M. S. Alzahrani, J. H. Dshalalow[1]
Department of Mathematical Sciences College
of Science Florida Institute of Technology
(Melbourne, FL 32901, USA
E-mail's: malzah01@my.fit.edu and eugene@fit.edu)

Fluctuation Analysis in a Queue with (L,N)-Policy and Secondary Maintenance Jobs. Continuous Time Parameter Process[2]

Investigation results on a queueing system with an auxiliary maintenance process launched during server's vacancy period and initiated by the exhausted queue are presented. The server operation strategy at different queue conditions is proposed. Time-sensitive analysis to investigate the queueing process at arbitrary periods of time is used. The results are obtained in explicit forms for several related models. Computational examples illustrate their analytical tractability. Various performance measures (the buffer load, switchover rate, and the number of jobs processed per unit time) are introduced and optimization problems are discussed.

Представлены результаты исследований систем очередей с дополнительным процессом обслуживания, который запускается в периоды простоя сервера и инициируется заполненной очередью. Предложена стратегия работы сервера при различных состояниях очереди. Использован времязависимый анализ для исследования процессов в системах очередей в произвольные периоды времени. Результаты получены в явной форме для нескольких моделей. Приведены примеры расчета, свидетельствующие о возможности их аналитической трактовки. Введены различные критерии производительности (загрузка буферов, скорость переключения и число заданий, обрабатываемых в единицу времени) и рассмотрены проблемы оптимизации.

Key words: queueing, semi-regenerative analysis, fluctuation theory, marked point process, N-policy, maintenance process, multiple vacations.

1. Introduction. In this article we investigate a flexible queue with a bivariate process modeling two servicing facilities. When the queue in the primary facility (PF) is dropped to zero, the server moves to the secondary facility (SF), where jobs, that are assembled in packets of random sizes, are waiting for being processed. The server works on the jobs, one at a time, and stays at the SF until the total quantity of the processed jobs crosses some fixed threshold L. The server may not be called off by abruptly breaking its work on a packet when the number

---

[1] The corresponding author

[2] This research is supported by the US Army Grant No. W911NF-07-1-0121

of processed jobs reaches L. It needs to finish that packet and only then return to the PF.

Meanwhile the PF accumulates a number of customers waiting for the server. However if the queue is below some N (N-policy), the server returns to the SF and continues processing waiting jobs there on the packet-by-packet basis. Upon completion of a packet the server will return to the PF, if the queue out there does cross N.

To calculate the joint distribution of the number of customers and jobs accumulated in the PF and processed at the SF, respectively, we used fluctuation analysis [1] applied to the system functioning during its two phases and with two active thresholds, L and N associated with phases I and II, respecively.

The unique feature of this model that differs from all other vacationing models is that there the servers are sent to do unspecified tasks during their vacationing time. In contrast, in our model the server is released to do a recorded task. An example of regular vacationing system is a basic N-policy system where the server remains on vacation until the buffer load reaches level N.

The present paper continues our studies initiated in [1] on the same system where we focused on time insensitive processes upon departures of customers and the beginnings of busy periods. Using the results obtained in [1] and time sensitive analysis we investigate the continuous time parameter queueing process.

The reader can be referred to our prequel paper [1] for the literature on the subject. In a nutshell, we mention that our methods relate to fluctuation theory [2, 3] and time sensitive analysis [4]. Topically, it comes relatively close to Tian and Zhang's monograph [5] on queues with vacations.

Al-Matar and Dshalalow [6] studied somewhat similar model, in which the server leaves the PF to work at the SF under the same assumptions. However, the time which the server spends at the SF as per [6] is limited directly by a fixed positive real T, rather than by the quantity of processed jobs controlled by the threshold L as it is in our case. The model studied in [6] offers a different way of controlling the efficiency of the vacating server involved in sequential processing.

The paper is laid out as follows. Section 2 gives a general background and provides a formal description of the model for phase I and phase II. Sections 3—5 treat continuous time parameter queueing process using time sensitive analysis and semi-regenerative techniques. Among other things, we evaluate the mean stationary service cycle and the probability generating function of the continuous time parameter queueing process in equilibrium. Section 6 deals with performance measures, such as the mean number of switchovers (between server's appointments at the PF and SF), the mean buffer load at the PF, and the mean quantity of jobs rendered at the SF, all per unit time. The paper concludes with numerical illustrations of some optimization problems.

2. The anatomy of the model and past results. ***The formalism of the model.*** When the system gets exhausted the server leaves for the SF to work on second priority jobs. They are organized in packets of random sizes. Each unit of a packet needs a random processing time from the equivalence class [δ] forming independent and identical distributed (iid) random variables (r.v.'s). If the *k*th packet contains X_k of jobs, which takes time Δ_k and if τ_n is the time needed to process n packets, then it can be formalized as follows:

$$\tau_n = \Delta_1 + \ldots + \Delta_n = \delta_{11} + \ldots + \delta_{X_1 1} + \ldots + \delta_{1n} + \ldots + \delta_{X_n n},$$

where $\delta_{ij} \in [\delta]$ is the processing time of the *i*th job from the *j*th packet.

***The system on phase* I.** The server stays at the SF after processing an amount of jobs specified by a threshold L.Namely, if for some $n = 1, 2, \ldots,$

$$A_n = X_1 + \ldots + X_n \geq L, \tag{1}$$

while $A_{n-1} < L$, the server will return to the system completing what we call phase I. We thus define $\nu := \inf \{n : A_n \geq L\}$ along with the following key r.v.'s: A_ν is the number of jobs (in excess of L) by the end of phase I, τ_ν is the time of server's return at the end of phase I (exit time I).

Meanwhile, the system will replenish with primary customers during the server's absence at the SF. Its number is specified as follows. The number of customers accumulated during phase I is $B_\nu = Y_1 + \ldots + Y_\nu$, where in particular, Y_1 is the number of arriving customers in interval $[0, \Delta_1]$, Y_i is the number of arriving customers in interval $(\tau_{i-1}, \tau_i]$, $i = 2, 3\ldots$.

Whether or not the server will enter phase II of maintenance depends on whether or not $B_\nu \geq N$, where N is yet another threshold associated with the PF and being a part of widely referred to as «*N*-Policy», only in our case within a more complex servicing system. This will be continued in the next subsection.

The first phase of the system is specified by the trivariate «interdependent» process on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$:

$$(\mathcal{A}, \mathcal{B}, \mathcal{T}) := \sum_{i=0}^{\infty} (X_i, Y_i)\, \varepsilon_{\tau_i},$$

(ε_a is the unit mass at a) is a bivariate marked point process with position dependent marking. The value τ_ν is the hitting time or exit time of the random walk from the set $[0, L) \times \mathrm{N} \times \mathrm{R}_+$. The first component $\mathcal{A}$ of the process is «active» and $\mathcal{B}$ and $\mathcal{T}$ are «passive» because only $\mathcal{A}$ is watched to cross L, while the rest of the components take their associated values at the crossing by $\mathcal{A}$.

The input process is compound Poisson of rate λ with

$$\mathcal{I} = \sum_{k=1}^{\infty} U_k \varepsilon_{t_k}, \;\; Ez^{U_k} = a(z), \;\; a := EU_k < \infty, \tag{2}$$

which is to be «observed» at times τ_i's with respective increments Y_i of the arriving customers. As mentioned above, τ_i's are successive completions of packs of jobs rendered by the server during its maintenance time. In our prequel paper [1] we calculated the joint functional

$$\gamma(u, z, \theta) = Eu^X z^Y e^{-\theta\Delta} = G[u\delta(\theta + \lambda - \lambda a(z))],$$

where the increments of $(\mathcal{A}, \mathcal{B}, \mathcal{T})$ $X_i \in [X], Y_i \in [Y], \Delta_i \in [\Delta]$ are copies from the associated equivalence classes, $\delta(\theta) = Ee^{-\theta\delta}$, $\hat{\delta} := E\delta < \infty$, is the LST and the expected value, respectively, of processing time $\delta_j \in [\delta]$ of a job and $G(u) = Eu^X$, $\hat{X} := EX < \infty$, is the pgf and the expected value, respectively, of a pack size.

