2003

Distributed Admission Control for IEEE 802.11 Ad Hoc Networks

Hua Zhu, *et al.*

Distributed Admission Control for IEEE 802.11 Ad Hoc Networks

H. Zhu, M. Li, I. Chlamtac, B. Prabhakaran

TR UTD/EE/~~12~~13/2003

Distributed Admission Control for IEEE 802.11 Ad Hoc Networks

Hua Zhu, Ming Li, Imrich Chlamtac, and B. Prabhakaran

The University of Texas at Dallas, P.O. Box 830688
Richardson, TX 75083-0688
{zhuhua, mingli, chlamtac, praba}@utdallas.edu

***Abstract*—IEEE 802.11 has very poor performance in terms of throughput and transmission delay when the traffic load reaches the saturation condition. Admission control must be provided in order to guarantee the service of existing traffic. Unfortunately, the normalized saturation throughput is variable corresponding to different traffic statistics (i.e. bit-rate and average packet length). Therefore it does not perform well if the station admits traffic simply based on certain threshold of the normalized throughput.**

Most existing analytical models for IEEE 802.11 MAC adopt quite strict assumptions of saturation conditions and simplified traffic scenarios. Nevertheless, it is more realistic to analyze the non-saturation condition under heterogeneous traffic scenarios. Moreover, an accurate analytical model under non-saturation condition is critical for the correctness of admission control decisions.

In this paper, (1) we propose a unified analytical model which is the first model capable of analyzing performance under both non-saturation and saturation conditions; (2) we then introduce a new performance criterion, saturation coefficient $C_{n,sat}$, which reflects the degree of saturation experienced by any specific station; (3) finally we propose a distributed admission control scheme for IEEE 802.11 based on this criterion. With this scheme, any station can make local decision on whether admitting/rejecting a new traffic. The accuracy of the proposed analytical model and performance of the proposed admission control scheme are validated by simulations.

***Index Terms*—IEEE 802.11, DCF, Admission Control, QoS.**

1. INTRODUCTION

IEEE 802.11 [19] has gained popularity in the last mile wireless networks, and it will increasingly be used for multimedia applications. However, since IEEE 802.11 was developed as a simple and cost effective technology for best effort services, the lack of QoS support is becoming a problem. Recently, service differentiation in IEEE 802.11 has been extensively studied, and multiple variations of MAC schemes [1][3][6][7][8][14][15] have been proposed, including the 802.11e MAC. A nice overview of service differentiation can be found in [18].

Service differentiation is helpful in providing better quality of service for multimedia data traffic under low or medium traffic load condition. However, due to the inefficiency of IEEE 802.11 MAC, service differentiation strategy does not perform well under high traffic load condition [25]. In this case, admission control becomes a must in order to guarantee QoS of existing traffic. Otherwise, the extremely large saturation delay may lead to the failure of supporting multimedia applications.

Only a few papers explored the admission control in IEEE 802.11. Barry et al. [20] proposed a Virtual MAC algorithm, which passively monitors the channel by virtual MAC frames and estimates local service level (i.e. throughput and delay) by the measurement of virtual frames. Valaee and Li [23] proposed another measurement based admission procedure using a sequence of probe packet for ad hoc networks. Instead of using probe packets, Shah et al. [21] proposed a measurement based admission control scheme using data packets to measure the network load. Kazantzidis et al. [22] proposed heuristic solution using a simple parameter, permissible throughput as the admission decision criterion. Li and Prabhakaran [24] investigated a novel framework for admission control with priority reservation and allocation, which is mainly focus on optimizing the usage of priority resources. Banchs and Pérez [7] proposed ARME, an extension of DCF, using token bucket based algorithm to detect whether the network is in overloading condition, and thus to improve the performance of system by adjusting contention appropriately.

Considering the serious performance degradation of existing stations under even a short period of saturation, the admission control scheme should be capable to predict saturation and prevent it from really happening under various traffic statistics. And this scheme should be dynamic because of the constant change of traffic characterisitcs.

Admission control scheme with an absolute throughput threshold does not perform well, because the saturation throughput is variable corresponding to different traffic statistics (i.e. bit-rate and average packet length). The normalized saturation throughput may range from lower than 40% to higher than 70% depends on the average packet length of traffic and the configuration of contention parameters at stations. If the threshold is set too low, there will be too much waste of valuable bandwidth. On the other hand, if the threshold is set too high, in certain circumstances, it will not be able to prevent the saturation from happening.

In the measurement based admission control schemes, decisions are made based on the current performance of existing stations. They cannot actually calculate the dynamic saturation throughput. It may be possible that existing stations, but after they accept new stations or traffic flows, saturation occurs. Although dynamic correction can be taken to remove the newly added station once the saturation condition occurs, the negative influence, such as notification to the sender or initiation of a negotiation procedure, may be significant.

In order to build an admission control scheme with the capability of prediction and prevention of saturation, we introduce an index, named saturation coefficient $C_{n,sat}$, ranged from 0 to 1 and with 1 indicating saturation. By checking $C_{n,sat}$ with regard to a threshold lower than 1, admission control schemes can easily predict and prevent saturation from happening. Since $C_{n,sat}$ describes how close a station is to saturation condition, an analytical model capable of analyzing non-saturation system behavior is required.

Several models have been proposed in literature. Bianchi proposed a Markov model to describe the complicated process of IEEE 802.11 DCF in a landmark paper [9]. Several extensions have been published since then. Wu et al [2] proposed an extension of Bianchi's model with retransmission limits. Two papers (Chatzimisios et al. [11], Velkov and Spasenovski [13]) extended Bianchi's models into MAC transmission delay analysis by different approaches. Finally Zhu and Chlamtac [17] extended the model into IEEE 802.11e EDCF for service differentiation. Other models, which are not related to Bianchi's model, can be also found in [4][12][16].

However, most existing analytical models for IEEE 802.11 MAC adopt quite strict assumptions of saturation condition and simplified traffic scenarios [4][5], which are unusual because most networks work under non-saturation condition. Hence in order to well describe the critical transition from non-saturation to saturation, it is desirable to provide a model that relaxes the assumptions of saturation condition and also fully considers the traffic variations.

In this paper, (1) we first extend the Bianchi's model to the first unified model capable of analyzing performance in both non-saturation and saturation condition; (2) we then introduce a new performance criterion, i.e. saturation coefficient $C_{n,sat}$, which reflects the local degree of saturation of any station; (3) finally we propose a novel distributed admission control scheme for IEEE 802.11 based on this criterion. With this scheme, any station can make local decision on whether admitting new traffics. The accuracy of the proposed analytical model and performance criterion are validated by simulation.

The remainder of the paper is organized as follows. Section 2 proposes the analytical model, which is valid for both non-saturation and saturation conditions. Section 3 proposes a novel distributed admission control scheme based on the analytical model. Section 4 validates the analytical model by simulation and shows the performance improvement of the proposed admission control scheme. Finally, we conclude the paper in Section 5.

