VoIP Call Admission Control in WLANs in Presence of Elastic Traffic

Boris Bellalta, Michela Meo, and Miquel Oliver

Abstract— **VoIP service over WLAN networks is a promising alternative to provide mobile voice communications. However, several performance problems appear due to *i)* heavy protocol overheads, *ii)* unfairness and asymmetry between the uplink and downlink flows, and *iii)* the coexistence with other traffic flows. This paper addresses the performance of VoIP communications with simultaneous presence of bidirectional TCP traffic, and shows how the presence of elastic flows drastically reduces the capacity of the system. To solve this limitation a simple solution is proposed using an adaptive Admission and Rate Control algorithm which tunes the BEB (Binary Exponential Backoff) parameters. Analytical results are obtained by using an IEEE 802.11e user centric queuing model based on a bulk service $M/G^{[1,B]}/1/K$ queue, which is able to capture the main dynamics of the EDCA-based traffic differentiation parameters (*AIFS*, *BEB* and *TXOP*). The results show that the improvement achieved by our scheme on the overall VoIP performance is significant.**

I. Introduction

Access Points (AP) and wireless cards implementing the Enhanced Distributed Channel Access (EDCA) [1] are already available under the Wireless MultiMedia (WMM) denomination. Based on the differentiation capabilities provided by the EDCA specification, some possible solutions for guaranteeing acceptable QoS levels can be achieved, consisting in optimizing the use of the transmission resources. This process requires an entity, the call admission and rate controller, capable to decide whether new traffic flows can be accepted and what resources they can use in order to guarantee the QoS levels of already active flows.

Nowadays, voice calls over Internet (VoIP) are becoming popular, with service providers growing around the world. Most of the VoIP users already have WLANs as access network to the Internet and therefore, the bidirectional VoIP stream goes through the AP from/to a laptop or a VoIP/WLAN phone. In this situation, the VoIP flow competes with typical TCP based flows (like P2P, FTP, e-mail, etc.) to get access to the transmission resources over the wireless channel; this competition causes serious performance problems for the voice data.

Boris Bellalta and Miquel Oliver are with the Departament de Tecnologia, Universitat Pompeu Fabra, Barcelona, Spain, (e-mail: boris.bellalta@upf.edu).

Michela Meo is with the Dipartimento di Elettronica, Politecnico di Torino, Italy, (e-mail: michela@polito.it).

Major part of research efforts evaluate the EDCA differentiation capabilities [2] and MAC protocol performance issues by using both simulation or/and analytical models [3], [4], focusing in some cases in admission control (CAC) and resource management [5], [6], [7].

Admission control schemes have been mainly evaluated through simulations at packet level, i.e., the improvements are only shown as a gain in terms of throughput or delay, [5], [7], [8] or [9]. On the contrary, the flow-level behavior has not been considered and basic metrics such as blocking, dropping probability and average flow transfer delay have not been evaluated. Brief results concerning these metrics are shown in [6].

EDCA models are based on the Distributed Coordination Function (DCF) models from Bianchi [10], Tay et al. [11] or Cali et al. [12], to include the different IEEE 802.11e enhancements: *AIFS*, *BEB* and *TXOP* parameters [3], [4], [13].

The performance perceived by VoIP calls in a hot-spot scenario with the simultaneous presence of persistent TCP connections is investigated in this work. To solve the performance limitations, a combined admission control and rate shaping scheme based on the use of different values of the minimum backoff contention window (CW_{min}) for the uplink TCP flows is proposed. Moreover, a preemptive and a non-preemptive versions of the proposed admission control are also evaluated. At packet level, in order to evaluate the performance of the CAC scheme and provide quantitative analytical results, a mathematical model of the IEEE 802.11e specification [1] is developed. The model is able to capture the main protocol dynamics caused by the different settings of the MAC parameters ($AIFS$, $TXOP$ and CW).

II. A Hot-spot Wireless Scenario

A single AP providing access to a fixed network to n Mobile Nodes (MNs) is considered. The MNs and the AP use the EDCA operation mode and the DSSS PHY specifications in the $2.4\ GHz$ band [14]. Ideal channel conditions are assumed, i.e., no packet is lost due to channel errors or the hidden terminal phenomenon. The system parameters are reported in Table I. This scenario is based on the E-MORANS (Extended Heterogeneous Mobile Radio Access Networks Reference Scenarios) WLAN single cell reference scenario [15].

A. Traffic behavior in current WLANs

Na et al. [16] analyze the traffic composition of several hot-spots. The authors report that more than 90% of the total traffic

TABLE I
SYSTEM PARAMETERS OF THE IEEE 802.11B SPECIFICATION [14]

Parameter	Value	Parameter	Value
R_{data}	2 $Mbps$	R_{basic}	1 $Mbps$
AIFS:A_j	$\{7,3,2,2\}$	R_{PHY}	1 $Mbps$
SIFS	10 μs	CW_{min}	32
TXOP (msec)	$\{0,0,6.016,3.264\}$	CW_{max}	1024
SLOT (σ)	20 μs	ACK	112 $bits$ @ R_{basic}
RTS	160 $bits$ @ R_{basic}	CTS	112 $bits$ @ R_{basic}
MAC header	240 $bits$ @ R_{data}	MAC FCS	32 $bits$ @ R_{data}
PLCP preamble	144 $bits$ @ R_{PHY}	PLCP header	48 $bits$ @ R_{PHY}
Retry Limit (R)	7	K (Queue length)	20 $packets$

is TCP-based, mainly due to HTTP transactions; however, the presence of peer-to-peer (P2P) traffic is also remarkable, with values that are even larger than those obtained by the email or FTP services. Another interesting observation is the asymmetry of the traffic flows: 85% of the total flows go from the fixed network to mobile nodes (downlink) and the remaining 15%, which is still a significant value, on the other direction, from mobile nodes to the fixed network (uplink).

Then, nowadays the Internet user behavior is evolving to use also VoIP and P2P with the already traditional Web browsing, file transfer or e-mail services. In this new situation, WLANs, which are used as access networks, show several performance limitations, such as the difficult simultaneous coexistence of TCP and UDP flows, as WLAN technology was not designed to support these heterogeneous traffic conditions. For example, the presence of TCP traffic in both the downlink (buffer losses) and specially in the uplink (AP starvation) leads to low performance of VoIP calls.

These problems have to be solved in order to deploy a successful VoIP service over WLANs. In the downlink, a simple classification/prioritization scheme can be used (four Access Categories (AC) are defined in EDCA [1]) where the TCP and VoIP packets can occupy separated buffers with priority to the VoIP packets. However, the main problems raise in presence of uplink TCP flows because nodes act independently from each others. The only possible solution is setting different MAC parameters (such as $AIFS$ [17], CW_{min} [18] or/and $TXOP$ [19]) at each mobile node in order to reduce the interaction of these TCP flows with the VoIP calls.

