An Analytic Model for the Delay in IEEE 802.11 PCF MAC based Wireless Networks

Biplab Sikdar

Abstract— **In this paper, we present an analytic model for evaluating the queueing delays at nodes using the IEEE 802.11 Point Coordination Function (PCF) MAC for real time, delay sensitive traffic. We develop a queueing model to obtain closed form expressions for the expected delay at each node which accounts for arbitrary (but fixed) packet sizes, polling rates, channel rates and the order in which the nodes are polled. The model is then further extended to account for the delays when the nodes use power management, and for cases when not all nodes are served in a frame. Our analytical results are verified through simulations. The model is also extended to evaluate the number of nodes that can be supported by a base station while satisfying an arbitrary delay requirement at all nodes and can be used as a mechanism for admission control by the base station.**

Index Terms— **Wireless LAN, queueing analysis, modeling.**

I. Introduction

The IEEE 802.11 MAC [5] has become ubiquitous and gained widespread popularity as a layer-2 protocol for wireless local area and in-home networks. With increasing deployment, the services supported by such networks have started to migrate from the traditional data applications to various forms of interactive multimedia involving voice and video [4], [11], [13] transmissions as well as multiplayer network gaming [3]. Supporting these real-time applications requires that the MAC layer provide sufficient delay guarantees and the Point Coordination Function has been included in 802.11 to achieve this objective. This paper analytically characterizes the delays experienced with the 802.11 PCF and provides a framework for doing admission control in such networks by determining the maximum number of users that can be supported by the 802.11 PCF while satisfying a given delay constraint.

While PCF is simpler and can provide stricter Quality of Service (QoS) guarantees as compared to the more widely deployed Distributed Coordination Function (DCF), it has received considerably lesser attention and deployment. While it is partly due to the fact that PCF is optional in the standards while DCF is not, other contributing reasons are overhead issues in extremely large networks and the fact that parameters settings and other specifications in the standards have been left largely open. Detailed studies and analytic tools to quantify the effect of various network settings on PCF's performance are necessary to to facilitate the widespread deployment and harness the performance benefits of the PCF and this paper is a step in that direction. Also, though DCF provides satisfactory performance in scenarios with limited number of users such as homes, in more crowded scenarios, DCF fails to provide the requirements of delay sensitive applications [10]. In these scenarios, PCF is a viable protocol choice due to its ability to provide strict delay guarantees (note that the Enhanced DCF in the 802.11e specifications provides only QoS differentiation, not delay guarantees). As public wireless hot spots and voice and video streaming applications become ubiquitous, MAC layer protocols with strict QoS support like PCF will become more important. Finally, PCF can help avoid denial of service attacks by outside users (and the associated delays) which can be hard to detect and control in DCF based systems.

The delay characteristics of the 802.11 PCF has been extensively studied using simulations [2], [12]. The effect of different polling strategies on PCF performance is presented in [14] while the performance of video transmission with PCF has been investigated in [8], [9]. However, these are all simulation studies, and to the best of our knowledge, no detailed queueing or analytic models for 802.11 PCF exists in literature. Also, to the best of our knowledge, no models exist to evaluate the delays in the case of nodes using power management. This paper addresses this issue by proposing a queueing model and closed form expressions for the expected delay at each node. In [11] the maximum number of nodes that can be supported by a 802.11 PCF network to support voice transmissions has been evaluated using simulations. In contrast, we use the expressions from our analytic model to provide a framework for doing admission control by the base station in order to support the given delay constraints.

This paper first proposes a detailed queueing model to evaluate the delays experienced by nodes using 802.11 PCF as the MAC protocol. Our model allows for arbitrary number of users in the network, their packet arrival rates and packet lengths and both unidirectional and bidirectional data transfers. The model evaluates the delays as a function of various 802.11 specific parameters like the superframe and beacon lengths, facilitating the estimation of the tradeoffs involving the values of these parameters and the system performance. Next, the paper provides a framework for determining the maximum number of stations that can be supported by a base station given a certain delay constraint and using it for admission control. Finally, the analysis is extended to the case where the nodes employ power management strategies where nodes may stay in the active mode (AM) or switch to the power save (PS) or doze mode in order to save energy. Our analysis has been validated using simulations.

Manuscript received August 24, 2005; revised December 28, 2005 and March 24, 2006; accepted March 28, 2006. The associate editor coordinating the review of this paper and approving it for publication was M. Zorzi. This work was supported by NSF under grant 0313095 and by Intel Corporation.
B. Sikdar is with the Rensselaer Polytechnic Institute, Troy, NY USA (e-mail: bsikdar@ecse.rpi.edu).
Digital Object Identifier 10.1109/TWC.2007.05666.

The rest of the paper is organized as follows. In Section II we give a brief overview of the 802.11 PCF. Section III presents our delay model for the unidirectional data transfer case while Section IV extends it for the bidirectional case. In Section V we present the admission control strategy and Section VI and VII extend the analysis for power management and short frame durations. Finally, Section VIII presents the validation results and Section IX presents the concluding remarks.

II. Background

In addition to the physical layer specifications, the IEEE 802.11 standard [5] specifies two methods for medium access: DCF and the PCF. While DCF uses a distributed mechanism for channel access and is not the focus of the paper, in PCF the nodes are polled by a “master” residing within the base station. The channel access mechanism alternates between the DCF and PCF modes when PCF is implemented. The duration of time the DCF is used for channel access is termed the contention period (CP) and the polled duration is called the contention-free period (CFP). The lengths of the CP and the CFP is controlled explicitly by the contention free period repetition interval (CFPri). We call a CFPri duration where the PCF and DCF alternate a “superframe”.

Each CFP begins with a beacon frame and the CFPs occur at a defined repetition rate as determined by the CFPrate parameter. The length of the CFP is controlled by the PC, with the maximum duration specified by the value of the CFPMaxDuration parameter, and the remainder of the frame spent for DCF. With PCF, the access to the channel is determined centrally by the base station, usually referred to as the Point Coordinator (PC) and provides a contention free transfer service. The PC gains control of the medium at the beginning of the CFP and maintains control for the entire CFP by waiting for a shorter time between transmissions than the stations using the DCF access mode. All stations other than the PC set their NAVs to the CFPMaxDuration at the start of each CFP. The PC transmits a CF-End or CF-End+ACK frame at the end of each CFP and on receiving either of these frames, a station resets its NAV. During the CFP, the base station polls the nodes for a *single* pending frame transmission according to a list ordering of their association with the base station, known as the polling list. The PC starts CF transmissions a SIFS interval after the beacon frame by sending a CF-Poll (no data), Data or Data+CF-Poll frame. If a station receives a CF-Poll (no data) frame from the PC, the station can respond to the PC after a SIFS interval with a CF-ACK (no data) or a Data+CF-ACK frame. If the PC receives a Data+CF-ACK frame from a station, it can send a Data+CF-ACK+CF-Poll frame to a different station where the CF-ACK part is used to acknowledge receipt of the previous data frame. If the PC transmits a CF-Poll (no data) frame and the destination station does not have any data to transmit, the station sends a Null Function (no data) frame back to the PC. If the PC fails to receive an ACK for a transmitted data frame, it waits for a PIFS interval and moves on to the next station in the polling list. Figure 1 shows the transmission of frames between the PC and stations. The figure shows the piggybacking of poll and ACK packets with data by the PC and the data and ACK by the stations as well as an unresponsive node (node 3) which does not respond to polls from the PC.

