Int. J. Communications, Network and System Sciences, **2010, 3, 631-637**
doi:10.4236/ijcns.2010.37085 Published Online July 2010 (http://www.SciRP.org/journal/ijcns/).

Scientific Research

Modeling and Analysis of Bandwidth Allocation in IEEE 802.16 MAC: A Stochastic Reward Net Approach

Shanmugam Geetha, Raman Jayaparvathy
Department of EEE, Coimbatore Institute of Technology, Coimbatore, India
*E-mail: geetha_thiagu@yahoo.com, jayaparvathy*14*@gmail.com*
Received April 20, 2010; *revised May* 27, 2010; *accepted July* 2, 2010

Abstract

In this paper, we present a stochastic reward net (SRN) approach to analyse the performance of IEEE 802.16 MAC with multiple traffic classes. The SRN model captures the quality of service requirements of the traffic classes. The model also takes into account pre-emption, priority and timeout characteristics associated with the traffic classes under consideration. The performance of the system is evaluated in terms of mean delay and normalized throughput considering the on-off traffic model. Our analytical model is validated by simulations.

Keywords: Wimax, IEEE 802.16, Stochastic Reward Net, Mean Delay, Throughput

1. Introduction

Over the last few years there has been tremendous increase in the use of broadband access. The deployment has boosted the usage of several multimedia applications such as Voice over IP (VoIP), online gaming and Video on Demand (VoD). However, in the rural and suburban areas, deployment of traditional wired technologies is too expensive. In such cases, broadband wireless access (BWA) based on IEEE 802.16 provides a promising solution [1,2]. One of the key features of IEEE 802.16 is that it supports multiple applications such as HDTV, video conference and conventional internet applications. The challenge for BWA networks is to simultaneously provide quality of service (QoS) to applications with very different characteristics. Hence, a proper resource allocation scheme for packet transmission is imperatively needed.

Performance evaluation of resource allocation mechanisms plays an important role in design of communication systems. Increasing complexity of networks and the way in which they are used, has made it difficult to construct models that are analytically tractable. SRNs are very useful in analytical modeling of complex networks. System operations can be precisely described by means of a graph which translates into a markovian model. Properties such as liveness and deadlock freeness make SRN a reliable analytical modeling tool.

SRN has been used extensively for performance modeling. Performance of opportunistic and non opportunistic schedulers was compared in [3] using analytical model developed with stochastic Petri net (SPN). A protocol of QoS has been developed using Petri net in [4]. The protocol has been verified for service guarantee and effective use of resources. Modeling power, analysis and verification of SPN has been discussed in [5]. Application of Petri net (PN) in performance and availability analysis is discussed in [6]. The authors in [7] presented a SRN approach to model IEEE 802.11 DCF with on-off traffic model. Performance metrics such as mean delay and average system throughput have been evaluated. Reconfigurable PN and their ability to model dynamic systems have been studied in [8].

Several approaches have been used for performance evaluation of IEEE 802.16 networks. Simulation approach has been followed in [9] for evaluating IEEE 802.16 system metrics such as mean delay and throughput. Analytical approach to study bandwidth allocation process has been presented in [10,11]. Packet scheduling scheme for QoS provisioning in WiMax networks is discussed in [12]. The proposed scheme in [12] has been verified using simulations. In [13], authors have proposed a Light WiMAX simulator (LWX) for evaluating performance of IEEE 802.16 bandwidth allocation algorithms. Simulation approach has been adopted in [14] to compare various scheduling schemes such as round-robin, token bucket-based and M-LWDF algorithms. Authors in [15] have proposed an intelligent bandwidth allocation of uplink (IBAU) for WiMax systems. IBAU mechanism is shown to decrease delay and increase throughput of the network. A survey on scheduling

schemes in IEEE 802.16e systems has been presented in [16]. Simulation methodologies to be adopted for MAC and PHY layers of IEEE 802.16 are presented in [17].

In this paper, we propose a SRN approach to model and analyze performance of the IEEE 802.16 MAC with multiple traffic classes. The proposed model incorporates prioritization and pre-emption of traffic classes. Packet drop due to waiting time exceeding threshold is also considered. We compute the average system throughput and mean delay suffered by the first packet (*i.e.*, the packet in the head of line (HOL) of each queue, through the proposed SRN formulation. Mean delay of subsequent packets is determined by modelling each queue as M/G/1 queue [7]. The mean service time for the computation is obtained from the mean delay suffered by the HOL packet. Our analytical model is validated by comparing the results with simulations carried out using event based simulator.

