A Queueing Model for Finite Load IEEE 802.11 Random Access MAC

Omesh Tickoo and Biplab Sikdar
Department of Electrical, Computer and Systems Engineering
Rensselaer Polytechnic Institute, Troy, NY, 12180
Email:tickoo,bsikdar@networks.ecse.rpi.edu

***Abstract*— This paper presents an analytic model for evaluating the MAC layer queueing delays at wireless nodes using the Distributed Coordination Function of IEEE 802.11 MAC speci£cations. Our model is valid for £nite loads and can account for arbitrary arrival patterns, packet size distributions and number of nodes. Each node is modeled as a discrete time $G/G/1$ queue and we obtain closed form expressions for the delay and queue length characteristics at each node. We derive the service time distribution for the packets at each node while accounting for a number of factors including the channel access delay due to the shared medium, impact of packet collisions, the resulting backoffs as well as the packet size distribution. Our analytical results are veri£ed through extensive simulations and are more accurate than existing models.**

I. Introduction

The IEEE 802.11 MAC [7] has become ubiquitous and gained widespread popularity as a layer-2 protocol for wireless local area networks. While efforts have been made to support the transmission of real time traf£c in such networks they primarily use centralized scheduling and polling techniques based on the point coordination function (PCF). For ad hoc scenarios, a more reasonable model of operation is that of random access and the distributed coordination function (DCF) where it is substantially more dif£cult to provide delay guarantees, and the performance of the MAC protocol can easily become the bottleneck due to factors like channel contention delays and collisions. In order to provide such guarantees, it is necessary to be able to characterize the delays and other performance metrics in these networks. In this paper we developing an analytic model for the delay and queue length characteristics in IEEE 802.11 MAC based networks in the random access mode with £nite load.

Existing work on the performance of the 802.11 MAC has focused primarily on its throughput and capacity [3], [10], adaptive backoff schemes [2], [14] and traf£c characteristics [11]. A simulation based comparison of the delays in 802.11b and 802.11e in the DCF mode is presented in [4]. Delay analysis for the PCF mode of operation has been proposed in [5], [13] but no such analysis been reported for the DCF case. In [12] a queueing model has been proposed for IEEE 802.11. However, it makes many simplifying assumptions resulting is various inaccuracies. In this paper, we improve the model of [12] by including explicitly modeling the impact of the network load on the loss rates and thus the delays.

As in [12], we model each node using a discrete time $G/G/1$ queue. However, unlike [12], we propose and use a detailed model which accounts for the effect of £nite load on the collision rates and the queue utilization. This improved characterization allows a more accurate model for the service time distribution to be developed. Our model provides closed form expressions for the queue length in the presence of arbitrary arrival patterns, packet size distributions and network load. The model accounts for the collision avoidance and exponential backoff mechanism of 802.11, the delays in the channel access due to other nodes transmitting and the delays caused by collisions. The results obtained from this model have been veri£ed through extensive simulations and are shown to be signi£cantly better than those in [12].

The rest of the paper is organized as follows. In Section II we present the detailed queueing model and Section III presents the simulation results to verify the model. Finally, Section IV presents the concluding remarks.

II. Queueing Model for the 802.11 DCF

In this section we introduce a discrete time $G/G/1$ queue for modeling nodes in a random access network based on the 802.11 MAC. We assume a network with N nodes and using the DCF of IEEE 802.11 to schedule their transmissions and arbitrary packet arrival process and packet length distribution. We assume the use of RTS and CTS messages for channel reservation. The analysis can be easily extended for the cases where such messages are absent.

