

A Model-based Admission Control for IEEE 802.11e Networks

Rosario G. Garroppo, Stefano Giordano, Stefano Lucetti, Luca Tavanti

Dept. of Information Engineering

University of Pisa

Pisa, Italy, I-56122

Email: {name.surname}@iet.unipi.it

Abstract—The recently approved IEEE 802.11e standard improves the support for real-time multimedia applications in Wireless LANs. Yet, to provide some level of guarantee in terms of QoS parameters, it is essential to use an admission control scheme. We have devised one such scheme, which bases the admissibility test on the time occupancy of the medium. The network state is assessed through an analytical model of the EDCA in non-saturation conditions. In the paper, beyond describing the proposed scheme, we also show its effectiveness in the case of voice over IP applications.

I. INTRODUCTION

Wireless LANs based on the IEEE 802.11 standard are by now a solid and widely deployed technology. They provide users an easy and flexible way to access the Internet, offering bit rates up to 54 Mbps. This simple technology, well suited to support data services, presents several hurdles for multimedia applications. This drawback, still negligible until a few years ago, is becoming more and more substantial. The demand for (real time) voice and video services over the Internet is rapidly increasing, and consequently the convergence of wireless LANs and Multimedia over IP (MoIP) applications is one of today's key research (and business) topic.

The same IEEE recognized this trend and faced it by starting the Task Group E. The recently approved 802.11e amendment enhances the standard with traffic differentiation capabilities [1]. With the most popular EDCA mode of operation¹, voice and video frames are given priority in accessing the medium, but no strict guarantee is given in terms of QoS parameters. The benefits brought to MoIP services are thus effective only as long as the number of users in the network is kept below a given threshold [2][3]. In fact, 802.11e provides for the use of admission control (a.c.), but it leaves the definition and implementation of the algorithm to the various manufacturers.

In this paper we present an admission control scheme for users of MoIP services in IEEE 802.11e networks. The scheme we propose operates on the basis of input parameters retrieved by an underlying analytical model of EDCA. Instead of developing our own model, we build on the one described in [4]. One interesting feature of this model is that it also accounts for the non-saturation condition, thus being closer to a real scenario. Most previous works are instead based on saturation models, and this has been pointed out to be one of

their main drawbacks (see e.g. [5] for a nice review of the related literature).

The analytical model is used to extract the probability of unsuccessful transmission, which is then used by our algorithm to estimate whether accepting a new user will cause an unbearable degradation of the quality of the ongoing communications. This assessment is performed on the exclusive basis of the temporal occupancy of the medium. We test whether, in a generic reference period, there will be time for all users to transmit their frames. We will show that this simple test, which does not involve any complex computation on the traditional QoS parameters (e.g. bandwidth, delay, delay jitter, packet loss), is sufficiently accurate to guarantee a satisfactory network performance.

The most interesting previous work on this topic is the one by Chen et al. [3]. The authors propose two call a.c. schemes that relies on the average delay estimates and the channel busyness ratio, defined as the portion of the time that the channel is busy in an observation period. An analytical model is built to derive the average delay estimate for traffic of different priorities in an unsaturated network. When deciding on the acceptance of a new real time flow, the a.c. algorithm considers its effect on the channel utilization and the delay of existing flows. The proposed G/M/1 and G/G/1 models deliver a rough upper bound for the average delay, which becomes looser as the number of flows increases. As a consequence, the proposed a.c. schemes can only suggest a pessimistic limit on the number of admissible users for small-size networks. Yet, they prove the effectiveness of using a time based metric for regulating the access.

The paper is organized as follows. In the next Section we accurately describe the proposed a.c. scheme. Then we expose a short overview of the reference analytical model. Application of the a.c. scheme to a practical case is reported in Section IV. Finally, we present some simulation results to confirm the effectiveness of our method.

II. THE ADMISSION CONTROL ALGORITHM

Let us indicate with T_{ref} an arbitrary reference period. The basic principle of operation of the proposed algorithm is to verify whether in a T_{ref} the medium occupancy time T_{occ} of the offered traffic (including collisions, retransmissions, etc.) keeps below T_{ref} itself. Upon the admission request of a

¹The first commercial products do not implement HCCA.

new flow, the algorithm evaluates T_{occ} in the hypothesis of acceptance of the incoming flow. The flow is actually admitted to the service only if $T_{occ} \leq T_{ref}$.

