Transient Analysis of a Series Configuration Queueing System

Yu-Li Tsai*, Daichi Yanagisawa, and Katsuhiro Nishinari

Abstract—**In this paper, we consider the transient analysis of a popular series configuration queueing system consisting of two service stations with blocking phenomena. This kind of queueing system contains wide values of applications for automobile industries. We assume arrivals obey Poisson process and exponential service times of each service station. Transient probabilities of the system can be obtained by applying Runge-Kutta method to solve the master equations. Performance measures including mean number in the system, mean waiting time in the system, blocking probability of the service station before the terminal service station and rejecting probability are defined to study the dynamic behavior of the system. Disposition strategies are suggested to make the system work in high performance process.**

Keywords—**Transient Analysis, Disposition Strategy, Rung-Kutta, Performance Analysis, Simulation**

I. Introduction

Open queueing networks are very important queueing systems for the optimization of industrial activities in modern economy. Several significant applications of open queueing networks include computer networks, global logistic networks, production line system, supply chain networks, telecommunication system etc. Recently, because of the development of computational facilities, the remarkable trends for applying real data analytics to improve theoretical predictions of specific service queueing systems become more and more important. Therefore, successfully figuring out the theoretical system performance values of a queueing system through exact analysis or numerical analysis is a prerequisite to make the systems work more efficiently by real data analytics. Moreover, the theoretical results can be further validated and improved by real data analysis.

Fig 1. Series configuration queueing system with two service stations.

Manuscript received April 21, 2017; This work was supported in part by the JSPS KAKENHI Grant Numbers 25287026 and 15K17583.

Y. L. Tsai*, the corresponding author, is with the Department of Advanced Interdisciplinary Studies, School of Engineering, The University of Tokyo, CO 153-8904 Japan (phone: +81-3-5452-5022; e-mail: tokyotsai@gmail.com).

D. Yanagisawa is with the Research Center for Advanced Science and Technology, The University of Tokyo. CO 153-8904 Japan (e-mail: tDaichi@mail.ecc.u-tokyo.ac.jp)

K. Nishinari is with the Research Center for Advanced Science and Technology, The University of Tokyo. CO 153-8904 Japan (e-mail: tknishi@mail.ecc.u-tokyo.ac.jp)

We study a very popular kind of queueing system containing abundant applications for the production line in automobile industries, called series configuration queueing systems. The major characteristic of this system is that there are no queues between each service station. The system consisting of two service stations is shown in **Figure 1**. Dynamic performance measures evolving with time including mean number in the system, mean waiting time in the system, blocking probability of the station-1 and rejecting probability of the system are evaluated through transient probabilities. We discover that some of the performance measures are convergent to the results of our previous studies on steady-state analysis studies. Simulation results reveal that there is still difference of operational efficiency of the system by setting different service rates for the service stations. We will suggest better disposition strategies for the system to keep higher performance by numerical results.

Hunt [1] first studied queueing systems operating in series configurations with blocking phenomena. He derived the maximum possible utilization of the systems constrained with different capacities including an infinite queue between service stations, no queue between service stations, a finite queue between stations, and the case of the unpaced belt-production line. Abate and Whitt [2] presented an approximation method to investigate the transient behavior of M/M/1 queueing system. The method can help determine whether steady-state descriptions are reasonable or not in the condition that the arrival and service rates are nearly constant over time interval. Abate and Whitt [3] showed how Laplace transform analysis can obtain insights about transient behavior of the M/M/1 queueing system. They further determined the asymptotic behavior of the system through a transform factorization. Bertsimas and Nakazato [4] investigated queueing systems with the class of mixed generalized Erlang distributions. They found simple closed form expressions for the Laplace transforms of the queue length distribution and the waiting time distribution. Abate and Whitt [5] gave the time-dependent moments of the workload process in the M/G/1 queue. They obtained results for the covariance function of the stationary workload process. Various time-dependent characteristics described in terms of the steady-state workload distribution are demonstrated. Choudhury et al. [6] proposed an algorithm for numerically inverting multidimensional transforms. This method can be applied to both continuous variables and discrete variables transformations. They applied the method to invert the two-dimensional transforms of the joint distribution of the duration of a busy period, the number served in the busy period, time-dependent transient queue-length and workload distributions in the M/G/1 queue. Kaczynski et al. [7] derived the exact distribution of the nth customer's mean waiting time in the system in an M/M/s

ISBN: 978-988-14048-4-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

system with k customers initially present. Algorithms for evaluating the covariance between mean waiting time in the system and for an M/M/1 with k customers at beginning of the state of the system was developed. Kim and Whitt [8] showed that the bias of the steady-state Little's law can be estimated and reduced by applying time-varying Little's law. Kim and Whitt [9] advocated a statistical approach to study characteristics of Little's law for queueing systems with non-stationary distributions. They presented their theoretical analysis with data from a call center and simulation experiments. Tsai et al. [10, 11] proposed general disposition strategies for series configuration queueing systems consisting of the arbitrary number of service stations. Stability conditions of steady-states for the system with two, three and four service stations are derived in exact form. Several numerical methods used to solve transient queueing systems can be found in [12] by Bolch and Greiner.

