Statistics and Applications {ISSN 2454-7395 (online)}
Volume 22, No. 2, 2024 (New Series), pp 169–180
http://www.ssca.org.in/journal

Study of Priority Based Network Nodes Using Quasi Birth and Death Process

Malla Reddy Perati and Abhilash Vollala
Department of Mathematics
Kakatiya University, Warangal, India

Received: 15 December 2022 ; Revised: 16 July 2023; Accepted: 07 December 2023

Abstract

In this paper, network node with self-similar priority based input traffic is modeled into finite buffer single server queuing system, and is analyzed through level dependent quasi birth-death (QBD) process with preemptive priority mechanism. Here, input process follows transient Markovian arrival process (MAP), and service time (packet lengths) follows Phase type (PH) distribution, which is more general than deterministic and exponential distributions. The queuing behavior of the system at arbitrary times through the performance metrics, namely, queue length, mean waiting time, and packet loss probability is investigated. For this, time dependent state probability vector of transition rate matrix is obtained using method of product integrals which in turn gives performance measures, and computational complexity of analysis is presented. This type of analysis is useful in dimensioning the network node to provide Quality of Service (QoS) guarantee.

Key words: Self-similar; Quasi birth and death process; Markovain arrival process; Phase type; Transition rate matrix; Waiting time; Loss probability.

AMS Subject Classifications: 60G18, 60K25, 68M20

1. Introduction

Performance of communication system depends on network nodes. The network nodes namely, switch, router, and multiplexer in B-ISDN (Broadband Integrated Switching Digital Network), play a vital role in communication, and therefore it is essential to analyze the performance of nodes for providing QoS. In general, analysis of network nodes is made by queueing methods, and this queueing based analysis has a long history of success in planning and dimensioning of networks. The fundamental studies of network traffic namely LAN (Leland *et al.*, 1994), WAN (Paxson and Floyd, 1995), and WWW (Crovella and Bestavros, 1997) at AT & T Bell labs disclosed that these traffic are self-similar, and degrade performance of system. It is clear that self-similar nature of traffic is emulated by homogeneous Markovian Modulated Poisson process which was superposition of Interrupted Poisson Process (IPP) or Switched Poisson Process (SPP) over different time scales. In the papers

Corresponding Author: Abhilash Vollala
Email: ashwesa.abhi@gmail.com

(Andersen and Nielsen, 1998; Yoshihara *et al.*, 2001; Shao *et al.*, 2005), performance analysis was made under steady state conditions, as such is not so useful for real time network traffic analysis. Recently, Abhilash and Malla Reddy (2022) proposed a fitting procedure for time dependent Markovian process, namely, MMPP with Sinusoidal arrival rates based on second order statistics, and proved that resultant MMPP emulates self-similar nature of network traffic in prescribed time scales. On the other hand, in B-ISDN, high demand causes congestion, and pertinent issues can be handled using priority queueing mechanism. Prioritization based on the importance is most common feature in all modern internet applications to offer QoS. Priority mechanism is a concept of scheduling of different classes of arrivals to a single server. It has wide range of applications not only in engineering, but in inventory of manufacturing industries and health care systems (Zhao and Alfa, 1995; Brahimi and Worthington, 1991; Cohen *et al.*, 1988). There are different priority disciplines like preemptive, non-preemptive and discretionary priority. Each discipline has a scheduling procedure. In the literature, there are number of supplements based on priority scheduling, the outline of few fundamental priority queueing models in continuous-time was evident in the papers (Miller, 1960; Kleinrock, 1976; Takagi, 1991) and references therein. White and Christie (1958) studied M/M/1 queues with multiclass arrivals using preemptive priorities and analysis is made by generating functions of state probabilities. Later, Marks (1973) proposed an algorithm for computing probabilities of queue length. Sandhu and Posner (1989) analyzed voice/data communication using priority M/G/1 queue. Boxma *et al.* (1999) worked on heavy traffic using M/G/1 queue with priority classes and regularly varying heavy tailed service time distributions. Sharma and Virtamo (2002) consider finite buffer queue with priorities to model the system in the internet and obtain algorithms for workload, waiting time, and packet loss. Takine and Hasegawa (1994) derived LST of waiting time of customers based on MAP/G/1 queue with state dependent service time distributions. Takahashi and Miyazawa (1994) gave relation between queue length and waiting time distribution in a priority queue with batch arrivals. Takada and Miyazawa (2002) obtain moments of buffer contents for a Markov modulated fluid queue with preemptions. Jin and Min (2007) propose a novel analytical model for priority queueing system with heterogeneous LRD input traffic. Tarabia (2007) investigated the impact of catastrophes on single server preemptive priority queue using generating functions. Sampath *et al.* (2013)studied performance of wavelength division multiplexing optical packet switch employing wave length conversion techniques under self-similar input traffic. Zhao *et al.* (2015) analyzed sojourn time of two classes of customers using MAP/PH/1 queue with discretionary priority based on service stages. Ravi Kumar *et al.* (2017) evaluated performance of self-similar traffic input model in terms of high priority and low priority packet loss probabilities using MMPP/PH/c/K queuing system. Also, Malla Reddy and Ravi Kumar (2014, 2016, 2021)explored performance of network routers (synchronous and asynchronous) with self-similar input traffic using various multiserver queueing systems employing priority mechanism in the papers. From above cited papers, one can observe that priority discipline was used in various contexts to analyze systems, but in all the above cases performance analysis was made under steady state with homogeneous arrival and service processes, which are not realistic. As mentioned earlier, in present work, a network node with self-similar input traffic is modeled into transient MAP/PH/1/N queue with preemptive priority mechanism. Time dependent analysis of the system is made by level dependent quasi birth and death process, and arrival process follows MMPP with sinusoidal arrival rates (which is a special case of MAP). Performance measures, namely, queue length, mean waiting time, and packet loss of high priority and low priority packets are presented numerically.

