The $M/GI/\infty$ System Subject
to Semi-Markovian Random Environment

Anatoly Nazarov$^{(\boxtimes)}$ and Galina Baymeeva

Tomsk State University, Tomsk, Russia
{nazarov.tsu,baymeevag}@gmail.com

Abstract. In this paper we consider an $M/GI/\infty$ queueing system
operating in a semi-Markovian random environment. That is, the arrival
rate and service-time distribution change according to the external semi-
Markov process state transitions. The service policy subject to environ-
ment transitions is as follows: the service-time distribution of the present
customers does not change until their service is finished. The purpose of
our study is to obtain the probability distribution of the number of cus-
tomers in the system under asymptotic condition of high arrival rate and
frequent environment transitions. To do this, we first apply the method
of supplementary variable and the original method of dynamic screening
to our system. We then conduct the asymptotic analysis of the system
to obtain the discrete probability distribution.

Keywords: Queueing theory · Random environment · Semi-Markov
process · Method of dynamic screening · Method of asymptotic analysis

1 Introduction

Queueing systems subject to external stochastic influences such as Markov mod-
ulation and random environment are of considerate interest in scientific litera-
ture. Such influences are often represented as system breakdowns or arrivals of
priority customers (including batch arrivals) that force the system to behave at a
different mode. Namely, several cases of Markov-modulated single-server queues
are studied in [1]. Represented as a road subject to traffic incidents, the $M/M/\infty$
system operating under batch partial failures is considered in [2]. The random
environment is assumed to have only two states: when there is an incident and
when there are no incidents on the road. In [4], authors consider a more general
case of $M/M/\infty$ queue in Markovian random environment with arbitrary finite
number of states. The expression for steady-state factorial moments is obtained.
In [5], the analysis of $M/G/\infty$ system in Markovian random environment is
given. As a result, transient mean and stationary variance of the number of
customers present in the system are obtained; a deeper analysis of exponential
service case is conducted; the asymptotic normality of the number of customers
probability distribution is shown under conditions of high arrival rate and fre-
quent environment transitions due to the central limit theorem. In turn, the

steady-state mean number of customers in $M/M/\infty$ in semi-Markovian random environment is obtained in [7] as well as the steady-state distribution of the number of customers for the environment with 2 states.

Depending on the model application, different service policies of present customers with respect to environment transition may be considered. For instance, in one of the earliest papers [3] a single-server queue with general service-time is considered subject to interruptions with generally distributed durations. Two different cases of customer behavior after interruption clearance are studied: first, the resume policy, when the service is continued as it was left before the interruption; second, the repeat policy, when the service starts over. In [10, 11], the $M/G/\infty$ system in semi-Markovian random environment is studied. These papers cover three cases of the present customers' reaction to environment state transitions. The first one is considered in the present paper — service-time distribution stays the same while the customer is in the system. This policy is also assumed in [5]. In the second case all customers are immediately cleared from the queue as environment state transition happens. The last case considers customers in service moving to a secondary queue which is an infinite-server system with bulk arrivals. This case is specifically analyzed in [10], and as a result the steady-state mean number of customers in the secondary queue is obtained.

Infinite-server queues are often used to approximate the behavior of systems with sufficiently large number of servers, such as banks, call-centers, supermarkets or digital distribution platforms. Such objects in reality are often affected by extraneous factors of stochastic nature which affect their performance. For instance, the change of bank rate set by the Central bank affects the conditions under which commercial banks give loans to their clients. These, in turn, significantly influence the intensity of clients' arrival. In this article we consider a mathematical model of such situation as an $M/GI/\infty$ queue operating in a random environment, for which the underlying process is a semi-Markov process with finite number of states. The arrival rate and service-time distribution change according to the environment state. Note that distribution of service-time customers which are currently being served does not change until the service-time is finished. Say the bank provided a credit to the client on certain conditions and during the repayment period there was a change of bank rate. The client will continue to repay his debt on those initial conditions — as mentioned in a loan agreement.

2 Problem Statement

We consider an $M/GI/\infty$ queueing system operating in semi-Markovian random environment. The system under discussion is an infinite-server queue with one stationary Poisson arrival process with parameter $\lambda_s N$ and the unlimited number of servers each having service-time distribution function $B_s(x)$, $s = \overline{1, K}$. We use a large parameter N that represents the condition of high arrival rate. Here $s = \overline{1, K}$ is the current state of a semi-Markov stochastic process $s(t)$ defined by the matrix product $\mathbf{P} \cdot \mathbf{A}(x)$. The matrix $\mathbf{P}$ here is a probability matrix of $s(t)$

state transitions and $\mathbf{A}(x)$ is a diagonal matrix with conditional sojourn time cdfs for every state $s = \overline{1, K}$ of $s(t)$ on its main diagonal. As there is always a free server in the system, there is no queue or loss option and each arriving customer is immediately placed at any free server and stays there for random time with distribution function $B_s(x)$. Note that we study the case when service-time distribution of a customer which is currently being served does not change until its service is finished.