From [1, Theorem 2] we have

Theorem 1. The joint functional $\Phi_\nu(u, z, \theta)$ of phase I satisfies

$$\Phi_\nu(u, z, \theta) = E[u^{A_\nu} z^{B_\nu} e^{-\theta\tau_\nu}] =$$

$$= 1 - [1 - G[u\delta(\theta + \lambda - \lambda a(z))]] \mathcal{D}_x^{L-1} \left\{ \frac{1}{1 - G[u\delta(\theta + \lambda - \lambda a(xz))]} \right\}, \tag{3}$$

where the operator $\mathcal{D}^k$ is defined as

$$k \mapsto \mathcal{D}_x^k \varphi(x, y) = \begin{cases} \lim\limits_{x \to 0} \dfrac{1}{k!} \dfrac{\partial^k}{\partial x^k} \left[\dfrac{1}{1-x} \varphi(x, y) \right], & k \ge 0, \\ 0, & k < 0, \end{cases}$$

if applied to a function $\varphi(x, y)$ analytic at zero in the first variable. □

After making some special assumptions on $\delta(\theta)$ and $G(u)$

(i) δ is exponential with parameter $d > 0$, i.e. $\delta(\theta) = Ee^{-\theta\delta} = \dfrac{d}{d + \theta}$,

(ii) The number of jobs in a packet is geometric with parameter p, i.e.

$$G(u) = Eu^X = \frac{pu}{1 - qu},$$

formula (1) reduced to

$$\Phi_\nu(u, z, \theta) = \frac{pdu}{d + \theta + \lambda - \lambda a(z) - dqu} \left(\frac{du}{d + \theta + \lambda - \lambda a(z)} \right)^{L-1}. \tag{4}$$

***The system on phase* II.** Now, upon exiting from phase I, or equivalently, returning to the system, the server evaluates the status of the buffer if its content exceeds $N - 1$, it immediately resumes its service at the PF. Otherwise, it enters phase II, which requires the server to commute to the SF working on one pack at a time and checking if the length exceeds $N - 1$. So, it continues its work at the SF up until it takes place. As already mentioned, it does not break its servicing on any pack which it has to finish even if the queue has exceeded $N-1$ in the middle of its work.

Consequently, phase II evolves in accordance with yet another random walk

$$(\widetilde{\mathcal{A}},\widetilde{\mathcal{B}},\widetilde{\mathcal{T}}):=\sum_{i=0}^{\infty}(\widetilde{X}_i,\widetilde{Y}_i)\varepsilon_{\widetilde{\tau}_i} \tag{5}$$

that proceeds somewhat similarly, except that the second component $\widetilde{\mathcal{B}}$ is now active while the rest of them, $\widetilde{\mathcal{A}}$, $\widetilde{\mathcal{T}}$ are passive. Furthermore, the initial values of $(\widetilde{\mathcal{A}},\widetilde{\mathcal{B}},\widetilde{\mathcal{T}})$, namely $\widetilde{X}_0,\widetilde{Y}_0,\widetilde{\Delta}_0=\widetilde{\tau}_0$ are now exit components of $(\mathcal{A},\mathcal{B},\mathcal{T})$. Consequently, the initial functional $E[u^{\widetilde{X}_0}z^{\widetilde{Y}_0}e^{-\theta\widetilde{\Delta}_0}]$ is equated to $\Phi_\nu(u,z,\theta)$ of (4). The random walk $(\widetilde{\mathcal{A}},\widetilde{\mathcal{B}},\widetilde{\mathcal{T}})$ of (5) now will be terminated once $\widetilde{Y}_0+\ldots+\widetilde{Y}_n$ exceeds $N-1$ for some n taking place at some $\widetilde{\tau}_n$. Define the second exit index $\rho=\min\{n:\widetilde{B}_n=\widetilde{Y}_0+\ldots+\widetilde{Y}_n\geq N\}$.

The functional of our further interest will be $\Psi_\rho(u,z,\theta):=E[u^{\widetilde{A}_\rho}z^{\widetilde{B}_\rho}e^{-\theta\widetilde{\tau}_\rho}]$ with the following key r.v.'s:

$\widetilde{A}_\rho$ is the number of jobs done at the SF by the end of phase II;

$\widetilde{B}_\rho$ is the number of customers in the buffer by the end of phase II;

$\widetilde{\tau}_\rho$ is the exit time II (from phase II).

According to [1, (23)—(25)], the exit functional satisfies the formula

$$\Psi_\rho(u,z,\theta):=E[u^{\widetilde{A}_\rho}z^{\widetilde{B}_\rho}e^{-\theta\widetilde{\tau}_\rho}]=$$

$$=\Phi_\nu(u,z,\theta)-[1-G[u\delta(\theta+\lambda-\lambda a(z))]]\mathcal{D}_y^{N-1}\left\{\frac{\Phi_\nu(u,yz,\theta)}{1-G[u\delta(\theta+\lambda-\lambda a(yz))]}\right\}. \tag{6}$$

To arrive at readily tractable form of $\Psi_\rho(u,z,\theta)$ we proceeded in [1] with one more assumption, namely, we reduced the input $\mathcal{I}$ to an ordinary Poisson, i.e. setting $a(z)=z$. Having done this, we had two compact expressions for $\Psi_\rho(u,z,\theta)$. In case of $L=1$,

$$\Psi_\rho(u,z,\theta)=\frac{pdu}{d+\theta+\lambda-\lambda z-dqu}+\left(\frac{\lambda z}{d+\theta+\lambda-du}\right)^N \tag{7}$$

and in case of $L\geq 2$,

$$\Psi_\rho(u,z,\theta)=\frac{p(du)^L}{d+\theta+\lambda-\lambda z-dqu}\times$$

$$\times\left[\frac{1}{(d+\theta+\lambda-\lambda z)^{L-1}}-\frac{1}{(d+\theta+\lambda)^{L-1}}\sum_{j=0}^{N-1}\binom{L+j-2}{j}\times\right.$$

$$\times\left(\left(\frac{\lambda z}{d+\theta+\lambda}\right)^{j}-\left(\frac{\lambda z}{d+\theta+\lambda-du}\right)^{N}\left(\frac{d+\theta+\lambda-du}{d+\theta+\lambda}\right)^{j}\right)\Bigg]. \tag{8}$$

We recall that formulas (7) and (8) being free of operator $\mathcal{D}$ were due to the three assumptions:

(i) $\delta(\theta)=Ee^{-\theta\delta}=\dfrac{d}{d+\theta}$ (transform of processing time of a job),

(ii) $G(u)=Eu^{X}=\dfrac{pu}{1-qu}$ (pgf of the size of a pack),

(iii) $a(z)=z$ (pgf of the number of customers in a batch).

In [1] we also investigated another special case of Ψ_ρ under the assumption that $N=1$ however, retaining the generality of the input stream. Thus under the assumptions

(i) $\delta(\theta)=Ee^{-\theta\delta}=\dfrac{d}{d+\theta}$,

(ii) $G(u)=Eu^{X}=\dfrac{pu}{1-qu}$,

(iii') $N=1$,

we had

$$\Psi_\rho(u,z,\theta)=E[u^{\tilde{A}_\rho}z^{\tilde{B}_\rho}e^{-\theta\tilde{\tau}_\rho}]=\frac{p(d+\theta+\lambda-\lambda a(z))\left(\dfrac{du}{d+\theta+\lambda-\lambda a(z)}\right)^{L}}{d+\theta+\lambda-\lambda a(z)-duq}-$$

$$-\frac{d+\theta+\lambda-\lambda a(z)-du}{d+\theta+\lambda-\lambda a(z)-duq}\,\frac{p(d+\theta+\lambda)\left(\dfrac{du}{d+\theta+\lambda}\right)^{L}}{d+\theta+\lambda-duq}\left(q+p\frac{1}{1-\dfrac{du}{d+\theta+\lambda}}\right). \tag{9}$$

