



Analysis of Power-Law Queues Using Maximum Entropy Framework: An Application to Performance Evaluation of Network Systems

Shachi Sharma¹

Accepted: 12 April 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Extensive research on traffic of modern communication networks, including Internet and wireless, establish the presence of power-law behavior. However, developing traffic models capturing power-law characteristics is a difficult task due to analytical intractability. Adopting maximum entropy approach, a theoretical model $Pow/Pow/1/\infty$ has been proposed recently. The paper extends this framework to build two new theoretical models $Pow/Pow/1/K$ and $Pow/Pow/K/K$ where both inter-arrival and service times follow power-law distribution. Closed form expressions of queue length distribution and various performance measures are derived in terms of traffic intensity ρ . An explicit expression for state probability distribution of $Pow/Pow/K/K$ is also derived and compared with well-known Erlang loss formula. Both queue length distribution and blocking probability are found to depict power-law implying that increasing buffers have little impact on improving the performance of systems in these cases. Numerical computations reveal that the mean queue length explodes as ρ tends to 0.5 resulting in longer delay. The variance of number of packets also depicts sharp rise as ρ approaches to 0.3. These results are in close agreement with the empirical studies carried out on real traffic traces in the past. The proposed models are useful in designing networks, dimensioning resources and developing congestion control algorithms.

Keywords Entropy · Erlang loss formula · Geometric mean constraint · Inter-arrival time · Maximum entropy principle · Power-law behavior · Quality of service · Network systems

1 Introduction

Ubiquity of power-law has been well established across variety of systems [1]. If X is a random variable following power-law distribution then there exists a positive parameter k such that the Complementary Cumulative Distribution Function (CCDF) behaves as

✉ Shachi Sharma
shachi@sau.int

¹ Computer Science, South Asian University, Maidan Garhi, New Delhi, Delhi 110068, India

$$P\{X > x\} \sim x^{-k}, \quad x > x_{\min}. \quad (1)$$

It can be noted from (1) that the graph of CCDF with x on log–log scale is a straight line with slope k . The presence of power-law also leads to long-tail behavior in a system [2]. One of the widely observed power-law distribution in computer communication networks is Pareto. Many empirical studies on various types of network traffic also validate the presence of power-law characteristics [3]. This is in contrast to traditional voice networks which exhibit exponentially distributed inter-arrival time. Thus, it becomes imperative to examine the impact of power-law distributed traffic on the performance of network systems. Such a study is essential for determining appropriate resource (for example buffers) dimensioning in network systems like routers and gateways. The requirement of proper number of buffers in routers has been emphasized by Ghosh et al. [4] by observing that improper and large buffers may result in longer queueing delays based on the analysis utilizing M/M/1 and M/D/1 queueing models.

1.1 Motivation

Several studies on various types of applications and networks traffic also suggest inapplicability of models based on exponential distribution. For example, statistical analysis of mobile data traffic reveals that inter reference time of file (time difference between successive requests of a file) as well as the inter-session time (time difference between two sessions of a client) is well described by lognormal distribution [5]. Investigation on the cyberlocker service traffic on cloud platform confirms that inter arrival time of cyberlocker traffic flows can be best modelled by hybrid lognormal and gamma distribution [6]. It is worth to mention the findings by Mitzenmacher [7] here, “*lognormal and power law distributions connect quite naturally, and hence, it is not surprising that lognormal distributions have arisen as a possible alternative to power law distributions across many fields*”. Power-law is ubiquitous. Access pattern of file downloads in global file hosting platforms at cloud [8] and DNS queries [9] also follow power-law. A similar behavior is observed for video traffic stream [10]. Empirical studies of web and peer-to-peer network traffic also establish that inter-arrival time follows two mode Weibull distribution and Weibull-Pareto distribution respectively [11]. Kokoszka et al. [12] analyze the inter-arrival time between consecutive anomalies in the Internet and tail probabilities are found to decay following power-law. A model to evaluate the performance of Wi-fi network using power-law distribution for ON periods has been built in [13]. The emergence of power-law is also reported in satellite network [14]. The most evolving trend in networking domain is Internet of Things (IoT). After collecting and analyzing sensors data from IoT-based smart home applications, generalized Pareto distribution is found to describe the inter-arrival time of packets in [15]. The requirement for developing simple models based on Pareto and Weibull distribution to capture traffic characteristics of access as well as backbone networks is emphasized in [16]. The presence of power-law in mobile network of an operator in France has been reported in [17]. A good summary of the appearance of power-law in diverse types of network is compiled in [18] recently.

The power-law behavior in network traffic has severe impact on the performance of network systems, for example, high queueing delays. The performance analysis of network systems is carried out using models from queueing theory like $M/G/1$, $M/G/\infty$ with service time G following Pareto or other long-tail distribution [19]. These models do not represent an actual real network system. Also, these models become analytically intractable

and only provide approximate asymptotic results, for example, upper bound on blocking probability is estimated from buffer overflow probability [20–22]. The blocking probability relates to the probability with which buffers inside a network system are full. The blocking probability is computed using finite buffer queueing models. Recently, bounds on queue and packet loss distribution are obtained for finite buffer system under the assumption of Poisson arrivals [23]. However, the Poisson assumption based queueing models are useful only for voice networks and classical data networks as they fail to capture the features of traffic exhibited in modern days' integrated communication networks. Thus, there is a requirement to develop appropriated models capturing the traffic characteristics of the modern networks.

1.2 Role of Maximum Entropy Framework in Networks' Performance Evaluation

An alternate mechanism for deriving the results of queueing theory is based on maximum entropy framework exploiting the Jaynes maximum entropy principle (MEP) which states that the most appropriate probability distribution given some prior information about the system is the one which maximizes entropy of the system [24]. Prior information is usually available in the form of moments. MEP using Shannon entropy has been successfully employed to recover the results of $M/M/1$ queueing systems [24] when arithmetic mean is the characterizing moment. While applying MEP to queueing systems, there is no assumptions made about the probability distribution of arrivals and services, rather the queueing system is treated as a black box and only the indices like mean number of customers in the system, mean utilization etc. are considered [25]. Kouvatso [26] has derived the expression for steady state probability distribution of $G/G/1$ queueing system using MEP when information about first two moments of arrival and service distributions are available. It is important to note that the power exponent k in (1) usually lies between 2 and 3 and the second order moment (i.e. variance) becomes infinite in this range. Hence, one cannot use the results of $G/G/1$ model to analyze the queueing system driven by power law distributed arrivals and service. Thus, other entropy measures has been explored to model these systems. Shachi and Karmeshu [27, 28] have used MEP with Tsallis entropy measure to obtain the results for broadband networks depicting power-law behavior. Amit and Karmeshu [29, 30] have applied MEP with Shannon entropy when geometric mean or shifted geometric mean of system size is specified as constraint to generate power-law behavior. The importance of geometric mean in relation to power-law has been discussed in the literature. As pointed out in [31], for majority of the power law distributions like Pareto, first moment or arithmetic mean is not defined and application of geometric mean thus becomes logical. Visser [32] concludes that the logarithmic average i.e. geometric mean plays an important role seems to have a connection with the fact that logarithmic scales are ubiquitous in classifying many natural and social phenomena. Following this, Amit and Karmeshu [29] note that geometric mean is an unbiased measure for lognormal distribution that approximates tail of many distributions showing power-law behavior.

