STATISTICAL MECHANICS APPROACH FOR STEADY-STATE ANALYSIS IN M/M/S QUEUEING SYSTEM WITH BALKING

IKUO ARIZONO∗
Okayama University
3-1-1 Tsushima-naka, Okayama 700-8530, Japan

YASUHIKO TAKEMOTO
Kindai University
3-4-1 Kowakae, Higashi-Osaka, Osaka, 577-8502, Japan

(Communicated by Shoji Kasahara)

Abstract. Behavior that a customer who has just arrived at a crowded queueing system leaves without joining the queue is known as the phenomenon of balking. Queueing systems with balking have been studied continually as one of significant subjects. In this paper, the theoretical approach for the steady-state analysis of the Markovian queueing systems with balking is considered based on the concept of the statistical mechanics. Here, it can be easily seen that the strength of balking is not constant but various in each queueing systems. Note that the strength of balking means how degree a customer who has just arrived at a crowded queueing system leaves without joining the queue. In our approach, under considering the difference of the strength of balking for each queueing systems, we have proposed a statistical mechanics model for analyzing the M/M/s queueing system with balking by introducing a parameter influencing the strength of balking. Further, we define a procedure for estimating the model parameter influencing the strength of balking. In addition, we consider a method of improving the performance of the M/M/s queueing system with balking by utilizing the statistical mechanics approach.

1. Introduction. In the queueing theory for analyzing congestion phenomena of systems, the behavior of balking has been sometimes considered, where the phenomenon of balking means the behavior that a customer who has just arrived at a crowded queueing system leaves without joining the queue. It is common recognition that such a phenomenon as leaving from each queueing system based on balking results in economic and mental losses [4]. And, the phenomenon of balking has been treated as one of stochastic factors in the steady-state analysis of queueing systems. Therefore, until now, the steady-state analysis of queueing systems under
the consideration of balking has been discussed continually. For example, the respective Markovian queueing systems with balking have been considered in Montazer-Haghighi et al. [8], Abou-El-Ata and Hariri [1], El-Sherbiny [5], Sztrik [11], and Jain et al. [7].

In the traditional analysis of queueing systems under the consideration of balking, it is typically assumed that any arrival rates $\lambda_n$ for the number of customers $n$ in the system are known and specified. For reference, Montazer-Haghighi et al. [8] have supposed that the rate of balking is constant when all servers are busy. On the other hand, Abou-El-Ata and Hariri [1], El-Sherbiny [5] and Sztrik [11] have expressed any arrival rates by a function depending on the number of customers $n$ in the system. In the traditional investigations, the steady-state probability distribution of queueing systems has been obtained under the specified arrival rates $\lambda_n$ as a function of the number of customers $n$ in the system. That is, the traditional investigations have been limited to the case where the information about every arrival rates are complete. This is the reason why the traditional investigations are based on the birth-and-death stochastic process. Conversely, we should note that the steady-state probability distribution cannot be derived unless all of arrival rates $\lambda_n$ are given by the specified function for the number of customers $n$.

It is natural that the change of arrival rates is derived from balking. In the traditional papers, the degree of the change in arrival rates has been given by only depending on the number of customers in the system. However, we can assume that the change of arrival rates is influenced by the characteristics and circumstances in the system in addition to the number of customers in the system. In concrete, the change of arrival rates should also depend on the characteristics and circumstances in the system, such as an item sold at a store and equipment of a waiting room. In this paper, the phenomenon that a customer who has just arrived at a crowded queueing system leaves without joining the queue is considered as the phenomenon occurred by balking. Then, we consider that the strength of balking brings the change of arrival rates. Thus, it is assumed that the phenomenon of balking occurs due to the influence of not only the number of customers in the system but also some factors such as the characteristics and circumstances of the system. As a consequence, since the change of arrival rates depends on the number of customers in the system and the characteristics and circumstances of the system, we should assume the strength of balking depending on such factors. Then, as factors of reducing arrival rates by balking, we consider the number of customers in the system, that is, system-common factor, and while the characteristics and circumstances of the system, that is, system-specific factor. In addition, we would call the degree of the influence reducing arrival rates by the system-specific factor especially “the strength of balking”, and introduce a parameter for indicating the strength of balking to construct our analysis model.

Based on the mentioned above, in this paper, we would first propose a model for analyzing the steady-state probability distribution based on the system-common factor, i.e, the number of customers in the system and the system-specific factor, i.e, the strength of balking. This analysis model would be presupposed to estimate the strength of balking through practical observation. Hence, secondly, the way of estimating the strength of balking through practical observation would be also considered. Note that these features in our considerations are different from the traditional investigations.
In this paper, based on the concept of the statistical mechanics, we propose the theoretical approach for the steady-state analysis of the Markovian queueing systems with balking. Concretely, a model for analyzing the steady-state probability distribution of the Markovian queueing systems with balking would be developed based on the strength of balking. Then, as mentioned above, we consider the decline of arrival rates by the strength of balking in addition to the number of customers in the system in the case of constructing the analysis model. In such a case, we introduce a parameter expressing the strength of balking for the factor to reduce the arrival rates. As a result, a statistical mechanics model for analyzing the M/M/s queueing system with balking would like to be proposed under the consideration of the strength of balking.

The concept of the statistical mechanics is explained based on both the entropy and potential energy [3]. The entropy is interpreted as the criterion for evaluating the microscopic irregularity of systems, and the potential energy is interpreted as the criterion for evaluating the instability of systems. In the statistical mechanics, the steady-state probability distribution can be explained based on the balance of the entropic force and the potential energy force defined by the change of states in the system.

In this paper, first, we construct the statistical mechanics model for analyzing the steady-state probability distribution of the M/M/s queueing system with balking under the consideration of the strength of balking. For the validity of our proposal, we verify that the constructed statistical mechanics model can consistently involve some existing analysis results for the steady-state probability distribution for the traditional Markovian queueing systems with and without balking as respective special cases. As a consequence, it is confirmed that the constructed statistical mechanics model in this paper have the correctness for the analysis of Markovian queueing models. Furthermore, secondly, the procedure for estimating the model parameter expressing the strength of balking in the constructed analysis model for the M/M/s queueing system with balking is explained. Through some numerical analysis, we show the usefulness of the proposed model based on the statistical mechanics. In addition, thirdly, two methods for optimizing the total service rate in the M/M/s queueing system system with balking is considered through some numerical analyses. Concretely, one is a method of improving the total service rate by increasing the number of servers s and the other is a method of improving the total service rate by increasing the service rate $\mu$ in each server.

The rest of the paper is as follows. In section 2, we explain the outline of the statistical mechanics. In section 3, the statistical mechanics model to analyze the steady-state probability distribution of the M/M/s queueing system with balking is constructed. It is demonstrated that the constructed statistical mechanics model is applicable to various types of Markovian queueing systems with and without balking. Further, in section 4, for practical usage, the estimation of the model parameter expressing the strength of balking in the constructed analysis model for the M/M/s queueing system with balking is investigated. In addition, in section 5, the optimization of the M/M/s queueing system with balking is considered by using the result of the estimation of the model parameter expressing the strength of balking. Finally, section 6 concludes this paper.

