QUALITY-OF-SERVICE PROVISIONING SYSTEM FOR MULTIMEDIA TRANSMISSION IN IEEE 802.11 WIRELESS LANS

Joseph Deng and Hsu-Chun Yen

Department of Electrical Engineering
National Taiwan University, Taipei, Taiwan, R.O.C.

Abstract

In this paper, we propose a polling with non-preemptive priority based access control scheme for the IEEE 802.11 protocol. Under such a scheme, modifying the DCF access method in the contention period supports multiple levels of priorities such that user mobility can be supported in wireless LANs. The proposed transmit-permission policy and adaptive bandwidth allocation scheme derive sufficient conditions such that all the time-bounded traffic sources satisfy their time constraints to provide various QoS (quality-of-service) guarantees in the contention free period while maintaining efficient bandwidth utilization at the same time. In addition, our proposed scheme is provably optimal for voice traffic, minimizing the average waiting time for voice over IP packets. In addition to theoretical analysis, simulations are conducted to evaluate the performance of the proposed scheme. As it turns out, our design indeed provides a good performance in the IEEE 802.11 wireless LANs environment.

Key Words
WLAN, QoS, Priority, VoIP

1. Introduction

In WLANs, the *medium access control* (*MAC*) protocol is the key component that provides the efficiency in sharing the common radio channel while satisfying the quality-of-service (QoS) requirements of various multimedia traffic. However, frames in DCF, the basic access method in the IEEE 802.11 MAC layer protocol [1], do not have priorities, and there is no other mechanism to enforce a guaranteed access delay bound. As a result, real-time applications such as voice or live video transmissions may suffer from unacceptable delay with this protocol. The second access mode of the IEEE 802.11 MAC layer protocol, Point Coordination Function (PCF), offers a "packet-switched connection-oriented" service, which is well suited for real-time traffic. However, in order to poll the stations an AP (access point) must maintain a polling list, which is implementation dependent. What this means is that end-to-end QoS requirements still cannot be satisfied in this scheme since it does not include any access control policy. It does not include any priority scheme to support user mobility, nor does it apply any bandwidth allocation strategy for handoff calls. Since the demand for the use of packet-switched techniques for transferring delay-sensitive data in wireless environments is inevitable for multimedia applications, several works have been investigated and discussed along this line of research [2]-[4]. Although the 802.11e standard is for providing QoS support for WLAN applications such as voice over wireless IP, the process of creating a definitive standard might be too slow for us waiting for it to be ratified. In this paper, we propose an advanced, pragmatic, and yet more complete polling with non-preemptive priority based access control scheme for the IEEE 802.11 protocol. Under such a scheme, by modifying the DCF access method in the contention period, the protocol offers multiple levels of priorities such that user mobility can be supported in wireless LANs. Besides, the proposed transmit-permission policy and adaptive bandwidth allocation scheme not only separate admitted inactivated users from newly requesting access users, but also derive sufficient conditions such that all the time-bounded traffic sources satisfy their time constraints to provide various QoS guarantees in the contention free period while maintaining efficient bandwidth utilization at the same time. Furthermore, the proposed scheduling algorithm for voice traffic is provably optimal in that it gives the minimum average waiting time for voice over IP packets. The proposed scheme is performed at each AP in a distributed manner. Such a scheme can be implemented in a broad class of algorithms with relatively minor modifications. The performance of our design for integrated traffic is examined in detail. As it turns out, our design indeed provides good performance improvements over the original IEEE 802.11 protocol.

2. The Proposed Scheme

Cross-strait trade between Taiwan and China has increased greatly since the end of year 1987 when Taiwanese government began to gradually loosen restrictions on Taiwan's indirect trade with China. However, the large amount of expenditure on communications conducted between the Taiwan-based headquarters and the branches in China costs the Taiwanese investors a lot of money. Hence, we implement a Quality-of-Service provisioning system for multimedia transmission in IEEE 802.11 Wireless LANs, with a view to cutting, to a great extent, the telecommunication costs incurred to cross-strait enterprises. In this section, we present the proposed scheme in detail. Figure 1 shows an overview of the proposed system architecture. Our method involves three basic components: (a) *a priority enforcement mechanism for request access*, (b) *a packet transmit-permission policy*, and (c) *an adaptive bandwidth management strategy*.