A. Modeling the Backoff Mechanism

In order to model the MAC layer queueing delays and losses, we £rst analyze the back-off mechanism associated with the exponential back-off mechanism of 802.11 MAC protocol's Collision Avoidance mechanism. In the following analysis, we denote the probability that an arbitrary packet transmission (i.e. an RTS transmission) results in a collision by p. The lower and upper bounds on the contention window associated with backoffs are denoted by CW_{min} and CW_{max} and we use the notation $m = \log_2(CW_{max}/CW_{min})$. Once a node goes into collision avoidance or the exponential back-off phase, we denote the number of slots that it waits beyond a DIFS period before initiating transmission by BC. This back-off counter is calculated from

$$BC = \text{int}\left(\text{rnd}() \cdot CW(k)\right) \tag{1}$$

where the function `rnd()` returns a pseudo-random number uniformly distributed in $[0,1]$ and $CW(k)$ represents the contention window after k unsuccessful transmission attempts. Note that in case the `int()` operation is done using a `ceil()` function, the effective range for BC becomes $1 \leq BC \leq CW(k)$ since the probability of `rnd()`$= 0$ is 0 assuming a continuous distribution. For the rest of this paper we assume that a `ceil()` function is used to do the `int()` operation.

The £rst attempt at transmitting a given packet is performed assuming a CW value equal to the minimum possible value of CW_{min} [7]. For each unsuccessful attempt, the value of CW is doubled until it reaches the upper limit of CW_{max} speci£ed by the protocol. Then, at the end of k unsuccessful attempts, $CW(k)$ is given by

$$CW(k) = \min\left(CW_{max}, 2^{k-1}CW_{min}\right) \tag{2}$$

Also, let the probability that a transmission attempt is unsuccessful, i.e., the probability of a collision be denoted by p. Then, the probability that $CW = W$ is given by

$$\Pr\{CW = W\} = \begin{cases} p^{k-1}(1-p) & \text{for } W = 2^{k-1}CW_{min} \\ p^m & \text{for } W = CW_{max} \end{cases} \tag{3}$$

where $k \leq m$. Note that the second case ($W = CW_{max}$) includes all cases where the number of collisions is greater than m. The probability that back-off counter $BC = i$, $1 \leq i \leq CW_{max}$, is then given by

$$\Pr\{BC = i\} = \begin{cases} \left[\sum_{k=0}^{m-1} \frac{p^k(1-p)}{2^k CW_{min}} + \frac{p^m}{CW_{max}}\right] & 1 \leq i \leq CW_{min} \\ \left[\sum_{k=j}^{m-1} \frac{p^k(1-p)}{2^k CW_{min}} + \frac{p^m}{CW_{max}}\right] & 2^{j-1}CW_{min} + 1 \leq i \leq 2^j CW_{min} \\ \frac{p^m}{CW_{max}} & 2^{m-1}CW_{min} + 1 \leq i \leq CW_{max} \end{cases} \tag{4}$$

In [10], [11] the collision probability p was derived for the saturated network case where each node always has a packet to send and each incoming packet is immediately backlogged. In this paper, we extend the model to obtain an expression for collision probabilities in the general case. From [10], [11], the average backoff window in the saturated case is given by

$$\overline{W} = \frac{1 - p - p(2p)^m}{1 - 2p} \frac{CW_{min}}{2} \tag{5}$$

Now consider a network with N nodes operating in discrete time where the packet arrival rate at each node is λ packets per slot, the channel service rate is μ packets per slot and the queue utilization at a node is denoted by ρ. Consider a tagged node which transmits in a given slot. Now, a collision occurs if one or more of the remaining $N-1$ nodes also transmit in this slot. Then, letting $P[NT]$ denote the probability that a node does not transmit in a slot, we have

$$p = 1 - P[NT]^{N-1} \tag{6}$$

Now, with using QE to represent queue not empty and QNE queue not empty for ease of notation, $P[NT]$ is given by

$$\begin{aligned} P[NT] &= P[NT \mid QE]P[QE] + P[NT \mid QNE]P[QNE] \\ &= 1 \cdot (1-\rho) + \rho P[NT \mid QNE]. \end{aligned}$$