Remark 1. Notice that in this special case, if we assume that threshold L is random, say with the pgf

$$g(w)=Ew^{L}, \tag{10}$$

then the functional $\Psi_\rho(u,z,\theta)$ in (9) can be interpreted as a conditional expectation $\Psi_\rho(u,z,\theta|L)=E[u^{\tilde{A}_\rho}z^{\tilde{B}_\rho}e^{-\theta\tilde{\tau}_\rho}|L]$. Consequently, from (9) and with (10) in mind, using the double expectation formula, we get

$$\hat{\Psi}_\rho(u,z,\theta):=E[E[u^{\tilde{A}_\rho}z^{\tilde{B}_\rho}e^{-\theta\tilde{\tau}_\rho}|L]]=\frac{p(d+\theta+\lambda-\lambda a(z))g\left(\dfrac{du}{d+\theta+\lambda-\lambda a(z)}\right)}{d+\theta+\lambda-\lambda a(z)-duq}-$$

$$-\frac{d+\theta+\lambda-\lambda a(z)-du}{d+\theta+\lambda-\lambda a(z)-duq}\,\frac{p(d+\theta+\lambda)\,g\left(\dfrac{du}{d+\theta+\lambda}\right)}{d+\theta+\lambda-duq}\left(q+p\frac{1}{1-\dfrac{du}{d+\theta+\lambda}}\right).$$

Embedded queueing process. We make the usual assumptions on the service process for the $M/G/1$ type of model that service times $\sigma_1, \sigma_2, \ldots$ are iid r.v. with a common LST $\beta(\theta)=Ee^{-\sigma_1}$, $\mathrm{Re}(\theta)\geq 0$, and $b:=E\sigma_1<\infty$ and that T_0, T_1, $T_2, \ldots$ are successive departures of the individually processed units. Furthermore, $\{Q_n := Q(T_n);\, n=0, 1, \ldots\}$ is the associated embedded process. The input is described in subsection (1), see formula (2).

According to [1], the process Q_n is ergodic if and only if $\rho=\lambda ab<1$ and its pgf $P(z)$ for the steady state distribution satisfies the Kendall formula

$$P(z)=p_0\beta(\lambda-\lambda a(z))\,\frac{\alpha(z)-1}{z-\beta(\lambda-\lambda a(z))}, \tag{11}$$

where $\alpha(z)=Ez^{\widetilde{B}_\rho}$ is the marginal functional of $\Psi_\rho(u,z,\theta)=E[u^{\widetilde{A}_\rho}z^{\widetilde{B}_\rho}e^{-\theta\widetilde{\tau}_\rho}]$ for the three cases of (7)—(9) from which under assumptions (*i-iii*) and (*i-iii*') we get $\alpha(z)$ as follows.

For the case of $L=1$, $a(z)=z$, and N is arbitrary,

$$\alpha(z)=\frac{pd}{\lambda-\lambda z+pd}+z^N. \tag{12}$$

For the case of $L\geq 2$, $a(z)=a$, and N is arbitrary,

$$\alpha(z)=\frac{pd^L}{(pd+\lambda-\lambda z)(d+\lambda-\lambda z)^{L-1}}-\frac{pd}{pd+\lambda-\lambda z}\left(\frac{d}{d+\lambda}\right)^{L-1}\times$$
$$\times\sum_{j=0}^{N-1}\left(\frac{L+j-2}{j}\right)\left(\frac{d}{d+\lambda}\right)^j(z^j-z^N). \tag{13}$$

For the case of $a(z)$ arbitrary and $N=1$,

$$\alpha(z)=\frac{pd^L}{(pd+\lambda-\lambda a(z))(d+\lambda-\lambda a(z))^{L-1}}-\frac{pd(1-a(z))}{pd+\lambda-\lambda a(z)}\left(\frac{d}{d+\lambda}\right)^{L-1}. \tag{14}$$

Also, in Kendall's formula,

$$p_0=\frac{1-\rho}{\alpha} \tag{15}$$

with

$$\alpha=E\widetilde{B}_\rho \tag{16}$$

and the three respective cases of α:

$L=1,\ a(z)=z,$

$$\alpha=\lim_{z\to 1}\frac{d\alpha(z)}{dz}=\frac{\lambda}{pd}+N; \tag{17}$$

$L\geq 2,\ a(z)=z,$

$$\alpha=\frac{\lambda}{d}\left(\frac{1}{p}+L-1\right)+\left(\frac{d}{d+\lambda}\right)^{L-1}\sum_{j=0}^{N-1}\binom{L+j-2}{j}\left(\frac{\lambda}{d+\lambda}\right)^{j}(N-j); \tag{18}$$

$N=1,$

$$\alpha=a\frac{\lambda}{d}\left(\frac{1}{p}+L-1\right)+a\left(\frac{d}{d+\lambda}\right)^{L-1}. \tag{19}$$

3. A semi-regenerative process. To further enhance the results obtained in part I of [1] as well as arrive at important performance measures and optimize the system we consider the continuous time parameter queueing process. In particular, it will enable us to relate the pertinent functionals to real time and bring the forthcoming performance measures such as the quantity of finished jobs at the SF, number of switchovers, and buffer load at the PF, all in a unit time interval. The available functionals so far obtained are «time insensitive».

As previously mentioned, the supplementary variable technique (introduced by D. Cox a long time ago) very often used in the contemporary literature on queueing has several shortcomings. For one, it works poorly with fluctuations and requires time insensitivity of the primary functionals like those of section 2. Furthermore, the method puts a stiff requirement on the service time distribution as being absolutely continuous and hence it leaves behind all discrete and mixed distributions. We begin with the notion of a semi-regenerative process and as $Q(t)$ definitely is [4, 6].

Definition 1. Let $(\Omega, \mathcal{F}(\Omega), \mathcal{F}_t, (P^x)_{x=0,1,\ldots})$ be a filtered probability space, T be a stopping time relative to the filtration and a process $Q=(Q(t))$ (valued in a discrete topological space) is $(\mathcal{F}_t)$-adapted. Q is said to have a locally strong Markov property at T if for each $t\geq 0$, $E^x[g(Q(t+T))|\mathcal{F}_T]=E^{Q_T}[g(Q(t))]$, for any Borel measurable and integrable function g. The process $Q\geq 0$ is called semi-regenerative if

(i) there is a point process $\{T_n; n=0,1,\ldots\}$ such that for each n, T_n is a stopping time relative to $(\mathcal{F}_t)$;

(ii) Q has a locally strong Markov property at T_n, $n=0,1,\ldots$;

(iii) Q is a.s. right-continuous;

(iv) $(Q(T_n), T_n):=(Q_n, T_n)$ is a time-homogeneous Markov renewal process (which is embedded in Q over $\{T_n\}$). □

Let Q be a semi-regenerative process. For a nonnegative integer value j, let

$$K_{ik}(t) := P^i\{Q(t) = k, T_1 > t\} = P\{Q(t) = k, T_1 > t | Q_0 = i\}. \tag{20}$$

Then, the functional matrix $K(t) := \{K_{ik}(t); i, j = 0, 1, ...\}$ is called the semi-regenerative kernel. The stationary probabilities (provided that the equilibrium condition is met) are

$$\pi_k = \lim_{t\to\infty} P\{Q(t) = k | Q(0) = i\} = \frac{1}{<\mathbf{p}, \mathbf{c}>} \sum_{j=0}^{\infty} p_j \int_{s=0}^{\infty} K_{jk}(u)\, du, \tag{21}$$

where $\mathbf{p} = (p_0, p_1, ...)$ is the invariant probability measure of the transition probability matrix P of $Q_n = Q(T_n)$ and $\mathbf{c} = (c_0, c_1, ...)^T$ with $c_j = E^j[T_1]$. The inner product

$$\mathcal{C} := <\mathbf{p}, \mathbf{c}> \tag{22}$$

is called the mean stationary service cycle. Denote matrix

$$H := (h_{ik}; j, k = 0, 1, ...) = \int_{s=0}^{\infty} K(u)\, du \tag{23}$$

and call it the integrated semi-regenerative kernel. It can be readily shown that the semi- regenerative kernel K is integrable. Let $h_i(z)$ be the pgf of the ith row of H. Then, from (20) and (23) we have $h_i(z) = \int_{t=0}^{\infty} E^i[z^{Q(t)} \mathbf{1}_{\{t<T_1\}}]\, dt$. Furthermore, with notation (22), equation (21) can be rewritten as

$$\Pi(z) = \frac{1}{c} \sum_{j=0}^{\infty} h_j(z)\, p_j, \tag{24}$$

where $\Pi(z)$ is the pgf of the $\mathbf{p}$. With the notation

$$\mathbf{h}(z) = (h_0(z), h_1(z), ...)^T \tag{25}$$

we can rewrite (24) in the compact form

$$\Pi(z) = \frac{1}{c} < \mathbf{p}, \mathbf{h}(z) >. \tag{26}$$

Since our model is of $M/G/1$ type, the queueing process $Q(t)$ is indeed semi-regenerative [4] relative to the Markov renewal process (Q_n, T_n). In the upcoming sections we will work on obtaining an explicit expression for $\mathbf{h}(z)$ and $\mathcal{C}$.