The paper is organized as follows: Queueing model description is given in section 2. In section 3, performance analysis of system is presented. In section 4, computation complexity of algorithm is presented, and numerical results are illustrated in section 5. Finally, conclusions are given in section 6.

2. Queuing model

It is assumed that the packet arrivals are of high priority (Type I) and low priority (Type II) packets. Assume that Type I packet arrivals follows the MMPP characterized by $(Q^I, \Lambda^I(t))$, where $Q^I, \Lambda^I(t)$ are matrices of order n_I. Where as, Type II packet arrivals follow the MMPP characteriazed by $(Q^{II}, \Lambda^{II}(t))$, where $Q^{II}, \Lambda^{II}(t)$ are matrices of order n_{II}. Here n_I, n_{II} represent number of states of underlying Markov chains of Type I and Type II arrivals, respectively. As in Andersen and Nielsen (1998); Yoshihara *et al.* (2001); Shao *et al.* (2005); Abhilash and Malla Reddy (2022), modeling of self-similar traffic involves superposition of two-state MMPPs (In particular IPPs). The i^{th} IPP of Type I and Type II arrival process are given as follows:

$$\begin{aligned} Q_i^I &= \begin{bmatrix} -c_{1i} & c_{1i} \\ c_{2i} & -c_{2i} \end{bmatrix}, \Lambda_i^I(t) = \begin{bmatrix} \lambda_i^I(t) & 0 \\ 0 & 0 \end{bmatrix} \\ and, Q_i^{II} &= \begin{bmatrix} -d_{1i} & d_{1i} \\ d_{2i} & -d_{2i} \end{bmatrix}, \Lambda_i^{II}(t) = \begin{bmatrix} \lambda_i^{II}(t) & 0 \\ 0 & 0 \end{bmatrix}, 1 \le i \le r \end{aligned} \tag{1}$$

The superposition of r IPPs and a Poisson process of Type I and Type II arrival process are, respectively, given as

$$\begin{aligned} Q^I &= Q_1^I \bigoplus Q_2^I \bigoplus \cdots \bigoplus Q_r^I \\ \Lambda^I &= \Lambda_1^I(t) \bigoplus \Lambda_2^I(t) \bigoplus \cdots \bigoplus \Lambda_r^I(t) \bigoplus \lambda_p^I(t) \\ Q^{II} &= Q_1^{II} \bigoplus Q_2^{II} \bigoplus \cdots \bigoplus Q_r^{II} \\ \Lambda^{II} &= \Lambda_1^{II}(t) \bigoplus \Lambda_2^{II}(t) \bigoplus \cdots \bigoplus \Lambda_r^{II}(t) \bigoplus \lambda_p^{II}(t) \end{aligned} \tag{2}$$

Here, $\bigoplus, \bigotimes$ represent Kronecker's sum and product respectively, and $\lambda_p^I(t), \lambda_p^{II}(t)$ are time dependent Poisson arrival rates of Type I and Type II arrivals. The superposition of MMPPs $(Q^I, \Lambda^I(t))$, $(Q^{II}, \Lambda^{II}(t))$ is turned into a MAP with representation of $(D_0, D_1(t), D_2(t))$, where $D_0 = D_0^I \bigoplus D_0^{II}$ denote transitions of no arrival in both types, $D_1(t) = \Lambda^I(t) \bigotimes I_{n_{II}}$ and $D_2(t) = I_{n_I} \bigotimes \Lambda^{II}(t)$ denotes transitions corresponding to Type I and Type II arrivals, respectively, where $D_0^{II} = Q^{II} - \Lambda^{II}(t), D_0^I = Q^I - \Lambda^I(t)$. The mean arrival rate of Type I and Type II packets are given by (Abhilash and Malla Reddy, 2022)

$$\lambda_m^I(t) = \frac{1}{t}\left(\pi \int_0^t D_1(x)\, dx\, e\right), \lambda_m^{II}(t) = \frac{1}{t}\left(\pi \int_0^t D_2(x)\, dx\, e\right) \tag{3}$$

where e represents column vector of 1's with appropriate size, and π is unique vector satisfying $\pi(Q^I + Q^{II}) = 0, \pi e = 1$. The system is modeled into a single server queue with finite buffer capacity. Server provides priority scheduled service for Type I and Type II packets with preemptive discipline. That is, if Type I packet arrives, when Type II packet is in service, service process is interrupted, and after completion of the service, if there are no Type I packets, it starts processing of left over Type II packet as it is new one. Otherwise, it would go for another Type I packet. Assume that service process of the Type I and Type II packets follows continuous-time PH distributions denoted by (α, T) and (β, S) respectively, with same dimension p , where, α, β are vectors of size $1 \times p$, T, S are $p \times p$ matrices, and $t^0 = -Te, s^0 = -Se$. The mean service time of Type I and Type II packets are obtained by $\mu^I = -\alpha T^{-1} e$, $\mu^{II} = -\beta S^{-1} e$, respectively. Buffer capacity of system is taken to be N . Finally, the resultant queueing system of network node is MAP/PH/1/N queue with preemptive priority. Let$N_{II}(t)(N_I(t))$ be the number of Type II (Type I) packets in the system at time t, including packet in service. The thresholds for Type I and Type II packets are K_1, K_2 respectively, where N equals to $K_1 + K_2$. The arrival phase of superposed MAP at time t is denoted by $A(t)$, and service phase of system is denoted by $B(t)$. Therefore, arrival process of system is characterized by a multi-dimensional continuous time Markov chain $F(t) = \{N^I(t), N^{II}(t), A(t), B(t), t \geq 0\}$, The state space of is given by:

$$
\begin{aligned}
F_1 &= \{(0, 0, a, 0), a = 1, \dots, n\} \\
F_2 &= \{(m_I, 0, a, b), m_{II} > 0; a = 1, \dots, n; b = 1, \dots, p\} \\
F_3 &= \{(0, m_{II}, a, b), m_{II} > 0; a = 1, \dots, n; b = 1, \dots, p\} \\
F_4 &= \{(m_I, m_{II}, a, b), m_I > 0, m_{II} > 0; a = 1, \dots, n; b = 1, \dots, p\}
\end{aligned}
$$