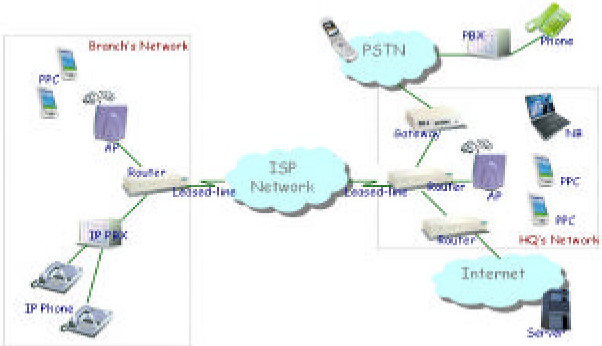

Figure 1. System Architecture.

• Priority Enforcement Mechanism for Request Access

In this section, we propose a novel method to modify the DCF access method to get many levels of priorities, capable of giving handoff requests higher priority over new connection requests. It should be noted that our approach substantially extends the idea originated in [3], in which a preliminary version of priority enforcement mechanism was proposed.

The basic idea behind our method is that prioritized access to the wireless medium is controlled through different backoff time periods. To this end, instead of using the one defined in [1], we change the backoff time generation function to $\lfloor \mathrm{ranf}() \bullet 2^{m+i} \rfloor + k \bullet 2^{n+i}$, where k is the level of priority, and m and n are the parameters used to decide the number of slots in individual priority levels and the number of slots between each priority levels, respectively. In this paper, the real-time handoff traffic requests have the highest priority among all other requests, and the second priority class is the admitted inactivated video traffic. The new requests and pure data traffic will reside on the lowest priority level, as illustrated in Table I.

As mentioned earlier, our scheme has the ability to expand or contract the backoff range arbitrarily by changing the parameters k, m and n. Hence, we propose an adaptive contention window mechanism for our scheme to dynamically expand and contract the contention window size according to the current load and achieve the theoretical capacity limits. This scheme is based on the results of the capacity analysis model of the IEEE 802.11 protocol originally proposed in [5]-[7] as well as the concept introduced in [8]. However, our scheme is simpler and more efficient and accurate, and it does not suffer from the problem of harmful fluctuation reported in [9]

Table I. Examples of backoff time of individual traffic

Consecutive time (i) / Backoff slot numbers / Types of requests (k, m, n)	1^{st}	2^{nd}	3^{rd}	4^{th}
Real-time handoff traffic (0, 1, 1)	0-3	0-7	0-15	0-31
Admitted inactivated video traffic (1, 1, 1)	4-7	8-15	16-31	32-63
Non-real-time/New request/Data traffic (2, 2, 1)	8-15	16-31	32-63	64-127

As reported in [6], the collision probability is low if the stations utilize the optimal p value, i.e., p_{opt}, meaning that $p\{N_{tr} > 1 \mid N_{tr} > 0\} << p\{N_{tr} = 1 \mid N_{tr} > 0\}$. As a consequence, $M \cdot p_{opt}$ is a tight upper bound of a in a system operating with the optimal channel utilization level, hence, $M \cdot p_{opt}$ can be used as a measure of the network contention level when the network utilizes the optimal contention window size corresponding to the ongoing network and traffic configuration. In other words, if the utilization factor, a, is bounded above by the constant, $M \cdot p_{opt}$, the system will be kept in a better situation from the viewpoint of channel utilization since the lower rate of collisions will result in a significant increase in throughput and decrease in mean access delay. Thus, when the utilization factor, a, exceeds the value of $M \cdot p_{opt}$, we can expand the contention window by allocating more slots than the old contention window had. On the other hand, it is also desirable to contract the contention window when the utilization factor becomes too small, so that the wasted bandwidth is not exorbitant.

