Admission Control and Bandwidth Allocation above Packet Level for IEEE 802.16 Wireless MAN

Haitang Wang, Bing He and Dharma P. Agrawal
ORB Center for Distributed and Mobile Computing
Department of ECECS, University of Cincinnati
Email: {wangh7, heb, dpa}@ececs.uc.edu

Abstract

IEEE 802.16 wireless MAN is expected to be the revolution of broadband wireless access technology. The grant/request mechanism in IEEE 802.16 MAC enables different QoS for different class of multimedia. Efficient resource management is essential in providing scalability in such large wireless network. The mixture of conventional voice traffic and fractal traffic make the resource management more complex in 802.16 wireless network than that in telephone networks. In this paper, by using Gaussian model for aggregated traffic in large network and Chernoff bound method, we analyze upper bound blocking probability above the packet level for all types of traffic in 802.16 wireless MAN. Based on the analysis result, we propose an admission control and bandwidth allocation mechanism above the packet level for 802.16 wireless MAN. We also compare our upper bound blocking probability methodology with Erlang B formula which is only useful for voice traffic and find out that result from our methodology can be used as approximate blocking probability when network is large and is more efficient than Erlang B formula in numerical computing.

1. Introduction

IEEE 802.16 standard and WiMAX (Worldwide Interoperability of Microwave Access) are standards-based wireless technologies that provide broadband wireless access over a long distance. Wireless broadband access setup is like cellular system using base station (BS) for serving multiple subscriber stations (SSs) up to a radius of several miles. Four classes of scheduling services are defined in IEEE 802.16: Unsolicited Grant Service (UGS), Real-time Polling Service (rtPS), Non-Real-Time Polling Service (nrtPS) and Best Effort (BE) [10]. Supporting Quality of Service (QoS) guarantees for these services transporting over 802.16 wireless MAN (Metropolitan Area Networks) is crucial. IEEE 802.16 MAC, defined as connection-oriented, is designed to support different QoS for these services. All services, including inherently connectionless service (packet service like IPv4, IPv6) are mapped to a connection. Once a new connection arrives at the SS, SS requests bandwidth per connection so that the BS uplink-scheduling algorithm could properly consider QoS while allocating bandwidth. The bandwidth is aggregated into a single grant to SS, not directly to the request connection. Typically, SS uses the bandwidth for the requested connection, but also can send the higher QoS data when QoS situation at the SS has changed since the last request [5]. The fair resource management is essential in providing scalability in 802.16 MAN MAC Layer but particularly difficult and complicated because of three characteristics of 802.16 MAN 1) The traffic of each service has its own spectrum of traffic statistics, like minimum guaranteed rate, maximum guaranteed rate, delay jitter, blocking probability etc. 2) The traffic belongs to some services such as rtPS, nrtPS, BE can exhibit significant rate variability and could demonstrate self-similarity or fractal [8]. 3) Bursty channel errors are present in the wireless links.

Traffic states in 802.16 wireless MAN can be characterized by packet, burst and call levels, which are different from original circuit switching network where traffic status are characterized only by call level. To facilitate efficient packet scheduling for transporting traffic in 802.16 wireless MAN, we need a good understanding of resource allocation above the packet level for enhancing network efficiency. The objective of this paper is to study the resource allocation above the packet level and induct an admission control and bandwidth allocation mechanism based on blocking probability for call level and burst level. Since the analytic traffic models for four classes of traffic in 802.16 wireless MAN are not exactly same, the resource allocation we are concerned about is different. For low bit rate connections such as the low bit rate UGS (voice) and BE (WWW) connections, we are concerned only with call level. For long

Proceedings of the 12th International Conference on Parallel and Distributed Systems (ICPADS'06)
0-7695-2612-8/06 $20.00 © 2006 IEEE

bursts connections which are caused by downloading large files, or long periods of high levels of VBR video activities such as nrtPS and rtPS connections, we are concerned burst level.

The rest of this paper is organized as follows. In section II, we describe the empirically derived analytic traffic models for UGS, rtPS, nrtPS and BE in 802.16 wireless MAN. Section III is the analysis part that focuses on evaluating the upper bound blocking probability at call level for UGS and BE and burst level for high bit rate fractional traffic such as rtPS and nrtPS. Based on the theoretical result from section III, we propose an admission control and bandwidth allocation mechanism above the packet level for 802.16 wireless MAN in section IV, the numerical results are also showed in this section. Section V concludes this paper.