2. Concept of statistical mechanics. In this section, the outline of the statistical mechanics is explained. Suppose the probability of state $n$ as $P_n$, and define the
probability vector $\vec{P}$ as $\vec{P} = (P_0, P_1, P_2, \ldots, P_n, \ldots)$. Then, the entropy $H(\vec{P})$ of the stochastic system characterized by the probability vector $\vec{P}$ is given by

$$H(\vec{P}) \equiv -\sum_{n=0}^{\infty} P_n \ln P_n. \quad (1)$$

Further, the force of changing the entropy of the system is called “the entropic force” in the concept of the statistical mechanics. The entropic force can be defined as $\partial H(\vec{P})/\partial \vec{P}$ in response to the change of the state probability distribution of the system.

On one hand, the potential energy $U(\vec{P})$ is defined as

$$U(\vec{P}) \equiv \sum_{n=0}^{\infty} g(n) P_n. \quad (2)$$

where $g(n)$ describes a potential function defined by state $n$ of the system. The absolute value of $g(n)$ means the magnitude of the force, and the sign of $g(n)$ indicates the direction of the force. Further, the force of changing the potential energy $U(\vec{P})$ is called “the energy force”. As with the entropic force, the energy force can be defined as $\partial U(\vec{P})/\partial \vec{P}$ in response to the change of the state probability distribution of the system.

The concept of the statistical mechanics mentions that the balance between the entropic force and the energy force lead the system to a steady-state. Therefore, the steady-state probability distribution of the system can be derived from the relationship of the balance between the entropic force and the energy force as follows:

$$\frac{\partial}{\partial \vec{P}} H(\vec{P}) = \omega_1 \frac{\partial}{\partial \vec{P}} U(\vec{P}), \quad (3)$$

where $\omega_1$ is introduced as an unknown coefficient to unify the scales of the entropic force and the energy force. And then, it is obvious that the following equation should be satisfied:

$$\sum_{n=0}^{\infty} P_n = 1. \quad (4)$$

3. Construction of statistical mechanics model for M/M/s queueing system with balking. In this section, we address the statistical mechanics approach to analyze the steady-state probability distribution of an M/M/s queueing system with balking. Suppose an M/M/s queueing system described on the service rate $\mu$ and the natural arrival rate $\lambda$ without balking, where $\mu$ and $\lambda$ are given. Then, the traffic intensity in the M/M/s queueing system is defined as $\rho = \lambda/(s\mu)$. Further, the relationship of $\rho < 1.0$ presents the prerequisite condition that the steady-state probability distribution exists at the M/M/s queueing system without balking. Note that such a condition is not necessarily required for the existence of the steady-state probability distribution under the phenomenon of balking.

In the M/M/s queueing system with balking, a customer who arrived at the system leaves without joining the queue depending on the number of customers $n$ in the system. Then, we describe the actual arrival rate under the situation that the number of customers in the system is $n$ as $\lambda_n$. In this case, it is naturally assumed that the phenomenon of balking does not occur in the situation that there is at least one available vacant server. Hence, the arrival rate $\lambda_n$ for $0 \leq n \leq s - 1$ can be assumed as $\lambda_n = \lambda$. On the other hand, it is also natural to suppose
the relation of \( \lambda_n \geq \lambda_{n+1} \) for \( n \geq s \) as common sense under balking. Then, we define the probability of state \( n \) as \( \lambda_n \) and the probability vector \( \bar{P} \) is presented as \( \bar{P} = (P_0, P_1, P_2, \cdots, P_n, \cdots) \). Additionally, by referring Arizono et al. \[2\], whole states in the M/M/s queueing system are divided into two groups for \( 0 \leq n < s \) and \( s \leq n \), and further, each state \( n \) is divided into some quasi-states. In the transition from state \( n+1 \) to state \( n \) in the group of \( 0 \leq n < s \), the total service rate varies depending on state \( n \), that is, the total service rate in the transition from state \( n+1 \) to state \( n \) is given as \((n+1)\mu\). In such a case, state \( n \) of \( 0 \leq n < s \) can be divided into quasi-states \((n,i)\) for \( 1 \leq i \leq s!/(s-i)! \) consisting of \( s!/(s-i)! \) cases by considering the number of cases in the transitions from state \( s \) to state \( n \), \( 0 \leq n < s \). Thus, the state probability \( P_n \) in the case of \( 0 \leq n < s \) is represented as

\[
P_n = \sum_{i=1}^{s!/(s-i)!} q_{n,i}, \quad 0 \leq n < s,
\]

where \( q_{n,i} \) means the state probability of the quasi-state \((n,i)\), \( 1 \leq i \leq s!/(s-i)! \). On the other hand, in the group of \( s \leq n \), the service rate in the transition from state \( n+1 \) to state \( n \) is constant as \( s\mu \). And, the transition from state \( n+1 \) to state \( n \) occurs if one of \( s \) servers completes its service. Then, we divide state \( n \) in the group of \( s \leq n \) into the quasi-states \((n,j)\) for \( 1 \leq j \leq s \) consisting of \( s \) cases by considering the number of cases in the transitions from state \( n+1 \) to state \( n \). Thus, the state probability \( P_n \) in the case of \( n \geq s \) is supposed as

\[
P_n = \sum_{j=1}^{s} q_{n,j}, \quad n \geq s.
\]

where \( q_{n,j} \) means the state probability of the quasi-state \((n,j)\), \( 1 \leq j \leq s \).

Furthermore, since the states of the M/M/s queueing system are divided into two groups of \( 0 \leq n < s \) and \( s \leq n \), we should define two kinds of the potential. Then, in the state of the group of \( 0 \leq n < s \), a customer arriving at the M/M/s queueing system does not have to wait to receive the service because there are some available vacant servers. Therefore, we can consider that the phenomenon of balking does not occur if a customer has arrived at the system in the case of \( 0 \leq n < s \). Then, assume the potential function \( g_1(n) \) that is interpreted as the function indicating the potential regarding the increase of busy servers. Hence, we can assume \( g_1(0) = 0.0 \), and define \( g_1(1) = 1.0 \) as the unit amount of the potential regarding the increase of busy servers. On one hand, in the case of \( n \geq s \), a customer who has arrived newly at the M/M/s queueing system must wait by necessity. In this case, although the number of busy servers is not increased by the customer who has just arrived at the M/M/s queueing system, the number of waiting customers is increased.

