East Asian Math. J.
Vol. 31 (2015), No. 5, pp. 707–716
http://dx.doi.org/10.7858/eamj.2015.051

QUEUEING SYSTEMS WITH N-LIMITED NONSTOP FORWARDING

Yutae Lee

Abstract. We consider a queueing system with N-limited nonstop forwarding. In this queueing system, when the server breaks down, up to N customers can be serviced during the repair time. It can be used to model an assembly line consisting of several automatic stations and a manual backup station. Within the framework of $Geo^X/D/1$ queue, the matrix analytic approach is used to obtain the performance of the system. Some numerical examples are provided.

1. Introduction

In many practical systems such as communications and manufacturing systems, the assumption that the server is reliable and always available to customers seems to be unrealistic. It is well known that the performance of unreliable system is highly influenced by server breakdowns [1]. For this reason, queueing systems with servers subject to breakdowns and repairs have been studied extensively. For related works, refer to [2, 3, 4, 5, 6, 7, 8, 9, 10], where customers who find the server broken down should wait in the queue, without being serviced, until the server is repaired.

This paper analyses a variant of a queueing system with unreliable server: discrete-time queueing system with N-limited nonstop forwarding. This queueing system operates as follows: customers arrive according to a batch geometric process. The server starts immediately the repair process whenever the server breaks down. Despite the server breakdown, up to N customers can be serviced during the repair time.

Many practical situations can be represented more accurately by the queueing system with N-limited nonstop forwarding. An example is found in manufacturing systems [11]. There is an assembly line consisting of several automatic

Received February 24, 2015; Revised April 20, 2015; Accepted September 21, 2015.

2010 *Mathematics Subject Classification.* 60K25, 68M20.

Key words and phrases. queueing system; nonstop forwarding; N-limited; computer network.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(No. NRF-2013R1A1A4A01013094).

(pISSN 1226-6973, eISSN 2287-2833)

stations and a backup station in the end of the line. The backup station is able to perform the work of any station in the line and is typically a manual station that is used to fix the assembly from a failed station [11]. When a breakdown occurs, the assembly line remains operational, while all jobs going through the line bypass the failed station and have to be fixed at the backup station. Due to the high cost at the manual backup station, if the breakdown station is not repaired until a specified amount of work is processed, the entire assembly line is shut down.

In this paper, within the framework of discrete-time single server $Geo^X/D/1$ queueing system, the stability condition and the service availability are given. The matrix analytic approach is used to obtain the steady-state joint probability distribution of the number of customers and system state. The mean delay is also derived. Finally, some numerical examples are provided.

2. $Geo^X/D/1$ queue with N-limited nonstop forwarding

We consider a discrete-time single server queueing system in which the time axis is divided into fixed-length contiguous intervals, referred to as slots. Customers arrive to the system from outside in accordance with a batch geometric process. The numbers of customers entering the system during the consecutive slots are assumed to be i.i.d. non-negative discrete random variables with an arbitrary probability distribution. Let a_k be the number of customers that arrive during slot k. It is assumed that the service of a customer can start only at a slot boundary. Owing to the synchronous type of service, a customer cannot be put into service in the slot that it has arrived. Its service can start no earlier than at the beginning of the next full slot. The service times of customers are assumed to be a one slot. The system has one buffer of infinite capacity to accommodate arriving customers. Customers are served in FIFO (First In First Out) order. The exact location of arrival instants within the slot length is not specified here. It is even irrelevant as long as the system is observed at slot boundaries only.

It is assumed that the server is subject to breakdowns and repairs. The server broken down starts immediately the repair process. The lifetime of the server is assumed to be geometrically distributed with parameter α, where α is the failure probability that a server breakdown occurs in a slot. The repair times of the server follow a geometrical distribution with parameter β, where β is the repair probability that a repair is completed in a slot. Even when the server breaks down, the system can continue to forward some customers. After every breakdown of the server, up to N customers can be serviced during the repair time. The interarrival times, the lifetimes, and the repair times are assumed to be mutually independent of each other.

