Annales Univ. Sci. Budapest., Sect. Comp. 23 (2004) 229-238

A COMPARISON OF A CLASSICAL RETRIAL $M/G/1$ QUEUEING SYSTEM AND A LAKATOS–TYPE $M/G/1$ CYCLIC–WAITING TIME QUEUEING SYSTEM

K.V. Mykhalevych (Kiev/London, Ukraine)

Abstract. In several papers of Lakatos (see e.g. [3, 4]) a new queueing discipline was introduced and investigated, as follows: If the channel is busy or at least one customer is on the orbit then an arriving customer is admitted to the service channel with a delay of the magnitude kT, where T is a constant whereas k is minimal under the condition that the FIFO (FCFS) queueing discipline is still kept. Unlike this, the usual retrial queue is such that "first_returned from the orbit - first served" discipline holds.

The paper solves the following problem: which of the two disciplines would be preferable, should one assume that the cycle length T is small? The comparison is made on the basis of a cost function σ measuring the cost associated with the delay of a customer during time x. The paper proves if $\sigma(x)$ is a strictly increasing convex function, then the Lakatos type discipline leads to a smaller mean cost per customer than the usual retrial one, the cycle length T being small enough.

Queueing systems in which demands arriving into the system, when all the service channels are busy and there are no free waiting places, can come back for the service after a period of time, are named retrial queueing systems. Demands which come back for later service are said to be in the orbit. For the last two decades the retrial queueing systems theory has been significantly developed, see [1, 2].

The general retrial queueing system is described, for instance, in [1]. As a rule, the retrial queueing system model provides for the following service discipline: when a channel becomes free, the first demand to find the free channel is taken into service, irrespective of it arrives from the orbit or from the primary input flow.

A special type of retrial systems was considered by the Hungarian mathematician Laszlo Lakatos. This is the so-called cyclic-waiting time queueing system. For the first time, such a model of a single-channel queueing system with constant orbit time, without losses, without waiting places and with an unlimited orbit capacity was considered by Lakatos [3]. This model arose as part of an aircraft landing process, and in connection with testing of a simulation model. The Lakatos model is characterized by the demands being serviced in tum, i.e. the system has FCFS (first come, first served) service discipline. For the cyclic-waiting time system this means that when all the service channels are busy, the demand is rejected and goes into the orbit. The demands arriving after it cannot be serviced before it. Thus, the service is carried out in tum, in the order of the actual arrival times of the demands.

Some modifications, generalizations and limitations of Lakatos-type retrial queueing systems were investigated in Ukraine, for instance, see [5, 6].

Note that the sphere of applications of retrial queueing systems is wide enough: computer networks (local and global), telephone systems, computer systems, aircraft landing systems, customer service systems and so on.

Three queueing systems with Poissonian input flow are being considered in this work:

1) Q_T: an M/G/1 queueing system with service time distribution function $B_T(x) = B(x - T)$, $T \geq 0$;
2) R_T: the mentioned above M/G/1 retrial queue with constant orbit time which is equal to T, and service time distribution function $B(x)$;
3) L_T: the mentioned above M/G/1 Lakatos-type queueing system with constant orbit time which is equal to T, and service time distribution function $B(x)$.

Let λ be the parameter of input flow. Denote t_n - the arrival time of the n-th demand, $n \geq 0$; $X_n = t_n - t_{n-1} = n$-th inter-arrival time, $n \geq 1$; S_n - the service time of the n-th demand. Then

$$B(x) = P\{S_n \leq x\}, \qquad n \geq 0.$$

Denote by $\tau = \int\limits_0^\infty x dB(x)$ - the mean service time; $\beta(s) = \int\limits_0^\infty e^{-sx} dB(x)$ - the Laplace-Stieltjes transform of service time distribution function.

Let $\sigma(x)$ be a financial loss function, i.e. this is the loss which is associated with a demand during the time period $x \geq 0$. Suppose $\sigma(0) = 0$, $\sigma(x)$ is an increasing, strictly convex function for $x \geq 0$, i.e.

$$2\sigma(x) < \sigma(x - h) + \sigma(x + h)$$

for $h > 0$, $x \geq h$.