The rest of the paper is organized as follows. The summary of notations used in our model and problem description are introduced in the beginning of next section. Formulation of dynamic behavior of the system is also described in the Section 2. Section 3 contains numerical results of the transient analysis of the system and disposition strategies. Finally, we conclude with discussions of our works and indicate possible directions for future research in section 4.

II. Problem Formulation and Notations

In our analysis, a series configuration queueing system consisting of two service stations operates independently and simultaneously. Poisson arrival process with mean arrival rate λ. The time to serve a customer in each station obeys exponential distribution with mean service time $\frac{1}{\mu}$. Customers should enter each service station to receive services in order to complete the service and leave the system. The existence of blocking phenomena after the service in the station-1 is because there is no queue between service stations. This phenomenon happen in the condition that a customer has completed the service in the station-1, but another customer is still receiving service in the station-2. The finite capacity of the queueing system is denoted as K. Each service station can serve a customer at a time and the service time is independent of the number of customers. This system obeys the first come first serve (FCFS) discipline.

The notations μ_1 and μ_2 denote the service rate of the station-1 and the station-2, respectively. Moreover, we use $P_{n_1,n_2,n_3}(t)$ to denote the transient probability $P_{n_1,n_2,n_3}(t)$ of n_1 customer in the station-2 and n_2 customer in the station-1 and n_3 customer in the queue. For instance, the steady-state probability $P_{1,b,3}(t)$ means that there is a customer who is blocked in the station-1, since the customer in the station-2 is still receiving the service. There are 3 customers waiting in the queue.

III. Modeling Framework

We apply continuous-time Markov process to model the series configuration queueing system consisting of two service stations. The capacity of the system is supposed to be equal or larger than 5 (i.e. $K \geq 5$). According to the quasi-birth-and-death process, the dynamic behavior of the system can be described as following system of differential equations:

$$\frac{dP_{0,0,0}(t)}{dt} = -\lambda P_{0,0,0}(t) + \mu_2 P_{1,0,0}(t), \tag{1}$$

$$\frac{dP_{1,0,0}(t)}{dt} = -(\lambda+\mu_2) P_{1,0,0}(t) + \mu_2 P_{1,b,0}(t) + \mu_1 P_{0,1,0}(t), \tag{2}$$

$$\frac{dP_{1,b,0}(t)}{dt} = -(\lambda+\mu_2) P_{1,b,0}(t) + \mu_1 P_{1,1,0}(t), \tag{3}$$

$$\frac{dP_{0,1,i}(t)}{dt} = -(\lambda+\mu_1) P_{0,1,i}(t) + \lambda P_{0,0,i}(t) + \mu_2 P_{1,1,i}(t), \quad i = 0,1,2,\ldots,K-4. \tag{4}$$

$$\frac{dP_{1,1,i}(t)}{dt} = -(\lambda+\mu_1+\mu_2) P_{1,1,i}(t) + \lambda P_{1,0,i}(t) + \mu_2 P_{1,b,i+1}(t) + \mu_1 P_{0,1,i+1}(t), \quad i = 0,1,2,\ldots,K-4. \tag{5}$$

$$\frac{dP_{1,b,i+1}(t)}{dt} = -(\lambda+\mu_2) P_{1,b,i+1}(t) + \lambda P_{1,b,i}(t) + \mu_1 P_{1,1,i+1}(t), \quad i = 0,1,2,\ldots,K-3. \tag{6}$$

$$\frac{dP_{0,1,K-3}(t)}{dt} = -(\lambda+\mu_1) P_{0,1,K-3}(t) + \lambda P_{0,1,K-4}(t) + \mu_2 P_{1,1,K-3}(t), \tag{7}$$

$$\frac{dP_{1,1,K-3}(t)}{dt} = -(\lambda+\mu_1+\mu_2) P_{1,1,K-3}(t) + \lambda P_{1,1,K-4}(t) + \mu_2 P_{1,b,K-2}(t) + \mu_1 P_{0,1,K-2}(t), \tag{8}$$

$$\frac{dP_{1,b,K-2}(t)}{dt} = -\mu_2 P_{1,b,K-2}(t) + \lambda P_{1,b,K-3}(t) + \mu_1 P_{1,1,K-2}(t), \tag{9}$$

$$\frac{dP_{0,1,K-2}(t)}{dt} = -(\lambda+\mu_1) P_{0,1,K-2}(t) + \lambda P_{0,1,K-3}(t) + \mu_2 P_{1,1,K-2}(t), \tag{10}$$

$$\frac{dP_{1,1,K-2}(t)}{dt} = -(\mu_1+\mu_2) P_{1,1,K-2}(t) + \lambda P_{1,1,K-3}(t) + \mu_1 P_{0,1,K-1}(t), \tag{11}$$

$$\frac{dP_{1,1,K-1}(t)}{dt} = -\mu_1 P_{1,1,K-1}(t) + \lambda P_{0,1,K-2}(t). \tag{12}$$

