Queueing Analysis of Polled Service Classes in the IEEE 802.16 MAC Protocol

Biplab Sikdar

***Abstract*— This paper considers the performance of the polling based service classes of IEEE 802.16 based broadband wireless access networks and develops queueing models to evaluate their delay distributions and loss rates. Both single and multiple carrier OFDMA operations are considered and models are proposed for two polling strategies. The models can be used to provide probabilistic service guarantees and explore the impact of various system parameters on the performance, thereby aiding in system design. The models are verified using simulations.**

***Index Terms*— Wireless broadband access, MAC protocol, IEEE 802.16**

I. Introduction

The IEEE 802.16 standard for point to multipoint broadband wireless access is an emerging technology for ubiquitous broadband wireless access supporting fixed, nomadic, portable and fully mobile operations offering integrated voice, video and data services. The IEEE 802.16e standard supports five scheduling service classes for quality of service (Unsolicited Grant Service (UGS), real-time, non-real-time and extended-real-time Polling Service (rtPS, nrtPS and ertPS) and Best Effort (BE)) and includes a request-grant mechanism for uplink transmissions from a Subscriber Station (SS) to its Base Station (BS).

While existing literature has evaluated many aspects of IEEE 802.16, analytic models for polled services classes are largely absent. Simulation studies to evaluate the performance of various service classes are presented in [1], [2]. The binary exponential backoff and random access mechanism of IEEE 802.16 are modeled in [3], [4]. Delay bounds for orthogonal frequency division multiple access with time division multiple access (OFDMA-TDMA) and OFDMA systems for some specific burstiness control schemes are developed in [5]. Connection-level characteristics of IEEE 802.16 under call admission control and bandwidth allocation schemes proposed by the authors are presented in [6], [7].

Unlike existing literature, this paper focuses on developing queueing models specific to the case of polling based service classes in IEEE 802.16. This MAC layer delay is an important factor in the overall performance and capacity utilization of the system and accurate characterization of this delay is critical to meeting performance goals of delay-sensitive applications. This paper analyzes different polling schemes and presents comparative results based on both our analysis as well as simulations. The analytical models derive expressions for the packet delay distribution and packet blocking rates at each SS as a function of various systems parameters. Our models can be used for determining optimal frame lengths and other system settings, number of supportable connections for a given delay constraint, and admission control.

The rest of the paper is organized as follows. Section II presents the queueing model for the case where the SSs are polled at the end of the uplink subframe and Section III considers polling at the beginning of the uplink subframe. Section IV extends the analysis to the IEEE 802.16 OFDMA PHY. Finally, Section V presents the simulation results and Section VI concludes the paper.

Manuscript submitted on 08 January 2009, revised on 01 May 2009, and accepted 29 June 2009.

B. Sikdar is with the Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY USA.

II. Delay Analysis: Polling at end of Uplink Subframe

We consider a single BS serving n SSs through a TDMA/TDD, single carrier air-interface. Each frame is divided into uplink and downlink subframes, as per the IEEE 802.16 standards. The standard however does not specify any scheduling algorithm and leaves it to be vendor specific. We assume that a single packet is transmitted by a SS in a frame if it made a bandwidth request in the previous frame.

This section considers the following polling scheme: nodes are polled sequentially at the end of every uplink subframe. The packet interarrival times at a SS are assumed to distributed according to a Markov modulated Poisson process (MMPP) with an arbitrary number of states, r. An MMPP based arrival process is used in this paper because of their versatility in modeling traffic types such as voice, video as well as long range dependent traffic [8], [9]. The MMPP is characterized by the transition rate matrix $\mathbf{R}$ and the diagonal rate matrix $\mathbf{\Lambda}$ that contains the arrival rates at each state:

$$\mathbf{R} = \begin{bmatrix} -\sigma_1 & \sigma_{12} & \cdots & \sigma_{1r} \\ \sigma_{21} & -\sigma_2 & \cdots & \sigma_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{r1} & \sigma_{r2} & \cdots & -\sigma_r \end{bmatrix} \tag{1}$$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_r \end{bmatrix} \tag{2}$$

The steady state probability vector q of the Markov chain satisfies $q\mathbf{R} = 0$ and $qe = 1$ where e is a unit vector. The average arrival rate at the SS is then given by $\lambda = q\mathbf{\Lambda}$. The utilization factor of the queue at each SS is approximated in this paper as $\rho = \lambda T_s$. We denote the durations of the uplink and downlink subframes by T_{UL} and T_{DL}, respectively. The total frame duration is denoted by T_S. The time taken to poll a single station is denoted by T_P and time taken to transmit a packet by L. Each SS has a finite buffer that holds K packets.

Consider a tagged packet arriving at SS i, $1 \leq i \leq n$. At the instant of its arrival, the queue at the SS may be in one of two states: **1. S0:** The queue is empty (an arbitrary arrival sees an empty queue with probability $1-\rho$). **2. S1:** The queue is non-empty (with probability ρ).