• The Packet Transmit-permission Policy

Serving for the purpose of deciding whether a network accepts a new connection or not, the design of a packet transmission policy is one of the important challenges of traffic control in wireless networks. The policy is also used by the AP to determine which mobile gets permission to transmit a packet, especially in the realm of providing QoS. In this section, we propose a packet transmit-permission policy for IEEE 802.11 protocol to support integrated multimedia traffic. Our scheme is an enhanced version of the transmitting policy originally proposed in [10], and substantially extended from [11].

At the IP layer, the MTU (maximum transmission unit) is set to be 1500 bytes, which is the maximum MSDU (the packet delivered to the MAC layer by the higher layer) size for the 100basedT Ethernet. Similarly, the 802.11 standard also provides a fragmentation mechanism, which allows the MAC layer to split an MSDU into more MPDUs (packets delivered by the MAC layer to the PHY layer). Hence, to formalize our problem, we assume that all real-time traffic packets have the same size in this paper. Besides, two types of real-time traffic are considered. The first is voice traffic which is characterized by two parameters (r_c , d), where r_c is the rate of the source and d is the maximum tolerable jitter (packet delay variation) for this stream. The second is video traffic which is characterized by three parameters (r_v , b , d), where r_v is the average rate of the source, b is the maximum burstiness of the source, and d is the maximum tolerable delay (packet transfer delay) for this stream.

In one Basic Service Area (BSA) of IEEE 802.11 infrastructure network architecture, the AP implements a token buffer for each real-time source. In token buffers for voice sources, the smaller the rate of voice source, r_c , is, the higher the priority is. In token buffers for video sources, the priority is assigned in a similar way. That is, the one with the smallest maximum delay constraint has the highest priority among all video sources. We depict the packet transmit-permission policy in Fig. 2. In order to gain control of the medium, the AP performs the function of the point coordinator by transmitting a beacon frame at the beginning of the CFP after sensing the medium to be idle for a PIFS period. Once the AP has the control of the medium, it performs the following tasks.

1) The AP first scans the token buffers of voice sources according to the preset priority order. If a token is found, it removes one from this token buffer and polls this voice terminal. On receiving a poll the station transmits its packet after a SIFS interval. Then, the AP generates the next token for this voice source after $\frac{1}{r_c} - (PIFS + 2 \cdot SIFS + CFPoll + t_p + ACK)$ second if the piggyback was set while transmitting the packet, where t_p is the time to transmit a real-time traffic packet.
2) If no tokens are found in the token buffers of voice sources, the AP continues to scan the token buffers for video sources according to the preset priority order. If a token is found, it polls this video source. And it will not remove the token if the piggyback was set while this video source transmit it packet. If the piggyback was not set and it is not the last packet (End-of-File) either, the AP removes the token, and then generates the next token for this video source after h seconds if there is no new token generated for this video source within h , where h will be defined later.
3) If there is no token found in all token buffers, the AP will not know which, if any, of the mobiles have packets to transmit, then, it can end the CFP by transmitting a CF-End frame, and, for assuring the time constraint of admitted real-time traffic, the Ap shall announce the beginning of the next CFP interval by observing the token buffer of highest priority among its polling list.

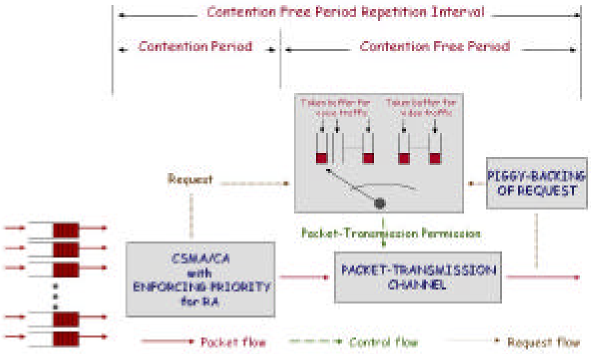

Figure 2. Proposed packet transmit-permission policy.

