

STATIONARY ANALYSIS OF INFINITE QUEUEING SYSTEM WITH TWO-STAGE NETWORK SERVER

R. SUDHESH¹ AND A. VAITHIYANATHAN^{2,*}

Abstract. As the world's major economies and technologies have matured, they are dominated by service-focused approach leading to study and analysis of service models for improved understanding and efficiency. Research in this direction has been done on various parameters of the finite queues using different approaches. The study discussed in this paper deals with the stationary behavior of two-stage queueing system with infinite capacity where any arriving customer is serviced in two stages in a mutually exclusion fashion. The steady state system size probabilities for the infinite capacity queueing system with two stages of service are obtained in recursive form. Further, numerical interpretations are presented to depict the system behavior for values of the parameters.

Mathematics Subject Classification. 60k25.

Received December 18, 2019. Accepted July 4, 2020.

1. INTRODUCTION

Extensive study of the world's best service companies which aims to create a win-win-win situation reveals that providing services in successive stages is a common feature. The service ranges from marketing to management, travel to telecommunication and any day-to-day services. A real-life example of two stage service is the purchase process in supermarkets, where the first stage is the self-service of customers collecting their items and the second stage is the payment at the checkout. Moreover, vegetable packing process in a company undergoes two stages for exporting a special fruit. Firstly, the product is sent for quality checking. Then, if the quality of the product is desirable, it is further sent for the second stage of processing. Haoyu *et al.* [8] presented a two-stage service replica strategy to improve the execution efficiency of the business process by shortening the response time of request to single service and reducing interaction time among distributed services.

Information technology adopts such services in multiple locations for multiple functions. The usefulness and application of such a system can be numerous in software architectures [1], communication system and network servers. Analysis of queueing system of a router with two stages of service represents typical network servers where network packets arrive and get queued to be served sequentially in two stages in a mutually exclusion fashion. That is arrivals will be served by the first stage followed by a second stage, with only one stage being

Keywords. Generating function, infinite queueing system, discrete time, two-stage service.

¹ Department of Mathematics, Bharathidasan Institute of Technology (BIT) Campus, Anna University, Tiruchirappalli 620024, India.

² Department of Mathematics, IFET College of Engineering, Villupuram 605108, India.

*Corresponding author: vaithi05@gmail.com

active at a time. Another example is the dynamism of network servers and client machines where the first stage is processing the arriving packets at the kernel level and the second stage processing web requests. Any existing network cloud is a network of routers which aids in directing the packets to the appropriate destination. The primary work of the router is the forwarding of packets. The major event in this forwarding process is the transfer of datagrams from a router's incoming link through the input port to an outgoing link through the output port, based on the entries in the routing table. It performs the routing function associated with the network layer and the switching function associated with the data link layer. Each port has a buffer which constitutes a set of queues which stores the incoming packets and outgoing packets of the port.

The input port performs the physical layer functionality of terminating an incoming physical link to a router and the data link layer functionality needed to interoperate with the data link layer functionality on the other side of the incoming link. It also performs a lookup and forwarding function so that a datagram forwarded into the switching fabric of the router emerges at the appropriate output port. Control packets or packets carrying routing protocol information are forwarded from the input port to the routers processor. This presents a scenario wherein multiple ports along with their input and output queues are gathered together on a single line card within a router. This can be modeled as a twin queuing system with a two-stage service, where the services are in a mutually exclusive fashion. The first service denotes the arrival of the packet and opening the packet to read the destination address. Forwarding the packet to the appropriate output port by comparing the destination address with the entries in the routing table constitutes the second service. An infinite queue or infinite buffer eliminates the loss of packets and ensures that the data reaches the destination intact.

In discrete time queue, Yang *et al.* [23] discussed the behavior of customers in the *Geo/Geo/1* queue with server breakdowns and repairs. Moreover, the stationary distribution and mean sojourn time of an arriving customer of the model derived. Equilibrium behavior of Markovian queues with vacation strategy has received considerable attention [12,22]. Krishna and Lee [11] discussed the study of the control of a server which alternates between two phases, *i.e.*, individual and batch phases. Yang *et al.* [24] presented the steady state system size distribution for the two-stage queue with a single server and N -policy by using matrix geometric approach. In [3] and [10], authors presented a two-stage state space approach to solve the state probabilities and sojourn time analysis for $M/G/1$ queueing model. In [5], the time-dependent system size probabilities are derived when the service is accomplished through two parallel channels with different service rate. Yu [25] discussed a model under consideration can be viewed as an $M/M/1$ queue in a random environment. Equilibrium mixed strategies are derived for the almost unobservable and fully unobservable queues. Zarrinpoor and Saber [26], formulated a model based on a two-stage robust optimization approach in which decisions are defined in two-stages such that a first-stage solution should be robust against the possible realizations of the disruption that can only be revealed in a second stage.