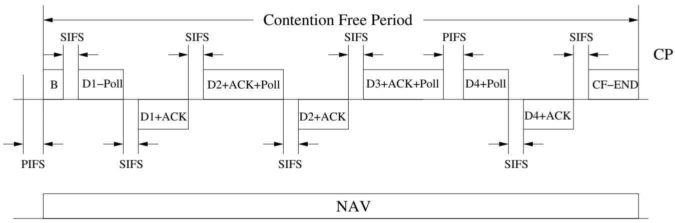

Fig. 1. PC to station frame transmissions in PCF.

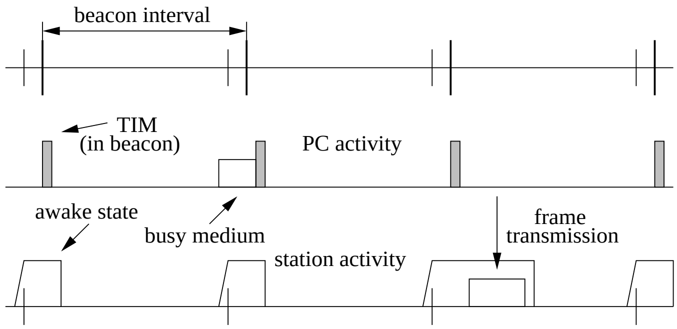

Fig. 2. Power management operation in IEEE 802.11 PCF.

The IEEE 802.11 standard also specifies a power management strategy wherein a station may either be in the active mode where it is fully powered and may receive frames at any time or be in the power save mode. In the PS mode, the station stays in the doze state where it is unable to transmit or receive and consumes very low power. Also, the station enters the awake state to receive selected beacons and transmit and receive frames. Stations inform the PC about their state using the Power Management bits within the Frame Control field of transmitted frames. The PC buffers frames destined for stations in the PS mode and stations with buffered frames are identified in a traffic indication map (TIM) which is included in each beacon generated by the PC. On receiving a TIM indicating buffered frames for it, a station stays awake until the buffered frame is received. If the More Data field in the Frame Control field of the last frame from the AP indicates more traffic is buffered, the node may enter the doze state during the contention period and wake again at the start of the next CFP. Figure 2 illustrates PC and station activity with power management where we show the TIM transmissions by the PC (in the middle row) every beacon interval (shown by the time axis in the top) and the activity of a node in the PS mode (bottom). The node stays awake only for the TIM transmission when it does not have any data to send or receive (for example in the first two TIMs) but stays awake for longer if any data is to be transferred (third TIM).

III. Analysis

In this section we present our model to evaluate the delays experienced by stations using the PCF mode to transmit their data and power management is not used and PCF and DCF alternate in each superframe. We first introduce the notation used in this paper and our assumptions.

An arbitrary number of nodes, M, use the PCF mode to transmit their packets. The packet inter-arrival times at the i^{th} node are assumed to be exponentially distributed with rate λ_i, $1 \leq i \leq M$. The results obtained in this paper under this Poisson traffic assumption may be thought of as a lower bound on the delays obtained under more bursty and correlated traffic models. The packet arrival process is assumed to be independent of the departure process and the queue length. We denote the duration of the superframe by T_S, the duration of the beacon by B, the length of a polling duration by V and the expected length of a packet from the i^{th} polled node by L_i, $1 \leq i \leq M$. Note that we include the lengths of the SIFS and CF-Poll in V and SIFS and CF-ACK in L_i. The utilization of the i^{th} station is denoted by ρ_i. Note that since each polled stations gets to transmit once in every superframe, the service rate of the i^{th} station is $\mu_i = 1/T_S$, $1 \leq i \leq M$. The utilizations are thus given by $\rho_i = \lambda_i/\mu_i = \lambda_i T_S$. In the derivations presented in this paper, we assume that the arrival rates and packet lengths are the same at each node, i.e., $\lambda_i = \lambda$, $\forall i$ and $L_i = L$, $\forall i$ and thus $\rho_i = \rho = \lambda T_S$, $\forall i$. We assume that in each CFP, at most one packet is transmitted by a node. This limited-1 polling mechanism ensures that the CFP does not dominate the superframe duration and nodes using DCF also get a fair chance to transmit their packets in a superframe.

We evaluate the expected delay experienced by an arbitrary packet arriving at the i^{th} polled node. We break the analysis into two parts: (1) the delay experienced when the packet arrived at an empty queue and (2) when the arrival occurred at a non empty queue. The probability that an arbitrary arrival finds the queue empty, $P[\text{EQ}]$, is given by

$$P[\text{EQ}] = 1 - \rho = 1 - \lambda T_S \tag{1}$$

and the probability that an arbitrary arrival finds the queue busy, $P[\text{NEQ}]$, is thus

$$P[\text{NEQ}] = 1 - P[\text{EQ}] = \rho = \lambda T_S. \tag{2}$$

Since the arrivals at each queue are independent and the probability that a queue is busy is given by ρ, the number of active queues, j, at any instant of time, out of M queues follows a Binomial distribution and is given by

$$P[j \text{ active}] = \binom{M}{j} \rho^j (1-\rho)^{M-j} \qquad j = 0, 1, \cdots, M \tag{3}$$

In this section we consider the case where there is only upstream traffic from the stations to the base station. The model for case of bidirectional traffic is presented in Section IV.