2. ANALYTICAL MODEL

The notations involved in the proposed analytical model are listed as follows.

CW_{min}/CW_{max}	minimum/maximum backoff window size
W_i	backoff window size in the i^{th} retransmission stage
S_n (bps)	traffic bit-rate
L_n (bytes)	average packet (payload) length
m	retry limit
m'	the number of retries to reach the maximum backoff window CW_{max}
n	index of station, $n \in [1, N]$
i	retransmission stage, $i \in [0, m]$
k	backoff counter, $k \in [0, W_i-1]$
$b_{n,i,k}$	stationary distribution of the Markov chain
$b_{n,empty}$	stationary empty queue probability at the n^{th} station
$C_{n,sat}$	saturation coefficient
τ_n	probability that the n^{th} station transmits in slot time given that its queue is not empty
p_n	conditional collision probability

The objective of the proposed analytical model is to introduce the saturation coefficient, $C_{n,sat}$, which indicates the degree of saturation for a station. In general, $C_{n,sat}$ depends on the existing traffic load of the system, the number of contending stations, and the average packet length. The higher the traffic load, the higher the $C_{n,sa}$. However, due to collision and backoff overheads, the maximum channel utilization is far less than 100% and $C_{n,sa}$ will reach one before the traffic load equals to the channel capacity. Therefore, we need to consider interactions between contending stations, which can be derived from transmission probabilities, τ_n's of all contending stations. Furthermore, given the same traffic load and number of stations, stations may be far from or close to the saturation condition depending on different traffic statistics. Also, shorter average packet lengths incur more contentions, while longer average packet lengths incur fewer contentions. For example, let there be two stations with identical flows of bit-rate of 500Kbps, changing the average packet lengths of all flows from 1000 bytes to 100 bytes will significantly increases the occurrences of collision. Thus, the degree of saturation of stations will be accordingly higher.

The transmission probabilities, τ_n's, have been calculated by Bianchi's DCF model. However, this calculation is inaccurate in practice because of the assumption of saturation condition, i.e. transmission queues of all stations are always nonempty. Furthermore, a station's probability of accessing the channel depends on traffic bit-rates of others when it is non-saturated. Therefore, under the non-saturation condition, the effects of traffic bit-rate variations among stations must be fully considered. By introducing an additional $b_{n,empty}$ state, which stands for the empty queue at the n^{th} station, and extending Bianchi's Markov model with traffic bit-rate variations, we relax the undesirable constraint of *saturation* condition, and propose a unified DCF model. This model is capable of analyzing performance under both non-saturation and saturation conditions, with various traffic bit-rates and average packet sizes at stations. With transmission probabilities, τ_n's, calculated from the proposed model, the saturation coefficient, $C_{n,sa}$, can be easily obtained. The proposed model is shown in Fig. 1.

Assume there are totally N stations. The traffic parameters of each station are traffic bit-rate, S_n (bps), and average packet (payload) length, L_n (bytes), with $n = 1,2\ldots N$. Stations do not necessarily have identical traffic parameters. And within a station, it might be possible that multiple applications are requesting transmissions concurrently. In this case, their packets contend the channel based on the FIFO policy. And, the traffic bit-rate of a station is the sum of bit-rates of all admitted traffic flows, excluding those flows that have not been accepted by the admission control scheme. Similarly, the average packet length is the overall average among admitted traffic flows.

Let CW_{min} and CW_{max} be the minimum and maximum backoff window size respectively. For convenience, we define $W=W_0=CW_{min}+1$, $W_i=2^iW$, and we have

$$W_{m'} = 2^{m'}W = CW_{\max} + 1 \qquad (1)$$

$$\begin{cases} W_i = 2^iW & i \le m' \\ W_i = 2^{m'}W & m' < i \le m \end{cases} \qquad (2)$$

where $(m+1)$ is the retry limit of the packet, and m' is the number of retries to reach the maximum backoff window CW_{max}. So far in 802.11 DCF short retry limit is 7, while the long retry limit is 4.

In Fig. 1, the state of the n^{th} station is described by the tuple $\{n, i, k\}$, where $n\in[1, N]$, $i\in[0, m]$, standing for the backoff stage (i.e. the number of retries), and $k\in[0, W_i\text{-}1]$, is the backoff delay in timeslots. Let $b_{n,i,k}$ be the stationary distribution of the Markov chain.

Each station has a transmission probability τ_n (i.e. the probability that the n^{th} station transmits in a randomly chosen slot time given the station queue is not empty) and a conditional collision probability p_n (i.e. the probability of collision seen by a packet from the n^{th} station being transmitted on channel). Moreover, we introduce the empty queue probability, $p_{n,empty}$, i.e. the probability that the n^{th} station's queue is empty at a randomly chosen slot time. $p_{n,empty}$ is not only determined by the traffic bit-rate and average packet length of a station itself, but also by the traffic statistics of other stations. Generally, higher traffic bit-rate yields higher $p_{n,empty}$ and vice versa. However, it is possible that for a specific station with low traffic bit-rate, its $p_{n,empty}$ is still high if the overall traffic load is high. Therefore, $p_{n,empty}$ provides a more accurate description of each specific station's loading condition, i.e., degree of saturation, than its traffic bit-rate.

In the Markov chain of the n^{th} station, the only non-null one-step transition probabilities are[1]

$$\begin{cases} P\{n,i,k \mid n,i,k+1\}=1 & k\in[0,W_i-2], i\in[0,m] \\ P\{n,0,k \mid n,i,0\}=(1-p_n)(1-p_{n,empty})/W_0 & k\in[0,W_0-1], i\in[0,m-1] \\ P\{n,i,k \mid n,i-1,0\}=p_n/W_i & k\in[0,W_i-1], i\in[1,m] \\ P\{n,0,k \mid n,m,0\}=(1-p_{n,empty})/W_0 & k\in[0,W_0-1] \end{cases} \quad (3)$$

Note that $b_{n,i-1,0}p_n= b_{n,i,0}$ for $(0\leq i\leq m)$, therefore

$$b_{n,i,0}=p_n{}^i b_{n,0,0} \quad (0\leq i\leq m) \quad (4)$$

We also have

$$b_{n,0,0}=(1-p_{n,empty})b_{n,empty} \quad (5)$$

[1] We adopt the similar short notation in [9]: $P\{n,i_1,k_1|n,i_0,k_0\}=P\{s(t+1)=i_1,b(t+1)=k_1|s(t)=i_0,b(t)=k_0$ for the n^{th} station$\}$.