Note that a queue is non-empty in a slot either if if is backlogged or if a new arrival occurs in that slot while the queue was empty. Now, considering the fact that we are interested in stable queues and backoff slots are two orders of magnitude smaller than typical data packet lengths, the probability of the latter case is quite small. Also, a backlogged queue will not transmit in a slot with probability $(\overline{W}-1)/\overline{W}$. Then, $P[NT \mid QNE]$ can be approximated by $(\overline{W}-1)/\overline{W}$. Consequently,

$$P[NT] = (1-\rho) + \rho\frac{\overline{W}-1}{\overline{W}} = 1 - \frac{\rho}{\overline{W}} \tag{7}$$

and combining Equations (5), (6) and (7) the loss rate p is given by

$$p = 1 - \left(1 - \rho\frac{(1-2p)}{1-p-p(2p)^m}\frac{2}{CW_{min}}\right)^{N-1}. \tag{8}$$

To determine ρ, we now characterize the average time to serve a packet. For each packet, the node spends $\overline{W}$ slots in backoff. Also, with the long term fairness of exponential backoff, in the case where all nodes have the same traf£c arrival rates, on an average $\rho(N-1)$ transmissions from other nodes occur between two transmissions from the tagged node. This contributes $\rho(N-1)T_S$ slots to the service time where T_S is the average length of a packet in units of backoff slots. The contribution due to the collisions of packets of other nodes is given by $\rho(N-1)pT_C/(1-p)$ where T_C is the time of a collision in units of slots. Finally, adding the time to transmit the packet of the tagged node and any collision that it may have, we get,

$$\frac{1}{\mu} = \rho(N-1)\left[T_S + T_C\frac{p}{1-p}\right] + \overline{W} + T_S + T_C\frac{p}{1-p} \tag{9}$$

Then using the fact that $\rho = \lambda/\mu$ for a stable system, we can substitute ρ in Equation (8) to obtain p by solving the following equation

$$p = \frac{\lambda\left[T_S + T_C\frac{p}{1-p}\right]}{1 - \lambda(N-1)\left[T_S + T_C\frac{p}{1-p}\right] - \frac{\lambda(1-p-p(2p)^m)}{1-2p}\frac{CW_{min}}{2}} \tag{10}$$

B. The Queueing Model

To obtain the delays experienced by packets, each node is modeled as a discrete time $G/G/1$ queue. The unit of time or the slot length corresponds to the length δ of a backoff slot. While packet lengths in real networks are not integral multiples of slot times, since δ is of the order of 20μsec, the error introduced by the discretization is quite small. We denote by $a(n)$ the probability that n messages arrive in a given slot at a given node with the corresponding probability generating function (pgf) $A(z)$. Also, $b(n)$ denotes the the probability that the service time of a packet takes n slots with the corresponding pgf $B(z)$. Now, $b(n)$ depends on the number of nodes contending for the channel as well as the packet length distribution and we now characterize its distribution.

The service time of a packet can be broken into two components: (1) the time till the node successfully accesses and reserves the channel for use and (2) the time required to transmit the packet (determined by the packet length distribution). To characterize the £rst component, we refer to Fig. 1. Between any two successful transmissions by a tagged node, other nodes may successfully transmit a number of packets or may be involved in a number of collision, each of which add to the channel access time of the tagged node.

We £rst characterize the number of backoff slots that the tagged node has to wait between two successful transmissions. When a packet comes in and £nds that the system is empty it is transmitted without going into backoff, and thus the probability that the number of backoff slots, BO, is zero is given by $P[BO = 0] = (1 - \rho)(1 - p)$. Now with probability ρ the packet goes into backoff at least once. Now, note that if the tagged node successfully transmits the packet in its £rst attempt (with probability $1 - p$) the number of backoff slots is uniformly distributed between $1, \cdots, CW_{min}$. In case of a successful transmission after a single collision (with probability $p(1-p)$), the pmf of the number of backoff slots is obtained through $U_{1,CW_{min}} * U_{1,2CW_{min}}$ where $U_{a,b}$ denotes a uniform distribution between a and b and $*$ represents the convolution operation. Following the same procedure for a sequence of successive collisions for the same packet, the probability the tagged node experiences i backoff slots, $i > 0$, is given by