2 Traffic Models

In 802.16 wireless MAN, each connection in the uplink direction is mapped to a scheduling service which belongs to UGS, rtPS, nrtPS and BE. ATM constant bit rate (CBR) and E1/T1 over ATM are the typical connections for UGS service. The rtPS is well suited for connections that carry services such as variable bit rate video and audio. The nrtPS is suitable for Internet access with minimum guaranteed rate such as FTP. The BE is suited for World Wide Web (WWW) traffic. The accurate traffic models for these four classes of traffic are important for 802.16 network design. In the conventional telecommunication network, the connection interarrival time and duration for voice traffic is modeled as exponential distribution and it had been proven a perfect model for performance evaluation both in theory and simulation. While for the other broadband traffic, the connection duration is not simple exponential distribution any more. More and more work in the recent years shows that the wide-area network traffic is much better modeled using statistically called self-similar processes [8]. In this part we study these accurate models used for traffic in 802.16 wireless MAN.

2.1 UGS

UGS has been characterized by a constant bit rate connection like voice traffic. For voice traffic, a Poisson process can model the connection arrival process with fix hourly rates within one-hour periods. That means the interarrival time between two connections is exponentially distributed. The connection duration (holding time) is also exponential distribution. In the packet level, packet interarrival time and packet length are constant depend on the bit rate. Figure 1 shows the traffic process aboce packet level for UGS.

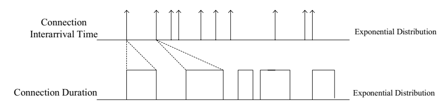

Figure 1. UGS traffic arrival process above packet level

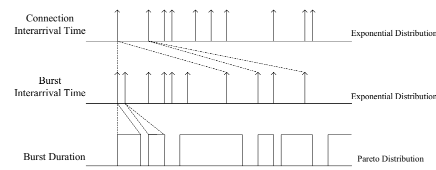

Figure 2. High speed fractal traffic arrival process above packet level

2.2 rtPS and nrtPS

The rtPS and nrtPS services are mostly the variable bit rate video and FTP. Both classes of traffic are high-speed fractal (self-similarity) traffic characterized by burst on many timescales. Therefore the model for this type of traffic should involve modeling the long bursts [2].The Poisson Pareto burst process (PPBP) is used to model the burst for such traffic [11]. It is basically a Poisson process with certain rate of Pareto-distributed overlapping bursts. Figure 2 shows the process above packet level for rtPS and nrtPS. Like UGS, the arrival process of rtPS and nrtPS connections can be modeled by Poisson process within one-hour intervals. Each of these arrivals reflects an individual user starting a new session. Within a connection, the arrival of bursts is a Poisson process with rate λ. The period represents the length of a burst has a Pareto distribution. During the burst duration period, the packet arrival process is constant with rate r. The Pareto distribution with shape parameter α and location parameter β has cumulative distribution function:

$$F(x) = P[X \leq x] = 1-(\frac{\beta}{x})^{\alpha} \qquad 1 < \alpha < 2, \beta > 0, x \geq \beta$$

By [11]the mean amount of work arriving within an interval of length t is $\lambda tr\alpha\beta/(\alpha-1)$,the variance of amount of work arriving in an interval of length t is

$$\sigma^2(t) = \begin{cases} 2r^2\lambda t^2(\frac{\alpha\beta}{2(\alpha-1)} - \frac{t}{\beta}) & 0 \leq t \leq \beta \\ 2r^2\lambda(\frac{\beta^3\alpha}{6(3-\alpha)} - \frac{\beta^2 t\alpha}{2(2-\alpha)} - \frac{\alpha\beta t^{3-\alpha}}{(1-\alpha)(2-\alpha)(3-\alpha)}) & t > \beta \end{cases}$$

2.3 BE

BE service is mainly about WWW traffic. The WWW traffic has also been proven to be self-similar [4]. The distri-

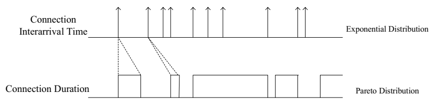

Figure 3. BE traffic arrival process above packet level

bution of document transfer size of WWW traffic is heavy-tailed with Pareto distribution. Every connection of WWW traffic represents a new request and the connection arrival is Poisson process. Figure 3 provides a picture of BE traffic arrival process above the packet level.