Let W_n denote the waiting time of the n-th demand (from the moment of arrival until the start of its service). Note that the busy period for retrial queues in this paper is the period of time when at least one demand is being serviced or in the orbit. Let N denote the number of demands serviced within the busy period.

Denote

$$\overline{N} = E\{N\},$$

$$\sigma_N = \sigma(W_0) + \ldots + \sigma(W_{N-1}),$$

$$\overline{\sigma}_N = E\{\sigma_N\},$$

$$\overline{\sigma} = \lim_{n\to\infty} \frac{1}{n} E\{\sigma(W_0) + \ldots + \sigma(W_{n-1})\}.$$

Note that all the quantities being introduced are marked with the system symbol when it is necessary, for instance $\overline{\sigma}_N[Q_T]$ refers to the Q_T queueing system.

From the well-known theory [9] for all the systems which are being considered

$$\text{(1)} \qquad \overline{\sigma} = \frac{\overline{\sigma}_N}{\overline{N}}, \qquad \overline{N} < \infty.$$

The aim of this paper is to prove the following theorem.

Theorem 1. *For $\alpha > 0$ let*

$$\text{(2)} \qquad \sigma(x) \leq ce^{\alpha x}, \qquad x \geq 0$$

and

$$\text{(3)} \qquad \lambda(\beta(-\alpha) - 1) < \alpha.$$

Then there exists a $T_0 > 0$, such that

$$\text{(4)} \qquad \overline{\sigma}[L_T] < \overline{\sigma}[R_T] < \infty$$

for $0 < T < T_0$.

Proof. Consider the structure of the first busy period for the Q_T system $[0, Z[Q_T]]$ and for the L_T system $[0, Z[L_T]]$ (see Fig. 1), having assumed that for both cases t_k and S_k are the same.

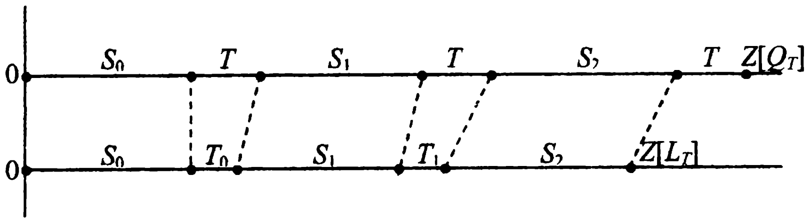

Figure 1.

Thus

$$Z[Q_T] = S_0 + T + S_1 + T + \ldots + S_{N[Q_T]-1} + T,$$
$$Z[L_T] = S_0 + T_0 + S_1 + T_1 + \ldots + S_{N[L_T]-1},$$

where T_i - the time until the return of the $i+1$-st demand from the orbit after service completion of i-th demand ($i \geq 0$). Thus

$$W_k[L_T] = \sum_{i=0}^{k}(S_i + T_i) - t_k, \quad 0 \leq k \leq N[L_T] - 1,$$
$$W_k[Q_T] = \sum_{i=0}^{k}(S_i + T) - t_k, \quad 0 \leq k \leq N[Q_T] - 1.$$

Since, it is evident that $T_1 \leq T$ and $N[L_T] \leq N[Q_T]$, then from these equalities we have $W_k[L_T] \leq W_k[Q_T]$, therefore the monotonity of $\sigma(x)$ implies

$$\sigma(W_k[L_T]) \leq \sigma(W_k[Q_T]).$$

Summation by k and taking the average leads to the following inequality

$$\overline{\sigma}_N[L_T] \leq \overline{\sigma}_N[Q_T]. \tag{5}$$

The following inequality can be obtained in a similar way

$$\overline{N}[L_T] \geq \overline{N}[Q_0]. \tag{6}$$

Thus

$$\overline{\sigma}[L_T] \leq \frac{\overline{\sigma}_N[Q_T]}{\overline{N}[Q_0]} = \frac{\overline{N}[Q_T]}{\overline{N}[Q_0]}\frac{\overline{\sigma}_N[Q_T]}{\overline{N}[Q_T]} = \frac{\overline{N}[Q_T]}{\overline{N}[Q_0]}\overline{\sigma}[Q_T]. \tag{7}$$