A. Arrival at an Empty Queue: State S0

We consider two subcases: arrival before (case C1) and after (case C2) SS i has been polled in the current frame. Consider case C1. Since the SS has not been polled yet, a reservation can be made in this frame for transmitting the tagged packet in the next frame. For arbitrary arrivals independent of the departure process in a slotted departure system, an arrival is equally likely to occur anywhere in a slot [10]. In our case, *given* that an arrival occurs in a frame, the arrival instance, t, *relative* to the start of the frame is thus uniformly distributed in $[0, T_s]$. SS i is polled $(n-i+1)T_P$ seconds before the frame ends. The time the SS waits before it sends the bandwidth request is thus $T_s-(n-i+1)T_P-t$. The probability distribution function (PDF) of t given that the arrival occurred before SS i was polled in the frame is

$$\begin{aligned} P[t \leq \tau \mid t \leq T_s-(n-i+1)T_P] &= \frac{P[t \leq \tau, t \leq T_s-(n-i+1)T_P]}{P[t \leq T_s-(n-i+1)T_P]} \\ &= \frac{\tau}{T_s-(n-i+1)T_P} \end{aligned} \quad (3)$$

which is an uniform distribution: $U[0, T_s-(n-i+1)T_P]$. If a random variable Y is uniformly distributed in the range 0 to a, then $a-Y$ is also uniformly distributed in the range 0 to a. Thus the PDF of $T_s-(n-i+1)T_P-t$ is also $U[0, T_s-(n-i+1)T_P]$. Following the bandwidth request, $(n-i+1)T_P$ seconds pass before the current frame ends. If j of the $i-1$ SSs that were polled before SS i also transmit data in the next frame, SS i has to wait an additional $T_{DL}+jL$ seconds in the next frame before it is served. Since an SS has a non-empty queue with probability ρ, the probability that there are j SSs who send packets is binomially distributed with parameters $B[i-1, \rho]$. Here we have made the approximation that all SSs with non-empty queues had bandwidth reserved for them in the previous frame. This assumption is fairly accurate as the load increases and as our results show in Section V, has very little effect at low loads. The Laplace-Stieltjes Transform (LST) of the service time in this case, $X_{i,S0,C1}$, is

$$\begin{aligned} H_{X_{i,S0,C1}}(s) &= \text{LST}\left[U[0, T_s-(n-i+1)T_P]+(n-i+1)T_P \right. \\ &\quad \left. +T_{DL}+B[i-1,\rho]L+L\right] \quad (4) \\ &= \frac{1-e^{-s(T_s-(n-i+1)T_P)}}{s(T_s-(n-i+1)T_P)} \frac{(1-\rho+\rho e^{-sL})^{i-1}}{e^{s((n-i+1)T_P+T_{DL}+L)}} \quad (5) \end{aligned}$$

where the first term in the equation above is the LST of $U[0, T_s-(n-i+1)T_P]$, the second term is the LST of the constants $(n-i+1)T_P+T_{DL}+L$ and the third term is the LST of $B[i-1, \rho]L$.

For case C2, the packet first has to wait till the current frame is over (T_S-t). The SS makes a bandwidth request in the next frame and is allocated a transmission slot in the subsequent frame. The PDF of t given that it arrived after SS i was polled, i.e., it arrived after the first $T_s-(n-i+1)T_P$ seconds of the frame is

$$\begin{aligned} P[t \leq \tau \mid t > T_s-(n-i+1)T_P] &= \frac{P[t \leq \tau, t > T_s-(n-i+1)T_P]}{P[t > T_s-(n-i+1)T_P} \\ &= \frac{\tau-T_s+(n-i+1)T_P}{(n-i+1)T_P} \end{aligned} \quad (6)$$

which is distributed as $U[T_S-(n-i+1)T_P, T_S]$. If a random variable Y has the distribution $U[a,b]$, then $b-Y$ has the distribution $U[0, b-a]$. Thus T_S-t is distributed as $U[0, (n-i+1)T_P]$. Again, the number of SSs before SS i that send data in the frame in which the tagged packet is transmitted is binomially distributed with parameters $B[i-1, \rho]$. The LST of the service time in this case, $X_{i,S0,C2}$, is then

$$\begin{aligned} H_{X_{i,S0,C2}}(s) &= \text{LST}\left[U[0,(n-i+1)T_P]+T_S+T_{DL}+ \right. \\ &\quad \left. B[i-1,\rho]L+L\right] \quad (7) \\ &= \frac{1-e^{-s(n-i+1)T_P}}{s(n-i+1)T_P} \frac{(1-\rho+\rho e^{-sL})^{i-1}}{e^{s(T_S+T_{DL}+L)}} \quad (8) \end{aligned}$$