4. Time sensitive analysis. The below statement is due to Theorem 6 of [4].

Proposition 1. In a queue with a marked Poisson input stream and general service, suppose $Q_0 = 0$ and thus the first service cycle $[0, T_1]$ consists of the first service period preceded by a random walk process that lasts from $T_0 = 0$ to $\tilde{\tau}_\rho$. Upon this event, the queue length equals $\tilde{B}_\rho$ and the first service begins. (Customers, thereafter, continue entering the system.) The Laplace functional of the continuous time parameter queueing process $Q(t)$ observed over the first service period, jointly with the first passage time $\tilde{\tau}_\rho$ satisfies the formula

$$\phi(z, \vartheta, \theta) := \int_{t=0}^{\infty} e^{-\theta t} E[z^{Q(t)} e^{-\vartheta \tilde{\tau}_\rho} \mathbf{1}_{\{\tilde{\tau}_\rho \le t < \tilde{\tau}_\rho + \sigma_1\}}] dt =$$

$$= \frac{1 - \beta(\theta + \lambda - \lambda a(z))}{\theta + \lambda - \lambda a(z)} E[z^{\tilde{B}_\rho} e^{-(\vartheta + \theta)\tilde{\tau}_\rho}],$$

where $\beta(\theta) = Ee^{-\theta\sigma_1}$. □

While this is a worthwhile information on the process in the random interval $[\tilde{\tau}_\rho, \tilde{\tau}_\rho + \sigma_1)$, it is not yet what we exactly need as per $h_i(z) :=$ $:= \int_{t=0}^{\infty} E[z^{Q(t)} \mathbf{1}_{\{t < T_1\}} | Q_0 = i] dt$ treated in the next section. We thus need to observe the queue over the entire interval $[0, T_1]$ known as the first service cycle, where $T_1 = \tilde{\tau}_\rho + \sigma_1$. As mentioned earlier, the first service cycle will be partitioned into $[0, \tilde{\tau}_\rho] \bigcup [\tilde{\tau}_\rho, \tilde{\tau}_\rho + \sigma_1)$.

Theorem 2 [6]. Under the condition of Proposition 1, the following formula holds true

$$\int_{t=0}^{\infty} e^{-\theta t} E[z^{Q(t)} \mathbf{1}_{\{t < \tilde{\tau}_\rho + \sigma_1\}}] dt =$$

$$= \frac{Ez^{Q_0} - E[z^{\tilde{B}_\rho} e^{-\theta\tilde{\tau}_\rho}]}{\theta + \lambda - \lambda a(z)} + \frac{1 - \beta(\theta + \lambda - \lambda a(z))}{\theta + \lambda - \lambda a(z)} E[z^{\tilde{B}_\rho} e^{-\theta\tilde{\tau}_\rho}],$$

where $\beta(\theta) = Ee^{-\theta\sigma_1}$. □

We now utilize Theorem 2 in the light of functional (24) where $T_1 = \tilde{\tau}_\rho + \sigma_1$ is the end of the first service cycle. (As per our agreement, we set $\tilde{\tau}_\rho = 0$ if $Q_0 > 0$ and the first service will start immediately following $T_0 = 0$.)

Theorem 3. The inner product of the invariant probability measure $\mathbf{p}$ and $\mathbf{h}(z)$ of (25) satisfies the formula $< \mathbf{p}, \mathbf{h}(z) > = \frac{1}{\lambda\alpha} \Delta_p(z) K(z)$, where $\Delta_p(z) =$ $= \frac{1 - \alpha(z)}{1 - a(z)}$, $\alpha(z) = Ez^{\tilde{B}_\rho}$, and

$$K(z) = (1 - \rho)\beta(\lambda - \lambda a(z)) \frac{1 - z}{\beta(\lambda - \lambda a(z)) - z}.$$

P r o o f. For $Q_0 = i = 0$ and $\theta = 0$, we have from the following formula

$$h_i(z) = \frac{Ez^0 - \alpha(z)}{\lambda - \lambda a(z)} + \Delta(z)\alpha(z) = \frac{1}{\lambda}\Delta_p(z) + \Delta(z)\alpha(z), \tag{27}$$

where

$$\Delta(z) := \frac{1 - \beta(\lambda - \lambda a(z))}{\lambda - \lambda a(z)}. \tag{28}$$

Now, for $Q_i = i > 0$ we have $\tilde{\tau}_\rho = 0$ and $\tilde{B}_\rho = Q_i = i$. Setting $\theta = 0$ from (27),we have $h_i(z)$ reduce to

$$h_i(z) = z^i \Delta(z),\ i > 0. \tag{29}$$

Since the queueing process $Q(t)$ under investigation is semi-regenerative, the pgf $\Pi(z)$ of $Q(t)$ in the steady state according to (26), satisfies the functional equation

$$\Pi(z) = \sum_{i=0}^{\infty} \pi_i z^i = \frac{1}{\mathcal{C}} < \mathbf{p}, \mathbf{h}(z) >. \tag{30}$$

With (27) — (30) we have

$$< \mathbf{p}, \mathbf{h}(z) >= p_0 h_0(z) + \sum_{i=1}^{\infty} p_i h_i(z) =$$

$$= p_0\left(\frac{1}{\lambda}\Delta_p(z) + \Delta(z)\alpha(z)\right) + \sum_{i=1}^{\infty} z^i \Delta(z)\, p_i =$$

$$= \Delta(z)\, P(z) + p_0\left[\frac{1}{\lambda}\Delta_p(z) - [1 - \alpha(z)]\Delta(z)\right] =$$

$$= \Delta(z)\, P(z) + p_0 \frac{1}{\lambda}\Delta_p(z)\beta(\lambda - \lambda a(z))$$

by (11) for $P(z)$

$$< \mathbf{p}, \mathbf{h}(z) > = p_0 \frac{1}{\lambda}\Delta_p(z)\beta(\lambda - \lambda a(z))\frac{1 - z}{\beta(\lambda - \lambda a(z)) - z}, \tag{31}$$

where $p_0 = \frac{1-\rho}{\alpha}$, and α satisfies formulas (14)—(16).

5. The service cycle $\mathcal{C}$ and the PGF $\Pi(z)$. The interval $(T_n, T_{n+1}]$ may contain the nth service time and the period of absence dependent on what state the queue was at T_n. The mean value of the length of this interval in equilibrium is the mean stationary service cycle

$$\mathcal{C} = <\mathbf{p}, \mathbf{c}> = \sum_{i=0}^{\infty} p_i c_i,$$

where $c_i = E[T_{n+1} - T_n | Q_n = i]$. This is a part in formula (30) which is left behind, and that is what we will be concerned with in this section.

Proposition 2. The mean stationary service cycle $\mathcal{C}$ equals

$$\mathcal{C} = \frac{1}{\lambda a}. \tag{32}$$

P r o o f. Obviously,

$$c_i = \begin{cases} E\,\tilde{\tau}_\rho + b, & i = 0, \\ b, & i > 0, \end{cases}$$

where $b = E\sigma_1$. Thus,

$$\mathcal{C} = <\mathbf{p}, \mathbf{c}> = p_0 E\,\tilde{\tau}_\rho + b = \frac{1-\rho}{\alpha} E\,\tilde{\tau}_\rho + \frac{\rho}{\lambda a}. \tag{33}$$

Now, if we prove that

$$E\,\tilde{\tau}_\rho = \frac{\alpha}{\lambda a} = \frac{E\tilde{B}}{\lambda a}. \tag{34}$$

then (33) will immediately yield (32). We turn to the marginal functional $\Psi_\rho(1, z, \theta) = Ez^{\tilde{B}_\rho} e^{-\theta \tilde{\tau}_\rho}$ which by (6), for $u = 1$ is

$$\Psi_\rho(1, z, \theta) = \Phi_\nu(1, z, \theta) -$$
$$-[1 - G[\delta(\theta + \lambda - \lambda a(z))]]\, \mathcal{D}_y^{N-1} \left\{ \frac{\Phi_\nu(1, yz, \theta)}{1 - G[\delta(\theta + \lambda - \lambda a(yz))]} \right\},$$

where from (3)

$$\Phi_\nu(1, z, \theta) = E[z^{B_\nu} e^{-\theta \tau_\nu}] =$$
$$= 1 - [1 - G[\delta(\theta + \lambda - \lambda a(z))]]\, \mathcal{D}_x^{L-1} \left\{ \frac{1}{1 - G[\delta(\theta + \lambda - \lambda a(xz))]} \right\}. \tag{35}$$