The rest of the paper is organized as follows: Section 2 presents a brief overview of IEEE 802.16 MAC. System model is presented in Section 3. Section 4 discusses the performance evaluation. Results and discussion are presented in Section 5. Conclusions are drawn in Section 6.

2. IEEE 802.16 MAC

IEEE 802.16 system consists of two kinds of fixed stations: subscriber station (SS) and base station (BS). All communication in the network is regulated by BS. Two direction of communication path exists between BS and SS: uplink (from SS to BS) and downlink (from BS to SS). IEEE 802.16 MAC defines QoS signaling mechanisms and functions that control BS and SS data transmissions. Two modes of sharing the wireless medium is possible: Point-to-Multipoint (PMP) and Mesh. In PMP, BS serves a set of SS in a broadcast manner. Coordination of transmissions from SSs is done by BS. In mesh mode, organization of nodes is in ad hoc manner and communication exists between SS. In this paper, we focus on PMP mode.

The IEEE 802.16 MAC defines four different scheduling service flows in order to meet the QoS requirements of multimedia applications [9]. *Unsolicited Grant Service* (UGS) is designed to support real-time applications, with strict delay requirements which generate fixed-size packets at periodic intervals such as T1/E1. *Real-time Polling Service* (rtPS) is designed to support real-time applications with less stringent delay requirements, which generate variable size packets at periodic intervals, such as VoIP with silence suppression. *Non-real-time Polling Service* (nrtPS) support non-real-time variable bit rate services, such as FTP. *Best Effort* (BE) traffic does not have QoS guarantees, such as HTTP. Since rtPS, nrtPS and BE traffic classes have varying bandwidth requirements; bandwidth allocation for these classes is performed dynamically. As UGS is allocated fixed and reserved bandwidth, dynamic reassignment of bandwidth is not required.

SS maintains separate connection for each service flow. The allocation of bandwidth by the BS to SS is based on two modes: grant per subscriber station (GPSS) and grant per connection (GPC). In GPSS, the SS obtains aggregate bandwidth for all its individual flow and in turn reallocates the bandwidth to each flow individually. In GPC, the bandwidth allocation by BS is made on per flow basic. We assume GPSS mode of operation in this paper.

3. System Model

A typical IEEE 802.16 network consists of multiple BSs. Each BS covers several SSs. Every SS is associated with multiple queues corresponding to different traffic classes. We model a single SS with three queues corresponding to rtPS, nrtPS and BE traffic classes as shown in **Figure 1**. The SS is assigned aggregate bandwidth by the BS. The three queues contend for bandwidth from the SS. The objective is to obtain the mean delay and normalized throughput of each traffic class for varying load conditions. The analytical model is required to take into account prioritization, pre-emption and dropping of packets (with waiting time exceeding the threshold) corresponding to various traffic classes.

Packets arrive at each of the queues in random epochs of time. Data packets arriving at a queue gets buffered till they gain access to channel. Newly arriving packets are added to the queue on a first come first serve (FCFS) basis. Delay of a packet is defined as the time spent by a packet till it is successfully transmitted. Normalized throughput of a given traffic class is defined as the ratio of successful packets transmitted to total packets generated. Average system throughput is the sum of throughputs of individual traffic class.

The following assumptions are made in the model.

- There are 3 different traffic classes in the system, namely rtPS, nrtPS and BE denoted as $class_1$, $class_2$ and $class_3$ respectively.
- We consider data-only traffic with on-off traffic model. Data bursts consist of active and idle periods. (Practically, a data burst represents data packet of variable

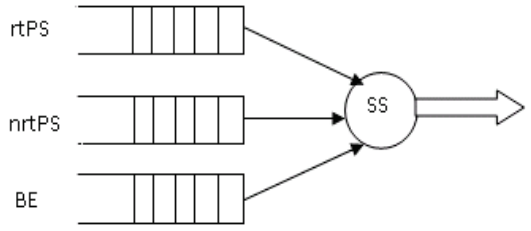

Figure 1. System model.

length, for example an IP packet with zero idle time between finite set of consecutive packets.[7])

- Data bursts arrival at any queue follows a Poisson process with mean arrival rate λ_i.
- Service times of data bursts are exponentially distributed with mean $1/\mu_i$ seconds.
- The SSs are assumed to have negligible mobility.