And for each $k \in (0, W_i\text{-}1]$, we have

$$b_{n,i,k} = \frac{W_i - k}{W_i} \begin{cases} (1-p_n)(1-p_{n,empty})\sum_{j=0}^{m-1} b_{n,j,0} + b_{n,m,0} & i=0 \\ p_n b_{n,i-1,0} & 0 < i \le m \end{cases} \quad (6)$$
$$= \frac{W_i - k}{W_i} \begin{cases} (1-p_{n,empty} + p_n^m p_{n,empty}) b_{n,0,0} & i=0 \\ b_{n,i,0} & 0 < i \le m \end{cases}$$

By using the normalization condition for stationary distribution, we have

$$\begin{aligned} 1 &= b_{n,empty} + \sum_{i=0}^{m}\sum_{k=0}^{W_i-1} b_{n,i,k} \\ &= b_{n,empty} + \sum_{i=1}^{m} b_{n,i,0} \sum_{k=0}^{W_i-1} \frac{W_i - k}{W_i} + (1-p_{n,empty} + p_n^m p_{n,empty}) b_{n,0,0} \sum_{k=0}^{W-1} \frac{W_i - k}{W_i} \quad (7) \\ &= b_{n,0,0}\left[\frac{1}{p_{n,empty}} + \left(\sum_{i=1}^{m} p_n{}^i \frac{W_i + 1}{2}\right) + (1-p_{n,empty} + p_n^m p_{n,empty})\frac{W+1}{2}\right] \end{aligned}$$

With all contention parameters known, a general function of $b_{n,0,0}$ with respect to p_n and $p_{n,empty}$ for any station can be found in (8) for ($m \le m'$) and ($m > m'$) respectively.

$$b_{n,0,0} = \begin{cases} \dfrac{2(1-2p_n)(1-p_n)(1-p_{n,empty})}{\left[2Wp_n(1-(2p_n)^m)(1-p_n) + p_n(1-2p_n)(1-p_n{}^m) + (1-2p_n)(1-p_n)(1-p_{n,empty} + p_n^m p_{n,empty})(W+1)\right](1-p_{n,empty})+1} & m \le m' \\ \dfrac{2(1-2p_n)(1-p_n)(1-p_{n,empty})}{\left[2Wp_n(1-(2p_n)^{m'})(1-p_n) + p_n(1-2p_n)(1-p_n{}^m) + W2^{m'}p_n{}^{m'+1}(1-2p_n)(1-p_n{}^{m-m'}) + (1-2p_n)(1-p_n)(1-p_{n,empty} + p_n^m p_{n,empty})(W+1)\right](1-p_{n,empty})+1} & m > m' \end{cases} \quad (8)$$

The transmission probability τ_n that the n^{th} station transmits in a randomly chosen slot time given that its queue is not empty can be expressed as

$$\tau_n = \sum_{i=0}^{m} b_{n,i,0} = \frac{1-p_n{}^{m+1}}{1-p_n} b_{n,0,0} \qquad 1 \le n \le N \qquad (9)$$

In the stationary state, the n^{th} station transmits a packet with probability τ_n, so we have

$$p_n = 1 - \prod_{j=1, j\ne n}^{N} (1-\tau_j) \qquad 1 \le n \le N \qquad (10)$$

2.1. Saturation Coefficient

As we have discussed in the beginning of this section, the mathematical expression of $C_{n,sat}$ can be given as a function of bit-rate S_n, transmission probability τ_n, and the average packet length L_n, which is shown as follows.

$$
\begin{aligned}
C_{n,sat} &= \min\left[\left(\frac{\sum_{j=1}^{N} S_j}{bandwidth}\cdot\left(1+\sum_{j=1,j\neq n}^{N}\tau_j-\tau_n\sum_{j=1,j\neq n}^{N}\tau_j+\tau_n\sum_{j=1,j,k\neq n,j\neq k}^{N}\tau_j\tau_k+\cdots\right)\cdot\frac{\log(1024)}{\log(L_n)}\right),\ 1\right] \quad 1\leq n\leq N \\
&\cong \min\left[\left(\frac{\sum_{j=1}^{N} S_j}{bandwidth}\cdot\left(1+\sum_{j=1,j\neq n}^{N}\tau_j\right)\cdot\frac{\log(1024)}{\log(L_n)}\right),\ 1\right] \quad 1\leq n\leq N
\end{aligned}
\tag{11}
$$

Note that $p_{n,empty}$ is actually $1\text{-}C_{n,sat}$. Thus,

$$p_{n,empty}=1-C_{n,sat} \tag{12}$$

In (11), a station's influence on saturation coefficients, $C_{n,sat}$'s of other stations depends on its transmission probability, τ_n. For any two transmitting stations, if ($\tau_j>\tau_i$), then the j^{th} station has more impact on other stations' $C_{n,sat}$'s than the i^{th} station. For a specific station, the overall influence of all other stations is approximated as the summation of τ_j terms by ignoring high order terms. Obviously, we take the minimum since the saturation coefficient cannot be greater than 1. Intuitively, we assume that $C_{n,sat}$ increases logarithmically with the increase of the average packet length, which is validated by extensive simulations.

Equations (8), (9), (10) and (11) represent a nonlinear system with the three unknown vectors (τ_1, $\tau_2,\ldots\tau_N$), ($p_1, p_2,\ldots p_N$) and ($p_{1,empty}, p_{2,empty},\ldots p_{N,empty}$), which can be solved numerically. Note that we must have all τ_n's, p_n's and $p_{n,empty}$'s $\in(0,1)$.

2.2. Throughput

In order to further validate our model, we use the similar method, which is derived by [9], to calculate the throughput of each station. However, because we introduce the traffic variations, therefore there must be necessary adjustments on the collision probability and throughput equations.

Let P_b denote the probability that the channel is busy in a slot time. Then

$$P_b = 1 - \prod_{j=1}^{N} (1 - \tau_j) \tag{13}$$

Let $P_{n,s}$ denote the probability that the transmission of the n^{th} station is successful in a slot time. So we have

$$P_{n,s} = \tau_n \prod_{j=1, j\neq n}^{N} (1 - \tau_j) \tag{14}$$

Without loss of generality, let us assume that from the first to the last station, we have the longest to the shortest average packet length. Of course there may be two or more successive stations with the same value of average packet length.

Let $P_{n,c}$ denote the probability that a transmission with collision occurs in a slot time that only involves from the n^{th} to the N^{th} stations (i.e. the largest payload in collision belongs to the n^{th} station.) And let $T_{n,c}$ denote the corresponding average time that the channel is sensed busy because of a collision in which a traffic flow, which belongs to the n^{th} station, has the largest payload length. In that case, $P_{n,c}$ can be calculated, e.g.

$$
\begin{cases}
P_{n=1,c} = \tau_1 - P_{n=1,s} \\
P_{n=2,c} = \tau_2 \times \prod_{j=1}^{1}(1-\tau_j) - P_{n=2,s} \\
P_{n=3,c} = \tau_3 \times \prod_{j=1}^{2}(1-\tau_j) - P_{n=3,s} \\
\cdots \\
P_{n=N,c} = \tau_N \times \prod_{j=1}^{N-1}(1-\tau_j) - P_{n=N,s}
\end{cases}
\qquad (15)
$$

$$
\Rightarrow P_{n,c} = \begin{cases}
\tau_1 - P_{1,s} & n=1 \\
\tau_n \times \prod_{j=1}^{n-1}(1-\tau_j) - P_{n,s} & 1 < n \le N
\end{cases}
$$

To verify the correctness of (15), we have

$$
(1-P_b) + \sum_{n=1}^{N} P_{n,s} + \sum_{n=1}^{N} P_{n,c} = 1
$$

The normalized throughput of the n^{th} station, S_n, is

$$
S_n = \frac{E[payload \quad in \quad a \quad slot \quad time]}{E[length \quad of \quad a \quad slot \quad time]} = \frac{P_{n,s} E[P_n](1-p_{n,empty})}{(1-P_b)\sigma + \sum_{n=1}^{N} P_{n,s} T_{n,s} + \sum_{n=1}^{N} P_{n,c} T_{n,c}} \qquad (16)
$$

In the numerator of the above equation, E[P_n] is the average packet length. In the denominator of the above equation, σ is the duration of an empty slot time.