Denote

$$\psi(z, \theta) := -[1 - G[\delta(\theta + \lambda - \lambda a(z))]]\, \mathcal{D}_y^{N-1} \left\{ \frac{\Phi_\nu(1, yz, \theta)}{1 - G[\delta(\theta + \lambda - \lambda a(yz))]} \right\}. \tag{36}$$

We start with the mean number of customers EB_ν accumulated at the PF upon the exit from phase I. From (35)

$$\lim_{z\to 1}\frac{d}{dz}\Phi_\nu(1,z,0)=EB_\nu=$$

$$=\left[\lim_{z\to 1}\frac{d}{dz}G[\delta(\lambda-\lambda a(z))]\right]\mathcal{D}_x^{L-1}\left\{\frac{1}{1-G[\delta(\lambda-\lambda a(x))]}\right\}=$$

$$=G'(\delta(0))\,\delta'(0)(-\lambda a)\,\mathcal{D}_x^{L-1}\left\{\frac{1}{1-G[\delta(\lambda-\lambda a(x))]}\right\}=$$

$$=\lambda a\hat{X}\hat{\delta}\mathcal{D}_x^{L-1}\left\{\frac{1}{1-G[\delta(\lambda-\lambda a(x))]}\right\}. \tag{37}$$

Now, we compare it with the mean duration time $E\tau_\nu$ of phase I:

$$\lim_{\theta\to 0}(-1)\frac{d}{d\theta}\Phi_\nu(1,1,\theta)=E\tau_\nu=$$

$$=-\lim_{\theta\to 0}\frac{d}{d\theta}G(\delta(\theta))\,\mathcal{D}_x^{L-1}\left\{\frac{1}{1-G[\delta(\lambda-\lambda a(x))]}\right\}=$$

$$=G'(1)(-\delta'(0))\,\mathcal{D}_x^{L-1}\left\{\frac{1}{1-G[\delta(\lambda-\lambda a(x))]}\right\}=$$

$$=\hat{X}\hat{\delta}\mathcal{D}_x^{L-1}\left\{\frac{1}{1-G[\delta(\lambda-\lambda a(x))]}\right\}. \tag{38}$$

From (37) and (38) we easily conclude that

$$EB_\nu=\lambda a E\tau_\nu. \tag{39}$$

Now we turn to $\psi(z,\theta)$ of (36):

$$\lim_{z\to 1}\frac{d}{dz}\psi(z,0)=G'(\delta(0))\,\delta'(0)(-\lambda a)\,\mathcal{D}_y^{N-1}\left\{\frac{\Phi_\nu(1,y,0)}{1-G[\delta(\lambda-\lambda a(y))]}\right\}=$$

$$=\lambda a\hat{X}\hat{\delta}\mathcal{D}_y^{N-1}\left\{\frac{\Phi_\nu(1,y,0)}{1-G[\delta(\lambda-\lambda a(y))]}\right\}.$$

followed by

$$-\lim_{\theta\to 0}\frac{d}{d\theta}\psi(1,\theta)=-G'(\delta(0))\,\delta'(0)\,\mathcal{D}_y^{N-1}\left\{\frac{\Phi_\nu(1,y,0)}{1-G[\delta(\lambda-\lambda a(y))]}\right\}=$$

$$=\hat{X}\hat{\delta}\mathcal{D}_y^{N-1}\left\{\frac{\Phi_\nu(1,y,0)}{1-G[\delta(\lambda-\lambda a(y))]}\right\}=\frac{1}{\lambda a}\lim_{z\to1}\frac{d}{dz}\psi(z,0). \tag{40}$$

Since $E\widetilde{B}_\rho = EB_\nu + \psi_z(1,0)$ and $E\widetilde{\tau}_\rho = E\tau_\nu - \psi_\theta(1,0)$, due to (39) and (40) we have (34) and thus the statement.

Remark 2. Formula (32) is known to hold for the basic system $M^X/G/1$ with no impact of the idle period. Surprisingly, the same trend applies for a far more complex queue and yet it is as insensitive in spite of the presence of two phases. Another remarkable fact is that Proposition 2 was proved with no special assumptions made on service time at the SF as well as the job packs distribution. We note that formula (32) for the mean stationary service cycle holds for a different system studied in [6]. We also utilized a similar idea when proving Proposition 2. □

Since the queueing process $Q(t)$ under investigation is semi-regenerative relative to the departure process $\{T_n\}$, the pgf of $Q(t)$ in its steady state $\Pi(z)$ according to (26) satisfies the functional equation

$$\Pi(z)=\sum_{i=0}^{\infty}\pi_i z^i=\frac{1}{\mathcal{C}}<\mathbf{p},\mathbf{h}(z)>,$$

where by (32), $\mathcal{C}=\frac{1}{\lambda a}$. Combining the latter with Theorem 3 we have

$$\Pi(z)=\lambda a<\mathbf{p},\mathbf{h}(z)>=\lambda a\frac{1}{\lambda\alpha}\Delta_p(z)K(z)=\frac{a}{\alpha}\Delta_p(z)K(z), \tag{41}$$

where

$$\Delta_p(z)=\frac{1-\alpha(z)}{1-a(z)}, \tag{42}$$

$$K(z)=(1-\rho)\beta(\lambda-\lambda a(z))\frac{1-z}{\beta(\lambda-\lambda a(z))-z}. \tag{43}$$

Here $\alpha(z)$ satisfies formulas (12)—(14) for three special cases and α is from (17)—(19), respectively.

From (41)—(43) there is probably no customer in the system at any time of equilibrium:

$$\pi_0=\Pi(0)=\lambda a\,\frac{1-\rho}{\alpha}\frac{1}{\lambda}\beta(\lambda)\frac{1}{\beta(\lambda)}=ap_0. \tag{44}$$

The latter is much less significant compared to the standard $M^X/G/1$ system, because this probability has no relevance with the period when the server is idle for two reasons. Firstly it is never really idle and secondly, the period when there is no customer in the system is a small portion of the entire maintenance period consisting of two phases.

The above can be summarized as follows.

Theorem 4. The continuous time parameter queueing process $Q(t)$ has its steady state under the necessary and sufficient condition that $\rho = \lambda ab < 1$. The corresponding stationary distribution $\pi = (\pi_0, \pi_1, \ldots)$ of $Q(t)$ in the form of the pgf $\Pi(z)$ satisfies formula (41)—(44).

6. The main performance measures. ***The expected number of processed jobs per unit time.*** In this subsection we will evaluate the expected number of processed jobs $\widetilde{A}_\rho$, by the end of phase II, under the same assumptions (*i*—*iii*) for the first two cases in Section 2. For the case of $L = 1$, from (7), we have from the marginal pgf

$$Eu^{\widetilde{A}_\rho} = \frac{pu}{1-qu} + \left(\frac{\lambda}{d+\lambda-du}\right)^N$$

the expected value

$$E\widetilde{A}_\rho = \frac{1}{p} + \frac{d}{\lambda} N. \tag{45}$$

For the case of $L \geq 2$, from (8), we form the marginal pgf

$$Eu^{\widetilde{A}_\rho} = \frac{pu^L}{1-qu} - \frac{pu^L}{1-qu}\left(\frac{d}{d+\lambda}\right)^{L-1} \times$$

$$\times \sum_{j=0}^{N-1} \binom{L+j-2}{j} \left(\left(\frac{\lambda}{d+\lambda}\right)^j - \left(\frac{\lambda}{d+\lambda-du}\right)^N \left(\frac{d+\lambda-du}{d+\lambda}\right)^j \right)$$

the expected value

$$E\widetilde{A}_\rho = \frac{1}{p} + L - 1 + \frac{d}{\lambda}\left(\frac{d}{d+\lambda}\right)^{L-1} \sum_{j=0}^{N-1} \binom{L+j-2}{j} \left(\frac{\lambda}{d+\lambda}\right)^j (N-j). \tag{46}$$