Gopalan and Kannan [7] dealt with the analysis of transient behaviour of a two-server queueing network subject to interstage inspection and rework. The authors have developed a stochastic model and obtained the explicit expressions for some of the system characteristic using state-space method and regeneration point technique. In [17, 18], Vedat Saglam *et al.* analyzed tandem queueing model with two stages and derived the transient probability and loss probabilities of this model. Shin and Moon [13], provided an approximation technique for multi-server two-stage networks with splits and blocking. Many countries are suffering severe flood problems due to inadequate investment in their drainage system. Xinhua He and Wenfa Hu [9], investigated distribution feature of rainstorms and draining process and used a two-stage single queue method to model urban drainage system. Zhou *et al.* [28], developed a two stage tandem queueing network and introduced a Markov arrival process to characterize the correlation of the demand in the business sector. In [6], Giorno *et al.* performed the asymptotic analysis for the single server queueing model and derived the transient probabilities for this model by considering constant arrival rates and state dependent service rates. Sudhesh [20] discussed the transient analysis for the model "Single server with system disaster and customer impatience" with the help of generating functions and continued fractions. For a detailed list of references for two-stage queueing models and their application to various sectors, see review papers from [2, 4, 15, 16, 21, 27].

An excellent study on the analysis of a finite queueing system with two-stages of service had been presented by Salah [19]. To analyze the behavior of such system, presented two equivalent analytical models in which derived equations for key features such as throughput, packet loss, packet delay and server CPU utilization and some important performance measures. In [14], Patil and Turck presented a two-queue Markov model for a wireless sensor node. The introduction of “Value of Information” (VOI) and its non-additive data gathering process plays an important role in decision making. The authors formulated the problem as Markov decision process and found the exact solution by policy iteration.

This paper aims to study the equilibrium behavior of the two-stage network server. We obtain a closed form steady state system size probabilities for the infinite capacity single server queueing system with two stages of service. In this study, the underlying Kolmogorov difference equations are first transformed into a set of linear algebraic equations by employing partial generating functions. Using the normalization equation, we derived the steady-state probabilities in recursive form. Numerical illustrations are provided to understand the system behavior.

2. MODEL DESCRIPTION

In this model, the arriving customer gets service sequentially in two stages in a mutually exclusion fashion. Consider the state space $S = \{(m, k), k \geq 0, m \in \{0, 1\}\}$, where k represents the number of packets in the system and m denotes the type of stage the CPU is performing. The queueing system has infinite buffer size. Initially, the system is empty at the state $(0, 0)$. The states $(1, k)$ and $(0, k)$ denotes the states where the CPU is busy handling stage 1 and stage 2 respectively. Consider the tasks carried out at stage 2 which include network layer functionalities of examining the packet headers for errors, looking up forwarding tables, and then forwarding the packet. On the other hand, stage 1 is responsible for merely data link layer functionalities and notifying the arrival of a new packet. Arrival occurs according to geometric distribution with parameter λ and the service completions occurs according to geometric distribution with parameter μ_1 and μ_2 for the state $(1, k)$ and $(0, k)$ respectively. For network servers and hosts, $\mu_1 < \mu_2$ in practice, since the service time of stage 1 $(1, k)$ is on average shorter than that of stage 2 $(0, k)$ as the amount of processing at stage 2 is more involved. we assume, initially that the system is empty at the state $(0, 0)$. Denote $1 - \lambda$, $1 - \mu_1$ and $1 - \mu_2$ as $\bar{\lambda}$, $\bar{\mu}_1$ and $\bar{\mu}_2$ respectively. The state transition rate diagram is depicted in Figure 1.

The one step transition probabilities of the states are as follows:

Case 1: if $X_n = (0, k), k = 2, 3, 4, \dots$,

$$X_{n+1} = \begin{cases} (0, k) & \text{with probability } \bar{\lambda}\bar{\mu}_2 + \lambda\mu_2 \\ (0, k + 1) & \text{with probability } \lambda\bar{\mu}_2 \\ (1, k - 1) & \text{with probability } \bar{\lambda}\mu_2. \end{cases}$$

Case 2: if $X_n = (0, 1)$

$$X_{n+1} = \begin{cases} (0, 1) & \text{with probability } \bar{\lambda}\bar{\mu}_2 + \lambda\mu_2 \\ (0, 2) & \text{with probability } \lambda\bar{\mu}_2 \\ (0, 0) & \text{with probability } \bar{\lambda}\mu_2. \end{cases}$$

Case 3: if $X_n = (0, 0)$

$$X_{n+1} = \begin{cases} (0, 0) & \text{with probability } \bar{\lambda} \\ (1, 1) & \text{with probability } \lambda. \end{cases}$$