Now, the probabilities of cases C1 and C2 are given by $P[C1] = \frac{T_S-(n-i+1)T_P}{T_S}$ and $P[C2] = \frac{(n-i+1)T_P}{T_S}$, respectively. The LST of the service time in state S0, $X_{i,S0}$, is then given by

$$\begin{aligned} H_{X_{i,S0}}(s) &= P[C1]H_{X_{i,S0,C1}}(s)+P[C2]H_{X_{i,S0,C2}}(s) \\ &= \frac{(1-\rho+\rho e^{-sL})^{i-1}}{sT_S e^{s(T_{DL}+L)}} \left[\frac{1-e^{-s(T_s-(n-i+1)T_P)}}{e^{s(n-i+1)T_p}}+ \right. \\ &\quad \left. \frac{1-e^{-s(n-i+1)T_P}}{e^{sT_S}}\right] \quad (9) \end{aligned}$$

B. Arrival at a Non-Empty Queue: State S1

Let the number of packet seen by a tagged arrival at a non-empty queue be N_{NQ}. The service time of the tagged packet begins when the last of the N_{NQ} enqueued packets departs the queue. A bandwidth request is sent for the tagged packet in the frame in which it comes to the head of the line (HOL) and the tagged packet is transmitted in the next frame. Let j of the $i-1$ SSs before SS i also transmit in the frame where the tagged packet comes to the HOL and starts its service (j is binomially distributed with parameters $B[i-1, \rho]$). Then the time remaining in this frame when the tagged packet starts its service is $T_S-T_{DL}-jL-L$. In the next frame, if j' of the $i-1$ SSs also transmit a packet, SS i has to wait for $T_{DL}+j'L$ seconds before it begins its service. The total service time in this case is $X_{i,S1} = T_S-T_{DL}-jL-L+T_{DL}+j'L+L = T_S-jL+j'L$. Since both j and j' are binomially distributed with parameters $B[i-1, \rho]$, the LST of the service time for this case is

$$\begin{aligned} H_{X_{i,S1}}(s) &= \text{LST}\left[T_S-B[i-1,\rho]L+B[i-1,\rho]L\right] \quad (10) \\ &= e^{-sT_S}(1-\rho+\rho e^{sL})^{i-1}(1-\rho+\rho e^{-sL})^{i-1} \quad (11) \end{aligned}$$

C. Overall Service Time, Delay Distribution and Loss Rates

Combining the service times for cases S0 and S1, the LST of the service time of an arbitrary arrival at SS i, X_i, is given by

$$H_{X_i}(s) = (1-\rho)H_{X_{i,S0}}(s) + \rho H_{X_{i,S1}}(s) \tag{12}$$

where $H_{X_{i,S0}}(s)$ and $H_{X_{i,S1}}(s)$ are given in Eqn. (9) and Eqn. (11) respectively. The average service time is denoted by Θ and given by

$$\begin{aligned}\Theta &= -\frac{d}{ds}H_{X_{i,S1}}(s)\Big|_{s=0} \\ &= (1+\rho)\frac{T_S}{2} + (1-\rho)\left[(n-i+1)T_p + T_{DL} + (i-1)\rho L + L\right]\end{aligned}$$

To obtain the distribution of the packet delays and loss rates, the queue at each SS is modeled as a MMPP/G/1/K queue whose service time distribution is given by Eqn. (12). We use the analysis for the MMPP/G/1/K queue from [11] and list the equations below for completeness.

Consider the imbedded Markov chain consisting of the service completion instants at the queue. Let $\pi(k)$ (respectively, $p(k)$) be the $r-$dimensional vector whose $j-$th element is the limiting probability at the imbedded epochs (at an arbitrary time instant) of having k packets in the queue and being in the phase j of the MMPP, $k = 0, 1, \cdots, K-1$ ($k = 0, 1, \cdots, K$). Consider the matrix sequence $\{\mathbf{C}_k\}$ defined as

$$\mathbf{C}_{k+1} = \left[\mathbf{C}_k - \mathbf{U}\mathbf{A}_k - \sum_{\nu=1}^{k}\mathbf{C}_\nu \mathbf{A}_{k-\nu+1}\right]\mathbf{A}_0^{-1} \tag{13}$$

for $k = 1, 2, \cdots, K-2$ with $\mathbf{C}_0 = \mathbf{I}$, $\mathbf{C}_1 = (\mathbf{I} - \mathbf{U}\mathbf{A}_0)\mathbf{A}_0^{-1}$ and $\mathbf{I}$ being a $r \times r$ identity matrix. The $(k,l)-$th element of the matrix $\mathbf{A}_\nu$ denotes the conditional probability of reaching phase l and having ν arrivals at the end of a service time, starting from phase k. The matrices $\mathbf{A}_\nu$ can be easily calculated using an iterative procedure [12]. The probability vectors $\pi(k)$ can then be calculated using