4. Performance Evaluation

In this section, we present a SRN model to evaluate the performance of the system considered in Section 3. Performance metrics considered are normalized throughput and mean delay suffered by a packets belonging to each traffic class.

4.1. Stochastic Reward Net Model

SRN model for a SS with three queues is shown in **Figure 2**. The model incorporates priority, pre-emption and timeout characteristics of the queues. **Tables 1-3** lists the various places, transitions and the meaning associated with each of them.

Transition usr_i generates packets at a given rate λ_i and deposits them into place q_i. An inhibitor arc with

Table 1. List of places.

Place	Meaning
cap	Total available bandwidth
usg_i	Number of channels currently in use
q_i	Packets in buffer

Table 2. List of timed transitions.

Timed Transition	Meaning
usr_i	Packet arrival at rate λ_i
end_i	Departure of packets after service at rate u_i
$time\_o_i$*	Removal of time out packets at rate $u_{to\ i}$

*$i = 1, 2$

Table 3. List of immediate transitions.

Immediate Transitions	Meaning
$chchk_i$	Priority transition checking availability of channel
$pre\_empt_{i,j}$	Enable pre-emption

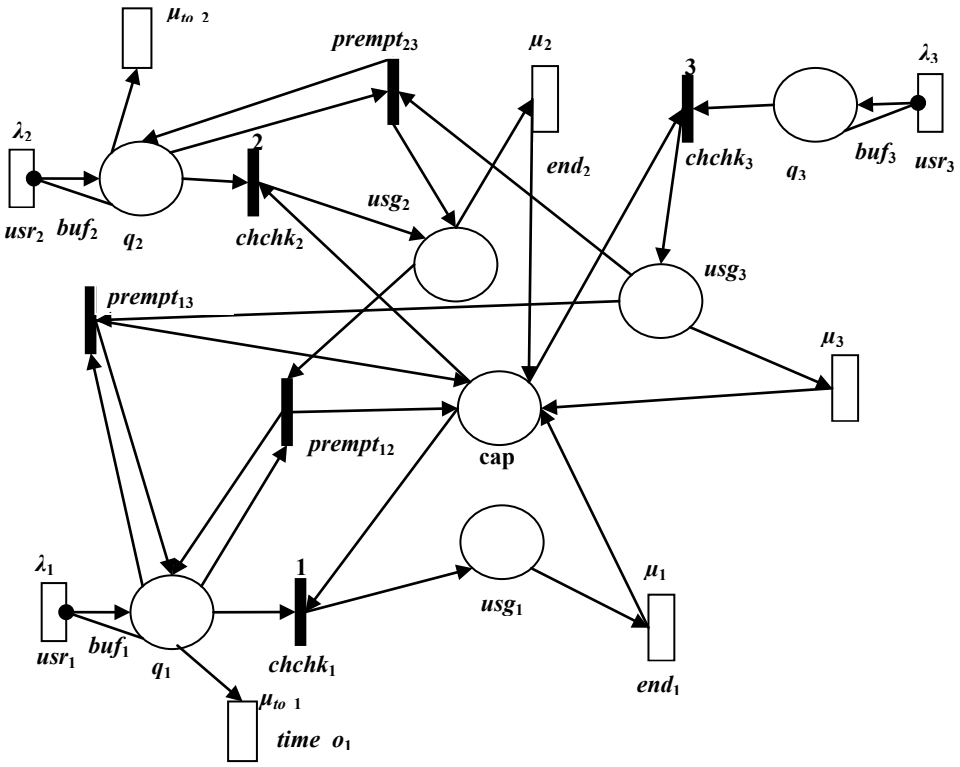

Figure 2. SRN model.

cardinality buf_i is needed to ensure that the number of packets waiting to enter the current queue is finite. If all channels are busy, the data packets are buffered in q_i with buffer size buf_i.