Then, we suppose that the potential regarding the increase of busy servers does not change as \( g_1(n) = g_1(s-1) \) for \( n \geq s \). And, separately from the potential \( g_1(n) \) that does not change in the case of \( n \geq s \), we assume the potential function \( g_2(n) \) that increases as the number of waiting customers. Then, we define \( g_2(s) = 1.0 \) as the unit amount of the potential regarding the increase of customers waiting for service. Remark that it is not necessary to consider the influence of balking in the potential function \( g_1(n) \), because the phenomenon of balking does not occur in the situation that there is at least one available vacant server, that is, in the case of \( 0 \leq n \leq s-1 \). Whereas, in the case of \( n \geq s \), it is necessary to consider the influence
of balking, i.e., the strength of balking, because there is no available vacant server, and the phenomenon of balking occurs.

In consequence, based on the quasi-state probabilities mentioned above, the entropy in the M/M/s queueing system can be defined as

\[ H(\vec{P}) = -\sum_{n=0}^{s-1} \frac{s^n}{n!} \ln^{s-n} \sum_{i=1}^{s-1} q_{n,i} \ln q_{n,i}. \]

Further, based on the potential function \( g_1(n) \), the potential energy \( U_1(\vec{P}) \) due to the increase of busy servers can be defined as follows:

\[ U_1(\vec{P}) = \sum_{n=0}^{s-1} \frac{s^n}{n!} g_1(n) q_{n,i} + \sum_{n=s}^{\infty} \sum_{j=1}^{s} g_1(s-1) q_{n,j}. \]

And, the potential energy \( U_2(\vec{P}) \) due to the increase of customers waiting for service can be defined as

\[ U_2(\vec{P}) = \sum_{n=s}^{\infty} \sum_{j=1}^{s} g_2(n) q_{n,j}. \]

Note that the potential energy \( U_1(\vec{P}) \) is defined according to the number of busy servers as equation (8) based on the quasi-state probabilities \( q_{n,i} \) and potential \( g_1(n) \). The first term expresses the potential energy according to the number of busy servers in the case of \( 0 \leq n < s \) and the second term expresses the potential energy in the case of no vacant server. On one hand, the potential energy \( U_2(\vec{P}) \) is defined according to the number of waiting customers as equation (9) based on the quasi-state probability \( q_{n,j} \) and potential \( g_2(n) \). It is sure that the potential energy \( U_2(\vec{P}) \) according to the number of waiting customers is defined in just only the cases that \( n \geq s \). And then, note that the potential energy \( U_2(\vec{P}) \) is defined starting from the state of \( n = s \) which there is no vacant server. The potential energy in the whole system is defined as the sum of the potential energies \( U_1(\vec{P}) \) and \( U_2(\vec{P}) \) in equations (8) and (9).

In the concept of the statistical mechanics, the steady-state is induced by the balance of the change in the forces of the entropy and potential energy. Here, the change of the state in the M/M/s queueing system can be shown by the change of the probability distribution. Therefore, the entropic force and energy force are associated with the change of the probability distribution of the M/M/s queueing system. The balance of the change in the forces of the entropy and potential energy can be evaluated by the partial differential regarding the probability vector \( \vec{P} \) for individual functions \( H(\vec{P}), U_1(\vec{P}) \) and \( U_2(\vec{P}) \). Consequently, based on the concept of the statistical mechanics, the steady-state in the M/M/s queueing system is explained as the following relationship:

\[ \frac{\partial}{\partial \vec{P}} H(\vec{P}) = \omega_1 \frac{\partial}{\partial \vec{P}} U_1(\vec{P}) + \omega_2 \frac{\partial}{\partial \vec{P}} U_2(\vec{P}), \]

where \( \omega_1 \) and \( \omega_2 \) are given as undetermined multipliers according to the respective dimensions (units) of the energy forces associated with \( U_1(\vec{P}) \) and \( U_2(\vec{P}) \) to the dimension (unit) of the entropic force associated with \( H(\vec{P}) \). In addition, because
the queueing system is a probabilistic system, we consider the following constraint:

\[
\sum_{n=0}^{\infty} P_n = \sum_{n=0}^{s-1} \frac{s!}{n!} P_0 + \sum_{n=s}^{\infty} q_{n,i} + \sum_{n=s}^{\infty} q_{n,j} = 1.
\] (11)

It is obvious that equation (11) is derived from the necessary condition for any probability system. Further, equation (10) presents that the steady-state can be induced based on the balance of the change in the forces of the entropy and potential energy under the concept of the statistical mechanics. And then, it is employed as the necessary condition to construct the statistical mechanics model to analyze the steady-state in the M/M/s queueing system with balking. That is, the statistical mechanics model to analyze the steady-state in the M/M/s queueing system with balking has been constructed as follows:

\[
\Lambda(\vec{P}, \omega_1, \omega_2, \omega_3) = H(\vec{P}) - \omega_1 U_1(\vec{P}) - \omega_2 U_2(\vec{P}) - \omega_3 \left( \sum_{n=0}^{s-1} \frac{s!}{n!} P_0 + \sum_{n=s}^{\infty} q_{n,i} + \sum_{n=s}^{\infty} q_{n,j} - 1 \right),
\]

where \( \omega_3 \) mean the Lagrange multipliers. Note that this model is analogous to the method of the equilibrium statistical physics where the (Gibbs) free energy is minimized [6, 10]. And then, \( \omega_1 \) and \( \omega_2 \) introduced as unknown coefficients to unify the scales of the entropic force and the energy force are also operated in the same way as the Lagrange multiplier \( \omega_3 \) in equation (12). Based on equation (12), we can obtain the steady-state probability distribution of the M/M/s queueing system with balking such as \( P_n \), \( q_{n,i} \) and \( q_{n,j} \). As mentioned above, equation (12) is a model to be solved for deriving the steady-state probabilities of the M/M/s queueing system with balking and equations (10) and (11) are first-order conditions regarding equation (12) as a Lagrange function.

As a result, we have derived the steady-state probability distribution from equation (12) as

\[
P_n = \begin{cases} 
\frac{a^g_1(n)}{n!} P_0, & 0 \leq n < s, \\
\frac{s!}{a^g_1(s-1)} \rho^2_2(n) P_0, & s \leq n.
\end{cases}
\] (13)

Here, \( a \) is defined as \( \lambda/\mu \), and \( P_0 \) is represented as

\[
P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{a^g_1(n)}{n!} + \sum_{n=s}^{\infty} \frac{s!}{a^g_1(s-1)} \rho^2_2(n)}.
\] (14)

Furthermore, the actual arrival rate \( \lambda_n \) under the steady-state in the statistical mechanics model is derived as

\[
\lambda_n = \begin{cases} 
\lambda a^g_1(n+1) - a^g_1(n) - 1, & 0 \leq n < s, \\
\lambda \rho^2_2(n+1) - \rho^2_2(n) - 1, & s \leq n.
\end{cases}
\] (15)

The derivation process of equations (13)-(15) is given in Appendix.

As you can see from equation (15), in the case of \( 0 \leq n < s \), the arrival rate under \( g_1(n) = n \) is presented as \( \lambda_n = \lambda \). So, the potential function \( g_1(n) \) regarding the increase of busy servers has been defined as follows:

\[
g_1(n) = \begin{cases} 
n, & 0 \leq n < s, \\
s - 1, & s \leq n.
\end{cases}
\] (16)
Further, we construct that the statistical mechanics model for the M/M/s queueing system with balking can reproduce the traditional results of the steady-state probability distribution of the M/M/s queueing systems (including the M/M/1 queueing system) without balking and the typical result of the steady-state probability distribution of the M/M/1 queueing system with balking.