Let packet header be H = PHY_{hdr}+MAC_{hdr} and let the average propagation delay be δ. Based on the average payload length of the n^{th} station, we have the following expressions of $T_{n,s}$ and $T_{n,c}$. Note that the packet header H must include the additional MAC overhead due to the fragmentation if necessary.

$$
\begin{cases}
T_{n,s}^{bas} = DIFS + H + E[P_n] + \delta + SIFS + ACK + \delta \\
T_{n,c}^{bas} = DIFS + H + E[P_n^*] + SIFS + ACK
\end{cases}
\qquad (17)
$$

$$
\begin{cases}
T_{n,s}^{rts} = DIFS + RTS + SIFS + \delta + CTS + SIFS \\
\qquad + \delta + H + E[P_n] + \delta + SIFS + ACK + \delta \\
T_{n,c}^{rts} = DIFS + RTS + SIFS + CTS
\end{cases}
\qquad (18)
$$

where *bas* and *rts* means basic access method and RTS/CTS access method (mandatory for all data frames) respectively. E[P_n*] is the average length of the longest packet payload in a collision. Based on the definition of $P_{n,c}$ used in (15), E[P_n*]=E[P_n]. The ACK, RTS and CTS are calculated as the transmission time of the corresponding frame length and the PHY overhead.

The aggregated throughput of N stations is

$$S = \sum_{n=1}^{N} S_n \qquad (19)$$

2.3. MAC Transmission Delay

In this paper, the MAC frame transmission delay is defined as the time interval between two successive successful frame transmissions for a station. It is possible that these two frames are not consecutive if a frame is dropped after exceeds the retry count.

By this definition, the average frame delay of the n^{th} station, E[D_n], can be calculated as

$$E[D_n] = \frac{E[P_n]}{S_n} \qquad (20)$$

3. Distributed Admission Control Scheme

In general, admission control can be implemented in centralized or distributed way. However, the centralized admission control is not preferred in IEEE 802.11 because of two reasons: (1) centralized admission control exposes location-dependent error, and the central control station may be the bottleneck of the system; (2) in multi-hop ad hoc networks, a station may be several hops away from the central control station, making the signaling overhead unacceptable. Therefore in IEEE 802.11, a fully distributed admission control scheme is desirable. Also, it should be noted that there is no support of service differentiation in DCF. Thus, in our scheme, we do not distinguish real-time traffic and best-effort traffic at the time of making decision. However, it is possible to extend our scheme to EDCF, which provides priority based service differentiation.

The proposed analytical model provides an ideal quantitative performance criterion, $C_{n,sat}$, for admission control. In this section, we introduce a distributed admission control scheme, which is compatible with both infrastructure and ad hoc modes. Every station will make decision on whether to admit/reject incoming/outgoing traffics based on the local information of saturation coefficient $C_{n,sat}$, which is calculated at each station locally. To make this calculation possible, the necessary information, which includes the number of contending stations and the corresponding total bit-rates of admitted traffics of each contending station, need to be disseminated along with regular MAC frame transmissions. There are two ways for a station to obtain the necessary information, either by a cross layer protocol that the upper layer (e.g. IP layer) information can be obtained, or by piggybacking the information directly in RTS/CTS/DATA/ACK frames of IEEE 802.11 MAC. The advantage of piggyback is that it involves less modification of the standard DCF and it does not require frequent accesses to upper layers. Therefore in this paper, we adopt the piggyback method in designing the admission control scheme.

To realize the piggyback technique, RTS/CTS/DATA/ACK frames [1] are modified by inserting additional fields, which are shown in Fig. 2 and Table 1. In each type of frame, we insert a 2-bit bit-rate flag immediately after the duration field. This flag indicates whether there are additional fields of bit-rate and average packet length information and whether it is the start of a new flow or the end of an existing flow. Also, in the following part, we use *frame* in MAC layer while *packet* in IP layer.

3.1. Frame Format

- For RTS/DATA frame:

1) If the flag is zero, it means no additional field. There will be 6-bits padding, followed by the remaining regular fields.
2) If the flag is one, it indicates that the frame is the start (i.e. from the first packet) of a new flow from

that source station, and the following bits are the bit-rate (Kbps) and the average packet length (10bytes) of the new flow. And this frame is the request for admission from a new traffic flow of the source station.

3) If the flag is two, invalid.
4) If the flag is three, it indicates that the frame is the end (i.e. from the last packet) of an existing flow. And the following bits are the bit-rate (Kbps) and the average packet length (10bytes) of the existing flow.

- For CTS/ACK frame:

1) If the flag is zero, it means no additional field. There will be 6-bits padding, followed by the remaining regular fields.
2) If the flag is one, it indicates that the transmission request for the new flow from the corresponding source station has been accepted. And the following bits are the bit-rate (Kbps) and the average packet length (10bytes) of the new flow.
3) If the flag is two, it indicates that the transmission request for the new flow from the corresponding source station has been rejected. And the following bits are the bit-rate (Kbps) and the average packet length (10bytes) of the new flow.
4) If the flag is three, it indicates that the frame is the end (i.e. from the last packet) of an existing flow. And the following bits are the bit-rate (Kbps) and the average packet length (10bytes) of the existing flow.

The average packet length field is a fixed 8 bits-long, with 1-255 represents the average packet length of 10-2550bytes. The variable length of the bit-rate field (in Kbps) after the bit-rate flag depends on the version of IEEE 802.11. For example, for IEEE802.11b, the total channel bandwidth is 2Mbps. So the 11-bits for length of the bit-rate field will be enough. So we have 2-bit flag, 11-bits bit-rate field, 8-bits pktLen field, plus 3-bits padding, total three bytes. For IEEE 802.11a, the total bandwidth is 54Mbps. Therefore the bit-rate field may be longer. However, three bytes (i.e. 2-bit flag, 8-bits pktLen, 14-bits bit-rate field) should be also enough for IEEE 802.1a in practice, since 14-bits is able to represent the maximum 16Mbps, which is enough for a single station. Since all the information is piggybacked, there is no extra signaling message involved and the overhead for admission control is minimized.