B. EDCA-based WLAN

The EDCA mode of operation of the IEEE 802.11e medium access control classifies each traffic flow in an Access Category (AC). Four AC are defined, each one associated to a single MAC transmission queue. Letting $AC_{i,j}$ be the access category j of the i MN, the basic MAC parameters of each access category are labeled as: Arbitration Interframe Space $AIFS_{i,j}$, Minimum Contention Window $CW_{min,i,j}$, Maximum Contention Window $CW_{max,i,j}$ and Transmission Opportunity $TXOP_{i,j}$. The default values of each parameter are reported in [1].

According to the basic access (BA) mechanism, when node i has no packets to transmit and receives a packet from network layer, it sends the packet to the corresponding $AC_{i,j}$ queue. At the same time, the node starts to sense the channel to determine its state, that can be either *busy* or *free*. If the channel is detected busy, the node waits until the channel is released. When the channel is detected free for a period of time larger than the $AIFS_{i,j}$ duration, a new backoff instance is generated, which consists on a counter set to a random value. The random value is picked from an uniform distribution in the range $CW_{i,j}(k) = [0, min\left(2^k CW_{min,i,j} - 1, CW_{max,i,j} - 1\right)]$, where k is the current packet transmission attempt. For each packet to be transmitted, k is initially set to 0 and it is increased by one at each failed transmission until a maximum number of retransmissions, called Retry Limit, is reached, and the packet is dropped.

The backoff counter is decreased by one for each time-slot in which the channel is sensed free, until the countdown reaches zero, instant in which the node starts the packet transmission on the channel. If, during the backoff countdown, the channel is sensed busy, the backoff is suspended until the channel is detected free again. The $AIFS_{i,j}$ value is computed using a non-negative integer $A_{i,j}$ specific for each $AC_{i,j}$: $AIFS_{i,j} = SIFS + A_{i,j}\sigma$ (where σ is an empty SLOT duration). Once a node gets the channel, it can transmit up to $B_{i,j}$ $MPDU$ packets ($TXOP_{i,j}$ limit). This limit is expressed in time units (ms) and corresponds to the consecutive time that a node can transmit few (large) or several (small) packets, where $B_{i,j}$ is computed by considering the average duration of the packets transmitted by node i.

A channel collision occurs if two nodes transmit at the same time, i.e., a backoff instance from two nodes reach 0 at the same time. After the data packet is transmitted to the channel by the sender, the receiver waits for a SIFS (Short Inter-Frame Spacing) time and sends a MAC layer ACK to acknowledge the correct reception of the data packet. In the case the sender does not receive the ACK frame, it starts the retransmission procedure. After discarding or successfully transmitting a packet, if more packets are ready to be transmitted, the node starts the transmission procedure again. Otherwise, it waits for a new packet from the network layer. Another EDCA feature is the use of the different ACK policies (no ACK transmission

or ACKs aggregation) which can also be used to improve the system performance.

Alternative to the BA mechanism, nodes can employ a RTS/CTS protocol to access the channel, so as to reduce the hidden terminal effect.

III. A MODEL OF THE IEEE 802.11E EDCA

An user-centric model of the IEEE 802.11e-EDCA operation mode is proposed. Each mobile node is approximated by a finite length queue with bulk and network-dependent service times. Each mobile node is assumed to carry a single traffic flow of category $AC_{i,j}$ (or equivalently, each node i has only one $AC_{i,j}$ active at the same time). For the sake of simplicity, henceforth the subscript j is omitted as a single AC is considered to be active in each MN.

A. A mobile node

Packets with average length L_i arrive to node i with average rate α_i. Both the time between packet arrivals and the service time are assumed to be exponentially distributed. Each node (the AP included) is modeled by an $M/G^{[1,B_i]}/1/K_i$ queue with bulk services times and queue length K_i (which includes the packet/s in service) measured in packets (Figure 1).

B. The $M/G/1^{[1,B]}/K$ queue

The $M/G^{[1,B_i]}/1/K_i$ queue is used to capture the multiple packet transmission behavior each time the MN_i gets the channel. Several assumptions are done:

1) The number of packets transmitted at bulk $m+1$ is variable and depends on the number of packets remaining at the queue after the m bulk transmission.
2) If after the $m-$th bulk transmission the queue is empty, next transmission always comprises a single packet.
3) The bulk of packets is not dequeued until they are completely transmitted.

1) Arrival and equilibrium stationary distributions: The steady state probability that q packets are in the MN_i queue at any arbitrary instant is denoted by $\pi_{q,i}^s$. The bulk service time X_i^b, where the super-index b refers to the length of the burst, is used to compute the departure rates from each state, $\beta_i^b = 1/X_i^b$. The steady state probabilities are computed by solving the $\pi_i^s Q_i = 0$ linear equation system, where Q_i is the infinitesimal generator matrix for the model of the queue. From the $PASTA$ property, the arrival distribution equals the stationary distribution, $\pi^a = \pi^s$. Therefore, from the arrival and equilibrium distribution, several performance metrics are computed

$$
\begin{cases}
\rho_i = 1 - \pi_0^s \\
P_{b,i} = \pi^a(K_i) = \pi^s(K_i) \\
E[Q_i] = \sum_{i=0}^{K} \pi_i^s \cdot i \\
E[D_i] = \frac{E[Q_i]}{\alpha_i(1-P_{b,i})} \\
P_{L,i} = P_{b,i} + (1-P_{b,i})P_{d,i}
\end{cases} \quad (1)
$$

where ρ_i is the queue utilization, $P_{b,i}$ is the packet blocking probability, $E[Q_i]$ is the average queue occupation, $E[D_i]$ is the average queuing delay and $P_{L,i}$ is the probability to loose a packet, with $P_{d,i}$ as the probability that a packet is dropped at the MAC layer because the number of retransmissions has exceeded the retry limit.
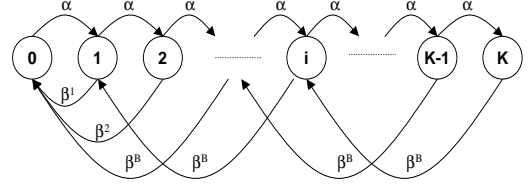