1.3 Contributions

It is worth highlighting that the packets size in networks does not remain fixed. In the context of wireless Long Term Evolution (LTE) networks, packets lengths are found to follow Pareto distribution [33]. The same pattern is expected to present in 5G wireless networks [34]. Thus, it becomes essential to develop models considering power-law distributed

service time as well. To the best of our knowledge, there is no literature available that considers power-law distributed arrival and service processes and obtain closed-form expressions for performance measures. These models facilitate in performance evaluation of systems such as routers, gateways, servers on cloud etc. in modern computer communication networks. Motivated by these findings, an infinite capacity $Pow/Pow/1/\infty$ where both inter-arrival time of packets as well as service time follow power-law distribution has been proposed by Sharma [35] using maximum entropy approach with geometric mean and utilization as constraints. This paper extends the framework presented in [35] to finite buffer $Pow/Pow/1/K$ and $Pow/Pow/K/K$ models. The closed-form expressions for queue length distribution and various performance measures are obtained in terms of traffic intensity. The results of performance analysis are found to be in close agreement with past empirical studies. This paper is an extension of [35] and makes following contributions:

1. A single server finite buffer model $Pow/Pow/1/K$ is proposed and closed-form formula for blocking probability is derived in terms of traffic intensity.
2. The above model is extended to include multiple servers viz. $Pow/Pow/K/K$, the loss system and modified Erlang loss formula is obtained.
3. The applications of both newly proposed models in computing buffer requirements is illustrated.

The paper is organized into five sections. The $Pow/Pow/1/\infty$ model is reproduced from [35] in Sect. 2 and closed form expressions for steady-state queue length distribution and various performance measures in terms of traffic intensity are presented. The finite buffer queueing system $Pow/Pow/1/K$ is proposed in Sect. 3. The multi-server generalization of the finite buffer model is presented in Sect. 4. The last Sect. 5 discusses the findings and application of the proposed models in congestion control algorithm's design as well as concludes the paper.

2 $Pow/Pow/1/\infty$ Model

The model as proposed in [35] is presented in this section for completeness. Consider a network system such as a router or where packets arrive following a stochastic process such that inter-arrival time of the packets is explained by power-law distribution. Since, the packets may vary in size and some studies have confirmed it to be characterized by power-law distribution [33, 34], the processing time of the packets at the system can also be modeled by power-law distribution. Let random variable N represent the number of packets inside the system. The probability that there are n packets at a network system at any given instant of time is $P(N = n) = p_n$. Given this, the Shannon entropy for the system is given by

$$S = - \sum_{n=0}^{\infty} p_n \ln p_n \quad (2)$$

Assuming that prior information about the number of packets is available in the form of geometric mean G of non-zero queue size as

$$\sum_{n=1}^{\infty} (\ln n)p_n = \ln G \tag{3}$$

and system utilization

$$U = 1 - p_0 = \rho \tag{4}$$

where ρ is the traffic intensity. The normalization constraint is

$$\sum_{n=0}^{\infty} p_n = 1. \tag{5}$$

Employing MEP, the most objective probability distribution of number of packets is obtained by maximizing Shannon entropy (2) subject to (3), (4) and (5) as constraints. The optimization problem can be solved by Lagrange’s multiplier method. The Lagrangian function in this case becomes

$$L = - \sum_{n=0}^{\infty} p_n \ln p_n + \kappa \left[\ln G - \sum_{n=1}^{\infty} (\ln n)p_n \right] + \mu [1 - U - p_0] + \gamma \left[1 - \sum_{n=0}^{\infty} p_n \right]. \tag{6}$$

Differentiating (6) with respect to p_n and equating to zero gives

$$-(1 + \ln p_n) - \kappa \ln n - \gamma = 0, \quad n \neq 0 \tag{7}$$

and

$$-(\ln p_0 + 1) - \mu - \gamma = 0. \tag{8}$$

Solving (7) and (8) using constraints (4) and (5) results in

$$p_n = Z^{-1} U n^{-\kappa} \tag{9}$$

where $Z = \sum_{n=1}^{\infty} n^{-\kappa}$ is normalization constant. In order to compute Z , we consider $\left\{ \frac{1}{n^\kappa}, \kappa > 1 \right\}$, which is a strictly decreasing sequence of positive numbers. Thus, taking $f(x) = \frac{1}{x^\kappa}$, the normalization constant can be approximated by integral as $Z = \sum_{n=1}^{\infty} n^{-\kappa} \approx \int_1^{\infty} x^{-\kappa} dx$ transforming (9) into

$$p_n = U \kappa \left(1 - \frac{1}{\kappa} \right) n^{-\kappa}, \quad \kappa > 1, \quad n \geq 1. \tag{10}$$

The Lagrange’s parameter κ can be computed by substituting (10) in (3) viz.,

$$U(\kappa - 1) \sum_{n=1}^{\infty} (\ln n)n^{-\kappa} = \ln G. \tag{11}$$

Replacing summation by integral in (11), the expression for κ becomes

$$\kappa = 1 + \frac{U}{\ln G}. \tag{12}$$

Postulating $\frac{1}{\kappa} = \rho$, (10) becomes

$$p_n = (1 - \rho)n^{-1/\rho}, \quad n \geq 1, \quad \rho < 1. \quad (13)$$

This is steady state probability distribution of the number of packets at a network system, also known as queue length distribution and is a power-law distribution. This is similar to the asymptotic behavior of the stationary queue length distribution obtained by [19] for $M/G/1$ with service time following power-tail distribution. Using (4), one can easily express p_n in terms of p_0 as

$$p_n = n^{-1/\rho}p_0, \quad n \geq 1, \quad \rho < 1. \quad (14)$$

2.1 Performance Measures

For analytical tractability of the performance evaluation of the $Pow/Pow/1/\infty$ queue, the fluid flow approximation is considered such that the number of packets arriving to the queue is assumed to be so large that it appears to be continuous flow of fluid. Under this approximation, discreteness of the random variable N can be neglected and it can be treated as a continuous random variable X , (chapter 2 of [36]). Fluid flow is a good approximation under heavy-traffic condition.