Here, F_1 represents idle state of server with arrival at phase at a. F_2 represent $m_I(> 0)$ Type I packets and no Type II packets in queue. F_3 represent $m_{II}(> 0)$ Type II packets and no Type I packet in queue. F_4 represent there are $m_I(> 0)$, $m_{II}(> 0)$ of Type I and Type II packets are in queue. In three cases, arrival is in phase a, and service is in phase b. If stages of $F(t)$ are arranged in lexicographical order. The level dependent block tridiagonal generator matrix of system occupancy at time t is given by

$$
Q(t) = \begin{bmatrix}
\overline{A_0(t)} & \overline{A_1(t)} & 0 & 0 & \dots & 0 & 0 & 0 \\
\overline{B_2(t)} & A_0(t) & A_1(t) & 0 & \dots & 0 & 0 & 0 \\
0 & A_2(t) & A_0(t) & A_1(t) & \dots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \dots & A_2(t) & A_0(t) & A_1(t) \\
0 & 0 & 0 & 0 & \dots & 0 & A_2(t) & A_0(t) + A_1(t)
\end{bmatrix}
$$

where all block matrices in $Q(t)$ are square matrices of finite order, and are defined as follows,

$$
\overline{A_0(t)} = \begin{bmatrix}
D_0 & D_2(t) \otimes \beta & 0 & 0 & \dots & 0 & 0 & 0 \\
I \otimes s^0 & D_0 \bigoplus S & D_2(t) \otimes I & \dots & 0 & 0 & 0 & \\
0 & I \otimes s^0\beta & D_0 \bigoplus S & D_2(t) \otimes I & \dots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \dots & I \otimes s^0\beta & D_0 \bigoplus S & D_2(t) \otimes I \\
0 & 0 & 0 & 0 & \dots & 0 & I \otimes s^0\beta & M
\end{bmatrix}
$$

$$
A_0(t) = \begin{bmatrix}
D_0 \bigoplus T & D_1(t) \otimes I & 0 & 0 & 0 & 0 \\
0 & D_0 \bigoplus T & D_1(t) \otimes I & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & D_0 \bigoplus T & D_1(t) \otimes I \\
0 & 0 & 0 & 0 & 0 & D_0 \bigoplus T + D_1(t) \otimes I
\end{bmatrix}_{(K_2+1)\times(K_2+1)}
$$

$$\overline{A_1(t)} = \begin{bmatrix} D_1(t) \otimes \alpha & 0 & 0 & 0 \\ 0 & D_1(t) \otimes I & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & D_1(t) \otimes I \end{bmatrix}_{(K_2+1)\times(K_2+1)}$$

$$\overline{B_2(t)} = \begin{bmatrix} I \otimes t^0 & 0 & 0 & 0 \\ 0 & I \otimes t^0\alpha & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & I \otimes t^0\alpha \end{bmatrix}_{(K_2+1)\times(K_2+1)}$$

$$A_1(t) = diag[D_1(t) \bigotimes I, D_1(t) \bigotimes I, \ldots, D_1(t) \bigotimes I]_{K_2+1},$$

$$A_2(t) = diag[I \bigotimes t^0\alpha, I \bigotimes t^0\alpha, \ldots, I \bigotimes t^0\alpha]_{K_2+1},$$

where $M = D_0 \bigoplus S + D_2(t) \otimes I$, I is an identity matrix of an appropriate order and $\overline{A_0(t)}$ is of order $(K_2 + 1) \times (K_2 + 1)$.

3. Performance analysis

Let $\pi(t) = (\pi_0(t), \pi_1(t), \ldots, \pi_{K_1}(t))$ be transient state probability vector of $Q(t)$. That is, $\pi(t)$ satisfies (Stewart, 1994)

$$\frac{d}{dt}\pi(t) = Q(t)\pi(t) \tag{4}$$

$$\Longrightarrow \pi(t) = \pi(0)\ exp\left(\int_0^t Q(x)\,dx\right) \tag{5}$$

By using Theorem. 2.4.3 in (Slavík, 2007), one can get

$$\pi(t) = \pi(0)\prod_{k=0}^{n}(I + Q(t_k)h) \tag{6}$$

where $h = t_k - t_{k-1}$, n is number of partitions of the interval $(0, t]$, and $\pi(0)$ is state probability vector at time $t = 0$. Here, each $\pi_j(t)$ is vector corresponding to the set of states with j Type I packets, and is in the form of $\pi_j(t) = (\pi_{j0}(t), \pi_{j1}(t), \ldots, \pi_{jK_2}(t)$. Each $\pi_{jk}(t)$ represents the probability that there exist j Type I packets, and k Type II packets are in the system. The performance measures are given as follows:
The Mean waiting time of Type I packets (Zhao *et al.*, 2015)

$$MWT_{TypeI} = \frac{E[N^I(t)]}{\lambda_m^I(t)} = \frac{1}{\lambda_m^I(t)}\sum_{j=1}^{K_1} j\pi_j(t)e \tag{7}$$