Let $M(k)$ be the number of customers in the system at the beginning of slot k. Let $S(k)$ be the server state at the beginning of slot k: if the server is under repair and the system has forwarded n customers after the server's breakdown, then $S(k) = N - n$ for $n = 0, 1, 2, ..., N$; if the server is normal, then $S(k) = N + 1$. Then $\{(M(k), S(k)), k = 0, 1, 2, ...\}$ is a Markovian process with state space $\{(m, n), m = 0, 1, 2, ..., n = 0, 1, ..., N + 1\}$. We have the transition matrix of the process $\{(M(k), S(k)), k = 0, 1, 2, ...\}$:

$$\mathbf{P} = \begin{pmatrix} a_0\mathbf{B} & a_1\mathbf{B} & a_2\mathbf{B} & a_3\mathbf{B} & \cdots \\ a_0\mathbf{C} & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & \cdots \\ \mathbf{0} & a_0\mathbf{C} & \mathbf{A}_1 & \mathbf{A}_2 & \cdots \\ \mathbf{0} & \mathbf{0} & a_0\mathbf{C} & \mathbf{A}_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where the matrices $\mathbf{A}_i$, $\mathbf{B}$, and $\mathbf{C}$ are $(N+2) \times (N+2)$ matrices given by

$$\mathbf{A}_i = \begin{pmatrix} a_{i-1}(1-\beta) & 0 & 0 & \cdots & 0 & 0 & a_{i-1}\beta \\ a_i(1-\beta) & 0 & 0 & \cdots & 0 & 0 & a_i\beta \\ 0 & a_i(1-\beta) & 0 & \cdots & 0 & 0 & a_i\beta \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & a_i\beta \\ 0 & 0 & 0 & \cdots & a_i(1-\beta) & 0 & a_i\beta \\ 0 & 0 & 0 & \cdots & 0 & a_i\alpha & a_i(1-\alpha) \end{pmatrix}.$$

$$\mathbf{B} = \begin{pmatrix} 1-\beta & 0 & 0 & \cdots & 0 & 0 & \beta \\ 0 & 1-\beta & 0 & \cdots & 0 & 0 & \beta \\ 0 & 0 & 1-\beta & \cdots & 0 & 0 & \beta \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1-\beta & 0 & \beta \\ 0 & 0 & 0 & \cdots & 0 & 1-\beta & \beta \\ 0 & 0 & 0 & \cdots & 0 & \alpha & 1-\alpha \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1-\beta & 0 & 0 & \cdots & 0 & 0 & \beta \\ 0 & 1-\beta & 0 & \cdots & 0 & 0 & \beta \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \beta \\ 0 & 0 & 0 & \cdots & 1-\beta & 0 & \beta \\ 0 & 0 & 0 & \cdots & 0 & \alpha & 1-\alpha \end{pmatrix}.$$

3. Steady state analysis

A necessary and sufficient condition to ensure the existence for the stationary probability vector of the process $\{(M(k), S(k)), k = 0, 1, 2, ...\}$ is provided. Let

$$\mathbf{A} \equiv a_0\mathbf{C} + \sum_{i=1}^{\infty} \mathbf{A}_i.$$

Then $\mathbf{A}$ is given by

$$\mathbf{A}=\begin{pmatrix} 1-\beta & 0 & 0 & \cdots & 0 & 0 & \beta \\ 1-\beta & 0 & 0 & \cdots & 0 & 0 & \beta \\ 0 & 1-\beta & 0 & \cdots & 0 & 0 & \beta \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \beta \\ 0 & 0 & 0 & \cdots & 1-\beta & 0 & \beta \\ 0 & 0 & 0 & \cdots & 0 & \alpha & 1-\alpha \end{pmatrix},$$

which is finite and irreducible. The stability condition for the system is given [12] by the following inequality

$$\pi\sum_{i=1}^{\infty} i\mathbf{A}_{i+1}\mathbf{e} < \pi a_0\mathbf{C}\mathbf{e},$$

where $\pi \equiv (\pi_0, \pi_1, \pi_2, ..., \pi_{N+1})$ is the stationary probability vector of $\mathbf{A}$ and $\mathbf{e}$ is a column vector whose elements are all equal to 1. Solving the linear equations $\pi\mathbf{A} = \pi$ and $\pi\mathbf{e} = 1$, we get

$$\begin{aligned} \pi_0 &= \frac{\alpha}{\alpha+\beta}(1-\beta)^N, \\ \pi_i &= \frac{\alpha\beta}{\alpha+\beta}(1-\beta)^{N-i}, \quad i = 1, 2, ..., N, \\ \pi_{N+1} &= \frac{\beta}{\alpha+\beta}. \end{aligned}$$

Then, the stability condition of the system is given by

$$\sum_{i=1}^{\infty} ia_i < 1 - \frac{\alpha}{\alpha+\beta}(1-\beta)^N, \tag{1}$$

where the right-hand side of (1) is the probability that the service is available regardless of server breakdown, $\lim_{k\to\infty} \mathrm{P}\{S(k) \neq 0\}$.