$$\pi(0)\left[\sum_{\nu=0}^{K-1}\mathbf{C}_\nu + (\mathbf{I}-\mathbf{U})\mathbf{A}(\mathbf{I}-\mathbf{A}+eq)^{-1}\right] = q \tag{14}$$

and $\pi(k) = \pi(0)\mathbf{C}_k$, $k = 1, 2, \cdots, K-1$. The vectors $p(k)$ are then obtained using $p(0) = \xi\pi(0)(\mathbf{\Lambda}-\mathbf{R})^{-1}\Theta^{-1}$ and

$$p(k) = \xi\left[\pi(k) + \sum_{\nu=0}^{k-1}\pi(\nu)\mathbf{U}^{k-1-\nu}(\mathbf{U}-\mathbf{I})\right](\mathbf{\Lambda}-\mathbf{R})^{-1}\Theta^{-1} \tag{15}$$

for $k = 1, 2, \cdots, K-1$ and $p(K) = q - \sum_{\nu=1}^{K-1} p(\nu)$ where $\xi = [1 + \pi(0)(\mathbf{\Lambda}-\mathbf{R})^{-1}\Theta^{-1}e]^{-1}$. The packet blocking probability is given by

$$P_b = 1 - \sum_{\nu=0}^{K-1} p(\nu) \tag{16}$$

Finally, the LST of the cumulative distribution function of the packet waiting time, $W(s)$ is given by

$$W(s) = \frac{1}{1-P_b}\left[p(0) + \xi\Theta^{-1}\sum_{\nu=1}^{K-1}G_\nu(s)H_{X_i}^{K-1-\nu}(s)\mathbf{T}_{K-1-\nu}(s)\right] \tag{17}$$

where $G_j(s) = \pi(0)[\mathbf{I} - \mathbf{U}H_{X_i}(s)] - H_{X_i}^j(s)\pi(j)$, $\mathbf{T}_j(s) = \mathbf{F}(s)[-\mathbf{\Lambda}\mathbf{F}(s)]^j$ and $\mathbf{F}(s) = [s\mathbf{I} + \mathbf{R} - \mathbf{\Lambda}]^{-1}$. Moments of the packet waiting time can be easily obtained from Eqn. (17).

III. Delay Analysis: Polling at the Start of the Uplink Subframe

In this section we analyze the case where stations are polled at the start of the uplink subframe. The analysis follows along the same lines as in Section II and the details have been omitted. The same definitions as in Section II are used for the two states S0 and S1 and their subcases C1 and C2.

A. Arrival at an Empty Queue: State S0: With SSs polled at the beginning of the uplink subframe, the time from the start of a frame till SS i is polled is $T_{DL}+(i-1)T_P$ and the arrival instant t thus has the uniform distribution $U[0, T_{DL}+(i-1)T_P]$. The time from the arrival till the poll, $T_{DL}+(i-1)T_P-t$, is thus also distributed as $U[0, T_{DL}+(i-1)T_P]$ and the time remaining in the frame after SS i is polled is $T_S-T_{DL}-(i-1)T_P$. In case C1, the bandwidth request is sent in this frame itself and the packet is transmitted in the next frame. The number j of the $i-1$ SSs that also transmit a packet before SS i in the next frame is binomially distributed with parameters $B[i-1, \rho]$, resulting in a delay of $T_{DL} + nT_p + jL$ seconds in the frame before SS i is served. The LST of the service time in this case, $X_{i,S0,C1}$, is given by

$$\begin{aligned}H_{X_{i,S0,C1}}(s) &= \text{LST}\,[U[0, T_{DL} + (i-1)T_P] + T_S + \\ &\qquad (n-i+1)T_P + B[i-1,\rho]L + L] \qquad (18)\\ &= \frac{1-e^{-s(T_{DL}+(i-1)T_P)}}{s(T_{DL}+(i-1)T_P)}\frac{(1-\rho+\rho e^{-sL})^{i-1}}{e^{s(T_S+(n-i+1)T_P+L)}} \qquad (19)\end{aligned}$$

Following along the same lines, the LST of the service time for case C2, $X_{i,S0,C2}$, is given by

$$\begin{aligned}H_{X_{i,S0,C2}}(s) &= \text{LST}\,[U[0, T_S - T_{DL} - (i-1)T_P] + \\ &\qquad T_S + T_{DL} + nT_P + B[i-1,\rho]L + L] \qquad (20)\\ &= \frac{1-e^{-s(T_S-T_{DL}-(i-1)T_P)}}{s(T_S-T_{DL}-(i-1)T_P)}\frac{(1-\rho+\rho e^{-sL})^{i-1}}{e^{s(T_S+T_{DL}+nT_P+L)}} \qquad (21)\end{aligned}$$