Fig. 1. $M/G^{[1,B]}/1/K$, $B < i$

2) Departure stationary distribution: To compute the π^d distribution, let ξ_m be the random variable denoting the number of packets in the queue after the completion of the $m-$th bulk transmission. Then, the steady state probability of q packets after a departure is

$$
\pi_q^d = \lim_{m \to \infty} Prob\{\xi_m = q\}, \quad \forall q = 0 \ldots K \quad (2)
$$

To find π_q^d the embedded ξ_m Markov process is considered, with space state $S = \{D_0, D_1, \ldots D_{K-1}, D_K\}$, where state D_q means that there are q packets remaining in the queue. The transition probabilities from state D_f to state D_g are

$$
p_{g,f} = lim_{m \to \infty} Prob\{\xi_{m+1} = g | \xi_m = f\}, \quad 0 \leq g, f \leq K \quad (3)
$$

The probability of h packets arrivals during the transmission of a bulk of length B packets, $A_{h,B}$, is related to the service time distribution of the burst, $f_b(B,t)$, which is different for each B value and is approximatively equal to the sum of B exponential distributed packets, which results in an $Erlang-B$ distribution. However, for simplicity, is assumed that all burst are exponentially distributed, so

$$
A_{h,B} = \int_{t=0}^{\infty} f_a(j,t) f_b(B,t) dt
$$
$$
= \int_{t=0}^{\infty} \frac{(\alpha t)^j}{j!} e^{-\alpha t} \beta^B e^{-\beta^B t} dt \quad (4)
$$

where $f_a(h,t)$ is the arrival process distribution which is assumed to follow a Poisson process. An example of the transition probability matrix P of the embedded Markov Chain for $B = 2$ and $K = 4$ is shown in Eq. 5. Notice that $\pi_K^d = 0$ as the packets are dequeued only after they are transmitted. Thus, the queue is never full at a departure epoch.

$$
P = \begin{pmatrix}
A_{0,1} & A_{1,1} & 1 - A_{0,1} - A_{1,1} & 0 \\
A_{0,1} & A_{1,1} & 1 - A_{0,1} - A_{1,1} & 0 \\
A_{0,2} & 1 - A_{0,2} - A_{0,1} & 0 & 0 \\
0 & 1 & 0 & 0
\end{pmatrix}
$$
$$
(5)
$$

Once, the departure distribution is found, by solving the

linear system $\pi^{\mathbf{d}}\mathbf{P} = \pi^{\mathbf{d}}$, the average number of packets transmitted in a bulk and obviously, the average bulk length in bits or in time units can be computed.

$$E[B_i] = \pi_0^d + \sum_{k=1}^{K} k \cdot \pi_k^d \qquad (6)$$

Notice that the derivation of $\pi^{\mathbf{d}}$ is computationally costly. A solution is to approximate $\pi^{\mathbf{d}}$ with $\pi^{\mathbf{s}}$ which leads to worst case results.

C. The MAC protocol

Letting ζ_i be the average number of backoff slots, including the BEB and AIFS-blocked slots, at each transmission attempt by node i, the steady state probability that the node transmits in a random slot given that a packet is ready in its transmission queue can be approximated by

$$\tau_i = \frac{lim_{t\to\infty}Pr(Q_i(t)>0)}{\zeta_i+1} = \frac{1-\pi_0^s}{\zeta_i+1} = \frac{\rho_i}{\zeta_i+1} \qquad (7)$$

The ζ_i parameter includes the blocked slots for each $AIFS$ interval (A_i) and the average number of back-off slots selected at each transmission attempt, EB_i, which can be computed from the expression presented by Tay et al. [11] as

$$EB_i = \frac{1-p_i-p_i(2p_i)^{m_i}}{1-2p_i}\frac{CW_{min,i}}{2} - \frac{1}{2} \qquad (8)$$

or using the expression derived by Bianchi [10]. The expression presented by Wu in [20], that includes the impact of retry limit, can also be used. However, the effects of the retry limit can be neglected in non-saturation conditions.

In order to compute the value of ζ_i an approach similar to the one in [21] is considered. It is taken into consideration that the number of transmissions observed by a node during its back-off is $p_{tr,i} \cdot EB_i$, which implies an extra number of blocked slots that the node has to wait approximatively equal to $p_{tr,i} \cdot EB_i \cdot A_i$, then

$$\begin{cases} \zeta_i \approx (EB_i+A_i) + p_{tr,i} \cdot EB_i \cdot A_i \\ p_{tr,i} = 1 - \prod_{j\neq i}(1-\tau_j) \end{cases} \qquad (9)$$

where $p_{tr,i}$ is the probability that at least another node transmits in a given slot.

The service time, i.e., the time interval from the instant in which a packet enters in service until it is completely transmitted or discarded, is given by,

$$X_i^{b_i} = (M_i-1)\left(\zeta_i\gamma_i + ET_{c,i}^{ba||rts}\right) + \zeta_i\gamma_i + ET_{s,i}^{b_i,ba||rts} \qquad (10)$$

where M_i is the average number of required transmissions, γ_i is the average slot duration, $ET_{s,i}^{b_i,ba||rts}$ is the average duration of a $E[B_i]$ packets burst transmission, which is computed from the $\pi^{\mathbf{d}}$ distribution, using the BA or the RTS/CTS access mechanism, and $ET_{c,i}^{ba||rts}$ is the average duration of a collision of node i.

The value of $ET_{c,i}^{ba||rts}$ is approximated by,

$$\begin{cases} ET_{c,i}^{ba} \approx \frac{\sum_{j\neq i}\tau_j \max(T_{s,i}^{1,ba},T_{s,j}^{1,ba})}{\sum_{j\neq i}\tau_j} \\ ET_{c,i}^{rts} = T_c^{rts} \end{cases} \qquad (11)$$

where it is neglected the fact that more than two packets collide simultaneously. Note that if the RTS/CTS access scheme is used, $ET_{c,i}^{rts}$ is constant and equal for all nodes.