A. Arrivals at an Empty Queue

Consider an arrival at the i^{th} polled station whose queue is currently empty and we call this arrival the "tagged arrival". If this station has not yet been polled in the current superframe when the packet arrives, the packet gets served in the current superframe. Otherwise, the packet gets served in the following superframe. Now, it is well known that with exponential arrivals independent of the departure process in a slotted departure system (for example a classical M/D/1 queue), an

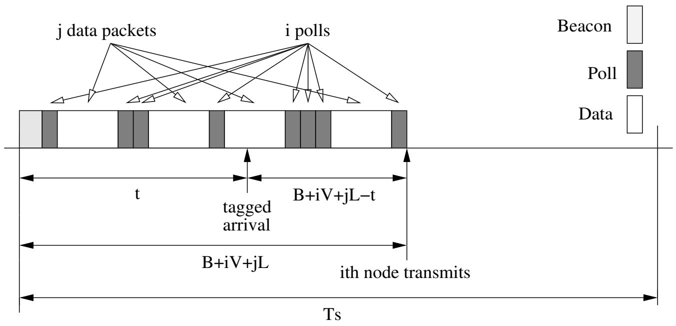

Fig. 3. Packet delays when arriving packet finds an empty queue.

arrival is equally likely to occur anywhere in a slot or frame [7], [1]. In our case, *given* that an arrival occurs in a given superframe, the arrival instance is thus uniformly random variable over $[0, T_s]$ *relative* to the start of the superframe. Consider the case where j of the $i-1$ nodes polled before the i^{th} node in a given superframe have data to send. In this case, a period of $B + (i-1)V + jL$ seconds elapse in the superframe before the i^{th} node is polled and $B + iV + jL$ seconds elapse before it has to reply to the poll. Thus if the tagged arrival occurs in this duration, it gets served in this superframe. Otherwise it waits for the next superframe.

Since the arrival instant, t, of any packet relative to the start of its superframe is uniformly distributed ($U[0, T_S]$), the probability that the tagged packet arrived at node i in the first $B + iV + jL$ seconds is given by

$$P[t \leq B + iV + jL] = \frac{B + iV + jL}{T_S}. \tag{4}$$

In this case (which we call case C1), the packet waits till the i^{th} node is polled and is then transmitted, as shown in Figure 3. The time the packet waits before it begins service, $X_{i,j,C1}$, is thus $X_{i,j,C1} = B + iV + jL - t$ after which it receives service for another L seconds before departing the system. We will now characterize the distribution of $X_{i,j,C1}$. The probability distribution function (PDF) of t given that the arrival occurred in the first $B + iV + jL$ seconds of the superframe is given by

$$\begin{aligned} P[t \leq \tau \mid t \leq B + iV + jL] &= \frac{P[t \leq \tau, t \leq B + iV + jL]}{P[t \leq B + iV + jL]} \\ &= \frac{\tau}{B + iV + jL} \end{aligned} \tag{5}$$

which is an Uniform distribution in the range 0 to $B + iV + jL$. Now, note that if a random variable Y is uniformly distributed in the range 0 to a, then the random variable $a - Y$ is also uniformly distributed in the range 0 to a. Following this observation, since the conditional PDF of t is uniformly distributed in the range 0 to $B + iV + jL$, the conditional PDF of $X_{i,j,C1} = B + iV + jL - t$ is also an Uniform distribution in the range 0 to $B + iV + jL$, i.e., $U[0, B + iV + jL]$. The expected value of $X_{i,j,C1}$ is thus

$$E[X_{i,j,C1}] = E[U[0, B + iV + jL]] = \frac{B + iV + jL}{2}. \tag{6}$$

In the case where the packet does not arrive in the first $B + iV + jL$ seconds of the superframe (which we call case C2), i.e. $t > B + iV + jL$, the packet has to wait till the remaining part of the superframe ($T_S - t$) is over and node i is polled in the following superframe. The PDF of t given that the arrival occurred after the first $B + iV + jL$ seconds of the superframe is given by

$$P[t \leq \tau \mid t > B+iV+jL] = \frac{\tau - B + iV + jL}{T_S - B + iV + jL} \quad (7)$$

which is an Uniform distribution in the range $B+iV+jL$ to T_S, i.e., $U[B+iV+jL, T_S]$. Again we note that if a random variable Y follows a Uniform distribution $U[a,b]$, then $b-Y$ is uniformly distributed in the range 0 to $b-a$, i.e. $U[0, b-a]$. Thus the duration of the remaining part of the superframe, $T_S - t$, is also uniformly distributed and is $U[0, T_S - B - iV - jL]$.

In the following superframe, if there are k nodes with data to send among the $i-1$ nodes polled before the i^{th} node, the tagged packet has to wait for $B + iV + kL$ seconds before its service begins. Since the probability that there are k nodes with data among $i-1$ nodes follows a Binomial distribution as given in Eqn. (3), the probability mass function (pmf) of this waiting time, X_{FR}, is given by

$$P[X_{FR}=x] = \begin{cases} \binom{i-1}{k}\rho^k(1-\rho)^{i-k-1} & x = B+iV+kL \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

with $0 \leq k \leq i-1$ and the expected value of X_{FR} is given by

$$E[X_{FR}] = B + iV + (i-1)\rho L. \quad (9)$$

Thus the amount of time, $X_{i,j,C2}$, before the packet begins its service is $X_{i,j,C2} = T_S - t + X_{FR}$. The expected value of $X_{i,j,C2}$ is thus

$$\begin{aligned} E[X_{i,j,C2}] &= E[U[0, T_S - B - iV - jL]] + E[X_{FR}] \\ &= \frac{T_S - B - iV - jL}{2} + B + iV + (i-1)\rho L \quad (10) \end{aligned}$$

To find the expected waiting time in the systems when an arrival occurs at an empty queue given that j of the $i-1$ nodes before the i^{th} node send data in the current superframe, we combine the waiting times of the above two cases. This expected waiting time, $D_{i,j,EQ}$, is given by

$$D_{i,j,EQ} = E[X_{i,j}] + L \quad (11)$$

with

$$E[X_{i,j}] = E[X_{i,j,C1}]P[C1] + E[X_{i,j,C2}]P[C2] \quad (12)$$

where $E[X_{i,j,C1}]$ and $E[X_{i,j,C2}]$ are given in Eqns. (6) and (10) respectively and $P[C1]$ and $P[C2]$ are the probabilities that the arrival occurs in the first $B + iV + jL$ seconds of the superframe or not, respectively. As discussed earlier in this section, these are given by

$$P[C1] = \frac{B + iV + jL}{T_S} \quad \text{and} \quad P[C2] = 1 - \frac{B + iV + jL}{T_S} \quad (13)$$

Putting these values in Eqn. (12), $E[X_{i,j}]$ can be simplified to

$$\begin{aligned} E[X_{i,j}] &= \frac{T}{2} + \frac{(B + iV + jL)^2}{T_S} - (B + iV + jL) \\ &\quad + E[X_{FR}]\frac{T_S - B - iV - jL}{T_S} \quad (14) \end{aligned}$$

which can now be used in Eqn. (11) to obtain $D_{i,j,EQ}$. The expected delay at the i^{th} node, $D_{i,EQ}$ is then obtained by unconditioning Eqn. (11) on j. Recall that j denotes the number of nodes among the $i-1$ polled ahead of node i had packets to send in an arbitrary superframe and has a Binomial pmf given in Eqn. (3) with $M = i-1$. Thus $D_{i,EQ}$ is given by

$$\begin{aligned} D_{i,EQ} &= \sum_{j=0}^{i-1} (E[X_{i,j}] + L) \binom{i-1}{j} \rho^j (1-\rho)^{i-j-1} \\ &= \frac{T_S}{2} + \frac{\rho L^2 (i-1)(1-\rho)}{T_S} + L \quad (15) \end{aligned}$$

B. Arrivals at a Non-Empty Queue

We now consider the case when an arbitrary arrival to the i^{th} polled node finds the queue non-empty and we denote the number of packets in the queue found by this packet by N_{NQ}. Consider again the case where j of the $i-1$ nodes polled before node i have packets to send in the current superframe. Then the probabilities of the events C1 and C2 in this case are given by

$$P[C1] = \frac{B+iV+(j+1)L}{T_S} \qquad P[C2] = 1 - \frac{B+iV+(j+1)L}{T_S} \quad (16)$$

where the $j+1$ terms comes from the fact that in addition to the j nodes, node i is also transmitting.