2.4 Gaussian Model

It has been proved in [9] [3] [1] [11] [6] that due to the central limit theorem, as the number of independent sources contributing to an aggregate flow increases, traffic weakly converges to a Gaussian process. In another words, as λ increases to above thousand, the behavior of the Poisson process for UGS and PPBP model for rtPS and nrtPS are both approach that of a Gaussian process. In 802.16 wireless MAN, from BS's view, the traffic is aggregated by numbers of end users from amount of SSs. Thus traffic process at BS in 802.16 wireless MAN, not matter what traffic model for each class of traffic, can be approximately as Gaussian model at BS's side due to the property of wide area network. This conclusion is very useful for our theoretical analysis in the later section.

3 Blocking probability Guarantee at call level or burst level

In order to guarantee QoS of four classes of services in 802.16 wireless MAN, admission control should be made at the connection level or burst level to help efficient time slot allocation at packet level. How bandwidth should be allocated for different connections with certain QoS requirement? Before we give the answer to this question, let's analyze the blocking probability at connection level or burst level for UGS, rtPS, nrtPS and BE. For low bit rate connections that belong to UGS and BE services, we are concerned with the connection level. While for large bit rate fractal traffic such as rtPS and nrtPS where the duration of a connection is related to the bursts occurring within the connection, we are concerned with blocking probability at the burst level.

Erlang formula has been widely used to calculate the blocking probability in telephone network. Erlang formula is derived by modeling the number of voice call in the system as a Markov chain process since the voice call holding time is exponential distribution which is memoryless. In 802.16 wireless MAN, a BS charges thousands of SSs, and in very one SS, there are hundreds and thousands of end users, which makes the N (number of channels needed)in the Erlang formula very large. It consequently leads to numerical difficulties in the use of Erlang formula for calculating the blocking probability of UGS service. Moreover, as we stated before, the connection/burst duration for fractal traffic (rtPS, nrtPS, and BE) is not exponential distribution anymore, thus we cannot use Markov Chain method to analyze the performance. This means in this case, the Erlang B formula is not useful in calculating the blocking probability for fractal traffic. In this section, we use Chernoff bound method to derive upper bound blocking probability for four classes of traffic in 802.16. The method is proved in theory and simulation (in the later section) to be a very good approximation and efficient compared to Erlang B for blocking probability of UGS traffic when the number of SSs is up to thousand. And

Consider the number of SSs in a BS area is $N(1 \le j \le N)$, the number of classes of traffic is $4(1 \le i \le 4)$. The amount of traffic of class i arriving at SS j requesting for transmission from BS per second is $a_{ij}(1 \le i \le 4, 1 \le j \le N)$, where $a_{ij}(1 \le i \le 4, 1 \le j \le N)$ are independent random variables. $a_i = \sum_{j=1}^{n} a_{ij}$is the total amount of traffic of class i arriving at BS from all its SSs. $a_i(1 \le i \le 4)$ are also independent random variables. Assume the bandwidth demanded per second for traffic i to ensure the blocking probability less than or equal to p_i is C_i, then

$$P(a_i \ge C_i) \le p_i \tag{1}$$

Let us denote the pdf of a_i as $f_{a_i}(w)$. The upper bound blocking probability at BS of class can be estimated by the Chernoff Bound [7]. That is

$$P(a_i \ge C_i) = \int_{C_i}^{\infty} f_{a_i}(w)dw = \int_{C_i}^{\infty} u(w-C_i) f_{a_i}(w)dw \tag{2}$$

Obviously, for $t \ge 0$ the unit step function if bounded by the following exponential relation:

$$u(w - C_i) \le e^{t(w-C_i} \tag{3}$$

Substituting (3) in (2), we have

$$P(\textstyle\sum_{j=1}^{N} a_{ij} \ge C_i) \le e^{-tC_i} \int_{c_i}^{\infty} e^{tw} f_{a_i}(w)dw \quad t \ge 0 \tag{4}$$

Let us denote the moment generating function of $f_{a_i}(w)$ is $M_{a_i}(t)$, then (4) is

$$P(a_i \ge C_i) \le e^{-tC_i} M_{ai}(t) \quad t \ge 0 \tag{5}$$

We simply differentiate the left side of (5) and set it equal to zero. Thus, we find the optimal t^* that yields the tightest upper bound for blocking probability, which is equal to p_i, that is

$$p_i = inf_{t\geq 0}\{e^{-tC_i}M_{a_i}(t)\} \qquad (6)$$

The most important step to get the upper bound for blocking probabiltiy is to figure out the moment generating function $M_{ai}(t)$ of distribution of traffic arriving at BS.