Let assume $Y = \sum_{i=1}^{K_1}\pi_j(t)$, and $Y = \{Y_{:0}(t), Y_{:1}(t), \ldots, Y_{:K_2}(t)\}$, where each $Y_{:m_{II}}(t)$ is a row vector corresponding to m_{II} Type II customers in the system. The Mean waiting time of Type II packets is

$$MWT_{TypeII} = \frac{E[N^{II}(t)]}{\lambda_m^{II}(t)} = \frac{1}{\lambda_m^{II}(t)}\left(\sum_{k=1}^{K_2} k\left(Y_{:k}(t) + \pi_{0k}(t)\right)e\right) \tag{8}$$

Since, buffer capacity is finite, If Type I (Type II) packet arrives, and finds that there are $K_1\,(K_2)$ packets in system, then the packet is lost. The loss probability of Type I and Type II packets (Zhao *et al.*, 2015) in small time duration Δ are, respectively

$$P^{I}_{loss} = \frac{1}{\lambda^{I}_{m}(t+\Delta)}\left[\pi_{K_1}(t)\left(\int_t^{t+\Delta}\left(D_1(x)\bigotimes I\right)\,dx\right)e\right] \tag{9}$$

$$P^{II}_{loss} = \frac{1}{\lambda^{II}_{m}(t+\Delta)}\left[(Y_{:K_2}(t)+\pi_{0K_2}(t))\left(\int_t^{t+\Delta}\left(D_2(x)\bigotimes I\right)\,dx\right)e\right] \tag{10}$$

4. Computational complexity

In this section, one can present computational complexity of performance measures (Malla Reddy and Ravi Kumar, 2016; Chen *et al.*, 2007; Wang *et al.*, 2000), namely, mean waiting time and packet loss probability of Type I and Type II packets, which are given in Eqs.(7-10). The complexity of $MWT_{TypeI}, MWT_{TypeII}$ is of the order $\mathcal{O}$ $((K_2+1)nm)$, $\mathcal{O}$ $((K_1+1)nm)$ respectively, due to it involves product of several row and column vectors. The complexity of $P^{I}_{loss}, P^{II}_{loss}$ is of the order $\mathcal{O}$ $((K_2+1)^2n^2m^2)$, $\mathcal{O}$ $((K_1+1)^2n^2m^2)$ respectively. But, the Eqs. (7-10) involves transient state probability vector of generator matrix $Q(t)$ (with dimensions$((K_1+1)(K_2+1)nm)$, which is obtained by using method of Product integrals, and it is given in Eq. 6. Since, the problem of finding state probability vector involves addition and product of matrix $Q(t)$ several times. The computational complexity of finding product $\prod_{i=0}^{n}(I+Q(t_i)h)$ is of the order $\mathcal{O}\left(((K_1+1)(K_2+1)nm)^{2.37}\right)$ (using Coppersmith-Winograd algorithm), and complexity of addition is of the order $\mathcal{O}((K_1+1)^2(K_2+1)^2n^2m^2)$. Therefore, the overall computation complexity of the algorithm according to Big-O analysis is of order $\mathcal{O}\left(((K_1+1)(K_2+1)nm)^{2.37}\right)$.

5. Numerical results

In this section, performance measures of the system are presented numerically. For arrival process of Type I and Type II packets, the numerical values given in Table 1 and 2 are used. The number superposed MMPPs are taken to be 2, and sinusoidal arrival rates are taken in the form of $a + b_j sint$, where a is whole arrival rate and b_j varies in between $(0, 1)$. For Type I packet arrivals transition rates are given in Table 1 and arrival rates are same for three samples (based on traffic parameters) of Type I packets, these are $\lambda^{I}_1(t) = 1 + 0.3 \times \sin t$, $\lambda^{I}_2(t) = 1 + 0.7 \times \sin t$. For Type II packet arrivals transition rates are given in Table 2, and arrival rates are same for three samples (based on traffic parameters), these are $\lambda^{II}_1(t) = 1 + 0.4 \times \sin t$, $\lambda^{II}_2(t) = 1 + 0.8 \times \sin t$. Assume that service distribution follows two phase distribution, i.e, Erlang distribution(E_2) with varying service rates. Figs. 1-6 show that waiting time and packet loss for Type I and Type II packets increases as traffic intensity increases at every time instant, and also represent that waiting time and packet loss increase as Hurst parameter (H) increases. From Figures 7 and 8, one can observe that mean waiting time increases, and packet loss decreases as threshold of Type I increases at every particular instant of time for $H = 0.9$.

Table 1: Values of traffic parameters and fitting parameters of Type I arrival rates

Sample	Parameters of Self-similar Input Traffic	r=2	
		c_{11}	c_{21}
Sample 1	$H = 0.7, \lambda_w(t) = 1, and\, \sigma^2 = 0.6$	0.196	0.001
Sample 2	$H = 0.8, \lambda_w(t) = 1, and\, \sigma^2 = 0.6$	0.0102	0.000188
Sample 3	$H = 0.9, \lambda_w(t) = 1, and\, \sigma^2 = 0.6$	0.005198	0.0005

Table 2: Values of traffic parameters and fitting parameters of Type II arrival rates

Sample	Parameters of Self-similar Input Traffic	r=2	
		d_{11}	d_{21}
Sample 1	$H = 0.7, \lambda_w(t) = 1, and\, \sigma^2 = 0.6$	0.23	0.0013
Sample 2	$H = 0.8, \lambda_w(t) = 1, and\, \sigma^2 = 0.6$	0.05	0.00092
Sample 3	$H = 0.9, \lambda_w(t) = 1, and\, \sigma^2 = 0.6$	0.003	0.000272