It is assumed that the stability condition (1) is satisfied. Let

$$\mathbf{x} \equiv (\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, ...)$$

be the steady-state distribution of the nubmer of customers and the system state of the M/G/1-type process $\{(M(k), S(k)), k = 0, 1, 2, ...\}$, i.e.,

$$\begin{aligned} x_{i,j} &\equiv \lim_{k\to\infty} \mathrm{P}\{M(k) = i, S(k) = j\}, \\ \mathbf{x}_i &\equiv (x_{i,0}, x_{i,1}, ..., x_{i,N+1}). \end{aligned}$$

For the solution of M/G/1-type processes, several algorithms exist [12, 13, 14, 15, 16, 17]. These algorithms starts with the computation of matrix $\mathbf{G}$ as the solution of the following equation:

$$\mathbf{G} = a_0\mathbf{C} + \sum_{i=1}^{\infty} \mathbf{A}_i \mathbf{G}^i. \tag{2}$$

The matrix $\mathbf{G}$ is obtained by solving iteratively the equation (2) or by using the cyclic-reduction algorithm [13]. The stationary probability vector $\mathbf{x}$ is computed recursively using either Ramaswami's recursive formula [14] or its fast FFT version [15]. ETAQA [17] is the other available alternative for the solution of M/G/1-type processes.

4. Mean delay

Using the stationary probability vector presented above, we now determine the mean delay $\mathrm{E}(D)$. Let $d_{i,j}$ be the mean remaining delay of a tagged customer at the beginning of a slot when the number of customers that will be served before the tagged customer is $i-1$ and the server state is j for $i = 1, 2, ...,$ and $j = 0, 1, ..., N+1$. Letting

$$\mathbf{d}_i \equiv \left(d_{i,0}, d_{i,1}, ..., d_{i,N+1}\right)^t,$$

$$\mathbf{F} \equiv \begin{pmatrix} 1-\beta & 0 & \cdots & 0 & \beta \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

we obtain $\mathbf{d}_1 = \mathbf{F}\mathbf{d}_1 + \mathbf{e}$ and $\mathbf{d}_n = \mathbf{C}\mathbf{d}_{n-1} + \mathbf{F}\mathbf{d}_n + \mathbf{e}$ for $n > 2$. Hence,

$$\mathbf{d}_n = \sum_{i=0}^{n-1} \left[(\mathbf{I}-\mathbf{F})^{-1}\mathbf{C}\right]^i (\mathbf{I}-\mathbf{F})^{-1}\mathbf{e}, \quad n \geq 1.$$

Then, the mean delay $\mathrm{E}(D)$ of a tagged customer is obtained by

$$\begin{aligned} \mathrm{E}(D) \;=\; & \sum_{k=1}^{\infty} b(k) \left[\sum_{j=0}^{N} x_{0,j}\left\{\beta d_{k,N+1} + (1-\beta)d_{k,j}\right\}\right. \\ & + x_{0,N+1}\left\{\alpha d_{k,N} + (1-\alpha)d_{k,N+1}\right\} \\ & + \sum_{i=1}^{\infty}\left\{ x_{i,0}\left(\beta d_{k+i,N+1} + (1-\beta)d_{k+i,0}\right)\right. \\ & \quad + x_{i,N+1}\left(\alpha d_{k+i-1,N} + (1-\alpha)d_{k+i-1,N+1}\right) \\ & \quad \left.\left. + \sum_{j=1}^{N} x_{i,j}\left(\beta d_{k+i-1,N+1} + (1-\beta)d_{k+i-1,j-1}\right)\right\}\right], \end{aligned}$$