Here, it is well-known that the steady-state probability distribution of the M/M/s queueing systems (including the M/M/1 queueing system) without balking is given as

\[
\lambda_n = \lambda, \quad 0 \leq n,
\]

\[
P_n = \begin{cases} 
\frac{\lambda^n}{n!} P_0, & 0 \leq n < s, \\
\frac{\lambda^n}{n!} \rho^{n-s+1} P_0, & s \leq n,
\end{cases}
\]

\[
P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{\lambda^n}{n!} + \sum_{n=s}^{\infty} \frac{\lambda^n}{n!} \rho^{n-s+1}}.
\]

Note that the steady-state probability distribution in the M/M/1 queueing system without balking can be derived as follows:

\[
P_n = (1 - \rho) \rho^n.
\]

Furthermore, in the traditional papers for the M/M/1 queueing system with balking by Abou-El-Ata and Hariri [1], El-Sherbiny [5], Natvig [9], and Sztrik [11], the arrival rate \( \lambda_n \) has been typically assumed as follows:

\[
\lambda_n = \lambda_n + 1.
\]

Then, under the arrival rate in equation (21), the steady-state probability \( P_n \) in the M/M/1 queueing system with balking has been obtained as

\[
P_n = \rho^n e^{-\rho}.
\]

It is easily found that equation (22) is the probability function of the Poisson distribution. Equation (21) has been employed frequently without justifiable reason. Although there is no justifiable reason to use the arrival rate presented by equation (21), it seems that the mathematically preferable conclusion that the Poisson distribution can be derived might be the reason why equation (21) has been employed frequently.

In constructing the statistical mechanics model, we consider that the constructed statistical mechanics model for the M/M/s queueing system with balking has to consistently involve the past analysis results for the steady-state probability distribution for the traditional Markovian queueing systems with and without balking. Hence, the potential function \( g_2(n) \) regarding the increase of waiting customers has been defined as follows:

\[
g_2(n) = (n - s + 1) - \frac{r \ln(n - s + 1)!}{\ln \rho}, \quad s \leq n.
\]

where \( r \geq 0 \) denotes the parameter influencing the strength of balking. Then, we can obtain the following derivation:

\[
\lambda \rho^{g_2(n+1) - g_2(n) - 1} = \frac{\lambda}{(n - s + 2)^r}.
\]
Therefore, we can understand that the potential function \( g_2(n) \) in equation (23) is expansively applicable in the case of \( \rho = 1.0 \).

Under the potential functions \( g_1(n) \) and \( g_2(n) \) defined respectively as equations (16) and (23), the arrival rate \( \lambda_n \) can be rewritten as

\[
\lambda_n = \begin{cases} 
\lambda, & 0 \leq n < s, \\
\frac{\lambda}{(n-s+2)r}, & s \leq n. 
\end{cases}
\] (25)

Note that the arrival rate \( \lambda_n \) is not dependent to the service rate \( \mu \) from equation (25). Moreover, the steady-state probability distribution is presented as follows:

\[
P_n = \begin{cases} 
\frac{a^n}{s!} P_0, & 0 \leq n < s, \\
\frac{a^{s-1} \rho^{n-s+1}}{(n-s+1)!} P_0, & s \leq n, 
\end{cases}
\] (26)

and

\[
P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{a^n}{s!} + \sum_{n=s}^{\infty} \frac{a^{s-1} \rho^{n-s+1}}{(n-s+1)!}}.
\] (27)

From equations (26) and (27), it is seen that the steady-state probability distribution for the M/M/s queueing system with balking has been presented under the consideration of the model parameter \( r \) influencing the strength of balking. Additionally, note that equations (26) and (27) have included some existing results as special cases. For example, in the case of \( r = 0.0 \), the well-known traditional results of the M/M/s queueing system without balking in equations (18) and (19) have been reproduced by equations (26) and (27) based on the constructed statistical mechanics model. Further, under the condition of \( r = 1.0 \) and \( s = 1 \), we can easily obtain equations (21) and (22) in the typical M/M/1 queueing system with balking from equations (25)-(27). In this way, we have successfully constructed the statistical mechanics model for the M/M/s queueing system with balking under the consideration of the model parameter \( r \) influencing the strength of bulking. As a consequence, the steady-state probability distribution for the M/M/s queueing system with balking can be evaluated under the consideration of the model parameter \( r \) influencing the strength of balking by using equations (26) and (27).

4. Numerical verifications and applications. We illustrate some numerical results based on the proposed statistical mechanics model for analyzing the steady-state probability distribution in the M/M/s queueing system with balking. In the following, the arrival rate \( \lambda \) is set as \( \lambda = 20.0 \). Some numerical calculations have been performed in the cases of \( s = 1 \) and \( s = 3 \), and under various \( r \) and \( \rho \). These results have been shown in Tables 1-4.

Figure 1 illustrates the relationships between the number of customers \( n \) in the system and the arrival rate \( \lambda_n \) in the case of \( s = 1 \) using equation (25). From Figure 1, since the arrival rate \( \lambda_n \) satisfies the relationship of \( \lambda_n \geq \lambda_{n+1} \), the influence of balking considered in this paper has been successfully explained. In the case of \( r = 0.0 \), we can confirm that balking does not occur since \( \lambda_n \) is constant. On the other hand, it is found that the influence of balking is quite strong in the case of \( s = 1 \) and \( r = 1.0 \). Note that the traditional papers such as Abou-El-Ata and Hariri [1], El-Sherbiny [5], Natvig [9], and Sztrik [11] have treated only deterministically the situation under \( r = 1.0 \). That is, we can say that the situation in the traditional papers has not been always in real situation and so the steady-state probability distribution in the traditional papers has corresponded to only a specific situation.
Table 1. The arrival rate $\lambda_n$ against various $r$ in the case of $s = 1$

| $n$ | $r = 0.00$ | $r = 0.25$ | $r = 0.50$ | $r = 0.75$ | $r = 1.00$ | $r = 1.25$ |
|-----|------------|------------|------------|------------|------------|------------|
| 0   | 20.00      | 20.00      | 20.00      | 20.00      | 20.00      | 20.00      |
| 1   | 20.00      | 16.82      | 14.14      | 11.89      | 10.00      | 8.41       |
| 2   | 20.00      | 15.20      | 11.55      | 8.77       | 6.67       | 5.07       |
| 3   | 20.00      | 14.11      | 10.00      | 7.07       | 5.00       | 3.54       |
| 4   | 20.00      | 13.37      | 8.94       | 5.98       | 4.00       | 2.67       |
| 5   | 20.00      | 12.78      | 8.16       | 5.22       | 3.33       | 2.13       |
| 6   | 20.00      | 12.30      | 7.56       | 4.65       | 2.86       | 1.76       |
| 7   | 20.00      | 11.89      | 7.07       | 4.20       | 2.50       | 1.49       |
| 8   | 20.00      | 11.55      | 6.67       | 3.85       | 2.22       | 1.28       |
| 9   | 20.00      | 11.25      | 6.32       | 3.56       | 2.00       | 1.12       |
| 10  | 20.00      | 10.98      | 6.03       | 3.31       | 1.82       | 1.00       |