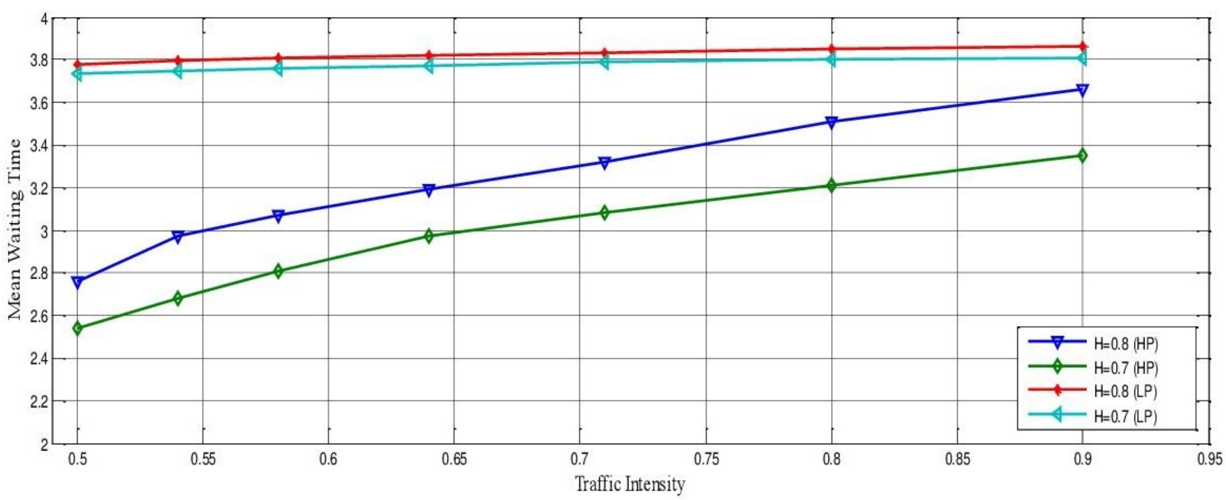

Figure 1: Traffic intensity vs mean waiting time with $N = 10, t = 1$

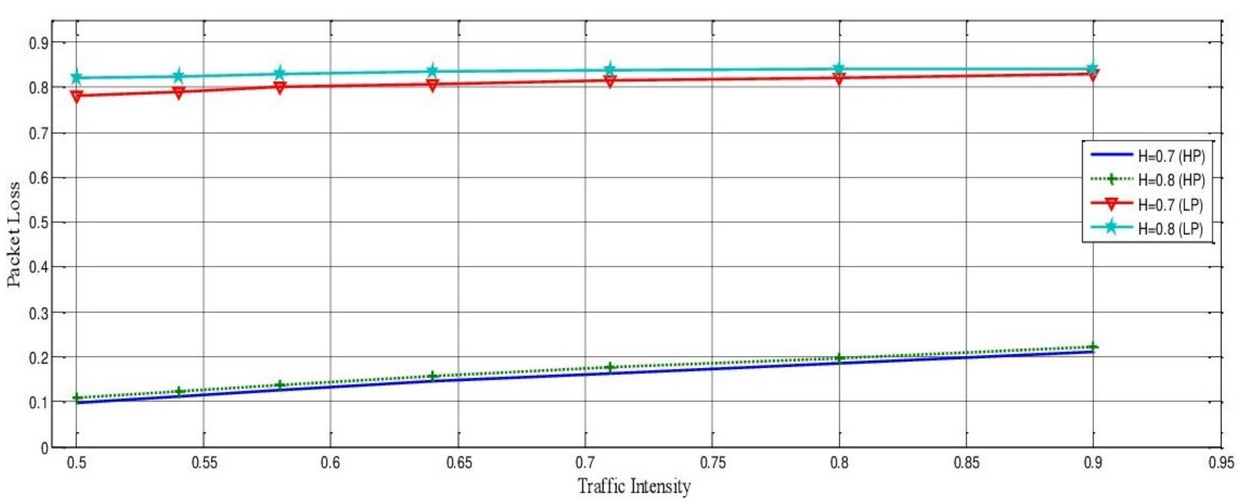

Figure 4: Traffic intensity vs packet loss with $N = 10, t = 3, \Delta = 0.5$

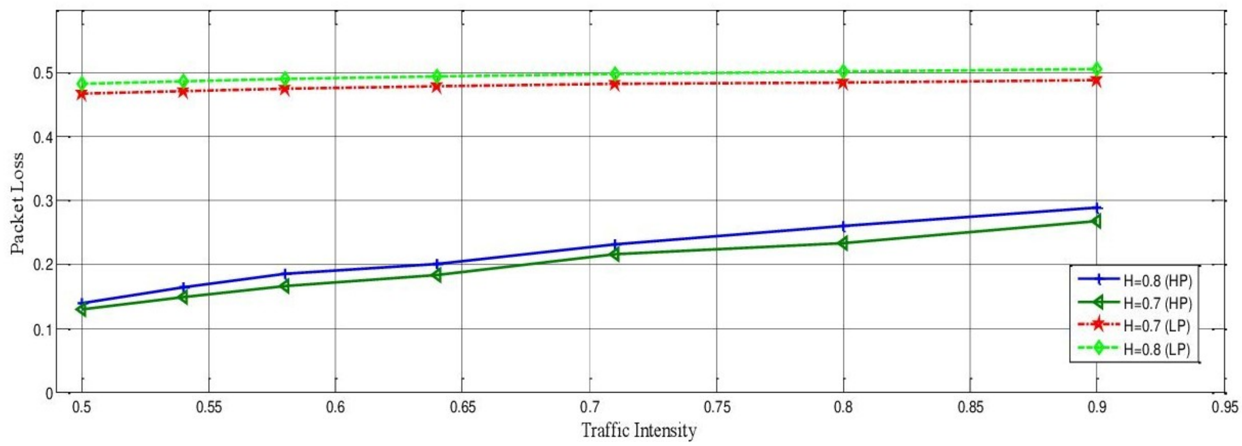

Figure 2: Traffic intensity vs packet loss with $N = 10, t = 1, \Delta = 0.5$

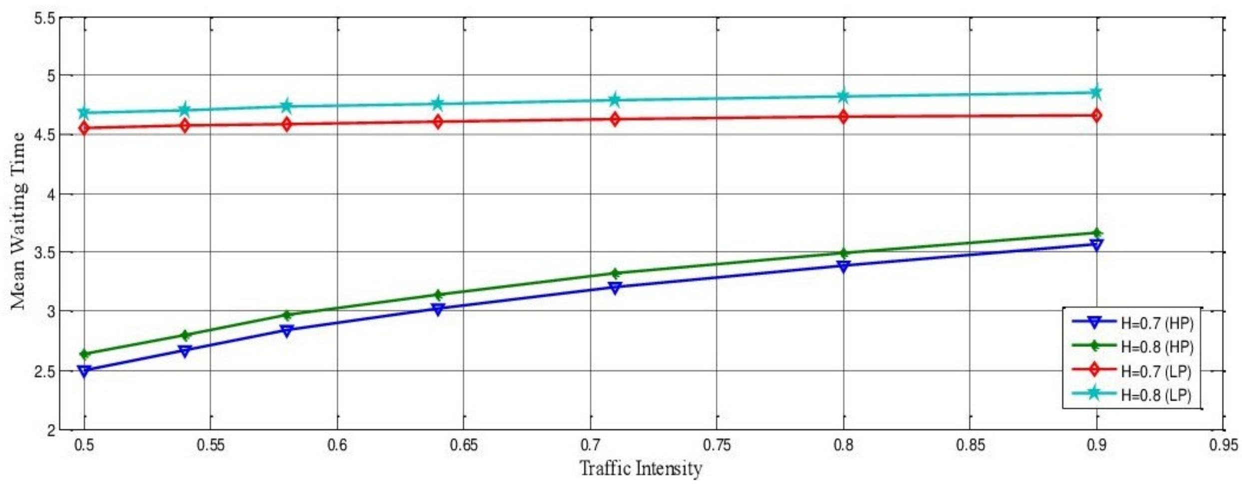

Figure 3: Traffic intensity vs mean waiting time with $N = 10, t = 3$

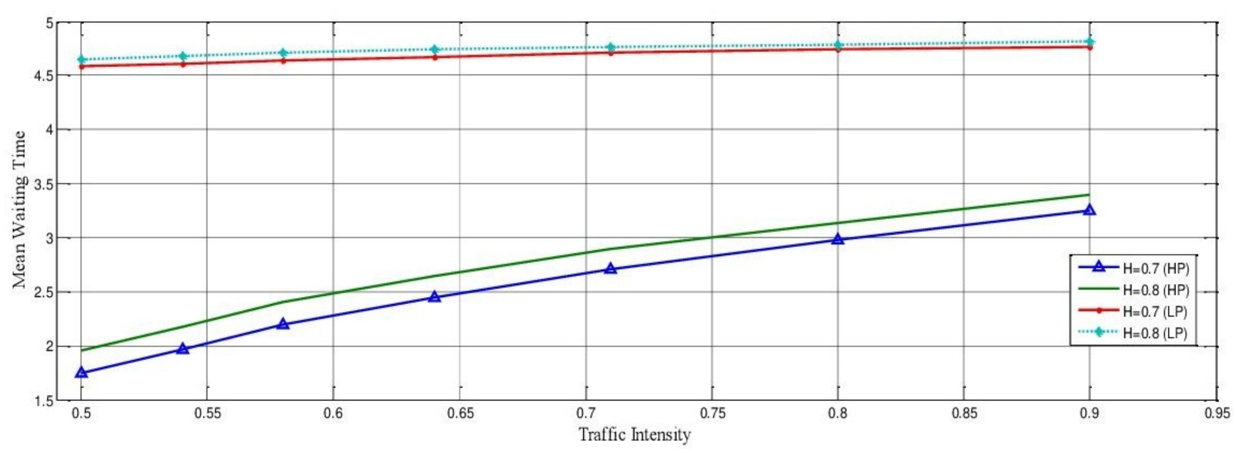

Figure 5: Traffic intensity vs mean waiting time with $N = 10, t = 5$

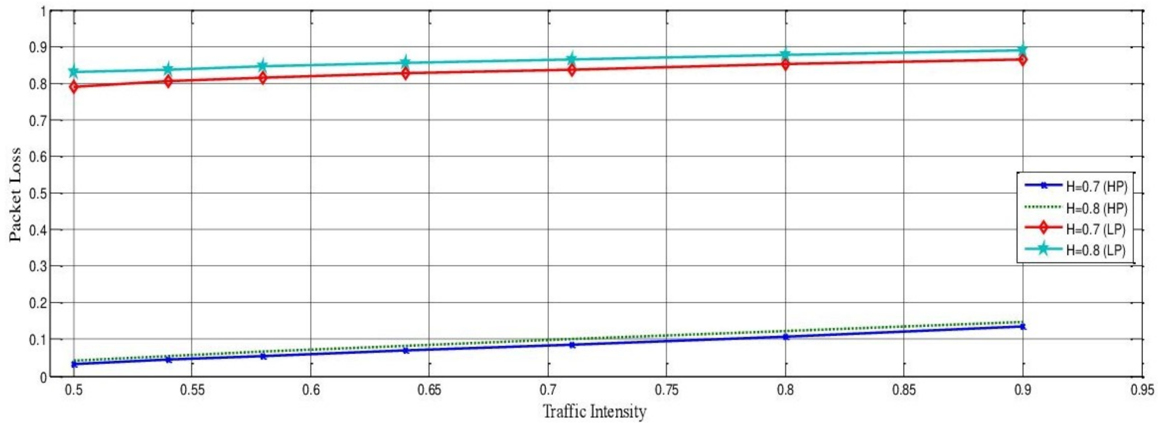

Figure 6: Traffic intensity vs packet loss of Type II packets with $N = 10, t = 5, \Delta = 0.5$