The normalization condition of the system at each time step t is

$$P_{0,0,0}(t) + P_{1,0,0}(t) + \sum_{i=0}^{K-1} P_{0,1,i}(t) + \sum_{i=0}^{K-2} P_{1,1,i}(t) + \sum_{i=0}^{K-2} P_{1,b,i}(t) = 1. \tag{13}$$

We apply Runge-Kutta method to evaluate transient probabilities of the system. Meanwhile, we assume that the system is empty at the beginning of the system state. This means that the initial conditions of the system are given by

$$P_{0,0,0}(0) = 1;\ P_{1,0,0}(0) + \sum_{i=0}^{K-1} P_{0,1,i}(0) + \sum_{i=0}^{K-2} P_{1,1,i}(0) + \sum_{i=0}^{K-2} P_{1,b,i}(0) = 0. \tag{14}$$

Theorem 1. The stability conditions of the series configuration queueing system consisting of two service stations can be referred in [10],

(1) For $\mu_1 \neq \mu_2$

$$\lambda < \frac{\mu_1(\mu_1\mu_2 + \mu_2^2)}{\mu_1^2 + \mu_1\mu_2 + \mu_2^2}.$$

(2) Special case: $\mu_1 = \mu_2 = \mu$

$$\lambda < \frac{2}{3}\mu.$$

- Performance measures

Performance measures for the system consisting of two service stations are defined by

ISBN: 978-988-14048-4-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

(1) Mean number of customers in the system

$$L(t) = [P_{1,0,0}(t) + P_{0,1,0}(t) + P_{1,b,0}(t)] + \sum_{n=2}^{K-1} [P_{1,b,n-1}(t) + P_{1,1,n-2}(t) + P_{0,1,n-1}(t)] \cdot n + P_{1,1,K-2}(t) \cdot K. \quad (15)$$

(2) Mean number of customers in the queue

$$L_q(t) = \sum_{n=1}^{K-2} [P_{1,b,n}(t) + P_{1,1,n}(t) + P_{0,1,n}(t)] \cdot n. \quad (16)$$

(3) Blocking probability of the customer in the station-1

$$P_b(t) = \sum_{n=0}^{\infty} P_{1,b,n}(t). \quad (15)$$

(4) Rejecting probability of the system

$$P_r(t) = P_{1,b,K-2}(t) + P_{1,1,K-2}(t) + P_{0,1,K-1}(t). \quad (16)$$

(5) Mean waiting time in the system (Little's Law)

$$W(t) = \frac{L(t)}{\lambda_{eff}(t)}, \quad (17)$$

where $\lambda_{eff}(t) \equiv \lambda[1 - P_r(t)]$ is the effective mean arrival rate.

(6) Mean waiting time in the queue (Little's Law)

$$W_q(t) = \frac{L_q(t)}{\lambda_{eff}(t)}. \quad (18)$$

IV. NUMERICAL RESULTS

In this section, numerical experiments for the queueing system consisting of two service stations are performed. Both performance metrics of the system with equivalent service rates (i.e. $\mu_1 = \mu_2 = \mu$) and different service rates are presented to study the dynamic behavior of the system. The disposition strategies will be proposed to make the system work in a better operational efficient way through our results of simulations.

- **Same service rates for each service station**

We first want to study the effects of the capacity of the queue on decay rate of different performance measures. We fix $\mu_1 = \mu_2 = 1$, $\lambda = 0.666$ and set $K = 200$ and $K = 50$. It is investigated how mean number in the system, mean waiting time in the system, blocking probability and rejecting probability of the system evolve with time when the finite capacity of the queue is 200, as shown in **Figure 2 ~ Figure 5.** It is observed that the decay speed of each performance converges to the steady-states is pretty slow, because it almost takes 50000 (time steps) from the transition states to become steady states. On the other hand, we discover that the speed of convergence is faster in the cases that we set lower capacity (K = 50) for the system, as shown in **Figure 6** ~ **Figure 9.** It is noted that the steady-state of blocking probability in both cases approach to 0.33. This result is consistent with our simulations in previous work [10].