The probabilities of the cases C1 and C2 are given by $P[C1] = \frac{T_{DL}+(i-1)T_P}{T_S}$ and $P[C2] = \frac{T_S-T_{DL}-(i-1)T_P}{T_S}$, respectively. Combining cases C1 and C2, the LST of the service time, $X_{i,S0}$, is given by

$$H_{X_{i,S0}}(s) = \frac{(1-\rho+\rho e^{-sL})^{i-1}}{sT_S e^{s(T_S+nT_p+L)}}\left[\frac{1-e^{-s(T_{DL}+(i-1)T_P)}}{e^{-s(i-1)T_p}} + \frac{1-e^{-s(T_S-T_{DL}-(i-1)T_P)}}{e^{sT_{DL}}}\right] \tag{22}$$

B. Arrival at a Non-Empty Queue: State S1: In this case, the service time of the tagged packet begins when the last of the enqueued packets seen by the tagged packet on arrival, departs the queue. Let j and j' of the $i-1$ SSs before SS i also transmit in the frame where the tagged packet starts its service and the next frame, respectively. The total service time in this case is $X_{i,S1} = T_S - T_{DL} - nT_P - jL - L + T_{DL} + nT_P + j'L + L = T_S - jL + j'L$. Since both j and j' are binomially distributed with parameters $B[i-1,\rho]$, the LST of the service time for this case is

$$H_{X_{i,S1}}(s) = e^{-sT_S}(1-\rho+\rho e^{sL})^{i-1}(1-\rho+\rho e^{-sL})^{i-1} \quad (23)$$

C. Overall Service Time, Delay Distribution and Loss Rates: Combining cases S0 and S1, the LST of the service time of an arbitrary arrival at SS i, X_i, is given by $H_{X_i}(s) = (1-\rho)H_{X_{i,S0}}(s)+\rho H_{X_{i,S1}}(s)$ where $H_{X_{i,S0}}(s)$ and $H_{X_{i,S1}}(s)$ are given in Eqn. (22) and Eqn. (23) respectively. The distribution for the waiting time and the expected blocking rates can then be evaluated using Eqns. (17) and (16) and the methodology of Section II-C.

IV. Multichannel Scenario

This section extends the analysis to IEEE 802.16 operation over an OFDMA PHY. The OFDMA PHY is modeled as a set of m orthogonal groups of subchannels (each consisting of multiple subcarriers) in the frequency domain. A SS is assigned one such group when it wants to transmit data and at most one packet is served from a SS in one frame. For both polling scenarios, the analysis closely follows the structure developed in previous sections and we only consider polling at the start of the uplink subframe for illustrative purposes. The main difference is that m SSs may transmit at the same time in the multichannel scenario. Thus the time before SS i is polled relative to the start of polling is $\lfloor \frac{i-1}{m} \rfloor T_P$ and if j SSs transmit their data before SS i, SS i has to wait for $\lfloor \frac{j}{m} \rfloor L$ seconds before it transmits its own packet. The rest of the analysis stays the same and the details are thus omitted to avoid repetition.

A. Arrival at an Empty Queue: State S0: The time till SS i is polled in the frame is $T_{DL}+\lfloor \frac{i-1}{m} \rfloor T_P$. In case C1, the arrival time t is then distributed as $U[0, T_{DL}+\lfloor \frac{i-1}{m} \rfloor T_P]$. The remaining time in the frame after SS i is polled is $T_S - T_{DL} - \lfloor \frac{i-1}{m} \rfloor T_P$. In the next frame, SS i has to wait $T_{DL}+\lceil \frac{n}{m} \rceil T_P + \lfloor \frac{j}{m} \rfloor L$ seconds where j is binomially distributed with parameters $B[i-1,\rho]$. The LST of the distribution of the random variable $\lfloor \frac{j}{m} \rfloor L$ is given by

$$H_{BF}(s) = I_{1-\rho}(n-m+1,m) + \sum_{j=1}^{\lfloor \frac{i-1}{m} \rfloor} [-I_{1-\rho}(n-jm+1,jm) + I_{1-\rho}(n-(j+1)m+1,(j+1)m)]^{-jsL} \quad (24)$$

where $I_{1-\rho}(a,b)$ is the incomplete regularized beta function defined as $I_x(a,b) = \frac{\int_0^x y^{a-1}(1-y)^{b-1}dy}{\int_0^1 y^{a-1}(1-y)^{b-1}dy}$. The LST of the service time, $X_{i,S0,C1}$, is then given by

$$\begin{aligned} H_{X_{i,S0,C1}}(s) &= \text{LST}\left[U\left[0, T_{DL}+\left\lfloor \frac{i-1}{m} \right\rfloor T_P\right] + T_S - \left\lfloor \frac{i-1}{m} \right\rfloor T_P + \left\lceil \frac{n}{m} \right\rceil T_P + \left\lfloor \frac{B[i-1,\rho]}{m} \right\rfloor L + L\right] && (25) \\ &= \frac{1-e^{-s(T_{DL}+\lfloor \frac{i-1}{m} \rfloor T_P)} H_{BF}(s)}{s(T_{DL}+\lfloor \frac{i-1}{m} \rfloor T_P) e^{-s(T_S - \lfloor \frac{i-1}{m} \rfloor T_P + \lceil \frac{n}{m} \rceil T_P + L)}} && (26) \end{aligned}$$