In the following theorems, we provide sufficient conditions for all the voice packets to satisfy their maximum jitter constraints (theorem 1) and for all the video packets to satisfy their maximum delay constraints (theorem 3), while optimizing the overall utilization of network bandwidth simultaneously (theorem 2).

Assume there are n_c voice sources (indexed by $i = 1,\ldots,n_c$), and n_v video sources (indexed by $j = 1,\ldots,n_v$). We denote (r_{ci} , d_i) as the traffic parameters of the i^{th} voice source, (r_{vj}, b_j, d_j) as the traffic parameters of the j^{th} video source, and p_i as the time needed for handoff for source i .

Theorem 1:

Let $\delta_1^* = \mathrm{PIFS} + 2 \cdot \mathrm{SIFS} + \mathrm{CFPoll} + t_p + \mathrm{ACK}$ and

$$\delta_i^* = \mathrm{PIFS} + 2 \cdot \mathrm{SIFS} + \mathrm{CFpoll} + t_p + \mathrm{ACK} + \sum_{k=1}^{i-1} \left\lceil \frac{r_{ck}}{r_{ci}} \right\rceil \cdot (\mathrm{PIFS} + 2 \cdot \mathrm{SIFS} + \mathrm{CFPoll} + t_p + \mathrm{ACK}) ,$$

$i = 2,\ldots,n_c$ and t_p be the time to transmit a packet. If $d_i^* < \frac{1}{r_{ci}}$ and $d_i^* \le d_i$ for all i = 1, 2, …, n_c , then all the packets generated by new-call voice sources meet their jitter constraints. Furthermore, if $d_i^* + p_i < \frac{1}{r_{ci}}$ and

$d_i^* + p_i \le d_i$ for the i^{th} source which is handoffed from other cells, then the packet generated by the i^{th} source after handoff meets its jitter constraint.

Theorem 2:

Suppose n_c voice sources are scheduled in the given priority order. The average waiting time is minimized if $r_{ci} \le r_{cj}$ for all $i < j$.

Theorem 3:

Let $\bar{b}_0 = (PIFS + 2\cdot SIFS + CFPoll + t_p + ACK)\cdot(n_c + 1)$,

$\bar{r}_{v0} = (PIFS + 2\cdot SIFS + CFPoll + t_p + ACK)\cdot\sum_{i=1}^{n_c} r_{ci}$,

$\bar{b}_j = (PIFS + 2\cdot SIFS + CFPoll + t_p + ACK)\cdot(b_j + 1)$,

$\bar{r}_{vj} = (PIFS + 2\cdot SIFS + CFPoll + t_p + ACK)\cdot r_{vj}$,

$$d_1^* = h_1 + \frac{\bar{b}_0 + \bar{b}_1}{1 - \bar{r}_{vo}},$$

and

$$d_j^* = \eta_j + \frac{\sum_{k=0}^{j}\bar{\beta}_k + (PIFS + 2\cdot SIFS + CFpoll + t_p + ACK)\cdot\sum_{k=1}^{j-1}(r_{vk}\cdot d_k^*)}{1 - \sum_{k=0}^{j-1}\bar{r}_{vk}}$$

, where $j = 2,\ldots,n_v$. If $\sum_{k=0}^{n_v}\bar{r}_{vk} \le 1$ and $d_j^* \le d_j$ for all j , then the delay constraints are satisfied for all the new-call video sources. Furthermore, if $d_j^* - h_j \le d_j - p_j$ for j^{th} source which is handoffed from other cells, then the packet generated by the j^{th} source after handoff meets its delay constraint.