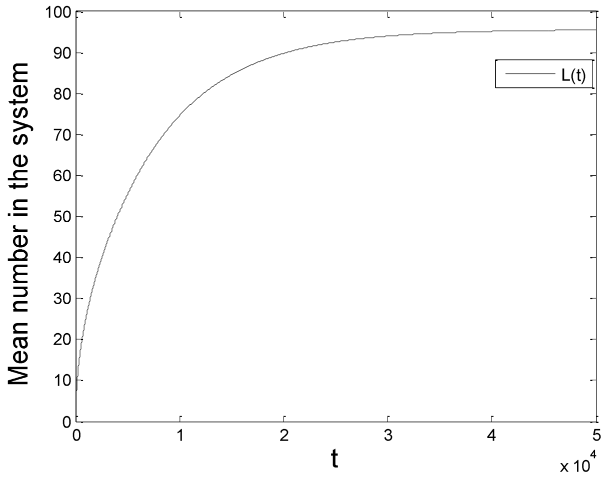

Fig 2. Mean number in the system (K = 200)

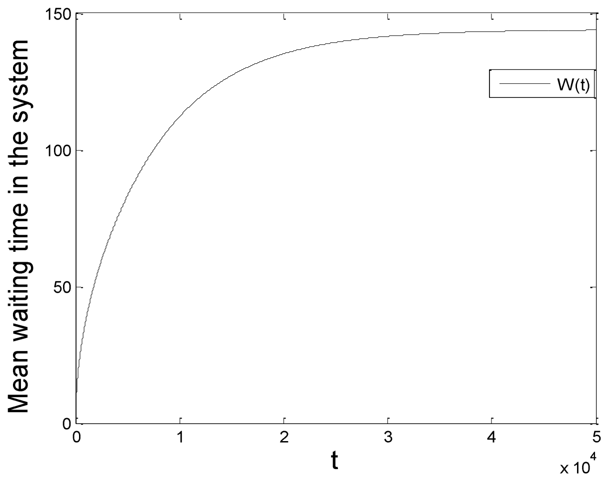

Fig 3. Mean waiting time in the system (K = 200)

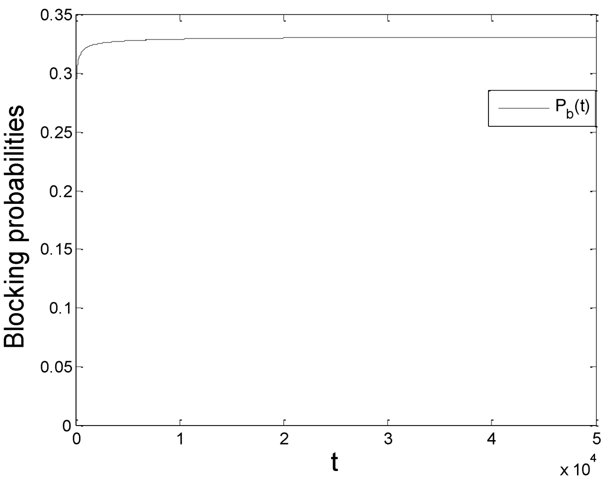

Fig 4. Blocking probability (K = 200)

ISBN: 978-988-14048-4-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

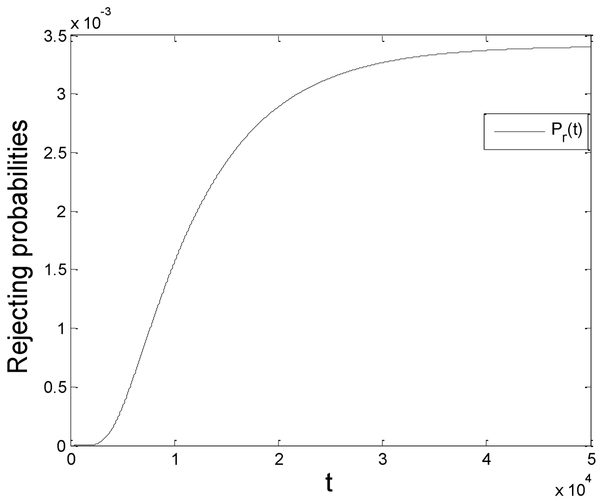

Fig 5. Rejecting probability (K = 200)

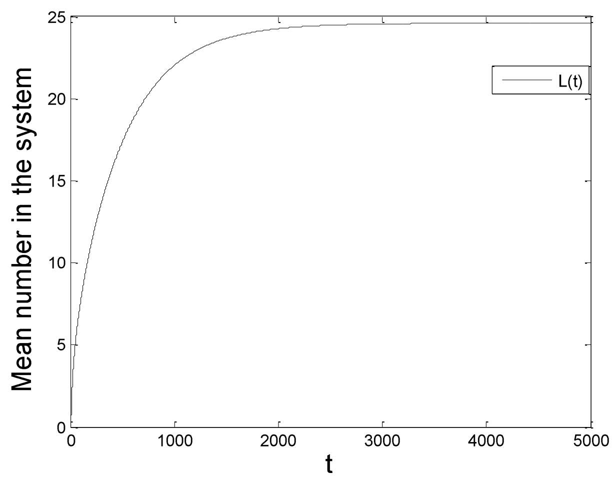

Fig 6. Mean number in the system (K = 50)