Having applied a well-known formula for the $M/G/1$ queueing system we get

(8) $$\overline{N}[Q_T] = \frac{1}{1-\lambda(\tau+T)}, \qquad \overline{N}[Q_0] = \frac{1}{1-\lambda\tau}.$$

Formulas (7) and (8) imply

Lemma 1. *For* $\rho = \lambda\tau < 1$

(9) $$\limsup_{T\to 0} \overline{\sigma}[L_T] \le \limsup_{T\to 0} \overline{\sigma}[Q_T].$$

We will use the following well-known lemma from the theory of Laplace-Stieltjes transforms.

Lemma 2. *Let* $(F_T(x),\ \ T \ge 0)$ *be a family of distributions on* R_+*;*

$$G_T(s) = \int_0^\infty e^{-sx} dF_T(x), \quad T \ge 0,$$

and let $\sigma(x)$ *be a monotonic function satisfying the condition*

(10) $$|\sigma(x)| \le ce^{\alpha x}, \quad x > 0.$$

If for any $A > 0$

(11) $$G_T(-\alpha + it) \to G_0(-\alpha + it), \quad T \to 0,$$

uniformly in $|t| \le A$*, then*

(12) $$\int_0^\infty \sigma(x) dF_T(x) \to \int_0^\infty \sigma(x) dF_0(x), \quad T \to 0.$$

Denote by $F_T(x)$ the distribution function of stationary virtual waiting time. Then

(13) $$\overline{\sigma}[Q_T] = \int_0^\infty \sigma(x) dF_T(x)$$

if the RHS of (13) is limited for a given $T \ge 0$.

The formula (10) implies

$$\overline{\sigma}[Q_T] \le c \int_0^\infty e^{\alpha x} dF_T(x) = cG_T(-\alpha). \tag{14}$$

From the Pollaczek-Khinchin formula

$$G_T(s) = \frac{1-\rho}{1 - \frac{\lambda}{s}(1 - \beta_T(s))}, \tag{15}$$

where

$$\beta_T(s) = \beta(s) e^{-sT}. \tag{16}$$

Condition (3) implies that the denominator in (15) is separated from zero, and moreover (11) holds. Lemma 2 implies

$$\overline{\sigma}[Q_T] \to \overline{\sigma}[Q_0], \quad T \to 0, \tag{17}$$

then from (3) and (9)

$$\limsup_{T \to 0} \overline{\sigma}[L_T] \le \overline{\sigma}[Q_0]. \tag{18}$$

Consider the first busy period in the R_T queueing system. It consists of some number $N = N[R_T]$ of service times $S_0, \dots, S_{N-1}$ and of intervals between them, each of them no longer than T. Denote by D_0, D_1, D_2 the numbers of demands arrived during the service times S_0, S_1, S_2 and define an event $\Gamma_{d_1 d_2 d_3 n}$ to be one in which

(i) $D_i = d_i, \quad i = 0, 1, 2;$

(ii) $N = n;$

(iii) no more demands arrived during the service intervals.

Thus

$$\sum_{d_0, d_1, d_2, n} P_{R_T} \{\Gamma_{d_1 d_2 d_3 n}\} \le 1, \tag{19}$$

here and further the symbol R_T means that this refers to the system R_T. Thus

$$\overline{\sigma}_N[R_T] \ge \sum_{d_0, d_1, d_2, n} E_{R_T} \{\sigma_N : \Gamma_{d_1 d_2 d_3 n}\}. \tag{20}$$

Reflect the first busy period in the R_T system to the first busy period in the Q_0^* system, which is the $M/G/1$ queueing system governed by λ, $B(x)$ and the same (not necessarily FCFS) service order as in the initial system. For the reflection we will shift left the intervals S_k and the moments t_i of arriving by excluding the intermediate intervals, see Fig.2.