For computing quality of service (QoS) parameters, the average number of packets in the system, known as mean queue length, is required. The mean queue length for this model is

$$\bar{N} = \sum_{n=1}^{\infty} np_n = (1 - \rho) \sum_{n=1}^{\infty} n^{-\left(\frac{1}{\rho}-1\right)}. \quad (15)$$

Using fluid flow approximation, (15) becomes

$$\bar{N} = \frac{\rho(1 - \rho)}{(1 - 2\rho)}, \quad \rho < 1/2. \quad (16)$$

One can also compute the closed form expression for variance of number of packets σ^2 in the system under fluid flow approximation as

$$\sigma^2 = \rho(1 - \rho) \left[\frac{1}{1 - 3\rho} - \frac{\rho(1 - \rho)}{(1 - 2\rho)^2} \right], \quad \rho < 1/3. \quad (17)$$

The variance of number of packets become infinite as ρ tends to $1/3$.

One important QoS parameter is overflow probability which is the probability of exceeding a given buffer size ω viz.,

$$P(n > \omega) = \sum_{n=\omega+1}^{\infty} (1 - \rho)n^{-1/\rho} \quad (18)$$

which under heavy traffic can be simplified into

$$P(n > \omega) = \left(\frac{\rho}{1 - \rho} \right) (\omega + 1)^{-\left(\frac{1-\rho}{\rho}\right)}. \quad (19)$$

It is clear from (19) that the overflow probability also follows power-law in buffer sizes asymptotically

Table 1 Probabilities in $Pow/Pow/1/\infty$ Queue

n	ρ			
	0.1	0.2	0.5	0.9
2	8.79×10^{-4}	0.02500	0.12500	0.04629
5	9.22×10^{-8}	2.56×10^{-4}	0.02000	0.01672
10	9.00×10^{-11}	8.00×10^{-6}	0.00500	0.00774
20	8.79×10^{-14}	2.50×10^{-7}	0.00125	0.00358
50	9.22×10^{-18}	2.56×10^{-9}	0.00020	0.00129
100	9.00×10^{-21}	8.00×10^{-11}	5.00×10^{-5}	0.00059

Table 2 Comparison of approximated and exact mean queue length of $Pow/Pow/1/\infty$

ρ	Approximated mean queue length	Exact mean queue length
0.1	0.1125	0.9018
0.2	0.2667	0.8659
0.3	0.5250	0.9906
0.4	1.2000	1.5674
0.45	2.4750	2.8012
0.49	12.4950	12.7909
0.499	124.9995	125.2888
0.4999	1.2500×10^3	1.2503×10^3

$$P(n > \omega) \sim \omega^{-\left(\frac{1-\rho}{\rho}\right)}. \tag{20}$$

This result is in conformity with the one obtained by [37]. The average waiting time of the packets at network system can be computed by using Little’s law and is given by

$$\overline{W} = \frac{1}{\lambda} \left[\frac{\rho(1-\rho)}{1-2\rho} \right]. \tag{21}$$

Here, λ is the mean arrival rate of packets at the network system.

The performance measures are analyzed numerically to get deeper insight into $Pow/Pow/1/\infty$ queueing system.

2.2 Numerical Results

This subsection contains the numerical results evaluating performance of $Pow/Pow/1/\infty$ system. Table 1 shows the probability of number of packets, as obtained from (13) for various values of ρ . As expected, the probability is initially low and then increases with ρ .

Since the performance measures are derived using fluid flow approximation, it becomes imperative to examine its accuracy. The comparison of exact mean queue length (15) is carried out with approximated mean queue length (16) and results are presented in Table 2. The approximated mean queue length starts matching with exact mean queue length as traffic intensity parameter ρ increases. This is expected as fluid flow approximation provides accurate results under heavy traffic.

The mean number of packets (16) viz. mean queue length is plotted for various values of traffic intensity ρ in Fig. 1.

The mean queue length for $Pow/Pow/1/\infty$ queue remains lower for small values of ρ . As the traffic intensity crosses 0.3, the mean queue length increases fast. The mean queue length increases abruptly towards infinite as ρ approaches to 0.5 in case of $Pow/Pow/1/\infty$ system. These results comply to earlier empirical studies [38]. Hence, the proposed model captures the real network traffic burstiness.

The variance of number of packets (17) shows interesting property as shown in Fig. 2. The variance is not significant for lower values of ρ but increases sharply as ρ approaches 0.3. This is due to bursty nature of the input traffic.

The behavior of overflow probability (19) is plotted for various values of traffic intensity in Fig. 3 on log–log scale. The plot is a straight line for all values of traffic intensity confirming the presence of power-law behavior in overflow probability. For a given value of overflow probability, (19) can be used to compute number of buffers ω . Norros [39] provides a closed form expression for calculating buffer requirement, referred as Norros formula, when input traffic is governed by fractional Brownian motion process and service rate is constant. The Norros formula is

$$b = \frac{\rho^{\frac{1}{2(1-H)}}}{(1-\rho)^{\frac{H}{1-H}}}, \quad 1/2 \leq H < 1. \quad (22)$$

Here, H is the Hurst parameter measuring the degree of long range dependence. The buffer requirement as obtained from (19) for overflow probability of 0.01 is compared with Norros formula for $H = 0.938$ in Fig. 4. Both $Pow/Pow/1/\infty$ and Norros formula result in approximately same number of buffers for lower value of traffic intensity. However, as the traffic intensity increases, the buffer requirement is higher for $Pow/Pow/1/\infty$ system as compared to Norros formula. The reason is because Norros formula has been derived