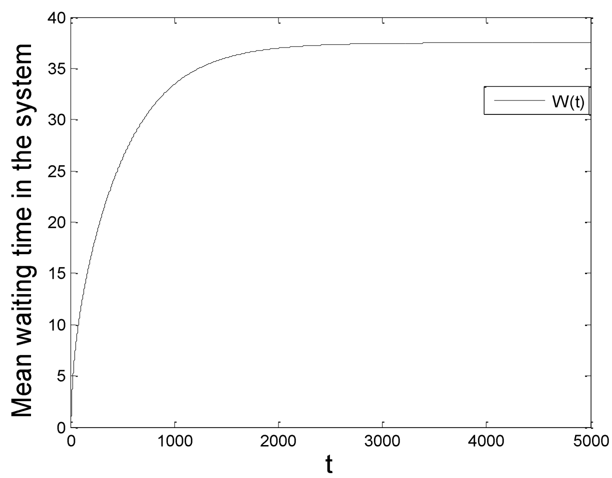

Fig 7. Mean waiting time in the system (K = 50)

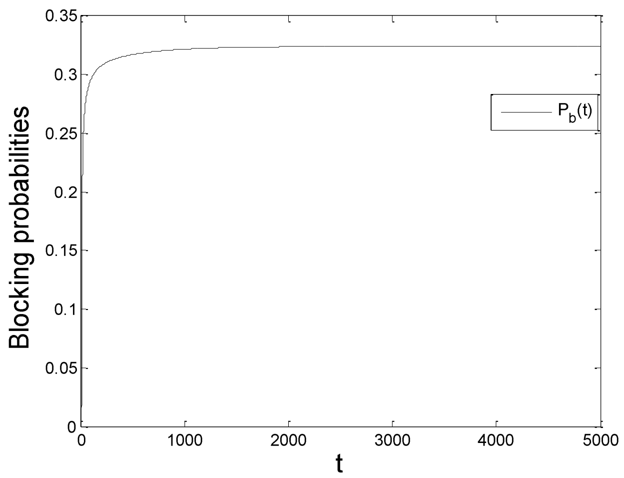

Fig 8. Blocking probability (K = 50)

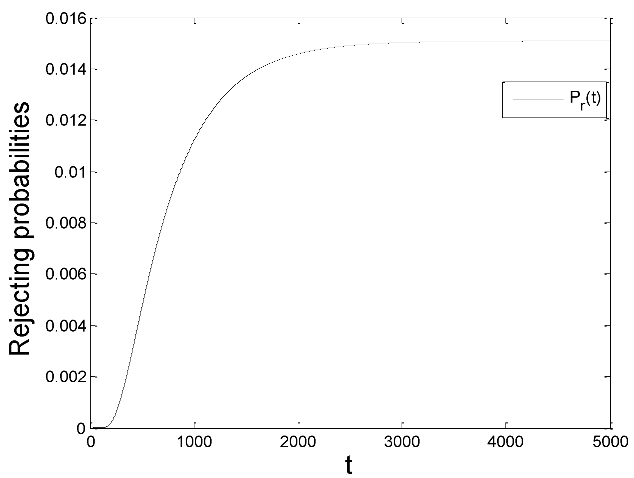

Fig 9. Rejecting probability. (K = 50)

We continue to study the effects of various mean arrival rates on performance measures of the system. We set $\mu_1 = \mu_2 = 1$, $K = 50$ and varies mean arrival rate λ from 1~5. It can be easily observed that the convergent speed of each performance measure increases as mean arrival rate increases, as shown in **Figure 10** ~ **Figure 13**., respectively. Furthermore, it is noted that the difference of mean waiting time in the system is not large for each case with different mean arrival rate in the early stage of transient states, as shown in **Figure 11.** It seems that the mean waiting time in the system in the stationary states are similar except $\lambda = 1$ in **Figure 11.** The blocking probabilities still approach 0.33 in all cases when the transient states become steady, as shown in **Figure 12**. It is discovered that the convergent speed of blocking probabilities are almost equivalent.