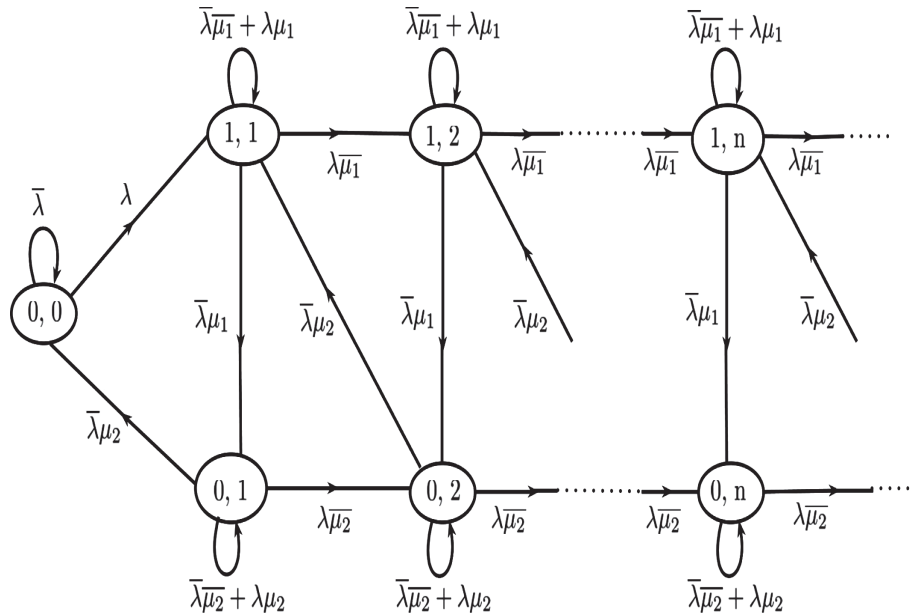


FIGURE 1. State transition diagram for an infinite queue with two stages of service.

Case 4: if $X_n = (1, k), k = 1, 2, 3, \dots,$

$$X_{n+1} = \begin{cases} (1, k) & \text{with probability } \bar{\lambda}\bar{\mu}_1 + \lambda\mu_1 \\ (1, k + 1) & \text{with probability } \lambda\bar{\mu}_1 \\ (0, k) & \text{with probability } \bar{\lambda}\mu_1. \end{cases}$$

Let $\pi_{j,k}$ be the steady state probability of the server in the state j and there are k customers in the system, where $j = 0, 1$ and $k = 1, 2, \dots$ and $\pi_{0,0}$ be the empty system size probability.

A system of difference equations can be written as follows

$$\pi_{0,0} = \bar{\lambda}\mu_2\pi_{0,1} + \bar{\lambda}\pi_{0,0}, \tag{2.1}$$

$$\pi_{1,1} = \lambda\pi_{0,0} + \bar{\lambda}\mu_2\pi_{0,2} + [\bar{\mu}_1\bar{\lambda} + \lambda\mu_1] \pi_{1,1}, \tag{2.2}$$

$$\pi_{1,k} = \lambda\bar{\mu}_1\pi_{1,k-1} + \bar{\lambda}\mu_2\pi_{0,k+1} + [\bar{\mu}_1\bar{\lambda} + \lambda\mu_1] \pi_{1,k}, \quad k \geq 2 \tag{2.3}$$

$$\pi_{0,1} = \bar{\lambda}\mu_1\pi_{1,1} + [\bar{\mu}_2\bar{\lambda} + \lambda\mu_2] \pi_{0,1} \tag{2.4}$$

$$\pi_{0,k} = \bar{\lambda}\mu_1\pi_{1,k} + \lambda\bar{\mu}_2\pi_{0,k-1} + [\bar{\mu}_2\bar{\lambda} + \lambda\mu_2] \pi_{0,k}, \quad k \geq 2 \tag{2.5}$$

since π is a stationary distribution

$$\sum_{k=0}^{\infty} \pi_{0,k} + \sum_{k=1}^{\infty} \pi_{1,k} = 1. \tag{2.6}$$

Define the partial stationary probability generating function of the system as

$$G_0(z) = \sum_{k=0}^{\infty} \pi_{0,k}z^k, \quad G_1(z) = \sum_{k=1}^{\infty} \pi_{1,k}z^k, \quad |z| \leq 1. \tag{2.7}$$

On applying (2.7) in (2.1)–(2.5), we get

$$[1 - \lambda\bar{\mu}_2z - \bar{\lambda}\bar{\mu}_2 - \lambda\mu_2] G_0(z) = \bar{\lambda}\mu_1 G_1(z) + (1 - \bar{\lambda}\bar{\mu}_2 - \lambda\mu_2) \pi_{0,0} - \lambda\bar{\mu}_2z$$

and

$$\frac{\bar{\lambda}\mu_2}{z} G_0(z) = [1 - \lambda\bar{\mu}_1z - \bar{\lambda}\bar{\mu}_1 - \lambda\mu_1] G_1(z) + \left(1 - \bar{\lambda} + \frac{\bar{\lambda}\mu_2}{z} - \lambda z\right) \pi_{0,0}.$$