The third case of $E\widetilde{A}_\rho$ deals with a bulk input stream and $N = 1$. From formula (8) we can get the marginal pgf

$$Eu^{\widetilde{A}_\rho} = \frac{pu}{1-qu} u^{L-1} - \frac{pd^L}{(d+\lambda)^{L-1}} \frac{u^L(1-u)}{(1-qu)(d+\lambda-dqu)} \left(q + \frac{p(d+\lambda)}{d+\lambda-du}\right)$$

which yields

$$E\widetilde{A}_\rho = \frac{1}{p} + L - 1 + \frac{d}{\lambda}\left(\frac{d}{d+\lambda}\right)^{L-1}. \tag{47}$$

Comparing $E\widetilde{A}_\rho$ for the three cases in (45)—(47) with the respective values of α (the mean number of customers accumulated in the queue upon the exit from the second phase) in (17)—(19) we notice that in the first two cases with $a(z)=z$ (and $a=1$) and from (34)—(48) we have

$$\frac{1}{d}E\widetilde{A}_\rho = \frac{1}{\lambda}\alpha = E\widetilde{\tau}_\rho, \tag{48}$$

where we recall d^{-1} is the mean processing time of one job and λ^{-1} is the mean interarrival time between single customers that enter the system. Formula (48) can be interpreted as follows. The left-hand side of (48) gives the mean time needed to process all jobs during two phases of server's maintenance. The right-hand side gives the time it takes for α customers to arrive in the system and this is also the mean duration of the two phases.

For the third case we have

$$\frac{1}{d}E\widetilde{A}_\rho = \frac{1}{\lambda a}\alpha = E\widetilde{\tau}_\rho, \tag{49}$$

being essentially the same result as (48) only for $N = 1$ and $a(z)$ unrestricted. Apparently, (49) holds true for the very general case (without assumptions (*i*—*iii*) and (*i*—*iii'*)) which can be proved.

The above expected values of $\widetilde{A}_\rho$ will be used to find the number of processed jobs per unit time in the secondary system using the following arguments. The marked point process $(Q,T)=\{Q_n, T_n : n=0,1,...\}$ is the Markov renewal process. The matrix $\{R^x(j,t): x, j,=0,1,...\}$ is the Markov renewal function. Now, to calculate the required number of processed jobs we observe that

$$R^x(0,t) = E^x\left[\sum_{n\geq 0} \mathbf{1}_{\{0\}}(Q_n)\mathbf{1}_{[0,t]}(T_n)\right]$$

is the expected number of entrances of the system in phases I, II (or equivalently, server's departures from the primary system) during the interval of time [0,t]. If $\widetilde{A}_\rho^{[k]}$ denotes the total number of processed units at the SF during the kth server's sojourn time in the secondary system, then obviously $\widetilde{A}_\rho^{[k]} \in [\widetilde{A}_\rho]$. Now, $\widetilde{A}_\rho^{[k]}$'s are independent and identically distributed with a common mean $E\widetilde{A}_\rho$, which satisfies formulas (45)— (47).

Consequently, the expected number of processed units in interval [0, t] (by using Wald's type equation) will be $E[\widetilde{A}_\rho]R^x(0,t)-l$, where l is the mean number leftovers of the jobs unfinished from the very last cycle overlapping with [0, t]. Obviously, $l \leq E\widetilde{A}_\rho$ and it is the finite number. Thus, $\lim\limits_{t\to\infty}\frac{1}{t}(E[\widetilde{A}_\rho]R^x(0,t)-l)=$

$= \lim\limits_{t\to\infty} \frac{1}{t} (E[\widetilde{A}_\rho] R^x(0,t))$. With this in mind, and from the theory of Markov renewal processes [7, Theorem 5.4.3] we conclude that the mean number of completely processed jobs in the secondary system per unit time interval (over the infinite horizon) is

$$\hat{J} = \lim_{t\to\infty} \frac{1}{t} (E[\widetilde{A}_\rho] R^x(0,t)) = \frac{E\widetilde{A}_\rho}{\mathcal{C}} p_0 = \lambda a p_0 E\widetilde{A}_\rho,$$

with $p_0 = \frac{1-\rho}{\alpha}$ and α satisfying formulas (17)—(19). Using (48) and (49) we thus have

$$\hat{J} = d(1-\rho) = d(1-\lambda ab). \tag{50}$$

Formula (50) shows that the mean number of jobs processed per unit time over the finite horizon does not depend on L, N and on many other parameters.

The mean switchovers rate. A switchover is a change of server's mode from servicing to vacancies, maintenance, or any other form of absence of the server from the PF. It takes place when the server exits for a busy period at the PF, after the queue becomes empty, and it moves to the SF. When the server resumes its service at the PF, the associated busy cycle ends and the new one begins. A busy cycle consists of a busy period and maintenance period that contains two phases. Each busy cycle thus includes exactly one switchover. (We can always double it if necessary to address entrances to busy and vacant modes.)

While server's secondary work is mandatory and even profitable, in some applications, the large number of switchovers is undesirable and they may be induced by a light input traffic. If this is the case, the server may be better off to stay longer at the SF to make sure that enough customers will accumulate at the PF. This can be achieved by means of increasing the thresholds L and N. In both cases, the number of processed jobs at the SF will also increase.

As with other performance measures, the number of switchovers should be related to a fixed time interval, say $[0, t]$, implying a need of the time sensitive analysis. We turn again to the Markov renewal function introduced in the above Section:

$$R^x(0,t) = E^x\left[\sum_{n\ge 0} \mathbf{1}_{\{0\}}(Q_n)\mathbf{1}_{[0,t]}(T_n)\right].$$

Recall that $R^x(0,t)$ gives the total number of entrances of the queue (upon departures) in state $\{0\}$ over the interval $[0, t]$, given that the queue started from state $\{x\}$. As was shown above

$$\lim_{t\to\infty} \frac{R^x(0,t)}{t} = \frac{p_0}{\mathcal{C}}. \tag{51}$$

The latter has the following interpretation. The mean number of entrances of the queueing process into state $\{0\}$upon departures per unit time collected over the infinite horizon is proportional to the steady state probability p_0. Therefore, from (51) we obtain

$$\lim_{t\to\infty}\frac{R^x(0,t)}{t}=\frac{p_0}{\mathcal{C}}=\frac{\lambda a}{\alpha}(1-\rho)$$

the switchover rate. If ξ is the cost of one switchover, then

$$\hat{\xi}=\xi\frac{\lambda a}{\alpha}(1-\rho) \tag{52}$$

is the mean penalty rate for all system's switchovers.

The buffer load rate. Let $q(k)$ be the expense due to the presence of k customers per unit time in the system. Then,

$$Q[0,t]=E^x\left[\int_{u=0}^{t}q(Q(u))\,du\right] \tag{53}$$

gives the mean expense due to all customers present in the system in interval [0, t] . We can represent $Q[0, t]$ as follows, by using the Fubini's theorem:

$$Q[0,t]=\sum_{k\ge1}E^x\left[\int_{u=0}^{t}\mathbf{1}_{\{k\}}(Q(u))\,q(Q(u))\,du\right]=$$

$$=\sum_{k\ge1}q(k)E^x\left[\int_{u=0}^{t}\mathbf{1}_{\{k\}}(Q(u))\,du\right]=\sum_{k\ge1}q(k)\int_{u=0}^{t}P^x\{Q(u)=k\}du. \tag{54}$$

Now use the result from [8, p. 98, Theorem A.1]:

$$\lim_{t\to\infty}\frac{1}{t}\int_{u=0}^{t}P^x\{Q(u)=k\}\,du=\pi_k,\ k=0,1,\ldots$$

Applying (54) to (53) provided the limits are interchangeable we have

$$Q=\lim_{t\to\infty}\frac{Q[0,t]}{t}=\sum_{k\ge1}q(k)\,\pi_k,$$

where Q is the penalty rate for all customers that have ever occupied the system. If $q(k)=qk$, where q is the penalty rate for one customer per unit time, we have

$$Q=q\Pi'(1). \tag{55}$$

We will evaluate $\Pi'(1)$ from (41)—(44). After some lengthy calculations we arrive at

$$\Pi'(1)=\frac{\alpha''(1)}{2\alpha}-\frac{\tilde a_2-a}{2a}+\rho+\frac{\lambda}{2(1-\rho)}(\lambda a^2 b_2-\tilde a_2 b+ab), \tag{56}$$

where $\tilde a_2$ is the second moment of U_k (i.e. $\tilde a_2-a=a''(1)$), b_2 is the second moment of σ_1. Notice that for the two special cases (of $L = 1$ and $L\geq 2$ (see Section 2) with $a(z)=z$ we have $a=1$ and $a''(1)=0$. Thus (56) will reduce to

$$\Pi'(1)=\frac{\alpha''(1)}{2\alpha}+\rho+\frac{\lambda^2 b_2}{2(1-\rho)}. \tag{57}$$