• The Adaptive Bandwidth Management Strategy

Recall that user mobility (handoff) can be supported in our proposed method since a real-time handoff mobile might be given higher priority over new connection requests. As shown in Fig. 3, the total bandwidth is divided into three parts: channels I, II, and channel III. We allocate channel I for real-time traffic and channel II for handoff real-time traffic in contention free period. By allowing the handoff real-time traffic to use bandwidth exclusively with preemptive priority over other traffic in channel II, the handoff real-time traffic might has a larger share of pie in bandwidth utilization to reduce the dropping probability. Likewise, the remaining real-time traffic has precedence over the new request/data traffic for using network resources in channel I. Channel III is only reserved for new requests and data traffic to guarantee a minimum bandwidth for data traffic in the contention period. However, after bandwidth is allocated, network conditions may change. Therefore, the proposed strategy can also adjust the amount of allocated bandwidth based on the measured dropping probability, blocking probability, and bandwidth utilization. The algorithm to control the size of the allocated bandwidth is summarized in the following.

Function Adaptive_Bandwidth_Allocation

IF monitored dropping probability > threshold_D **THEN**
IF bandwidth utilization < μ **THEN**
size of allocated bandwidth II= min {max {size of allocated bandwidth I, size of allocated bandwidth II} × up_γ , total bandwidth }
ELSE
size of allocated bandwidth II= min {max {size of allocated bandwidth I, size of allocated bandwidth II} × up_γ , total bandwidth× threshold_channel_II_max }
ELSE
IF monitored blocking probability > threshold_B **THEN**
IF bandwidth utilization < μ **THEN**
size of allocated bandwidth I= min {size of allocated bandwidth I × up_ γ , total bandwidth × threshold_channel_I_max }
ELSE
size of allocated bandwidth I= min {size of allocated bandwidth I × up_ γ , total bandwidth× threshold_channel_I_medium }
ELSE
IF bandwidth utilization < μ **THEN**
size of allocated bandwidth II= max {size of allocated bandwidth II × down_ γ , total bandwidth× threshold_channel_II_min }
size of allocated bandwidth I= max {size of allocated bandwidth I × down_ γ , total bandwidth× threshold_channel_I_min }

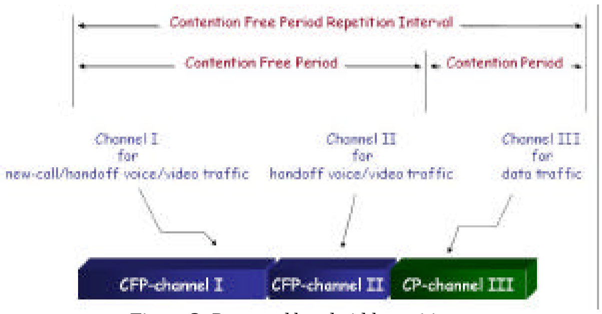

Figure 3. Proposed bandwidth partition.

3. Simulations and Performance Evaluation

In this section, we evaluate the performance of the proposed scheme.

• Simulation Environment

Our simulation model is built using the Simscript tool [12]. The model represents a BSS in the IEEE 802.11 WLANs with all stations in the BSS (Basic Service Set) capable of directly communicating with the remaining parties. Three types of traffic are considered in the simulation.

1. Pure data:
The arrival of data frames from a mobile' s higher-layer to MAC sublayer is Possion. Frame length is assumed to be exponentially distributed with mean length 1024 octets.

2. Voice traffic:
We used the mio8380's [13] built-in audio codec which is based on GSM610 format to generate voice traffic patterns. For each connection, frames are sent out every 10ms, and each connection is exponentially distributed with mean time 3 minutes.

3. Video traffic:
For video traffic, we use the mio8380's built-in CanaryA camera and domaintek MPEG4 encoder/decoder to generate video traffic source patterns. By means of CannaryA camera driver API, we can get YUV420 or RGB24 raw image data for the domaintek MPEG4 video encoder. The length of video stream is 1000 frames with 8 frames per second and 29 PPI (how many P frames follows per I frame).

