality of Service (QoS), *IEEE Std 802.11e/D8.0*, Feb. 2004.

[9] P. Giacomazzi, G. Saddemi, "The equivalence of the non saturated IEEE 802.11e Enhanced Distribution Channel Access and Early Deadline First Scheduler." *Submitted to IEEE Trans. on Communications Letters.*

[10] E. Knightly, "Second moment resource allocation in multiservice networks," *in Proceedings of ACM SIGMETRICS '97*, pages 181-191, Seattle, WA, June 1997.

[11] E. Knightly, "Enforceable quality of service guarantees for bursty traffic streams," *in Proceedings of IEEE INFOCOM '98*, Mar. 1998.

[12] A. Kuzmanovic, E.W. Knightly, "Measuring service in multi-class networks," *in Proceedings of IEEE INFOCOM 2001*, vol. 3, 22-26, Apr. 2001, pp. 1281-1289.

[13] P. Giacomazzi, "Statistical Traffic Envelopes for Markov-Modulated Poisson Packet Sources", *to be published in Proceedings of IEEE GLOBECOM 2007.*