In case the i^{th} node has not yet been served when the tagged packet arrives (case C1), one of the N_{NQ} packets currently waiting in the queue at node i gets served during this superframe. If we denote the instant of the tagged packet's arrival in the superframe by t, it has to wait for $T_S - t$ seconds before the current superframe ends. The tagged packet then has to wait for another $N_{NQ} - 1$ packets to depart, with one departure in one superframe or $(N_{NQ}-1)T_S$ seconds before the start of the superframe where it receives service. We denote the waiting time in the final superframe by X_{FR} and its distribution and expectation is given in Eqn. (8) and (9) respectively. Thus the total time before the packet begins service in this case, $X_{i,j,C1}$, is given by $X_{i,j,C1} = T_S - t + (N_{NQ} - 1)T_S + X_{FR}$.

Following the derivation in Eqn. (5), the PDF of t given that the arrival occurred in the first $B + iV + (j+1)L$ seconds has the Uniform distribution $U[0, B + iV + (j+1)L]$. Thus $T_S - t$ follows the Uniform distribution $U[T_S - B - iV - (j+1)L, T_S]$. The expected value of $X_{i,j,C1}$ is thus

$$E[X_{i,j,C1}] = E[T_S - t] + E[(N_{NQ} - 1)T_S] + E[X_{FR}]$$

In the case where the tagged arrival occurs after the i^{th} node has been served in the current round (case C2), at the end of the current superframe, there are still N_{NQ} packets ahead of the tagged packet. Thus at the end of a further $N_{NQ}T_S$ seconds, the superframe in which the tagged packet gets served starts. The total time before the packet begins service in this case, $X_{i,j,C2}$, is then given by $X_{i,j,C2} = T_S - t + N_{NQ}T_S + X_{FR}$. Now, following the derivation of Eqn. (7), the PDF of t given that the arrival occurred after the first $B + iV + (j+1)L$ seconds of the superframe has the Uniform distribution $U[B + iV + (j+1)L, T_S]$. Thus $T_S - t$ is also uniformly distributed and is $U[0, T_S - B - iV - (j+1)L]$. The expected value of $X_{i,j,C2}$ is thus

$$E[X_{i,j,C2}] = E[T_S - t] + E[N_{NQ}T_S] + E[X_{FR}]$$

Combining the two cases above, the expected waiting time at the i^{th} node, $D_{i,j,NEQ} = E[X_{i,j}] + L$, is given by

$$D_{i,j,NEQ} = \frac{T_S}{2} + E[N_{NQ}]T_S + E[X_{FR}] - B - iV - jL$$

Unconditioning the above equation on j and recalling that j follows the Binomial distribution of Eqn. (3) with $M = i-1$, the expected delay at the i^{th} node, $D_{i,NEQ}$ is given by

$$\begin{aligned} D_{i,NEQ} &= \sum_{j=0}^{i-1} D_{i,j,NEQ} \binom{i-1}{j} \rho^j (1-\rho)^{i-j-1} \\ &= \frac{T_S}{2} + E[N_{NQ}]T_S \end{aligned} \quad (17)$$

C. Overall Delay

The expressions for the delays of the previous two sections can now be combined to obtain the expression for the delay experienced by an arbitrary arrival. The expected packet delay at node i is given by

$$\begin{aligned} D_i &= D_{i,EQ}P[\text{EQ}] + D_{i,NEQ}P[\text{NEQ}] \quad (18) \\ &= \frac{T_S}{2} + \rho E[N_{NQ}]T_S + \left[\frac{\rho L^2(i-1)(1-\rho)}{T_S} + L\right](1-\rho) \end{aligned}$$

where $P[\text{EQ}]$, $P[\text{NEQ}]$, $D_{i,EQ}$ and $D_{i,NEQ}$ are given in Eqns. (1), (2), (15) and (17) respectively. Note however, that the expression $E[N_{NQ}]$ is the expected number of packets seen an arrival *given* that the queue is non-empty. The expected number in the queue seen by an arbitrary arrival, $E[N] = \sum_{i=1}^{\infty} iP[N = i]$ is related to $E[N_{NQ}]$ by

$$E[N_{NQ}] = \sum_{i=0}^{\infty} \frac{iP[N = i, \text{NEQ}]}{P[\text{NEQ}]} = \sum_{i=1}^{\infty} \frac{iP[N = i]}{\rho} = \frac{E[N]}{\rho} \quad (19)$$

where $P[N = i, \text{NEQ}]$ represents the joint probability that there are i packets in the queue and the queue is non-empty. Also from Little's Law $E[N] = \lambda D_i$. Thus we have $E[N_{NQ}] = \lambda D_i/\rho$ and substituting this in Eqn. (19) we have the final expression for D_i

$$D_i = \frac{1}{1-\lambda T_S}\left[\frac{T_S}{2} + \left(\frac{\rho L^2(i-1)(1-\rho)}{T_S} + L\right)(1-\rho)\right] \quad (20)$$

IV. Bi-Directional Traffic

We now consider the case where the base station also has traffic to send to the nodes. The derivations for this case are exactly the same as for the unidirectional case (and the details are thus omitted) except for one small change. Now instead of at most $i-1$ nodes which may transmit their packets before node i is polled, we can have at most $2i-1$ nodes. This is because the base station may also have traffic for each of the $i-1$ nodes as well as the i^{th} node before node i is polled. With this change, the pmf of the number of active queues, J before the i^{th} node follows the Binomial distribution

$$P[J = j] = \binom{2i-1}{j} \rho^j (1-\rho)^{2i-j-1} \qquad 0 \le j \le 2i-1 \quad (21)$$

Using this change in the results for the previous section, the expected delay at the i^{th} given that the arrival finds an empty queue, $D_{i,EQ}$ is given by

$$D_{i,EQ} = \frac{T_S}{2} + \frac{\rho L^2(2i-1)(1-\rho)}{T_S} + L \quad (22)$$

while the delay given that the packet arrives at a non-empty queue stays the same. Thus the final expression for the packet delay at the i^{th} node is given by

$$D_i = \frac{1}{1-\lambda T_S}\left[\frac{T_S}{2} + \left(\frac{\rho L^2(2i-1)(1-\rho)}{T_S} + L\right)(1-\rho)\right] \quad (23)$$

V. Admission Control

In this section, we use the expressions derived in the previous two sections to develop a very simple rule which can be used at a base station for doing admission control. The goal is to develop an expression which can calculate the number of nodes that the base station can accommodate for a given delay constraint that needs to be satisfied at each of the nodes. Note that while increasing the superframe size increases the number of nodes that may be polled and served in the superframe, it also increases the delay between successive polls at a given node.