We denote the number of connections/bursts of class i arriving at SS j per second as λ_{ij}, and the duration of a connection/burst of class i as t_i (unit: second). We assume transmission rate during that random time is a constant. Let r_i be the rate of class i. By Little's law [7], the average amount of connections/bursts in a stable system (over some time interval) is equal to their average arrival rate, multiplied by their average time in the system. Thus, if the blocking probability is p_i, the average amount of traffic of class i arrived at SS $j(a_{ij})$ is $\lambda_{ij}(1-p_i)t_ir_i$. Since, λ_{ij} and t_i are two independent variables, the mean of random variable $a_{ij}$is $\mu_{ij} = r_i(1-p_i)E(\lambda ij)E(t_i)$ and variance $\sigma_{ij}^2 = r_i^2(1-p_i)^2(E(\lambda_{ij}^2)E(t_i^2) - E(\lambda_{ij})^2E(t_i)^2)$ As we mentioned in the last section, from the BS's view, the distribution of amount $(a_i = \sum_{j=1}^{N} a_{ij})$ of traffic of class i arriving at BS per unit time is approximately a Gaussian distribution with mean $\mu_i = \sum_{j=1}^{N}\mu_{ij}$and variance $\sigma_i^2 = \sum_{j=1}^{N}\sigma_{ij}^2$. Thus, (6) is equal to

$$inf_{t\geq 0}\{e^{\sum_{j=1}^{N}\mu_{ij}t + \sum_{j=1}^{N}\sigma_{ij}^2 t^2/2 - tC_i}\} = p_i \qquad (7)$$

Simplify (7), we get

$$t^* = \frac{C_i - \sum_{j=1}^{N}\mu_{ij}}{\sum_{j=1}^{N}\sigma_{ij}^2} \qquad (8)$$

Substituting (8)in (7), yields the relationship of allocated bandwidth and the upper bound of blocking probability for class i.

Furthermore, we get

$$C_i = \sum_{j=1}^{N}\mu_{ij} \pm \sqrt{2\sum_{j=1}^{N}\sigma_{ij}^2(-\lg p_i)} \qquad (9)$$

Clearly, if the upper bound blocking probability p_i of traffic i increase, the bandwidth needed is decreased. Furthermore, for the two classes of traffic whose arrival processes have the same distribution and with the same mean arrival rate, the traffic with larger variance actually needs more bandwidth. Thus, (9) is reduced to

$$C_i = \sum_{j=1}^{N}\mu ij + \sqrt{2\sum_{j=1}^{N}\sigma_{ij}^2(-\lg p_i)} \qquad (10)$$

Using (10), we could determine the amount of traffic needed for all kinds of traffic to ensure the given upper bound blocking probability.

```
Min = 0;
Max = 1;
Value = Min;
While(true)
{
   Value = p_1;
   Mid = (Min+Max)/2;
   p_1 = Mid;
   if ((∑_{i=1}^4 C_i ≤ C)&&|p_1 − Value| < 10^{−4})
      Break;
      Return(p_1, C_i);
   else
      if (∑_{i=1}^4 C_i ≤ C): Max = Mid;
      if (∑_{i=1}^4 C_i > C): Min = Mid;
}
```

Table 1. Binary Search algorithm for calculating C_i and p_i

4 Admission control and bandwidth allocation above packet level

Using formula derived in the last section, we propose an admission control and bandwidth allocation mechanism for multimedia traffic in 802.16 wireless MAN. Simulation and theoretical results are compared at the later part of this section.

We use a binary search approach to solve the problem that given a total bandwidth, fairly allocating bandwidth to each class of multimedia traffic in 802.16 MAN, such that their upper bound connection/burst blocking probability is minimized. We assume the blocking probability of four class of traffic has linear relationship, i.e.$p_1 : p_2 : p_3 : p_4 = k_1 : k_2 : k3 : k_4$. The algorithm is illustrated in TABLE 1.

Our admission control and resource allocation mechanism functions as following. The total bandwidth (C) is completely portioned for four classes of traffic, and the partition value $C_i(i = 1, 2, 3, 4)$ is calculated by above binary search algorithm using Chernoff bound. If a new connection of UGS and BE arrives at SS, it send request to BS for bandwidth. It is granted bandwidth if the new aggregated bandwidth including this connection is less than $C_i(i = 1, 4)$. Else, it is blocked. If a new connection of rtPS and nrtPS arrives at SS, the connection is always admitted, but the burst within the connection will be blocked when the total used bandwidth larger than $C_i(i = 2, 3)$.