Here $\alpha''(1)$ can be found under the same assumptions of the two discussed cases in Section 2:

for the case of $L = 1$, we will use formula (12)

$$\alpha''(1)=2\left(\frac{\lambda}{pd}\right)^2+N(N-1) \tag{58}$$

and for the case of $L\geq 2$, we will use formula (13)

$$\alpha''(1)=\left(\frac{\lambda}{pd}\right)^2((p+2)(L-1)p+2+(L-1)^2 p^2)+$$
$$+\frac{2\lambda}{pd}\left(\frac{\lambda}{d+\lambda}\right)^{L-1}\sum_{j=0}^{N-1}\binom{L+j-2}{j}\left(\frac{\lambda}{d+\lambda}\right)^{j}(N-j)+$$
$$+\left(\frac{\lambda}{d+\lambda}\right)^{L-1}\sum_{j=0}^{N-1}\binom{L+j-2}{j}\left(\frac{\lambda}{d+\lambda}\right)^{j}(N(N-1)-j(j-1)). \tag{59}$$

In case of bulk input stream and $N = 1$, we will use formula (14)

$$\alpha''(1)=\frac{pd\lambda a''(1)[1+(L-1)p]+p\lambda^2(a)^2(L-1)(p-1)+\lambda^2 a^2[1+(L-1)p][2+(L-1)p]}{p^2 d^2}+$$
$$+a''(1)+\frac{2a^2\lambda}{pd}\left(\frac{d}{d+\lambda}\right)^{L-1}. \tag{60}$$

For further details please see the next calculation.

Calculation of $\alpha''(1)$. In the special case of the ordinary input, i.e., $a(z)=z$, the marginal functionals $\alpha(z)=\Psi_\rho(1,z,0)$ of (12) and (14) will be used to calculate $\alpha''(1)$.

For the case of $L = 1$, recall that

$$\alpha(z)=\frac{pd}{\lambda-\lambda z+pd}+z^N.$$

Thus we get $\alpha''(1)=2\left(\frac{\lambda}{pd}\right)^2+N(N-1)$.

For the case of $L\geq 2$, with

$$\alpha(z)=\frac{pd^L}{(pd+\lambda-\lambda z)(d+\lambda-\lambda z)^{L-1}}-\frac{pd}{pd+\lambda-\lambda z}\left(\frac{d}{d+\lambda}\right)^{L-1}\times$$
$$\times\sum_{j=0}^{N-1}\binom{L+j-2}{j}\left(\frac{\lambda}{d+\lambda}\right)^j(z^j-z^N)$$

we have

$$\alpha'(z)=\frac{pd^L\lambda\left[1+(L-1)\frac{pd+\lambda-\lambda z}{d+\lambda-\lambda z}\right]}{(pd+\lambda-\lambda z)^2(d+\lambda-\lambda z)^{L-1}}+$$
$$+\frac{-pd\lambda}{(pd+\lambda-\lambda z)^2}\left(\frac{d}{d+\lambda}\right)^{L-1}\sum_{j=0}^{N-1}\binom{L+j-2}{j}\left(\frac{\lambda}{d+\lambda}\right)^j(z^N-z^j)+$$
$$+\frac{pd}{pd+\lambda-\lambda z}\left(\frac{d}{d+\lambda}\right)^{L-1}\sum_{j=0}^{N-1}\binom{L+j-2}{j}\left(\frac{\lambda}{d+\lambda}\right)^j(N_z^{N-1}-jz^{j-1})$$

and then after some laborious calculations,

$$\alpha''(1)=\left(\frac{\lambda}{pd}\right)^2((p+2)(L-1)p+2+(L-1)^2p^2)+$$
$$+\frac{2\lambda}{pd}\left(\frac{d}{d+\lambda}\right)^{L-1}\sum_{j=0}^{N-1}\binom{L+j-2}{j}\left(\frac{\lambda}{d+\lambda}\right)^j(N-j)+$$
$$+\left(\frac{d}{d+\lambda}\right)^{L-1}\sum_{j=0}^{N-1}\binom{L+j-2}{j}\left(\frac{\lambda}{d+\lambda}\right)^j(N(N-1)-j(j-1)).$$

For the case of $N=1$ and bulk input stream, we will get the marginal functional from the joint functional (14),

$$\alpha(z)=\frac{pd^L}{(pd+\lambda-\lambda a(z))(d+\lambda-\lambda a(z))^{L-1}}-\frac{pd(1-a(z))}{pd+\lambda-\lambda a(z)}\left(\frac{d}{d+\lambda}\right)^{L-1}$$

to yield

$$\alpha'(z)=\frac{pd^L\lambda a'(z)\left[1+(L-1)\frac{pd+\lambda-\lambda a(z)}{d+\lambda-\lambda a(z)}\right]}{(pd+\lambda-\lambda a(z))^2(d+\lambda-\lambda a(z))^{L-1}}+\frac{a'(z)p^2d^2}{(pd+\lambda-\lambda a(z))^2}\left(\frac{d}{d+\lambda}\right)^{L-1}$$

and then

$$\alpha''(1) = \frac{pd\lambda\alpha''(1)[1+(L-1)p] + p\lambda^2(a)^2(L-1)(p-1) + \lambda^2 a^2[1+(L-1)p][2+(L-1)p]}{p^2 d^2} +$$
$$+ a''(1) + \left(\frac{2a^2\lambda}{pd}\right)\left(\frac{d}{d+\lambda}\right)^{L-1}.$$

Example 1. In this example we consider an optimization problem for the case of $N=1$ and $a(z)$ arbitrary. As an objective function, we consider a linear combination

$$\mathcal{O}(L) = \xi \frac{\lambda a}{\alpha}(1-\rho) + q\Pi'(1) \tag{61}$$

of the switchover and buffer load rates according to formulas (52) and (55). (We recall that the jobs rate turned out to be invariant of L and N). In this case, $\Pi'(1)$ comes from (56) and (60). For the illustration purpose we pick out the following parameters: $\lambda = 0{,}25$, $a = 5$, $\tilde{a}_2 = 7$, $d = 0{,}75$, $p = 0{,}25$, $\xi = 4000$, $\rho = 0{,}25$, $b_2 = 2\rho/\lambda$ and use them to calculate the optimal value of $\mathcal{O}(L)$ $(L \le 50)$ from (61):

```
MATLAB Coding
MaxL=50;
Lambda=.25;
d=.75;
p=.25;
xi=4000;      % Switchovers Rate Penalty
rho=.25;
b2=2*(rho/Lambda);
a=5;
b=rho/(Lambda*a);
a2=7;   % a"(1)= a2-a
JC=-5;              % Job Coefficient
q=7;                % Buffer Load Rate Penalty
for L=1:MaxL

Alpha(L)=a*(Lambda/d)*((1/p)+L-1)+a*((d/(d+Lambda))^(L-1));
AlphaDP(L)=((p*d*Lambda*(a2-a)*(1+(L-1)*p)+p*(Lambda^2)*(a^2)*(L-1)*(p-1)
+(Lambda^2)*(a^2)*(1+(L-1)*p)*(2+(L-1)*p))/((p^2)*(d^2)))+(a2-a)
+(((2*(a^2)*Lambda)/(p*d))*((d/(d+Lambda))^(L-1)));

Job(L)=JC*(Lambda*a*((1-rho)/Alpha(L))*((1/p)+L-1+(d/Lambda)*((d/(d+L))^(L-1))));

Switch(L)=xi*(Lambda*a/Alpha(L))*(1-rho);
```