In addition, it is seen that the arrival rate $\lambda_n$ decreases faster by a larger model parameter $r$. This fact means that the influence of balking becomes stronger when the model parameter $r$ is large. Further, in Figure 2, some steady-state probability

Table 2. The steady-state probability $P_n$ under some $\rho$ in the case of $s = 1$

| $n$ | $\rho = 0.8$ | $\rho = 1.0$ | $\rho = 1.2$ |
|-----|--------------|--------------|--------------|
| 0   | 0.386257     | 0.288225     | 0.209605     |
| 1   | 0.309606     | 0.288225     | 0.251526     |
| 2   | 0.174800     | 0.203806     | 0.213427     |
| 3   | 0.080737     | 0.117668     | 0.147867     |
| 4   | 0.032295     | 0.058834     | 0.088720     |
| 5   | 0.011554     | 0.026311     | 0.047612     |
| 6   | 0.003774     | 0.010742     | 0.023325     |
| 7   | 0.001141     | 0.004060     | 0.010579     |
| 8   | 0.000323     | 0.001435     | 0.004488     |
| 9   | 0.000086     | 0.000478     | 0.001795     |
| 10  | 0.000022     | 0.000151     | 0.000681     |

Table 3. The arrival rate $\lambda_n$ against various $r$ in the case of $s = 3$

| $n$ | $r = 0.00$ | $r = 0.25$ | $r = 0.50$ | $r = 0.75$ | $r = 1.00$ | $r = 1.25$ |
|-----|------------|------------|------------|------------|------------|------------|
| 0   | 20.00      | 20.00      | 20.00      | 20.00      | 20.00      | 20.00      |
| 1   | 20.00      | 20.00      | 20.00      | 20.00      | 20.00      | 20.00      |
| 2   | 20.00      | 16.82      | 14.14      | 11.89      | 10.00      | 8.41       |
| 3   | 20.00      | 15.20      | 11.55      | 8.77       | 6.67       | 5.07       |
| 4   | 20.00      | 14.11      | 10.00      | 7.07       | 5.00       | 3.54       |
| 5   | 20.00      | 13.37      | 8.94       | 5.98       | 4.00       | 2.67       |
| 6   | 20.00      | 12.78      | 8.16       | 5.22       | 3.33       | 2.13       |
| 7   | 20.00      | 12.30      | 7.56       | 4.65       | 2.86       | 1.76       |
| 8   | 20.00      | 11.89      | 7.07       | 4.20       | 2.50       | 1.49       |
| 9   | 20.00      | 11.55      | 6.67       | 3.85       | 2.22       | 1.28       |
| 10  | 20.00      | 11.25      | 6.32       | 3.56       | 2.00       | 1.12       |

In addition, it is seen that the arrival rate $\lambda_n$ decreases faster by a larger model parameter $r$. This fact means that the influence of balking becomes stronger when the model parameter $r$ is large. Further, in Figure 2, some steady-state probability
Table 4. The steady-state probability $P_n$ under some $\rho$ in the case of $s = 3$

| $n$ | $\rho = 0.8$ | $\rho = 1.0$ | $\rho = 1.2$ |
|-----|-------------|-------------|-------------|
| 0   | 0.092113    | 0.050987    | 0.028157    |
| 1   | 0.221072    | 0.152961    | 0.101365    |
| 2   | 0.265287    | 0.229442    | 0.182457    |
| 3   | 0.212229    | 0.229442    | 0.218948    |
| 4   | 0.120055    | 0.162240    | 0.185784    |
| 5   | 0.055451    | 0.093669    | 0.128715    |
| 6   | 0.022180    | 0.046835    | 0.077229    |
| 7   | 0.007936    | 0.020945    | 0.041445    |
| 8   | 0.002592    | 0.008551    | 0.020304    |
| 9   | 0.000784    | 0.003232    | 0.009209    |
| 10  | 0.000222    | 0.001143    | 0.003907    |

Figure 1. The relationships between the number of customers $n$ and the arrival rate $\lambda_n$ in the case of $s = 1$

distributions in the case of $s = 1$ and $r = 0.5$ are exemplified. From Figure 2, it is seen that the larger the traffic intensity $\rho$ is, the more gently the configuration of the steady-state probability distribution against the number of customers $n$ is changed.

Also, note that we can obtain the steady-state probability distribution of the M/M/1 queueing system with balking in the case of $\rho \geq 1.0$. This results from the reason why the increase in the number of customers reduces the arrival rate and make the actual traffic intensity smaller than 1. See Table 1.

Furthermore, Figure 3 illustrates the relationships between the number of customers $n$ and the actual arrival rate $\lambda_n$ in the case of $s = 3$. From Figure 3, it can be confirmed that balking does not occur in $n < s$. Also, the arrival rate $\lambda_n$ decreases faster by a larger model parameter $r$ in $n \geq s$. Additionally, Figure 4 illustrates some steady-state probability distributions in the case of $s = 3$ and $r = 0.5$. Note
Figure 2. The steady-state probability distributions in the case of \( s = 1 \) and \( r = 0.5 \)

Figure 3. The relationships between the number of customers \( n \) and the arrival rate \( \lambda_n \) in the case of \( s = 3 \)

that we can obtain the steady-state probability distribution of the \( M/M/s \) queueing system under any combination of the number of servers and the traffic intensity.

Through Figures 2 and 4, the influence and effect of the strength of balking can be understood in both situations of \( \rho < 1.0 \) and \( \rho > 1.0 \). Therefore, the usefulness of the analysis model constructed based on the concept of the statistical mechanics is found.

Although the potential functions \( g_1(n) \) and \( g_2(n) \) defined respectively as equations (16) and (23) are not the one and only one, we can investigate conveniently some features in \( M/M/s \) queueing systems with balking based on the constructed
analysis model including the potential functions $g_1(n)$ and $g_2(n)$. Through knowing the model parameter $r$ influencing the strength of balking, we can consider the action for improving the queueing system performance. The following describes the estimation of the model parameter $r$ influencing the strength of balking based on the constructed analysis model for the M/M/s queueing system with balking.