It immediately follows that,

$$G_0(z) = \frac{\bar{\lambda}\mu_1(\lambda z + \bar{\lambda}\mu_2 - \lambda z^2) - (z - \lambda\bar{\mu}_1z^2 - \bar{\lambda}\bar{\mu}_1z - \mu_1\lambda z)(1 - \lambda\bar{\mu}_2z - \bar{\lambda}\bar{\mu}_2 - \lambda\mu_2)}{\bar{\lambda}^2\mu_1\mu_2 - (z - \lambda\bar{\mu}_1z^2 - \bar{\lambda}\bar{\mu}_1z - \mu_1\lambda z)(1 - \lambda\bar{\mu}_2z - \bar{\lambda}\bar{\mu}_2 - \lambda\mu_2)} \pi_{0,0}$$

$$G_1(z) = \frac{(1 - \lambda\bar{\mu}_2z - \bar{\lambda}\bar{\mu}_2 - \lambda\mu_2)(\lambda z - \lambda z^2)}{\bar{\lambda}^2\mu_1\mu_2 - (z - \lambda\bar{\mu}_1z^2 - \bar{\lambda}\bar{\mu}_1z - \mu_1\lambda z)(1 - \lambda\bar{\mu}_2z - \bar{\lambda}\bar{\mu}_2 - \lambda\mu_2)} \pi_{0,0}.$$

Then we can obtain $G_0(1)$ and $G_1(1)$ as follows

$$G_0(1) = \frac{\lambda\mu_2 - \mu_1\mu_2(1 + \lambda)}{\lambda(\mu_1 + \mu_2) - \mu_1\mu_2(1 + \lambda)} \pi_{0,0}$$

$$G_1(1) = \frac{\lambda\bar{\lambda}\mu_2}{\mu_1\mu_2(1 + \lambda) - \lambda(\mu_1 + \mu_2)} \pi_{0,0}.$$

Using equation (2.6), i.e., $G_0(1) + G_1(1) = 1$, we can get the following result:

$$\pi_{0,0} = \frac{\lambda(\mu_1 + \mu_2) - \mu_1\mu_2(1 + \lambda)}{\lambda^2\mu_2 - \mu_1\mu_2(1 + \lambda)}$$

with the condition

$$\frac{\lambda}{\mu_1} < 1 \text{ and } \frac{\lambda}{\mu_2} < 1. \quad (2.8)$$

From equations (2.1), (2.4) and (2.2), we deduce that

$$\pi_{0,1} = \frac{\lambda}{\bar{\lambda}\mu_2} \pi_{0,0}$$

$$\pi_{1,1} = \left[\frac{1 - \bar{\lambda}\bar{\mu}_2 - \lambda\mu_2}{\bar{\lambda}\mu_1} \right] \frac{\lambda}{\bar{\lambda}\mu_2} \pi_{0,0}$$

$$\pi_{0,2} = \left[\left(\frac{1 - \bar{\lambda}\bar{\mu}_1 - \lambda\mu_1}{\bar{\lambda}\mu_2} \right) \left(\frac{1 - \bar{\lambda}\bar{\mu}_2 - \lambda\mu_2}{\bar{\lambda}\mu_1} \right) \frac{\lambda}{\bar{\lambda}\mu_2} - \frac{\lambda}{\bar{\lambda}\mu_2} \right] \pi_{0,0}.$$

From equation (2.3) and (2.5), we get

$$\pi_{0,k+1} = \frac{1 - \bar{\lambda}\bar{\mu}_1 - \lambda\mu_1}{\bar{\lambda}\mu_2} \pi_{1,k} - \frac{\lambda\bar{\mu}_1}{\bar{\lambda}\mu_2} \pi_{1,k-1}, \quad k \geq 2$$

$$\pi_{1,k} = \frac{1 - \bar{\lambda}\bar{\mu}_2 - \lambda\mu_2}{\bar{\lambda}\mu_1} \pi_{0,k} - \frac{\lambda\bar{\mu}_2}{\bar{\lambda}\mu_1} \pi_{0,k-1}, \quad k \geq 2.$$

For the sake of notation convenience, we define

$$\frac{\lambda}{\bar{\lambda}\mu_2} = k_1, \quad \frac{1 - \bar{\lambda}\bar{\mu}_1 - \lambda\mu_1}{\bar{\lambda}\mu_2} = k_2, \quad \frac{\lambda\bar{\mu}_1}{\bar{\lambda}\mu_2} = k_3, \quad \frac{1 - \bar{\lambda}\bar{\mu}_2 - \lambda\mu_2}{\bar{\lambda}\mu_1} = k_4, \quad \frac{\lambda\bar{\mu}_2}{\bar{\lambda}\mu_1} = k_5.$$