That being said, for considered model we define a two-component process $\{i(t), s(t)\}$, where $i(t)$ with values $i \geq 0$ is the number of customers in the system at time t. Apparently, this process is non-Markovian. To deal with it, we first apply the original method of dynamic screening and the method of supplementary variable.

3 Method of Dynamic Screening

The method of dynamic screening can be used for the analysis of both queueing systems and networks. Further applications may be found in [14–16]. We apply this method to our system in the following way.

Given that at a certain time t_0 the system is empty, we pick a moment T and track the customer arrivals during the time interval (t_0, T). The customer will be referred to as "screened" at time t with probability

$$S_s(t) = 1 - B_s(T - t), s = \overline{1, K}, t_0 < t < T,$$

if it arrived at the system at time $t < T$ and was not fully serviced until the time T. Thus, the screened customers will be in the system taking up its servers at time T.

Let us denote by $n(t)$ the number of customers that were screened until time t. Stochastic process $n(t)$ is a screened point process with its points being the screened customers. The following identity always takes place:

$$i(T) = n(T). \tag{1}$$

We need to choose time t_0 so that at all times $t < t_0$ there are no screened customers, i.e.

$$S_s(t) = 1 - B_s(T - t) = 0, s = \overline{1, K}, t < t_0.$$

Since $B_s(x)$ is a cumulative distribution function, it is obvious enough to put $t_0 = -\infty$.

We write the possible state transitions of $n(t)$ and their probabilities assuming $n(t) = n, n \geq 0$ as follows:

$$n(t + \Delta t) = \begin{cases} n + 1, & \text{with prob. } \lambda_s \Delta t S_s(t) + o(\Delta t), \\ n, & \text{with prob. } 1 - \lambda_s \Delta t S_s(t) + o(\Delta t), \end{cases} \quad s = \overline{1, K}$$

Equality (1) allows us to analyze a point process $n(t)$ instead of $i(t)$. Characteristics of the process $n(t)$ at time T coincide with the characteristics of value $i(T)$.

4   Kolmogorov Differential Equations

In order to deal with semi-Markovian process $s(t)$, we first need to apply the method of supplementary variable. We define $z(t)$ as residual sojourn time of $s(t)$ process in the current state, i.e. the interval from t until the next environment transition. It follows that the three-dimensional process $\{s(t), n(t), z(t)\}$ is a Markovian one. Therefore, we define the probabilities of system and environment state at time t as follows:

$$P(s, n, z, t) = P\left\{s(t) = s, n(t) = n, z(t) < \frac{z}{N}\right\}, s = \overline{1, K}, n \geq 0. \qquad (2)$$

Here the big parameter N justifies the condition of frequent environment transitions that compensates high arrival rate. The matrices that define the process $s(t)$ are determined as follows:

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ p_{21} & p_{22} & \cdots & p_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ p_{K1} & p_{K2} & \cdots & p_{KK} \end{pmatrix}, \mathbf{A}(x) = \begin{pmatrix} A_1(x) & 0 & \cdots & 0 \\ 0 & A_2(x) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_K(x) \end{pmatrix}.$$

Let τ_s be the sojourn time of $s(t)$ in state $s = \overline{1, K}$. Then functions $A_s(x)$ are defined in the following way:

$$A_s(x) = P\left\{\frac{\tau_s}{N} < x\right\} = P\left\{\tau_s < Nx\right\}, s = \overline{1, K},$$

which means that $A_s(x)$ are the distribution functions of N-fold sojourn time of $s(t)$ in state $s = \overline{1, K}$.