T_{occ} is the key element of our algorithm. Its calculation bases on the following considerations. During a T_{ref} , the j^{th} flow offers frames at rate λ_j , for a total number of offered frames $\lambda_j \cdot T_{ref}$. The transmission of each of these frames occupies the channel for a time whose mean $E[T_j]$ can be analytically evaluated. $E[T_j]$ will also include all the overhead related to backoff, retries, etc. Thus the a.c. scheme can compute T_{occ} simply as:

$$T_{occ} = \sum_{j=1}^M \lambda_j T_{ref} E[T_j], \quad (1)$$

with M being the total number of flows².

Estimation of sufficiently realistic and accurate $E[T_j]$ is the key point of the a.c. scheme. In particular, we are interested in measuring $E[T_j]$ from the channel point of view. Not always does this correspond to the time as seen by each single station, and consequently computing T_{occ} as the sum of all the individual station-measured contributions will not be correct. Specifically, it will result in a gross overestimate of T_{occ} . This happens because there is no difference, from the channel point of view, whether the medium is occupied by one or more than one station, since the channel sees this event just as “the medium is busy”.

For example, if we consider a collision, two (or more) stations do actually transmit their frame at the same time, hence occupying the channel for virtually half (or one third, fourth, etc.) of the time each. Similarly, when two or more stations are contending for the channel, the time spent in backing off is a shared virtual “occupancy” of the medium, as the backoff counter is decremented for all the contending stations.

To account for this point, the number of stations that cause those events should be included in the computation of $E[T_j]$. We actually do that by means of two parameters, α and β . We indicate with α the mean number of colliding stations in a generic time slot, given that a collision occurred, and with β the average number of stations competing to access the medium to transmit a queued frame. Both α and β can be estimated using the EDCA model developed in [4]; an example is shown in Section IV.

It is now possible to define a model, in Fig. 1, to evaluate $E[T_j]$. Let us introduce the index i , which designates the Access Category (AC) to which the j^{th} flow is mapped, and let p_i be the probability of unsuccessful transmission for a frame belonging to AC_i . As depicted in Fig. 1, $E[T_j]$ depends on p_i and the values $E[T_j(k)]$, which are the mean medium occupancy time of a frame transmission (of flow j and AC i) that requires exactly k retransmissions³.

²Let us sort the flows so that values of j from 1 to $M - 1$ index existing flows, and $j = M$ indexes the incoming flow.

³To lighten the notation, we omit the index i in $E[T_j]$ and $E[T_j(k)]$; yet their dependence on i is implicit in the mapping of the flow j to AC_i .

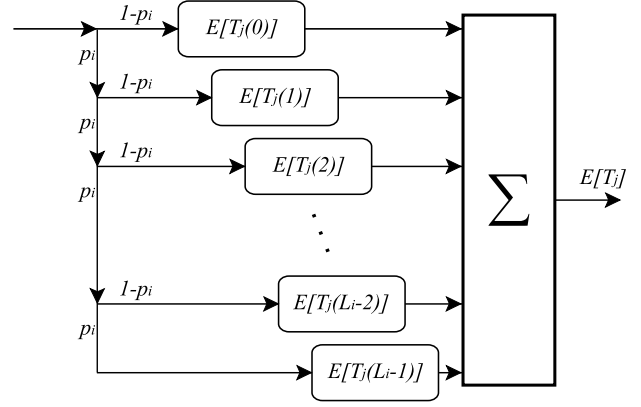


Fig. 1. Model for the evaluation of $E[T_j]$.

The terms $E[T_j(k)]$ can be evaluated making reference to the single transmissions. In this regard, we recall that a frame transmission attempt is composed of a sequence of periods. There is at first an inter-frame space (IFS), then the proper frame transmission (including PHY and MAC headers), possibly preceded by a backoff interval (e.g. due to contention). In case of correct reception, there is a SIFS followed by an ACK; in case of collision⁴, the medium can be assumed to be idle immediately after the end of the longest frame. In both cases, a backoff procedure concludes the transmission cycle.