A node frozes its backoff counter every time the channel is sensed busy and releases it after the channel is sensed free for an $AIFS$ period. Therefore, the time between two backoff counter decrements is a random variable which depends on the behavior of the other nodes. By letting γ_i be the average time between two backoff counter decrements, or equivalently, the average slot duration,

$$\gamma_i = p_{e,i}\sigma + p_{s,i}(ET_{s,i}^{ba||rts,*}+\sigma) + p_{c,i}(ET_{c,i}^{ba||rts,*}+\sigma) \quad (12)$$

where $ET_{s,i}^{ba||rts,*}$ and $ET_{c,i}^{ba||rts,*}$ are the average durations of an observed successful transmission or a collision for node i when it is performing a backoff instance. By neglecting the probability that more than two stations collide,

$$\begin{cases} ET_{c,i}^{ba,*} \approx \frac{\sum_{j\neq i}\sum_{k>j,k\neq i}\max(T_{s,j}^{1,ba},T_{s,k}^{1,ba})\left(\tau_j\tau_k\prod_{u\neq\{j,k,i\}}(1-\tau_u)\right)}{\sum_{j\neq i}\sum_{k>j,k\neq i}\left(\tau_j\tau_k\prod_{u\neq\{j,k,i\}}(1-\tau_u)\right)} \\ ET_{c,i}^{rts,*} = T_c^{rts} \end{cases} \qquad (13)$$

and

$$ET_{s,i}^{ba||rts,*} \approx \frac{\sum_{j\neq i}ET_{s,j}^{b_j,ba||rts}\left(\tau_j\prod_{u\neq\{i,j\}}(1-\tau_u)\right)}{\sum_{j\neq i}\left(\tau_j\prod_{u\neq\{i,j\}}(1-\tau_u)\right)} \qquad (14)$$

where $ET_{s,j}^{b_j,ba||rts}$ is the average duration of a successful transmission from a node j. Notice that this average duration is related also to the queue occupation and can be averaged from the $\pi^{\mathbf{d}}$ distribution.

The probabilities $p_{e,i}$, $p_{s,i}$ and $p_{c,i}$ are related to the channel status (empty, successful transmission and collision) in a given slot when a node is performing its backoff.

$$\begin{cases} p_{e,i} = \prod_{j\neq i}(1-\tau_j) \\ p_{s,i} = \sum_{z\neq i}\tau_z\prod_{j\neq z\neq i}(1-\tau_j) \\ p_{c,i} = 1 - p_{e,i} - p_{s,i} \end{cases} \qquad (15)$$

In order to validate the model and compare the performance of the DCF and $EDCA$ in a particular scenario, a single-hop network with two traffic types has been considered: elastic traffic, for which nodes are saturated and the frame length is equal to 1500 Bytes, and rigid traffic, with bandwidth of 100 Kbps and frame length equal to 400 Bytes. The network comprises n nodes, each node uses the BA access scheme and carries a single traffic flow. R is used to refer to unsaturated, or rigid, flows and E is used to refer to saturated, or elastic, flows. Analytical results are compared against simulations performed using a detailed simulator of the EDCA IEEE 802.11e MAC protocol built using the COST (Component Oriented Simulation Toolkit) simulation engine [22].

In Figure 2 the aggregate throughput is plotted for four E flows when the number of R flows increases. Each rigid flow uses a TXOP limit of $B_R = 4$ packets, a $CW_{min,R} = 8$ and $A_R = 2$ while the E flows use $B_E = 1$ packets, $CW_{min,E} = 32$ and $A_E = 3$.

Notice how the DCF is already saturated for a single R

flow while EDCA allow up to 8 rigid flows. Obviously, using EDCA the E throughput is lower.
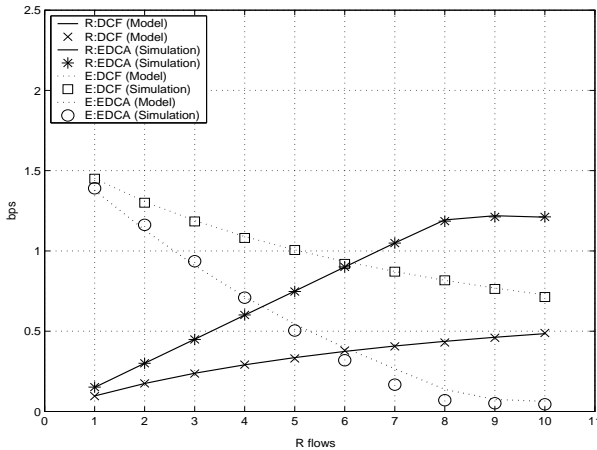

Fig. 2.   Aggregate Throughput for elastic and rigid flows

IV. Call Admission Control

In previous section, it has been pointed out that the use of EDCA is more suitable for WLAN networks where rigid (e.g., VoIP) and elastic (BE) flows coexist. However, in order to avoid the system saturates, thus causing an overall WLAN degradation, admission control is needed to block new flows, or drop already active ones, when the system cannot allocate enough resources for new flows.

One basic desirable feature of an admission control mechanism is the ability to adapt the rate of elastic flows in order to release resources which can be allocated to a new arriving rigid flow (and the opposite procedure when a rigid flow departs). In EDCA-based WLANs this rate control can be done by setting different MAC parameters for each AC. Currently, the adequate parameter selection still requires further research as there are no general tuning criteria, a simple parameter tuning algorithm can be found in [23].

Therefore, in this section, an admission and rate control scheme based on tuning the BEB parameters (CW_{min}) of the MN AC_BE (the AC used by BE flows) is proposed. The $CW_{min,BE}$ parameter is updated at each rigid flow arrival / departure with the goal of maximizing the BE throughput but at the same time providing the required protection to VoIP (VO) flows (which use the AC_VO access category). Moreover, the proposed admission control drops best-effort (BE) flows when the rate of BE flows can not be decreased enough to allocate the new arriving VoIP call.

A. CAC Architecture

The call admission control entity is located at the AP. When an application wants to use the cell resources, it sends an $ADDTS$ (Add Traffic Specification) packet to the AP with the traffic profile required by the flow [1]. Using the information provided by the application, the call admission control decides if the new state of the network is feasible. If so, it sends a positive response to the request. Otherwise, it sends a negative response and the new flow is rejected, preserving the grade of service of the active flows already in the system. However, another possibility is to drop one or several of the already active flows to allocate resources for the incoming request. In that case, a novel (not specified in the current standard) ADDTS drop response is sent to the selected MN/s and a positive one for the new arriving flow.

To differentiate TCP and VoIP downstream flows, the AP uses two different AC queues: voice queue (AC_VO) for VoIP packets and best-effort queue (AC_BE) for TCP packets. The service prioritization for the voice queue is given only by a short $AIFS$ as has been considered that the CW_{min} and the $TXOP$ burst parameters are equal for both queues with values 32 and 1 respectively. It is referred with $\rho_{s,d}$ to the AC_VO queue utilization and with $\rho_{e,d}$ to the AC_BE queue utilization. To simplify the analysis, it is assumed that TCP packets are served only when the AC_VO queue is empty, then the probability to transmit downstream TCP packets is $1 - \rho_{s,d}$. Note that, since the upstream feedback traffic is proportional to the downstream TCP traffic, a minimal impact of uplink TCP ACKs over the VoIP packets is assumed.