A way to assign priority is to give each transition an integer priority level. Transition $chchk_i$ are modelled as priority transitions. Lower integer value indicates higher priority level. A priority transition is enabled only if no other higher priority transition is enabled. Since, $chchk_1$ is assigned lowest value; $class_1$ has highest priority to gain access to channel, followed by $class_2$ and $class_3$. Firing $chchk_i$ transfers a packet from q_i to usg_i indicating the packet is being served. After completion of service time, transition end_i is fired and the channel is returned to the central pool. Note that $chchk_i$ are modelled as immediate transitions since they represent activity that does not imply time dependency. Although the action of assigning a channel implies time, the time is neglected from the point of view of traffic modelling.

In order to model pre-emption using SRN, it is required to check the simultaneous presence of a packet in place usg_{i+1} and q_i. The meaning of the above condition is that a lower priority packet is being served, when higher priority packet is waiting for resource. Transition $prempt_{i,j}$ are immediate transitions used to model pre-emption. $prempt_{i,j}$ is enabled when packets are available in places q_i and usg_j at the same time, where subscript i and j correspond to higher priority and lower priority traffic class respectively. Arc connecting $prempt_{i,j}$ indicates removal of packet from usg_j, and returning the channel to central pool of channels. Hence, firing $prempt_{i,j}$ pre-empts $class_j$ and enables $class_i$ to access the resource.

The channels available in the central pool of resource are shared by the traffic classes on arrival of data packets and returned to the pool on completion of service. At higher traffic loads, the available channels become insufficient to meet the bandwidth requirement. Under such conditions, packets in buffer wait for availability of resource. Traffic classes, $class_1$ and $class_2$, belong to delay sensitive application with maximum threshold on tolerable delay. Packets exceeding the threshold are dropped. Dropping of packets exceeding the delay limit is incorporated in the model using timed transitions $time\_o_i$. Firing rate of $time\_o_i$ is set to μ_{to_i}, $1/\mu_{to\_i}$ is the maximum tolerable delay for packets belonging to traffic $class_i$. Firing $time\_o_i$ removes a packet from q_i indicating the packet drop. Probability of packet drop depends on the available channels, transmission rates of packets, buffer size etc. Since, $class_3$ traffic is not associated with any such delay limit, we do not include time out feature for $class_3$.

4.2. Mean Delay and Normalized Throughput

The underlying continuous time markov chain (CTMC) of the SRN model discussed can be obtained from extended reachability graph (ERG) [7]. To obtain the desired performance metrics, one has to solve the CTMC. Complexity of CTMC increases with the size of the system. Solution of complex CTMC can be obtained by using standard software packages such as SHARPE [18], SPNica [19] or TimeNET [20]. The average number of packets in each place, and hence the steady state probability of occupancy of each state in the CTMC be determined using the software tools. In this paper, we use SHARPE to construct the SRN model and obtain the performance metrics.

The average throughput of a transition T is defined as the average rate at which packets are deposited by the transition in its output places. If $\hat{O}(t)$ is the average number of packets deposited by transition T in all of its output places up to a time t, then the throughput of a transition T, η_T is defined as

$$\eta_T = \lim_{t\to\infty} \frac{\hat{O}(t)}{t} \tag{1}$$

Since we consider three different traffic classes, the throughput of traffic class i, is given by

$$\eta_i = \frac{\eta_{end_i}}{\eta_{usr_i}} \tag{2}$$

Average system throughput, η is given by,

$$\eta = \sum_{i=1}^{3} \eta_i \tag{3}$$

The mean delay, $\hat{D}_H$, experienced by a HOL packet of traffic class i, is the sum of the mean packet holding time and the sum of mean waiting times in places q_i and usg_i. Let the average number of packets in place P be $\#\hat{P}$. $\hat{D}_H$ can be computed using Little's Theorem [21] as,

$$\hat{D}_{H_i} = \frac{\#(q_i)}{\eta_{usr_i}} + \frac{\#(usg_i)}{\eta_{chchk_i}} + \frac{1}{\mu_i} \tag{4}$$

where μ_i is the mean packet holding time for traffic class i. The buffer in each queue is modelled as M/G/1 queue with mean service time $\hat{D}_{H_i}$. The mean packet delay, $\hat{D}$ can be determined by applying the *Pollackzek-Kinchine* mean value formula [22] as

$$\hat{D}_i = \hat{D}_{H_i}\left[1 + \frac{\rho_{b_i}}{2\left(1-\rho_{b_i}\right)}\left(1+C_{R_i}^2\right)\right] \tag{5}$$

where . If delay of HOL packet is represented by random variable, R_i, then

$$C_{D_i}^2 = \frac{E\left[R_i^2\right]}{\hat{D}_{H_i}^{\,2}} \tag{6}$$

For small loads, $E\left[R_i^2\right]$ can be obtained as

$$E\left[R_i^2\right] = 2\left(\frac{\#usg_i}{\eta_{chchk_i}}\right)^2 \quad (7)$$