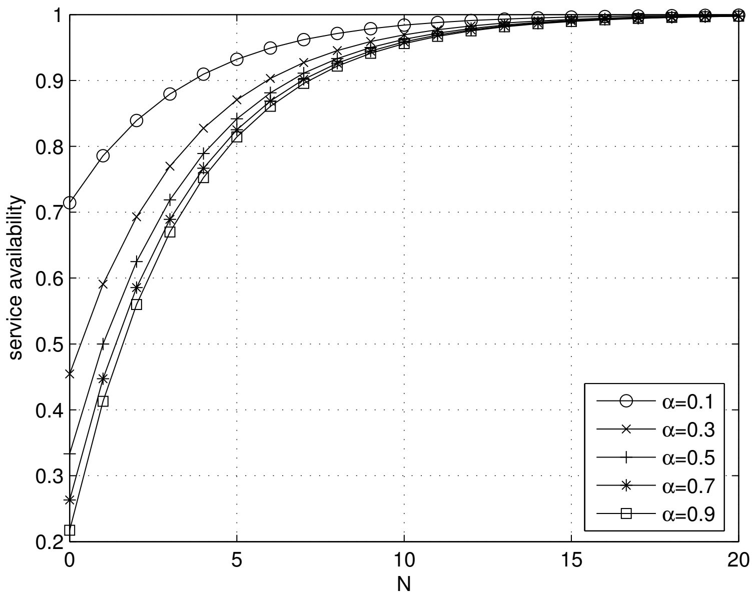

FIGURE 1. Service availability when $\beta = 0.25$.

where $b(k)$ is the probability of a tagged customer being in the kth position of its batch, which is given by

$$b(k) = \frac{1}{\sum_{i=1}^{\infty} i a_i} \sum_{j=k}^{\infty} a_j .$$

5. Numerical examples

Some numerical examples are presented. Note that the system with $N = 0$ corresponds to the system without nonstop forwarding and the system with $N = \infty$ corresponds to the system without server breakdown.

Figure 1 reveals the effect of the value N on the service availability, given by the right-hand side of (1), when $\beta = 0.25$. From the figure, we can see that, as the value N increases, the service availability also increases and converges to 1, which is the availability of the system without server breakdown. Moreover, to achieve target availability, the system with the larger failure probability α requires the larger N. For example, to achieve 99% availability, the system with $\alpha = 0.1$ requires $N \geq 12$, while the system with $\alpha = 0.7$ requires $N \geq 16$.

The results of mean delay are shown in Figures 2, 3, and 4. We consider the case that the number of customer arrivals in a slot has a Poisson distribution

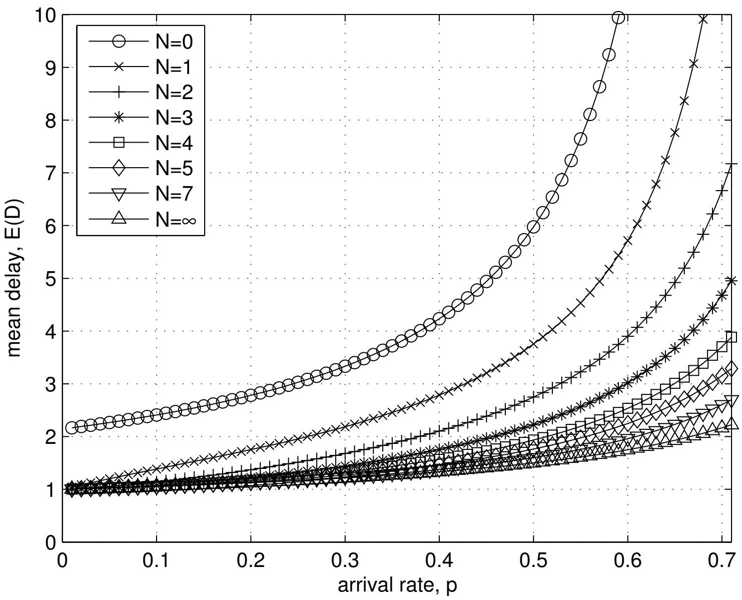

FIGURE 2. Mean delay E(D) vs. arrival rate p when $\alpha = 0.1$ and $\beta = 0.25$.

with rate p. i.e.,

$$a_k = e^{-p}\frac{p^k}{k!}.$$

We choose $\alpha = 0.1$ and $\beta = 0.25$ for Figure 2, $\beta = 0.25$ and $p = 0.5$ for Figure 3, and $\alpha = 0.1$ and $p = 0.5$ for Figure 4. As shown in the figures, the mean delay increases as the arrival rate p increases, as the failure probability α increases, or as the repair probability β decreases. Moreover, while the mean delay decreases and converges to the mean delay of the system without server breakdown as the value of N increases.