$$
\begin{aligned}
P[BO = i] = \rho\Big[& (1 - p)U_{1,CW_{min}}(i) + p(1 - p) \\
& \Big[U_{1,CW_{min}} * U_{1,2CW_{min}}(i)\Big] + \cdots + p^m(1 - p) \\
& \Big[U_{1,CW_{min}} * U_{1,2CW_{min}} * \cdots * U_{1,2^m CW_{min}}(i)\Big] \\
& + p^{m+1}(1 - p)\Big[U_{1,CW_{min}} * \cdots * U_{1,2^m CW_{min}} * \\
& U_{1,2^m CW_{min}}(i)\Big] + \cdots \Big] \qquad (11)
\end{aligned}
$$

with the corresponding pgf $BO(z)$. Note that the maximum number of retransmission attempts allowed for each packet is governed by the long retry count (SLRC) (short retry count (SSRC) for transmissions without the RTS-CTS exchange) which forms the limit on the summation above. However, its effect may be neglected since the term $p^k(1 - p)$ becomes negligibly small as k increases.

Since the average window size is $\overline{W}$ (Equation (5)) and a queue is active with probability ρ, the probability that a node attempts a transmission in an arbitrary slot is given by $\rho/\overline{W}$. Then, the probability that a given slot is active, q, is given by

$$
q = 1 - \left(1 - \frac{\rho}{\overline{W}}\right)^N \qquad (12)
$$

Then, given that the tagged node experiences i backoff slots before it successfully transmits a packet, the pmf of the number of active slots within the backoff slots is given by

$$
P[j \text{ slots active}|BO = i] = \begin{pmatrix} i \\ j \end{pmatrix} q^j(1 - q)^{i-j} \qquad (13)
$$

for $j = 0, \cdots, i$. Unconditioning on i, we have

$$
P[j \text{ slots active}] = \sum_{i=j}^{\infty} \begin{pmatrix} i \\ j \end{pmatrix} q^j(1 - q)^{i-j} P[BO = i] \quad (14)
$$

Also, the probability that a slot results in a collision given that it is active, q_c, is given by

$$
q_c = \frac{1 - \left(1 - \frac{\rho}{\overline{W}}\right)^N - \frac{\rho N}{\overline{W}}\left(1 - \frac{\rho}{\overline{W}}\right)^{N-1}}{1 - \left(1 - \frac{\rho}{\overline{W}}\right)^N} \qquad (15)
$$

and thus the probability that out of j active slots k result in collisions is given by

$$
P[k \text{ collisions}|j \text{ active slots}] = \begin{pmatrix} j \\ k \end{pmatrix} q_c^k(1 - q_c)^{j-k} \quad (16)
$$

Now, each collision adds T_C slots to the service time where $T_C = DIFS + \tau_{RTS}$ with τ_{RTS} being the time required to transmit a RTS packet. Note that in situations where RTS-CTS packets are not used to reserve the channel, the duration of a collision is given by $T_C = DIFS + \tau_{pkt}$ where τ_{pkt} is the packet transmission time. Also, each successful transmission by other nodes between the two successful transmissions of the tagged node adds a time proportional to the packet length of the transmitted packet to the service time at the tagged node. In our analysis we allow for general packet length distributions and the probability that a packet transmission takes n slots (which is dependent on the packet length and the channel rate) is denoted by $l(n)$ with the corresponding pgf $L(z)$. Then, the contribution of j successful transmissions to the service time of the tagged node is given by

$$
P[\sum^{j} \text{pkt time} = i] = l * l * \cdots * l(i) = l^{(j)}(i) \qquad (17)
$$

where $l^{(j)}()$ represents the $j-$fold convolution of $l(n)$. The contribution of the successful transmissions of the other competing stations and the collisions, X, to service time of the tagged node is then given by

$$
P[X = n] = \begin{cases} \begin{pmatrix} j \\ k \end{pmatrix} q^k(1 - q)^{j-k}\cdot & n = kT_C + i \\ l^{(j-k)}(i)P[\text{SA} = k] & \\ 0 & \text{otherwise} \end{cases} \qquad (18)
$$

where $P[\text{SA} = k]$ represents the probability that there are k active slots and is given by Equation (14). The above expression evaluates the probability of the event where there are k slots active between two transmissions from the tagged node, j of which result in collisions contributing kT_C slots to the service time while the $k - j$ successful transmissions contribute i slots. Note that the above expression needs to be

