

Chapter 1

Queueing Systems with Unreliable Servers in a Random Environment

Larisa Afanasyeva and Elena Bashtova

1.1 Introduction

We consider a single-server queueing system with unreliable server operating in a random environment. One would like to point out that the systems with unreliable servers have been intensively investigated for a long time now. Corresponding models are systems with interruptions of the service. This direction of research is represented by a vast collection of literature. The setting of the problems and the solutions can be found in the paper of Krishnamoorthy et al. [7].

In this paper one assumes that the breakdowns of the server are connected to a certain external factor. The external environment is a regenerative stochastic process and the breakdowns of the server occur in accordance with the points of regeneration of this process. The similar model was investigated in the pioneering paper [5]. It was assumed that a breakdown can appear only if the server is occupied by a customer. The notion of completion time, which is the generalization of the service time was introduced. This notion made it possible to apply results for a queueing system $M|G|1|\infty$ with a reliable server to investigate a system subjected to interruptions, i.e. with unreliable server.

Key elements of our analysis are the coupling of renewal processes [4] based on the structure of the random environment, construction of auxiliary processes, and relations between their characteristics and characteristics of the basic process. Note that more general models were investigated in [1]. All the statements of this paper are fulfilled for the model under consideration. But here we focus on another problems. Namely, we find the stationary distribution of the virtual waiting time process and prove the limit theorem for this distribution in heavy traffic situation.

L. Afanasyeva • E. Bashtova (✉)
Lomonosov Moscow State University, Moscow, Russia
e-mail: l.g.afanaseva@yandex.ru; bashtovaelena@rambler.ru

The paper is organized as follows. In the next section the queueing system and the random environment are described and basic relations are established. The ergodic condition is also given. In Sect. 1.3 Laplace–Stieltjese transform (LST) for the stationary distribution of the virtual waiting time is obtained and heavy traffic situation is investigated. Section 1.4 is devoted to two examples. The first one concerns the system operating in a Markov random environment. The model can be applied for analysis of systems with a preemptive priority discipline. The second example arose as a mathematical model of unregulated crossroads [2, 6]. The random environment is described by the number of customers in a queueing system with infinite number of servers.

1.2 Model Description: Basic Relations

A single-server queueing system with a Poisson input $A(t)$ with an intensity λ and an unreliable server is considered. In such a system the service of a customer is subjected to interruptions that are caused by a random environment $U(t)$ not depending on $A(t)$. It is assumed that $U(t)$ is a regenerative stochastic process and $\{u_j\}_{j=1}^{\infty}$ is a sequence of its regeneration periods (see, e.g., [4, 10]). Besides, we suppose that

$$u_j = u_j^{(1)} + u_j^{(2)}$$

where $u_j^{(1)}$ and $u_j^{(2)}$ are independent random variables and

$$\mathbf{P}(u_j^{(1)} \leq x) = 1 - e^{-\alpha x}, \quad \mathbf{P}(u_j^{(2)} \leq x) = G(x).$$

Let us introduce the sequences

$$s_n^{(2)} = \sum_{j=1}^n u_j^{(2)}, \quad s_0^{(2)} = 0, \quad s_n^{(1)} = s_{n-1}^{(2)} + u_n^{(1)}, \quad n = 1, 2, \dots$$

so that

$$0 = s_0^{(2)} < s_1^{(1)} < s_1^{(2)}, \dots$$

and

$$u_n^{(1)} = s_n^{(1)} - s_{n-1}^{(2)}, \quad u_n^{(2)} = s_n^{(2)} - s_n^{(1)}.$$

The breakdowns of the server occur at time moments $s_n^{(1)}$ and the server is repairing till the moments $s_n^{(2)}$, $n = 1, 2, \dots$. We suppose that the service was

interrupted by the breakdown of the server is continued after reconstruction from the point at which it was interrupted. Service times $\{\eta_j\}_{j=1}^{\infty}$ are independent identically distributed random variables not depending on $A(t)$ and $U(t)$.