The VoIP and TCP upstream flows use respectively the AC_VO and AC_BE queues of mobile nodes. They use burst of size $B = 1$, different and fixed values of $AIFS$ and different values of the CW_{min} parameter (variable for TCP uplink flows and determined dynamically by the admission control entity). Let Φ_{CW} be the set of all possible CW_{min} values that can be used by uplink BE flows, with $\Phi_{CW} = \{32, 64, 128, 256, 1024\}$. When the CAC receives a new request of a VoIP call or a TCP uplink flow, it computes the suitable CW_{min} value for the new and remaining active uplink BE flows and broadcasts the new CW_{min} value in next beacon frame. If there are no VoIP flows in the system, it is assumed that all nodes and the AP use the default value of $CW_{min} = 32$.

B. Performance Evaluation

Results are shown as the comparison of the performance achieved by the WLAN using a non adaptive CAC scheme called *AC-S* (static), according to which all flows always use a fixed value of CW_{min} equal to 32, and using the proposed CAC scheme, called *AC-A* (adaptive). Moreover, two preemption policies are considered for both schemes: Non-Preemption (NP) and Preemption (P), where uplink TCP flows can be dropped to allocate their bandwidth resources to new arriving VoIP flows.

1) Source Traffic Models:

- *The G.729 flow traffic model.* The VoIP calls use the $G.729$ codec with a source bandwidth of 8 $Kbps$. Each flow is modeled by a Poisson process with packet sending rate of $\alpha_s = 50$ packets/second. The RTP, UDP and IP header are added to 20 Bytes of voice data, resulting in a packet length of $L = 60$ Bytes at the MAC layer entrance. Thus, a raw bandwidth of $B_s = 24$ Kbps has to be managed by the MAC layer for each VoIP flow.
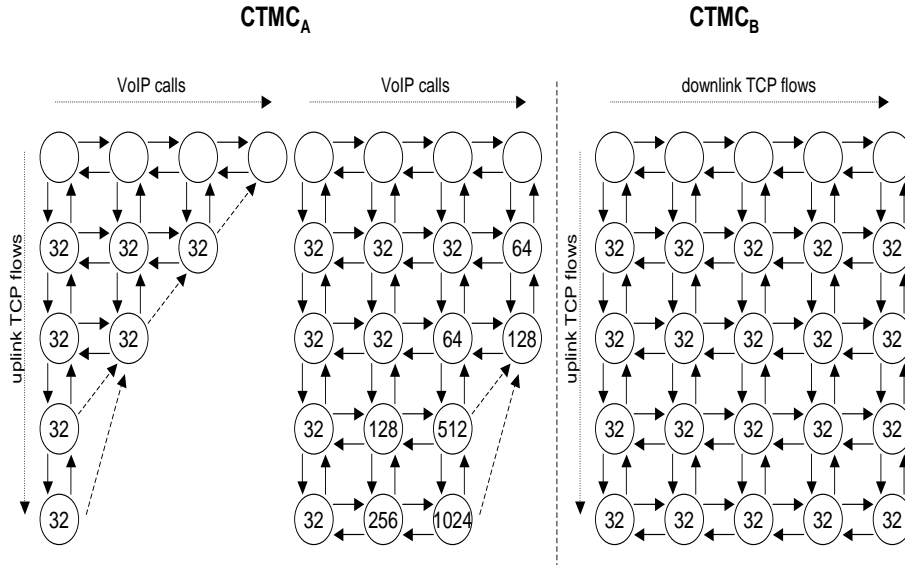
Fig. 3. CTMCs Example. Feasible states and CW_{min} values for uplink TCP flows. Dashed lines show the preemption transitions.

- *A bidirectional TCP flow model for WLANs*. The TCP traffic model used can be found in [24]. The TCP packet length considered is set to 1500 Bytes, including the TCP header). Basic modeling assumptions are the following. First, the downlink queue is always saturated. Second, the average packet length transmitted by the AP is computed from $EL_d = \phi_d L_{tcp} + \phi_u L_{ack}$ where ϕ_d and ϕ_u are the probability that a packet sent by the AP is a data or an ACK packet. These probabilities are computed from: $\phi_d = 1 - \phi_u$ and $\phi_u = E\tau_{e,u} \cdot n_{e,u}/(E\tau_{e,u}n_{e,u} + \tau_{e,d})$. Moreover, the $E\tau_{e,u}$ parameter is the MN expected uplink slot transmission probability (considering only the nodes with uplink TCP data packets) and $\tau_{e,d}$ is the AP slot transmission probability. Finally, MNs with uplink TCP data packets are always saturated. Nodes with uplink ACKs have $\alpha_{ack,u} = \alpha_{e,d}/n_{e,d}$, where $\alpha_{e,d} = \phi_d/X_d(EL_d)$, where $X_d(EL_d)$ is the AP service time.

2) Model of a cell: Under the assumption of exponential distributions of flow arrivals and departures, the system can be described by a three-dimensional Continuous Time Markov Chain (CTMC). In order to solve the model, the CTMC can be broken into two bi-dimensional CTMCs. The first CTMC ($CTMC_A$) comprises the situation where the VoIP calls compete with uplink TCP flows and second CTMC ($CTMC_B$) the situation where downlink TCP flows compete with uplink TCP flows.

The partial results of both CTMC are averaged using the approximation that with probability $\rho_{s,d}$ the system works in the situation described by $CTMC_A$ and with probability $1 - \rho_{s,d}$ the system behavior can be modeled by $CTMC_B$. Notice that $\rho_{s,d}$ is the AP AC_VO queue utilization.

While the number of BE flows can grow to infinity, the maximum number of streaming flows is limited by the bandwidth requirements of the voice calls to N_{voip}^{th}. In order to solve these infinite bi-dimensional CTMCs, the state space must be truncated. Without loss of generality, a realistic minimum bandwidth B_e^{min} required for a BE flow is introduced, which gives a maximum number of $N_{e,u}^{th}$ ($N_{e,d}^{th}$) uplink (downlink) BE flows. The CTMCs state space is described by

$$S_A = \{(n_{e,u}, n_s)| \ S_{e,u}^{tcp}(n_{e,u}, n_s)/n_{e,u} \geq B_e^{min}$$
$$, \ S_{s,d}^{voip}(n_{e,u}, n_s) > 0.97 \cdot n_s B_s\}$$
$$S_B = \{(n_{e,u}, n_{e,d})| \ S_{e,u}^{tcp}(n_{e,u}, n_{e,d})/n_{e,u} \geq B_e^{min},$$
$$S_{e,d}^{tcp}(n_{e,u}, n_{e,d})/n_{e,d} \geq B_e^{min}\}$$

$$(16)$$

Notice that if the preemptive policy (P) is considered transitions between state $S_A(n_{e_u}, n_s)$ and $S_A(n_{e,u} - d_{e,u}, n_s + 1)$ are considered, where $d_{e,u}$ is the number of dropped uplink TCP flows. For the detailed procedures to compute blocking and dropping probabilities in multi-service wireless networks, refer to [25].