The system of Kolmogorov differential equations that defines the probabilities (2) is written as follows:

$$\frac{1}{N} \frac{\partial P(s, n, z, t)}{\partial t} - \frac{\partial P(s, n, z, t)}{\partial z} + \frac{\partial P(s, n, 0, t)}{\partial z} =$$
$$\lambda_s S_s(t) \{P(s, n-1, z, t) - P(s, n, z, t)\} + \qquad (3)$$
$$A_s(z) \sum_{k=1}^{K} p_{ks} \frac{\partial P(k, n, 0, t)}{\partial z}, s = \overline{1, K}, n \geq 0$$

Here we use the denotation

$$\frac{\partial P(s, n, 0, t)}{\partial z} = \left. \frac{\partial P(s, n, z, t)}{\partial z} \right|_{z=\infty}.$$

Provided $z \to \infty$, the initial condition to such system's solution is defined as follows:

$$P(s, n, t_0) = \begin{cases} r(s), & \text{if } n = 0, \\ 0, & \text{if } n > 0, \end{cases} s = \overline{1, K} \qquad (4)$$

Here $r(s)$ are the stationary probabilities of embedded Markov chain states of $s(t)$, $s = \overline{1, K}$. The partial characteristic functions of the process $\{s(t), n(t), z(t)\}$ are defined as follows:

$$H(s, u, z, t) = \sum_{n=0}^{\infty} e^{jun} P(s, n, z, t), s = \overline{1, K}$$

Here $j = \sqrt{-1}$ is the imaginary unit. We rewrite the system (4) using partial characteristic functions in the following way:

$$\frac{1}{N} \frac{\partial H(s, u, z, t)}{\partial t} - \frac{\partial H(s, u, z, t)}{\partial z} + \frac{\partial H(s, u, 0, t)}{\partial z} =$$

$$\lambda_s S_s(t)(e^{ju} - 1)H(s, u, z, t) + \tag{5}$$

$$A_s(z) \sum_{k=1}^{K} p_{ks} \frac{\partial H(k, u, 0, t)}{\partial z}, s = \overline{1, K}$$

We then use the following vector and matrix denotations:

$$\mathbf{H}(u, z, t) = \big(H(1, u, z, t)\ H(2, u, z, t) \cdots H(K, u, z, t) \big),$$

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K \end{pmatrix}, \mathbf{S}(t) = \begin{pmatrix} S_1(t) & 0 & \cdots & 0 \\ 0 & S_2(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_K(t) \end{pmatrix},$$

to rewrite the system (4) as follows:

$$\frac{1}{N} \frac{\partial \mathbf{H}(u, z, t)}{\partial t} - \frac{\partial \mathbf{H}(u, z, t)}{\partial z} +$$

$$\frac{\partial \mathbf{H}(u, 0, t)}{\partial z} [\mathbf{I} - \mathbf{PA}(z)] = (e^{ju} - 1)\mathbf{H}(u, z, t)\boldsymbol{\Lambda}\mathbf{S}(t). \tag{6}$$

Here $\mathbf{I}$ is the identity matrix. Our goal is to obtain the solution to system (6) as $z \to \infty$ that satisfies the initial condition derived from (4):

$$\mathbf{H}(u, t_0) = \mathbf{r}. \tag{7}$$

The row vector $\mathbf{r}$ here is the stationary probability distribution of the embedded Markov chain of the process $s(t)$ and solves the following system of matrix-vector equations:

$$\begin{cases} \mathbf{rP} = \mathbf{r}, \\ \mathbf{re} = 1. \end{cases} \tag{8}$$

5   Method of Asymptotic Analysis

Method of asymptotic analysis for queueing systems is the analysis of equations that define any of the system's characteristics or parameters [13]. It allows

us to obtain the explicit distribution, parameters and moments under certain asymptotic conditions.

We obtain the solution to system (6) under asymptotic conditions of high arrival rate and frequent environment transitions, that is, as $N \to \infty$.

5.1   First-Order Asymptotic Analysis

Let us define substitutions for system (6) as follows:

$$\varepsilon = \frac{1}{N}, u = \varepsilon w, \mathbf{H}(u, z, t) = \mathbf{F}_1(w, z, t, \varepsilon).$$

Then (6) can be rewritten as

$$\varepsilon \frac{\partial \mathbf{F}_1(w, z, t, \varepsilon)}{\partial t} - \frac{\partial \mathbf{F}_1(w, z, t, \varepsilon)}{\partial z}$$
$$+ \frac{\partial \mathbf{F}_1(w, 0, t, \varepsilon)}{\partial z} [\mathbf{I} - \mathbf{PA}(z)] \tag{9}$$
$$= (e^{j\varepsilon w} - 1)\mathbf{F}_1(w, z, t, \varepsilon)\mathbf{\Lambda S}(t).$$

As $\varepsilon \to 0$, the following equality holds:

$$\frac{\partial \mathbf{F}_1(w, z, t)}{\partial z} = \frac{\partial \mathbf{F}_1(w, 0, t)}{\partial z} [\mathbf{I} - \mathbf{PA}(z)]. \tag{10}$$