6. Conclusion

This paper considered a variant of a queueing system with unreliable server: queueing system with N-limited nonstop forwarding. In this queueing system, the server starts immediately the repair process whenever the server breaks down. Despite the server breakdown, up to N customers can be serviced during the repair time. It can be used to model an assembly line consisting of several automatic stations and a manual backup station. Within the framework of $Geo^X/D/1$ queue with N-limited nonstop forwarding, the stability condition

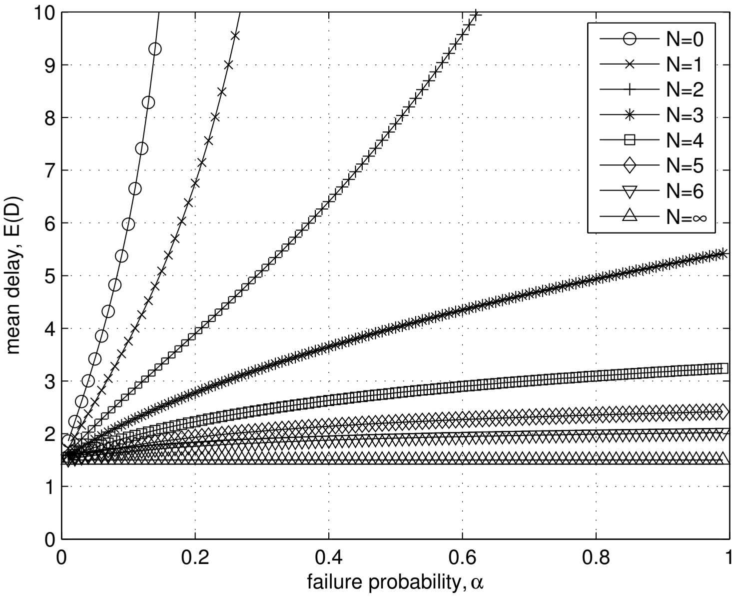

FIGURE 3. Mean delay E(D) vs. failure probability α when $\beta = 0.25$ and $p = 0.5$

and the service availability were given. The matrix analytic approach was used to obtain the steady-state distribution of the number of customers and system state. The mean delay was also derived. Finally, some numerical examples were provided.

References

[1] W. L. Wang and G. Q. Xu, *The well-posedness of an M/G/1 queue with second optional service and server breakdown*, Computers & Mathematics with Applications **57** (2009), no. 5, pp. 223–227.

[2] D. Perry and M. J. M. Posner, *A correlated M/G/1-type queue with randomized server repair and maintenance modes*, Operations Research Letters **26** (2000), no. 3, pp. 137–147.

[3] N. P. Sherman and J. P. Kharoufeh, *An M/M/1 retrial queue with unreliable server*, Operations Research Letters **34** (2006), no. 6, pp. 697–705.

[4] A. Econimou and S. Kanta, *Equilibrium balking strategies in the observable single-server queue with breakdowns and repairs*, Operations Research Letters **36** (2008), no. 6, pp. 696–699.

[5] Y. Lee, *Discrete-time bulk-service queue with Markovian service interruption and probabilistic bulk size*, J. Appl. Math. & Informatics **28** (2010), no. 1–2, pp. 275–282.

[6] G. Falin, *An M/G/1 retrial queue with an unreliable server and general repair times*, Perform. Evaluation **67** (2010), no. 7, pp. 569–582.