For both voice and best-effort flows the user population is considered to be infinite with steady state arrival rates $\lambda_{e,u}$ and $\lambda_{e,d}$ for BE flows and λ_s for VoIP calls. The BE flow duration is a function of the bandwidth observed by the BE flows and the flow length (amount of data to be transmitted) FL_e, with departure rate equal to $\mu_{e,x} = S_{e,x}^{tcp}(.)/(n_{e,x}FL_e)$, while VoIP calls have a fixed average duration equal to $1/\mu_s$. An example of both the CTMCs is depicted in Figure 3. It can be observed how by properly tuning the $CW_{min,BE}$ parameter the number of feasible states grows.

C. Parameters

It is expected that the *AC-A* provides better performance than the *AC-S* as the number of coexisting VoIP calls and
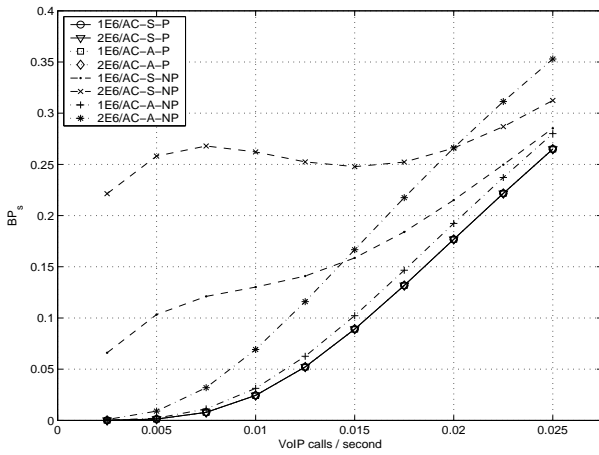
Fig. 4. BP_s - Blocking probability VoIP calls
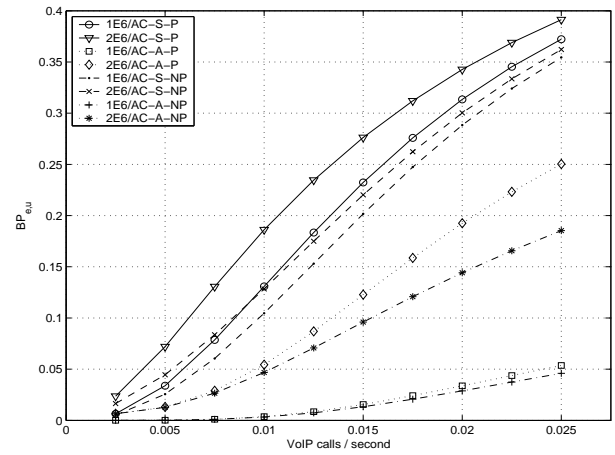
Fig. 5. $BP_{e,u}$ - Blocking probability for BE uplink flows

TCP flows grows. In order to quantify these expected gains, next scenarios are evaluated:

- Scenario A: *Arrival rate increment of VoIP calls.* VoIP calls arrive to the system with rate λ_s and have an average duration equal to $1/\mu_s = 240$ seconds. Both downlink and uplink TCP flows have flow length equal to $FL_e = 1$ or $FL_e = 2$ Mbits, and arrive to the system with $\lambda_{e,d} = 0.5$ and $\lambda_{e,u} = 0.5$ flows / second, respectively.

- Scenario B: *Arrival rate increment of uplink TCP flows.* VoIP calls arrive to the system with rate $\lambda_s = 54$ calls / hour and have duration equal to $1/\mu_s = 240$ seconds. Both downlink and uplink TCP flows have flow length equal to $FL_e = 1$ or $FL_e = 2$ Mbits, and arrive to the system with $\lambda_{e,d} = 0.5$ and variable $\lambda_{e,u}$ flows / second, respectively.

In all cases, the flow arrival and departure rates follow a Poisson process, the flow length has an exponential distribution and, for the sake of simplicity, B_e^{min} is computed to allow a maximum number $N_{e,u}^{th} = N_{e,d}^{th} = 10$ of uplink and downlink active BE flows in the system in all situations. The maximum number of G.729 VoIP calls is equal to 7.

D. Numerical Results

Results obtained using the *AC-A* can be better understood if the two following aspects are taken into consideration: *i)* the number of feasible states of the CTMC grows (i.e., the number of coexisting uplink TCP flows and VoIP calls grows) and *ii)* the *AC-A* reduces the instantaneous throughput of TCP uplink flows, increasing their latency, that is the time they are active in the system.

First, the effect of increasing the VoIP call arrival rate is investigated in terms of blocking (BP_s) probability for VoIP calls and blocking ($BP_{e,u}$) and dropping ($DP_{e,u}$) probability for uplink TCP flows. Results are shown in Figures 4, 5 and 6, respectively.

By using preemption, the blocking probability for VoIP calls is equal for both the CAC schemes and it only depends on the VoIP traffic load. This is motivated by the dropping of
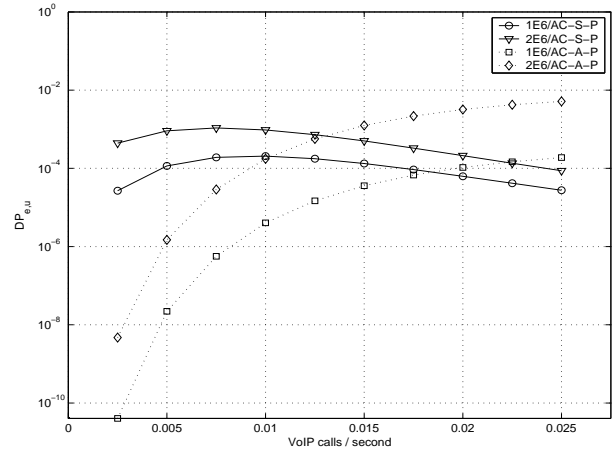
Fig. 6. $DP_{e,u}$ - Dropping probability for BE uplink flows

$d_{e,u}$ uplink TCP flows each time a new VoIP call arrives to the system and there are available resources for the new VoIP call but they are used by the $d_{e,u}$ uplink TCP flows. However, the gain achieved in terms of BP_s is compensated with the dropping probability of uplink TCP flows, which is plotted in Figure 6. It can be seen that *AC-S* shows a clear tendency to decrease the dropping probability as λ_s increases, showing better $DP_{e,u}$ than the *AC-A*. Moreover, the use of preemption also increases the blocking probability of uplink TCP flows for all λ_s values in both, *AC-S* and *AC-A* schemes. Notice, also, how the BE flow length has a lower impact when the *AC-A* is used with respect to the case of *AC-S*.

