

COMPARATIVE STUDY OF FINITE CAPACITY QUEUE WITH COMPLEX BUFFER MANAGEMENT SCHEMES

Shakeel Ahmad¹, Bashir Ahmad², Inayatullah Babar³ and Arjamand Bano⁴

^{1,2} *Institute of Computing and Information Technology, Gomal University, Dera Ismail Khan (NWFP) Pakistan*

³ *Department of Computer Science and Information Technology University of Engineering and Technology Peshawar (NWFP) Pakistan*

⁴ *Mathematics Department, Gomal University, Dera Ismail Khan (NWFP) Pakistan*

ABSTRACT

This paper presents comparative study of finite capacity queue with complex buffer management schemes (i.e. partial buffer sharing and complete buffer partitioning Schemes) for real time multimedia based web traffic. Maximum Entropy Methodology is used to address performance issues associated with these Active Queue Buffer management schemes. This study examines the queueing behaviour of traffic generated by real time and non real time applications. In this context, finite capacity queue with complex but effective buffer management schemes have been analytically analysed. In particular this research focuses on the comparative analysis of the following two queueing systems: (a) GE/GE/1/N censored queue with a single server, finite capacity and multiple classes under a partial buffer management scheme (PBS). (b) GE/GE/1/N censored queue with a single server, finite capacity and multiple classes under a complete buffer partitioning (CBP) management scheme. The external bursty traffic and service times have been modeled using the generalised exponential (GE) distribution. The analysis has been carried out using the principle of maximum entropy (PME). The forms of the state probabilities, as well as basic performance measures such as mean queue length and blocking (or packet loss) probabilities are analytically established at equilibrium via appropriate mean value constraints. QNAP-2 based simulations have carried out to validate the results obtained from the analytical models under a wide range of parameterisation. These results specifically focus on the impact of buffer threshold positions and different class based buffer partitioning on various performance measures. In order to maintain the fair treatment to all arriving traffic, the single server works under the first-come-first-served (FCFS) service discipline. These results depict that various grades of QoS for different types of traffic can be achieved by adjusting their corresponding threshold positions (under PBS scheme) or by allocating a particular sizes of buffer (under complete partitioning scheme).

Keywords

Partial buffer sharing, complete buffer partitioning queues, bursty traffic, principle of maximum entropy, GE distribution and performance modelling.

INTRODUCTION

Over the recent year's tremendous growth have been found in Internet and wireless technologies and also have been observed marvelous success in the past few years. Its importance in every day life is a fact and can't be denied. In general traditional communication networks have been forced to merge with the data communication networks. As the today's network seems to evolve in popularity, size, and diversity in future, it is therefore of great importance to develop more advanced as well as robusted methodologies and techniques to deliver manageable, stable and reliable services in order to fulfill the ever-growing demands for the Quality-of-Service (QoS) requirements. Generally, as the number of users and the size of the Internet increases, users because of congestion problems are likely to experience increase in

response time due to longer delay, more packets loss and other performance degradation due to congestion. Currently, the network providers try to resolve and address these issues by just keeping the network utilization low, which may not, seems to be a cost-effective solution. Current trends towards rapid growth of network communication clearly indicates that IP and packet switching will continue to dominate the Internet in future, thus handling congestion more efficiently to improve the network utilization, while providing a satisfactory level of service to the users, is known to be a practical, but challenging problem (Fernandez, 2002).

Multimedia applications are generally delay sensitive and needs preferential treatment in order to satisfy a desired level of Quality of Service (QoS) constraints. Many enterprises

demand applications development using software which integrates the support for real-time applications with the support for conventional computations. Such demands have posed various challenges to network community for robust design of the communication infrastructure. In that, significant developments have been made to design networks with the ability to guarantee the QoS for the real time data (Bulteman, 2002).

Traffic generated by the multimedia applications is generally very sensitive to the transmission. End-to-end delay and delay jitter are usually introduced due to random queueing in the network routers. Traditionally finite capacity queues with tail drop (TD) mechanisms have been employed in the network routers. Such queues temporarily accommodate the arriving packets when the server is busy. The arriving packets are dropped when the queue reaches its maximum capacity. Although this technique is simple, it suffers various problems, e.g., lock out, global synchronisation and full queue (Braden, 1998). The main problem among these is the full queue which causes longer delays and makes this technique an inappropriate choice for time sensitive applications.