We then represent the function $\mathbf{F}_1(w, z, t)$ as a product

$$\mathbf{F}_1(w, z, t) = \mathbf{r}(z)\Phi_1(w, t). \tag{11}$$

Substitution (11) applied to (10) gives the following equation that defines row-vector $\mathbf{r}(z)$:

$$\mathbf{r}(z) = \int_0^z \mathbf{r}'(0) [\mathbf{I} - \mathbf{PA}(x)] \, dx \tag{12}$$

To determine the value $\mathbf{r}'(0)$, we make the following substitution

$$\mathbf{r}'(0) = C\mathbf{r}, \ C = const. \tag{13}$$

Note that according to (8)

$$\lim_{z \to \infty} \mathbf{r}(z) = C \int_0^\infty \mathbf{r} [\mathbf{I} - \mathbf{PA}(x)] \, dx$$
$$= C \int_0^\infty \mathbf{r} [\mathbf{I} - \mathbf{A}(x)] \, dx = C\mathbf{r}\mathbf{A}.$$

Apparently, the matrix $\mathbf{A}$ here is the diagonal matrix containing means $A_s, s = \overline{1, K}$ of distribution functions from $\mathbf{A}(x)$ on its main diagonal. According to (8), the constant C is derived as follows:

$$C = \frac{1}{\mathbf{rAe}} = \frac{1}{a}.$$

Finally, we write the expression for $\mathbf{r}(z)$:

$$\mathbf{r}(z) = \frac{1}{a} \int_0^z \mathbf{r} \left[\mathbf{I} - \mathbf{PA}(x)\right] dx.$$

Note that

$$\mathbf{r}(z)\Big|_{z=\infty} = \frac{\mathbf{rA}}{a}.$$

Now we set $z = \infty$ in (5.1) and make substitution (11):

$$\varepsilon \frac{1}{a} \mathbf{rA} \frac{\partial \Phi_1(w,t)}{\partial t} + \Phi_1(w,t) \mathbf{r} \left[\mathbf{I} - \mathbf{P}\right]$$
$$= (e^{j\varepsilon w} - 1)\Phi_1(w,t) \frac{1}{a} \mathbf{rA\Lambda S}(t).$$

Post-multiplication by $\mathbf{e}$ of both parts of the latter equation gives us the following first-order ordinary differential equation:

$$\frac{\partial \Phi_1(w,t)}{\partial t} = \frac{1}{a} \frac{e^{j\varepsilon w} - 1}{\varepsilon} \Phi_1(w,t) \mathbf{rA\Lambda S}(t)\mathbf{e}. \tag{14}$$

As $\varepsilon \to 0$, the function $\Phi_1(w,t)$ that solves the equation above and satisfies the initial condition derived from (7) is as follows:

$$\Phi_1(w,t) = exp\left\{jw\kappa_1(t)\right\},$$

$$\kappa_1(t) = \frac{1}{a} \int_{-\infty}^t \mathbf{rA\Lambda S}(\tau)\mathbf{e}d\tau.$$

Finally, we can write

$$\mathbf{H}(u,t) = \mathbf{F}_1(w,t,\varepsilon) \approx \mathbf{F}_1(w,t) = \frac{\mathbf{rA}}{a}\Phi_1(w,t) = \frac{\mathbf{rA}}{a}exp\{jw\kappa_1(t)\},$$

where $w = Nu$. It follows that

$$M\{e^{jun(t)}\} = \mathbf{H}(u,t)\mathbf{e} \approx h_1(u,t) = exp\{ju\kappa_1(t)N\}.$$

Since (1) takes place, we can finally conclude:

$$M\{e^{jui(T)}\} = M\{e^{jun(T)}\} = \mathbf{H}(u,T)\mathbf{e}$$
$$\approx h_1(u,T) = exp\{ju\kappa_1(T)N\}.$$

Let us calculate the value $\kappa_1(T)$:

$$\kappa_1(T) = \int_{-\infty}^{T} \mathbf{r}\mathbf{A}\mathbf{\Lambda}\mathbf{S}(t)\mathbf{e}dt = \int_{-\infty}^{T} \sum_{s=1}^{K} \mathbf{r}(s)A_s\lambda_s S_s(t)dt$$

$$= \sum_{s=1}^{K} \mathbf{r}(s)A_s\lambda_s \int_{-\infty}^{T} \{1 - B_s(T - t)\}dt$$

$$= \sum_{s=1}^{K} \mathbf{r}(s)A_s\lambda_s \int_{0}^{\infty} \{1 - B_s(\tau)\}d\tau = \sum_{s=1}^{K} \mathbf{r}(s)A_s\lambda_s b_s,$$

where b_s are the service-time means, $s = \overline{1, K}$. Thus

$$\kappa_1(T) = \sum_{s=1}^{K} \mathbf{r}(s)\lambda_s \int_{0}^{\infty} \{1 - B_s(\tau)\}d\tau = \sum_{s=1}^{K} \mathbf{r}(s)A_s\lambda_s b_s = \mathbf{r}\mathbf{A}\mathbf{\Lambda}\mathbf{B}\mathbf{e},$$

where $\mathbf{B}$ is a diagonal matrix containing service-time means b_s.