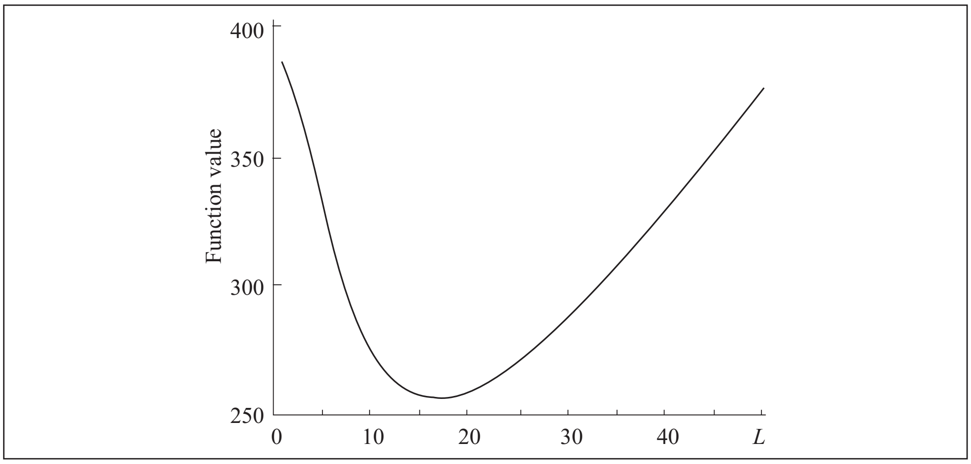

Fig. 1

```
Buffer(L)=q*((AlphaDP(L)/(2*Alpha(L)))-((a2-a)/(2*a))+rho+((Lambda/(2*(1-rho)))*(Lambda*
(a^2)*b2- a2*b+a*b)));
    Result(L)=Switch(L)+Buffer(L)-Job(L);
end
Optimal=min(Result(:))
L=find(Result==Optimal)
plot(Result)
```

The following results are due to the above MATLAB program.

The Optimal Value of $\mathcal{O}(L) = 257{,}0946$ is reached at $L = 17$ (see Fig. 1). See more details on calculation of optimal values in the technical report [9].

Example 2. In this example we consider an optimization problem for the case of L and N arbitrary and $a(z) = z$. As an objective function, we consider the same linear combination as in Example 1

$$\mathcal{O}(L) = \xi \frac{\lambda a}{\alpha}(1-\rho) + q\Pi'(1) \tag{62}$$

of the switchover and buffer load rates according to the same formulas (52) and (55). Only $\Pi'(1)$ comes from (57)—(59). For the illustration purpose we pick out the following parameters: $\lambda = 0{,}23$, $a = 1$, $d = 2$, $p = 0{,}25$, $\xi = 10000$, $\rho = 0{,}9$, $b_2 = 2\rho/\lambda$ and use them to calculate the optimal value of $\mathcal{O}(L)\,(L \le 20, N \le 20)$ from (62):

```
MATLAB Coding:
MaxL=20;
MaxN=20;
```

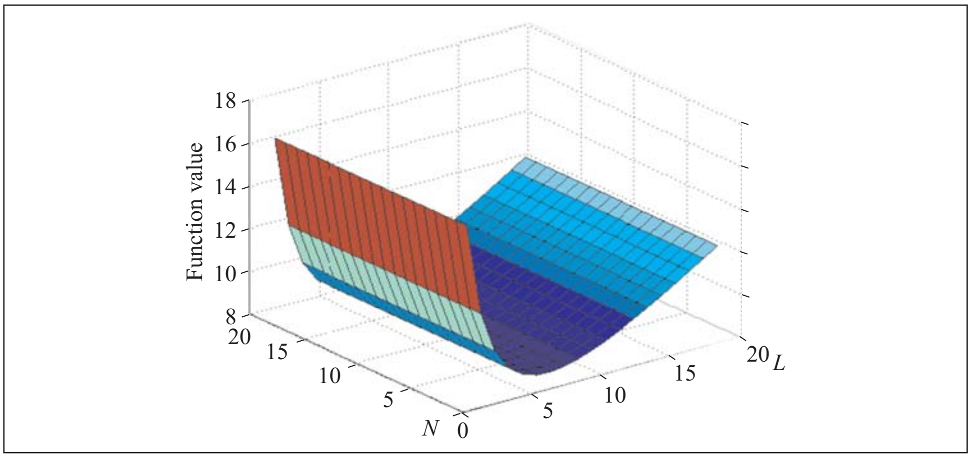

Fig. 2

```
Lam=.03;
d=2; p=.25;
xi=10000;               % Switchovers Rate Penalty
rho=.9;
b2=2*(rho/Lam);
q=1;                    % Buffer Load Rate Penalty

for L=2:MaxL
  for N=2:MaxN
    PartA=0;
    PartB=0;
    for j=0:N-1
       PartA=PartA+nchoosek(L+j-2,j)* ((Lam/(d+Lam))^j)*(N-j);
       PartB=PartB+nchoosek(L+j-2,j)* ((Lam/(d+Lam))^j)*(N*(N-1)-j*(j-1));
    end
    PartAarray(L-1,N-1)=PartA;
    PartBarray(L-1,N-1)=PartB;

    Alpha(L-1,N-1)=(Lam/d)*((1/p)+L-1)+((d/(d+Lam))^(L-1))*PartAarray(L-1,N-1);
    AlphaDP(L-1,N-1)=((Lam/(p*d))^2)*((p+2)*(L-1)*p+2+((L-1)^2)*p^2)
    +2*(Lam/(p*d))*((d/(d+Lam))^(L-1))*PartAarray(L-1,N-1)
    +((d/(d+Lam))^(L-1))*PartBarray(L-1,N-1);

    Switch(L-1,N-1)=xi*Lam*(1-rho)/Alpha(L-1,N-1);
    Buffer(L-1,N-1)=q*(AlphaDP(L-1,N-1)/(2*Alpha(L-1,N-1))+rho+((Lam^2)*b2)/(2*(1-rho)));
    Result(L-1,N-1)=(Switch(L-1,N-1)+Buffer(L-1,N-1));
```

```
  end
end
Optimal=min(Result(:));
y=find(Result==Optimal);
N=fix(y/(MaxL-1))+2
L=y-((N-2)*(MaxL-1))+1
surf(Result)
```

The following results are due to the above MATLAB program.

The Optimal Value of $\mathcal{O}(L)=$ 8,4260 is reached at $L=5$ and $N=8$ (see Fig. 2). See more details on calculation of optimal values in the technical report [8].

Наведено результати досліджень систем черг з доповняльним процесом обслуговування, який запускається в періоди простоїв сервера та ініціюється заповненою чергою. Запропоновано стратегію роботи сервера при різних станах черги. Використано часозалежний аналіз для дослідження процесів у системах черг в довільні періоди часу. Результати отримано в явній формі для декількох моделей. Наведено приклади розрахунків, які свідчать про можливість їхньої аналітичної трактовки. Введено різні критерії продуктивності (загрузка буферів, швидкість переключання та число завдань, оброблювальних за одиницею часу) і розглянуто проблеми оптимізації.

1. *Alzahrani M., Dshalalow J. H.* Fluctuation analysis of a queue with (L, N)-policy and secondary maintenance. Discrete time parameter process // Electronic modeling. — 2011. — **33**, N 4. — P. 15—34.
2. *Dshalalow J. H.* First excess level analysis of random processes in a class of stochastic servicing systems with global control// Stoch. Anal. Appl. — 1994. — **12**, N 1. — P. 75—101.
3. *Takács L.* On fluctuations problems in the theory of queues // Adv. Appl. Probab. — 1976. — **8**, N 3. — P. 548—583.
4. *Al-Matar N., Dshalalow J. H.* A game-theoretic approach in single-server queues with maintenance. Time sensitive analysis // Commun. in Appl. Nonlin. Analysis. — 2010. — **17,** N 1. — P. 65—92.
5. *Tian N., Zhang Z. G.* Vacation queueing models. — NY : Springer, 2006.
6. *Al-Matar N., Dshalalow J. H.* Time sensitive functionals in classes of queues with sequential maintenance//Stochastic models. — Accepted for publication.
7. *Çinlar E.* Introduction to stochastic processes. — Prentice-Hall, 1975.
8. *Dshalalow J. H.* On exit times of a multivariate random walk with some applications to finance// Nonlinear Analysis. — 2005. — **63**. — P. 569—577.
9. *Alzahrani M., Dshalalow J. H.* Fluctuation analysis of a queue with (L, N)-policy and secondary maintenance. Continuous time parameter process and numerics. Technical report, TR—AD2a.— 2011.—P. 1—30.

Submitted 06.01.11