In the following, $B_{i,k}$ is the average backoff timer of AC_i at the k^{th} backoff stage. In particular, $B_{i,k}$ can be expressed as $T_{slot} W_{i,k} / 2$, being T_{slot} the basic IEEE 802.11 time slot and $W_{i,k}$ the contention window (for AC_i at the k^{th} backoff stage). $E[T_j(k)]$ is obtained from the knowledge of the mean medium occupancy time in case of a successful transmission $E[T_{succ,j}(k)]$ and in case of collision $E[T_{coll,j}(k)]$, whose expressions are:

$$\begin{aligned} E[T_{succ,j}(k)] &= AIFS[i] + T_{PHY} + T_{MAC} + T_{DATA,j} \\ &\quad + SIFS + T_{ACK} + \frac{B_{i,0}}{\beta} \\ E[T_{coll,j}(k)] &= AIFS[i] + T_{PHY} + T_{MAC} + T_{DATA}^* \\ &\quad + T_{ACK_Timeout} + \frac{B_{i,k+1}}{\beta} \end{aligned}$$

Here, $AIFS[i]$ is the Arbitration IFS of AC_i , T_{PHY} and T_{MAC} are the durations of the physical and MAC headers, $T_{DATA,j}$ is the time to transmit the payload (MSDU) of flow j , T_{ACK} is the time to transmit the ACK, and $T_{ACK_Timeout}$ is the timeout for the reception of the ACK. In particular, $T_{DATA,j}$ can be expressed as D_j / R , where D_j is the MSDU size and R the transmission rate, assumed to be constant in the absence of rate adaptation algorithms. In case of collision, we should consider the time related to the longest collided data frame (T_{DATA}^*). As already outlined, β accounts for the number of stations that are doing backoff.

⁴We assume that collisions are the sole source of transmission errors.

$E[T_j(k)]$ can then be evaluated for $k = 0 \dots L_i - 1$ (being L_i the retry limit for AC i) according to the following expressions:

$$\begin{aligned} E[T_j(0)] &= E[T_{succ,j}(0)] \\ E[T_j(1)] &= \frac{E[T_{coll,j}(0)]}{\alpha} + E[T_{succ,j}(1)] \\ E[T_j(2)] &= \frac{E[T_{coll,j}(0)] + E[T_{coll,j}(1)]}{\alpha} + E[T_{succ,j}(2)] \\ &\dots \\ E[T_j(L_i - 1)] &= \frac{\sum_{k=0}^{L_i-2} E[T_{coll,j}(k)]}{\alpha} + \\ &+ p_i \frac{E[T_{coll,j}(L_i - 1)]}{\alpha} + (1 - p_i)E[T_{succ,j}(L_i - 1)] \end{aligned}$$

Given these and the scheme in Fig. 1, it is finally easy to obtain $E[T_j]$:

$$E[T_j] = \frac{p_i B_{i,0}}{\beta} + \sum_{k=0}^{L_i-1} (1 - p_i)^{1-\delta(k-L_i+1)} p_i^k E[T_j(k)], \quad (2)$$

where $\delta(k)$ is 1 if $k = L_i - 1$ and 0 otherwise. The term $B_{i,0}/\beta$ accounts for the extra backoff that must be performed when a frame arrives at an idle node and the medium is busy. As explained in [4], the probability of this event is again p_i .

III. THE REFERENCE MODEL

The probability p_i can be computed using any analytical model of the 802.11e EDCA. Among the many existing, we have chosen the model proposed by Engelstad and Østerbø in [4]. This work basically improves [6], which in turn extended [7] to the 802.11e. Though it may present some shortcomings, nevertheless it offers several advantages. One major feature is that it accounts for a non-saturated channel. This makes the model much closer to reality. While having the channel the most utilized as possible may increase the throughput, it is also true that a saturated network based on a distributed access control (DCF and EDCA) is not able to transport real-time data with any satisfactory level of QoS. This is a consequence of the increased collision probability, which leads to large frame delivery delay and delay jitter, making many frames useless for the destination application (see e.g. [2], [3], [5], and [10]).

Engelstad-Østerbø's model introduces the utilization factor ρ_i , which is tied to the probability that there is a frame in the transmission queue of AC $_i$ at the time of a completed transmission. ρ_i can thus be used to account for the non-saturated network. The authors build a Markov chain in which the states are identified by the AC, the retransmission attempt, the backoff stage and the state of the transmission queue (either empty or not). The chain for the i^{th} AC is reported in Fig. 2. Without entering into the many details of the solution of the chain (see [4] and also [8]), we just report the final results.