From the above assumptions we get,

$$\pi_{0,1} = k_1\pi_{0,0}, \quad \pi_{1,1} = k_1k_4\pi_{0,0}, \quad \pi_{0,2} = k_1[k_2k_4 - 1]\pi_{0,0}. \quad (2.9)$$

Using the above notations, one can rewrite the difference equations as follows

$$\pi_{0,k+1} = k_2\pi_{1,k} - k_3\pi_{1,k-1}, \quad k \geq 2 \quad (2.10)$$

$$\pi_{1,k} = k_4\pi_{0,k} - k_5\pi_{0,k-1}, \quad k \geq 2. \quad (2.11)$$

Solving (2.10) and (2.11) for $k = 2$ and $k = 3$ with the values given in (2.9), we get

$$\pi_{1,2} = k_1[-k_5 + k_4(k_2k_4 - 1)]\pi_{0,0}$$

$$\pi_{1,3} = k_1[-k_2k_4k_5 - k_3k_4^2 + (k_2k_4^2 - k_5)(k_2k_4 - 1)]\pi_{0,0} = k_1\left[R_{0,1}^{(1,3)} - k_4R_{1,1}^{(1,3)} + (k_2k_4 - 1)R_{0,2}^{(1,3)}\right]\pi_{0,0}$$

$$\pi_{0,3} = k_1[-k_2k_5 - k_3k_4 + k_2k_4(k_2k_4 - 1)]\pi_{0,0} = k_1\left[R_{0,1}^{(0,3)} - k_4R_{1,1}^{(0,3)} + (k_2k_4 - 1)R_{0,2}^{(0,3)}\right]\pi_{0,0}.$$

From this we obtain the general solution for $k > 3$ in recursive form.

$$\pi_{i,k} = k_1\left[R_{0,1}^{(i,k)} - k_4R_{1,1}^{(i,k)} + (k_2k_4 - 1)R_{0,2}^{(i,k)}\right]\pi_{0,0}, \quad k > 3 \text{ and } i = 0, 1, \quad (2.12)$$

where $R_{0,1}^{(i,k)}$, $R_{1,1}^{(i,k)}$, $R_{0,2}^{(i,k)}$ satisfied the recursive relations,

$$R_{0,1}^{(i,k)} = k_5R_{1,1}^{(i,k-1)} - k_2k_5R_{0,2}^{(i,k-1)}$$

$$R_{1,1}^{(i,k)} = k_3R_{0,2}^{(i,k-1)}$$

$$R_{0,2}^{(i,k)} = k_2k_4R_{0,2}^{(i,k-1)} + R_{0,1}^{(i,k-1)} - k_4R_{1,1}^{(i,k-1)}, \quad k \geq 4, \quad i = 0, 1.$$

The above stationary probability exists only if λ , μ_1 and μ_2 satisfy the conditions defined in the equation (2.8).

3. NUMERICAL ILLUSTRATIONS

In order to gain more insight of the model behavior, some numerical experiments are provided in this section. In Figures 2–5, the Equilibrium probabilities $\pi_{0,k}$ and $\pi_{1,k}$, $k \geq 1$ are plotted against the number of customers in the system (k) with fixed arrival rate $\lambda = 0.1$ and service rate $\mu_2 = 0.2$ by varying the service rate μ_1 as 0.3, 0.4, 0.5 and 0.6. We note that when number of customers increases, the probabilities $\pi_{i,k}$ ($i = 0, 1$) decreases with the increase in service rates μ_1 and the probability $\pi_{1,k}$ decreases gradually for the various service rates μ_1 and finally attains the saturation point more quickly than $\pi_{0,k}$.

Similarly, Figures 6–9 exhibits the impact of varying the service rate μ_2 . We fix the arrival rate $\lambda = 0.1$ and the service rate $\mu_1 = 0.2$, we notice that the system size probability $\pi_{0,k}$ decreases with the increase of μ_2 and attains steady state faster than the probability $\pi_{1,k}$. From the Figures 2–9, Probability values are slowly decreasing for the state $(0, k)$ even though we increase the service rate μ_2 due to more task carried out in stage 2. On the other hand, probability values are decreasing for the state $(1, k)$ more faster, if we increase the service rate μ_1 due to less work assigned to stage 1.

Tables 1 and 2 shows that the increase in service rates μ_1 and μ_2 leads to decrease in the mean and variance of the system. When comparing the Tables 1 and 2, it is evident that whenever the service rates (either μ_1 or μ_2) increases, the average number of customers and variance for the system automatically gets reduced and reaches the steady state.