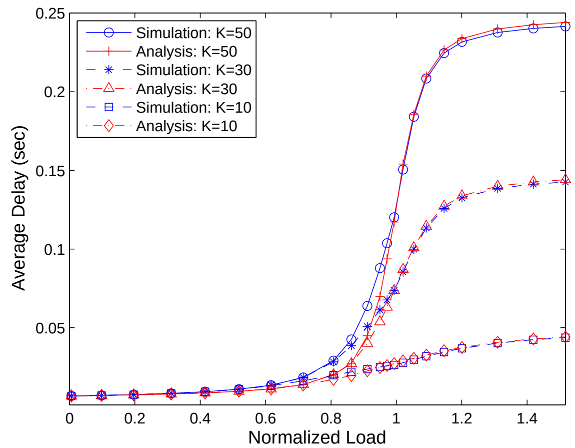

Fig. 1. Polling at end of uplink subframe: Average delay, $n = 5$, $i = 3$.

Similarly, the LST of the service time for case C2, $X_{i,S0,C2}$, is given by

$$\begin{aligned} H_{X_{i,S0,C2}}(s) &= \text{LST}\left[U\left[0, T_S - T_{DL} - \left\lfloor \frac{i-1}{m} \right\rfloor T_P\right] + T_S + T_{DL} + \left\lceil \frac{n}{m} \right\rceil T_P + \left\lfloor \frac{B[i-1,\rho]}{m} \right\rfloor L + L\right] && (27) \\ &= \frac{1-e^{-s(T_S - T_{DL} - \lfloor \frac{i-1}{m} \rfloor T_P)} H_{BF}(s)}{s(T_S - T_{DL} - \lfloor \frac{i-1}{m} \rfloor T_P) e^{s(T_S + T_{DL} + \lceil \frac{n}{m} \rceil T_P + L)}} && (28) \end{aligned}$$

The probabilities of cases C1 and C2 are given by $P[C1] = \frac{T_{DL}+\lfloor \frac{i-1}{m} \rfloor T_P}{T_S}$ and $P[C2] = \frac{T_S - T_{DL} - \lfloor \frac{i-1}{m} \rfloor T_P}{T_S}$, respectively. Combining cases C1 and C2, the LST of the service time, $X_{i,S0}$, is given by

$$H_{X_{i,S0}}(s) = \frac{H_{BF}(s)}{sT_S e^{s(T_S + \lceil \frac{n}{m} \rceil T_p + L)}} \left[\frac{1-e^{-s(T_{DL}+\lfloor \frac{i-1}{m} \rfloor T_P)}}{e^{-s\lfloor \frac{i-1}{m} \rfloor T_p}} + \frac{1-e^{-s(T_S - T_{DL} - \lfloor \frac{i-1}{m} \rfloor T_P)}}{e^{sT_{DL}}} \right] \quad (29)$$

B. Arrival at a Non-Empty Queue: State S1: Let j and j' of the i–1 SSs before SS i also transmit in the frame where the tagged packet starts its service and the next frame, respectively. The total service time in this case is $X_{i,S1} = T_S - T_{DL} - \lceil \frac{n}{m} \rceil T_P - \lfloor \frac{j}{m} \rfloor L - L + T_{DL} + \lceil \frac{n}{m} \rceil T_P + \lfloor \frac{j'}{m} \rfloor L + L = T_S - \lfloor \frac{j}{m} \rfloor L + \lfloor \frac{j'}{m} \rfloor L$. The LST of the service time for this case is

$$H_{X_{i,S1}}(s) = e^{-sT_S} H_{BF}(s) H_{BF}(-s) \quad (30)$$

C. Overall Service Time, Delay Distribution and Loss Rates: Combining cases S0 and S1, the LST of the service time of an arbitrary arrival at SS i, X_i, is given by $H_{X_i}(s) = (1-\rho)H_{X_{i,S0}}(s)+\rho H_{X_{i,S1}}(s)$ where $H_{X_{i,S0}}(s)$ and $H_{X_{i,S1}}(s)$ are given in Eqn. (22) and Eqn. (23) respectively. The distribution for the waiting time and the expected blocking rates can then be evaluated using Eqns. (17) and (16) and the methodology of Section II-C.

Normalized Load	Packet Drop Rate					
	$K = 10$		$K = 30$		$K = 50$	
	Simulation	Analysis	Simulation	Analysis	Simulation	Analysis
0.911	0.031	0.037	0.003	0.007	0.001	0.002
0.993	0.061	0.057	0.024	0.023	0.014	0.014
1.092	0.111	0.104	0.086	0.081	0.084	0.079

TABLE I

PACKET BLOCKING RATES FOR BUFFER SIZES OF 10, 30 AND 50 FOR POLLING AT THE END OF THE UPLINK SUBFRAME.