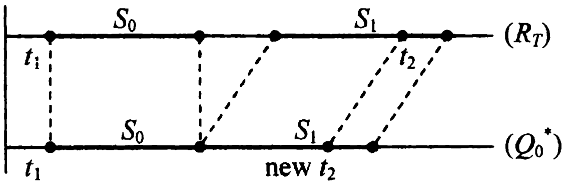

Figure 2.

Then

$$E_{R_T}\{\sigma_N : \Gamma_{d_1 d_2 d_3 n}\} \geq e^{-(n-1)\lambda T} E_{Q_0^*}, \tag{21}$$

since, firstly, the demand waiting times, together with the $\sigma(\ldots)$ values, can decrease when shifting S_k; and, secondly, the exponential multiplier is the probability of no new demands arriving in the intervals between S_k in the R_T system.

Consider the value $\sigma_N[Q_0^*]$ under the condition that an event $\Gamma_{d_1 d_2 d_3 n}$ has occured. Let the FCFS service order be broken at least once. Then the demands which arrived at the moments $t_i = x < t_j = y$ were taken for the service at the moments $u > v$, respectively. The waiting loss for these demands will be $I^* = \sigma(u - x) + \sigma(v - y)$. For another service order, the loss will be $I = \sigma(v - x) + \sigma(u - y)$. Thus

$$I^* - I = \int_v^u [\sigma'(t - x) - \sigma'(t - y)]dt > 0.$$

Therefore, the minimum loss will be for the FCFS service discipline, i.e. (21) implies

$$E_{R_T}\{\sigma_N;\ \Gamma_{d_1 d_2 d_3 n}\} \geq e^{-(n-1)\lambda T} E_{Q_0}\{\sigma_N;\ \Gamma_{d_1 d_2 d_3 n}\}. \tag{22}$$

Now consider an event Γ_{2003}, see Fig.3.

This event implies

$$\sigma_N[Q^*] = (\sigma(S_0 - t_1) + \sigma(S_0 + S_1 - t_2))I_0 + (\sigma(S_0 - t_2) + \sigma(S_0 + S_1 - t_1))I_1, \tag{23}$$

where I_0 is the indicator of the event (demand 1 was taken for service before demand 2), $I_1 = 1 - I_0$.

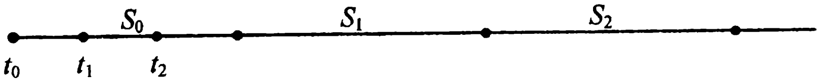

Figure 3.

Thus (22) and (23) imply

(24)
$$\overline{\sigma}_N[R_T] \geq e^{-2\lambda T} \times$$
$$\times E\{(\sigma(S_0 - t_2) + \sigma(S_0 + S_1 - t_1) - \sigma(S_0 - t_1) - \sigma(S_0 + S_1 - t_2))I_1; \Gamma_{2003}\} +$$
$$+ \sum_{d_0, d_1, d_2, n} e^{-(n-1)\lambda T} E_{Q_0}\{\sigma_N; \Gamma_{d_1 d_2 d_3 n}\}.$$

Note that

(25)
$$\sum_{d_0, d_1, d_2, n} E_{Q_0}\{\sigma_N;\ \Gamma_{d_1 d_2 d_3 n}\} = \overline{\sigma}_N[Q_0].$$

Therefore, the summation term in the RHS of (24) converges to $\overline{\sigma}_N[Q_0]$ when $T \to 0$. Now we will estimate the first term in the RHS of (24).

For fixed S_0, S_1, S_2

(26)
$$P\{\Gamma_{2003} \mid S_0, S_1, S_2\} = e^{-\lambda(S_0 + S_1 + S_2)}(\lambda S_0)^2/2.$$

The event Γ_{2003} (i.e. there are two arriving points in the interval S_0, there are no arriving points in S_1 and S_2) implies that t_1 and t_2 are distributed uniformly in the interval $(0, S_0)$ taking into account that $t_1 < t_2$.