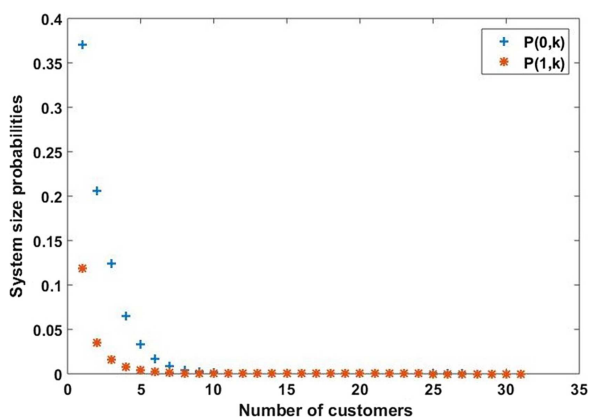


FIGURE 2. System size probabilities for $\lambda = 0.1$, $\mu_1 = 0.3$ and $\mu_2 = 0.2$.

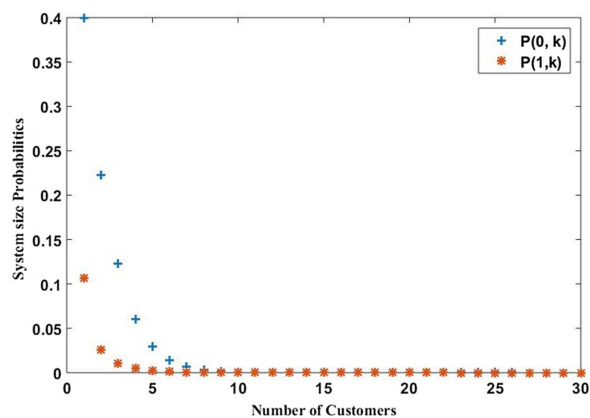


FIGURE 3. System size probabilities for $\lambda = 0.1$, $\mu_1 = 0.4$ and $\mu_2 = 0.2$.

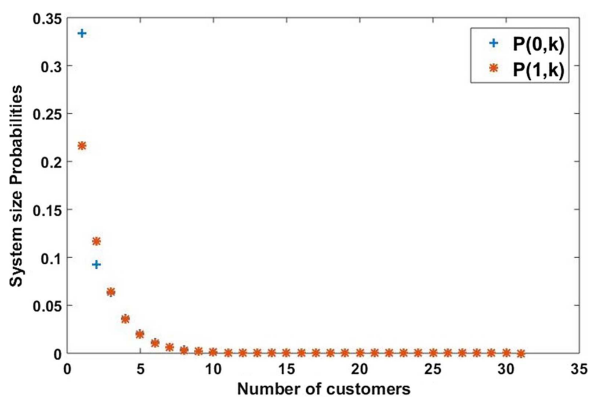


FIGURE 4. System size probabilities for $\lambda = 0.1$, $\mu_1 = 0.5$ and $\mu_2 = 0.2$.

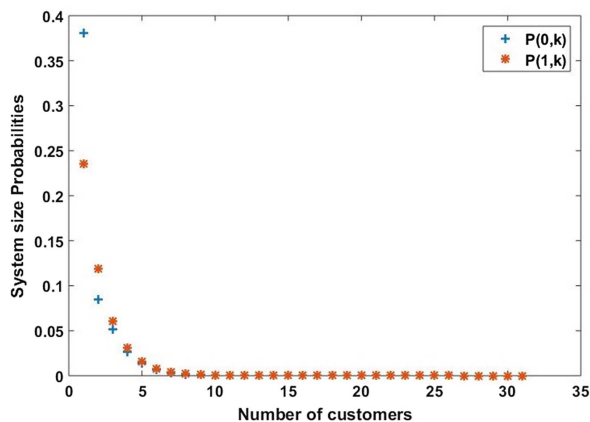


FIGURE 5. System size probabilities for $\lambda = 0.1$, $\mu_1 = 0.6$ and $\mu_2 = 0.2$.

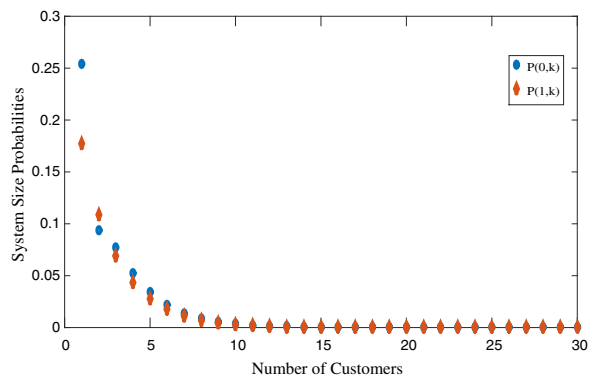


FIGURE 6. System size probabilities for $\lambda = 0.1$, $\mu_1 = 0.2$ and $\mu_2 = 0.3$.

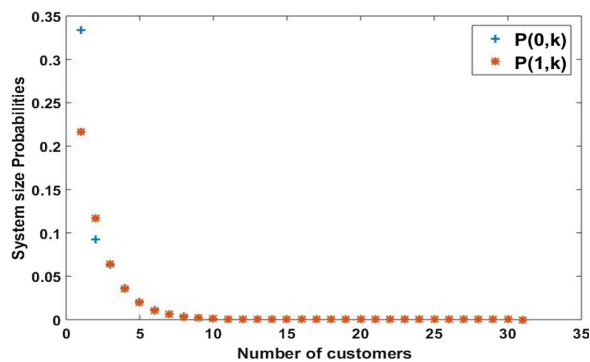


FIGURE 7. System size probabilities for $\lambda = 0.1$, $\mu_1 = 0.2$ and $\mu_2 = 0.4$.