5.2 Second-Order Asymptotic Analysis

In the equation (6) we make a substitution

$$\mathbf{H}(u, z, t) = \mathbf{H}_2(u, z, t)e\{ju\kappa_1(t)N\}. \tag{15}$$

The function $\mathbf{H}_2(u, z, t)$ here is the centered characteristic function as the following relation takes place:

$$\mathbf{H}_2(u, z, t)\mathbf{e} = \mathbf{H}(u, z, t)e^{-ju\kappa_1(t)N}\mathbf{e}$$
$$= M\left\{exp\left[ju(n(t) - \kappa_1(t)N)\right]\right\}.$$

The substitution (15) yields an equation which defines $\mathbf{H}_2(u, z, t)$:

$$\frac{1}{N}\frac{\partial \mathbf{H}_2(u, z, t)}{\partial t} - \frac{\partial \mathbf{H}_2(u, z, t)}{\partial z}$$
$$+ \frac{\partial \mathbf{H}_2(u, 0, t)}{\partial z}\left[\mathbf{I} - \mathbf{P}\mathbf{A}(z)\right] \tag{16}$$
$$= \mathbf{H}_2(u, z, t)\left\{(e^{ju} - 1)\mathbf{\Lambda}\mathbf{S}(t) - ju\kappa_1'(t)\mathbf{I}\right\}$$

We rewrite the latter system using substitutions

$$\varepsilon^2 = \frac{1}{N}, u = \varepsilon w, \mathbf{H}_2(u, t) = \mathbf{F}_2(w, z, t, \varepsilon)$$

in the following way:

$$\varepsilon^2 \frac{\partial \mathbf{F}_2(w, z, t, \varepsilon)}{\partial t} - \frac{\partial \mathbf{F}_2(w, z, t, \varepsilon)}{\partial z}$$
$$+ \frac{\partial \mathbf{F}_2(w, 0, t, \varepsilon)}{\partial t}\left[\mathbf{I} - \mathbf{P}\mathbf{A}(z)\right] \tag{17}$$
$$= \mathbf{F}_2(w, z, t, \varepsilon)[(e^{j\varepsilon w} - 1)\mathbf{\Lambda}\mathbf{S}(t) - j\varepsilon w\kappa_1'(t)\mathbf{I}]$$

As $\varepsilon \to 0$, the following relation takes place:

$$\frac{\partial \mathbf{F}_2(w,z,t)}{\partial z} = \frac{\partial \mathbf{F}_2(w,0,t)}{\partial z} \left[\mathbf{I} - \mathbf{PA}(z)\right].$$

It follows that the function $\mathbf{F}_2(w,z,t)$ may be represented as follows:

$$\mathbf{F}_2(w,z,t) = \mathbf{r}(z)\Phi_2(w,t). \tag{18}$$

In turn, the function $\mathbf{F}_2(w,z,t,\varepsilon)$ may be approximated with the following expression:

$$\mathbf{F}_2(w,z,t,\varepsilon) = \Phi_2(w,t)\left\{\mathbf{r}(z) + j\varepsilon w \mathbf{f}_2(z,t)\right\} + O(\varepsilon^2). \tag{19}$$

The row-vector function $\mathbf{f}_2(z,t)$ is to be defined. To do this, first we make a substitution (19) in the system (17). We also make the following approximation in (17):

$$e^{j\varepsilon w} - 1 = j\varepsilon w + O(\varepsilon^2),$$

and then set $\varepsilon \to 0$. These manipulations yield us the following equation that defines $\mathbf{f}_2(z,t)$:

$$\frac{\partial \mathbf{f}_2(z,t)}{\partial z} - \frac{\partial \mathbf{f}_2(0,t)}{\partial z}\left[\mathbf{I} - \mathbf{PA}(z)\right] + \Phi_2(w,t)\mathbf{r}(z)\left[\mathbf{\Lambda S}(t) - \kappa_1'(t)\mathbf{I}\right] = \mathbf{0}. \tag{20}$$