3.2.Source-Destination Admission Control Strategy

Admission control decisions are made based on the performance criterion – saturation coefficient, which is described in Section 2. If the saturation coefficient reaches one, it means that saturation occurs and stations should reject all data transmission requests. In practice, it is advisable that the admission control decisions should be made well before the saturation coefficient reaches one. Thereby, we introduce a saturation threshold, T_{sat}. When $C_{n,sat} \geq T_{sat}$, no new traffic will be accepted. Generally, admission control decisions can be either aggressive or conservative, depending on the selection of T_{sat}.

Admission decisions can be made at either source stations, or destination stations. Usually, it is straightforward to enforce admission control locally at the source station before it attempts data transmission. However, it is possible that when stations enter a broadcast region, there is no system-wise traffic information available. Therefore, transmission requests originated from these stations will be approved automatically without being aware of the presence of other existing traffics. In this case, to eliminate this possibility, it is necessary to enforce admission control at the destination station as well. Therefore, we propose source-destination dual admission control strategy.

To accommodate the fact that there may be multiple flows within a station, we define *existing traffic bit-rate* of a station as the sum of bit-rates of all traffic flows that have been accepted/admitted (i.e. admitted) for transmission from the corresponding station. Each station maintains a *local table* of existing traffic bit-rates of all stations in its broadcast region. However, a station only knows its own existing traffic bit-rate. In order to collect the information of other stations, a station needs to listen to the channel, and get the piggybacked information along with the source/destination station addresses, which are already included in the standard MAC frames. For all stations being in the broadcast region for enough long time, the completeness of information in their local table is very high. Therefore in most cases, admission decisions are correctly made at the source stations.

In the local table, the existing traffic bit-rates of other stations will only be updated when a station receives a CTS/ACK frame with bit-rate flag equal to one or three. If the flag is one, the value of bit-rate obtained from the following field will be added to the existing traffic bit-rate of the corresponding station in the local table. Similarly, if the flag is three, the value of bit-rate obtained from the following field will be deducted from the existing traffic bit-rate of the corresponding station in the local table.

Every time a station updates the local table, it will also re-calculate the new saturation coefficient. The computational complexity of the non-linear solution for the analytical model is associated with the number of stations that have accepted traffic flows. In a BSS (Basic Service Set) with relatively small number of stations, the exact value of the saturation coefficient can be calculated, while for the large station number, the empirical approximation, which is described in Section 4.2, will be used to calculate the local saturation coefficient.

Then, admission control procedures based on the proposed source-destination strategy are briefly described as follows.

- Source station decision: (the station has a new traffic flow to send)

  Depending on which access methods (basic, rts/cts) are used, the station will send the first RTS/DATA frame of the new traffic flow with bit-rate flag equal to one only if ($C_{n,sat}<T_{sat}$); while if ($C_{n,sat}\geq T_{sat}$), the station will not bother to send the request at all because itself is already in saturation condition. In order to minimize unnecessary computation and processing, the new traffic flow has to

wait for a certain time before the station retries.

Sending a request does not guarantee the admission of the new flow. The source station has to wait for the reply from the destination station. If the replied CTS/ACK has a flag equal to one, which means acceptance, the source station will transmit the following upper layer packets in this flow till the completion. If the replied CTS/ACK has a flag equal to two, which means rejection, the source station has to wait for a certain period before the next request.

- Destination station decision: (the station receives a admission request via a RTS/DATA frame with the bit-rate flag equal to one)

  The station will accept the request by reply with a CTS/ACK with bit-rate flag set to one if ($C_{n,sat}<T_{sat}$); otherwise it will reject the request by reply with a CTS/ACK with bit-rate flag set to two.

3.4.Extension to Multi-hop Ad Hoc Networks

By incorporating with routing protocols, our scheme can be extended to multi-hop ad-hoc networks with necessary modifications. In multi-hop ad-hoc networks, nodes have different neighbors within their broadcast regions and collect different local traffic information. Therefore, each node has different saturation coefficient, which reflects the degree of congestion in its broadcast regions. If a node is in saturated status, it will not be selected as the intermediate router for end-to-end paths of new flows. Furthermore, with these local saturation coefficients calculated at each node, routing protocols can decide the least congested paths along intermediate routers to destinations. This saturation coefficient based QoS routing approach will be our future work.

4. SIMULATION

4.1. *Validation of the Proposed Analytical Model*

The simulation is implemented by NS-2. The general DCF parameters are shown in Table 2 [10]. All other parameters can be derived from these basic parameters, e.g., DIFS=SIFS+2SlotTime. To validate the analytical model, we choose two scenarios, homogeneous traffic (i.e. all stations have identical traffic statistics) and heterogeneous traffic (stations may have different traffic statistics).

In homogeneous traffic scenario, we test both 40Kbps/station and 200Kbps/station VBR traffics, which are shown in the Fig. 3 and Fig. 4 respectively. For each traffic bit-rate, we test different average packet lengths (i.e. 164/1024/2000 bytes). Normally, the voice traffic is CBR with short packet length. For example, 160 byte-long packets (ITU-T G.711 speech codec) plus 4 byte-long compressed RTP/UDP/IP headers have been be used for good quality voice calls. The average packet length of 1024 bytes is often used to analyze general IP traffic, while the average packet length of 2000 bytes may be used for compressed video traffic.

From both Fig.3 and Fig. 4, we can see that the proposed analytical model is accurate under the non-saturation condition, while there is slight difference between the analytical model and simulation under the saturation condition, which is mainly caused by the simple assumption of $T_{n,c}$ in (17) and (18), without considering the NAV effect on other stations. However, since only the non-saturation part of the analytical model is useful for the admission control scheme, the model is acceptably accurate. In Fig. 3 (a1), (b1) and (c1), the normalized saturation throughputs are ranged from 0.35 (for 164 byte-long) to 0.65 (for 2000 byte-long). Also, the average packet length affects the *saturation point*, which is the minimum traffic load that triggers the saturation condition. The numerical results of τ_n, p_n and $p_{n,empty}$ are shown in Fig. 3 (a2), (b2) and (c2). We can see that the saturation point is located at where the empty queue probability $p_{n,empty}$

reaches zero (or we can say that the saturation coefficient reaches one).

With the same traffic load, the number of 40 Kbps-stations required in Fig.3 is larger than the number of 200 Kbps-stations required in Fig. 4. Since the more stations, the more contentions and collisions may occur, the throughput achieved in Fig. 4 is higher than that in Fig. 3.

In summary, both the bit-rate per station and the average packet length of stations affect the saturation point and the value of normalized saturation throughput. Therefore the admission control decision simply based on static throughput threshold is inappropriate under different traffic statistics. Instead, this inaccuracy can be eliminated by making admission control decision based on the saturation coefficient.