One of the most tailored research areas to deal with the above mentioned problems induced by congestion is Active Queue Management (AQM). Up till now many AQM schemes have been proposed but no one can claim and achieve optimal performance in all network and traffic scenarios. In fact under dynamic network and traffic conditions an AQM scheme without adaptability can hardly meet the QoS requirement. With different congestion affecting parameters like response time, packet drop probability etc. most of the AQM schemes are not considered to be an efficient in providing fairness to users, and thus think off inefficient to control non-responsive traffic to prevent Denial of Service (DoS) attacks. However, it is also important to differentiate the QoS requirements among various users or aggregations (classes) of users. Furthermore, only few among all existing AQM schemes can be directly adapted to support service differentiation such as the Differentiated Service (Diffserv) architecture. The above mentioned

congestion control and fairness issues are not only important for various applications that use wireless technologies for flexible and reliable data communication but also need due attention for the Internet infrastructure where wired backbones and centralized control entities are available.

The main aim of this research is to formulate and compare different models with multiple queue thresholds as well as different buffer partitioning schemes and examine the queueing behaviour for multimedia type traffic with specific QoS requirements under first come first served (FCFS) service discipline. To facilitate this aim, we study stable GE/GE/1/N censored queues with a single server, finite capacity and multiple classes under a complex but effective buffer management schemes .i.e. Partial Buffer Sharing (PBS) and Complete Buffer Partitioning (CBP) Management Schemes. The external bursty traffic and service time have been modeled using the Generalised Exponential (GE) distribution. The analysis has been carried out using the *Principle of Maximum Entropy* (PME). Credibility of the proposed models has been validated using QNAP-2 based simulation.

The paper is organised as follows: The maximum entropy (ME) solution for a stable GE/GE/1/N censored queue with PBS and CBP schemes is characterised in Section 2. State probabilities are presented in Section 3. Comparisons of numerical results involving Generalised Exponential (GE) inter-arrival and service time distributions are included in Section 4. Section 5 includes conclusions.

PRELIMINARIES

The Principle of ME

The principle of ME (Jaynes, 1957) provides a self-consistent method of inference for characterising an unknown but true probability distribution, subject to known (or known to exist) mean value constraints. The ME solution can be expressed in terms of a normalising constant and a product of Lagrangian coefficients corresponding to the constraints. In an information theoretic context (Jaynes, 1957) the ME solution corresponds to the maximum disorder of system states and, thus, is considered to be the least biased distribution estimate of all

solutions that satisfy the system's constraints. In sampling terms, it has been shown ((Jaynes, 1957) that, given the imposed constraints, the ME solution can be experimentally realised in overwhelmingly more ways than any other distribution. Major discrepancies between the ME distribution and the experimentally observed distribution indicate that important physical constraints have been overlooked. Conversely, experimental agreement with the ME solution represents evidence that the constraints of the system have been properly identified.

More details on entropy maximisation and its applications can be found in (Kouvatsos and Awan, 1998).

The GE Distribution

The GE distribution is an inter-event time distribution of the form (c.f., Fig. I)

$$F(t) = P(W \leq t) = 1 - \tau e^{-\sigma t}, t \geq 0$$

$$\tau = 2/(C^2 + 1),$$

$$\sigma = \tau \nu$$

where W is a mixed-time random variable (rv) of the interevent-time, whilst $(1/\nu, C^2)$ are the mean and squared coefficient of variation (SCV) of rv W . The GE distribution versatile, possessing pseudo memoryless properties which makes the solution of many GE-type queueing systems and networks analytically tractable ((Kouvatsos and Awan, 1998).

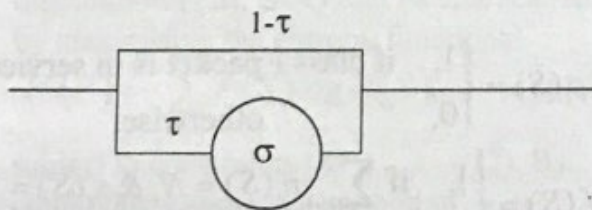


Fig.i. The GE Distribution with parameters τ and σ ($0 \leq \tau \leq 1$)