If preemption is not used, *AC-A* performs notably better for lower values of λ_s in terms of BP_s. However, for higher λ_s values, the BP_s of *AC-A* tends to be higher than using *AC-S* as is shown in Figure 4. This is motivated by the system tendency to be always occupied with uplink TCP flows, as their rate has been decreased, and there are no the possibility to drop them when a new VoIP call request arrives. However, for the uplink TCP flows, the proposed adaptive scheme always show better results.
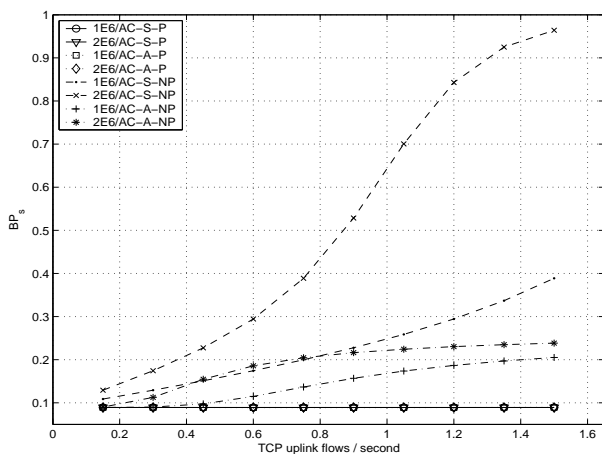
Fig. 7. BP_s - Blocking probability VoIP calls
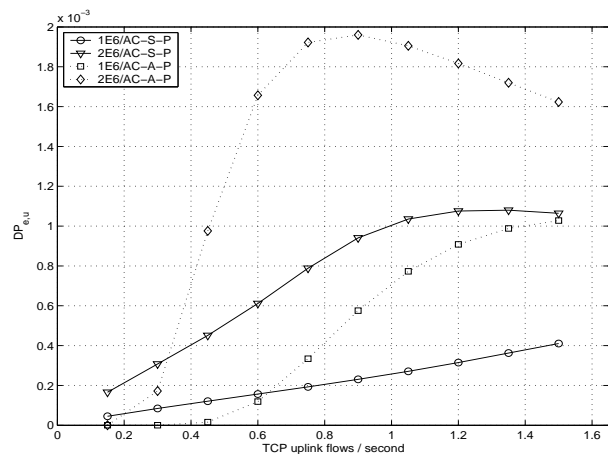
Fig. 9. $DP_{e,u}$ - Dropping probability for BE uplink flows
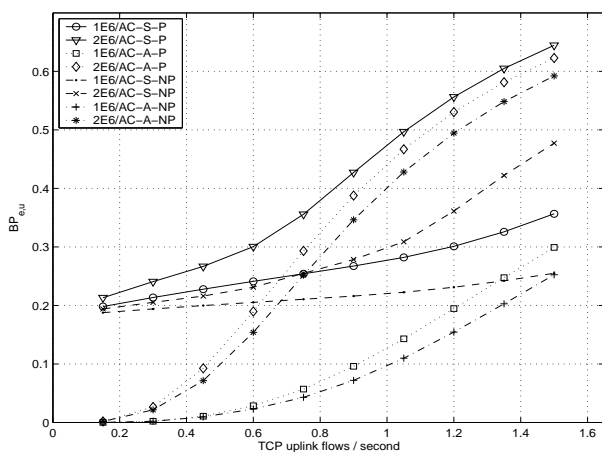
Fig. 8. $BP_{e,u}$ - Blocking probability for BE uplink flows

In order to test the sensitivity of the novel scheme to variations of the traffic intensity of the uplink TCP flows, the arrival rate of VoIP calls is kept constant and the TCP uplink flow offered load is changed. The results are plotted in Figures 7, 8 and 9.

Using preemption, the VoIP blocking probability is, as expected, insensitive to variations of uplink BE load. The cost, similar to the previous case, is a higher probability of dropping BE TCP flows. In terms of $BP_{e,u}$, both schemes show a higher blocking probability than without preemption. In that case, the *AC-A* exhibits substantially lower values for BP_s than the *AC-S* and slightly higher than the one obtained using preemption. The impact of the BE flow length is clearly noticeable in BP_s, especially using the *AC-S* scheme.

In all situations, the higher number of accepted VoIP calls using the *AC-A* results in an increment of the VoIP throughput and then, a higher utilization ($\rho_{s,d}$) of the downlink AC_VO queue, decreasing the transmission opportunities of the TCP downlink traffic, which increases also the TCP downlink blocking probability.

The benefits of preempt uplink TCP flows are shown clearly in terms of VoIP calls blocking probability. However, the cost

of dropping BE TCP flows has to be quantified in order to decide if it is acceptable.

V. CONCLUSIONS

The performance of the IEEE 802.11e MAC protocol (EDCA) has been investigated by means of a novel user-centric model based on a $M/G^{[1,B]}/1/K$ queue which captures the main impact of the use of different MAC parameters (AIFS, BEB and TXOP) for each Access Category. This model is used to evaluate an admission and rate control algorithm based on tuning the CW_{min} parameter of BE flows.

Results show how using the proposed CAC algorithm the performance of VoIP calls is significantly improved using both preemption and non-preemption of uplink TCP flows. However, it is remarkable that, for a set of scenarios, the adaptive scheme without preemption performs worse than the static one, showing the importance to evaluate any scheme at both packet and flow level in order to properly assess its possible drawbacks.

REFERENCES

[1] IEEE Std 802.11e. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications; Amendment: Medium Access Control (MAC) Quality of Service Enhancements. *IEEE Std 802.11e*, 2005.

[2] L. Scalia and I. Tinnirello. Inter-frame space (IFS) based service differentiation for IEEE 802.11 wireless LANs. In *IEEE BroadNets, Workshop on Broadband Wireless Multimedia, San José, California*, October 2004.

[3] J. W. Robinson and T. S. Randhawa. Saturation Throughput Analysis of IEEE 802.11e Enchanced Distributed Coordination Function. In *IEEE Journal on Selected Areas in Communications, Vol, 22, No. 5*, June 2004.