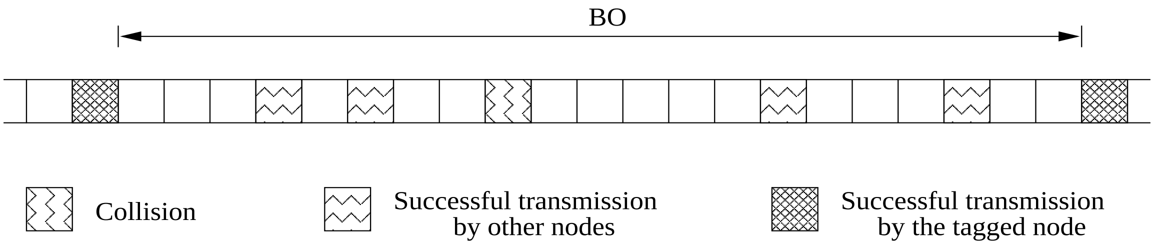


Fig. 1. Interleaving of transmissions and collisions contributing to the service time.

evaluated for all possible values of i, j and k which result in a given value of n. The pgf of the £nal service time, $B(z)$, which comprises of the backoff slots (BO), the delay due to other stations transmitting (X) and the length of the packet to be served (l) is then given by

$$B(z) = BO(z)X(z)L(z) \tag{19}$$

Using standard discrete time queueing theory [1], the pgf of the system occupancy of the $G/G/1$ queue at random slot boundaries (beginning of a slot), $U(z)$, is given by

$$U(z) = [1 - A'(1)B'(1)] \frac{(z-1)B(A(z))}{z - B(A(z))} \tag{20}$$

and the pgf of the integer part of the system time (where system time is de£ned as the total time spent in the system from the arrival instant to the service completion time) can be shown to be

$$V_{int}(z) = \frac{[1 - A'(1)B'(1)]\,(z-1)B(z)\,[1 - A(B(z))]}{A'(1)\,[1 - B(z)]\,[z - A(B(z))]} \tag{21}$$

Allowing arrivals to occur at any point in the slot, we denote the distance of the arrival point from the start of the slot by F with mean $\overline{F}$. This adds a fractional component to the system time of $V_{frac} = 1 - \overline{F}$. The total system time is then given by $V = V_{int} + V_{frac}$ whose mean can be expressed as

$$\overline{V} = 1 - \overline{F} + B'(1) + \frac{[A'(1)]^2\,B''(1) + A''(1)B'(1)}{2\,[1 - A'(1)B'(1)]} \tag{22}$$

The average queue size at each node can then be obtained using Little's law and is given by $\overline{Q} = A'(1)\overline{V}$. Equation (22) can now be solved to obtain the number of nodes that can be supported for arbitrary arrival traf£c patterns while providing a speci£ed delay guarantee.

III. Simulation Results

To validate our analytic model, we conducted extensive simulations using the simulator *ns-2* [6] for different network topologies, number of nodes as well as the load on the network. In this section, we report on our simulation results for the case of 10 and 20 nodes and omit the others since they are similar. The simulations for the results reported in this section were carried out for a rectangular region of 1500×500 meters and the nodes were randomly distributed over this region. The routing protocol used for the simulations was Ad-hoc On-demand Distance Vector routing (AODV) [9] and we also veri£ed our results for routing using Destination Sequenced Distance Vector (DSDV) [8]. The interface queues at each mode used a Droptail policy and the interface queue length was set at 50 packets. All sources and receivers have an omni-directional antenna of height 1.5m with transmitter and receiver gains of 1 each. The simulations were run for a simulated time of 1800 seconds. All other parameter settings for the physical and MAC layers for these simulations are given in Table 1.