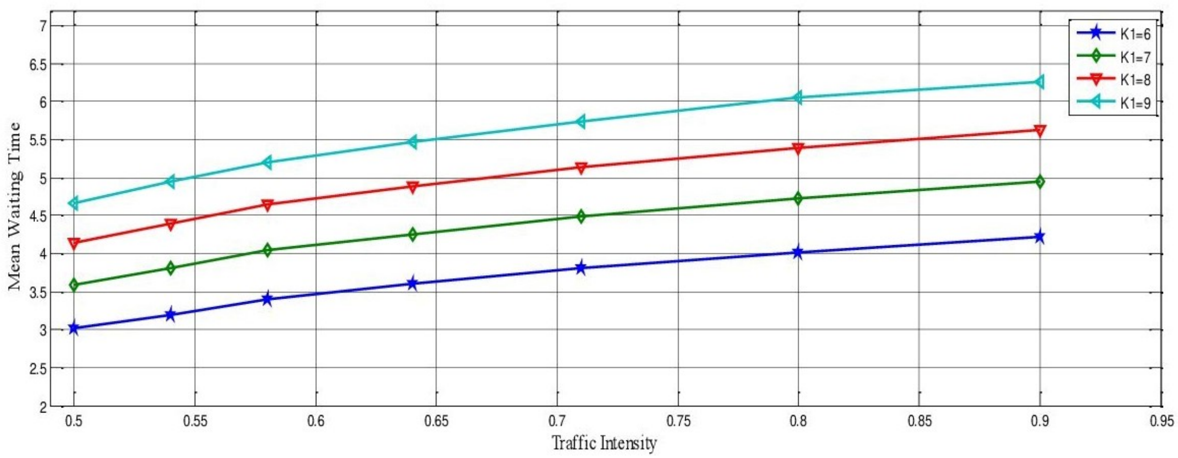

Figure 7: Traffic intensity vs mean waiting time of Type I packets with $t = 3, H = 0.9$

6. Conclusions

In this paper, network nodes with self-similar input priority based traffic are modeled into transient MAP/PH/1 queueing system, and its performance analysis is made by using level dependent quasi-birth and process with preemptive priority mechanism. The system is approximated by a finite buffer and transient state probability vector is obtained by the method of product integrals. Numerical results show that how traffic intensity, Hurst parameter, and threshold effects mean waiting time and packet loss of HP and LP packet arrivals at different time instants.

Acknowledgements

