Special Retrial Systems with Exponentially and Uniformly Distributed Service Time

Péter Kárász

Budapest Polytechnic, John von Neumann Faculty of Informatics

P.O.Box 267, H–1300 Budapest, Hungary

karasz.peter@nik.bmf.hu

Abstract: Based on a real problem connected with the landing of aeroplanes we investigate a special queueing system where special conditions prevail. In these systems, called Lakatos-type queueing systems, the service of a client may start upon arrival or at times differing from it by multiples of cycle-time T *. In this paper we investigate a system which serves two different types of customers. First-type customers can be serviced only when the system is free, wheras second-type customers may join a queue in the case of a busy server. Customers form Poisson-processes, and their service time distributions can be either exponential or uniform. Service discipline is FIFO (FCFS). We calculate the generating functions of transition probabilities and determine equilibrium distribution of the corresponding Markov-chain, and give conditions of ergodicity for each case.*

Keywords: Retrial systems, Lakatos-type queueing system

1 Introduction

Systems where customers arrive, wait for service, then leave after getting serviced often occur in real life. The wide range of possible applications explains the important role of queueing theory in the description of several everyday processes.

A new class of queueing systems was considered by Falin in [1]. This class of queues are characterized by the following feature: if a customer arrives when all potential servers are busy, it leaves the service area, but repeats its request after some random time until it gets serviced. This feature appears in many computer and commuication networks problems.

In his works Lakatos has extensively investigated such type of retrial systems, where the service of a request can be started upon arrival (in case of a free system) or at times differing from it by multiples of the cycle time T (in the case of a busy server). This happens at aeroports, where aeroplanes can start landing on arrival if the runway is free. Otherwise they have to start a circular manoeuvre, and can put another request for landing if they have reached the starting point of the trajectory. In contrast with Falin's retrial queues, here the *FIFO (FCFS)* rule is used, because of possible fuel shortage.

In [2] Lakatos considered a system with Poisson-arrivals and exponentially distributed service time; whereas in [3] service time distribution is uniform. Koba found sufficient condition of ergodicity for a more general case, which she published in [4]. As a generalisation, Lakatos examined a special system which serves customers of two different types in [5]. In the system only one customer of first type can be present, it can only be accepted for service in the case of a free server, whereas in all other cases the requests of such customers are turned down. There is no such a restriction on customers of second type; they are serviced immediately or join a queue in case of a busy server. Both types of customers form Poisson processes, i.e. interarrival times are independent, identically (exponentially) distributed. Service time distributions are exponential, and also independent. In [6] we investigated the same system, but service time distributions were uniform, and and in [7] we also included numerical results by simulation. Numerical investigation was also carried out in [8] with the same type of system, but service-time distribution is discrete (geometric). In [9] classical and Lakatos-type systems are compared.

In this paper we are going to consider the same system but the service time of customers is either exponentially or uniformly distributed. We summerize results of former papers and give results when service time distributions are of different type.

2 Description of the system

In conventional queuing systems the service process runs continuously; after having completed the service of a customer, we immediately take the next one. In retrial systems, this is not so. Therefore, to elaborate the mathematical description of the system we make the following assumptions. In the system there might be idle periods, when the service of a request is completed, but the next one has not reached its starting position. We consider these periods as part of the service time, making the service process continuous in such way. We also make a restriction on the boundaries of the intervals of the uniform distribution: they are multiples of the cycle-time. This assumption does not violate the generality of the theory, but without it formulae are much more complicated.

For the description of the system we are going to use the *embedded Markov-chain technique*. Let us consider the number of customers in the system at moments just before the service of a new customer begins. In other words, if t_k denotes the moment when the service of the k-th entity starts, we consider the sequence, whose states correspond to the number of customers at $t_k - 0$, and is denoted by ξ_k. For the sake of definiteness, at $t = 0$ let the system be free. It is not difficult to see that the sequence of random varibales $\{\xi_k, k \geq 1\}$ defined this way forms a Markov-chain. This is a consequence of three facts: service times are independent, interarrival times are independent and exponentially distributed.

For this chain we introduce the following transition probabilities:

- a_{ji} — the probability of appearance of i customers of second type at the service of a j-th type customer $(j = 1,2)$ if at the beginning there is only one customer in the system;
- b_i — the probability of appearance of i customers of second type at the service of a second-type customer, if at the beginning of service there are at least two customers in the system;
- c_i — the probability of appearance of i customers of second type after free state.