In the real world, it is very common that stations do not have identical traffic patterns. The correctness of the proposed analytical model is also validated in heterogeneous traffic scenarios. In Fig. 5(a), the number of 40Kbps-stations is five times of that of 200Kbps stations. Similarly, Fig. 5(b), the number of 20Kbps-stations is five times of that of 100Kbps stations. The model is still accurate under the non-saturation condition.

4.2. Empirical Equation of the Saturation Coefficient

In the analytical model, the saturation coefficients are solved by non-linear equations, which is inefficient to be implemented in practical admission control schemes. Instead, it is more feasible to find an empirical equation of the saturation coefficients. From both Fig. 3 and 4, we can see that, under the non-saturation condition, the empty queue probability decreases linearly with a constant slope when the traffic load increases. Therefore, under the non-saturation condition, the increment of the empty queue probability will be a constant given the same traffic statistics (i.e. bit-rate and average packet length) of the newly accepted flow.

Let $\Delta C_{n,sat}$ be the increment of the saturation coefficient due to the newly accepted flow, bit-rate of the newly added flow be ΔS, and the corresponding average packet length be L. Given ΔS and L, this slope,

$f(\Delta S,L)$, is a constant, and can be expressed as

$$f(\Delta S,L)=\frac{\Delta C_{n,sat}}{\Delta S}=\frac{p_{n,empty}^{old}-p_{n,empty}^{new}}{\Delta S}=\frac{\Delta p_{n,empty}}{\Delta S} \qquad (21)$$

$f(\Delta S,L)$ represents how fast the saturation coefficient reaches one. If $\Delta S_1>\Delta S_2$, we have $f(\Delta S_1,L)<f(\Delta S_2,L)$; if $L_1>L_2$, we have $\Delta p_{n,empty,1}<\Delta p_{n,empty,2}$, thus $f(\Delta S,L_1)<f(\Delta S,L_2)$.

As in practice, traffic flows are categorized into a number of classes based on the bit-rate and the average packet length. In our simulation, we divide the traffic into *Low*(40 Kbps)/*High*(200 Kbps) bit-rate and *Short*(164 byte)/*Medium*(1024 byte)/*Long*(2000 byte) average packet length. Therefore there are totally six classes. For each type of class, $f(\Delta S,L)$ can be pre-calculated by the nonlinear solution and set at the station as other contention parameters. The stations will simply search the correct $f(\Delta S,L)$ based on ΔS and L of the newly accepted traffic, and add the corresponding $\Delta C_{n,sat}$ to the current saturation coefficient. The computational complexity is linear with the number of incoming traffics. The pre-calculated $f(\Delta S,L)$'s for our simulation are shown in Table 3.

4.3. Performance of the Admission Control Scheme

In this section, simulations are conducted to evaluate the performance of the proposed admission control scheme in single-hop ad-hoc networks. All stations are assumed to be in the same broadcast region and each station only initiates one flow. Our simulation topology is a ring composed by totally 15 stations. Each station sends a flow to the next one in the ring and it is also the receiver of the flow originated from the previous station. All flows have the same bit rate of 200Kbps, and average packet size of 2000byte, respectively. From t=0 second, every 1 second, a new flow arrives until t=14 seconds and the simulation duration is 17 seconds. The performance of the first flow is shown in Fig. 6, in terms of throughput and end-to-end delay. It can be seen that without admission control, DCF experiences unbounded large end-to-end delay and low throughput. However, with the proposed admission control scheme, new flows are rejected once the saturation coefficient exceeds the specified saturation threshold. Therefore, QoS requirements of accepted flows can be guaranteed. We also compare the effects of choosing saturation thresholds of 0.8 and 0.9. With T_{sat} of 0.8 and 0.9, the numbers of accepted flows are 8 and 9, respectively. Thus, certain performance degradation is expected when T_{sat}=0.9 comparing to the case of T_{sat} =0.8, although the corresponding rejection rate is slightly lower. Notice that when T_{sat} =0.9, the system is actually in quasi-saturation, which is characterized by unstable delay and throughput (dash line in Fig. 6). However, QoS of existing flows are guaranteed when T_{sat}=0.8. Therefore, considering the tradeoff between rejection rate and QoS guarantee, we recommend T_{sat} of 0.8.

5. CONCLUSION

In this paper, we proposed the first analytical model capable to analyze both non-saturation and saturation condition. Based on that model, we then introduced saturation coefficient (i.e. $C_{n,sat}$, which reflects the degree of saturation of any station) as the performance criterion for the admission control. Finally we developed a novel fully distributed source-destination admission control scheme, which avoids the problem of enforcing admission control at new source stations. The required modification of the standard IEEE 802.11 is quite small, without introducing much additional overhead. And the admission control scheme has very low computation complexity. The accuracy of the proposed analytical model and performance of the proposed admission control scheme have been shown by simulation.

Further research will be located at several areas: (1) to further extend the signaling part of the admission control scheme for multi-hop ad hoc networks; (2) to support EDCF with prioritized traffic; (3) to design a QoS routing protocol for IEEE 802.11 ah hoc networks that utilizes the saturation coefficient information to select optimal end-to-end paths.

REFERENCES

[1] J. Deng and R.S. Chang, "A Priority Scheme for IEEE 802.11 DCF Access Method," IEICE Trans. Commun., Vol. E82-B, NO.1, Jan. 1999.

[2] Haitao Wu, Yong Peng, Keping Long, Shiduan Cheng, Jian Ma, "Performance of Reliable Transport Protocol over IEEE 802.11 Wireless LAN: Analysis And Enhancement," in Proceedings of *IEEE INFOCOM 2002*, New York, NY, June 23--27 2002.

[3] H.L. Truong, G. Vannuccini, "The IEEE 802.11e MAC for Quality of Service in Wireless LANs," in Proceedings of SSGRR 2003w, L 'Aquila, Italy, Jan.6-12, 2003.

[4] F. Calí, M. Conti, E. Gregori, "IEEE 802.11 wireless LAN: Capacity analysis and protocol enhancement," in Proc. INFOCOM'98, pp. 142-149, San Francisco, CA, March 29-April 2, 1998.

[5] G. Anastasi, L. Lenzini, "QoS provided by the IEEE 802.11 wireless LAN to advanced data applications: a simulation analysis," ACM/Baltzer Wireless Networks, Vol.6, No.2, pp. 99-108, 2000.

[6] Imad Aad, Claude Castelluccia, "Differentiation Mechanisms for IEEE 802.11," In Proceedings of the IEEE Infocom 2001 Conference, Anchorage, Alaska USA , April 2001.

[7] Albert Banchs, Xavier Pérez, "Providing throughput guarantees in ieee 802.11 wireless LAN," in WCNC2002, Vol. 1, pp. 130-138, 2002.

[8] Anders Lindgren, Andreas Almquist, Olov Schelén, "Quality of service schemes for IEEE 802.11: a simulation study," Proceedings of the Ninth International Workshop on Quality of Service (IWQoS), 2001.