The choice of the GE distribution is further motivated by the fact that measurements of actual inter-arrival or service times may be generally limited and so only few parameters can be computed reliably. Typically, only the mean and variance may be relied upon, and thus, a choice of a distribution which implies least bias (i.e., introduction of arbitrary and, therefore, false assumptions) is that of GE-type distribution. For example, in the context of ATM networks, this model is particularly applicable in cases of traffic with low level of

correlation or where smoothing schemes are introduced at the adaptation level (e.g., for a stored video source) with the objective of minimising or even eliminating the problem of traffic correlation (Ball *et al.*, 1996). Moreover, under renewability assumptions, the GE distribution is most appropriate to model simultaneous-job arrivals at output port queues generated by different bursty sources (e.g., voice or high resolution video) with known first two moments. In this context, the burstiness of the arrival process is characterised by the SCV of the inter-arrival time or, equivalently, the size of the incoming bulk.

Partial Buffer Sharing (PBS) shown as:

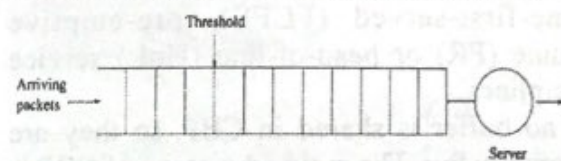


Fig. ii: Arrangement of a simple threshold based PBS scheme

PBS is a buffer management scheme which works by setting a descending sequence of thresholds N_i ($N_i > 0, i = 1, 2, \dots, R$) corresponding to R priority classes of a single server queue with finite capacity N_i . Different job loss and QoS requirements under various load conditions can be met by adjusting the threshold values. The highest priority jobs of class 1 can join the queue simply if there is space. However, lower priority jobs of class i ($i = 2, \dots, R$) can only join the queue if the total number of jobs in the queue is less than a threshold value N_i ($N_i \leq N_{i-1}$). Once the number of jobs waiting for service reaches N_i , all lower priority jobs of class k , ($k = i+1, \dots, R$) will be lost on arrival but higher priority jobs of class j ($j = 1, \dots, i-1$) will continue to join the queue until it reaches threshold value, N_j ($j = 1, \dots, i-1$). The motivation behind this arrangement is principally to try and meet the diverse QoS requirements and this is achieved by improving the loss performance of the high priority traffic while degrading the performance of the low priority. This arrangement shown by Fig. iii assumes that the buffer comprises a single FIFO queue.

Complete Buffer Partitioning (CBP) Scheme shown as:

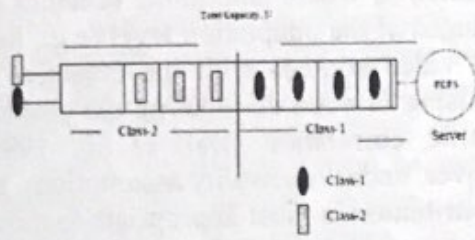


Fig.iii. The CBP management scheme with traffic classes

CBP is a buffer management scheme in which the buffer is completely partitioned into R (number of classes) portions. Jobs of any class can join only its own allocated buffer space. Jobs can be served either according to first-come-first-served (FCFS), pre-emptive resume (PR) or head-of-line (HoL) service disciplines.

As no buffer is shared in CBP, so they are inherently fair. The main advantage of CBP is its ability to achieve relatively high throughput at very high traffic load (Thareja, 1994) because of strict buffer partitioning policy in CBP scheme, it may prevent some unbalanced packets from hogging all the buffers.

ME ANALYSIS OF GE/GE/1/N QUEUE WITH COMPLETE BUFFER PARTITIONING SCHEME

This section presents the analysis of a single server GE/GE/1/N system to model a finite capacity queue with partial and complete buffer partitioning. The analysis models the bursty external traffic with compound Poisson process (CPP) and the transmission times of this traffic is represented by the GE distribution under FCFS service discipline. The total buffer capacity is N ($N > 2$) and to control the delay and delay jitter by reducing the queue length in case of:

PBS vector N represents a sequence of thresholds $\{(N_1, N_2, \dots, N_R), 0 < N_i \leq N_{i-1}, i = 2, \dots, R\}$

i. $N_{i-1}, i = 2, \dots, R\}$ to give space priorities to different classes of multimedia type traffic.

ii. **CBP** the vector N represents buffer partitioning $\{(N_1, N_2, \dots, N_R)\}$ to give separate buffer to different classes of traffic.