As the process runs, the busy period can start with a customer of either type. During the service of this customer only second-type customers are accepted for service, they join the queue, and requests of first-type customers are refused. This explains the need for introducing c_i, which will be determined with the help of a_{ji}, depending on the type of customer being serviced. If there are no other requests present, when the service of the next customer (which is obviously second-type) begins, the system turns into state 1, and probabilities of turning into other states from this one are given by a_{2i}. Probabilities of all other transitions are b_i.

The generating functions of these transition probabilities are respectively:

$$A_j(z) = \sum_{i=0}^{\infty} a_{ji} z^i \quad (j = 1,2), \quad B(z) = \sum_{i=0}^{\infty} b_i z^i, \quad C(z) = \sum_{i=0}^{\infty} c_i z^i.$$

3 Results

Let us consider a queuing system with two types of customers forming Poisson-processes with parameters λ_1 and λ_2. First type customers are accepted for service only if the system is free, in all other cases their requests for service are rejected. However, there is no such a restriction on customers of second type, if the server is busy they join a queue. The service of this type may start upon arrival or at moments differing from it by multiples of the cycle-time T. We define a Markov-chain, whose states correspond to the number of customers in the system just before starting a service. Service time distributions of customers of either type may be exponential with parameter μ_j or uniform in the interval $\lfloor \alpha_j, \beta_j \rfloor$ (α_j and β_j are multiples of T), thus we are going to consider four cases $(j = 1,2)$.

Theorem 1

The matrix of transition probabilities of the defined Markov-chain has the form

$$\begin{bmatrix} c_0 & c_1 & c_2 & c_3 & \dots \\ a_{20} & a_{21} & a_{22} & a_{23} & \dots \\ 0 & b_0 & b_1 & b_2 & \dots \\ 0 & 0 & b_0 & b_1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \qquad (1)$$

The elements of the matrix are determined by their generating functions below. The type of service time distribution is indicated in the upper index: $A_j^{\{\exp,\text{uni}\}}(z)$ indicates the type of service time distribution of j-th-type customers, and $B^{\{\exp,\text{uni}\}}(z)$ indicates the type of service time distribution of second-type customers.

$$A_j^{\exp}(z) = \sum_{i=0}^{\infty} a_{ji} z^i = \frac{\mu}{\lambda_2 + \mu} + \frac{\lambda_2 z}{\lambda_2 + \mu} \frac{e^{\lambda_2 (z-1)T}\left(1 - e^{-\mu T}\right)}{1 - e^{(\lambda_2(z-1)-\mu)T}}, \qquad (2)$$

$$A_j^{\text{uni}}(z) = \sum_{i=0}^{\infty} a_{ji} z^i = \frac{e^{-\lambda_2 \alpha} - e^{-\lambda_2 \beta}}{\lambda_2(\beta - \alpha)} \left(1 - z \frac{1 - e^{-\lambda_2 T}}{1 - e^{-\lambda_2 T z}}\right) + \\ + z \frac{e^{\lambda_2(z-1)\alpha} - e^{\lambda_2(z-1)\beta}}{\lambda_2(\beta - \alpha)} \left(\frac{1 - e^{-\lambda_2 T}}{1 - e^{-\lambda_2 T z}} - \frac{\lambda_2 T}{1 - e^{-\lambda_2 (z-1)T}}\right), \qquad (3)$$

$$B^{\exp}(z) = \sum_{i=0}^{\infty} b_i z^i = \frac{1 - e^{\lambda_2(z-2)T}}{\left(1 - e^{-\lambda_2 T}\right)(2 - z)} - \\ - \frac{\lambda_2}{\lambda_2(2 - z) + \mu} \cdot \frac{1 - e^{\lambda_2(z-1)T}}{1 - e^{-\lambda_2 T}} \cdot \frac{1 - e^{(\lambda_2(z-2)-\mu)T}}{1 - e^{(\lambda_2(z-1)-\mu)T}}, \qquad (4)$$

$$B^{\text{uni}}(z)=\sum_{i=0}^{\infty} b_i z^i = \frac{e^{\lambda_2(z-1)\alpha}-e^{\lambda_2(z-1)\beta}}{\left(1-e^{\lambda_2(z-1)T}\right)(\beta-\alpha)\left(1-e^{-\lambda_2 T}\right)}\times$$
$$\times\left(\frac{1-e^{\lambda_2(z-2)T}-e^{\lambda_2(z-1)T}+e^{\lambda_2(2z-3)T}}{\lambda_2(z-2)^2}+T\frac{e^{\lambda_2(z-2)T}-e^{\lambda_2(z-1)T}}{z-2}\right), \tag{5}$$

$$C(z)=\sum_{i=0}^{\infty} c_i z^i = \frac{\lambda_1}{\lambda_1+\lambda_2}A_1(z)+\frac{\lambda_2}{\lambda_1+\lambda_2}A_2(z). \tag{6}$$

Outlines of proof.