To investigate the model we employ coupling method as it was done for more general model in [1]. Introduce the following notation

$$B(x) = \mathbf{P}(\eta_j \leq x),$$

$$b(s) = \int_0^{\infty} e^{-sx} dB(x), \quad g(s) = \int_0^{\infty} e^{-sx} dG(x),$$

$$b = \mathbf{E}\eta_1, \quad a = \frac{1}{\alpha}, \quad g_1 = \mathbf{E}u_1^{(2)}$$

and assume that $b < \infty$ and $g_1 < \infty$.

Let $W(t)$ be the virtual waiting time process and $q(t)$ be the number of customers in the system at a time t .

Theorem 1.1. *Processes $W(t)$ and $q(t)$ are ergodic if and only if traffic coefficient*

$$\rho = \lambda(1 + \alpha g_1)b < 1. \quad (1.1)$$

This result follows from Theorem 3 in [1]. The proof is based on the following representation that will be also applied later on.

We introduce two auxiliary processes $\tilde{W}(t)$ and $\tilde{W}^*(t)$ by the relations

$$\tilde{W}(t) = W\left(t + S_{N_1(t)}^{(2)}\right), \quad \tilde{W}^*(t) = W\left(t + S_{N_2(t)}^{(1)}\right) \quad (1.2)$$

where

$$S_k^{(i)} = \sum_{j=1}^k u_j^{(i)}, \quad N_i(t) = \sup\{k : S_k^{(i)} \leq t\}, \quad i = 1, 2.$$

We see that $\tilde{W}(t)(\tilde{W}^*(t))$ is obtained from $W(t)$ by the deletion of time intervals when the server is restored (is in the working state). To express $W(t)$ by means of $\tilde{W}(t)$ and $\tilde{W}^*(t)$ we define the event

$$C_t = \bigcup_{n=0}^{\infty} \{t \in [s_n^{(2)}, s_{n+1}^{(1)}]\} \quad (1.3)$$

and random variables

$$r_t = (t - S_N(t))\chi(C_t), \quad r_t^* = (t - u_{N(t)+1}^{(1)} - S_N(t))\chi(\bar{C}_t)$$

where

$$N(t) = \sup\{k : S_k \leq t\}, \quad S_k = S_k^{(1)} + S_k^{(2)}.$$

Then one can easily obtain from (1.2) and (1.3)

$$W(t) = \tilde{W}(r_t + S_{N(t)}^{(1)})\chi(C_t) + \tilde{W}^*(r_t^* + S_{N(t)}^{(2)})\chi(\bar{C}_t). \quad (1.4)$$

Note that the event C_t means that the server is in the working state at a time t . Let

$$V(t) = |\tilde{W}(r_t + S_{N(t)}^{(1)}) - \tilde{W}^*(r_t^* + S_{N(t)}^{(2)})|.$$

It follows from Lemma 1 in [1] that for any fixed $t \geq 0$

$$\lim_{y \rightarrow \infty} \mathbf{P}(\limsup_{T \rightarrow \infty} V(tT < y)) = 1. \quad (1.5)$$

This relation was employed in [1] for the proof of the ergodic theorem as well as for the asymptotic analysis of $W(t)$ and $q(t)$ in heavy traffic situation. Functional limit theorems for these processes were also established. All the statements from [1] are valid for our model but here we focus on the limit distribution

$$\Phi(x) = \lim_{t \rightarrow \infty} \mathbf{P}(W(t) \leq x).$$

In view of Theorem 1.1 this distribution exists if and only if $\rho < 1$. First we obtain the expression for

$$\varphi(s) = \int_0^{\infty} e^{-sx} d\Phi(x)$$

and then we investigate the behavior of the function $\Phi(x)$ in the heavy traffic situation ($\rho \uparrow 1$). Our proofs are based on the results for a queueing system $M|G|1|\infty$ with a reliable server (see, e.g., [8]) and relations (1.4) and (1.5).