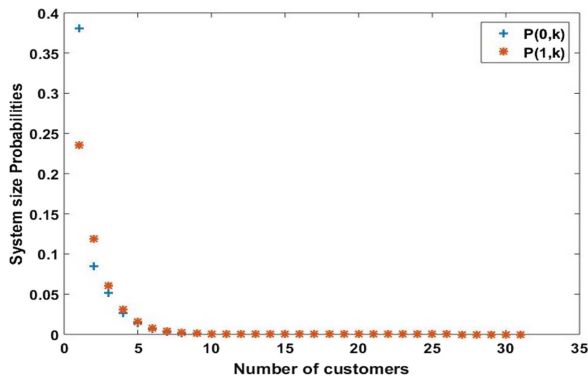


FIGURE 8. System size probabilities for $\lambda = 0.1$, $\mu_1 = 0.2$ and $\mu_2 = 0.5$.

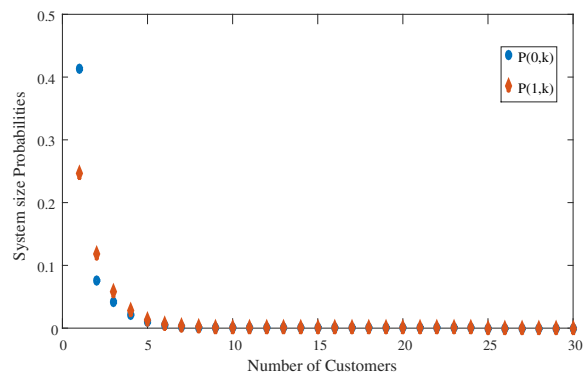


FIGURE 9. System size probabilities for $\lambda = 0.1$, $\mu_1 = 0.2$ and $\mu_2 = 0.6$.

TABLE 1. Expected system size and Variance for various service rates μ_1 corresponding to $\lambda = 0.1$ and $\mu_2 = 0.2$.

μ_1	$E(X_n), X_n = (0, k)$	$E(X_n), X_n = (1, k)$	$Var(X_n), X_n = (0, k)$	$Var(X_n), X_n = (1, k)$
0.3	1.399714576	0.794262625	6.371259001	3.351868266
0.4	1.107419423	0.450350631	3.802295520	1.362051214
0.5	0.990226313	0.313786005	2.971480007	0.799953292
0.6	0.927021691	0.240631163	2.570608803	0.55233752

TABLE 2. Expected system size and Variance for various service rates μ_2 corresponding to $\lambda = 0.1$ and $\mu_1 = 0.2$.

μ_2	$E(X_n), X_n = (0, k)$	$E(X_n), X_n = (1, k)$	$Var(X_n), X_n = (0, k)$	$Var(X_n), X_n = (1, k)$
0.3	0.947954845	1.298486621	4.363703683	5.642525291
0.4	0.566893276	1.054421623	1.993015616	3.451294688
0.5	0.407407397	0.965608454	1.264052422	2.784708743
0.6	0.318816984	0.921177585	0.920892765	2.477819707

4. CONCLUSION

An explicit solution is given for a single server queueing network model which consists of two stages of service with infinite capacity. In this model, steady state system size probabilities are obtained using partial generating function and expressed in recursive form. Numerical illustration were presented to show how sensitive the system probabilities are versus changes in parameters of the system. Performance measures such as mean variance are tabulated for different values of μ_1 and μ_2 . One can also extend this approach to the network with multi stages of service for both finite and infinite capacity.

Acknowledgements. The authors are grateful to the referees for their useful suggestions that lead to the considerable improvement in the presentation of the paper.