Because of earlier explanations the matrix of transition probabilities is straightforward. However, we underline that probabilities a_{1i} do not appear in it explicitly, as customers of first type can only be accepted when the system is free. These probabilities are represented through probabilities c_i.

For the discription of the system we use the embedded Markov-chain technique, i.e. we consider the number of customers in the system just before the service of a new customer begins. This actually means the number of second-type customers, as first type customers are refused when the server is busy. We find the transition probabilities of this chain.

We have to consider two cases, whether there are customers waiting or not. First we consider the case when only one customer of j-th type is present in the system. Let u denote the service time of this customer and v denote the time elapsed between the beginning of its service and the appearance of a new one. In order to be able to determine how many new requests appear in the meantime, we have to determine the distribution of the remaining time. We calculate the probability $P(0<u-v<t)$. If $F_2(u)$ denotes the distribution function of the service time of second-type customers,

$$P(0<u-v<t)=\int_0^{\infty}\int_v^{v+t}\lambda_2 e^{-\lambda_2 v}\,\mathrm{d}F_2(u)\,\mathrm{d}v.$$

Using this we are able to determine the probability of t falling between two multiples of T. The period between the entry of the second request till the beginning of its service is $\left\lceil \frac{u-v}{T}\right\rceil T$, where $\lceil x\rceil$ denotes the 'upper' integral part of x (the least integer which is not less than x). Considering that $\pi_0=\int_0^{\infty} e^{-\lambda_2 x}\,\mathrm{d}F_2(x)$ is the probability that no other requests appear during the service of the present customer, the generating functions of transition probabilities a_{ji} are:

$$A_j(z) = \pi_0 + z\sum_{k=0}^{\infty}\sum_{i=0}^{\infty} P((i-1)T < t < iT)\frac{(\lambda_2 iTz)^k}{k!}e^{-\lambda_2 iT}.$$

Expanding sums we get (2) and (3). The busy period can start with a customer of j-th type with probability $\frac{\lambda_j}{\lambda_1+\lambda_2}$, this explains (6).

Now we are going to determine the transition probabilities of all other states. In this case, at the instant when the service of the first request begins, the second one is already present, too. Let $x = u - \left\lfloor \frac{u}{T} \right\rfloor T$ and y mean the deviation of interarrival times $\bmod T$. It can easily be seen that y has truncated exponential distribution with distribution function $\frac{1-e^{-\lambda_2 y}}{1-e^{-\lambda_2 T}}$. The period between the starting moments of services of two consecutive requests is

$$t_c = \left\lfloor \frac{u}{T} \right\rfloor T + y, \text{ if } x \le y; \quad \text{and} \quad t_c = \left(\left\lfloor \frac{u}{T} \right\rfloor + 1\right)T + y, \text{ if } x > y.$$

Let us divide the service time into intervals of length T and fix y. Each such interval is divided into two parts by y (the first part has length y, the second part $T-y$). The probabilities of appearance of k requests during the investigated period are $\frac{(\lambda_2 t_c)^k}{k!}e^{-\lambda_2 t_c}$. Let $\left\lfloor \frac{u}{T} \right\rfloor = i$, and ξ be a random variable denoting the number of requests appearing during the investigated period.. The generating function of the number of requests entering the system provided that the $\bmod T$ interarrival time equals y is

$$E(z^{\xi} \mid y) = \sum_{k=0}^{\infty}\sum_{i=0}^{\infty}\left[\int_{iT}^{iT+y}\frac{[\lambda_2(iT+y)z]^k}{k!}e^{-\lambda_2(iT+y)}\,\mathrm{d}F_2(u) + \right.$$
$$\left. + \int_{iT+y}^{(i+1)T}\frac{[\lambda_2((i+1)T+y)z]^k}{k!}e^{-\lambda_2((i+1)T+y)}\,\mathrm{d}F_2(u)\right].$$

Multiplying this expression by $\frac{\lambda_2 e^{-\lambda_2 y}}{1-e^{\lambda_2 T}}$ and integrating with respect to y from 0 to T, we finally obtain the generating function (4) and (5) of transition probabilities b_i.