In our setting, video, voice and data are assumed to be mixed in the ratio of 1:1:1. Performance is measured in terms of the average access delay, the loss probability, the dropping probability, the bandwidth utilization, among others. The default values used in the simulation are listed in Table II. The values for the simulation parameters are chosen carefully in order to closely reflect the realistic scenarios as well as to make the simulation feasible and reasonable.

Table II. Default attribute values used in the simulation.

Attribute	Value	Meaning & Explanation
Channel rate	2 Mb/s	Data rate for the wireless channel
Mobiles	10	10 mobile hosts in a basic service set
Slot_Time	$20\,us$	Time needed for each time slot
SIFS	$10\,us$	Time needed for each short interframe space
DIFS	$50\,us$	Time needed for each DCF interframe space
MAC header	272 bits	Header length of MAC layer header
PHY header	128 bits	Header length of physical layer header
RTS	160 bits + PHY header	Frame length of each request-to-send frame
CTS	112 bits + PHY header	Frame length of each clear-to-send frame
ACK	112 bits + PHY header	Frame length of each Acknowledgement
BER	10^{-6}	Bit error rate
m	0.8	Minimum bandwidth utilization wanted
II	5 %	Size of allocated bandwidth II

• Simulation Results

In the following, the performances of the proposed scheme and the conventional IEEE 802.11 protocol are compared based on simulations. Figures 4 and 5 show the dropping probability of real-time handoff connections and blocking probability of real-time new connections for the proposed scheme and conventional IEEE 802.11 protocol. Since the handoff dropping probability is the first measure used to adjust the allocated bandwidth and we also allow the handoff real-time traffic to use bandwidth exclusively with preemptive priority over other traffics in the reserved region, channel II. The dropping probability will be kept under the threshold, threshold_D, usually. It seems counterintuitive that the proposed scheme provides lower blocking probability in light load. This is because that the contention free period can start as soon as the request table just becomes empty, then the AP will end the current contention free period, and once the contention period start, the real-time traffic is not allowed to be served until the next contention period in the conventional IEEE 802.11 protocol.

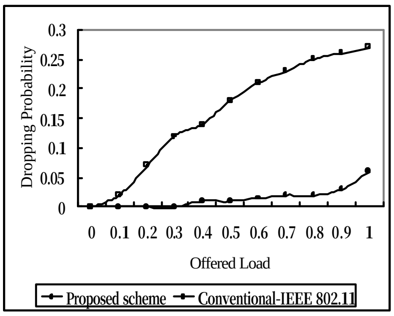

Figure 4. Dropping probability of real-time handoff connections

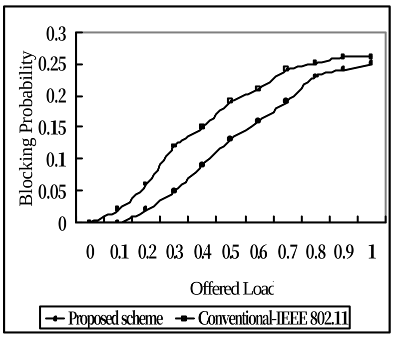

Figure 5. Blocking probability of real-time new connections

Figures 6 and 7 compare the average access delays of voice and video traffic from the proposed scheme and the conventional IEEE 802.11 protocol, respectively. We can see that although there is not much difference in the values of the performance measures when load is light, however, the proposed scheme provides significantly better performance than the conventional IEEE 802.11 protocol at heavy load.