Let each node have the constraint that its expected packet delay should be less than δ and the maximum number of hosts that can be supported by the base station be M_{UD} and M_{BD} for the unidirectional and bidirectional cases respectively. Thus for admission control, the base station just needs to evaluate M_{UD} and M_{BD} and add a new station to its polling list only if the current number of hosts is less than M_{UD} and M_{BD} for unidirectional and bidirectional cases, respectively. To evaluate M_{UD} and M_{BD} we first note that the expressions for the delays at the nodes for these two cases as given by Eqns. (20) and (23) respectively and are increasing functions of i, the order in which the node is polled by the base station. Thus the last station to be polled has the largest delay and if this node satisfies the delay constraint, all other nodes will do so too. Thus the delay at node M_{UD} and M_{BD} are given by

$$D_{M_{UD}} = \frac{1}{1-\lambda T_S}\left[\frac{T_S}{2} + \left(\frac{\rho L^2(M_{UD}-1)(1-\rho)}{T_S} + L\right)(1-\rho)\right] \quad (24)$$

$$D_{M_{BD}} = \frac{1}{1-\lambda T_S}\left[\frac{T_S}{2} + \left(\frac{\rho L^2(2M_{BD}-1)(1-\rho)}{T_S} + L\right)(1-\rho)\right] \quad (25)$$

respectively. Since in order to satisfy the delay constraint we should have $D_{M_{UD}} \le \delta$ and $D_{M_{BD}} \le \delta$, we have

$$M_{UD} = \left\lfloor \frac{[2\delta(1-\lambda T_S) - T_S - 2(1-\rho)L]\,T_S}{2\rho L^2(1-\rho)^2} \right\rfloor + 1 \quad (26)$$

$$M_{BD} = \left\lfloor \frac{[2\delta(1-\lambda T_S) - T_S - 2(1-\rho)L]\,T_S}{4\rho L^2(1-\rho)^2} + \frac{1}{2} \right\rfloor \quad (27)$$

Also, if it is required that all nodes on the polling list be polled in each superframe, the additional conditions $T_S \leq B + M_{UD}V + M_{UD}L$ and $T_S \leq B + M_{UD}V + 2M_{UD}L$ should also be satisfied for the unidirectional and bidirectional cases respectively.

VI. Delay with Power Management

We evaluate the expected delay experienced by an arbitrary packet arriving at the i^{th} polled node by extending the analysis in Sections III and IV. We break the analysis into two parts: (1) the delay experienced when the packet arrives while the station is in the PS mode and (2) when the arrival occurs while the station is in the AM. In the active mode, the operation of nodes is identical to the case where no power management is used and we can reuse the results of Sections III and IV to evaluate the delays in this case. We now evaluate the delays when the arrivals occur while the station is in the PS mode. In the discussion below we assume bidirectional traffic.

A station may go into the sleep mode if there are no packets queued up for it at the PC or in its own queue. Also, even if there are packets queued up, the station may go into the sleep mode at the end of the CFP and wake up again for the next beacon. In the latter case, the station gets served in every superframe and while there are energy savings, the delay stays the same as in the active mode. We thus include the analysis for this case in the analysis for the active mode. Thus in the rest of the paper, in the *sleep* mode, we only consider the scenario where the station goes into sleep because it has no outstanding packets queued up. In these cases, the station goes into the doze mode and wakes up for every S^{th} beacon and thus stays in the doze mode for $ST_S - B$ seconds before waking up to receive a beacon.

To evaluate the probability that an arbitrary arrival finds the node in the sleep mode, we now characterize the fraction of time a node stays in the sleep mode. We consider the behavior of the node at every beacon that it receives. If at the end of the beacon (and the TIM) a station does not have any packets to transmit and the PC does not have any packet queued up for it, the station goes in the sleep mode for a duration of $ST_S - B$ seconds. That is, in the duration ST_S seconds between the reception of two beacons, the node stays active for the beacon of duration B seconds and dozes for $ST_S - B$ seconds. Also, the probability that both the node's queue as well as its corresponding queue in the PC are empty is $(1-\rho)^2$. On the other hand, if at least one queue is non empty (with probability $1-(1-\rho)^2$), the node stays awake for the CFP and also for the next beacon and the fraction of time the node stays in the sleep mode in the T_S seconds between two successive beacon receptions is thus zero. The node's state is thus a two-state semi-Markov process corresponding to the active and sleep states with $(1-\rho)^2(ST_S - B)$ and $(1-(1-\rho)^2)(T_S + B)$ being expected time spent in each state, respectively. The probability that a station is in the *sleep* mode at any arbitrary instant of time is then given by fraction of the time spent in the sleep state:

$$P[\text{PS}] = \frac{(1-\rho)^2(ST_S - B)}{(1-\rho)^2 ST_S + (1-(1-\rho)^2)T_S} \tag{28}$$

and the probability that a station is in the active mode, $P[\text{AM}]$, is thus $P[\text{AM}] = 1 - P[\text{PS}]$. Note that a station may also enter the sleep mode in the middle of a CFP after it and the PC transmit packets to each other and their queues become empty. The equation above does not explicitly account for this and approximates this case by considering it equivalent to the sleep state entered just after the beacon transmission. However, simulation results show that the effect of this approximation is marginal.

A. Arrivals in the Power Save Mode

When an arrival occurs while the station is in the PS mode, it has to wait till the end of the sleep period before its service starts. The sleep period corresponding to each node is of duration $ST_S - B$ seconds. If we denote the instant of the tagged packet's arrival relative to the start of the sleep period by t, it has to wait for $ST_S - B - t$ seconds before the sleep period ends. In addition, it must wait for the other arrivals before it in the current sleep period to be served and if there are κ such packets, a waiting time of κT_S seconds is introduced before the start of the superframe where the tagged packet receives service. We denote the wait in the final superframe by X_{FR}. Thus the total time before the packet begins service in this case, $X_{i,PS}$, is given by $X_{i,PS} = ST_S - B - t + \kappa T_S + X_{FR}$.