Physical Layer		802.11 MAC	
Propagation	2 ray gnd	RTS size	44 bytes
Channel	Wireless	CTS size	38 bytes
Rx Threshold	3.652e-10	DIFS	50 μsec
Bandwidth	2 Mbps	SIFS	10 μsec
Frequency	914 MHz	Slot size	20 μsec
Loss Factor	1.0		

TABLE I
Simulation Settings

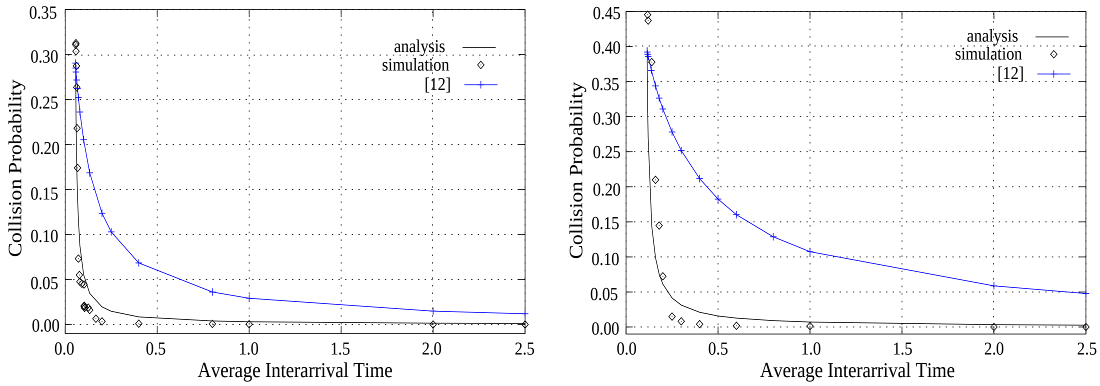
Each node was the source for one ¤ow as well as the sink for another ¤ow. Thus the 10 node case corresponds to 10 ¤ows while the 20 node case had 20 active ¤ows. The arrival process at each node, ($a(n)$), was assumed to follow the distribution

$$a(n) = \begin{cases} 1-p & n = 0 \\ p & n = 1 \end{cases} \tag{23}$$

resulting in an average inter-arrival time of $1/p$. The sources used UDP as the transport protocol and the packet sizes were assumed to be 1000 bytes.

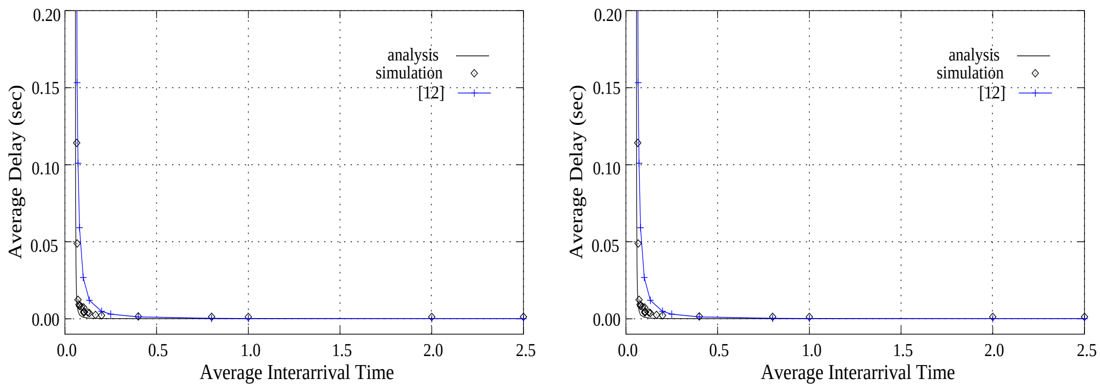
In Figure 2 we compare the simulation results for the collision probabilities with those obtained from our analysis. We also the corresponding analytic results for the collision rates from [12]. We note that our results are a large improvement over existing results and match quite well with the simulations. We also see that while for the 10 node case we have a good match with the simulation results, for the 20 node case we have a little deviation.