ISBN: 978-988-14048-4-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

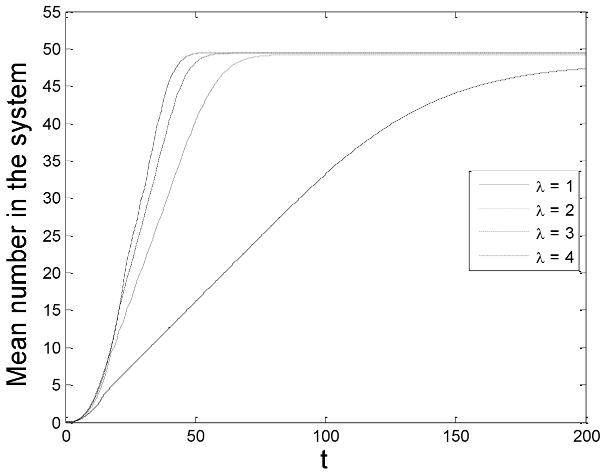

Fig 10. Mean number in the system (K = 50)

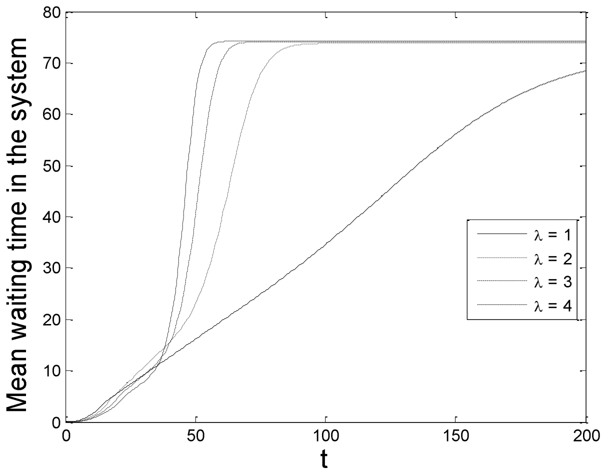

Fig 11. Mean waiting time in the system (K = 50)

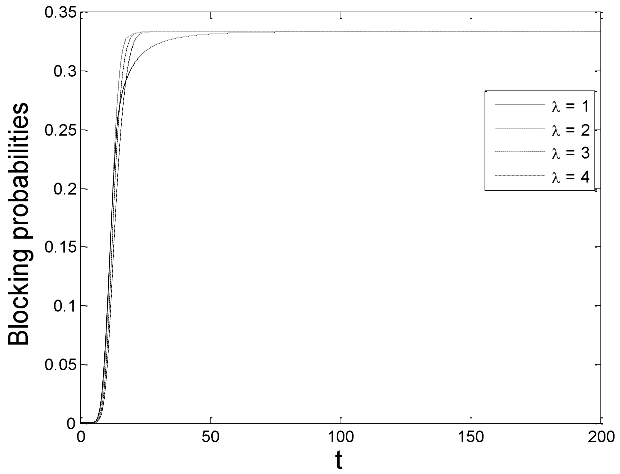

Fig 12. Blocking probability (K = 50)

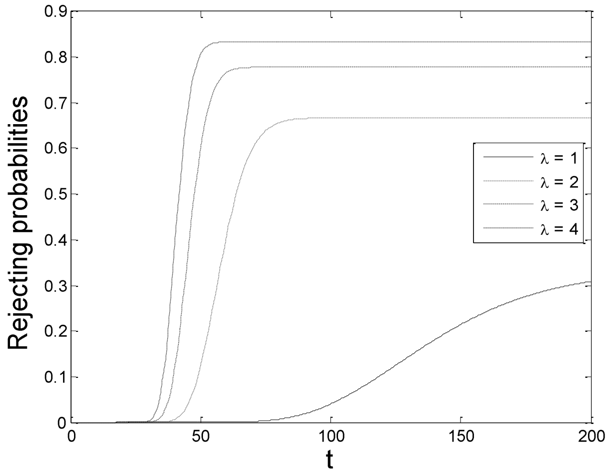

Fig 13. Rejecting probability (K = 50)

- **Controlling the service rates of the two service stations**

Next, we study the impact of setting different service rates for the service stations on operational efficiency of the system (i.e. mean waiting time in the system), and suggest disposition strategy for the system operating more efficiently. We consider the following cases:

Case 1
We choose $\mu_1 = 1,\ \mu_2 = 2$ and $\mu_1 = 2,\ \mu_2 = 1$ and set $\lambda = 0.2$.

Case 2
We choose $\mu_1 = 1,\ \mu_2 = 2$ and $\mu_1 = 2,\ \mu_2 = 1$ and set $\lambda = 0.4$.

Case 3
We choose $\mu_1 = 1,\ \mu_2 = 2$ and $\mu_1 = 2,\ \mu_2 = 1$ and set $\lambda = 0.66$.

Case 4
We choose $\mu_1 = 1,\ \mu_2 = 2$ and $\mu_1 = 2,\ \mu_2 = 1$ and set $\lambda = 0.7$.

Case 5
We choose $\mu_1 = 1,\ \mu_2 = 2$ and $\mu_1 = 2,\ \mu_2 = 1$ and set $\lambda = 0.74$.

Case 6
We choose $\mu_1 = 1,\ \mu_2 = 2$ and $\mu_1 = 2,\ \mu_2 = 1$ and set $\lambda = 0.78$.