Here $\mathbf{0}$ is the row-vector filled with zeros. It follows that as $z \to \infty$, we have the relation

$$\mathbf{f}_2(t) = \int\limits_0^\infty \left\{\frac{\partial \mathbf{f}_2(0,t)}{\partial z}\left[\mathbf{I} - \mathbf{PA}(x)\right] - \mathbf{r}(x)\left[\mathbf{\Lambda S}(t) - \kappa_1'(t)\mathbf{I}\right]\right\} dx. \tag{21}$$

The right part of the latter relation is the improper integral. In order for it to converge, it is necessary that the integrand function converges to 0 as the variable of integration approaches ∞. That is, the following relation stands for $\mathbf{f}_2(0,t)$:

$$\frac{\partial \mathbf{f}_2(0,t)}{\partial z}\left[\mathbf{I} - \mathbf{P}\right] = \frac{\mathbf{rA}}{a}\left[\mathbf{\Lambda S}(t) - \kappa_1'(t)\mathbf{I}\right]. \tag{22}$$

The equation above is the non-homogeneous underdetermined system of linear equations. We represent its solution as a sum of general solution to homogeneous system and a partial solution to non-homogeneous system:

$$\frac{\partial \mathbf{f}_2(0,t)}{\partial z} = c(t)\mathbf{r} + \mathbf{g}(t), \tag{23}$$

where $c(t)$ is an arbitrary scalar function of t. We write the additional condition for the function $\mathbf{g}(t)$ as follows:

$$\mathbf{g}(t)\mathbf{e} = 0, \tag{24}$$

Let us now define the explicit expression for (21):

$$\mathbf{f}_2(t) = \int_0^\infty \left\{ \frac{\partial \mathbf{f}_2(0,t)}{\partial z} \left[\mathbf{I} - \mathbf{P} + \mathbf{P}(\mathbf{I} - \mathbf{A}(z)) \right] - \mathbf{r}(z) \left[\mathbf{\Lambda S}(t) - \kappa_1'(t)\mathbf{I} \right] \right\} dz$$

$$= \int_0^\infty \left[\frac{1}{a}\mathbf{r}\mathbf{A} - \mathbf{r}(z) \right] \left[\mathbf{\Lambda S}(t) - \kappa_1'(t)\mathbf{I} \right] dz$$

$$+ \int_0^\infty \frac{\partial \mathbf{f}_2(0,t)}{\partial z} \mathbf{P} \left[\mathbf{I} - \mathbf{A}(z) \right] dz$$

$$= \frac{1}{a}\mathbf{r}\mathbf{A} \int_0^\infty \left\{ \mathbf{I} - \mathbf{A}^{-1} \int_0^z \left[\mathbf{I} - \mathbf{A}(x) \right] dx \right\} dz \left[\mathbf{\Lambda S}(t) - \kappa_1'(t)\mathbf{I} \right] + \frac{\partial \mathbf{f}_2(0,t)}{\partial z}\mathbf{P}\mathbf{A}.$$

Note that $\mathbf{A}^{-1} \int_0^z \left[\mathbf{I} - \mathbf{A}(x) \right] dx$ is a diagonal matrix that contains distribution functions of both elapsed and residual sojourn time of $s(t)$ at each of its states. Then denoted by $\overline{\mathbf{A}}$ is the diagonal matrix that contains means of such cdfs respectively. Finally, we rewrite the expression for $\mathbf{f}_2(t)$ as follows:

$$\mathbf{f}_2(t) = \frac{1}{a}\mathbf{r}\mathbf{A}\overline{\mathbf{A}} \left[\mathbf{\Lambda S}(t) - \kappa_1'(t)\mathbf{I} \right] + \frac{\partial \mathbf{f}_2(0,t)}{\partial z}\mathbf{P}\mathbf{A}. \tag{25}$$

Now we show that the row-vector function $\mathbf{f}_2(t)$ does not actually depend on the arbitrary scalar function $c(t)$ that is present in (23). To do that, we consider the following term that is present in (25):

$$\frac{\partial \mathbf{f}_2(0,t)}{\partial z}\mathbf{P}\mathbf{A} \left[\mathbf{\Lambda S}(t) - \kappa_1'(t)\mathbf{I} \right] \mathbf{e}$$

$$= [c(t)\mathbf{r} + \mathbf{g}(t)] \mathbf{P}\mathbf{A} \left[\mathbf{I} - \frac{1}{a}\mathbf{e}\mathbf{r}\mathbf{A} \right] \mathbf{\Lambda S}(t)\mathbf{e}$$

$$= \mathbf{g}(t)\mathbf{P}\mathbf{A} \left[\mathbf{I} - \frac{1}{a}\mathbf{e}\mathbf{r}\mathbf{A} \right] \mathbf{\Lambda S}(t)\mathbf{e}$$

$$+ c(t)\mathbf{r}\mathbf{P}\mathbf{A}\mathbf{\Lambda S}(t)\mathbf{e} - c(t)\frac{1}{a}\mathbf{r}\mathbf{P}\mathbf{A}\mathbf{e}\mathbf{r}\mathbf{A}\mathbf{\Lambda S}(t)\mathbf{e}.$$