1.3 Limit Distribution of Virtual Waiting Time

Theorem 1.2. *Let $\rho < 1$. Then the following relation takes place*

$$\varphi(s) = \tilde{\varphi}(s) \left(\frac{1}{1 + \alpha g_1} + \frac{\alpha g_1}{1 + \alpha g_1} v(s) \right) \quad (1.6)$$

where

$$\tilde{\varphi}(s) = (1 - \rho) \left(1 - (\lambda + \alpha) \frac{1 - \tilde{b}(s)}{s} \right)^{-1}, \quad (1.7)$$

$$\tilde{b}(s) = \frac{\lambda}{\lambda + \alpha} b(s) + \frac{\alpha}{\lambda + \alpha} g(\lambda(1 - b(s))), \quad (1.8)$$

$$v(s) = \frac{1 - g(\lambda(1 - b(s)))}{g_1 \lambda(1 - b(s))}. \quad (1.9)$$

Proof. Denote by

$$\tilde{\Phi}(x) = \lim_{t \rightarrow \infty} \mathbf{P}(\tilde{W}(t) \leq x), \quad \tilde{\Phi}^*(x) = \lim_{t \rightarrow \infty} \mathbf{P}(\tilde{W}^*(t) \leq x),$$

where $\tilde{W}(t)$ and $\tilde{W}^*(t)$ are defined by (1.2). Let $\varphi(s)$ and $\tilde{\varphi}(s)$ be LST of the distribution functions $\Phi(x)$ and $\tilde{\Phi}(x)$, respectively. It follows from (1.4) that

$$\mathbf{E}e^{-sW(t)} = \mathbf{E}e^{-s\tilde{W}(r_t + S_{N(t)}^{(1)})} \chi(C_t) + \mathbf{E}e^{-s\tilde{W}^*(r_t^* + S_{N(t)}^{(2)})} \chi(\bar{C}_t). \quad (1.10)$$

Now employing (1.5) and well-known limit theorems from the renewal theory (see, e.g., [4, 10]) one can take a limit (as $t \rightarrow \infty$) in (1.10). It gives the relation

$$\varphi(s) = \frac{1}{1 + \alpha g_1} (\tilde{\varphi}(s) + \alpha g_1 \tilde{\varphi}^*(s)).$$

To find $\tilde{\varphi}(s)$ we consider an auxiliary queueing system $M|G|1$ with input intensity $\tilde{\lambda} = \lambda + \alpha$ and $\tilde{b}(s)$, defined by (1.8), as the LST of the service time distribution. Since

$$-\tilde{b}'(0) = \frac{\lambda b}{\lambda + \alpha} (1 + \alpha g_1)$$

the traffic coefficient of the system $\tilde{\rho} = (\lambda + \alpha)(-\tilde{b}'(0)) = \rho < 1$. Therefore the system is ergodic and LST of the limit distribution of the virtual waiting time in this system is given by (1.7) (see, e.g., [8]).

Since the input flow $A(t)$ of the initial system is a Poisson process and $u_n^{(1)}$ has an exponential distribution with a parameter α the process $\tilde{W}(t)$ is the virtual waiting time process in the auxiliary system $M|G|1$ with input intensity $\lambda + \alpha$. Besides, one can easily verify that LST of the service time distribution is defined by (1.8).

To find $\tilde{\varphi}^*(s)$ let us denote by γ a random variable with the distribution function $g_1^{-1} \int_0^x (1 - G(y)) dy$ not depending on the sequence $\{\eta_j\}_{j=1}^\infty$ as well as on input $A(t)$. Then the LST of the distribution of the total service time of customers arriving during time interval $(0, \gamma)$ is given by the relation

$$v(s) = \mathbf{E} e^{-s \sum_{j=1}^{A(\gamma)} \eta_j} = \frac{1 - g(\lambda(1 - b(s)))}{g_1 \lambda(1 - b(s))}.$$

Since $A(t)$ is a Poisson process, then for any sequence $\{t_n\}_{n=1}^\infty$, $t_n \xrightarrow{n \rightarrow \infty} \infty$ we have

$$\lim_{n \rightarrow \infty} \mathbf{P}(\tilde{W}(t_n) \leq x) = \Phi(x).$$

In view of results from the renewal theory it means that

$$\tilde{\varphi}^*(s) = \tilde{\varphi}(s)v(s).$$

To describe the heavy traffic situation we introduce a family of queueing systems $\{S_\varepsilon\}$ with input flow $A_\varepsilon(t)$ with intensity

$$\lambda_\varepsilon = \frac{1 - \varepsilon}{b(1 + \alpha g_1)}$$

and traffic coefficient $\rho_\varepsilon = 1 - \varepsilon \rightarrow 1$ as $\varepsilon \rightarrow 0$.