REFERENCES

- [1] S. Balsamo and V. De Nitto Persone, A review on queueing network models with finite capacity queues for software architectures performance prediction. *Perform. Eval.* **51** (2003) 269–288.
- [2] S.K. Bar-Lev, O. Boxma, W. Stadjie, F.A. Van der Duyn Schouten and C. Wiesmeyer, Two-stage queueing network models for quality control and testing. *Eur. J. Oper. Res.* **198** (2009) 859–866.
- [3] M.S. Chen and H. Yen, A two-stage approach in solving the state probabilities of the multi-queue $M/G/1$ model. *Int. J. Syst. Sci.* **47** (2016) 1230–1244.
- [4] F. Dong, J.R. Deglise-Hawkinson, M.P. Van Oyen and D.J. Singer, Dynamic control of a closed two-stage queueing network for outfitting process in shipbuilding. *Comput. Oper. Res.* **72** (2016) 1–11.
- [5] P.C. Garg, A study of a time dependent queueing system with two parallel channels. *Microelectron. Reliab.* **28** (1988) 7–9.
- [6] V. Giorno, A.G. Nobile and E. Pirozzi, A state-dependent queueing system with asymptotic logarithmic distribution. *J. Math. Anal. App.* **458** (2018) 949–966.
- [7] M.N. Gopalan and S. Kannan, Expected number analysis of a two-server queueing network subject to inter-stage inspection and rework. *Comput. Oper. Res.* **22** (1995) 935–946.
- [8] L. Haoyu and L. Jin, A two-stage service replica strategy for business process efficiency optimization in community cloud. *Ch. J. Electron.* **26** (2017) 80–87.
- [9] X. He and W. Hu, A two-stage queue model to optimize layout of urban drainage system considering extreme rainstorms. *Math. Probl. Eng.* **2017** (2017) 6380521.
- [10] T. Katayama, A note on sojourn time analysis of a two-stage queueing system. *Stochastic Models* **15** (1999) 379–394.
- [11] C.M. Krishna and Y.H. Lee, A study of two phase service. *Oper. Res. Lett.* **9** (1990) 91–97.
- [12] L. Li, J. Wang and F. Zhang, Equilibrium customer strategies in markovian queues with partial breakdowns. *Comput. Ind. Eng.* **66** (2013) 751–757.
- [13] D.H. Moon, Analysis of multi-server two-stage queueing network with split and blocking Yang Woo Shin. *IFAC Proc. Vol.* **47** (2014) 1667–1671.
- [14] K. Patil and K. Turck, A two-queue model for optimising the value of information in energy-harvesting sensor networks. *Perform. Eval.* **5316** (2017) 30009–30015.
- [15] N. Qi, N.I. Miridakis, T.A. Tsiftsis and R. Yao, On the fundamental queue analysis for relay-assisted two-stage communication networks. In: *IEEE Wireless Communications and Networking Conference Workshop (WCNCW)*. IEEE (2019) 1-6.
- [16] N. Qi, N.I. Miridakis, M. Xiao, T.A. Tsiftsis, R. Yao and S. Jin, Traffic-aware two-stage queueing communication networks: queue analysis and energy saving. Preprint [arXiv:1909.05891](https://arxiv.org/abs/1909.05891) (2019).
- [17] V. Sağlam and M. Zobu, A two-stage model queueing with no waiting line between channels. *Math. Probl. Eng.* **2013** (2013) 679369.
- [18] V. Sağlam, M. Sağır, E. Yücesoy and M. Zobu, The analysis, optimization, and simulation of a two-stage tandem queueing model with hyperexponential service time at second stage. *Math. Probl. Eng.* **2015** (2015) 165219.
- [19] K. Salah, Analysis of a two-stage network server. *Appl. Math. Comput.* **217** (2011) 9635–9645.
- [20] R. Sudhesh, Transient analysis of a queue with system disasters and customer impatience. *Queue. Syst.* **66** (2010) 95–105.
- [21] S. Xuan, D. Man, W. Wang, J. Zhang, W. Yang and X. Du, Two-stage mixed queueing model for web security gateway performance evaluation. In: *26th International Conference on Computer Communication and Networks (ICCCN)*. IEEE (2017) 1–7.
- [22] M. Yan, W. Liu and J. Li, Equilibrium balking behavior in the $Geo/Geo/1$ queueing system with multiple vacations. *Appl. Math. Model.* **37** (2013) 3861–3878.
- [23] T. Yang, J. Wang and F. Zhang, Equilibrium Balking strategies in the $Geo/Geo/1$ queues with server breakdowns and repairs. *Quality Technol. Quant. Manage.* **11** (2014) 231–243.
- [24] W. Yang, T. Sung Kim, H. Min Park and D. Eun Lim, Analysis of a two-stage queue with a single server and N -policy. *Am. J. Math. Manage. Sci.* **35** (2016) 261–270.
- [25] Z. Yu and S. Liu, Equilibrium strategies of the unobservable $M/M/1$ queue with balking and delayed repairs. *Appl. Math. Comput.* **290** (2016) 56–65.
- [26] N. Zarrinpoor and F. Mohammad Saber, Design of a reliable hierarchical location-allocation model under disruptions for health service networks: a two-stage robust approach. *Comput. Ind. Eng.* **109** (2017) 130–150.
- [27] N. Zarrinpoor, M.S. Fallahnezhad and M.S. Pishvaei, Design of a reliable hierarchical location-allocation model under disruptions for health service networks: a two-stage robust approach. *Comput. Ind. Eng.* **109** (2017) 130–150.
- [28] W. Zhou, W. Huang and R. Zhang, A two-stage queueing network on form postponement supply chain with correlated demands. *Appl. Math. Model.* **38** (2014) 2734–2743.