At each station, the probability τ_i of a transmission attempt in a generic time slot for the i^{th} AC is given by:

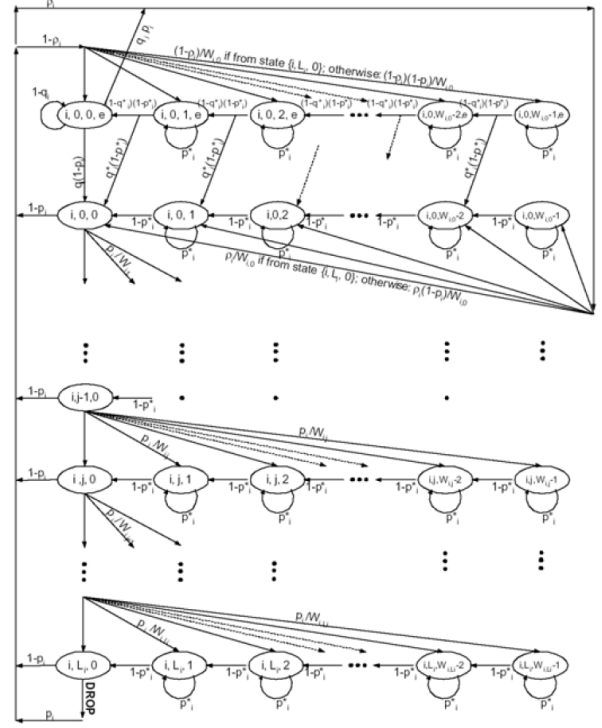


Fig. 2. The Markov chain of the Engelstad and Østerbø model.

$$\begin{aligned} \frac{1}{\tau_i} &= \frac{1 - 2p_i^*}{2(1 - p_i^*)} + \frac{W_{i,0}(1 - p_i)(1 - (2p_i)^{m_i})}{2(1 - p_i^*)(1 - 2p_i)(1 - p_i^{L_i+1})} \\ &+ \frac{1 - p_i}{1 - p_i^{L_i+1}} \frac{1 - \rho_i}{q_i} \left(1 + \frac{(W_{i,0} - 1)p_i q_i}{2(1 - p_i^*)}\right) \\ &+ W_{i,0} \frac{(2p_i)^{m_i}(1 - p_i^{L_i-m_i+1})}{2(1 - p_i^*)(1 - p_i^{L_i+1})} \end{aligned} \quad (3)$$

It can be noted that the transmission probability depends on several parameters. Beyond the already defined ρ_i and p_i , p_i^* is the probability that the backoff counter is not decremented (i.e. the channel is sensed busy), q_i is the probability that at least one frame arrives in the transmission queue during the following time slot under the condition that the queue is empty, q_i^* is the probability that a frame arrives while the backoff is frozen, $W_{i,0}$ is the initial contention window, and m_i is the backoff stage at which the contention window has reached its maximum. The probabilities can be evaluated with the following expressions:

$$p_b = 1 - \prod_{i=0}^{C-1} (1 - \tau_i)^{n_i} \quad (4)$$

$$p_s = \sum_{i=0}^{C-1} n_i (1 - p_i) \tau_i \quad (5)$$

$$p_i = 1 - \frac{1 - p_b}{\prod_{c=0}^i (1 - \tau_c)} \quad (6)$$

$$p_i^* = \min(1, p_i + \frac{A_i p_b}{1 - \tau_i}) \quad (7)$$

$$q_i = 1 - (p_s e^{-\lambda_i T_s} + (1 - p_b) e^{-\lambda_i T_e} + (p_b - p_s) e^{-\lambda_i T_c}) \quad (8)$$

$$\rho_i = \lambda_i \bar{s}_i \quad (9)$$

In the formulas: n_i is the number of stations contending for the channel for each AC; C is the total number of ACs (usually four); $A_i = AIFSN[i] - AIFSN[C]^5$; T_e , T_s and T_c denote respectively the real duration of an empty slot, of a slot containing a successfully transmitted frame and of a slot containing two or more colliding frames; p_b represents the probability that the channel is busy; p_s is the probability that a time slot contains a successful transmission. Eq. (6) refers to the case of virtual collision handling; however, if we assume that no more than one flow is generated at each station, a simpler form may be found. Furthermore, we can assume $q_i^* = q_i$ (according to [4] this is a good approximation), where q_i has been computed in the hypothesis of Poissonian traffic. Finally, the queue utilization factor ρ_i depends on the mean frame service time \bar{s}_i once it has reached the front of the transmission queue. The derivation of \bar{s}_i is rather complex, and therefore it has been skipped; the interested reader can refer to [8].

IV. ADMISSION CONTROL OF VOIP CALLS

As outlined in the Introduction, there is an increasing demand for voice services over wireless LANs. Therefore it seems natural to apply the proposed scheme to this kind of scenario. We consider a network with a variable number of associated stations, each sustaining a bidirectional VoIP call with a corresponding peer on the wired network. An 802.11e Access Point is the gateway between the wired and the wireless worlds. Voice is the only traffic offered to the network, and all stations use the same constant bit rate (CBR) codec.