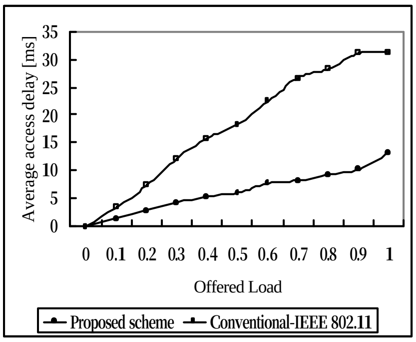

Figure 6. Average access delay of voice traffic

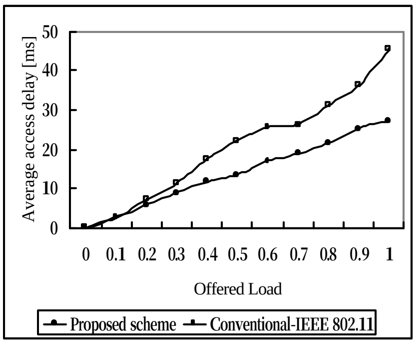

Figure 7. Average access delay of voice traffic

4. Conclusions

In this paper, we have proposed a pragmatic polling with non-preemptive priority-based access control scheme built on well-known protocols, offering easily implemented and yet flexible criteria for traffic prioritization in a wireless environment. By modifying the DCF access method in the contention period, our designed protocol supports multiple levels of priorities such that user mobility can be supported in wireless LANs. Besides, the proposed transmit-permission policy and adaptive bandwidth allocation scheme not only separate admitted inactivated users from newly requesting access users, but also derive sufficient conditions such that all the time-bounded traffic sources satisfy their time constraints to provide various deterministic QoS guarantees in the contention free period while maintaining efficient bandwidth utilization at the same time. Furthermore, the proposed scheduling algorithm for voice traffic is provably optimal in that it gives the minimum average waiting time for voice over IP packets. The proposed scheme is performed at each AP in a distributed manner. Through extensive simulations, we have demonstrated a satisfactory performance of our proposed scheme in a quantitative way.

References

[1] Wireless Medium Access Control and Physical Layer WG, IEEE Draft Standard P802.11, Wireless LAN, *IEEE Stds. Dept, D3*, Jan. 1996.

[2] J. L. Sobrinho and A. S. Krishnakumar, Quality-of-Service in Ad Hoc Carrier Sense Multiple Access Wireless Networks, *IEEE Journal on Selected Areas of Communications*, *17*(8), 1999, 1353-1368.

[3] D. J. Deng and R. S. Chang, A Priority Scheme for IEEE 802.11 DCF Access Method, *IEICE Trans. Commun.*, *E82-B*(1), 1999, 96-102.

[4] S. Chakrabarti and A. Mishra, QoS Issues in Ad Hoc Wireless Networks, *IEEE Communications Magazine*, *39*, 2001, 142-148.

[5] L. Bononi, M. Conti, and E. Gregori, Design and Performance Evaluation of an Asymptotically Optimal Backoff Algorithm for IEEE 802.11 Wireless LANs, *Proc. of the 33rd Hawaii International Conference on System Sciences*, 2000, 1-10.

[6] F. Calì, M. Conti, and E. Gregori, Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit, *IEEE/ACM Transactions on Networking*, *8*(6), 2000, 785-799.

[7] L. Alcuri, G. Bianchi and I. Tinnirello, Occupancy Estimation in the IEEE 802.11 Distributed Coordination Function, *Proc. Of ICS*, Hualien, Taiwan. December 2002.

[8] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms* (McGraw-Hill, 1990, 367-375).

[9] L. Bononi, M. Conti, and L. Donatiello, Design and Performance Evaluation of a Distributed Contention Control (DCC) Mechanism for IEEE 802.11 Wireless Local Area Networks, *Proc. Workshop WOWMOM'98, MOBICOM'98*, Dallas, Texas, 1998, 1-10.

[10] C. S. Chang, K. C. Chen, M. Y. You and J. F. Chang, Guaranteed Quality-of-Service Wireless Access to ATM Networks, *IEEE Journal on Selected Area of Communications*, *15*(1), 1997, 106-118.

[11] D. J. Deng and R. S. Chang, A Non-Preemptive Priority Based Access Control Scheme for Broadband Ad-Hoc Wireless ATM Local Area Networks, *IEEE Journal on Selected Areas of Communications*, *18*(9), 2000, 1731-1739.

[12] CACI Products Company, Simscript II.5, California 92037, Sep. 1997, http://www.caciasl.com/.

[13] MiTAC International Corporation, Mio8380, http://www.mitac.com/