For a system S_ε we mark by ε stochastic processes and functions introduced previously. So that $W_\varepsilon(t)$ is the virtual waiting time process for S_ε and

$$\Phi_\varepsilon(x) = \lim_{t \rightarrow \infty} \mathbf{P}(W_\varepsilon(t) \leq x).$$

Theorem 1.3. *If $b_2 = \mathbf{E}\eta_i^2 < \infty$, $g_2 = \mathbf{E}(u_n^{(2)})^2 < \infty$, then for any $x > 0$*

$$\lim_{\varepsilon \rightarrow 0} (1 - \Phi_\varepsilon(\varepsilon x)) = e^{-\frac{2x}{\sigma^2}}$$

where

$$\sigma^2 = \frac{b_2}{b} + \frac{\alpha g_2}{(\alpha g_1 + 1)^2}. \quad (1.11)$$

Proof. We apply relations (1.6–1.9) to obtain the result. First we note that

$$v_\varepsilon(\varepsilon s) = \frac{1 - g(\lambda_\varepsilon(1 - b(\varepsilon s)))}{g_1 \lambda_\varepsilon(1 - b(\varepsilon s))} \rightarrow 1 \quad \text{as } \varepsilon \rightarrow 0.$$

It is known (see, e.g., [3]) that for $M|G|1$ system there exists the limit

$$\tilde{\varphi}_\varepsilon(\varepsilon s) \rightarrow \frac{1}{1 + \frac{\tilde{b}_2}{2b}s}$$

where $\tilde{b}_2 = \tilde{b}''(0)$, $\tilde{b} = -\tilde{b}'(0)$. Therefore it is necessary only to find these constants from the relation (1.8).

1.4 Examples

Example 1. Here we assume that a random environment $U(t)$ is an ergodic Markov chain with the set of states $\{0, 1, 2, \dots\}$. We define the points of regenerations of $U(t)$ as the instants when $U(t)$ gets over the state zero. Then $u_n^{(1)}$ has an exponential distribution with a parameter $\alpha = 1/\mathbf{E}u_n^{(1)}$. Let $\{\pi_j\}_{j=0}^\infty$ be a stationary distribution of the process $U(t)$. Taking into account the equality

$$\pi_0 = \frac{\mathbf{E}u_n^{(1)}}{\mathbf{E}u_n^{(1)} + \mathbf{E}u_n^{(2)}} = \frac{1}{1 + \alpha g_1}$$

we see that a traffic coefficient of the system $M|G|1$ operating in a Markov random environment $U(t)$ is of the form

$$\rho = \frac{\lambda b}{\pi_0}.$$

Consider a birth and death Process $U(t)$. Let α_i be an intensity of birth, β_i be an intensity of death in the state i ($i = 0, 1, \dots$), $\beta_0 = 0$. Then $U(t)$ is ergodic Markov chain if and only if [8]

$$\sum_{j=1}^{\infty} \prod_{i=1}^j \frac{\alpha_{i-1}}{\beta_i} < \infty. \quad (1.12)$$

In this case

$$\pi_0 = \left(1 + \sum_{j=1}^{\infty} \prod_{i=1}^j \frac{\alpha_{i-1}}{\beta_i} \right)^{-1}$$

so that process $W(t)$ for a system operating in the random environment $U(t)$ is ergodic if and only if

$$\lambda b \left(1 + \sum_{j=1}^{\infty} \prod_{i=1}^j \frac{\alpha_{i-1}}{\beta_i} \right) < 1. \quad (1.13)$$

Consider the case $\alpha_i = \alpha$, $\beta_i = \beta$ so that $U(t)$ is the number of customers at a time t in a system $M|M|1|\infty$. It is well known (see, e.g., [8]) that the LST of the distribution of the busy period is of the form

$$g(s) = 2\beta(s + \alpha + \beta - \sqrt{(s + \alpha + \beta)^2 - 4\alpha\beta})^{-1}. \quad (1.14)$$

With the help of Theorems 1.2 and 1.3 one can find the LST of the stationary distribution of the virtual waiting time process for a system $M|G|1|\infty$ in a random environment $U(t)$ and analyze its asymptotic behavior in heavy traffic situation. It is evident that we consider, indeed, a system with preemptive priority discipline.