Notations

For each class i ($i = 1, 2, \dots, R$), let λ_i be the mean

arrival rate, C_{ai}^2 be the inter-arrival time SCV, μ_i be the mean service rate and C_{si}^2 be the service time SCV.

Focusing on a stable GE/GE/1/N/FCFS queue, let at any given time $n_i, n_i \leq N_i$ be the number of class i packets in the queue (waiting and/or receiving service)

$S = (n_1, n_2, \dots, n_R)$ be a joint queue state, where Q be the set of all
$$\sum_{i=1}^R n_i \leq N$$

feasible states $S \mathbf{n} = (n_1, n_2, \dots, n_R)$ be an aggregate joint queue state (n.b., $\mathbf{0} = 0, \dots, 0$)

Ω be the set of all feasible states \mathbf{n} .

Remarks

The arrival process for each class i ($i = 1, 2, \dots, R$) is assumed to be censored, i.e., a packet of class i will be lost if on arrival it finds N_i ($i = 1, 2, \dots, R$) packets at the queue.

For exploration purposes, the analysis that follows focuses on the FCFS with PBS and CBP is applicable in the performance modelling of networks for the effective mechanism of traffic congestion control and also for providing various QoS demands by different multimedia services.

Prior Information

For each state $S, S \in Q$ and class i ($i = 1, 2, \dots, R$) the following auxiliary functions are defined: $n_i(S)$ = the number of class i packets present in state S ,

$$s_i(S) = \begin{cases} 1, & \text{if class } i \text{ packet is in service} \\ 0, & \text{otherwise,} \end{cases}$$

$$f_i(S) = \begin{cases} 1, & \text{if } \sum_{i=1}^R n_i(S) = N_i \text{ \& } s_i(S) = 1 \\ 0, & \text{otherwise,} \end{cases}$$

Suppose what is known about the state probabilities $\{P(S)\}$ is that they satisfy the \leq Normalisation constraint

$$\sum_{S \in Q} P(S) = 1, \tag{4}$$

and that the following marginal mean value constraints per class i exist:

Server utilization, $U_i, (0 < U_i < 1)$

$$\sum_{S \in Q} s_i(S) P(S) = U_i, i = 1, 2, \dots, R; \tag{5}$$

Mean queue length, $L_i (U_i \leq L_i < N_i)$,

$$\sum_{S \in Q} n_i(S)P(S) = L_i, i = 1, 2, \dots, R; \quad (6)$$

Full buffer state probability,

$$\phi_i (0 < \phi_i < 1),$$

$$\sum_{S \in Q} f_i(S)P(S) = \phi_i, i = 1, 2, \dots, R; \quad (7)$$

satisfying the flow balance equations, namely

$$\lambda_i (1 - \pi_i) = \mu_i U_i \quad i=1, 2, 3, \dots, R; \quad (8)$$

where π_i is the blocking probability that an arriving packet of class i finds $N_i (i=1, \dots, R)$ packets in the queue (waiting or receiving service).

The choice of mean value constraints (4) - (7) is based on the type of constraints used for the ME analysis of stable multiple class queue without space priorities (Kouvatsos and Denazis, 1993). Note that if additional constraints are used, it is no longer feasible to capture a computationally efficient ME solution in closed-form. Conversely, the removal of one or more constraints from the set (4) - (7) will result into an ME solution of reduced accuracy.

A Universal Maximum Entropy Solution

A universal form of the state probability distribution $P(S), S \in Q$ can be characterised by maximising the entropy functional

$$H(P) = - \sum_s P(S) \log P(S), \quad (9)$$

subject to constraints (4) - (7). By employing Lagrange's method of undetermined multipliers, the ME solution is expressed by

$$P(S) = \frac{1}{Z} \prod_{i=1}^R g_i^{s_i(S)} x_i^{n_i(S)} y_i^{f_i(S)}, \forall S \in Q; \quad (10)$$

Where Z , the normalising constant, is clearly given by

$$Z = \sum_{S \in Q} \left(\prod_{i=1}^R g_i^{s_i(S)} x_i^{n_i(S)} y_i^{f_i(S)} \right), \quad (11)$$

and $\{g_i, x_i, y_i, i=1, 2, \dots, R\}$ are the Lagrangian coefficients corresponding to constraints (5) - (7), respectively.