This mechanism can guarantee the pre-required upper bound blocking probability for UGS and BE connections, and the burst blocking probability for rtPS and nrtPS.

In the following part, we compare the theoretical results

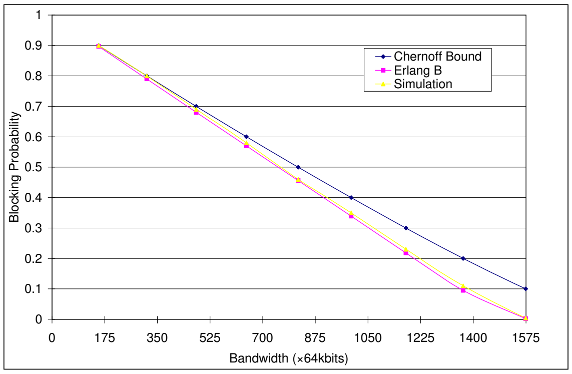

Figure 4. Blocking Probability vs. Bandwidth for UGS(Number of SSs: 500)

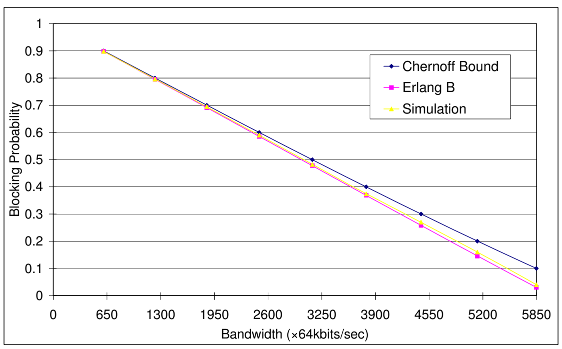

Figure 5. Blocking Probability vs. Bandwidth for UGS(Number of SSs: 2000)

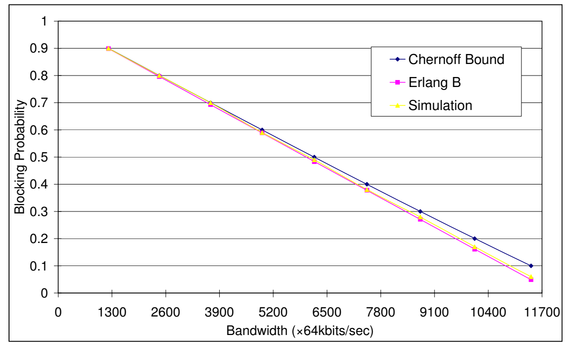

Figure 6. Blocking Probability vs. Bandwidth for UGS(Number of SSs: 4000)

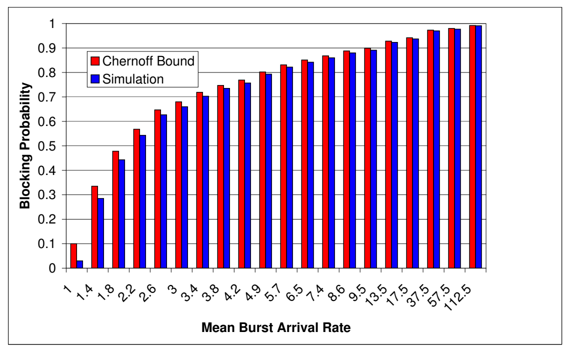

Figure 7. Blocking Probability vs. Burst Mean Arrival Rate ($\alpha = 1.5, \beta = 0.25$)

and simulation result to see the reliability of our propose mechanism.

Figure 4, 5,and 6 show the comparison of UGS connection blocking probabilities got from Chernoff bound, Erlang B formula and Simulation as a function of amount of bandwidth under different number of SSs in a BS area: 500, 2000, 4000. The mean arrival rate of UGS connection is 1 per second. The duration of UGS connection is 3. We can see that blocking probabilities got from simulation and Erlang B formula are tightly upper-bounded by Cheroff method. Moreover, comparing the three figures we can see that as the number of SSs increasing, the blocking probabilies got from Chernoff bound method is more closely to those of from Erlang B and simulation. This is because we use Gaussian distribution to approximate the traffic arrival at BS, with the number of SSs increasing, the distribution of amount of traffic arrived at BS is more like a Gaussian distribution. Thus, the probability computed by Chernoff bound method is more accurate. We conclude that when the network is small such as hundreds of end users in a BS area, Erlang B is good formula to calculate the blocking probability. But when the network is large such as in 802.16 wireless MAN, there are thousands of SSs and millions of end users in a BS area, Erlang B formula takes tons of time in numerical computing. In this case, Chernoff bound method is efficient and good approximate method for calculating the blocking probability.