Following the arguments in Section III, the arrival instant t of the tagged arrival relative to the start of the sleep period is uniformly distributed and is $U[0, ST_S - B]$. Thus $ST_S - B - t$ also follows the same uniform distribution and is $U[0, ST_S - B]$. Now, given that packet inter-arrival times are exponentially distributed, the pmf of the number of arrivals κ before the tagged packet is given by

$$P[\kappa = k \mid t] = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \tag{29}$$

and thus $E[\kappa \mid t] = \lambda t$ and. To evaluate the distribution of X_{FR}, we note that if there are k nodes with data to send among the $2i-1$ nodes (we consider the downstream queues at the PC for each station as a node) polled before node i, the packet has to wait for $B + iV + kL$ seconds before its service begins. Since k follows the Binomial distribution of Eqn. (3) with $M = 2i-1$, the pmf of X_{FR} is given by

$$P[X_{FR} = x] = \begin{cases} \binom{2i-1}{k} \rho^k (1-\rho)^{2i-k-1} & x = B + iV + kL \\ 0 & \text{otherwise} \end{cases} \tag{30}$$

with $0 \leq k \leq 2i-1$ and the expected value of X_{FR} is $E[X_{FR}] = B + iV + (2i-1)\rho L$. The total time before the packet receives service is then

$$\begin{aligned} E[X_{i,PS}] &= E[ST_S - B - t] + E\left[E[\kappa \mid t]\right] T_S + E[X_{FR}] \\ &= \frac{ST_S - B}{2} + \lambda T_S \frac{ST_S - B}{2} + B + iV + (2i-1)\rho L \end{aligned}$$

The expected delay at the i^{th} node is given by

$$D_{i,PS} = E[X_{i,PS}] + L. \tag{31}$$

B. Arrivals in the Active Mode

When arrivals occur in the active mode, the operation of the nodes is identical to the case where no power management is used. We can thus use the expressions from Section IV to evaluate the delay, $D_{i,AM}$, in this case. Thus we have

$$D_{i,AM} = \frac{1}{1-\lambda T_S}\left[\frac{T_S}{2}+\left(\frac{\rho L^2(2i-1)(1-\rho)}{T_S}+L\right)(1-\rho)\right] \tag{32}$$

Note that the expression for N_{NQ} used in the equation above now becomes only an approximation. The degree of approximation decreases as the load increases and as can be seen in the simulation results, the error introduced for low loads is quite small.

C. Overall Delay

The expressions for the delays of the previous two subsections can now be combined to obtain the expression for the delay experienced by an arbitrary arrival. The expected packet delay at node i is given by

$$D_i = D_{i,AM}P[\text{AM}] + D_{i,PS}P[\text{PS}] \tag{33}$$

where $D_{i,AM}$ and $D_{i,PS}$ and given in Eqns. (32) and (31), respectively and $P[\text{PS}]$ (and $P[\text{AM}] = 1 - P[\text{PS}]$) is given in Eqn. (28).

VII. Short CFP Durations

We now extend our model to the case where the CFP duration may not be large enough to serve all the nodes in the same superframe. We assume that at most m nodes may be served in a CFP, $m \leq M$. Note that the first m nodes in the polling list always get a chance to transmit in each superframe and thus there is no change in their delay model from the analysis in Section III (Eqn. (20)). However, nodes between $m+1$ to M in the polling list may have to wait for multiple CFPs (i.e. superframes) before they are served. Note that a strictly round robin polling scheme where the BS continues polling from where it left off in previous CFP would result in a "fair" and statistically same performance at each node. While our analysis can be easily extended to this case, in this paper we only concentrate on the strictly prioritized polling.

For nodes $m+1 \leq i \leq M$, we again break the analysis in two parts: arrivals in an empty and a non-empty queue. We first start with obtaining the probability that the i^{th} node receives service in an arbitrary CFP, denoted by $P[s]_i$. Note that $P[s]_i = 1$, $\mu_i = 1/T_S$ and $\rho_i = \rho = \lambda T_S$ for $1 \leq i \leq m$. For $m+i \leq i \leq M$, $\mu_i = P[s]_i/T_S$ and $\rho_i = \lambda T_S/P[s]_i$. Now, the probability of service $P[s]_i$ of node i, $m+1 \leq i \leq M$, depends on the probability that less than m of the $i-1$ nodes before it have packets to send in the superframe. Denote $i = m+k$. Now, the probability that j or less of the $i-1$ nodes and active is given by

$$Pr[j]_{i=m+k} = \sum_{l_0=0}^{j}\ \sum_{l_1=0}^{\min\{1,j-l_0\}} \cdots \sum_{l_{k-1}=0}^{\min\{1,j-l_0-\cdots-l_{k-2}\}} \binom{m}{l_0}\rho^{l_0} (1-\rho)^{m-l_0}\rho_{m+1}^{l_1}(1-\rho_{m+1})^{1-l_1}\cdots\rho_{m+k-1}^{l_{k-1}}(1-\rho_{m+k-1})^{1-l_{k-1}} \tag{34}$$

Note that the summations above evaluate the cases where the number of active nodes before the i^{th} node is less than or equal to j. Also, the first summation corresponds to the first m nodes while each subsequent summation corresponds to nodes $m+1$ to $i-1$ (i.e. node $m+k-1$). Now the probability of service, $P[s]_i$, is equal to the probability that less than $m-1$ of the preceding nodes have data to send in the superframe. Thus using Eqn. (34), $P[s]_i = Pr[m-1]_i$.

A. Arrivals at an Empty Queue

In the superframe the tagged packet arrives, the node does not receive any service with probability $1-P[s]_i$, and waits for an average of $T_S/2$ seconds before the superframe ends. Now in each subsequent superframe, the node is able to transmit a packet with probability $P[s]_i$ and the average number of superframes required to transmit is thus $1/P[s]_i$. Let X_{NR} be the time the packet spends in the final superframe before it is finally transmitted. Now, $m-1$ or fewer of the nodes before the i^{th} node in the polling list must have data to send in this superframe for node i to get a transmission attempt. From Eqn. (34), the probability that k or fewer preceding nodes are active given that $m-1$ or fewer nodes are active, $Pr[k|m-1]_i = \frac{Pr[k]_i}{Pr[m-1]_i}$, $0 \leq k \leq m-1$. Then

$$P[X_{NR}=x] = \begin{cases} \frac{Pr[k]_i - Pr[k-1]_i}{Pr[m-1]_i} & x = B+iV+kL \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

and $E[X_{NR}] = B + iV + L\sum_{j=0}^{m-1}\frac{Pr[k]_i-Pr[k-1]_i}{Pr[m-1]_i}$. The waiting time in this case of no service (NS) in the first CFP is then given by

$$E[X_i]_{NS} = \frac{T_S}{2} + \left(\frac{1}{P[s]_i} - 1\right)T_S + E[X_{NR}] \tag{36}$$