Example 2. Let a random environment $U(t)$ be the number of customers at a time t in a queueing system $M|G|\infty$ with a Poisson input with intensity α and $F(x)$ as a distribution function of service time with finite mean c . The points of regenerations of $U(t)$ are the instants when $U(t)$ gets over the state zero. Then the n th regeneration period u_n is of the form $u_n = u_n^{(1)} + u_n^{(2)}$. Here $u_n^{(1)}$ has an exponential distribution with a parameter α , $u_n^{(2)}$ represents the n th busy period. Besides, $u_n^{(1)}$ and $u_n^{(2)}$ are independent random variables. The LST of distribution of $u_n^{(2)}$ is defined with the help of the relation (see [9])

$$g(s) = \mathbb{E}e^{-su_n^{(2)}} = 1 - \frac{s\beta(s)}{\alpha(1 - \beta(s))} \quad (1.15)$$

where

$$\beta(s) = \alpha \int_0^{\infty} e^{-sx - \alpha \int_0^x \bar{F}(y) dy} \bar{F}(x) dx, \quad \bar{F}(x) = 1 - F(x).$$

One can verify that

$$g_1 = -g'(0) = \alpha^{-1}(e^{\alpha c} - 1).$$

and ergodicity condition is of the form

$$\rho = \lambda b e^{\alpha c} < 1.$$

If the distribution $F(x)$ has the second moment, we may apply (1.15) to calculate $g''(0)$ and describe the asymptotic behavior of the limit distribution of $W(t)$ with the help of Theorem 1.3. But calculations and formulas are too cumbersome to give them here. Therefore we consider $M|D|\infty$ with a constant service time c . Then

$$g(s) = \frac{s + \alpha}{se^{(s+\alpha)c} + \alpha}.$$

One can easily verify that the normalizing coefficient σ^2 from Theorem 1.3 is given by the equality

$$\sigma^2 = \frac{b_2}{b} + \frac{2}{\alpha}(e^{\alpha c} - 1 - \alpha c).$$

This model can be applied for description of the number of vehicles at the intersection roads. Some results in this direction were obtained in [2].

Acknowledgements The research was partially supported by RFBR grant 13-01-00653.

References

1. Afanasyeva, L.G., Bashtova, E.E.: Coupling method for asymptotic analysis of queues with regenerative input and unreliable server. *Queueing Syst.* **76**(2), 125–147 (2014)
2. Afanasyeva, L.G., Rudenko, I.V.: $GI|G|\infty$ queueing systems and their applications to the analysis of traffic models. *Theory Prob. Appl.* **57**(3), 1–26 (2013)
3. Borovkov, A.A.: *Stochastic Processes in Queuing Theory*. Springer, New York (1976)
4. Cox, D.R.: *Renewal Theory*. Methuen and Co, London; Wiley, New York (1962)
5. Gaver, D.: A waiting line with interrupted service, including priorities. *J. Roy. Stat. Soc.* **13**(24), (1962)
6. Gideon, R., Pyke, R.: Markov renewal modelling of Poisson traffic at intersections having separate turn lanes. *Semi-Markov Models Appl.* 285–310 (1999)
7. Krishnamoorthy, A., Pramod, P.K., Chakravarthy, S.R.: Queues with interruptions: a survey (2012). Doi:10.1007/s 11750-012-02566
8. Saaty, T.L.: *Elements of Queuing Theory with Applications*. Mc Graw Hill, New York (1961)
9. Stadjie, W.: The busy period of the queueing system $M|G|\infty$. *J. Appl. Probab.* **3**(22), 697–704 (1985)
10. Thorisson, H.: *Coupling, Stationary and Regeneration*. Springer, New York (2000)