As usual, denote by $\{a\}$ the fractional part of a. We can write that I_1 is an indicator of the event

$$\left\{\frac{S_0 - t_1}{T}\right\} < \left\{\frac{S_0 - t_2}{T}\right\}.$$

It is easy to show that

(27)
$$E\{I_1 \mid S_0\} \to \frac{1}{2}, \qquad T \to 0$$

uniformly in any interval $0 < \delta \leq S_0 < \infty$. Formulas (26), (27) imply

$$E\{\sigma(S_0 - t_2) + \sigma(S_0 + S_1 - t_1) - \sigma(S_0 - t_1) - \sigma(S_0 + S_1 - t_2)\} \geq \tag{28}$$
$$\geq \frac{\lambda^2}{4} e^{-2\lambda T} \int_0^\infty x^2 e^{-\lambda x} dB(x) \left(\int_0^\infty e^{-\lambda x} dB(x) \right)^2 + o(1),$$

when $T \to 0$. Together with (24) and (27) this implies

$$\liminf_{T \to 0} \overline{\sigma}_N[R_T] > \overline{\sigma}_N[Q_0]. \tag{29}$$

It is evident that $\overline{N}[R_T] \leq \overline{N}[Q_T]$; then (29) implies

$$\liminf_{T \to 0} \overline{\sigma}[R_T] = \liminf_{T \to 0} \frac{\overline{\sigma}_N[R_T]}{\overline{N}[R_T]} > \frac{\overline{\sigma}_N[Q_0]}{\overline{N}[Q_0]} \frac{\overline{N}[Q_0]}{\overline{N}[R_T]} \geq \overline{\sigma}[Q_0] \frac{\overline{N}[Q_0]}{\overline{N}[Q_T]},$$

or

$$\liminf_{T \to 0} \overline{\sigma}[R_T] > \overline{\sigma}[Q_0] \frac{1 - \lambda(\tau + T)}{1 - \lambda \tau}. \tag{30}$$

Then (18) and (30) imply (4).

Acknowledgements. The author considers his duty to thank Prof. I.N. Kovalenko for the supervision of this work, Mr. J.Ben Atkinson for the editing remarks, and also Prof. L. Lakatos for attention to it.

References

[1] **Yang T. and Templeton J.G.C.**, A survey on retrial queues, *Queueing Systems,* (2) (1987), 201-233.

[2] **Falin G.I.**, A survey on retrial queues, *Queueing Systems,* (7) (1990), 127-168.

[3] **Lakatos L.**, On a simple continuous cyclic-waiting problem, *Annales Univ. Sci. Budapest. Sect. Comp.*, **14** (1994), 105-113.

[4] **Lakatos L.**, A retrial system with time-limited tasks, *Theory of Stoch. Proc.*, **8(24)** (3-4) (2002), 249-255.

[5] **Коба Е.В.**, О системе обслуживания $GI/G/1$ с повторением заявок при обслуживании в порядке очереди, *Доповiдi НАН Українu,* (6)

(2000), 101-103. (Koba E.V., On $GI/G/1$ retrial queueing system with FIFO service discipline, *Dopovidi NAN Ukrainy*)

[6] **Коваленко И.Н.**, Вероятность потери в системе обслуживания $M/G/m$ с T-повторением вызовов в режиме малой нагрузки, *Доповiдi НАН України,* (5) (2002), 77-80. (Kovalenko I.N., A loss probability in $M/G/m$ queueing system with constant repeat time in a light traffic mode, *Dopovidi NAN Ukrainy*)

[7] **Фалин Г.И.**, Время ожидания в одноканальной системе обслуживания с повторением вызовов, *Вестник МГУ, Выч. мат. и кибернетика,* (4) (1977), 66-69. (Falin G.I., Waiting time in a single-channel retrial queueing system, *Vestnik Moscow University, Comp.*)

[8] **Бочаров П.П. и Печинкин А.В.**, *Теория массового обслуживания,* