JOURNAL OF INDUSTRIAL AND
MANAGEMENT OPTIMIZATION
Volume **17**, Number **1**, January **2021**

doi:10.3934/jimo.2019106

pp. **185–204**

AN APPROXIMATE MEAN QUEUE LENGTH FORMULA FOR QUEUEING SYSTEMS WITH VARYING SERVICE RATE

Jian Zhang*

Shanghai Jiao Tong University
800 Dongchuan Rd, Shanghai, China

Tony T. Lee, Tong Ye and Liang Huang

The Chinese University of Hong Kong(Shenzhen)
Shanghai Jiao Tong University
Zhejiang University of Technology

(Communicated by Wuyi Yue)

Abstract. In this paper, we analyze the delay performance of queueing systems in which the service rate varies with time and the number of service states may be infinite. Except in some simple special cases, in general, the queueing model with varying service rate is mathematically intractable. Motivated by the P-K formula for M/G/1 queue, we developed a limiting analysis approach based on the connection between the fluctuation of service rate and the mean queue length. Considering the two extreme service rates, we provide a lower bound and upper bound of mean queue length. Furthermore, an approximate mean queue length formula is derived from the convex combination of these two bounds. The accuracy of our approximation has been confirmed by extensive simulation studies with different system parameters. We also verified that all limiting cases of the system behavior are consistent with the predictions made by our formula.

1. Introduction.

In many newly emerging communication services, such as peer-to-peer (P2P) video streaming [10], file sharing [5], and cloud computing, both jobs and servers arrive and depart randomly. Thus, the number of available servers seen by a specific job arrival is a random variable, and the service rate for this job is time varying. To explore the behavior of this kind of service systems, the performance analysis of modern communication networks calls for a new kind of queueing model with varying service rate [11][19][15].

In almost all previously published works [7][12][3][6], queueing systems with time varying service rate were modeled as a single server queue with a finite number of service states, in which the state transition of the server can be described as a continuous-time Markov chain. They are solvable if the number of service states is small, say only two or three states, such as the model of a multi-rate wireless channel. However, those methods are totally failed when the number of service states is large or even infinite.

2020 *Mathematics Subject Classification.* Primary: 60K20, 60K25; Secondary: 68M20.

Key words and phrases. Continuous time Markov chain, M/G/1 queue, service rate fluctuation.

* Corresponding author: Jian Zhang.

The queueing system with a large number of service states is commonly seen in, for example, volunteer computing [1], also known as public-resource computing or peer-to-peer computing, which uses the spare computing power belonging to general public to perform scientific supercomputing tasks [4]. In this kind of systems, a supercomputing task is divided into a large number of small subtasks, which are then spread to all active volunteers around the world for parallel calculation. Thus, the total calculation (or service) capacity is the sum of the power of all the active volunteers, the number of which however may change over the time during such parallel process. The active volunteers can freely turn off the hosts and generate useless results, which reduces the service capacity. On the other hand, the volunteers may randomly join or rejoin the calculation for more assigned subtasks. When the computation of a subtask is completed, the volunteer sends back the result to the server [2]. Treating each supercomputing task as a job, we can consider the volunteer computing system as a kind of queueing system with multiple service states. It was reported in [20] that the number of volunteers involved in a project can even be larger than 15,000,000. This implies that the service capacity of the volunteer computing system could have a huge number of service states.

The major obstacle to analyzing the queueing model with varying service rate is the dependency on service times among different customers. In response to this issue, [12] introduced the concept of start-service probability, which provides the basis of the generalized P-K formula for two-state Markov channels derived in [7]. Despite that, the similar approach based on start-service probability is mathematically intractable for the queueing model with infinite number of states. Nevertheless, in this paper, based on the relationship between mean queue length and service rate variance, we provide a methodology to estimate the mean queue length of queueing systems with infinite number of service states.

1.1. Previous work.

The queuing model with variable service rate is an important analytical tool in studying multi-rate communication systems or computing systems. [16] and [14] investigated the behavior of power-aware servers in data centers, where the service rate of the server changes proportionally with the number of jobs waiting in the buffer. In particular, [16] considered the case where jobs arrive at the server in batches, and [14] studied the case where blocked jobs leave the server and retry after retrial times. [3] first introduced the two-state queueing model of a wireless channel. [6] derived the mean delay for the system with a two-state server, in which one of the service rates is zero. To cope with the dependencies among service times, a novel approach based on conditional moments of service time was proposed by [12]. However, their analysis is incomplete because the required start-service probabilities are only available for some extreme cases. The complete start-service probabilities and the closed-form mean delay formula for general two-state queueing model were derived in [7], which was extended to three-state queueing model by [18]. A matrix-geometric method for multi-state service process was developed in [13], which only provides numerical results and lacks of physical insight.

A different kind of queuing model with variable service rate is multi-server queuing system, which possesses two features: different servers can serve different jobs in parallel, and the number of servers in the system varies over the time. For example, usage problem of multiple service channels in broadband integrated services digital network (B-ISDN) [11] and transmission behavior of wireless channels in multiple-input-multiple-output (MIMO) systems [19], long term evolution-orthogonal frequency division multiplexing (LTE-OFDM) systems [15], or computer systems subject to technical obstacles [17] are such kind of multi-server queues with variable service rate. However, the number of servers in these models is limited and usually very small.

1.2. Our approach and contribution.

In this paper, based on the connection between the fluctuation of service rate and the mean queue length, we provide two bounds of mean queue length, and derive an approximate mean queue length formula for the queueing model. Our methodology and results are summarized as follows:

1. From simulation, we observe that mean queue length is increasing with the variance of the service rate when keeping service capacity constant. Based on this relationship, we provide two delay bounds for our queueing model by considering the two limiting cases of service rate fluctuation. When the variance of service rate approaches zero, the mean queue length reaches the lower bound. When the variance approaches infinity, the mean queue length reaches the upper bound.
2. From the convex combination of the lower bound and upper bound, we derive an approximate mean queue length formula. The accuracy of our approximation has been verified by extensive simulations, and all limiting cases of the model behavior agree with the predictions made by the formula.

The limiting analysis and approximate estimation of mean queue length developed for our queueing model can be extended and applied to other queueing systems with varying service rates. The innovative approaches proposed in this paper that are of particular interest include the extreme analysis of the service rate fluctuations, and the technique of approximate estimation based on conditional mean queue length and convex combination of extreme bounds.

The rest of the paper is organized as follows. In Section 2, we describe the queueing model and derive the differential equation associated with the number of jobs and number of servers in the system. A simple relationship between these two numbers is obtained from the differential equation. The stability condition of the queueing model is also provided in this section. In Section 3, we demonstrate the connection between the service rate fluctuation and the mean queue length, and provide two bounds of mean queue length. In Section 4, based on the conditional mean queue length and the convex combination of the two bounds, we derive an approximate mean queue length formula, and verify the accuracy of our approximation by using extensive simulations and limiting analysis. Section 5 draws a conclusion.

2. Markov Chain of the queueing model.

2.1. System description.

In this section, we consider a queueing system, such as the volunteer computing system, in which each job is served in parallel by a large number of servers and the servers may independently and randomly arrive and depart during the service. We assume that the server arrival process $\{n_s(t), t \geq 0\}$ is a Poisson process with rate λ_s, and the system is homogeneous, meaning that all servers are independent and identical. Also, the lifetime of a server is exponentially distributed with mean $\frac{1}{\mu_s}$, and the service rate of each server is μ_c jobs per unit of time. As Fig. 1 shows, the continuous time Markov chain of the server process can be viewed as an $M/M/\infty$ queue. In steady state, it is easy to show that the number of servers in the system, defined as $n_s = \lim\limits_{t\to\infty} n_s(t)$, is a Poisson random variable with parameter $\rho_s = \frac{\lambda_s}{\mu_s}$.

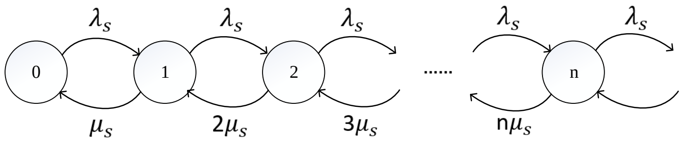


FIGURE 1. The continuous time Markov chain of the server process.

In the queueing model under consideration, the user can only process the download jobs one-by-one in a first-in-first-out (FIFO) manner. Furthermore, we assume that the job arrival process $\{n_c(t), t \geq 0\}$ is also a Poisson process with rate λ_c, which is independent of the server process. The service of jobs waiting in line follows first-come, first-served (FCFS) policy. At any instant of time, only one job will be simultaneously served by all servers. Therefore, the job service time depends on the number of servers in the system. If the number of servers is a constant within the service time of a job, then the job service time is exponentially distributed. However, as the number of servers may change within the service time of a job, the exact job service time distribution is unknown. Since the service times of jobs are dependent, the Kendall's notation of queueing systems cannot be extended to the queueing model, which is characterized by the following set of parameters:

- Job arrival rate: λ_c,
  Service rate of one server: μ_c,
  $\rho_c = \frac{\lambda_c}{\mu_c}$.
- Server arrival rate: λ_s,
  Mean server lifetime: $\frac{1}{\mu_s}$,
  $\rho_s = \frac{\lambda_s}{\mu_s}$.

Furthermore, we define the following notations used throughout this paper.

- $n_s(t)$: number of servers at time t,
  $n_s = \lim\limits_{t\to\infty} n_s(t)$: Poisson random variable of server number with parameter ρ_s when the system is in steady state.
- $n_c(t)$: number of jobs at time t,
  $n_c = \lim\limits_{t\to\infty} n_c(t)$: random variable of job number when the system is in steady state.
- $\mu(t) = n_s(t)\mu_c$: instantaneous service rate at time t,
  $\mu = \lim\limits_{t\to\infty} \mu(t)$: random variable of service rate when system is in steady state.

2.2. **Continuous-time Markov chain.** According to the previous description, Fig. 2 shows the continuous-time Markov chain of the queueing model with state space $\{(i,j), i = 0,1,2,\dots, j = 0,1,2,\dots\}$, where i is the number of jobs and j is the number of servers in the system in steady state.

Let $p_{i,j}$ denote the steady state probability that the system is in state (i,j), that is

$$p_{i,j} = Pr\{n_c = i, n_s = j\}, \tag{1}$$

where $i = 0,1,\dots$ is the number of jobs and $j = 0,1,\dots$ is the number of servers. From the state transition diagram shown in Fig. 2, we directly obtain the following

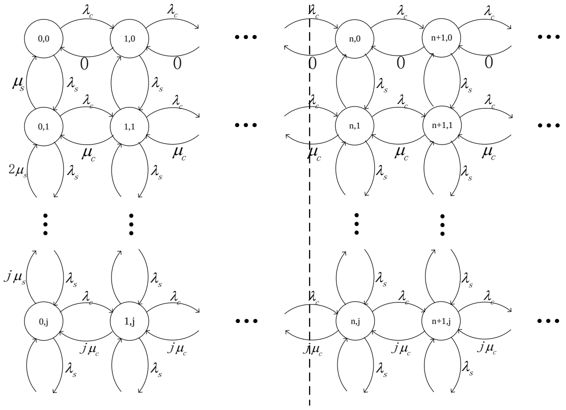


FIGURE 2. The continuous-time Markov chain of the queueing model.

set of balance equations:

$$(\lambda_s + \lambda_c)p_{0,0} = \mu_s p_{0,1} \quad i = 0, j = 0, \tag{2a}$$

$$(\lambda_s + \lambda_c)p_{i,0} = \lambda_c p_{i-1,0} + \mu_s p_{i,1} \quad i \geq 1, j = 0, \tag{2b}$$

$$(\lambda_s + \lambda_c + j\mu_s)p_{0,j} = \lambda_s p_{0,j-1} + j\mu_c p_{1,j} + (j+1)\mu_s p_{0,j+1} \quad i = 0, j \geq 1, \tag{2c}$$

$$(\lambda_s + \lambda_c + j\mu_s + j\mu_c)p_{i,j} = \lambda_s p_{i,j-1} + \lambda_c p_{i-1,j} + j\mu_c p_{i+1,j} + (j+1)\mu_s p_{i,j+1} \\ i \geq 1, j \geq 1. \tag{2d}$$

Define the following generating function of $p_{i,j}$:

$$F(z_1, z_2) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} p_{i,j} z_1^i z_2^j \quad |z_1| \leq 1, |z_2| \leq 1. \tag{3}$$

From the set of balance equations (2), we can derive the following differential equation of the generating function $F(z_1, z_2)$:

$$\begin{aligned}(\lambda_c - z_1\lambda_c + \lambda_s - z_2\lambda_s)F(z_1, z_2) = \quad & \mu_s(1 - z_2)\frac{\partial F(z_1, z_2)}{\partial z_2} \\ & + z_2\mu_c(\frac{1}{z_1} - 1)\frac{\partial F(z_1, z_2)}{\partial z_2} \\ & + (1 - \frac{1}{z_1})z_2\mu_c\frac{\partial F(0, z_2)}{\partial z_2}.\end{aligned} \tag{4}$$

Unfortunately, this Markov chain is irreversible. We know from [8] that there is no systematic way for solving irreversible Markov chain, and only some special cases can be solved in closed form. To the best of our knowledge, solving this differential equation to obtain a closed-form solution of $F(z_1, z_2)$ is mathematically intractable. That is, we cannot directly obtain the mean queue length of the system from (4). Nevertheless, this differential equation still provides us with some useful information regarding the performance of the queueing model.

First, applying $\frac{\partial}{\partial z_1}$ to equation (4) and then inserting $z_1 = z_2 = 1$, we obtain:

$$E[n_s] = \rho_c + \frac{\partial F(0, z_2)}{\partial z_2}|_{z_2=1}. \tag{5}$$

Similarly, applying $\frac{\partial^2}{\partial z_1^2}$ to equation (4) and then inserting $z_1 = z_2 = 1$, we obtain:

$$\rho_c E[n_c] = E[n_s n_c] - E[n_s] + \frac{\partial F(0, z_2)}{\partial z_2}|_{z_2=1}. \tag{6}$$

Combining (5) and (6), we have

$$\rho_c E[n_c] = E[n_s n_c] - \rho_c. \tag{7}$$

Since the number of jobs n_c is correlated with the number of servers n_s, the mean queue length of jobs $E[n_c]$ remains unsolvable from (7).

From the balance equations of the Markov chain separated by the dashed line shown in Fig. 2, we have

$$\lambda_c \sum_{j=0}^{\infty} p_{n-1,j} = \sum_{j=0}^{\infty} j\mu_c p_{n,j}. \tag{8}$$

Summing (8) over all $n \geq 1$, we have

$$\lambda_c = \sum_{j=0}^{\infty} j\mu_c (Pr\{n_s = j\} - p_{0,j}). \tag{9}$$

Define $\bar{\mu} = \sum_{j=0}^{\infty} j\mu_c Pr\{n_s = j\} = \rho_s\mu_c$ as the capacity of our queueing model. From (9), we have

$$\lambda_c = \bar{\mu} - \sum_{j=0}^{\infty} j\mu_c p_{0,j} < \bar{\mu}, \tag{10a}$$

or equivalently,

$$\rho_c < \rho_s, \tag{10b}$$

which gives rise to the stability condition that the job arrival rate λ_c is less than the capacity of the system, otherwise the system will be unstable and the queue length will approach to infinite.

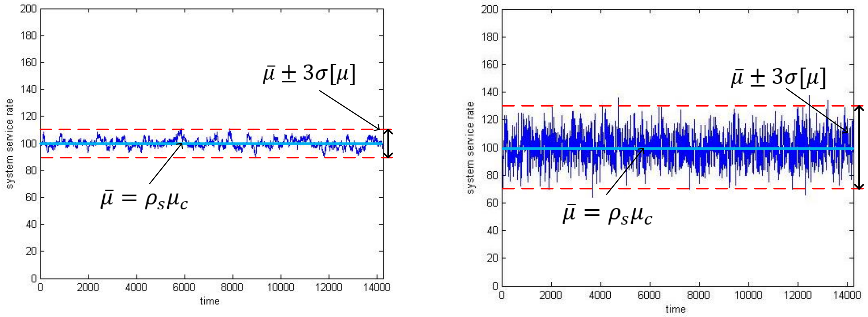
3. Service rate fluctuation and delay bound.

For our infinite-state queuing model, the lack of service time distribution and the dependency among service times of different jobs are major obstacles for deriving the mean queue length of jobs. The only known information related to the service time is the Poisson distribution of the number of servers, which determines the service rate distribution. In this section, we first investigate the influence of service rate fluctuations on the mean queue length of jobs, and then derive the mean queue lengths by considering two limiting cases of service rate.

3.1. Influence of service rate fluctuation.

In our queueing model, the number of active servers is a random variable and the service rate fluctuates over the time. Given that the number of servers n_s is a Poisson random variable with parameter ρ_s, we immediately obtain the following parameters related to the number of servers and service rate:

- Average number of servers: $E[n_s] = \rho_s = \frac{\lambda_s}{\mu_s}$,
- Standard deviation of server number: $\sigma[n_s] = \sqrt{\rho_s}$,
- System capacity: $\bar{\mu} = E[\mu] = \rho_s\mu_c$,
- Standard deviation of service rate: $\sigma[\mu] = \mu_c\sqrt{\rho_s} = \sqrt{\bar{\mu}\mu_c}$.



(a) $\mu_c = 0.1$

(b) $\mu_c = 1$

FIGURE 3. The fluctuation of service rate μ over the time with parameter $\frac{\mu_c}{\mu_s} = 10$, $\lambda_s = 10$.

If we keep parameter $\frac{\mu_c}{\mu_s}$ and λ_s constant and only change the value of μ_c, the service capacity would be a constant but the standard deviation $\sigma[\mu] = \sqrt{\bar{\mu}\mu_c}$ increases with μ_c. For example, if both parameters $\frac{\mu_c}{\mu_s} = 10$ and $\lambda_s = 10$ are fixed, the service capacity $\bar{\mu}$ equals a constant 100. Fig. 3 demonstrates two cases: $\mu_c = 0.1$ in Fig. 3(a) and $\mu_c = 1$ in Fig. 3(b). It is obvious that the service rate fluctuates more in (b) with larger μ_c. Since a larger μ_c and μ_s means each server has a relatively large unit service rate and the number of server is small. In this case, the arrival or departure of a server will lead to a larger change in the service rate than that of a smaller μ_c and μ_s. On the other hand, when both μ_c and μ_s are small, each server has a small unit service rate but the number of servers is large. Therefore, an arrival or departure of a single server hardly influences the service rate, which is the reason why the service rate is more static in Fig. 3(a).

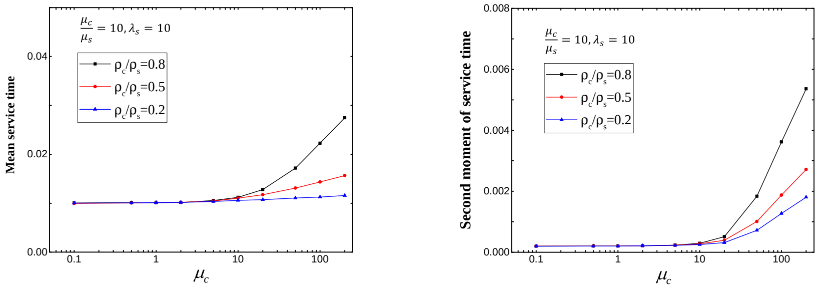
Intuitively, for a fixed capacity $\bar{\mu}$, both the mean and variance of service time will increase with respect to the fluctuation of service rate. Our simulation results clearly verify this property. As Fig. 4 shows, as the service capacity is fixed at $\bar{\mu} = \rho_s\mu_c = 100$, both the first and second moments of the service time increase with the variance of service rate $\sigma^2[\mu] = \bar{\mu}\mu_c$, especially when $\frac{\rho_c}{\rho_s}$ is large.

In a queueing system with a fixed arrival rate λ, the mean queue length would monotonically increase with both the mean and variance of the service time. The best example to illustrate this property is the well-known P-K formula of M/G/1 queue[9]:

$$L = \rho + \frac{\lambda^2 E[S^2]}{2(1-\rho)}. \tag{11}$$

Despite that the service times are dependent random variables, it should remain to be true that the mean queue length is monotonically increasing with both the mean and variance of the service time. Thus, the mean queue length should also increase with the variance of service rate when the capacity $\bar{\mu}$ is fixed. To confirm that this property holds , we first study the mean queue length under the two limiting cases of the variance $\sigma^2[\mu] = \bar{\mu}\mu_c$ of service rate: when $\mu_c \to 0$ and when $\mu_c \to \infty$.

3.2. **Limiting analysis of mean queue length.** For a fixed mean service rate $\bar{\mu} = \rho_s\mu_c = \frac{\lambda_s\mu_c}{\mu_s}$, the limiting case $\mu_c \to 0$, while keeping both parameters $\frac{\mu_c}{\mu_s}$ and



(a) Mean service time vs. μ_c (b) Second moment of service time vs. μ_c

Figure 4. The first and second moments of service time increase with the variance of the service rate.

λ_s constant, implies $\mu_s \to 0$ and $\rho_s \to \infty$. The physical meaning of the queueing model operated under this scenario can be interpreted as follows:

- $\rho_s \to \infty$ implies that the average number of servers is very large,
- $\mu_s \to 0$ implies that the average lifetime of a server $\frac{1}{\mu_s}$ is very long,
- $\mu_c \to 0$ implies that the capacity of each server, in terms of number of jobs served per unit of time, is very small.

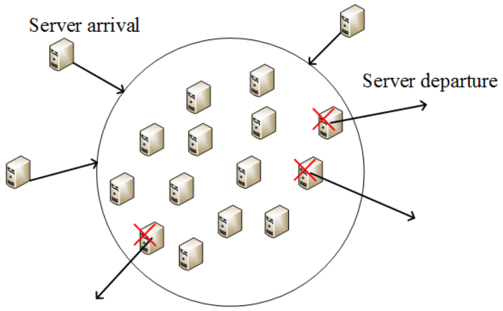


Figure 5. Service rate becomes a constant when system reaches equilibrium.

Fig. 5 illustrates this scenario, there is a large number of servers, and it seems that these servers always stay in the system because of their extremely long lifetime. However, the service rate (capacity) of each server is very small. Since the number of servers is too large to be effected by the number of arrivals or departures, the aggregate service rate $\mu = n_s\mu_c$ should converge to the constant rate $\bar{\mu} = \rho_s\mu_c$ in equilibrium. Thus, the system behaves like a single server queue with Poisson arrival rate λ_c. The mean queue length under this limiting scenario is derived in the following theorem.

Theorem 3.1. *The mean queue length L approaches to the following limit:*

$$L_1 = \lim_{\mu_c,\mu_s \to 0} L = \frac{\lambda_c}{\bar{\mu} - \lambda_c},$$

when both $\mu_c, \mu_s \to 0$ while keeping $\frac{\mu_c}{\mu_s}$ and λ_s constant.

Proof. When $\mu_c \to 0$, the standard deviation of service rate $\sigma[\mu] = \sqrt{\bar{\mu}\mu_c} \to 0$. From Chebyshev's inequality, we have

$$Pr\{|\mu - \bar{\mu}| \geq \varepsilon\} \leq \frac{E[(\mu - \bar{\mu})^2]}{\varepsilon^2} = \frac{\mu_c \bar{\mu}}{\varepsilon^2} \to 0, \quad \text{as } \mu_c \to 0, \tag{12}$$

which means the service rate $\mu = n_s \mu_c$ converges to the mean $\bar{\mu} = \rho_s \mu_c$ with probability 1. It follows that

$$-\varepsilon E[n_c] \leq E[(\mu - \bar{\mu})n_c] \leq \varepsilon E[n_c], \quad \text{as } \mu_c \to 0. \tag{13}$$

Since the stability condition implies the mean queue length is finite, i.e. $E[n_c] < \infty$ and the parameter $\varepsilon > 0$ can be arbitrarily small, therefore

$$\lim_{\mu_c, \mu_s \to 0} E[(\mu - \bar{\mu})n_c] = 0. \tag{14}$$

Hence, we have

$$\lim_{\mu_c, \mu_s \to 0} E[\mu n_c] = \bar{\mu} E[n_c], \tag{15a}$$

or equivalently,

$$\lim_{\mu_c, \mu_s \to 0} E[n_s n_c] = \rho_s E[n_c]. \tag{15b}$$

Combining equation (7) and (15b), the lower bound of queue length is given by

$$L_1 = \lim_{\mu_c, \mu_s \to 0} E[n_c] = \frac{\lambda_c}{\bar{\mu} - \lambda_c} = \frac{\rho_c}{\rho_c - \rho_s}. \tag{16}$$

□

The formula L_1 given in the above theorem is exactly the mean queue length of $M/M/1$ queue with arrival rate λ_c and service rate $\bar{\mu}$. Since the mean queue length increases with the fluctuation of the service rate, L_1becomes a lower bound of the mean queue length L that corresponds to the limiting case $\sigma[\mu] = \sqrt{\bar{\mu}\mu_c} \to 0$.

In contrast, for a fixed mean service rate $\bar{\mu} = \rho_s \mu_c = \frac{\lambda_s \mu_c}{\mu_s}$, the limiting case $\mu_c \to \infty$, while keeping both parameters $\frac{\mu_c}{\mu_s}$ and λ_s constant, implies $\mu_s \to \infty$ and $\rho_s \to 0$. The physical meaning of the queueing model operated under this scenario can be interpreted as follows:

- $\rho_s \to 0$ implies that the average number of servers is very small,
- $\mu_s \to \infty$ implies that the average lifetime of a server $\frac{1}{\mu_s}$ is very short,
- $\mu_c \to \infty$ implies that the capacity of each server, in terms of the number of jobs served per unit of time, is very large.

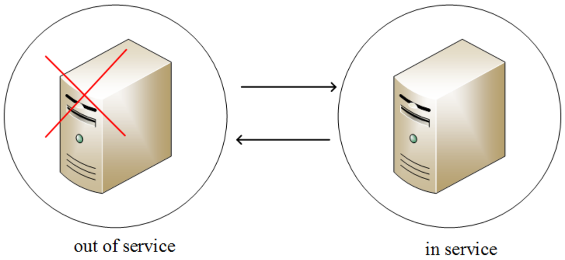


FIGURE 6. The two-state extreme scenario.

Fig. 6 illustrates this extreme scenario. The probability that there are two or more servers co-existing in the system is negligible because the lifetime of a server is so short. The only server in the system has a very large service rate (capacity) but a very short lifetime; thus, the system sometimes could be out of servers. Therefore, in this extreme scenario, our queueing model can be regarded as a queuing system with a two-state server, which is available in one state with a very large service rate μ_c but absent in the other state. The mean queue length of the system under this extreme scenario is derived in the following theorem.

Theorem 3.2. *The mean queue length L approaches to the following limit:*

$$L_2 = \lim_{\mu_c,\mu_s \to 0} L = (1 + \frac{\mu_c}{\mu_s}) \frac{\lambda_c}{\bar{\mu} - \lambda_c},$$

when both $\mu_c, \mu_s \to \infty$ while keeping $\frac{\mu_c}{\mu_s}$ and λ_s constant.

Proof. When $\mu_c \to \infty$, since the number of servers n_s follows Poisson distribution with parameter ρ_s, we have

$$Pr\{n_s = 0\} = 1 - \rho_s + o(\rho_s^2), \tag{17a}$$

$$Pr\{n_s = 1\} = \rho_s + o(\rho_s^2), \tag{17b}$$

$$Pr\{n_s \geq 2\} = o(\rho_s^2). \tag{17c}$$

Since the probability that the number of servers is larger than 1 is negligible as ρ_s approaches 0, which implies $p_{i,j} = 0$ for $j \geq 2$. Therefore, our queueing model will degenerate to a queueing model with a two-state server. The continuous time Markov chain depicted in Fig. 7 illustrates the transition diagram of the two service states.

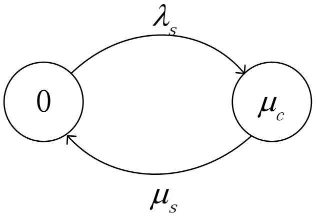


FIGURE 7. The transition diagram of the two service states.

A complete analysis of the queueing system with a two-state server is studied in [7][6]. The generating functions of the number of jobs at server state 0 and state 1 are respectively given by equation (8) in p.3533 of [7] as follows:

$$G_0(z) = \sum_{i=0}^{\infty} p_{i,0} z^i = \frac{(\frac{\lambda_s \mu_c}{\lambda_s + \mu_s} - \lambda_c)\mu_s}{z^2 \lambda_c^2 - z\lambda_c(\lambda_s + \mu_s + \lambda_c + \mu_c) + \mu_c(\lambda_s + \lambda_c)}, \tag{18a}$$

$$G_1(z) = \sum_{i=0}^{\infty} p_{i,1} z^i = -\frac{(\frac{\lambda_s \mu_c}{\lambda_s + \mu_s} - \lambda_c)(\lambda_s + \lambda_c - \lambda_c z)}{z^2 \lambda_c^2 - z\lambda_c(\lambda_s + \mu_s + \lambda_c + \mu_c) + \mu_c(\lambda_s + \lambda_c)}. \tag{18b}$$

The mean queue length L_2 can be readily obtained from the above generating functions and given as follows:

$$L_2 = \lim_{\mu_c,\mu_s \to \infty} G_0'(1) + G_1'(1).$$

We take the derivative of (18a) and (18b) and inserting $z = 1$, as the limit $\mu_c, \mu_s \to \infty$ the mean queue length is given by:

$$\begin{aligned} L_2 &= \lim_{\mu_c, \mu_s \to \infty} G_0'(1) + G_1'(1) \\ &= \frac{\lambda_c(\mu_s + \mu_c)}{\mu_c \lambda_s - \lambda_c \mu_s} \\ &= (1 + \frac{\mu_c}{\mu_s}) \frac{\lambda_c}{\bar{\mu} - \lambda_c}. \end{aligned} \tag{19}$$

□

Recall that the system capacity is given by $\bar{\mu} = \rho_s \mu_c = \lambda_s(\mu_c/\mu_s)$. Fig. 8 plots the mean queue length versus $\frac{\rho_c}{\rho_s}$ for a given $\frac{\mu_c}{\mu_s} = 10$ and $\lambda_s = 10$. For any value of $\frac{\rho_c}{\rho_s}$ on the x-axis in Fig. 8, the mean queue length is monotonically increasing with the variance of service rate $\sigma^2[\mu] = \bar{\mu}\mu_c$, which perfectly agrees with our assertion stated at the end of Section 3.1. Therefore, the two limiting cases, L_1 and L_2, given in the above two theorems are respectively the lower bound and upper bound of the mean queue length.

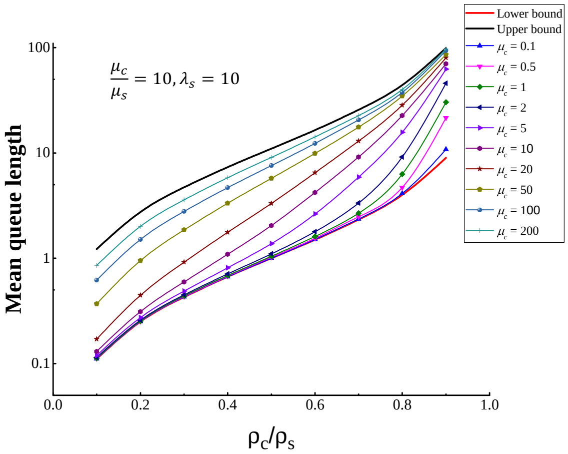


FIGURE 8. Mean queue length L is bounded by L_1 and L_2.

From Theorem 3.1 and Theorem 3.2, we know that the mean queue length L is bounded by L_1 and L_2 when λ_s and $\frac{\mu_c}{\mu_s}$ are fixed, where the formula L_1 is exactly the mean queue length of $M/M/1$ queue with arrival rate λ_c and service rate $\bar{\mu}$ and the formula $L_2 = (1 + \frac{\mu_c}{\mu_s})L_1$. Given that λ_s and $\frac{\mu_c}{\mu_s}$ are fixed, since inequality (10) is the necessary and sufficient condition of this $M/M/1$ queue, the same condition guarantees that the mean queue length L is finite and the system is stable.

4. **An approximation of mean queue length.** This section derives an approximate formula of the mean queue length for our queueing model. The derivation of the two bounds in the previous section indicates that the mean queue length is dependent on the fluctuation of the service rate. In traditional queueing analysis with constant service rate, we know that if the service rate is greater than the job

arrival rate, then the queue length is always finite and the system is stable. On the other hand, if the service rate is smaller than the job arrival rate, then the queue length rapidly grows to infinity and the system becomes unstable.

In our queueing model, however, the total service rate varies with time. That means sometimes the service rate is greater than the job arrival rate, and other times it is smaller than the job arrival rate. From last section, the mean queue length achieves the lower bound when the service rate converges to a constant. In this case, this constant service rate must be larger than the job arrival rate to make the system stable.

As the service rate fluctuates, if the probability that the service rate is smaller than the job arrival rate becomes larger, then the system would perform worse than the lower bound. Thus, the fraction of time that the service rate is smaller than the job arrival rate determines how much the system would perform worse than the lower bound. In this section, from the two bounds L_1 and L_2, we derive the following approximation of the mean queue length of our queueing model:

$$L = (1 + \frac{\mu_c}{\mu_s}\alpha)\frac{\lambda_c}{\bar{\mu} - \lambda_c}, \tag{20}$$

for some parameter $0 \leq \alpha \leq 1$.

4.1. **Mean queue length formula.** We first investigate the service rate fluctuation in time interval $[0, T]$. Since the number of servers $n_s(t)$ at time $t \in [0, T]$ may fluctuate around the mean ρ_s as Fig. 9 shows. Thus the time interval can be divided into two regions:

$$R_{underload} = \{t | \mu(t) > \lambda_c, t \in [0, T]\}, \tag{21a}$$

and

$$R_{overload} = \{t | \mu(t) \leq \lambda_c, t \in [0, T]\}. \tag{21b}$$

And the total amount of time can be noted as $T_{underload}$ and $T_{overload}$ respectively.

The queue length is always finite in the underload region $T_{underload}$, because the service rate $\mu(t)$ exceeds the job arrival rate λ_c. In the overload region $T_{overload}$, however, the queue length may grow rapidly when the service rate $\mu(t)$ is lower than the job arrival rate λ_c. In steady state, we have

$$Pr\{\mu > \lambda_c\} = \lim_{T \to \infty} \frac{T_{underload}}{T}, \tag{22a}$$

and

$$Pr\{\mu \leq \lambda_c\} = \lim_{T \to \infty} \frac{T_{overload}}{T}. \tag{22b}$$

We define the following notations used in the derivation of the mean queue length L:

- $L_{underload} = E[n_c | \mu > \lambda_c]$: conditional mean queue length of underload region,
- $L_{overload} = E[n_c | \mu \leq \lambda_c]$: conditional mean queue length of overload region,
- $a = Pr\{\mu \leq \lambda_c\}$: overload probability, and thus the underload probability equals $1 - a$,
- $b = Pr\{\mu \leq \bar{\mu}\}$: probability that service rate μ is smaller than service capacity $\bar{\mu}$.

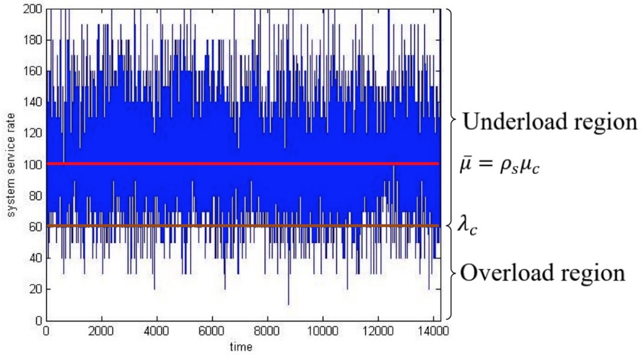


FIGURE 9. Overload and underload regions.

The mean queue length can be expressed as the combination of the conditional mean queue lengths as follows:

$$L = E[n_c] = (1-a)L_{underload} + aL_{overload}. \tag{23}$$

The above expression cannot help us to evaluate the mean queue length because the two conditional mean values $L_{underload}$ and $L_{overload}$ are unknown. Since we know that the mean queue length L is bounded by L_1 and L_2, therefore L can be expressed as follows:

$$\begin{aligned} L &= (1-\alpha)L_1 + \alpha L_2 \\ &= (1+\frac{\mu_c}{\mu_s}\alpha)\frac{\lambda_c}{\bar{\mu}-\lambda_c}, \end{aligned} \tag{24}$$

for some parameter $0 \le \alpha \le 1$. For a proper chosen parameter α, the linear combination (24) of L_1 and L_2, which is similar to expression (23), can serve as a good approximation of the mean queue length L.

Intuitively, the overload probability a is a measurement that indicates how much the system performance is worse than the lower bound. When the overload probability a is small, then the mean queue length L should be close to the lower bound L_1. In particular, when $a \to 0$, the parameter α should also approach to 0. On the other hand, for larger overload probability a, the parameter α should also be larger, indicating that the mean queue length L is closer to the upper bound L_2. Therefore, a proper choice of the parameter α is linearly proportional to the overload probability a:

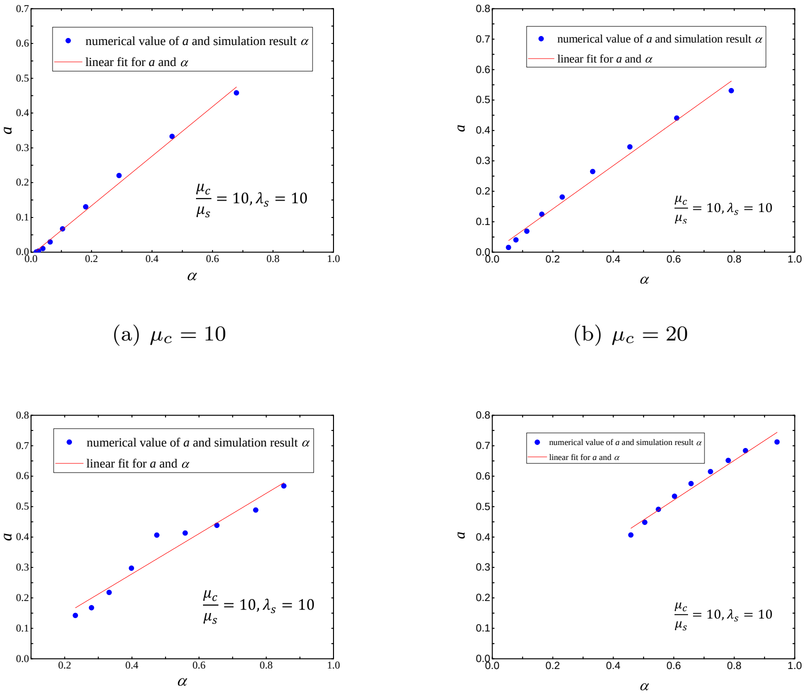
$$\hat{\alpha} = ka. \tag{25}$$

When the system becomes saturated as λ_c approaches $\bar{\mu}$, or equivalently ρ_c approaches ρ_s, we expect that the expression (24) of mean queue length L will reach the upper bound L_2. That is, the proportional constant k can be determined by the following limiting condition:

$$\lim_{\lambda_c \to \bar{\mu}} \hat{\alpha} = k \lim_{\lambda_c \to \bar{\mu}} Pr\{\mu \le \lambda_c\} = kPr\{\mu \le \bar{\mu}\} = 1, \tag{26}$$

thus, we have

$$k = \frac{1}{Pr\{\mu \le \bar{\mu}\}} > 1. \tag{27}$$



(a) $\mu_c = 10$ (b) $\mu_c = 20$

(c) $\mu_c = 50$ (d) $\mu_c = 100$

FIGURE 10. Overload probability a vs. parameter α.

It follows from (25) and (27) that

$$\hat{\alpha} = \frac{a}{b} = \frac{Pr\{\mu \leq \lambda_c\}}{Pr\{\mu \leq \bar{\mu}\}} = \frac{Pr\{n_s \leq \rho_c\}}{Pr\{n_s \leq \rho_s\}}. \tag{28}$$

The above choice of the parameter α satisfies the required condition (26), and the following mean queue length formula is readily obtained from (24):

$$L \cong (1 + \frac{\mu_c}{\mu_s}\frac{a}{b})\frac{\lambda_c}{\bar{\mu} - \lambda_c}, \tag{29}$$

where

$$a = Pr\{\mu \leq \lambda_c\} = \sum_{i=0}^{\rho_c} \frac{\rho_s^i}{i!} e^{-\rho_s} \triangleq \sum_{i=0}^{\lfloor \rho_c \rfloor} \frac{\rho_s^i}{i!} e^{-\rho_s} + (\rho_c - \lfloor \rho_c \rfloor) \frac{\rho_s^{\rho_s}}{\Gamma(\rho_s)} e^{-\rho_s}, \tag{30a}$$

and

$$b = Pr\{\mu \leq \bar{\mu}\} = \sum_{i=0}^{\rho_s} \frac{\rho_s^i}{i!} e^{-\rho_s} \triangleq \sum_{i=0}^{\lfloor \rho_s \rfloor} \frac{\rho_s^i}{i!} e^{-\rho_s} + (\rho_s - \lfloor \rho_s \rfloor) \frac{\rho_s^{\rho_s}}{\Gamma(\rho_s)} e^{-\rho_s}. \tag{30b}$$

Extensive simulations verify that the parameter α in (28) given by $\alpha = \frac{L - L_1}{L_2 - L_1}$ is indeed linearly proportional to the overload probability a, as Fig. 10 shows.

Similar to the convex combination (24) of the mean queue length, the two conditional mean queue lengths $L_{underload}$ and $L_{overload}$ can be respectively expressed as follows:

$$L_{underload} = (1 - \alpha_1)L_1 + \alpha_1 L_2, \tag{31}$$

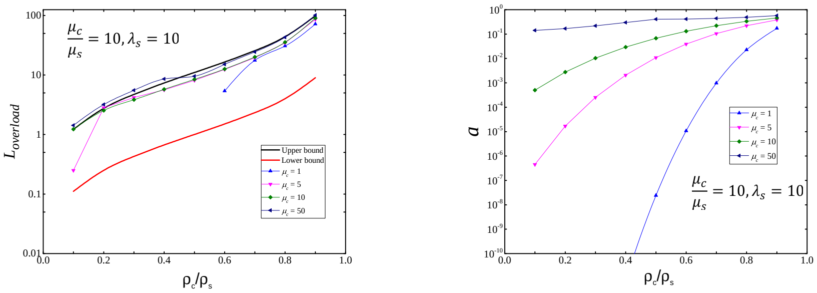
and

$$L_{overload} = (1 - \alpha_2)L_1 + \alpha_2 L_2. \tag{32}$$

When the system is in the overload region $T_{overload} = \{t|\mu(t) \leq \lambda_c, t \in [0, T]\}$, we expect that the mean queue length will quickly reach the upper bound L_2 because the loading is larger than the capacity of the system. In fact, our simulation results show that the following relation indeed holds in most cases:

$$L_{overload} \cong L_2, \tag{33}$$

that is, an appropriate choice of α_2 in expression (32) is $\alpha_2 = 1$. This approximation (33), however, may perform poorly when both unit service rate μ_c and job arrival rate λ_c are small. As the simulation results displayed in Fig. 11 demonstrate, the mean queue length in the overload region $L_{overload}$ is far below the upper bound L_2 when $\mu_c = 1$ and $\rho_c/\rho_s < 0.6$ or $\mu_c = 5$ and $\rho_c/\rho_s < 0.2$. However, it should be noticed that the system rarely comes across to the overload region under this condition. In fact, our simulation could not even collect any data in the overload region if $L_{overload}$ is much less than L_2. The reason is that the system does not stay in the overload region long enough to reach the upper bound L_2 if the fraction of overload time $T_{overload}$ is too small compared to that of $T_{underload}$. Therefore, the discrepancy between $L_{overload}$ and L_2 is negligible in most cases, and it becomes significant only when the overload probability $a = Pr\{\mu \leq \lambda_c\}$, defined by (22b), is very small.



(a) $L_{overload}$

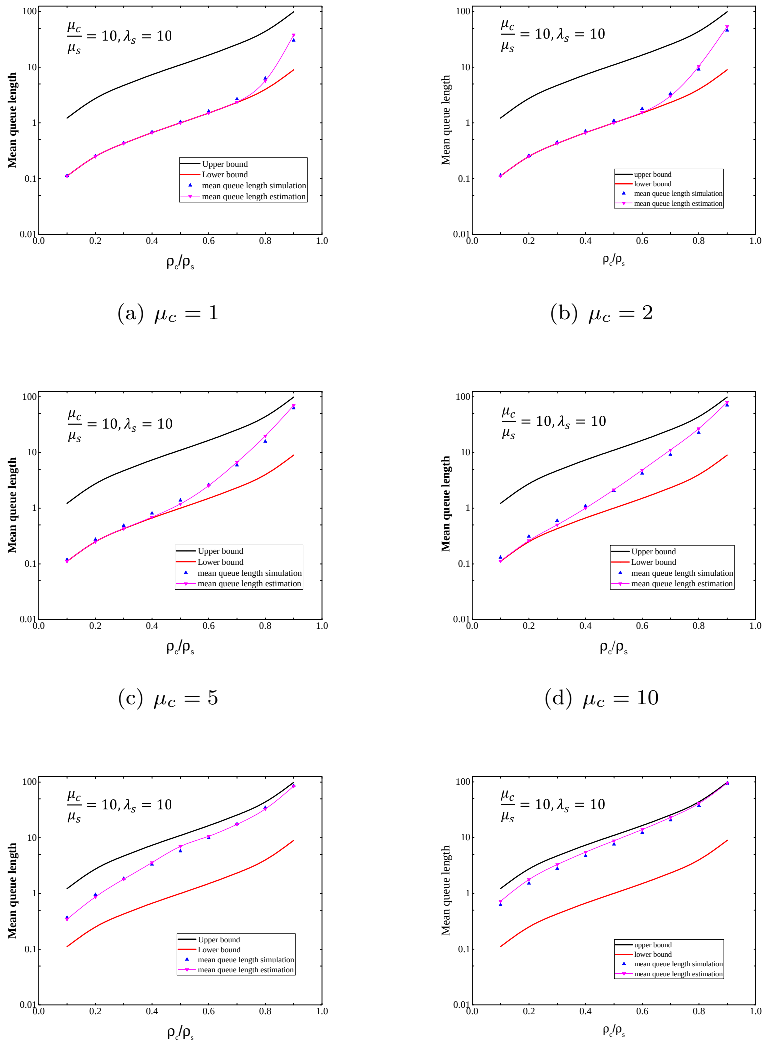
(b) Overload probability a

FIGURE 11. Mean queue length in overload region $L_{overload}$ and overload probability a.

Moreover, substituting (33) into (23) and combining the result with (29), we obtain the following approximation of the mean queue length in underload region:

$$L_{underload} \cong (1 + \frac{\mu_c}{\mu_s}\alpha_1)\frac{\lambda_c}{\bar{\mu} - \lambda_c} = (1 + \frac{\mu_c}{\mu_s}\frac{a(1-b)}{b(1-a)})\frac{\lambda_c}{\bar{\mu} - \lambda_c}. \tag{34}$$

For $\alpha = ka > a$, the parameter $\alpha_1 = \frac{\alpha - a}{1 - a}$ also satisfies the condition $0 < \alpha_1 < 1$.



(a) $\mu_c = 1$

(b) $\mu_c = 2$

(c) $\mu_c = 5$

(d) $\mu_c = 10$

(e) $\mu_c = 50$

(f) $\mu_c = 100$

FIGURE 12. Simulation and approximation results of mean queue lengths.

The robustness of the mean queue length formula (29) is demonstrated by simulation results displayed in Fig. 12, in which $\frac{\rho_c}{\rho_s} = \frac{\lambda_c \mu_s}{\lambda_s \mu_c}$ changes from 0.1 to 0.9 with a fixed parameter $\frac{\mu_c}{\mu_s} = 10$, $\lambda_s = 10$, while μ_c varies from 1 to 100 and λ_c changes from 10 to 90. The mean queue lengths estimated by formula (29) agree with the simulation results in all cases we considered, even when $L_{overload}$ is far below the upper bound, because the weight of $L_{overload}$ in terms of overload probability a in this case is extremely small, less than 10^{-4}, which helps to eliminate the discrepancy. For example, Fig. 11 shows that the $L_{overload}$ is much smaller than the upper bound L_2 when $\frac{\rho_c}{\rho_s}$ is below 0.6, and $\mu_c = 1$. However, as Fig. 12(a) shows, with the same set of parameters, the overall mean queue length L estimated by formula (29) still fits very well with the simulation result.

4.2. **Comparison with P-K formula of M/G/1 queue.** The expression (29) exhibits a strong similarity to the classical P-K formula of M/G/1 queues (11), which can be rewritten as follows:

$$L = \rho + \frac{\rho^2(1+C_b^2)}{2(1-\rho)} = [1 + \frac{\rho}{2}(C_b^2 - 1)]\frac{\rho}{1-\rho} \tag{35}$$

where $C_b^2 = Var[S]/E^2[S]$ is the coefficient of variation of the service time. The term $\frac{\rho}{1-\rho}$ is the mean queue length of M/M/1 queue with the same mean service rate, while the term $\frac{\rho}{2}(C_b^2 - 1)$ is an indicator of the fluctuation of the service time. For a fixed offered load ρ, the fluctuation of the service time and thus the mean queue length increases with C_b^2. As an example, C_b^2 of an M/D/1 queue is 0, while that of an M/M/1 queue is 1. Therefore, the mean queue length of an M/D/1 queue is less than that of an M/M/1 queue [9].

The mean queue length formula (29) of our queueing model and the P-K formula for M/G/1 queue (35) possess similar expressions. Recall that we derived the two bounds L_1 and L_2 of mean queue length L in Section 3 by considering the service rate fluctuations for a fixed service capacity $\bar{\mu} = \rho_s\mu_c = \frac{\lambda_s\mu_c}{\mu_s}$. The degree of service rate fluctuation is dependent on the average number of servers $\rho_s = E[n_s]$ and the capacity μ_c of each server. As we discussed in Section 3, the two bounds of queue length respectively correspond to the following two limiting cases of service rate fluctuations:

1. The lower bound L_1 of queue length corresponds to the minimum service rate fluctuation, when the average number of servers is very large, as $\rho_s \to \infty$, and the capacity of each server is very small, as $\mu_c \to 0$. In this case, the arrival or departure of any particular server almost does not affect the service rate, as Fig. 5 shows.
2. The upper bound L_2 of queue length corresponds to the maximum service rate fluctuation, when the average number of servers is very small, as $\rho_s \to 0$, and the capacity of each server is very large, as $\mu_c \to \infty$. In this case, the system has only one server at the best, and the arrival or departure of this server severely affects the service rate, as Fig. 6 shows.

Since the parameter α is defined in terms of the number of servers n_s in (28), the factor $\frac{\mu_c}{\mu_s}\alpha$ is served as an indicator of the service time fluctuation in the mean queue length formula (29) that can be rewritten as follows:

$$L = (1 + \frac{\mu_c}{\mu_s}\alpha)\frac{\lambda_c}{\bar{\mu} - \lambda_c} \cong (1 + \frac{\mu_c}{\mu_s}\frac{Pr\{n_s \le \rho_c\}}{Pr\{n_s \le \rho_s\}})\frac{\lambda_c}{\bar{\mu} - \lambda_c} \tag{36}$$

In the above expression, the factor $\frac{\mu_c}{\mu_s}\alpha$ is similar to that of the coefficient of variation of the service time C_b^2 in the P-K formula (35). However, due to the strong dependency among the service times of different jobs, the impact of the fluctuation of service time on the mean queue length of our queueing model is measured by the entire distribution of the service rate, instead of the first two moments of the service time in the P-K formula of M/G/1 queue.

4.3. **Verifications of limiting cases.** In this subsection, we verify all limiting cases of the system behavior. The results confirm that all limiting cases are consistent with the mean queue length formula (29).

1. Lower bound

In the limiting case $\mu_c \to 0$ for a fixed service capacity $\bar{\mu}$, the service rate becomes a constant, that is $\mu(t) = \bar{\mu}$ for all t. Since the stability condition $\rho_c < \rho_s$ holds if and only if $\lambda_c \in [0, \bar{\mu})$, we therefore must have $Pr\{\mu \leq \lambda_c\} = 0$. In this case $\alpha = \frac{Pr\{\mu \leq \lambda_c\}}{Pr\{\mu \leq \bar{\mu}\}} = 0$, then the mean queue length reduces to $L_1 = \frac{\lambda_c}{\bar{\mu} - \lambda_c}$.

2. Upper bound

In the limiting case $\mu_c \to \infty$ for a fixed service capacity $\bar{\mu}$, as Fig. 7 shows, the server arrival process can be described as a two-state Markov chain. The number of servers in the system is either $n_s = 0$ or $n_s = 1$, and the corresponding service rate is either $\mu = 0$ or $\mu = \mu_c$, while the service capacity $\bar{\mu} = \mu_c Pr\{n_s = 1\} < \mu_c$. Since the stability condition (10) implies $\lambda_c \in [0, \bar{\mu})$, then we must have

$$Pr\{\mu \leq \lambda_c\} = Pr\{\mu \leq \bar{\mu}\} = Pr\{n_s = 0\}. \tag{37}$$

It follows that $\alpha = \frac{Pr\{\mu \leq \lambda_c\}}{Pr\{\mu \leq \bar{\mu}\}} = 1$, in which case (29) reduces to the upper bound $L_2 = (1 + \frac{\mu_c}{\mu_s}\alpha)\frac{\lambda_c}{\bar{\mu} - \lambda_c}$.

3. $\lambda_c \to 0$

When the job arrival rate λ_c is very small, the overload probability can be interpreted as

$$\lim_{\lambda_c \to 0} a = Pr\{n_s = 0\} = e^{-\rho_s}.$$

There are two subcases of interest:

- If μ_s is relatively small, then the average number of servers ρ_s is very large, and we have
$$\alpha = \frac{a}{b} = \frac{e^{-\rho_s}}{Pr\{\mu \leq \bar{\mu}\}} \to \frac{0}{1} = 0, \tag{38}$$
which implies that the mean queue length approaches the lower bound L_1.
- When $\mu_s \to \infty$, then the average number of servers ρ_s is very small, we have
$$\alpha = \frac{a}{b} = \frac{e^{-\rho_s}}{Pr\{\mu \leq \bar{\mu}\}} \to \frac{e^{-\rho_s}}{e^{-\rho_s}} = 1, \tag{39}$$
which implies that the mean queue length approaches the upper bound L_2.

Thus, when the offer load is small, the system performance is mainly determined by the server process. This analysis generally agrees with the simulation result shown in Fig. 8.

4. $\lambda_c \to \bar{\mu}$

In this case, it is obvious that $\alpha = \frac{Pr\{\mu \leq \lambda_c\}}{Pr\{\mu \leq \bar{\mu}\}} \to 1$, which implies that the mean queue length of the system reaches the upper bound when the offer load is close to saturation. In Fig. 8, the simulation results also show that the mean queue length approaches the upper bound when $\frac{\rho_c}{\rho_s} \to 1$.

5. Conclusion.

In this paper, we propose a new kind of varying service rate queueing model in which the number of service rate may be infinite. However, it is mathematically intractable to obtain a closed-form solution of the steady-state probabilities of these Markov chains. Inspired by the P-K formula for M/G/1 queue, we show that the mean queue length of our model is significantly influenced by the variance of service rate. And we consider the two limiting cases as the bounds of mean queue length.

Furthermore, we provide a simple formula to estimate the mean queue length. Extensive simulation studies with different parameters fully verify the accuracy of our approximation, and all limiting cases of the system behavior we checked completely agree with the predictions made by our formula. The similarity between our approximation and the P-K formula for M/G/1 queue strongly supports our approach. Thus, this formula could serve as a useful tool in the study of the performance of varying service rate queueing model.

Using the queueing system with varying service rate to model complicated real network applications is intrinsically difficult. Except in some simple special cases, most of them are not solvable by using traditional queueing analysis. We expect that the approach developed in this paper will shed some light on the queueing model with variable service rate, and the proposed methodology can be extended and applied to many other research fields, such as mobile cloud computing or energy efficient Ethernet.

REFERENCES

[1] D. P. Anderson, BOINC: A system for public-resource computing and storage, *In Proceedings of the Workshop on Grid Computing*, (2004).

[2] D. P. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer, SETI@home: An experiment in public resource computing, *Communications of the ACM*, 45 (2002), 56-61.

[3] M. Eisen and M. Tainiter, Stochastic variations in queuing processes, *Operations Res.*, 11 (1963), 922–927.

[4] T. Estrada, M. Taufer and D. P. Anderson, Performance prediction and analysis of BOINC projects: An empirical study with EmBOINC, *Journal of Grid Computing*, 7 (2009), 537–554.

[5] B. Fan, D. Chiu and J. Lui, The Delicate Tradeoffs in BitTorrent-like File Sharing Protocol Design, *In Proceedings of the 2006 IEEE International Conference on Network Protocols*, (2006), 239–248.

[6] N. Gunaseelan, L. Liu, J. F. Chamberland and G. H. Huff, Performance analysis of wireless Hybrid-ARQ systems with delay-sensitive traffic, *IEEE Transactions on Communications*, 58 (2010), 1262–1272.

[7] L. Huang and T. T. Lee, Generalized Pollaczek-Khinchin formula for Markov channels, *IEEE Transactions on Communications*, 61 (2013), 3530–3540.

[8] F. P. Kelly, *Reversibility and stochastic networks*, Cambridge University Press, 2011.

[9] L. Kleinrock, *Queueing Systems, Volume 1, Theory*, John Wiley & Sons, New York, 1975.

[10] R. Kumar, Y. Liu and K. Ross, Stochastic Fluid Theory for P2P Streaming Systems, *In IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, (2007), 919–927.

[11] H. Li and T. Yang, Queues with a variable number of servers, *European J. Oper. Res.*, 124 (2000), 615–628.

[12] S. R. Mahabhashyam and N. Gautam, On queues with Markov modulated service rates, *Queueing Syst.*, 51 (2005), 89–113.

[13] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, John Hopkins Series in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 1981.

[14] T. Phung-Duc, W. Rogiest and S. Wittevrongel, Single server retrial queues with speed scaling: Analysis and performance evaluation, *J. Ind. Manag. Optim.*, 13 (2017), 1927–1943.

[15] B. A. Salihu, P. Li, L. Sang, Z. Li, Y. Gao and D. Yang, Network calculus delay bounds in multi-server queueing networks with stochastic arrivals and stochastic services, *In Global Communications Conference (GLOBECOM)*, IEEE (2015), 1–7.

[16] M. Yajima, and T. Phung-Duc, Batch arrival single server queue with variable service speed and setup time, *Queueing Syst.*, 86 (2017), 241–260.

[17] M. Yajima and T. Phung-Duc, A central limit theorem for a Markov-modulated infinite-server queue with batch Poisson arrivals and binomial catastrophes, *Performance Evaluation*, 129 (2019), 2–14.

[18] J. Zhang, Z. Zhou, T. T. Lee and T. Ye, Delay analysis of three-state Markov channels, *in Lecture Notes of Computer Science*, **10591** (2017), 101–117.

[19] J. Zheng, C. Luo and L. Yu, Performance analysis of stochastic multi server systems, *In Communications and Networking in China (ChinaCom), 2015 10th International Conference on*, IEEE (2015), 562–566.

[20] *BOINCstats*, Available from: http://boincstats.com.

Received October 2018; 1st revision March 2019; 2nd revision May 2019.

E-mail address: zhangjian19921022@163.com
E-mail address: tonylee@cuhk.edu.cn
E-mail address: yetong@sjtu.edu.cn
E-mail address: lianghuang@zjtu.edu.cn