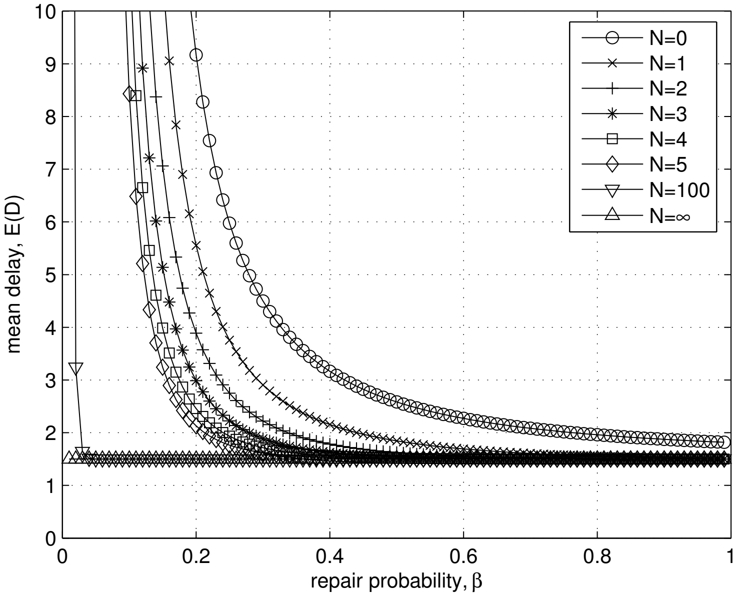

FIGURE 4. Mean delay E(D) vs. repair probability β when $\alpha = 0.1$ and $p = 0.5$

[7] F. Zhang and Z. Zhu, *A discrete-time unreliable Geo/G/1 retrial queue with balking customers, second optional service, and general retrial times*, Mathematical Problems in Engineering **2013** (2013), pp. 1–12.

[8] C. T. Do, N. H. Tran, C. S. Hong, S. Lee, J.-J. Lee, and W. Lee, *A lightweight algorithm for probability-based spectrum decision scheme in multiple channels cognitive radio networks*, IEEE Commun. Lett. **17** (2013), no. 3, pp. 509–512.

[9] G. Choudhury and J.-C. Ke, *An unreliable retrial queue with delaying repair and general retrial times under Bernoulli vacation schedule*, Applied Mathematics and Computation **230** (2014), pp. 436–450.

[10] N. Gharbi, B. Nemmouchi, L. Mokdad, and J. Ben-Othman, *The impact of breakdowns disciplines and repeated attempts on performance of small cell networks*, J. Computational Science **5** (2014), no. 4, pp. 633–644.

[11] Shin, F., Ram, B., Gupta, A., Yu, X., Menassa, R., *A decision tool for assembly line breakdown action*, Proc. 2004 Winter Simulation Conference, 2004, pp. 1122–1127.

[12] M. F. Neuts, *Structured Stochastic Matrices of M/G/1 Type and Their Applications*, Marcel Dekker, New York, NY, 1989.

[13] D. A. Bini, B. Meini, and V. Ramaswami, *Analyzing M/G/1 paradigms through QBDs: the role of the block structure in computing the matrix G*, In G. Latouche and P. Taylor, editors, *Advances in Matrix-Analytic Methods for Stochastic Models*, Notable Publications Inc. NJ, 2000, pp. 87–97.

[14] V. Ramaswami, *A stable recursion for the steady state vector in markov chains of M/G/1 type*, Comm. Statist. Stochastic Models **4** (1988), pp. 183–263.

[15] B. Meini, *An improved FFT-based version of Ramaswami's formular*, Comm. Statist. Stochastic Models **13** (1997), pp. 223–238.

[16] B. Meini, *Solving M/G/1 type Markov chains: Recent advances and applications*, Comm. Statist. Stochastic Models **14** (1988), no. 1–2, pp. 479–496.

[17] A. Riska and E. Smirni, *An exact aggregation approach for M/G/1-type Markov chains*, Proc. ACM International Conference on Measurement and Modeling of Computer Systems (ACM SIGMETRICS '02), Marina Del Rey, CA, 2002, pp. 86–96.

[18] J. Walraevens, D. Fiems, and H. Bruneel, *Performance analysis of priority queueing system in discrete time*, Network Performance Engineering, Lecture Notes in Computer Science **5233** (2011), pp 203–232.

Yutae Lee
Department of Information and Communication Engineering
Dongeui University
Busan 614-714, Korea
E-mail address: `ylee@deu.ac.kr`