[9] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," IEEE Journal on Selected Areas in Communication, Vol. 18, No. 3, March 2000.

[10] M. Gast, 802.11 *Wireless Networks: the Definitive Guide*. Sebastopol, CA, O'Reilly, 2002.

[11] P. Chatimisios, V. Vitsas, A.C. Boucouvalas, "Throughput and delay analysis of IEEE 802.11 protocol", in Proceedings of the 5th IEEE International Workshop on Network Appliances (IWNA) 2002, Liverpool, UK, Oct. 2002.

[12] Y. Ge, J. Hou, "An analytical model for service differentiation in IEEE 802.11," to appear in Proc. IEEE Int'l Conf. on Communications, Anchorage, Alaska, May 2003.

[13] Z. Hadzi-Velkov, B. Spasenovski, "Saturation throughput – delay analysis of IEEE 802.11 DCF in fading channels," IEEE International Conference on Communications 2003 (ICC'03), Vol. 1, pp. 121-126, Anchorage, Alaska, May 2003.

[14] S. Choi, J.D. Prado, S. Shankar, S. Mangold, "IEEE 802.11e contention-based channel access (EDCF) performance evaluation," in Proc. IEEE Int'l Conf. on Communications, Anchorage, Alaska, May 2003.

[15] Y. Xiao, "Backoff-based priority schemes for IEEE 802.11," in Proc. IEEE Int'l Conf. on Communications, Anchorage, Alaska, May 2003.

[16] F. Eshghi, A.K. Elhakeem, "Performance analysis of ad hoc wireless LANs for real-time traffic," IEEE Journal on Selected Areas in Communications, Vol. 21, No. 2, Feb. 2003.

[17] H. Zhu, I. Chlamtac, "An analytical model for IEEE 802.11e EDCF differential services," to be appeared on the 12th International Conference on Computer Communications and Networks (ICCCN'03), Dallas, TX, Oct. 2003.

[18] W. Pattara-Atikom, P. Krishnamurthy, "Distributed mechanisms for quality of service in wireless LANs," in the IEEE Wireless Communications, QoS in Next Generation Wireless Multimedia Communications Systems, Vol. 10, No. 3, pp. 26-34, June 2003.

[19] IEEE 802.11 WG, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specification, Aug. 1999.

[20] M. Barry, A.T. Campell, A. Veres, "Distributed control algorithms for service differentiation in wireless packet networks," in Proceedings of IEEE INFOCOM 2001.

[21] S.H. Shah, K. Chen, K. Nahrstedt, "Dynamic bandwidth management for single-hop ad hoc wireless networks," in Proceedings of the first IEEE International Conference on Pervasive Computing and Communications (PerCom'03). 2003.

[22] M. Kazantzidis, M. Gerla, S.-J. Lee, "Permissible throughput network feedback for adaptive multimedia in AODV MANETs," IEEE International Conference on Communications (ICC'01), Vol. 5, pp. 1352-1356, June 11-14, 2001.

[23] S. Valaee, B. Li, "Distributed call admission control in wireless ad hoc networks," in Proceedings of IEEE Vehicular Technology Conference (VTC 2002), Vancouver, British Columbia, Sept. 24-28, 2002.

[24] M. Li, B. Prabhakaran, S. Sathyamurthy, "On Flow Reservation and Admission Control for Distributed Scheduling Strategies in IEEE802.11 Wireless LAN," to be appeared on the Sixth ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM'03), San Diego, CA, Sept. 14-19, 2003.

[25] A. Lindgren, A. Almquist and O. Schelén, "Evaluation of quality of service schemes for IEEE 802.11 wireless LANs," in Proceedings of the 26th Annual IEEE Conference on Local Computer Networks (LCN 2001), Tampa, FL, November 15-16, 2001.

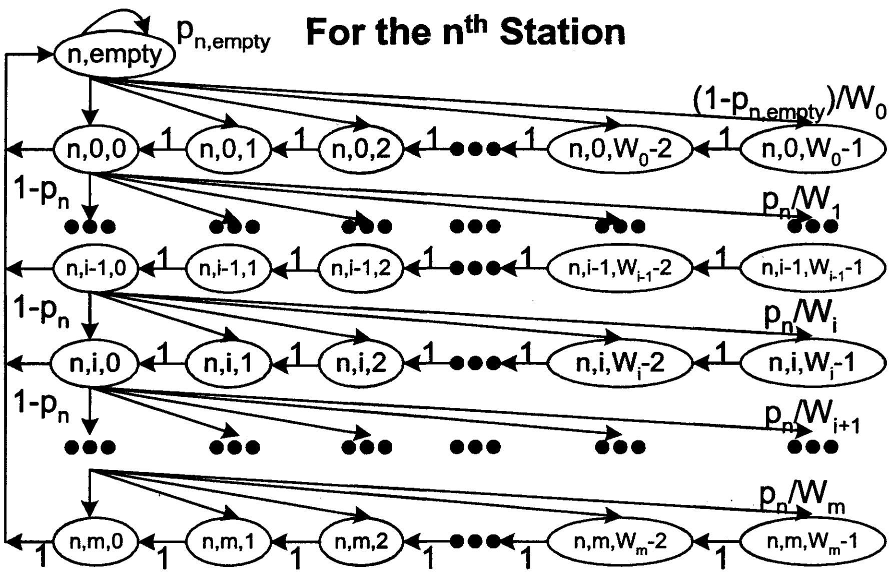

Fig. 1. Markov chain model of DCF (including the non-saturation condition and station-based traffic differentiation)

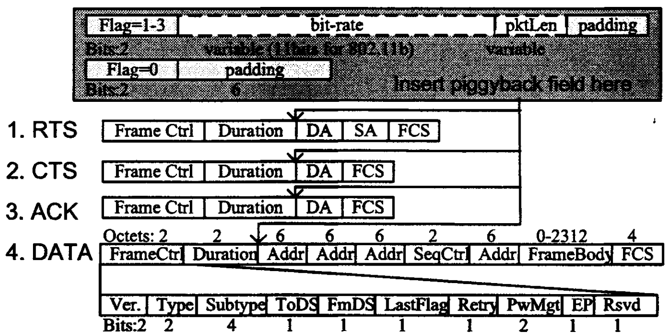

Fig. 2. Modified frame format

TABLE 1 FORMAT OF PIGGYBACK FIELDS

RTS/DATA Bit-rate flag	Indication
0	Regular frames, followed by 6-bits padding.
1	Source station requests for a new traffic flow, followed by bit-rate (Kbps) and pktLen (10bytes) of the new traffic flow.
2	(only valid for CTS/ACK)
3	The existing flow ends, followed by bit-rate (Kbps) and pktLen (10bytes) of the existing traffic flow.
CTS/ACK Bit-rate flag	**Indication**
0	Regular frames, followed by 6-bits padding.
1	Destination station accepts the new flow, followed by bit-rate (Kbps) and pktLen (10bytes) of the new traffic flow.
2 (only valid for CTS/ACK)	Destination station rejects the new flow, followed by bit-rate (Kbps) and pktLen (10bytes) of the new traffic flow.
3	The existing flow ends, followed by bit-rate (Kbps) and pktLen (10bytes) of the existing traffic flow.