As transient states become steady states, we discover that disposing different service rate for the service stations causes different operational efficiency of the system, as shown in **Figure 14.** ~ **Figure 19.**, respectively. It can also be observed that the speed of decay to the steady-state of the mean waiting time in the system decreases as mean arrival rate increases. We finally suggest that disposing higher service rate for the station-1 in order to keep the high performance operations of the series configuration queueing system consisting of two service stations. Transient analysis shows the same pattern of the disposition strategy as our previous works on steady-state analysis [10].

ISBN: 978-988-14048-4-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

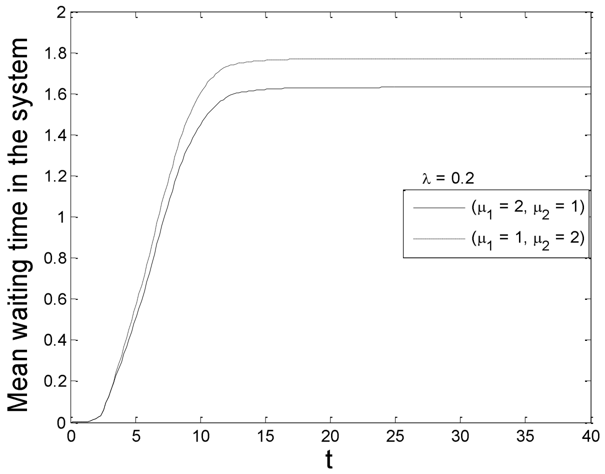

Fig 14. Mean waiting time in the system ($\lambda = 0.2$)

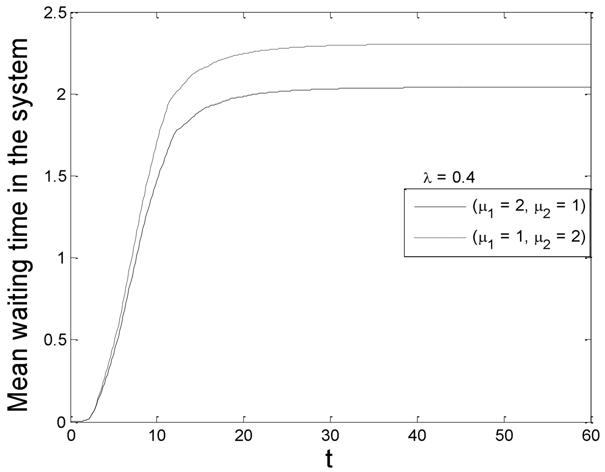

Fig 15. Mean waiting time in the system ($\lambda = 0.4$)

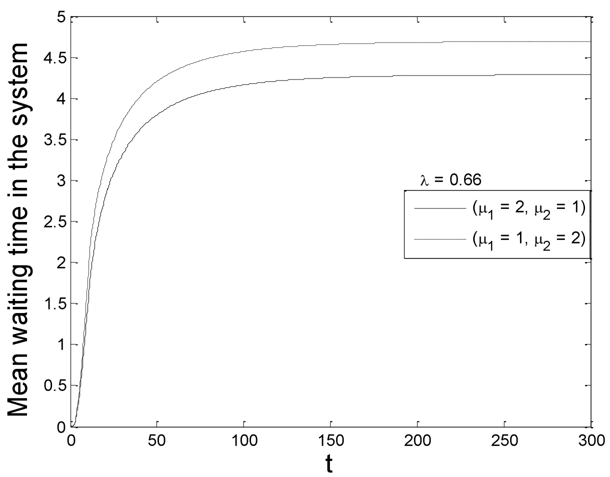

Fig 16. Mean waiting time in the system ($\lambda = 0.66$)

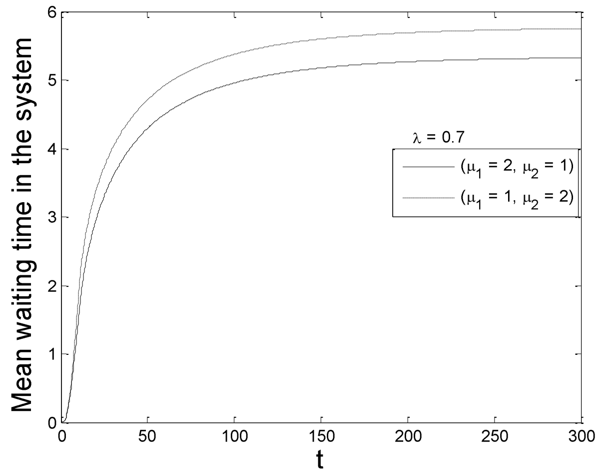

Fig 17. Mean waiting time in the system ($\lambda = 0.7$)