Next, we show the estimation of the model parameter $r$ in actual observation. We employ the following evaluation to estimate the model parameter $r$ influencing the strength of balking:

$$J = \sum_{n=0}^{K} \left( \frac{P^\dagger_n - P^{(r)}_n}{P^{(r)}_n} \right)^2,$$

(28)

where $P^\dagger_n$ denotes the observed state probability and $P^{(r)}_n$ denotes the theoretical steady-state probability under the parameter $r$ based on equation (28) under equations (18) and (19) which are the outcomes in this paper. Note that the observed state probability $P^\dagger_n$ can be obtained based on the observation of the frequency about the number of customers $n$ in the system through the repeated observation. Further, $K$ means the number of states appropriate for estimation, where $K$ is given from the result that some states of small probability $P^{(r)}_n$, $n > K - 1$ are unified. In conclusion, the model parameter $r$ influencing the strength of balking can be estimated optimally as the parameter $r \equiv r^*$ that minimizes $J$. Remark that the evaluation $J$ has been employed based on the concept of the goodness of fit test based on the chi-square statistic.

As an example, we show the procedure of estimating the model parameter $r$ influencing the strength of balking in the M/M/s queueing system with balking. Assume that the internal structure of the system is produced with the parameters setting of $\lambda = 80$, $\mu = 25$, $s = 4$ and $r = 0.75$. In such a case, $\lambda$, $\mu$, $s$ are given as known information, and $r$ is an unknown parameter to be estimated as $r^*$ based on observation data. At first, 300 random numbers are generated by using
the statistical mechanics model presented in this paper as the observed occurrence frequency of state $n$. Let us derive $P^\dagger_n$ based on above 300 random numbers for the observed occurrence frequency of state $n$. Based on the chi-squared method by using the evaluation $J$ in equation (28), the model parameter $r$ influencing the strength of balking can be estimated as the parameter $r \equiv r^*$ that minimizes $J$. Table 5 shows an example of state probabilities as theoretical values $P^\dagger_n$ under the parameters setting $(\lambda, \mu, s, r) = (80, 25, 4, 0.75)$, the probability distribution $P^\dagger_n$ by observation data and the probability distribution $P^\ast_n$ under the identified model based on the optimal estimate $r^* = 0.76$. Note that the chi-squared method depended on the observed probability distribution $P^\dagger_n$ has derived as $r^* = 0.76$ at first. Next, the reproduced probability distribution $P^\ast_n$ has been calculated by using $r^* = 0.76$. Third, for comparison, the theoretical probability distribution $P^\dagger_n$ under the parameters setting $(\lambda, \mu, s, r) = (80, 25, 4, 0.75)$ has been shown. From Table 5, it can be seen that the probability distribution $P^\ast_n$ reproduces $P^\dagger_n$ with high accuracy under fluctuating $P^\dagger_n$ by random number experiment.

Furthermore, in order to confirm the precision of the estimation method for $r^*$, we conduct the numerical analysis of 100 times under the same condition. We show the estimated result of $r^*$ in Table 6. Table 7 shows the average and standard deviation of estimated results of $r^*$ calculated by the results in Table 6 as the basic statistics. It can be found that the estimation method for $r^*$ based on $J$ has enough precision from Table 7. It is trivial that the increase of the frequency of observations brings the accuracy of the estimation. Table 8 shows the averages and standard deviations by the similar experiments in the cases that the number of generated random numbers are increased to 500, 1000 and 10000 in addition to 300.

5. Considerations of optimizing queueing system utilizing our proposed analysis. In this section, we consider the way to help that service providers make decisions for optimizing the queueing system by using the analysis result on the internal structure of the current queueing system. For example, by increasing the number of servers or by introducing high-performance machines, the service provider can improve the total service rate in an M/M/$s$ queueing system. If waiting duration is shortened by improving the total service rate in the queueing system, the phenomenon of balking will be less likely to occur, and the number of customers leaving from the queueing system by balking will decrease. To decrease the customers leaving from the queueing system by the phenomenon of balking leads to profit for the service provider. However, it is easy to imagine that cost is required to improve the total service rate in the queueing system. So, we will try to optimize the current queueing system by controlling the service rate of each servers or the number of servers. Specifically, by defining an evaluation function for the total service rate in the queueing system, we investigate the effect of the improvement of the total service rate. Then, we consider the optimal total service rate in the M/M/$s$ queueing system with balking.

Assume that $r^*$ in the M/M/$s$ queueing system which should be considered about the optimization of the total service rate has been estimated from observation data. We define the service rate of each server in the current queueing system as $\mu \equiv \mu^C$, current arrival rate expressed using $r^*$ represented by equation (25) as $\lambda_n \equiv \lambda^C_n$ and current steady-state probability distribution presented by equation (26) as $P_n \equiv P^C_n$, respectively. Further, it is assumed that an evaluation function
for the total service rate in the M/M/s queueing system is defined as

$$T = \left( \sum_{n=0}^{\infty} \lambda_n P_n \right) b - cs,$$

(29)

where $b$ means the profit for a service, and $c$ means operating cost per unit time for a server. It can be seen that the evaluation function $T$ is presented as the profit function for the total service rate in the M/M/s queueing system. Note that the probability $P_n$ in equation (29) has been derived as the function of the service rate $\mu$ and the number of servers $s$. Therefore, we can evaluate the profit function $T$ under various values of the service rate of each servers and/or the number of servers through changing $\mu$ and/or $s$. Further, the sum of multiplying the arrival rate $\lambda_n$ and the state probability $P_n$ means the average arrival rate $\bar{\lambda}$ in the steady-state. Therefore, consider that $\sum_{n=0}^{\infty} \lambda_n^C P_n^C$ is the average arrival rate before service improvement and $\sum_{n=0}^{\infty} \lambda_n P_n$ is the average arrival rate after improvement. In other words, the difference as $\sum_{n=0}^{\infty} \lambda_n P_n - \sum_{n=0}^{\infty} \lambda_n^C P_n^C$ describes the number of customers that can be acquired by increasing the number of servers or reinforcing the service rates, that is, can be prevented leaving from the queuing system by the phenomenon of balking.

We consider individual two methods to improve the total service rate in the M/M/s queueing system. As aforementioned, one is a method of improving the total service rate by increasing the number of servers $s$, and another is a method of improving the total service rate by reinforcing the service rate $\mu$ in each server. Note that although we can think a strategy that combines the above two methods, we will exclude it in this paper. Then, we assume the situation where $r^*$ has been estimated as $r^* = 0.75$ in the current queueing system that has the parameters setting of $(\lambda, \mu, s) = (80, 25, 4)$, and the profit for a service is given as $b = 30$.

First, let us consider the method of increasing the number of servers $s$ as the method of improving the total service rate, where assume that the operating cost per unit time for a server is given as $c = 100$. If the number of servers increases, the number of waiting customers will decrease. Accordingly, the number of customers leaving by balking will decrease. As a result, the number of customers arriving in the M/M/s queueing system increases and the profit increases. On one hand, the cost in order to increase the number of servers is required, so the service provider needs to make decisions taking into consideration the increase and decrease in the profit of the M/M/s queueing system. The profit $T$ and average arrival rate $\bar{\lambda}$ for each number of servers $s$ are shown in Table 9. Based on Table 9, the service provider can decide the optimal number of servers as $s = 6$.