Remarks

Although constraints (5) - (7) are not known priori, nevertheless it is assumed that these constraints exist. This information, therefore, has been incorporated into the ME formalism (4) - (9) in order to characterise the form of the joint state probability (10).

Aggregating (10) over all feasible states $S \in Q$ and after some manipulation, the joint aggregate ME queue length distribution $\{P(n), n \in \Omega\}$ for:

a. **CBP scheme is given by:**

$$P(0) = \frac{1}{Z} \quad (12)$$

$$P(k) = \sum_{j=1}^R \text{Pr ob}(Q_{i,k}) = \frac{1}{Z} \left(\prod_{j=1}^R x_j^k \right) \sum_{j=1}^R k_j g_j y_j^{\delta(k)} \left(\frac{N_j!}{\prod_{i=1}^R (k_i - N_j)} \right)$$

b. **PBS scheme is given by:**

$$P(0) = \frac{1}{Z}$$

$$P(k) = \sum_{i=1}^R \text{Pr ob}(Q_{i,k}) = \frac{1}{Z} \left(\prod_{j=1}^R x_j^{k_j} \right) \sum_{j=1}^R k_j \left(\frac{\sum_{i=1}^R k_i - N_j}{\prod_{i=1}^R (k_i - N_j)} \right) g_j y_j^{\delta(k)}$$

where, $\delta(k) = 1$, if $\sum_j k_j = N_i$ & $s_i(k) = 1$, or 0 , otherwise; and N_j are the threshold values for each class $j, j=1, 2, \dots, R$.

Blocking Probability

A universal form for the marginal blocking probabilities $\{\pi_i, i=1, 2, \dots, R\}$ of a stable multiple class GE/GE/1/N/FCFS queue with CBP and PBS can be approximately established, based on GE -type probabilistic arguments.

Consider a multiple class GE/GE/1/N/FCFS/PBS queue with non-zero inter-arrival time and service time stage selection probabilities

$$\sigma_i = (C_{ai}^2 + 1)/2 \quad \text{and} \quad r_i = (C_{si}^2 + 1)/2$$

respectively. Each arriving bulk of class $i(i=1,2, \dots ,R)$ joins the queue at Poisson arriving instants and finds the same aggregate number of packets as a random observer (n.b., this assumption is strictly true if the SCVs of the GE -type inter-arrival times per class are equal). Let us focus on a tagged packet within an arriving bulk of class $i(i=1,2, \dots ,R)$ which finds the queue in state $\mathbf{n}_j = (0, \dots, 0, n_j, n_{j+1}, \dots , n_R)$ where $n_k = 0, k = 1, 2, \dots, j-1$. Clearly, the total number of packets in the queue is $v = \sum_{k=j}^R n_k$ and the number of available buffer spaces is equal to $N-v$.

Using the probabilistic arguments, the blocking probabilities can be approximated:

a. In CBP scheme as:

$$\pi_i = \sum_{k=0}^N \delta_i(k) (1 - \sigma_i)^{[N-k]} P(k) \tag{16}$$

Where

$$\delta_i(k) = \begin{cases} \frac{r_i}{r_i(1 - \sigma_i) + \sigma_i}, & k = 0 \\ 1, & \text{otherwise} \end{cases}$$

b. In PBS scheme as:

$$\pi_i = \sum_{k=0}^N \delta_i(k) (1 - \sigma_i)^{[N-k]} P_{N_i}(k) \tag{17}$$

Where

$$\delta_i(k) = \begin{cases} \frac{r_i}{r_i(1 - \sigma_i) + \sigma_i}, & k = 0 \\ 1, & \text{otherwise} \end{cases}$$

Lagrangian Coefficients

The Lagrangian coefficients x_i and g_i for CBP and PBS can be approximated analytically by making asymptotic connections to an infinite capacity queues. Assuming x_i and g_i are invariant to the buffer capacity of size N , it can be established that

$$x_i = \frac{L_i - \rho_i}{L} \tag{18}$$

$$g_i = \frac{(1 - X)\rho_i}{(1 - \rho)x_i} \tag{19}$$

Where $X = \sum_{i=1}^R x_i, L = \sum_{i=1}^R L_i$ and L_i is the asymptotic marginal mean queue length of a multi-class GE/GE/1 queue (Kouvatsos and Awan, 1998). Note that statistics $L_i, i=1, 2, \dots, R$ can be determined by ((Kouvatsos and Awan, 1998).

$$L_i = \frac{\rho_i}{2} (C_a^2 + 1) + \frac{1}{2(1 - \rho)} \sum_{j=1}^R \frac{\Lambda_j}{\Lambda_j} \rho_j^2 (C_a^2 + C_g^2) \tag{20}$$

Where $\rho_i = \Lambda_i / \mu_i, \rho = \sum_{i=1}^R \rho_i$.