[4] Peter Clifford, Ken Duffy, John Foy, Douglas J. Leith, and David Malone. Modeling 802.11e for data traffic parameter design. In *WiOpt'06, Limassol, Cyprus*, April 2006.

[5] Dennis Pong and Tim Moors. Call Admission Control for IEEE 802.11 Contention Acess Mechanism. *IEEE Globecom 2003, San Francisco, USA*, December 2003.

[6] Deyun Gao and Jianfei Cai. Admission Control in IEEE 802.11e Wireless LANs . *IEEE Network*, July/August 2005.

[7] Hongqiang Zhai, Xiang Chen, and Yuguang Fang. A Call Admission and Rate Control Scheme for Multimedia Support over IEEE 802.11 Wireless LANs. *Quality of Service in Heterogeneous Wired/Wireless Networks (QShine'04), Dallas, USA*, October 2004.

[8] Kiran Kumar Gavini, Varsha Apte, and Sridhar Iyer. PLUS-DAC: A Distributed Admission Control Scheme for IEEE 802.11e WLANs. *International Conference on Networking (ICON), Kuala Lumpur, Malaysia*, November 2005.

[9] Liqiang Zhang and Sherali Zeadally. HARMONICA: Enhanced QoS Support with Admission Control for IEEE 802.11 Contention-based Access. *10th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS'04), Toronto, Canada*, May 2004.

[10] Giuseppe Bianchi. Performance Analysis of the IEEE 802.11 Distributed Coordination Function. *IEEE Journal on Selected Areas in Communications, Vol. 18, No. 3*, March 2000.

[11] Y.C. Tay and K.C. Chua. A Capacity Analysis for the IEEE 802.11 MAC Protocol. *Wireless Networks 7, 159-171*, March 2001.

[12] F. Cali, M. Conti, and E. Gregori. Dynamic Tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit. *IEEE/ACM Transactions on Networking, vol 18, pp. 785-799*, December 2000.

[13] Dapeng Wu Kejie Lu, Jianfeng Wang and Yuguang Fang. Performance of A Burst-Frame-Based CSMA/CA Protocol for High Data Rate Ultra-Wideband Networks: Analysis and Enhancement. In *The Third International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine 2006), Waterloo, Ontario, Canada*, August 7-9 2006.

[14] IEEE Std 802.11. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. *ANSI/IEEE Std 802.11*, 1999 Edition (Revised 2003).

[15] NewCOM AP-7 partners. NEWCOM single-cell WLAN reference scenario. *NewCOM NoE*, To appear (January 2007).

[16] C. Na, J. K. Chen, and T. S. Rappaport. Hotspot Traffic Statistics and Throughput Models for Several Applications. In *IEEE Globecom 2004, Dallas, Texas, USA*, November 2004.

[17] Chun-Ting Chou, Kang G. Shin., and Sai N. Shankar. Inter-frame space (IFS) based service differentiation for IEEE 802.11 wireless LANs. In *IEEE 58th Vehicular Technology Conference, VTC 2003-Fall*, October 2003.

[18] Albert Banchs, Xavier Pérez-Costa, and Daji Qiao. Providing Throughput Guarantees in IEEE 802.11e Wireless LANs. *ITC Specialist on Providing QoS in Heterogeneous Environments Seminar, Berlin, Germany*, September 2003.

[19] A. Ksentini, A. Guéroui, and M. Naimi. Adaptive Transmission Opportunity with Admission Control for IEEE 802.11e Networks. *ACM/IEEE MSWIM 2005, Montreal, Quebec, Canada*, October 2005.

[20] Haitao Wu, Yang Peng, Keping Long, Shiduan Cheng, and Jian Ma. Performance of Reliable Transport Protocol over IEEE 802.11 Wireless LAN: Analysis and Enhancement. *IEEE INFOCOM 2002, New York, USA*, June 2002.

[21] Paal E. Engelstad and Olav N. Østerbø. Non-Saturation and Saturation Analysis of IEEE 802.11e EDCF with Starvation Prediction. In *ACM/IEEE MSWIM 2005, Montreal, Quebec, Canada*, November 2005.

[22] Gilbert (Gang) Chen. Component Oriented Simulation Toolkit. *http://www.cs.rpi.edu/~cheng3/*, 2004.

[23] B. Bellalta, C. Cano, M. Oliver, and M. Meo. Modeling the IEEE 802.11e EDCA for MAC parameter optimization. *Het-Nets'06, IlKley, UK*, September 2006.

[24] B. Bellalta, M. Meo, and M. Oliver. Comprehensive Analytical Models to Evaluate the TCP Performance in 802.11 WLANs. *4th Wired/Wireless Internet Communications, Bern, Switzerland*, May 2006.

[25] Tejinder S. Randhawa and R.H.S Hardy. Performance Analysis of Multi-Service Cellular Networks with Mobile Users. In *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), Chicago, IL, USA*, September 23-28 2000.

Boris Bellalta received his M.Sc degree in telecommunications from the Technical University of Catalonia in 2002. Since 2003, he has been working towards his Ph.D. in Telecommunications Engineering at Pompeu Fabra University, working in flow-based CAC mechanisms in Multiservice Wireless Networks. His other research activities include Internet Traffic Modeling and MAC/PHY cross-layer design. Since 2003, he is an Assistant Professor at Pompeu Fabra University.

Michela Meo received the Laurea degree in Electronic Engineering in 1993, and the Ph.D. degree in Electronic and Telecommunication Engineering in 1997, both from Politecnico di Torino. Since November 1999, she is an Assistant Professor at Politecnico di Torino. She coauthored more than seventy papers and two book chapters. She edited six special issues of international journals. She was in the organization of many international workshops and conferences; she also acted as program and general chair of ACM MSWiM (International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems) and program chair of IEEE QoS-IP (International Workshop on QoS in Multiservice IP Networks). Her research interests are in the field of analysis and dimensioning of cellular networks, performance evaluation of transport and link layer protocols, traffic characterization, inter-vehicular networking.

Miquel Oliver received his PhD from the Polythecnic University of Catalunya (Spain) in 1999, and a Telecommuncations Engineering degree in 1994 also from Polythecnic University of Catalunya. He joined the Wireless Information Network Laboratory (WINLAB), at Rutgers University in 1999. Is associated professor from 2001, first at Polythecnic University of Catalunya, and nowadays at Universitat Pompeu Fabra (UPF) in Barcelona (Spain). He has been dean of the Telecommunications Engineering School at UPF. Dr. Oliver is the leader of the Research Group on Networking Technology and Strategies (NeTS), member of the Scientific Committee of the Advanced Internet Foundation i2CAT, and have been involved in different national and international research projects.