In case that the node receives service (S) in the first CFP, we again have the two cases C1 and C2 as in Section III-A corresponding to arrival before and after poll respectively, with probabilities given in Eqn. (13). Following the arguments of Section III-A and those for the derivation of $E[X_i]_{NS}$ above, the waiting times for cases C1 and C2 are given by

$$E[X_{i,j,C1}]_S = \frac{B+iV+jL}{2}$$
$$E[X_{i,j,C2}]_S = \frac{T_S-B-iV-jL}{2} + \left(\frac{1}{P[s]_i}-1\right)T_S + E[X_{NR}]$$

Combining the three cases above and adding the packet service time of L, the expected delay when the arrival occurs on an empty queue is given by

$$D_{i,EQ} = \sum_{j=0}^{m-1} E[X_{i,j}]\frac{Pr[j]_i - Pr[j-1]_i}{Pr[m-1]_i} \tag{37}$$

B. Arrivals at a Non-Empty Queue

We now consider the case when an arrival occurs at a non-empty queue and sees an average of $E[N_{NQ}]$ packets waiting which must be served before the tagged packet gets service. In

Parameter	Value
Transmission power	281.8 mW
Transmission range	250 meters
Slot time	20 μsec
SIFS	10 μsec
DIFS	50 μsec
PIFS	30 μsec
CFPriMax	30 msec
Channel bandwidth	2 Mbps
Beacon	209 μsec
CF-Poll	209 μsec
CF-End	209 μsec
CF-ACK	153 μsec
Packet size	520 B

TABLE I
SIMULATION SETTINGS

the case where the i^{th} node is not served in this superframe, following the derivation in the previous subsection

$$E[X_i]_{NS} = \frac{T_S}{2} + E[N_{NQ}]\frac{T_S}{P[s]_i} + \left(\frac{1}{P[s]_i} - 1\right)T_S + E[X_{NR}] \tag{38}$$

where X_{NR} is given in Eqn. (35). For the case when the i^{th} node gets served in this superframe, again we have the two cases C1 and C2 with probabilities given in Eqn. (16). Following the derivation of Section III-B and the subsection above, we have

$$E[X_{i,j,C1}]_S = \frac{2T_S - B - iV - (j+1)L}{2} + (E[N_{NQ}] - 1)\frac{T_S}{P[s]_i} + \left(\frac{1}{P[s]_i} - 1\right)T_S + E[X_{NR}] \tag{39}$$

$$E[X_{i,j,C2}]_S = \frac{T_S - B - iV - (j+1)L}{2} + E[N_{NQ}]\frac{T_S}{P[s]_i} + \left(\frac{1}{P[s]_i} - 1\right)T_S + E[X_{NR}] \tag{40}$$

Combining the three cases above and adding the packet service time of L, the expected delay when the arrival occurs on a non-empty queue, $D_{i,NEQ}$, is given by Eqn. (37) after substituting Eqns. (38), (39) and (40) for $E[X_i]_{NS}$, $E[X_{i,j,C1}]_S$ and $E[X_{i,j,C2}]_S$ respectively.

C. Overall Delay

Combining the expressions for the cases of arrivals at empty and non-empty queue, we have

$$D_i = D_{i,EQ}(1 - \rho_i) + D_{i,NEQ}\rho_i \tag{41}$$

and the expected delay for an arbitrary arrival, D_i, is obtained by substituting $E[N_{NQ}] = \lambda D_i/\rho_i$ in the equation above and solving for D_i, $m + 1 \leq i \leq M$.

VIII. SIMULATION RESULTS

In this section we validate the analytic models proposed in the previous sections by comparing them with simulation results. These simulations were carried out using our own simulation code. The simulations were carried out for different network sizes and parameter settings as indicated in in Table I. In the simulations, we considered a circular region of radius 240 meters with the base station at the center and all other nodes within its range.

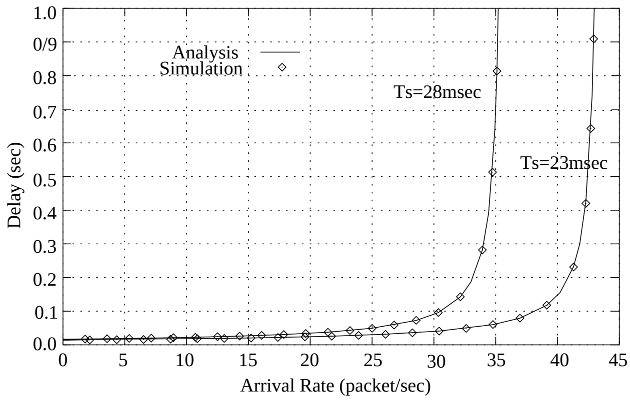

Fig. 4. Simulation and analytic results for unidirectional transfer with 10 nodes.

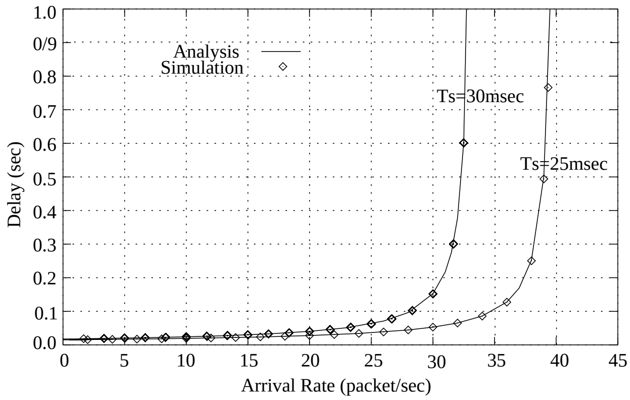

Fig. 5. Simulation and analytic results for bidirectional transfer with 5 nodes.

In Figure 4 we compare the simulation and analytic results (from Eqn. (20)) for the unidirectional traffic case when there are 8 nodes in the network. We show two cases corresponding to CFPri or T_S values of 23 msec and 28 msec and in both cases we note the close match between the simulation and analytic results. The delays are presented for the 5^{th} node in the polling list. We also note that having a shorter T_S supports higher arrival rates for a given delay requirement. In Figure 5 we compare the analytic model of Eqn. (23) for the bidirectional transfers with the simulations results for a network with 5 nodes and T_S values of 30 msec and 25 msec. The delays are presented for the 3^{rd} node in the polling list. We again note the close match which validates our model. Similar results were obtained for other network sizes, T_S lengths and packet sizes.