Application of the a.c. scheme in this context allows a number of simplifications. All the flows have identical features (same frame rate, payload size, etc.) and belong to the same AC, hence index i can be removed. The direct consequence of this is that $p^* = p$, since A_i becomes zero. Then, there is only one flow per station, so that the internal collisions are eliminated and the series in (4), (5), and (6) can be reduced to only one term. So (4)-(7) are now (n_i is replaced by the total number of stations N):

$$p_b = 1 - (1 - \tau)^N \quad (10)$$

$$p_s = N\tau(1 - p) \quad (11)$$

$$p = p^* = 1 - (1 - \tau)^{N-1} \quad (12)$$

⁵Let's order the ACs so that AC_0 has the lowest priority and AC_C the highest. $AIFSN[C]$ is thus the smallest AIFSN.

Equation (8) can be computed exploiting the CBR nature of the traffic. In detail, the terms $e^{-\lambda t}$, which denote the probability that the inter-arrival time is greater than t for a Poisson process, can be replaced by their equivalents for a process with periodic arrivals. We can reasonably assume that the average T_s , T_e and T_c are all smaller than T_{frame} (otherwise there will not be room even for a single flow) and that these events occur with the same probability across the whole frame arrival period (the probability distribution is therefore uniform). Hence:

$$q = q^* = 1 - \frac{p_s T_s + (1 - p_b) T_e + (p_b - p_s) T_c}{T_{frame}} \quad (13)$$

Finally:

$$\rho = \lambda \bar{s} \quad (14)$$

It is worth noting that \bar{s} can be computed using (2) where α and β have been removed (from the definition, \bar{s} measures the service time at each single station, hence it does not depend on medium sharing parameters).

Together with (2) and (3), (10)-(14) form a system of equations that can be solved numerically, once the values for the voice traffic AC have been substituted. The value of $E[T]$ is then fed to (1) to allow for the verification of the admissibility of the new call.

The last step to complete the computation of T_{occ} is finding an estimate of α and β . α can be computed using the probability that k stations transmit given that a collision occurs (i.e. $k \geq 2$). Indicating with P_k the probability that in a generic time slot there are exactly k transmitting stations (and $N - k$ silent stations), we get:

$$\alpha = \frac{\sum_{k=2}^N k P_k}{\sum_{k=2}^N P_k} = \frac{\sum_{k=2}^N k \binom{N}{k} \tau^k (1 - \tau)^{N-k}}{1 - (1 - \tau)^N - p_s} \quad (15)$$

Obviously, α is a function of the transmission probability τ . As for β , the average number of stations competing to transmit can be derived directly from the utilization factor ρ :

$$\beta = \rho N. \quad (16)$$

V. SIMULATION RESULTS

To verify the accuracy of the proposed a.c. scheme, we present some results obtained using the simple scenario described in the preceding Section. As a further step, we can choose $T_{ref} = T_{frame}$. In this way, the criterion for the a.c. algorithm simply becomes a verification that each flow has the time to transmit one frame in each frame generation period of the codec.

All flows were mapped to the highest priority AC, namely AC_VO. This category specifies the following parameters: $W_0 = 7$, $m = 1$, $L = 4$. The number of active calls was increased from one to system saturation, where no QoS guarantee is possible. Note that admitting a call requires that two flows (one in the uplink and one in the downlink) are admitted.

Two codecs were used in two different series of tests. The first is the ITU-T G.723.1, which presents a codec framing time T_{frame} of 30 ms and a payload D of 192 bit, leading to a net bit rate of 6.3 Kbps and an IP throughput of 17.2 Kbps. The other codec is the ITU-T G.729: it has $T_{frame} = 20$ ms and $D = 160$ bit, thus offering a net bit rate of 8 Kbps and an IP throughput of about 24 Kbps. In both cases no silence suppression is used, hence all sources generate data at constant bit rate (CBR).

To verify whether the a.c. scheme provides an accurate prediction of the maximum number of allowable calls with a satisfactory speech quality, we set up our simulations in order to evaluate the R-factor for each call [9]. Following the approach in [10], the R-factor can be used as an indicator of the perceived voice quality. It is based on a number of additive factors that depend, among other things, on codec and network performance metrics such as end-to-end delay and packet loss. It assumes values in the range 0-100, with values above 70 indicating an acceptable quality.