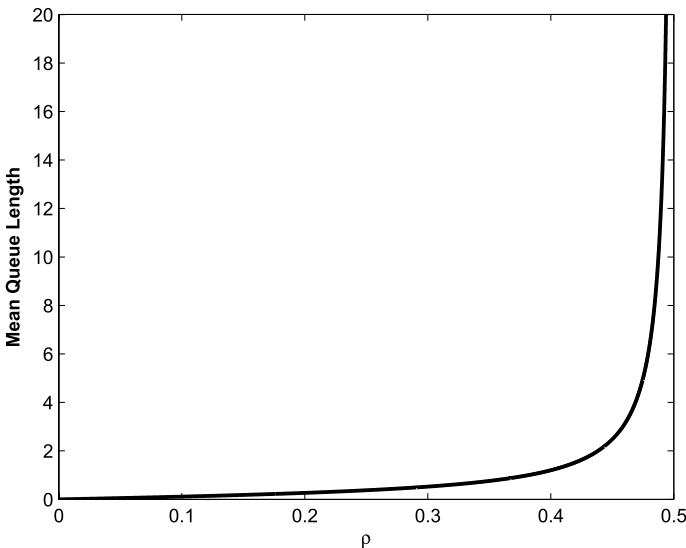


Fig. 1 Behavior of mean queue length of $Pow/Pow/1/\infty$ with ρ

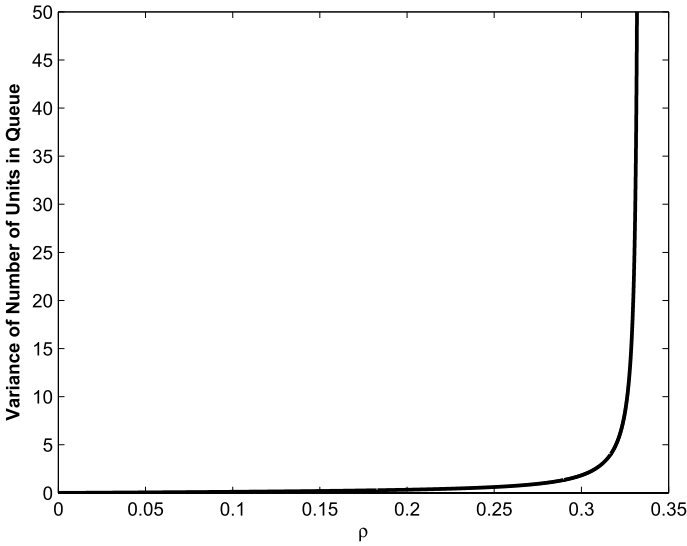


Fig. 2 Variance of number of packets in $Pow/Pow/1/\infty$ queue

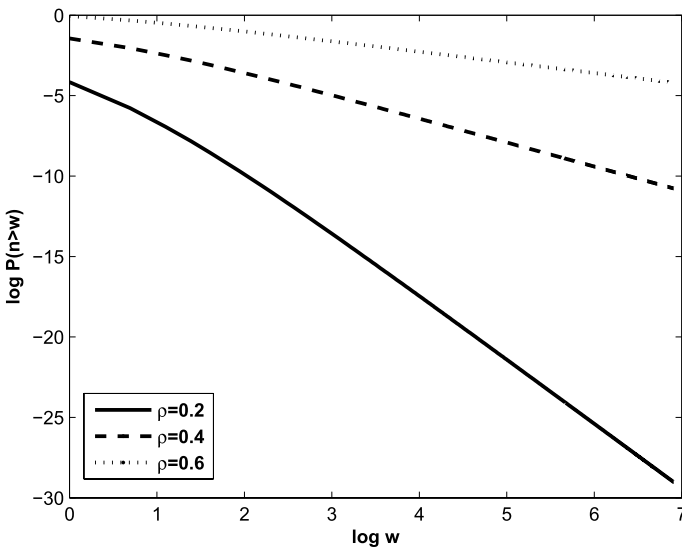


Fig. 3 Power-law behavior of overflow probability in $Pow/Pow/1/\infty$ queue

assuming constant service time, however, it is considered to be following power-law distribution in the proposed model.

The average waiting time (21) for different values of mean arrival rates is shown in Fig. 5. The average waiting time of packets in the system increases fast as ρ approaches to 0.5. This behavior of average waiting time is identical to the real network traffic where a sharp rise in mean delay is observed around 50% utilization [38].

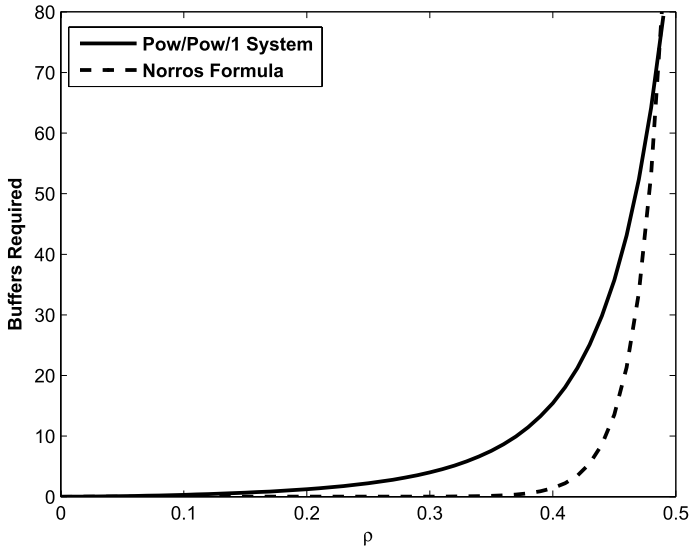


Fig. 4 Comparison of buffer requirements of $Pow/Pow/1/\infty$ with Norros formula

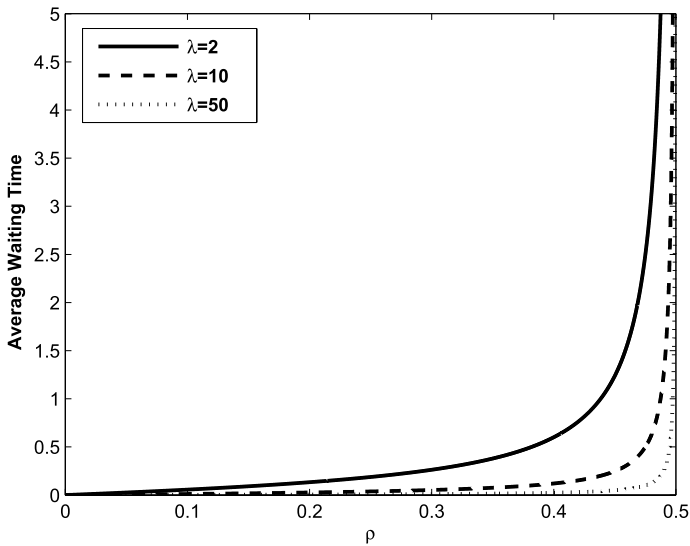


Fig. 5 Behavior of average waiting time in $Pow/Pow/1/\infty$ with ρ

3 Finite Buffer $Pow/Pow/1/K$ Model

In practice, network systems have finite number of buffers K where packets get stored and wait for processing. The performance of such a system can be analyzed by using $Pow/Pow/1/K$ queueing model. Following the same notations as used in Sect. 2, the

Shannon entropy of the finite buffer system becomes

$$S = - \sum_{n=0}^K p_n \ln p_n. \tag{23}$$

Since, geometric mean is considered to be characterizing moment for generating power-law [32], it is assumed that prior knowledge about the number of packets is available in the form of geometric mean

$$\sum_{n=1}^K (\ln n)p_n = \ln G. \tag{24}$$

The system utilization

$$U = 1 - p_0 = \rho. \tag{25}$$

and normalization constant

$$\sum_{n=0}^K p_n = 1. \tag{26}$$

are two other constraints. Following the same optimization procedure as outlined in Sect. 2, the queue-length distribution of finite buffer system can be derived as

$$p_n = C^{-1} U n^{-\mu} \tag{27}$$

where $C = \sum_{n=1}^K n^{-\mu}$ and μ is the Lagrange’s parameter corresponding to (24). To compute C , we again consider $\left\{ \frac{1}{n^\mu}, \mu > 1 \right\}$ which is a strictly decreasing sequence of positive numbers. Thus, taking $f(x) = \frac{1}{x^\mu}$, the summation can be approximated by integral in (27) transforming it into

$$p_n = \frac{U(\mu - 1)n^{-\mu}}{(1 - K^{1-\mu})}, \quad \mu > 1, n \geq 1. \tag{28}$$

Resembling the substitution for Lagrange’s parameter in previous Sect. 2, putting $\frac{1}{\mu} = \rho$, (28) becomes

$$p_n = \frac{(1 - \rho)n^{-1/\rho}}{1 - K^{-\frac{(1-\rho)}{\rho}}}, \quad n \geq 1, \rho < 1. \tag{29}$$

Drawing analogy from the application of MEP for deriving results of Poisson arrivals based queueing models [24], the form (29) can be regarded as the steady state probability distribution of the number of packets in *Pow/Pow/1/K* system, also known as queue length distribution. It is to be noted that queue length distribution follows power-law. The queue length distribution (29) can also be re-written in terms of p_0 using (25) as

$$p_n = \frac{n^{-1/\rho}}{1 - K^{-\frac{(1-\rho)}{\rho}}} p_0, \quad n \geq 1, \rho < 1. \tag{30}$$

The performance of the $Pow/Pow/1/K$ system is studied in the next subsection.

3.1 Performance Measures and Analysis

For deriving closed form solutions for various performance measures of the system, the fluid flow approximation is used as discussed in Sect. 2. The important QoS measure in case of finite buffer queueing system is blocking probability. Given a value of blocking probability, the network designers can use blocking probability closed-form expression to compute the buffer requirement at the network systems. Using (29), the blocking probability in case of $Pow/Pow/1/K$ is given by

$$P_K = \frac{(1 - \rho)K^{\frac{-1}{\rho}}}{1 - K^{\frac{-(1-\rho)}{\rho}}} \quad (31)$$

which also follows power-law. This is also clear from Fig. 6 where blocking probability (31) is plotted on log–log scale. Power-law behavior in blocking probability implies that increasing buffer sizes does not help to reduce the probability significantly.

The variation of blocking probability with buffer sizes for different value of utilization is shown in Fig. 7. The blocking probability is higher for higher values of utilization.

The closed-form of blocking probability (31) can be used to compute the required buffers at a network system. Some results of buffer sizes for corresponding blocking probability are presented in Table 3. Since buffers are integers, the calculated values are rounded off to nearest integer. It is observed that the buffer requirement is very less for low utilisation and increases sharply for higher utilization for specified blocking probability.

The expression for mean queue length can also be computed as

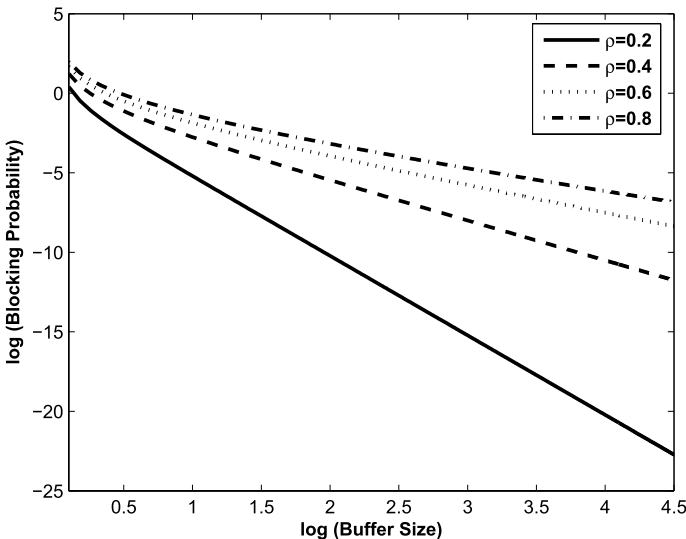


Fig. 6 Power-law behavior of blocking probability

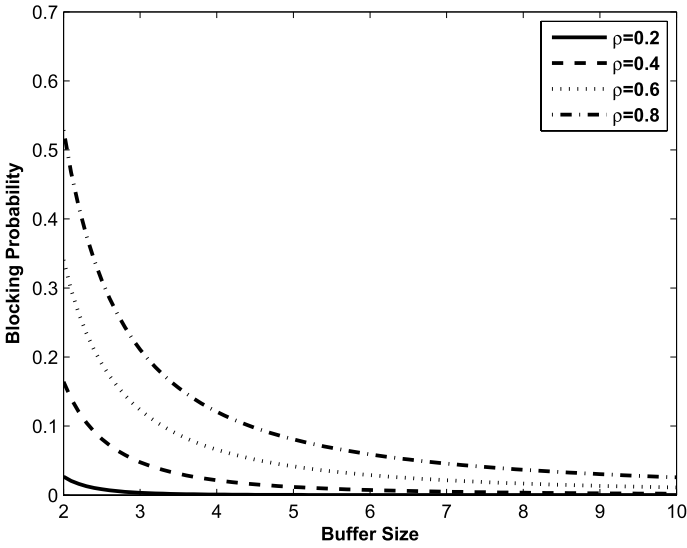


Fig. 7 Variation of blocking probability with buffer sizes

Table 3 Buffer size for given blocking probability

Blocking probability	Utilization		
	0.1	0.5	0.9
0.1	1	3	5
0.01	2	7	23
0.001	2	22	137
0.0001	2	71	887
0.00001	3	224	6117
0.000001	4	707	43,870

$$\bar{N} = \sum_{n=1}^K np_n = \sum_{n=1}^K \frac{(1-\rho)n^{-\frac{1}{\rho}+1}}{1 - K^{-\frac{(1-\rho)}{\rho}}} \tag{32}$$

which can be simplified into

$$\bar{N} = \frac{(1-\rho)\rho \left(1 - K^{-\frac{(1-2\rho)}{\rho}}\right)}{(1-2\rho) \left(1 - K^{-\frac{(1-\rho)}{\rho}}\right)}, \quad \rho \neq 1/2. \tag{33}$$

It would be imperative to examine the finite buffer power-law queueing system for multiple servers and it is discussed in the next section.

4 Multi-server Extension—Loss System: Pow/Pow/K/K

Finite buffer multi-server queueing system, also known as loss system, has well known applications for telecommunication systems. At a computer communication network system, ports represent servers. The channels are servers in case of wireless communication networks. Such systems are loss systems because the packets are lost if all the servers/ports/channels are busy. Using MEP, when arithmetic mean of number of packets is available as constraint, a model of loss system is developed in [40]. This model is useful when network traffic shows exponential behavior, for example, in case of voice traffic. However, present day communication network traffic depict different characteristics in the form of power-law behavior. The notations used in [40] are closely followed in this section. Consider a system with K servers, each server can be in two states viz. idle (0) or busy (1). Then, the state space of the system is $S = \{0, 1\}^K$ with $\mathbf{x} = \{x_1, x_2, \dots, x_K\} \in S$ as system state where $x_i = 1$ implies that the server i is busy. As discussed in [40], function f maps the state space S to a reduced state space M such that the number of busy servers or ports is

$$f(\mathbf{x}) = \sum_{i=1}^K x_i, \quad \mathbf{x} \in S. \quad (34)$$

The function f divides the state space S into equivalence classes

$$[n] = \{\mathbf{x} \in S : f(\mathbf{x}) = n\}, \quad n \in M = \{0, 1, \dots, K\}. \quad (35)$$

The probability of system been in state \mathbf{x} is given by p_x . It is assumed that the prior knowledge about the geometric mean of the number of packets in the system and utilization is available i.e.

$$\sum_{\mathbf{x} \in S} (\ln f(\mathbf{x})) p_x = \ln G \quad (36)$$

and

$$1 - p_0 = U = \rho \quad (37)$$

The normalization constant is

$$\sum_{\mathbf{x} \in S} p_x = 1. \quad (38)$$

When the Shannon entropy (23) is maximized with constraints (36), (37) and (38), the state probability distribution in steady state turns out to be

$$p_x = \frac{U [f(\mathbf{x})]^{-\alpha}}{\sum_{\mathbf{x} \in S} [f(\mathbf{x})]^{-\alpha}} \quad (39)$$

where α is the Lagrange's parameter. Since the cardinality of the equivalence classes $[n]$ is $\binom{N}{n}$, the probability distribution (39) on reduced state space can be expressed by

$$b(n) = \binom{K}{n} \frac{U \cdot n^{-\alpha}}{\sum_{i=1}^K \binom{K}{i} i^{-\alpha}} \tag{40}$$

Substituting $\rho = \frac{1}{\alpha}$, the steady state probability distribution (40) turns into

$$b(n) = \binom{K}{n} \frac{\rho \cdot n^{-\frac{1}{\rho}}}{\sum_{i=1}^K \binom{K}{i} i^{-\frac{1}{\rho}}}. \tag{41}$$

The probability that all the servers are busy i.e. blocking probability in this case can be derived from (41) as

$$b(K) = \frac{\rho \cdot K^{-\frac{1}{\rho}}}{\sum_{i=1}^K \binom{K}{i} i^{-\frac{1}{\rho}}} \tag{42}$$

which plays the same role in power-law queueing system as the well-known Erlang’s loss formula does for Poisson arrivals based queueing system. Thus, (42) can be regarded as generalized Erlang loss formula.

The state probability distribution $b(n)$ is plotted for $\rho = 0.1$ in Fig. 8. Interestingly, the state probability shows bimodal behavior corresponding to periods of low and high load. But as the ρ is increased, the bimodal behavior disappears, as shown in Fig. 9 indicating only high load on the system. The similar behavior is also observed in [27]. The comparison of state probability distribution (42) with Erlang loss model is presented in Table 4. The blocking probability is high for all values of maximum buffer size N and utilisation ρ in case of $Pow/Pow/K/K$ system as compared to Erlang loss formula. It implies that more buffers are required at a system in case network is driven by power-law traffic.

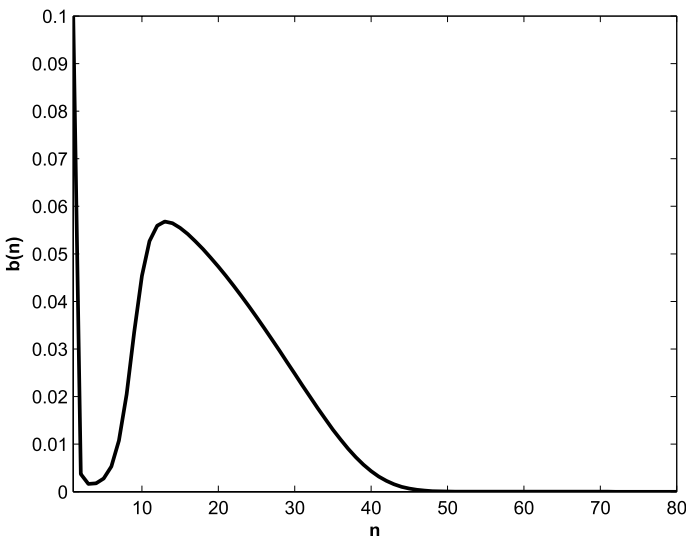


Fig. 8 Behavior of state probability distribution for $\rho = 0.1$

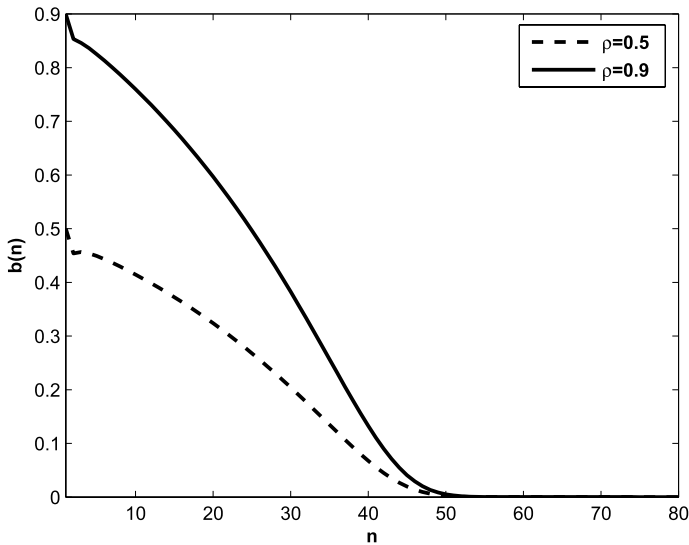


Fig. 9 Behavior of state probability distribution for $\rho = 0.5$ and $\rho = 0.9$