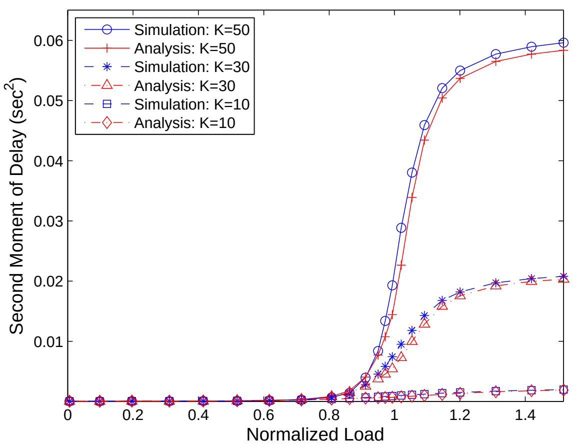

Fig. 2. Polling at end of uplink subframe: Second moment of delay, $n = 5$, $i = 3$.

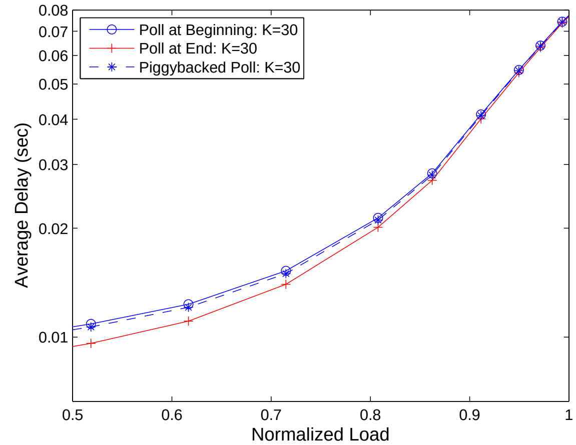

Fig. 4. Comparison with piggybacked operation: Average delay, $n = 5$, $i = 3$.

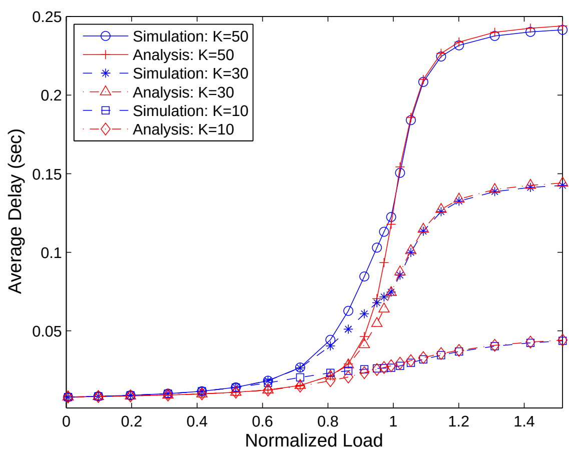

Fig. 3. Polling at beginning of uplink subframe: Average delay, $n = 5$, $i = 3$.

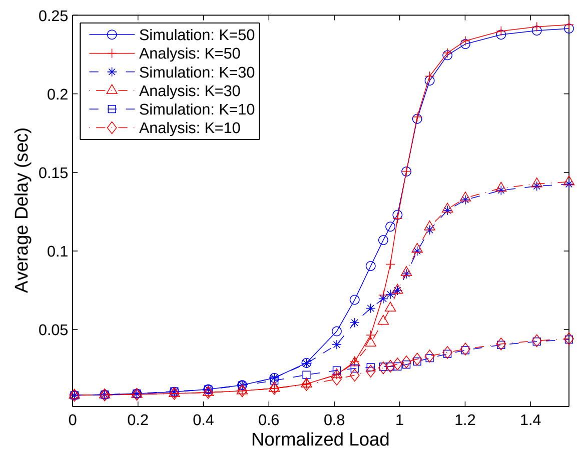

Fig. 5. Multichannel operation with polling at beginning of uplink subframe: Average delay, $n = 10$, $i = 7$.

V. SIMULATION RESULTS

This section verifies the accuracy of our models by comparing them against simulations. The simulations were carried out using a *NS-2* based IEEE 802.16 module developed by the WiMAX Forum. All results used the parameters: $T_S =$ 5ms, $T_{DL} = 3.75$ms, $T_{UL} = 1.25$ms and $T_P = 72\mu$s. The single (respectively, multi) channel operation had the following parameters: channel bandwidth: 3.5MHz (10MHz), Fast Fourier Transform (FFT) size: 256 (1024), oversampling: 8/7 (28/25), uplink data rate: 1.958Mbps (4.032Mbps), symbol time: 72μs (102.9μs), useful symbol time: 64μs (91.4μs) and 69 (48) symbols in the frame. Both scenarios used 16QAM 3/4 (quadrature amplitude) modulation. The simulations use a 2-state MMPP with transition rates of $\sigma_{12} = 3.15$ and $\sigma_{21} = 1.94$ and the ratio $\lambda_1 = 1.6\lambda_2$ for the arrival process [8].