By substituting the value of aggregate probabilities, $P_N(n), n=0, 1, \dots, N$, and blocking probabilities, $\pi_i, i=1, 2, \dots, R$, into the flow balance condition (8), the Lagrangian coefficients $\{y_i, i=1, 2, \dots, R\}$ can be easily derived.

NUMERICAL RESULTS

This section presents comparison of numerical results which have been conducted using the proposed analytical models and simulation (based on QNAP-2 at 95% confidence interval (Veran and Potier, 1985) to evaluate the effectiveness of the already discussed buffer management schemes for multiple traffic classes. In these experiments, Generalised Exponential (GE) - type traffic has been used which can be represented by the first two moments and exhibits the burstiness property of the traffic. These experiments use two heterogeneous video applications, representing a delay sensitive live video with 2.0 Mbps arrival rate and delay tolerant video streaming with 2.5 Mbps arrival rate. The service rate (or the link capacity) remains same as 4.0 Mbps. The service rate (or the link capacity) is 4.0 Mbps.

Fig:1-a: shows an effect of threshold position settings on throughput values for both types of video streams. Increasing values of buffer threshold will also increase the throughput for delay sensitive streams whilst decreasing that of delay tolerant traffic.

Where as Fig:1-b: shows an effect of partitioning difference settings on throughput

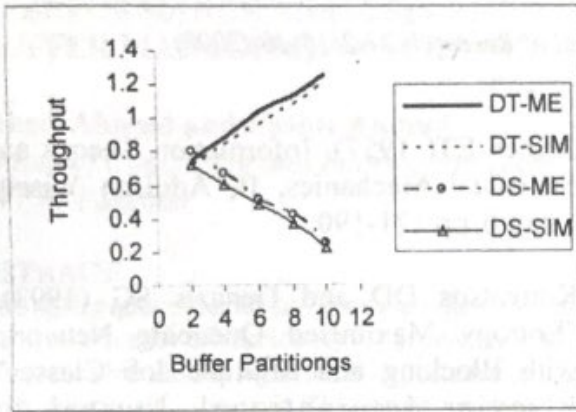


Fig1-b: CBP-Effect of Partitioning

values for both types of video streams. Increasing values of partitioning difference will decrease the throughput for delay

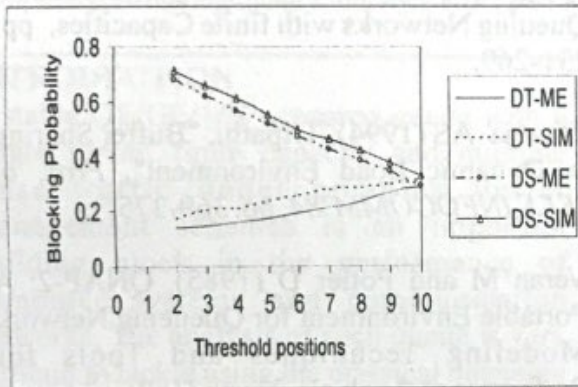


Fig-2a: PBS-Effect of threshold positions blocking probabilities

Fig-2a: shows an effect of threshold position settings on blocking probabilities for both types of video streams. Increasing values of buffer threshold will also increase the blocking probability for delay tolerant streams whilst decreasing that of delay sensitive traffic.

Where as Fig -2b: shows an effect of partitioning difference settings on blocking probabilities for both types of video streams. Increasing values of partitioning difference will decrease the blocking probability for delay tolerant streams whilst increasing that of delay sensitive traffic.