Table 4 Comparison of state probability distribution $b(N)$ with Erlang loss formula

ρ	K	$b(K)$	Erlang loss formula
0.1	10	9.9540×10^{-13}	1.0625×10^{-17}
	20	4.8316×10^{-16}	6.1097×10^{-40}
	50	9.5571×10^{-21}	2.8009×10^{-117}
0.5	10	7.4730×10^{-5}	4.6669×10^{-12}
	20	9.8604×10^{-8}	1.1788×10^{-28}
	50	1.0406×10^{-16}	4.5800×10^{-89}
0.9	10	8.4618×10^{-4}	1.5672×10^{-10}
	20	3.7092×10^{-7}	1.3294×10^{-25}
	50	3.6099×10^{-16}	1.9560×10^{-81}

5 Discussion and Conclusion

Extending the MEP framework proposed in [35], two finite buffer models are proposed in this paper where both inter-arrival and service times follow power-law distribution. First, an infinite buffer model $Pow/Pow/1/\infty$ has been discussed [35]. The numerical study and analysis of the model shows that such a queueing system explodes at very low utilization viz. mean number of packets shows sharp rise as ρ tends to 0.5, the variance of number of packets starts to explode as ρ approaches to 0.3. These results are in agreement with real traffic studies of network with LRD behavior [38].

A finite buffer queueing system $Pow/Pow/1/K$ has been proposed to generate closed form expressions for queue length distribution, QoS measures viz. blocking probability and mean queue length. The queue length distribution and blocking probability are found to

depict power-law. The numerical results validate that the blocking probability formula can be used to compute the buffer requirements. The analysis validates that the buffers will be occupied sharply as the utilization of the system increases to 50% introducing longer delays at low utilization. This is in conformity with the empirical studies carried out in the past [38]. The model is extended to multiple servers and state probability distribution for loss system $Pow/Pow/K/K$ are derived. The state probability distribution shows bimodal behavior at low utilization which turns into uni-modal at high utilization. A comparative study of numerical results of $Pow/Pow/K/K$ with Erlang loss model is carried out and it is observed that the blocking probability is very high in case $Pow/Pow/K/K$ system. These results are of significant importance to network designers facilitating them in dimensioning resources at network systems and in developing efficient admission and congestion control policies.

An exemplary application of the proposed models is in the congestion control at a network system. One of the famous congestion avoidance algorithm for gateway systems is random early detection (RED) [41]. The gateway system detects congestion by calculating mean queue size upon arrival of every packet. When the mean queue size crosses a pre-specified threshold, the gateway starts dropping packets with some dropping probability. In order to find the initial value of mean queue size, a procedure is proposed in [41] that allows minimum number of packets L to arrive at the gateway and then computes mean queue size. The mean queue length formula (33) can be used to compute mean queue sizes in the RED algorithm. This requires the knowledge of utilization ρ . One way to approximate the server utilization is by CPU utilization. All operating systems supports system calls to obtain utilization of CPU. Thus, the requirement to wait for L packets in RED does not exist. The expression for overflow probability (19) can be utilized for calculating dropping probability. Also, the probability of occupancy of buffers between two thresholds viz. \min_{th} and \max_{th} can be computed using (13) as

$$p_b = \int_{\min_{th}}^{\max_{th}} (1 - \rho)x^{-1/\rho} dx \quad (43)$$

A thorough comparative analysis of the aforementioned methods in RED algorithm is an area of future work.

Author Contributions The author Dr. SS has solely contributed to the work including conceptualization of idea, implementation, validation, writing the manuscript.

Funding Not applicable

Data Availability Not applicable.

Code Availability The code will be provided as per the requests.

Declarations

Conflict of interest The author declares that there is no conflict of interest.