Theorem 2

The generating function of ergodic distribution of this chain is:

$$P(z)=\sum_{i=0}^{\infty} p_i z^i = \frac{p_0\left(zC(z)-B(z)\right)+p_1 z\left(A_2(z)-B(z)\right)}{z-B(z)}, \tag{7}$$

where p_0 and p_1 are the first two probabilities of the ergodic distribution. They are connected with the relation $p_1 = \frac{1-c_0}{a_{20}} p_0$, where p_0 is determined by the condition $P(1)=1$, and equals:

$$p_0 = \frac{1-B'(1)}{1-B'(1)+C'(1)+k\left(A_2'(1)-B'(1)\right)}, \tag{8}$$

where $k = \frac{1-c_0}{a_{20}}$ establishes connection between the two types of customers.

Outlines of proof.

The matrix of transition probabilities has the form (1). With the help of this we can determine the probabilities of ergodic distribution denoted by p_l. They satisfy the equations

$$p_l = \sum_{k=2}^{l+1} p_k b_{l-k+1} + p_0 c_l + p_1 a_{2l} \quad (l \ge 1),$$
$$p_0 = p_0 c_0 + p_1 a_{20},$$

from which we receive the following expression for the generating function

$$P(z)=\sum_{i=0}^{\infty} p_i z^i = p_0 C(z)+p_1 A_2(z)+B(z)\left(\frac{P(z)}{z}-\frac{p_0}{z}-p_1\right),$$

which yields (7).

From the second equation for p_l, $p_1 = \frac{1-c_0}{a_{20}} p_0$ To determine p_0 we use the condition $P(1)=1$. From this we get (8), where $k = \frac{1-c_0}{a_{20}}$ is a constant factor between the first two probabilities of the ergodic distribution.

Theorem 3

The condition of existence of ergodic distribution is the fulfilment of the following inequalities.

If the service time of second-type customers is exponentially distributed (regardless of the distribution of service time of first-type customers):

$$\frac{\lambda_2}{\mu_2} < e^{-\lambda_2 T} \frac{1-e^{-\mu_2 T}}{1-e^{-\lambda_2 T}}. \tag{9}$$

If the service time of second-type customers is uniformly distributed (regardless of the distribution of service time of first-type customers):

$$\frac{\lambda_2(\alpha_2+\beta_2+T)}{2} < 1. \tag{10}$$

Outlines of proof.

Calculating the derivatives of generating functions, and substituting in (8), using the McLaurin-series of $e^{\lambda_2 T}$, it can be shown that $C'(1)+k(A_2'(1)-B'(1))$ is positive. Hence, the condition of ergodicity $0 < p_0 < 1$ simplifies into $1-B'(1) > 0$, which gives (9) and (10).

The influence of idle periods — while the system is waiting for the next entity to reach its starting position to be able to start its service — becomes less and less while $T \to 0$. At this limit transition, the conditions of ergodicity (9) and (10) tend to the classical conditions $\frac{\lambda_2}{\mu_2} < 1$ and $\frac{\lambda_2(\alpha_2+\beta_2)}{2} < 1$.

References

[1] Falin, G.I., Templeton, J.G.C.: Retrial Queues. *Monographs on Statistics and Applied Probability.* **75** (1997) Chapman & Hall

[2] Lakatos, L.: On a Simple Continuous Cyclic-Waiting Problem. *Annales Univ. Sci. Budapest., Sect. Comp.* **14** (1994) 105–113

[3] Lakatos, L.: On a Cyclic-Waiting Queuing System. *Theory of Stochastic Processes* Vol. 2 **18** no. 1-2 (1996) 176–180

[4] Коба, Е.В.: О системе обслуживания с повторением заявок при обслуживании в порядке очереди. *Доповіді НАН України* № 6 (2000) 101–103

[5] Lakatos, L.: A Special Cyclic-Waiting Queueing System with Refusals. *J. Math. Sci. (New York)* **111** (2002), no. 3, 3541–3544

[6] Kárász, P.: On a Class of Retrial Systems with Uniformly Distributed Service Time. *In Proc. of Magdeburger Stochastik-Tage 2002*, 110–112, 19-22 March 2002

[7] Kárász, P., Farkas, G.: Exact Solution for a Two-Type Customers Retrial System. *Computers and Mathematics with Applications.* (submitted, to appear)

[8] Farkas, G., Kárász P.: Investigation of a Discrete Cyclic-Waiting Problem by Simulation. *Acta Acad. Paed. Agriensis, Sectio Mathematicae* **27** (2000) 57–62

[9] Mykhalevych, K.V.: A Comparison of a Classical Retrial $M/G/1$ Queueing System and a Lakatos-Type $M/G/1$ Cyclic-Waiting Time Queueing System. *Annales Univ. Sci. Budapest., Sect. Comp.* **23** (2003) (to appear)