With (8) and $a = \mathbf{r}\mathbf{A}\mathbf{e}$ in mind, we conclude that the two latter terms cancel each other. Thus, the function $c(t)$ is not present in (25).

Now let us determine the function $\Phi_2(w,t)$. For this purpose, we again make substitution (19) in (17) and also the following approximation:

$$e^{j\varepsilon w} - 1 = j\varepsilon w + \frac{(j\varepsilon w)^2}{2} + O(\varepsilon^3).$$

As $\varepsilon \to 0$ and $z \to \infty$, this yields us the first-order ODE that defines $\Phi_2(w,t)$:

$$\frac{\partial \Phi_2(w,t)}{\partial t} = \frac{(jw)^2}{2}\Phi_2(w,t) \left\{ \kappa_1'(t) + 2\mathbf{f}_2(t) \left[\mathbf{\Lambda S}(t) - \kappa_1'(t)I \right] \mathbf{e} \right\}. \tag{26}$$

Its solution that satisfies the initial condition derived from (7) is of the following form:

$$\Phi_2(w,t) = exp\left\{\frac{(jw)^2}{2}\kappa_2(t)\right\}, \tag{27}$$

where

$$\kappa_2(t) = \kappa_1(t) + 2\int_{-\infty}^{t} \mathbf{f}_2(\tau)\left[\mathbf{\Lambda S}(\tau) - \kappa_1'(\tau)\mathbf{I}\right]\mathbf{e}d\tau. \tag{28}$$

Thus, the expression for the centered characteristic function $\mathbf{H}_2(u,t)$ is obtained and is written as follows:

$$\mathbf{H}_2(u,t) = \mathbf{F}_2(w,t,\varepsilon) \approx \mathbf{F}_2(w,t) = \frac{\mathbf{rA}}{a}\Phi_2(w,t)$$

$$= \frac{\mathbf{rA}}{a}exp\{\frac{(jw)^2}{2}\kappa_2(t)\} = \frac{\mathbf{rA}}{a}exp\{\frac{(ju)^2}{2}\kappa_2(t)N\}.$$

It follows that

$$\mathbf{H}(u,t) = \mathbf{H}_2(u,t)e^{ju\kappa_1(t)N} \approx \frac{\mathbf{rA}}{a}exp\left\{ju\kappa_1(t)N + \frac{(ju)^2}{2}\kappa_2(t)N\right\}, \tag{29}$$

$$M\{e^{jun(t)}\} = \mathbf{H}(u,t)\mathbf{e} \approx h_2(u,t) = exp\left\{ju\kappa_1(t)N + \frac{(ju)^2}{2}\kappa_2(t)N\right\}. \tag{30}$$

Considering (1), the following identities are true:

$$M\{e^{jui(T)}\} = M\{e^{jun(T)}\} = \mathbf{H}(u,T)\mathbf{e} \approx h_2(u,T)$$
$$= exp\{ju\kappa_1(T)N + \frac{(ju)^2}{2}\kappa_2(T)N\}, \tag{31}$$

where $\kappa_2(T)$ is of the following form:

$$\kappa_2(T) = \kappa_1(T) + 2\int_{-\infty}^{T} \mathbf{f}_2(t)\left[\mathbf{\Lambda S}(t) - \kappa_1'(t)\mathbf{I}\right]\mathbf{e}dt \tag{32}$$

According to the definition of functions $S_s(t) = 1 - B_s(T-t)$ it is clear that $\lim_{t\to\infty} S_s(t) = 0, s = \overline{1,K}$. Therefore, it is clear that the improper integral (32) is converging and thus can be calculated numerically given specific system and environment parameters.