The authors wish to acknowledge Council of Scientific and Industrial Research (CSIR), Government of India, for their funding under the Major Research Project (MRP) scheme (File. No: 25(0301)/19/EMR-II).

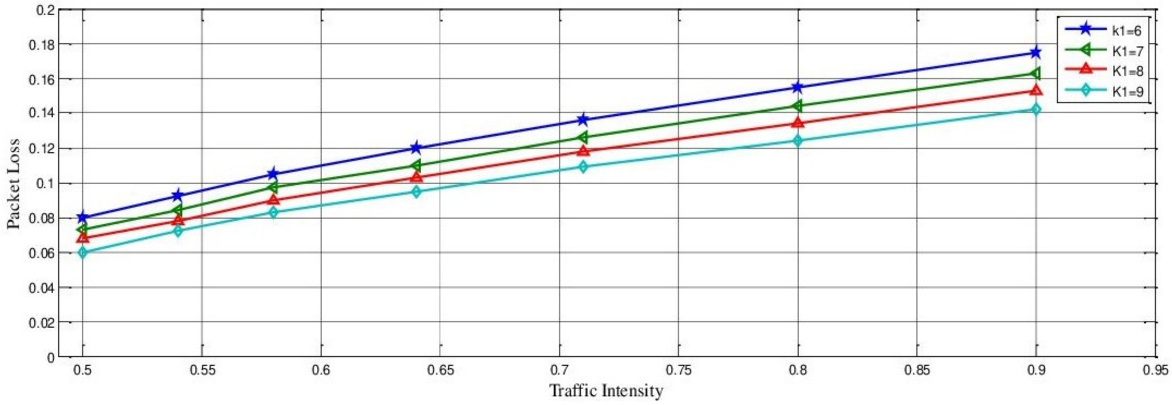

Figure 8: Traffic intensity vs packet loss of Type I packets with $t = 3, H = 0.9, \Delta = 0.5$

References

Abhilash, V. and Malla Reddy, P. (2022). Fitting model for self-similar traffic- time dependent markovian process and second order statistics. *Statistics and Applications*, **20**, 297–309.

Andersen, A. T. and Nielsen, B. F. (1998). A markovian approach for modelling packet traffic with long-range dependence. *IEEE journal of Selected Areas in Communications*, **16**, 719–732.