TABLE 2 IEEE 802.11 DCF PARAMETERS

General DCF Parameters	
Slot Time	20 μs
SIFS	10 μs
Max Propagation Delay	2 μs
Retry Limit (short/long)	7/4
RTS_threshold	2346 Bytes
Fragmentation threshold	2312 Bytes
DSSS PreambleLength	144 bits
DSSS PLCPHeaderLength	48 bits
PLCP Transmission Rate	1 Mbps (DBPSK)
PPDU Transmission Rate	2 Mbps (DQPSK)

TABLE 3 PRE-CALCULATED $f(\Delta S, L)$

BW = 2Mbps	ΔS: *Low* (40 Kbps)	ΔS: *High* (200 Kbps)
L: Short (164 byte)	2.208568/BW	1.792333/BW
L: Medium (1024 byte)	1.65682/BW	1.379641/BW
L: Long (2000 byte)	1.514873/BW	1.274952/BW

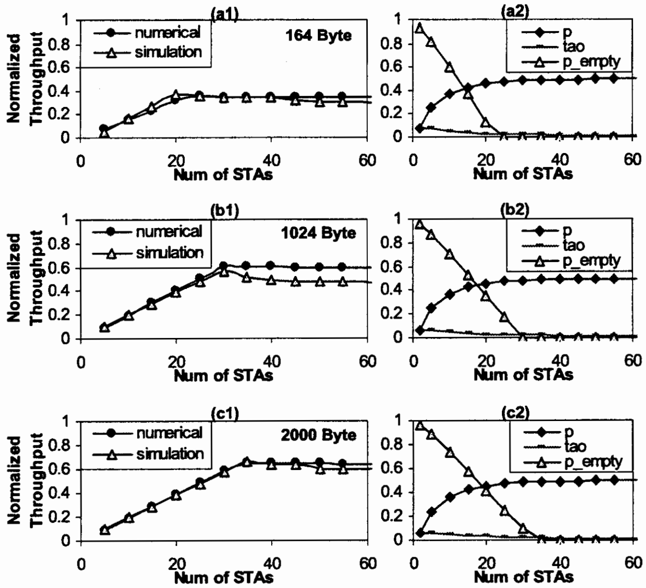

Fig. 3. Normalized throughput and other parameters from analytical model (each station has a bit-rate of 40 Kbps)
(a) average packet length 164 bytes; (b) average packet length 1024 bytes; (c) average packet length 2000 bytes.

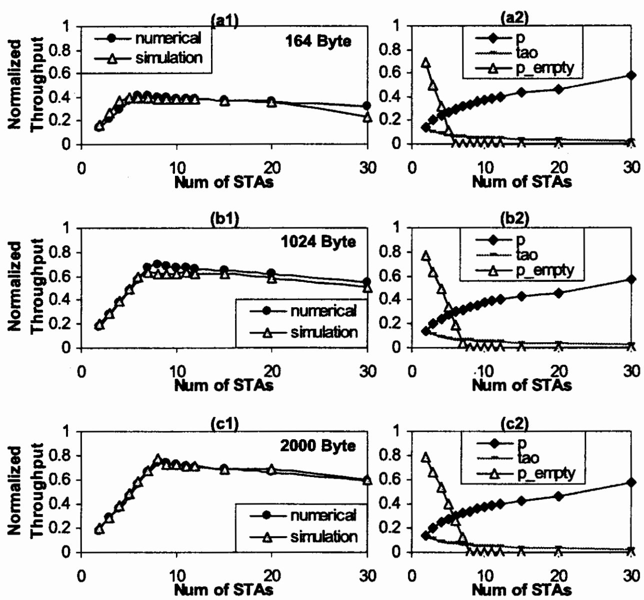

Fig. 4. Normalized throughput and other parameters from analytical model (homogeneous traffic, 200 Kbps/station)
(a) average packet length 164 bytes; (b) average packet length 1024 bytes; (c) average packet length 2000 bytes.

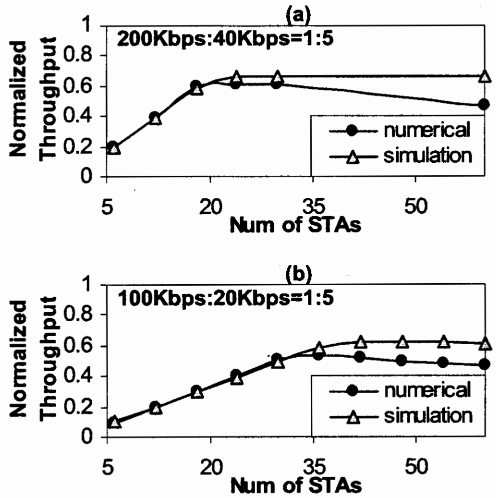

Fig. 5. Normalized throughput and other parameters from analytical model (heterogeneous traffic, average packet length 1024 byte)
The number of stations in (a) 200Kbps:40Kbps=1:5; (b) 100Kbps:20Kbps=1:5.

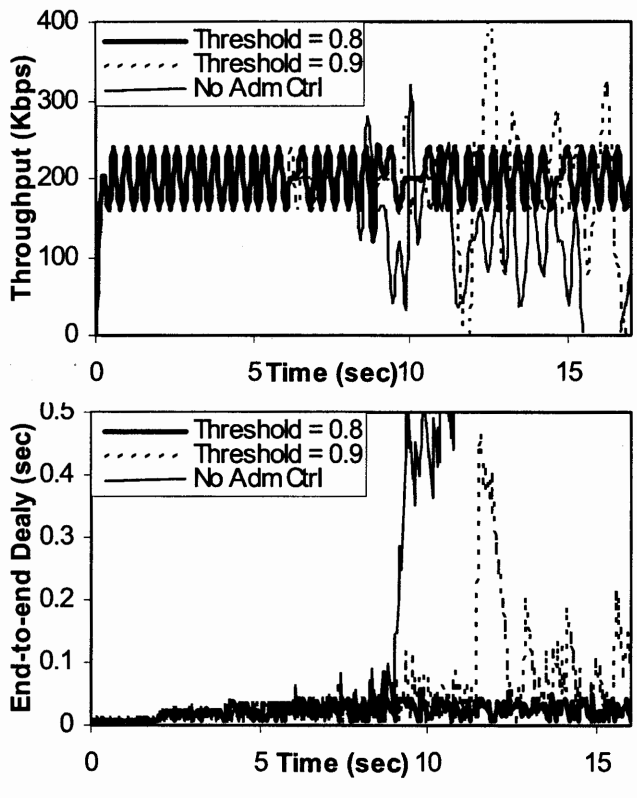

Fig. 6. Performance of admission control vs. time