Ethical Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

References

1. Karmeshu, Sharma, S., & Kumar, S. (2015). Generation of power law: Maximum entropy framework and superstatistics. In *4th international conference on man-machine interactions (ICMMI)* (pp. 45–59).
2. Mahanti, A., Carlsson, N., Mahanti, A., Arlitt, M., & Williamson, C. (2013). A tale of tails: Power-laws in internet measurements. *IEEE Network*, *27*(1), 59–64.
3. Downey, A. B. (2005). Lognormal and Pareto distribution in the internet. *Computer Communications*, *28*(7), 790–801.
4. Ghosh, D., Jagannathan, K., & Raina, G. (2020). Right buffer sizing matters: Some dynamical and statistical studies on Compound TCP. *Performance Evaluation*, *139*, 102095.
5. Song, Y. D., & Mahanti, A. (2019). Comparison of mobile and fixed device workloads in an academic web server. In *IEEE international symposium on measurements & networking*.
6. Mahanti, A., Carlsson, N., Arlitt, M., & Williamson, C. (2012). Characterizing cyberlocker traffic flows. In *37th annual IEEE conference on local computer networks*.
7. Mitzenmacher, M. (2003). A brief history of generative models for power-law and lognormal distributions. *Internet Mathematics*, *1*(2), 226–251.
8. Mahanti, A., Carlsson, N., & Williamson, C. (2012). Content sharing dynamics in the global file hosting landscape. In *IEEE 20th symposium on modeling, analysis and simulation of computer and telecommunication systems*.
9. Song, Y. D., Mahanti, A., & Ravichandran, S. C. (2019). Understanding Evolution and Adoption of Top Level Domains and DNSSEC. In *IEEE international symposium on measurements & networking*.
10. Yu, R., Christophersen, C., Song, Y. D., & Mahanti, A. Comparative analysis of adult video streaming services: Characteristics and workload. In *Network traffic measurement and analysis conference*.
11. Basher, N., Mahanti, A., Mahanti, A., Williamson, C., & Arlitt, M. (2008). A comparative analysis of web and peer-to-peer traffic. In *WWW08: Proceedings of the 17th international conference on World Wide Web*.
12. Kokoszka, P., Nguyen, H., Wang, H., & Yang, L. (2019). Statistical and probabilistic analysis of inter-arrival and waiting times of Internet2 anomalies. *Statistical Methods & Applications*, *29*, 727–744.
13. Mohamed, A. M., & Agamy, A. (2019). Effect of application types on the performance of Wifi networks. In *International conference on innovative trends in computer engineering (ITCE)*.
14. Bie, Y., Wang, L., Tian, Y., & Hu, Z. (2019). A combined forecasting model for satellite network self-similar traffic. *IEEE Access*, *7*, 52004–152013.
15. Majumdar, C., Lopez-Benitez, M., & Merchant, S. N. (2019). Experimental evaluation of the Poissonness of real sensor data traffic in the Internet of Things. In *2019 16th IEEE annual consumer communications & networking conference (CCNC)* (pp. 1–7).
16. Arfeen, A., Pawlikowski, K., McNickle, D., & Willig, A. (2019). Global and local scaling analysis of link streams in access and backbone core networks. *Computer Networks*, *149*, 154–172.
17. Mishra, S., Smoreda, Z., & Fiore, M. (2022). Second-level digital divide: A longitudinal study of mobile traffic consumption imbalance in France. In *Proceedings of the ACM Web Conference 2022* (pp. 2532–2540).
18. Kepner, J., Cho, K., Claffy, K. C., Gadepally, V., McGuire, S., Milechin, L., ... & Michaleas, P. (2022). New phenomena in large-scale internet traffic. [arXiv:2201.06096](https://arxiv.org/abs/2201.06096).
19. Roughan, M., Veitch, D., & Rumsewicz M. (1998). Computing queue-length distributions for power-law queues. In *Proceedings, IEEE INFOCOM*.
20. Addie, R. G., Neame, T. D., & Zukerman, M. (2009). Performance analysis of a Poisson-Pareto queue over the full range of system parameters. *Computer Networks*, *53*(7), 1099–1113.
21. Kim, H., & Shroff, N. (2001). Loss probability calculations and asymptotic analysis for finite buffer multiplexers. *IEEE/ACM Transactions on Networking*, *9*(6), 735–768.
22. Lin, G. C., Suda, T., & Ishizaki, F. (2005). Loss Probability for a Finite Buffer Multiplexer with the M/G/∞ Input Process. *Telecommunication Systems*, *29*, 181–197.
23. Ciucu, F., Poloczek, F., & Tu, A. R. (2019). Queue and loss distributions in finite-buffer queues. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, *3*(2), 31:1-31:29.
24. Kapur, J. N., & Kesavan, H. K. (1992). *Entropy optimization principles with application*. London: Academic Press Inc.
25. He, D., Li, R., Huang, Q., & Lei, P. (2014). Maximum entropy principle based estimation of performance distribution in queueing theory. *PLoS ONE*, *9*, 1. <https://doi.org/10.1371/journal.pone.0106965>
26. Kovatsos, D. D. (1988). A maximum entropy analysis of the G/G/1 queue at equilibrium. *Journal of Operational Research Society*, *39*(2), 183–200.
27. Sharma, S., & Karmeshu. (2008). Bimodal packet distribution in loss systems using maximum Tsallis entropy principle. *IEEE Transactions on Communications*, *56*(9), 1530–1535.

28. Sharma, S., & Karmeshu. (2009). Power-law characteristics and loss probability: Finite buffer queueing systems. *IEEE Communication Letters*, 13(12), 971–973.
29. Singh, A. K., & Karmeshu. (2014). Power-law behavior of queue size: Maximum entropy principle with shifted geometric mean constraint. *IEEE Communication Letters*, 18(8), 1335–1338.
30. Singh, A. K., et al. (2015). Analysis of finite buffer queue: Maximum entropy probability distribution with shifted fractional geometric and arithmetic means. *IEEE Communication Letters*, 19(2), 163–166.
31. Rostovtsev, A. (2005). On a geometric mean and power-law statistical distributions. [arXiv:cond-mat/0507414](https://arxiv.org/abs/cond-mat/0507414)
32. Visser, M. (2013). Zipfs law, power laws and maximum entropy. *New Journal of Physics*, 15, 1–13.
33. Peng, X., Bai, B., Zhang, G., Lan, Y., Qi, H., & Towsley, D. (2018). Bit-level power-law queueing theory with applications in LTE networks. IN *2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates* (pp. 1–6). <https://doi.org/10.1109/GLOCOM.2018.8647562>
34. Navarro-Ortiz, J., Romero-Diaz, P., Sendra, S., Ameigeiras, P., Ramos-Munoz, J. J., & Lopez-Soler, J. M. (2020). A survey on 5G usage scenarios and traffic models. *IEEE Communications Surveys & Tutorials*, 22(2), 905–929. <https://doi.org/10.1109/COMST.2020.2971781>
35. Sharma, S. (2022). Performance analysis of Pow/Pow/1/∞ queue under heavy network traffic using maximum entropy approach. In *IEEE international conference on advanced networks and telecommunications systems*.
36. Kleinock, L. (1976). *Queueing Systems: Computer Applications* (1st ed., Vol. 2). Berlin: Wiley-Interscience.
37. Choudhury, G. L., & Whitt, W. (1997). Long-tail buffer content distributions in broadband networks. *Performance Evaluation*, 30(3), 177–190.
38. Erramilli, A., Narayan, O., & Willinger, W. (1996). Experimental queueing analysis with long-range dependent packet traffic. *IEEE/ACM Transactions on Networking*, 4(2), 209–223.
39. Norros, I. (1994). A storage model with self-similar input. *Queueing Systems*, 16, 387–396.
40. Paz, J. M., Mark, B. L., & Kobayashi, H. (1995). A maximum entropy approach to the analysis of loss systems. In *Proceedings of IEEE Singapore international conference on networks and international conference on information engineering* (pp. 436–441).
41. Flyod, S., & Jacobson, V. (1993). Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1(4), 397–413.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Shachi Sharma (Senior Member, IEEE) received her Ph.D. degree from the School of Computer and System Sciences, Jawaharlal Nehru University, New Delhi, India, in 2007. Her total work experience of more than 22 years is a mix of academia and industry. She worked with Center for Development of Telematics (C-DOT) India, Alcatel-Lucent India Private Limited and Bharti Telesoft (now Tech Mahindra), India before completing her PhD degree. Post Ph.D., she was a research scientist with IBM Research Laboratory, New Delhi, India from 2008 to 2017. She was a visiting faculty with Indraprastha Institute of Information Technology (IIIT) Delhi, India, in 2018. Thereafter, she joined Department of Computer Science, South Asian University, New Delhi, India. Her research work is substantiated by many international research publications and U.S. patents. Her research interests include stochastic modeling and applications of maximum entropy framework in diverse areas of Computer Science. Presently, she is working on designing appropriate traffic models for computer communication networks.