Polling at End of Uplink Subframe: Figs. 1 and 2 demonstrate the closeness in the simulation and analytic results for the first and second moment of the packet delay when SSs are polled at the end of the uplink subframe, for different buffer sizes. The corresponding packet blocking rates are shown in Table I. The slight difference in the analytic and simulation results for the delay for moderate loads is because our model approximates the probability that a SS has a non-empty queue and bandwidth was reserved in the previous frame by ρ, the probability that the queue is empty.

Polling at Beginning of Uplink Subframe: Fig. 3 shows the close match between the average packet delay for various buffer sizes when SSs are polled at the beginning of the uplink subframe. Results for the delay's second moment and the blocking rates show similar trends and accuracy as those for polling at the end of the uplink subframe, and have been omitted due to constraints on the number of figures and tables.

Piggybacked Bandwidth Requests: With piggybacked operation, a SS may send bandwidth requests embedded in any data packet they transmit. An arrival is more likely to miss the poll in the frame of its arrival and thus have a longer wait if SSs are polled at the start of the uplink subframe. Polling at the end maximizes the likelihood of an arrival reserving bandwidth in the frame in which it arrives. Our models thus form upper and lower bounds on the delay for piggybacked operation, as verified in Fig. 4.

Multichannel Scenario: The comparisons between simulation and analysis for the multichannel OFDMA operation are presented in Fig. 5. A 10MHz channel was used. It was assumed that $m=5$ groups of subchannels were available for polled SSs and again the simulation and analytic results match closely.

VI. CONCLUSIONS

This paper presented queueing models to evaluate the performance of polling based operation of IEEE 802.16 networks in terms of the packet delays. We considered both single carrier as well as OFDMA based PHY layers under different polling strategies. Expressions for the delay distribution and packet blocking rates were obtained and the models were verified using simulations.

REFERENCES

[1] C. Cicconetti, A. Erta, L. Lenzini and E. Mingozzi, "Performance Evaluation of the IEEE 802.16 MAC for QoS Support," *IEEE Transactions on Mobile Computing,* vol. 6, no. 1, pp. 26-38, Jan. 2007.

[2] H. Lee, T. Kwon, D.-H. Cho, G. Lim and Y. Chang, "Performance Analysis of Scheduling Algorithms for VoIP Services in IEEE 802.16e Systems," *Proceedings of IEEE VTC,* pp. 1231-1235, Melbourne, Australia, May 2006.

[3] A. Vinel, Y. Zhang, M. Lott and A. Tiurlikov, "Performance Analysis of the Random Access in IEEE 802.16," *Proceedings of IEEE PIMRC,* pp. 1596-1600, Berlin, Germany, September 2005.

[4] H. Lee and J. Seo, "Queueing performance of IEEE 802.16 random access protocol with bulk transmissions," *Proceeding of IEEE ICC,* pp. 5963-5968, Beijing, China, May 2007.

[5] Y.-J. Chang, F.-T. Chien and C. Kuo, "Delay Analysis and Comparison of OFDM-TDMA and OFDMA under IEEE 802.16 QoS Framework," *Proceedings of IEEE GLOBECOM,* pp. 1-6, San Francisco, CA, November 2006.

[6] D. Niyato and E. Hossain, "Queue-aware uplink bandwidth allocation and rate control for polling service in IEEE 802.16 broadband wireless networks," *IEEE Trans. Mobile Computing,* vol. 5, no. 6, pp. 668-679, June 2006.

[7] D. Niyato and E. Hossain, "A queueing theoretic and optimization-based model for radio resource management in IEEE 802.16 broadband wireless networks," *IEEE Transactions on Computers,* vol. 55, no. 11, pp. 1473-1488, November 2006.

[8] S. Shah-Heydari and T. Le-Ngoc, "MMPP models for multimedia traffic," *Telecommunication Systems,* vol. 15, no. 3-4, pp. 273-293, December 2000.

[9] A. Andersen and B. Nielsen, "A Markovian approach for modeling packet traffic with long range dependence," *IEEE Journal on Selected Areas in Communications,* vol. 16, no. 5, pp. 719-732, June 1998.

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

[11] A. Baiocchi and N. Blefari-Melazzi, "Steady-state analysis of the MMPP/G/1/K queue," *IEEE Transactions on Communications,* vol. 41, no. 4, pp. 531-534, April 1993.

[12] W. Fischer and K. Meier-Hellstern, "The Markov-modulated Poisson-process (MMPP) cookbook," *Performance Evaluation,* vol. 18, no. 2, pp. 149-171, September 1993.