Boxma, O., Cohen, J., and Deng, Q. (1999). Heavy-traffic analysis of the m/g/1 queue with priority classes. *In: Proceedings of ITC*, **16**, 1157–1167.

Brahimi, M. and Worthington, D. (1991). Queueing models for out-patient appointment systems - a case study. *The Journal of the Operational Research Society*, **42**, 733–746.

Chen, C., Chang, C., Malla Reddy, P., Shao, S., and Wu, J. (2007). Performance analysis of wdm optical packet switches employing wavelength conversion under markovian modeled self-similar traffic input. *Workshop on High Performance Switching and Routing*, **1**, 1–6.

Cohen, M., Kleindorfer, and Lee, H. (1988). Service constrained (s,s) inventory systems with priority demand classes and lost sales. *Management Science*, **34**, 482–499.

Crovella, M. and Bestavros, A. (1997). *Self-similarity in World Wide Web traffic: evidence and possible causes*. IEEE.

Jin, X. and Min, G. (2007). Performance analysis of priority scheduling mechanisms under heterogeneous network traffic. *Journal of Computer and System Sciences*, **73**, 1207–1220.

Kleinrock, L. (1976). *Queueing Systems Volume II: Computer Applications*. John Wiley and Sons, New York.

Leland, W., Taqqu, M., Willinger, W., and Wilson, D. (1994). On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, **2**, 1–15.

Malla Reddy, P. and Ravi Kumar, G. (2014). Inverstigating priority based performance analysis of optical packet switch under asynchronous self-similar variable length packet traffic input with voids. *Springer Proceedings in Mathematics and Statistics, Switzerland*, **92**, 413–425.

Malla Reddy, P. and Ravi Kumar, G. (2016). Performance analysis of wavelength division multiplexing asynchronous internet router employing space priority mechanism under self-similar traffic input-multi-server queueing system with markovian input and erlang-k services. *Applied Mathematics*, **07**, 1707–1725.

Malla Reddy, P. and Ravi Kumar, G. (2021). Performance analysis of asynchronous priority-based internet router under self-similar traffic input - queueing system with markovian input and hyper-exponential services. *International Journal of Operational Research*, **40**, 239–260.

Marks, B. (1973). State probabilities of M/M/1 priority queues. *Operations Research*, **21**, 974–987.

Miller, R. (1960). Priority queues. *Annals of Mathematical Statistics*, **31**, 86–103.

Paxson, V. and Floyd, S. (1995). Wide area traffic: The failure of poisson modelling. *IEEE/ACM Transactions on Networking*, **3**, 226–244.

Ravi Kumar, G., Raj Kumar, L., and Malla Reddy, P. (2017). Loss behaviour analysis of asynchronous internet switch under self-similar traffic input using MMPP/PH/c/K queueing system employing pbs mechanism. *International Journal of Communication Networks and Distributed Systems*, **19**, 257–269.

Sampath, K., Malla Reddy, P., and Adilakshmi, T. (2013). Performance study of WDM OPS with space priority mechanism under self-similar variable length input traffic. *Proceedings of IEEE ICON-2013 held at Singapore*, **1**, 1–6.

Sandhu, D. and Posner, M. (1989). A priority M/G/1 queue with application to voice/data communication. *European Journal of Operational Research*, **40**, 99–108.

Shao, S. K., Malla Reddy, P., Tsai, M. G., Tsao, H., and Wu, J. (2005). Generalized variance-based markovian fitting for self-similar traffic modelling. *IEEE transactions on communications*, **88**, 1493–1502.

Sharma, V. and Virtamo, J. (2002). A finite buffer queue with priorities. *Performance Evaluation*, **47**, 1–22.

Slavík, A. (2007). *Product Integration, its History and Applications*. Matfyz press.

Stewart, W. J. (1994). *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, Princeton, New Jersey.

Takada, H. and Miyazawa, M. (2002). A markov modulated fluid queue with batch arrivals and preemptions. *Stochastic Models*, **18**, 529–652.

Takagi, H. (1991). *Queueing Analysis: A Foundation of Performance Evaluation, Volume 1: Vacation and Priority Systems, Part 1*. North-Holland.

Takahashi, Y. and Miyazawa, M. (1994). Relationship between queue-length and waiting time distributions in a priority queue with batch arrivals. *Journal of the Operations Research Society of Japan*, **37**, 48–63.

Takine, T. and Hasegawa, T. (1994). The workload in the MAP/G/1 queue with state dependent services: its application to a queue with pre-emptive resume priority. *Communications in Statistics - Stochastic Models*, **10**, 183–204.

Tarabia, A. (2007). Two-class priority queueing system with restricted number of priority customers. *AEU-International Journal of Electronics and Communications*, **61**, 534–539.

Wang, Y. C., Liu, C. W., and Lu, C. C. (2000). Loss behavior in space priority queue with batch markovian arrival process-discrete-time case. *Performance Evolution*, **41**, 269–293.

White, H. and Christie, L. (1958). Queueing with pre-emptive priorities with breakdown. *Operations Research*, **6**, 79–95.

Yoshihara, T., Shoji, K., and Takahashi, Y. (2001). Practical time-scale fitting of self-similar traffic with markov modulated poisson process. *Telecommunication systems*, **17**, 185–211.

Zhao, N., Zhaotong, L., and Kan, W. (2015). Analysis of a MAP/PH/1 queue with discretionary priority based on service stage. *Asia-Pacific Journal of Operational Research*, **32**, 1–22.

Zhao, Y. and Alfa (1995). Performance analysis of a telephone system with both patient and impatient customers. *Telecommunication Systems*, **4**, 201–215.