Obviously, the asymptotic steady-state probability distribution of the number of customers in the system defined by (31) is normal with first and second cumulants $\kappa_1(t)N$ and $\kappa_2(t)N$ respectively. It is known that

$$M\{i(T)\} \approx \kappa_1(T)N, D\{i(T)\} \approx \kappa_2(T)N. \tag{33}$$

Inverse Fourier transform of (31) gives the probability density function of the normally distributed random variable:

$$p(x) = \frac{1}{\sqrt{2\pi\kappa_2(T)N}} exp\left\{-\frac{(x-\kappa_1(T)N)^2}{2\kappa_2(T)N}\right\}. \tag{34}$$

It is necessary to switch from this continuous distribution to discrete as follows:

$$P(i) = Cp(i), i \geq 0, \tag{35}$$

where the constant value C is defined considering the normalizing condition:

$$\sum_{i=0}^{\infty} P(i) = C \sum_{i=0}^{\infty} p(i) = 1. \tag{36}$$

Due to (36), C is given as follows:

$$C = 1/\sum_{i=0}^{\infty} p(i) \tag{37}$$

6   Conclusion

Thus, the Gaussian approximation of the probability distribution of the number of customers in the system $M(\lambda_s)/G(B_s(x))/\infty$ is obtained during the asymptotic analysis under conditions of high arrival rate and frequent environment transitions. Using the method of dynamic screening, we considered a non-stationary Markov point process $n(t)$ instead of non-Markovian $i(t)$ which is the number of customers in the system. Then, according to the method of supplementary variable we defined the residual sojourn time $z(t)$ in the present state of the environment process $s(t)$ to be able to analyze it with theory of Markov processes tools. After deriving the system of differential equations in terms of vector characteristic functions of the number of customers in the system, we conducted the asymptotic analysis of the system in question.

Earlier we considered a problem of $M/G/\infty$ queue operating in Markovian random environment with the same service policy when service-time distribution does not change while the customer is in the system. Similarly, we obtained the steady-state probability distribution of the number of customers in the system. However, the Markov case narrows down the application area significantly. Thus, in this paper we considered a more general case with random environment being semi-Markovian.

References

1. Prabhu, N.U., Zhu, Y.: Markov-modulated queueing systems. Queueing Syst. **5**, 215–245 (1989)
2. Baykal-Gursoy, M., Xiao, W.: Stochastic decomposition in $M/M/\infty$ queues with Markov modulated service rates. Queueing Syst. **48**, 75–88 (2004)
3. Keilson, J.: Queues subject to service interruption. Ann. Math. Statist. **33**, 1314–1322 (1962)
4. O'Cinneide, C.A., Purdue, P.: The $M/M/\infty$ queue in a random environment. J. Appl. Prob. **23**, 175–184 (1986)
5. Blom, J., Kella, O., Mandjes, M., Thorsdottir, H.: Markov-modulated infinite-server queues with general service times. Queueing Syst. **76**, 403–424 (2014)
6. D'Auria, B.: $M/M/\infty$ queues in semi-Markovian random environment. Queueing Syst. **58**, 221–237 (2008)
7. Falin, G.: The $M/M/\infty$ queue in random environment. Queueing Syst. **58**, 65–76 (2008)
8. Fralix, B.H., Adan, I.J.B.F.: An infinite-server queue influenced by a semi-Markovian environment. Queueing Syst. **61**, 65–84 (2009)
9. D'Auria, B.: Stochastic decomposition of the $M/G/\infty$ queue in a random environment. Oper. Res. Lett. **35**, 805–812 (2007)
10. Purdue, P., Linton, D.: An infinite-server queue subject to an extraneous phase process and related models. J. Appl. Prob. **18**, 236–244 (1981)
11. Linton, D., Purdue, P.: An $M/G/\infty$ queue with m customer types subject to periodic clearing. Opsearch **16**, 80–88 (1979)
12. Nazarov, A.A., Baymeeva, G.V.: The study of $M/G/\infty$ in random environment
13. Nazarov, A.A., Moiseeva, S.P.: Method of Asymptotic Analysis in Queueing Theory. NTL, Tomsk (2006) (in Russian)
14. Nazarov, A.A., Moiseev, A.N.: Analysis of an open non-Markovian $GI - (GI|\infty)^K$ queueing network with high-rate renewal arrival process. Prob. Inf. Transm. **49**, 167–178 (2013)
15. Moiseev, A.N., Nazarov, A.A.: Asymptotic analysis of a multistage queuing system with a high-rate renewal arrival process Optoelectronics. Instrum. Data Process. **50**(2), 163–171 (2014)
16. Moiseev, A., Nazarov, A.: Asymptotic analysis of the infinite-server queueing system with high-rate semi-Markov arrivals. In: IEEE International Congress on Ultra Modern Telecommunications and Control Systems (ICUMT 2014), pp. 507–513. IEEE Press (2014)