In Figure 6 we plot the number of nodes that can be supported by the unidirectional and bidirectional cases as obtained using simulations and the analytic results of Section V for different superframe sizes. The curves represent a scenario with a delay requirement δ of 150 msec which corresponds to the bound specified for excellent quality voice transmissions [6]. The packet arrival rate was 33 packets per second with

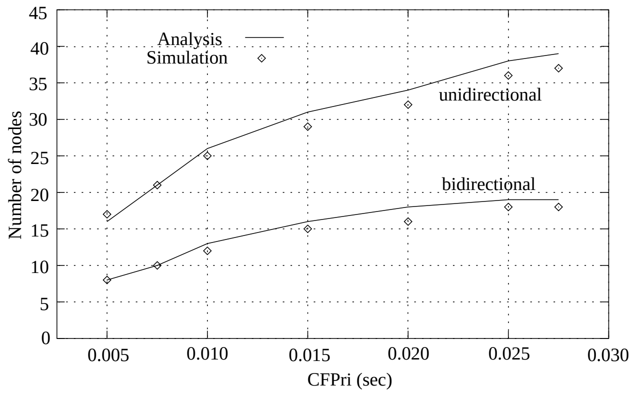

Fig. 6. Number of allowable nodes for voice traffic.

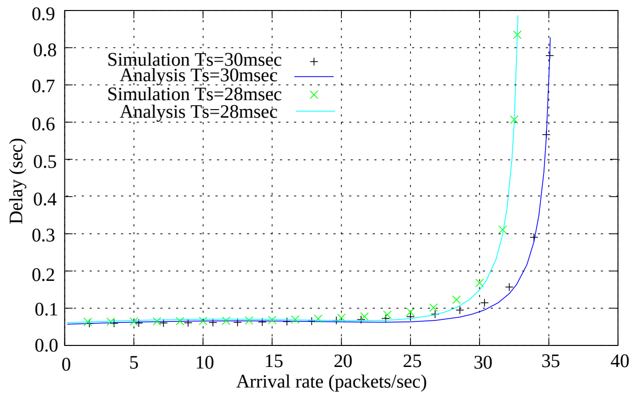

Fig. 7. Simulation and analytic results for nodes with power management for T_S values of 28 msec and 30 msec.

a payload of 24 bytes which is the data rate specified by the G.723.1 codec at 6.4Kbps. For both these cases we see the close match between our analysis and simulations.

In Figure 7 we compare the simulation and analytic results (from Eqn. (33)) for nodes with power management when there are 5 nodes in the network. We show two cases corresponding to T_S values of 28 msec and 30 msec and in both cases we note the close match between the simulation and analytic results. We also note that having a shorter T_S supports higher arrival rates for a given delay requirement. Finally in Figure 8 we compare the simulation and analysis results for the case considered in Section VII where not all nodes can be served in one CFP. We consider a scenario with 15 nodes and T_S of 25 msec which allows for atmost 8 nodes to be served in a CFP. The figure shows the delays at the 10th, 13th and 15th node in the polling list and we note that as expected, the delays are higher for nodes further down the polling order.

IX. Conclusions

In this paper, we presented an analytic model to evaluate the delays is wireless networks using the IEEE 802.11 PCF as the MAC layer protocol, with and without power management. Closed form expressions for the delays at each node were obtained using a queueing model for the cases with both unidirectional as well as bidirectional traffic. The model is able to account for arbitrary (but fixed) packet sizes, polling frequencies, channel rates and the order in which a node is polled. The accuracy of these expressions was verified using

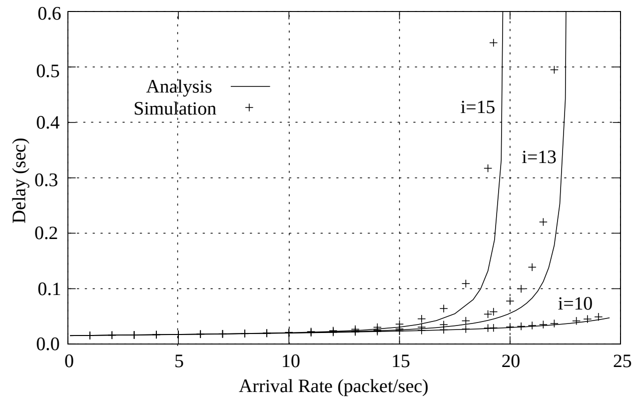

Fig. 8. Simulation and analytic results when not all nodes can be served in a superframe ($m = 8$ and $M = 15$).

simulations. In addition, a simple admission control strategy for the base stations was proposed and its results were verified using simulations.

References

[1] D. Bertsekas and R. Gallaher, *Data Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[2] B. Crow, I. Widjaja, J. Kim, and P. Sakai, "Investigation of the IEEE 802.11 Medium Access Control (MAC) sublayer functions," in *Proc. IEEE INFOCOM 1997*, pp. 126-133.

[3] V. DeLeon and R. Berry, "Bringing VR to the desktop: are you game?" *IEEE Multimedia*, vol. 7, no. 2, pp. 68-72, April 2000.

[4] A. Ganz, A. Phonphoem, N. Llopis, I. Kim, and K. Wongthavarawat, "Converged voice, video and data wired-wireless LANs testbed," in *Proc. IEEE MILCOM 1999*, pp. 1297-1301.

[5] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE standards 802.11, Jan. 1997.

[6] Series G: Transmission Systems and Media, Digital Systems and Networks: One-way Transmission Time, ITU-T Standards, May 2000.

[7] L. Kleinrock, *Queueing Systems, Vol. I*. John Wiley and Sons, 1975.

[8] S. Sharma, K. Kim, C. Oh, and A. Ahmad, "Performance evaluation of IEEE 802.11 MAC protocol for multimedia services," in *Proc. APCC/ICCS 1998* pp. 181-185.

[9] T. Suzuki and S. Tasaka, "Performance evaluation of integrated video and data transmission with the IEEE 802.11 standard MAC protocol," in *Proc. IEEE GLOBECOM 1999*, pp. 58-586.

[10] O. Tickoo and B. Sikdar, "Queueing analysis and delay mitigation in IEEE 802.11 random access MAC based wireless networks," in *Proc. IEEE INFOCOM 2004*, pp. 1404-1413.

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

[12] M. Visser and M. El Zarki, "Voice and data transmission over an 802.11 wireless network," in *Proc. IEEE PIMRC 1995*, pp. 648-652.

[13] Y. Xiao, H. Li, and S. Choi, "Protection and guarantee for voice and video traffic in IEEE 802.11e wireless LANs," in *Proc. IEEE INFOCOM 2004*, pp. 2153-2163.

[14] J. Yeh and C. Chen, "Support of multimedia services with the IEEE 802.11 MAC protocol," in *Proc. IEEE ICC 2002*, pp. 600-604.

Biplab Sikdar (S'98, M'02) received the B. Tech degree in electronics and communication engineering from North Eastern Hill University, Shillong, India, the M. Tech degree in electrical engineering from Indian Institute of Technology, Kanpur and Ph.D in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, USA in 1996, 1998 and 2001, respectively. He is currently an Assistant Professor in the Department of Electrical, Computer and Systems Engineering of Rensselaer Polytechnic Institute, Troy, NY, USA. His research interests include wireless MAC protocols, network routing and multicast protocols, network security and queueing theory. Dr. Sikdar is a member of IEEE, Eta Kappa Nu and Tau Beta Pi.