Figure 8 shows the comparison of burst/connection blocking probability got from Chernoff bound and simulation as function of mean interarrival rate for fractal traffic (rtPS, nrtPS and BS) when the number of SSs equals to 2000, the shape parameter is 1.5, and location parameter is 0.25. We can see that as the mean interarrival rate increases, the blocking probability increases with the decreased increasing rate. The blocking probability from Chernoff Bound is closer to the simulation result as the mean arrival rate increases, which is the same as the results from UGS.

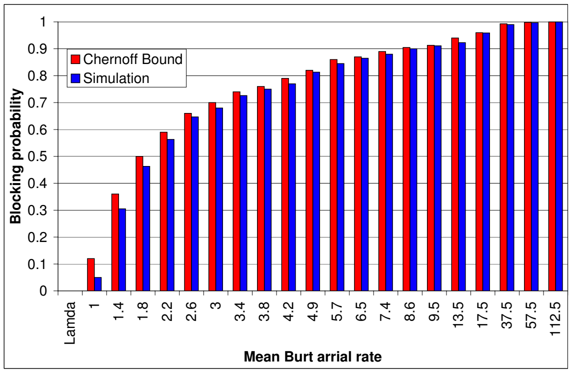

Figure 8. Blocking Probability vs. Burst Mean Arrival Rate ($\alpha = 1.6, \beta = 0.4$)

5 Conclusion

In this paper, we have examined method of using Chernoff bound to calculate the upper bound blocking probability for fractal traffic and non-fractal traffic in large wireless network, and we have found it to be a very good approximation for calculating the blocking probability for fractal traffic in large size wireless network and a good replacement for non fractal traffic when the Erlang B computation waste too much time. We have used the method of computing upper bound blocking probability and binary search to propose an admission control and bandwidth allocation mechanism for 802.16 wireless MAN, and this mechanism had been proven to guarantee the blocking probability of each class of services very well in theory and simulation.

Acknowledgment

This work has been supported by the Ohio Board of Regents Doctoral Enhancements Funds.

References

[1] R. Addie, P. Mannersalo, and I. Norros. Performance formulae for queues with gaussian input. In *Proceedings of ITC 16*, page 1169C1178, June 1999.

[2] R. G. Addie and T. D. N. M. Zukerman. Broadband traffic modeling: Simple solutions to hard problems. *IEEE Communications Magazine*, 8:88–95, Augeust 1998.

[3] J. Choe and N. B. Shroff. A central limit theorem based approach to analyze queue behavior in atm networks. *IEEE/ACM Transactions on Networking*, 6:659C671, Oct. 1998.

[4] M. Crovella and A. Bestavros. Explaining world wide web traffic self-similarity. In *Computer Science Department, Boston Univerity, Revised Technical Report TR-95-015*, October 1995.

[5] C. Eklund, R. B. Marks, K. L. Stanwood, and S. Wang. IEEE standard 802.16: A technical overview of the $wirelessman^{TM}$ air interface for broadband wireless access. *IEEE Communications Magazine*, 40:1–15, June 2002.

[6] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva. On the weak convergence of long range dependent traffic process. *Journal of Statistical planning and Inference*, 80:155–171, 1999.

[7] L. Kleinrock. *Queueing Systems*, volume 1.

[8] W. E. Leland. On the self-similar nature of ethernet traffic (extended version). *IEEE Communications Magazine*, 40:1–15, June 2002.

[9] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. Robbins. Performance models of statistical multiplexing in packet video communications. *IEEE Transactions on Communications*, 36:834–844, July 1988.

[10] IEEE 802.16 working group. *IEEE Std 802.16-2004 (Revision of IEEE Std 802.16-2001): IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems*. IEEE Press.

[11] M. Zukerman, T. D. Neame, and R. G. Addie. Internet traffic modeling and future technology implications. In *Proceedings of IEEE INFOCOM*, 2003.