Next, let us consider the case where the service rate of each server can be improved depending on the cost when the number of servers remains constant as $s = 4$. The service rate can increases with cost, but it is natural that there is an upper limit of the improvement of the service rate. Here, suppose that the relationship between the service rate $\mu$ and the operation cost $c$ for each server is defined as

$$\mu = 80(1 - e^{(-c/200)}),$$

(30)
Table 5. An example of the estimation as $r^* = 0.76$

| $n$ | $P^1_n$ | $P^2_n$ | $P^3_n$ |
|-----|---------|---------|---------|
| 0   | 0.044992| 0.053333| 0.045067|
| 1   | 0.143973| 0.156667| 0.144213|
| 2   | 0.230357| 0.210000| 0.230741|
| 3   | 0.245715| 0.246667| 0.246123|
| 4   | 0.196572| 0.200000| 0.196899|
| 5   | 0.093506| 0.086667| 0.093014|
| 6   | 0.032816| 0.036667| 0.032287|
| 7   | 0.009282| 0.006667| 0.009006|
| 8 or more | 0.002787| 0.003333| 0.002650|

Table 6. The estimated results of $r^*$

| $r^*$ | $0.76$ | $0.60$ | $0.70$ | $0.76$ | $0.97$ |
|-------|-------|-------|-------|-------|-------|
| $0.94$ | $0.77$ | $0.79$ | $0.80$ | $0.64$ |
| $0.76$ | $0.73$ | $0.65$ | $0.64$ | $0.66$ |
| $0.66$ | $0.76$ | $0.66$ | $0.76$ | $0.61$ |
| $0.80$ | $0.84$ | $0.86$ | $0.75$ | $0.77$ |
| $0.94$ | $0.65$ | $0.72$ | $0.91$ | $0.76$ |
| $0.79$ | $0.82$ | $0.76$ | $0.93$ | $0.88$ |
| $0.70$ | $0.83$ | $0.76$ | $0.75$ | $0.74$ |
| $0.57$ | $0.75$ | $0.86$ | $0.58$ | $0.61$ |
| $0.73$ | $0.75$ | $0.67$ | $0.68$ | $0.68$ |
| $0.73$ | $0.93$ | $0.61$ | $0.84$ | $0.79$ |
| $0.63$ | $0.72$ | $0.77$ | $1.05$ | $0.59$ |
| $0.72$ | $0.80$ | $0.74$ | $0.70$ | $0.81$ |
| $0.93$ | $0.84$ | $0.77$ | $0.64$ | $0.97$ |
| $0.83$ | $0.65$ | $0.70$ | $0.72$ | $0.92$ |
| $0.76$ | $0.73$ | $0.84$ | $0.80$ | $0.86$ |
| $0.66$ | $0.73$ | $0.75$ | $0.69$ | $0.74$ |
| $0.64$ | $0.78$ | $0.88$ | $0.89$ | $0.70$ |
| $0.69$ | $0.52$ | $0.71$ | $0.75$ | $0.86$ |
| $0.83$ | $0.59$ | $0.62$ | $0.74$ | $0.65$ |

Table 7. The basic statistics for $r^*$ in Table 6

|          | average | standard deviation |
|----------|---------|-------------------|
|          | 0.7527  | 0.1021            |

where, by assuming that $\mu$ never falls below the initial value of 25, $c$ has been set to 75 and more. Figure 5 illustrates the relationship between the cost $c$ and the profit $T$. From the result in Figure 5, we have obtained $c = 101$ as the optimal cost for each server. Then, we have as follows: $\mu = 31.71955$, and $\bar{\lambda} = 72.04416$ and $T = 1757.325$.

As the above numerical results for the optimization of the M/M/s queueing systems with balking, we have been able to improve the M/M/s queueing system under the respective situations based on the model parameter $r^*$ influencing the
Table 8. The averages and standard deviations by the similar experiments

| the number of observation | 300      | 500      | 1000     | 10000    |
|---------------------------|----------|----------|----------|----------|
| average                   | 0.7527   | 0.7446   | 0.7477   | 0.7516   |
| standard deviation        | 0.1021   | 0.0909   | 0.0466   | 0.0188   |

Table 9. The profit $T$ and average arrival rate when increasing the number of servers

| $s$ | $T$       | $\bar{\lambda}$ |
|-----|-----------|------------------|
| 4   | 1611.263  | 67.04209         |
| 5   | 1696.563  | 73.21876         |
| 6   | 1713.902  | 77.13008         |
| 7   | 1659.655  | 78.65517         |
| 8   | 1584.739  | 79.49130         |
| 9   | 1494.792  | 79.82639         |

Figure 5. The relationship between the cost $c$ and the profit $T$

strength of balking estimated through the constructed statistical mechanics model. On the other hand, suppose that we consider to improve the service rate in an M/M/1 queueing system with balking by deterministically utilizing $\lambda_n$ of equation (21) that is the traditional and typical arrival rate. In such a case, it is easily seen that we could not optimize even the M/M/1 queueing system. Hence, it can be understand that the statistical mechanics model constructed under considering the strength of balking has necessity and usefulness.

6. Conclusions. In this paper, based on the concept of the statistical mechanics, we have considered the mathematical modeling for analyzing the steady-state probability distribution of the M/M/s queueing system with balking. As a result, the statistical mechanics model for the M/M/s queueing system with balking has been proposed, and then the theoretical approach for analyzing the steady-state of the
M/M/s queueing system with balking has been successfully constructed. In the process, it has been verified that the proposed approach includes some existing analysis results for the steady-state probability distribution for the traditional Markovian queueing system with and without balking. Additionally, through the mathematical verification, it has been shown that the constructed statistical mechanics model is applicable to various types of Markovian queueing systems. Therefore, we have concluded that the constructed statistical mechanics model in this paper has general usefulness to analyze Markovian queueing systems.

Moreover, we have shown that the model parameter \( r \) influencing the strength of balking in the statistical mechanics model can be estimated by using simple observation data. Also, some methods for optimizing the total service rate in the M/M/s queueing system with balking have been considered using numerical examples. Concretely, it has been demonstrated that, in the case of improving the M/M/s queueing system with balking, the optimal number of servers \( s \) or the optimal service rate for each server \( \mu \) can be decided based on the statistical mechanics model for the M/M/s queueing system with balking.

The proposed approach has been successful to the case where, under the phenomenon of balking, the arrival rate \( \lambda \) is allowed to be dependent of the number of customers. In association with this, the phenomenon called “reneging” that customers lining up in a queue leave from the system in midway without waiting for own service to come is known \([1, 5, 8]\). The phenomenon of reneging may be considered to be an expansive phenomenon of balking. Also, the influence of the number of customers in a system may have an impact on the service rate \( \mu \). For example, a service provider may encourage oneself to enhance the service rate when the number of customers in the system becomes a lot. We would like to consider these interesting issues as some future subjects.