5. Results and Discussion

We evaluate the system performance in terms of mean delay and normalized throughput for increasing traffic load, ρ, given by $\sum_{i=1}^{3}\rho_i$, where ρ_i corresponds to traffic load of class$_i$ for $i = 1, 2$ *and* 3. $\rho_i = \lambda_i / \mu_i$, where λ_i is the arrival rate and μ_i is the service rate of each traffic class. Simulation parameters are shown in **Table 4**. Input traffic parameter settings are given in **Table 5.**

We compare the analysis and simulation results for three traffic classes in terms of mean delay and normalized throughput. From the results we find the simulation results match with the analysis, thus validating our analytical approach. We also analyse the performance of the system with varying buffer sizes.

Figure 3 presents a comparison of mean delay for three traffic classes with increasing traffic load. It is observed that the mean delay increases with traffic load. Mean delay suffered by packet of class$_1$ is least followed by class$_2$ and class$_3$. The increase in mean delay is more pronounced for class$_3$ since class$_3$ has the least priority among the competing traffic classes. At higher loads, class$_3$ packets are starved are resources which results in increased mean delay.

We further analyse the system with increased buffer size. **Figure 4** shows the comparison of mean delay for *buf* = 15. From the figure it is observed that with increasing buffer size there is no significant increase the mean delay of class$_1$ traffic because the packets belonging to class$_1$ have to wait for minimum amount of time to gain access to the channel. Further, since class$_1$ and class$_2$ packets are associated with a maximum tolerable delay,

Table 4. Simulation parameters.

Cell Radius	1 km
Duplexing Schemes	TDD
Ratio of Uplink slots to downlink in TDD	50%
Total available bandwidth	50 Mbps
Simulation time	500 s

Table 5. Input traffic parameters.

Traffic Class	Latency (ms)	Packet Size (Bytes)	Packet Interval (ms)	Traffic load (Kbps)	Mean Service Time (ms)
rtPS	8	240	2.6	2.8-20	0.6
nrtPS	10	120	3	2-10	0.5
BE	-	120	5	2-14	0.3

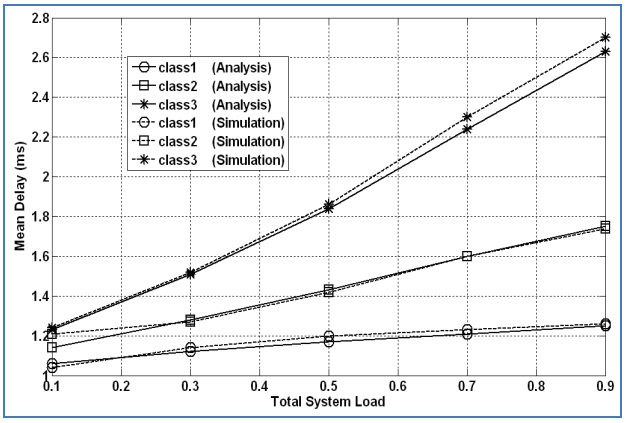

Figure 3. Comparison of mean delay (*buf* = 1).

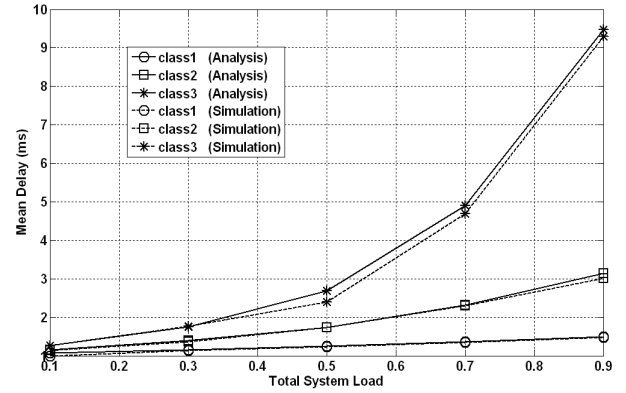

Figure 4. Comparison of mean delay (*buf* = 15).

packets exceeding the tolerable delay are dropped. Dropped packets introduce a decrease in throughput as observed in **Figure 7.**

Mean delay of class$_2$ and class$_3$ for varying buffer sizes is presented in **Figures 5** and **6**. From **Figure 6** we find that for a traffic load of 0.8, mean delay with *buf* = 1 and *buf* = 5 are 2.5 and 5.6 respectively resulting in 55% increase. For the same traffic load mean delay with *buf* = 10 and *buf* =15 are 6.9 and 7.2 respectively producing only a 4% increase. We observe that increase in buffer size does not produce a corresponding increase mean delay, particularly for higher values of buffer sizes. The reason is that the available bandwidth is insufficient to serve all packets in buffer. Hence, the number of packets successfully transmitted, which amounts to mean delay, does not increase significantly with increase in buffer size. Further, existing packets in buffer prevent additional packets entering the system.

Figures 7 and **8** present the normalized throughput of the three traffic classes with buffer size 1 and 15 respectively. From the graphs, it is observed that for a given buffer size, class$_1$ has the highest throughput followed by class$_2$ and class$_3$. Further, throughput of all traffic classes decrease with increase in traffic load. Comparing **Figures 7**

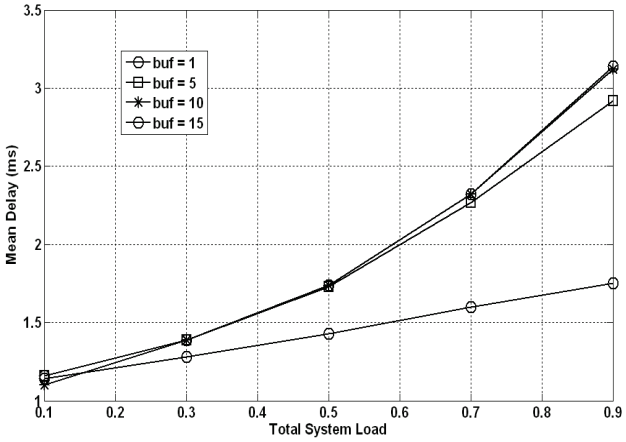

Figure 5. Mean delay of class$_2$ traffic for varying buffer size.

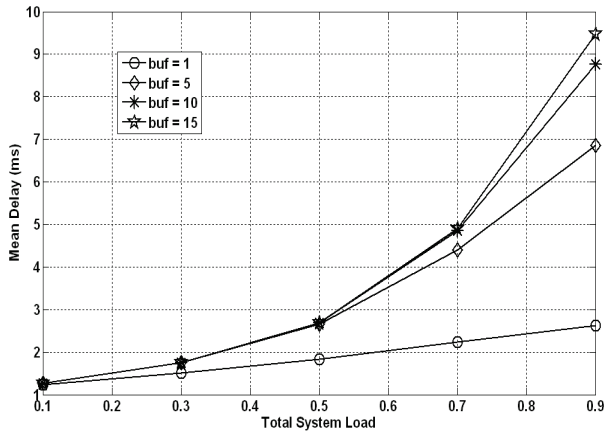

Figure 6. Mean delay of classs$_3$ traffic for varying buffer size.

and **8**, we find that increase in buffer size from 1 to 15 increases the throughput significantly. Decrease in throughput of class$_1$ traffic at higher traffic load is attributed to insufficient bandwidth. Also, class$_2$ and class$_3$ traffic suffer additional decrease in throughput due to pre-emption.

In **Figure 9**, presents the throughput of class$_3$ packets with increasing buffer sizes. From the graph it is observed that increasing buffer size from 1 to 5 increases the throughput significantly. But, further increase in buffer size from 10 to 15 does not produce any considerable increase in throughput. Further increase in buffer size results in saturation of the system with no further increase in throughput.

6. Conclusions

We presented a SRN formulation for performance evaluation of bandwidth allocation in IEEE 802.16 network considering multiple traffic classes. The model includes

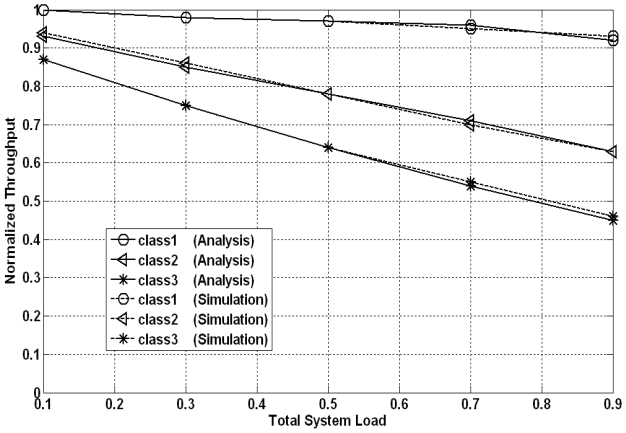

Figure 7. Comparison of normalized throughput (*buf* = 1).

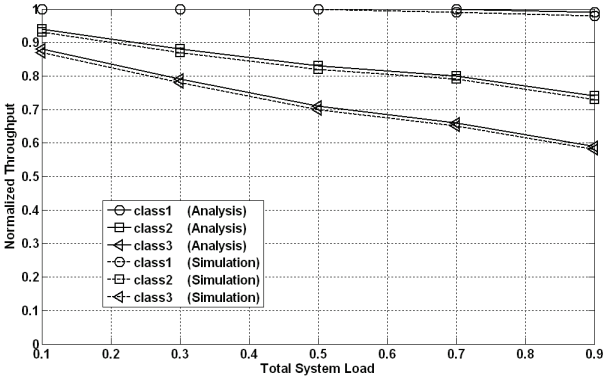

Figure 8. Comparison of normalized throughput (*buf* = 15).

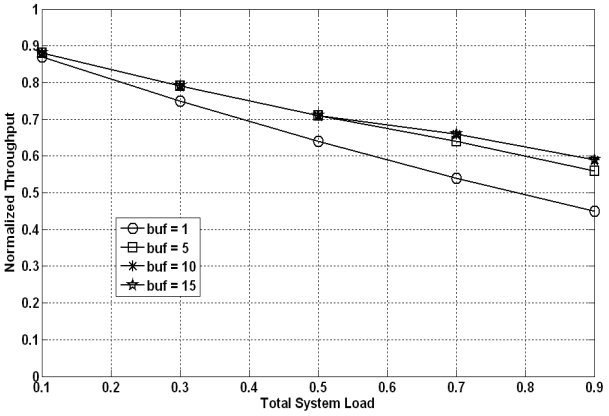

Figure 9. Normalized throughput of BE traffic class for varying buffer size.

priority, pre-emption and time-out characteristics of traffic classes. Performance of the system is evaluated in terms of mean delay and normalized throughput. Our model is validated by using simulations. The model can be extended to include more than three traffic classes. The model can be generalized to incorporate multiple SSs.

7. References

[1] IEEE 802.16-2004, "IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed Broadband Wireless Access Systems," IEEE, October 2004.

[2] IEEE 802.16e-2005, "Amendment and Corrigendum to IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed Broadband Wireless Access Systems," IEEE, February 2006.

[3] L. Lei, C. Lin, J. Cai and X. Shen, "Performance Analysis of Wireless Opportunistic Schedulers Using Stochastic Petri Nets," *IEEE Transactions on Wireless Communications*, Vol. 8, No. 4, 2009, pp. 2076-2087.

[4] D. Lee and J. Baik, "QoS Protocol Verification Using Petri-Net for Seamless Mobility in a Ubiquitous Environment: A Case Study," *International Conference on Software Engineering*, *Artificial Intelligence*, *Networking*, *and Parallel/Distributed Computing*, Phuket, August 2008, pp. 617-622.

[5] P. J. Haas, "Stochastic Petri Nets for Modelling and Simulation," *Proceedings of the* 2004 *Winter Simulation Conference*, Washington, DC, 2004, pp. 101-112.