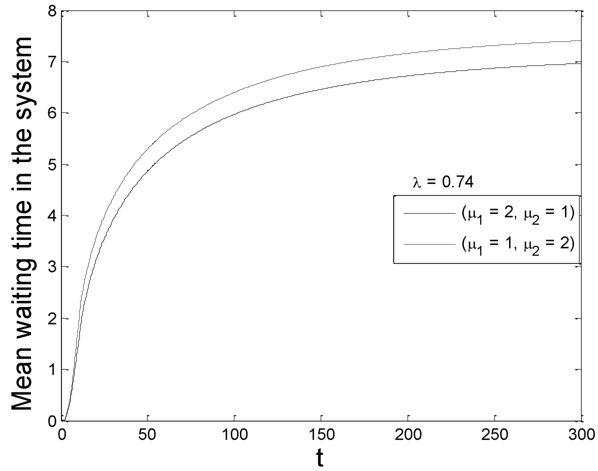

Fig 18. Mean waiting time in the system ($\lambda = 0.74$)

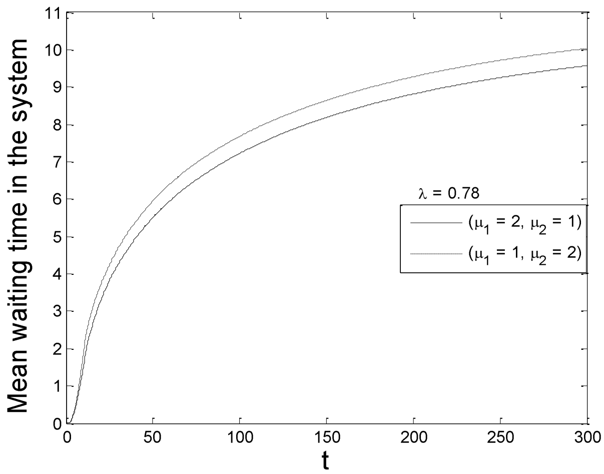

Fig 19. Mean waiting time in the system ($\lambda = 0.78$)

ISBN: 978-988-14048-4-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

V. Conclusion

We have successfully solved the transient probabilities of the series configuration queueing system consisting of two service stations. We also calculated important transient performance measures, such as mean number in the system, mean waiting time in the system, blocking probability and rejecting probability of the system to investigate the dynamic behavior of the system. Moreover, the convergent speed of each performance measure of the system depends on the capacity of the system and the mean arrival rate according to the numerical results.

In order to keep better operational efficiency of the system, we suggest setting higher service rate for the station-1 of the series configuration queueing system with two service stations. This consideration is consistent with the steady-state analysis for the same system in our previous works [10]. We have also shown the same results of disposition strategies for the transient analysis of the system in this study.

We will conduct statistical analysis of real data collected from industries to validate our theoretical analysis. Development of refined numerical method to keep numerical stability for the analysis of systems consisting of more than two service stations is worth for future research.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers 25287026 and 15K17583.

References

[1] G.C. Hunt, "Sequential arrays of waiting lines," Operations Research, Vol. 4 pp. 674–683, 1956.
[2] J. Abate, W. Whitt, "Transient behavior of the M/M/1 queue: srarting at the origin," Queueing Systems, Vol. 2 pp. 41–65, 1987.
[3] J. Abate and W. Whitt, "Transient behavior of the M/M/1 queue via Laplace transforms," Advanced in Applied Probability, Vol. 20 pp. 145–178, 1988.
[4] D.J. Bertsimas and D. Nakazato, "Transient and busy period analysis of the GI/G/1 queue: the method of stages," Queueing Systems, Vol. 10 pp. 153-184, 1992.
[5] J. Abate and W. Whitt "Transient behavior of the M/G/1 workload process," Operations Research, Vol. 42 pp. 750–764, 1994.
[6] G.L. Choudhury, D.M. Lucantoni and W. Whitt "Multidimensional transform inversion with applications to the transient M/G/1 Queue," The Annals of Applied Probability, Vol. 4 pp. 719-740, 1994.
[7] W.H. Kaczynski, L.M. Leemis and J.H. Drew "Transient queueing analysis," INFORMS Journal on Computing, Vol. 24 pp. 10-28, 2012.
[8] S.H. Kim and W. Whitt "Estimating waiting times with the time-varying Little's Law," Probability in the Engineering and Informational Science, Vol. 27 pp. 471-506, 2013.
[9] S.H. Kim and W. Whitt "Statistical analysis with Little's Law," Operations Research, Vol. 61 pp. 1030-1045, 2013.
[10] Y.L. Tsai, D. Yanagisawa and K. Nishinari "Disposition strategies for open queueing networks with different service rates," Engineering Letters, Vol. 46 pp. 418-428, 2016.
[11] Y.L. Tsai, D. Yanagisawa and K. Nishinari "Performance analysis of series configuration queueing system with four service stations," Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2016, IMECS 2016, 16-18 March, 2016, Hong Kong, pp. 931-935
[12] G. Bolch and S. Greiner, Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications. New Jersey: Wiley-Interscience, 2006, ch. 3.

ISBN: 978-988-14048-4-8
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)