Figure 3 compares the simulation and analytic results for the average delays for the 10 and 20 node cases along with the results obtained from the analysis of [12]. For both scenarios, we see the close match between the analytic and the simulation results with the analysis of this paper being more closer to the simulations as compared with those of [12]. Also, as expected, the system saturates more quickly for the 20 node cases at approximately half the load of the 10 node case. Similar results were also obtained for other topologies and network sizes, validating the analytic model for the delay in an 802.11 based



(a) 10 nodes (b) 20 nodes

Fig. 2. Comparison of the collision probabilities.



(a) 10 nodes (b) 20 nodes

Fig. 3. Comparison of the average packet delays.

network.

IV. Conclusions

In this paper we present an improvement to the queueing model presented in [12] to evaluate the performance of the IEEE 802.11 MAC in terms of its delays and queue lengths and evaluate its capability to support delay sensitive traf£c. The improvement is achieved by explicitly modeling the impact of £nite loas on the loss rates and the queue utilizations. The queueing model for each node in the network accounts for the intricacies of the MAC protocol and its behavior as a function of the number of users in the network. Each node is modeled as a discrete time $G/G/1$ queue and we allow for arbitrary number of nodes, arrival patterns and packet size distributions. Our analytic results have been veri£ed using extensive simulations and show signi£cant improvements over the results of [12].

References

[1] H. Bruneel and B. Kim, *Discrete-Time Models for Communication Systems Including ATM*, Kluwer Academic Publishers, Boston, 1993.

[2] F. Cali, M. Conti and E. Gregori, "IEEE 802.11 protocol: design and performance evaluation of an adaptive backoff mechanism," *IEEE journal on selected areas of Communications*, vol. 18, no. 9, pp. 1774-1786, September 2000.

[3] H. Chhaya and S. Gupta, "Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocols," *Wireless Networks*, vol. 3, pp. 217-234, 1997.

[4] S. Choi, J. del Prado, S. Nandgopalan and S. Mangold, "IEEE 802.11e contention-based channel access (EDCF) performance evaluation," *Proceedings of IEEE ICC*, Anchorage, Ak, May 2003.

[5] C. Coutras, S. Gupta and N. Shroff, "Scheduling of real-time traf£c in IEEE 802.11 wireless LANs," *Wireless Networks*, vol. 6, no. 6, pp. 457-466, November 2000.

[6] K. Fall and K. Varadhan, editors, "ns notes and documentation," The VINT Project, UC BERKELY, LBL, USC/ISI, and Xerox PARC, November 1997.

[7] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Speci£cations*, IEEE standards 802.11, January 1997.

[8] C. Perkins and P. Bhagwat, "Highly dynamic Destination Sequenced Distance-Vector routing (DSDV) for mobile computers," *Proceedings of ACM SIGCOMM*, pp. 234-244, August 1994.

[9] C. Perkins and E. Royer. "Ad-hoc On-Demand Distance Vector routing," *Proceedings of IEEE Workshop on Mobile Computing Systems and Applications*, pp. 90-100, New Orleans, LA, February 1999.

[10] Y. Tay and K. Chua, "A capacity analysis for the IEEE 802.11 MAC protocol," *Wireless Networks*, vol. 7, no. 2, pp. 159-171, March, 2001.

[11] O. Tickoo and B. Sikdar, "On the impact of IEEE 802.11 MAC on traf£c characteristics," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, pp. 189-203, February 2003.

[12] O. Tickoo and B. Sikdar, "Queueing analysis and delay mitigation in IEEE 802.11 random access MAC based wireless networks," *Proceedings of IEEE IMFOCOM*, Hong Kong, China, March 2004.

[13] M. Veeraraghavan, N. Cocker and T. Moors, "Support of voice services in IEEE 802.11 wireless LAN," *Proceedings of IEEE INFOCOM*, pp. 488-497, Anchorage, Alaska, April 2001.

[14] J. Weinmiller, H. Woesner, J.-P. Ebert and A. Wolisz, "Modi£ed backoff algorithms for DFWMAC's distributed coordination function," *Proceedings of the 2nd ITG Fachtagung Mobile Kommunikation*, pp. 363-370, Neu-Ulm, Germany, September 1995.