**Appendix: Derivation of equations (13)-(15).** From equation (12), we have

\[
\frac{\partial \Lambda(\vec{P}, \omega_1, \omega_2, \omega_3)}{\partial q_{n,i}} = -\ln q_{n,i} - 1 - \omega_1 g_1(n) - \omega_3 = 0, \quad (31)
\]

\[
\frac{\partial \Lambda(\vec{P}, \omega_1, \omega_2, \omega_3)}{\partial q_{n,j}} = -\ln q_{n,j} - 1 - \omega_1 g_1(s-1) - \omega_2 g_2(n) - \omega_3 = 0. \quad (32)
\]

Hence, the following formulae can be obtained:

\[
q_{n,i} = e^{-1-\omega_3-\omega_1 g_1(n)} = \alpha \beta g_1(n), \quad (33)
\]

\[
q_{n,j} = e^{-1-\omega_3-\omega_1 g_1(s-1)-\omega_2 g_2(n)} = \alpha \beta g_1(s-1) \gamma g_2(n). \quad (34)
\]

where \( \alpha = e^{-1-\omega_3}, \beta = e^{-\omega_1} \) and \( \gamma = e^{-\omega_2} \). Besides, by substituting equations (33) and (34) respectively into equations (5) and (6), we have

\[
P_n = \frac{s!}{n!} \alpha \beta g_1(n), \quad 0 \leq n < s, \quad (35)
\]

\[
P_n = s \alpha \beta g_1(s-1) \gamma g_2(n), \quad s \leq n. \quad (36)
\]

Then, from equation (35) and \( g_1(0) = 0.0 \), it can be easily see as

\[
P_0 = s! \alpha. \quad (37)
\]
Further, based on equations (35), (36) and (4), $\alpha$ can be represented as

$$\alpha = \frac{1}{\sum_{n=0}^{s-1} \prod_{i=0}^{n} \gamma_{i(n)} + \sum_{n=s}^{\infty} s \beta g_{1(s-1)} \gamma_{2(n)}}.$$  

(38)

And then, the relations of $g_1(0) = 0.0$, $g_1(1) = 1.0$, equation (35) and $\lambda P_0 = \mu P_1$ yield the following result:

$$\beta = \frac{\lambda}{\mu} \equiv a.$$  

(39)

While, based on the Markov analysis for M/M/s queueing systems with the service rate $\mu$ for each server and arrival rates $\lambda_n$, it is natural that the following relation is easily found:

$$P_n = \frac{\prod_{k=0}^{n-1} \lambda_{k-1}}{n! \mu^n} P_0, \quad 0 \leq n < s,$$

(40)

$$P_n = \frac{\prod_{k=0}^{n-1} \lambda_{k-1}}{s! (s-n-1) \mu^n} P_0, \quad s \leq n,$$

(41)

where note that $\lambda_0$ means $\lambda$. Since the relations of $\lambda_{s-2} P_{s-2} + s \mu P_s = \{ \lambda_{s-1} + (s-1) \mu \} P_{s-1}$ in the Markov process and $\lambda_{s-2} P_{s-2} = (s-1) \mu P_{s-1}$ under equation (38) are satisfied respectively, we have the following result by using the relation of $g_2(s) = 1.0$ and equations (35) and (36):

$$\lambda_{s-1} = (s \mu) \gamma.$$  

(42)

Then, it can be considered that the phenomenon of balking does not occur if a customer has arrived at the system in the case of $0 \leq n < s$, that is, we have $\lambda_n = \lambda$ in $0 \leq n < s$. Eventually, we can obtain

$$\gamma = \frac{\lambda}{s \mu} \equiv \rho.$$  

(43)

Furthermore, under the relation of $\{ \lambda_{n-1} + (n-1) \mu \} P_{n-1} = \lambda_{n-2} P_{n-2} + n \mu P_n$, the relation of $\lambda_{n-1} P_{n-1} = n \mu P_n$ is satisfied in general. By using this relation and equations (35), (39), we have

$$\lambda_{n-1} = \mu a g_{1(n)} - g_{1(n-1)} = \lambda_0 g_{1(n)} - g_{1(n-1)}, \quad 0 \leq n < s.$$  

(44)

On one hand, in the case of $n \geq s + 1$, we have $\lambda_{n-1} P_{n-1} = s \mu P_n$. Based on this relation, and equations (36) and (43), the following relation can be obtained as

$$\lambda_{n-1} = s \mu \rho g_{2(n)} - g_{2(n-1)} = \lambda_0 g_{2(n)} - g_{2(n-1)}, \quad n \geq s.$$  

(45)

Note that equation (45) can include the case of $n = s$ because we can assume $g_2(s-1) = 0$ in the case of $n = s$ and we have $g_2(s) = 1.0$.

As consequence, equations (13)-(15) have been derived by using equations (35)-(38), and equations (44) and (45).

**Acknowledgments.** We would like to thank the Editor and referees very much for their helpful, useful, and constructive comments and suggestions regarding this article. Further, this work was supported by JSPS KAKENHI Grant Number 18K04611: “Evaluation of system performance and reliability under incomplete information environment”. We would like to appreciate the grant for our research.
REFERENCES

[1] M. O. Abou-El-Ata and A. M. A. Hariri, The M/M/c/N queue with balking and reneging, Computers & Operations Research, 19 (1992), 713–716.

[2] I. Arizono, Y. Cui and H. Ohta, An analysis of M/M/s queueing systems based on the maximum entropy principle, Journal of the Operational Research Society, 42 (1991), 69–73.

[3] D. Chandler, Introduction to Modern Statistical Mechanics, Oxford University Press, Oxford, England, UK, 1987.

[4] C. Chen, Z. Jia and P. Varaiya, Causes and cures of highway congestion, IEEE Control Systems Magazine, 21 (2001), 26–32.

[5] A. A. El-Sherbiny, The truncated heterogeneous two-server queue: M/M/2/N with reneging and general balk function, International Journal of Mathematical Archive, 3 (2012), 2745–2754.

[6] W. Greiner, L. Neise and H. Stöcker, Thermodynamics and Statistical Mechanics, Springer-Verlag, New York, 1995.

[7] N. K. Jain, R. Kumar and B. Kumar Som, An M/M/1/N queuing system with reverse balking, American Journal of Operational Research, 4 (2014), 17–20.

[8] A. Montazer-Haghighi, J. Medhi and S. G. Mohanty, On a multiserver Markovian queueing system with balking and reneging, Computers & Operations Research, 13 (1986), 421–425.

[9] B. Natvig, On the transient state probabilities for a queueing model where potential customers are discouraged by queue length, Journal of Applied Probability, 11 (1974), 345–354.

[10] C. Preston, Gibbs States on Countable Sets, Cambridge University Press, London, England, UK, 1974.

[11] J. Sztrik, Basic Queueing Theory, Faculty of Informatics, University of Debrecen, Hungary, 2012.

Received January 2020; revised June 2020.

E-mail address: arizono@okayama-u.ac.jp
E-mail address: takemoto@mech.kindai.ac.jp