[6] Y. Ma, J. J. Han and K. S. Trivedi, "Composite Performance and Availability Analysis of Wireless Communication Networks," *IEEE Transactions on Vehicular Technology*, Vol. 50, No. 5, 2001, pp. 1216-1223.

[7] R. Jayaparvathy, S. Anand, S. Dharmaraja and S. Srikanth, "Performance Analysis of IEEE 802.11 DCF with Stochastic Reward Nets," *International Journal of Communication Systems*, Vol. 20, No. 3, 2007, pp. 273-296.

[8] M. Llorens and J. Oliver, "Structural and Dynamic Changes in Concurrent Systems: Reconfigurable Petri Nets," *IEEE Transactions on Computers*, Vol. 53, No. 9, 2004, pp. 1147-1158.

[9] 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, 2007, pp. 26-38.

[10] Q. Ni, A. Vinel, Y. Xiao, A. Turlikov and T. Jiang, "Investigation of Bandwidth Request Mechanisms under Point-to-Multipoint Mode of WiMAX Networks," *IEEE Communications Magazine*, Vol. 45, No. 5, 2007, pp. 132-138.

[11] Y. P. Fallah, F. Agharebparast, M. Minhas, H. M. Alnuweiri and V. C. M. Leung, "Analytical Modelling of Contention-Based Bandwidth Request Mechanism in IEEE 802.16 Wireless Networks," *IEEE Transactions on Vehicular Technology*, Vol. 57, No. 5, 2008, pp. 3094-3107.

[12] M. Sarkar and H. Sachdeva, "A QoS Aware Packet Scheduling Scheme for WiMAX," *Proceedings of IAENG Conference on World Congress on Engineering and Computer Science*, *Berkeley*, California, USA, October 2009.

[13] Y.-C. Lai and Y.-H. Chen, "Designing and Implementing an IEEE 802.16 Network Simulator for Performance Evaluation of Bandwidth Allocation Algorithms," *Proceedings of the* 11*th IEEE International Conference on High Performance Computing and Communications*, Seoul, 2009, pp. 432-437.

[14] A. Bestetti, G. Giambene and S. Hadzic, "Fair Traffic Scheduling for WiMAX Systems," 6*th International Symposium on Wireless Communication Systems*, Tuscany, September 7-10, 2009, pp. 254-258.

[15] S. Z. Tao and A. Gani, "Intelligent Uplink Bandwidth Allocation Based on PMP Mode for WiMAX," *Proceedings of the* 2009 *International Conference on Computer Technology and Development*, Malaysia, 2009, pp. 86-90.

[16] C. So-In, R. Jain and A.-K. Tamimi, "Scheduling in IEEE 802.16e Mobile WiMAX Networks: Key Issues and a Survey," *IEEE Journal on Selected Areas in Communications*, Vol. 27, No. 2, 2009, pp. 156-171.

[17] R. Jain, C. So-In and A.-K. Tamimi, "System-Level Modeling of IEEE 802.16E Mobile Wimax Networks: Key Issues," *IEEE Wireless Communications*, Vol. 15, No. 5, 2008, pp. 73-79.

[18] R. A. Sahner, K. S. Trivedi and A. Puliafito, "Performance and Reliability Analysis of Computer Systems: An Example-Based Approach Using the SHARPE Software Package," Kluwer Academic Publishers, Dordrecht, 1996.

[19] R. German, "Markov Regenerative Stochastic Petri Nets with General Execution Policies: Supplementary Variable Analysis, and a Prototype Tool," *Proceedings of the* 10*th International Conference on Modeling Techniques and Tools for Computer Performance Evaluation*, Palma de Mallorca, Spain, September 1998, pp. 255-266.

[20] R. German, C. Kelling, A. Zimmerman and G. Homel, "TimeNET: A Toolkit for Evaluating Non-Markovian Stochastic Petrinets," *Performance Evaluation*, Vol. 24, No. 1-2, 1995, pp. 69-87.

[21] L. Kleinrock, "Queuing Systems: Volume I, Theory," Kluwer Academic Press, Dordrecht, 1995.

[22] R. Jayaparvathy, S. Dharmaraja and S. Srikanth, "Stochastic Petri Nets in Performance Evaluation of IEEE 802.11 WLANs," *Sixth International Conference of the Association of the Asia Pacific Operational Research Societies*, New Delhi, India, December 2003, pp. 142-150.