We performed our simulations using OPNET Modeler [11]. In each simulation run we varied the number of calls in the network and measured the R-factor of the worst call. The results have been averaged over ten runs. Tables I and II reports the results for the G.729 and G.723.1 codecs. n is the number of calls; R is the R-factor, computed from the simulation data following the rules given in [9]; $E[T]$ and T_{occ} (both in ms) have been derived using the procedures described in Sections II-IV.

TABLE I
SIMULATION AND A.C. RESULTS FOR G.729

$(T_{frame} = 20ms)$			
n	R	$E[T]$	T_{occ}
11	83.2	0.849	18.68
12	82.9	0.838	20.11
13	12.0	0.829	21.55

TABLE II
SIMULATION AND A.C. RESULTS FOR G.723.1

$(T_{frame} = 30ms)$			
n	R	$E[T]$	T_{occ}
16	78.2	0.860	27.50
17	74.9	0.851	28.91
18	23.7	0.842	30.32

The R-factor in Table I suggests that up to twelve G.729-based calls can be admitted with a satisfactory level of service. The model instead would have rejected the twelfth call, as T_{occ} results greater than T_{frame} (20 ms). However, the error is very small (less than 1%), and can be ascribed to the approximations necessary to make the model manageable.

An even better result has been registered with G.723.1 (see Table II). The a.c. scheme agrees with the results of the simulation, accepting the 17th call, for which the R-factor is still above the target level, and rejecting the 18th.

As a final remark, it is worth highlighting that the a.c. decision has been driven by simple average delay estimates.

The achieved results are nevertheless quite accurate. Work is however ongoing to investigate the improvement that a finer analysis of the delay distribution can bring to the a.c. scheme.

VI. CONCLUSION

We have presented an admission control scheme that tests the acceptance of a new flow on the basis of the mean time occupancy of the medium. The key points of our scheme are the assessment of this time, which is computed as seen by the channel rather than from the single stations, and the use of an analytical model of the 802.11e EDCA mode in non-saturation conditions.

The proposed scheme has been applied and evaluated in a specific case, the increasingly important scenario of VoIP services over wireless LANs. The outcome of the simulations has proved its efficiency.

ACKNOWLEDGMENT

This work was supported by the TWELVE Project, funded by the Italian MIUR Ministry by means of the COFIN research program.

The authors would like to thank Saverio Petrunaro for his help in building and running the simulations.

REFERENCES

- [1] IEEE Std 802.11e-2005, Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements, September 2005.
- [2] R. G. Garroppo, S. Giordano, S. Lucetti, L. Tavanti, "Admission Region of Multimedia Services for EDCA in IEEE 802.11e Access Networks", G. Kotsis and O. Spaniol (Eds.), *Lecture Notes in Computer Science*, Volume 3427, pp. 105-120.
- [3] X. Chen, H. Zhai, X. Tian, Y. Fang, "Supporting QoS in IEEE 802.11e wireless LANs", *IEEE Transactions on Wireless Communications*, Volume 5, Issue 8, August 2006, pp 2217-2227.
- [4] P. E. Engelstad, O. N. Østerbø, "Delay and Throughput Analysis of IEEE 802.11e EDCA with Starvation Prediction", *Proc. of the 30th IEEE Conference on Local Computer Networks*, Nov. 2005, pp 647-655.
- [5] D. Gao, J. Cai, K. N. Ngan, "Admission Control in IEEE 802.11e Wireless LANs", *IEEE Network*, Volume 19, Issue 4, July 2005, pp. 6-13.
- [6] Y. Xiao, "Performance analysis of IEEE 802.11e EDCF under saturation condition", *Proc. of the IEEE International Conference on Communications*, June 2004, pp 170-174.
- [7] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function", *IEEE Journal on Selected Areas in Communications*, Volume 18, Issue 3, March 2000, pp 535-547.
- [8] P. E. Engelstad, O. N. Østerbø, "The Delay Distribution of IEEE 802.11e EDCA and 802.11 DCF", *Proc. of the 25th IEEE International Performance, Computing, and Communications Conference*, April 2006, pp 87-96.
- [9] ITU-T Recommendation G.107, "The E-model, a computational model for use in transmission planning", March 2003.
- [10] R. G. Garroppo, S. Giordano, S. Lucetti, L. Tavanti, "A Measurement Based Channel Aware Scheduler to Lessen Voip Capacity Degradation In 802.11 Networks", *Proc. of the IEEE International Conference on Communications*, June 2006.
- [11] OPNET Modeler, available at: <http://www.opnet.com/products/modeler>.