CONCLUSIONS

Comparative study of partial buffer sharing and complete buffer partitioning based queue for bursty external traffic has been presented. These models can be used to reduce end-to-end delay for traffic generated by these

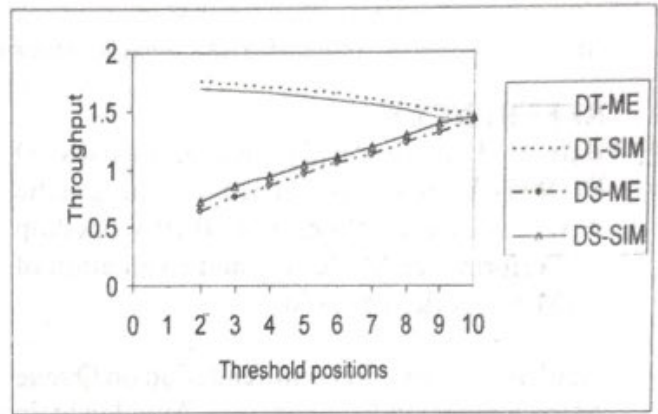


Fig-1a: PBS-Effect of threshold positions on throughput

sensitive streams whilst increasing that of delay tolerant traffic.

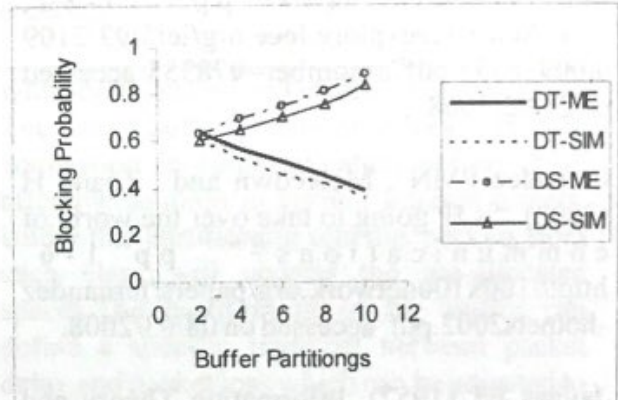


Fig2-b: CBP-Effect of Partitioning Difference on Blocking on Probabilities

applications. In this context, Product-form approximation, based on the principle of ME, for a stable GE/GE/1/N queue with FCFS scheduling discipline under PBS and CBP schemes have been proposed as a useful performance evaluation tool. These schemes effectively control the allocation of buffer to various traffic classes according to their delay constraints. Closed form analytical expressions for state probabilities have been derived. Proposed models have been implemented using the GE-type external traffic represent the bursty nature of the multimedia traffic. Typical numerical examples have been included to show the impact of buffer threshold and partitioning on different performance measures for delay sensitive video streams and delay tolerant data packets.

REFERENCES

Ball F, Hutchinson D and Kouvatsos D D(1996). VBR Video Traffic Smoothing at the AAL SAR Level, Proc. of 4th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, pp. 1-10

Braden B (1998). Recommendation on Queue Management and Congestion Avoidance in the Internet IETF RFC, 2309 April 1998, <http://www.ietf.org/rfc/rfc2309.txt> accessed on 20/03/2008.

Bulteman DCA(2002). SMIL 2.0.2. Examples and Comparisons, IEEE Multimedia 9(1), pp. 74-84, <http://www.ieeexplore.ieee.org/iel5/93/21090/00978355.pdf?arnumber=978355> accessed on 24/02/2008

Fernandez PMN , McKeown and Zhang H (2002). "is IP going to take over the world of communications?". pp 1-6. <http://100x100network.org/papers/fernandez-hotnets2002.pdf> accessed on 08/03/2008.

Jaynes ET (1957). Information Theory and Statistical Mechanics, 2nd edition, Addison Wisely. Chap11, pp. 620-630.

Jaynes ET(1957). Information Theory and Statistical Mechanics, II, Addison Wisely. Chap-6, pp.171-190.

Kouvatsos DD and Denazis SG (1993), "Entropy Maximised Queueing Networks with Blocking and Multiple Job Classes", Elsevier International Journal on Performance Evaluation: 17:189-205

Kouvatsos DD and Awan I (1998). "MEM for Arbitrary Closed Queueing Networks with RS-Blocking and Multiple Job Classes", *Annals of Operations Research* 79, Special Issue on Queueing Networks with finite Capacities, pp. 231-269.

Thareja AS (1994). Tripathi, "Buffer Sharing in Dynamic Load Environment", *Proc. of IEEE INFOCOMM '84*, pp. 369-375.

Veran M and Potier D (1985). QNAP-2: A Portable Environment for Queueing Network Modeling Techniques and Tools for Performance Analysis, North Holland, Chap-2, pp.25-63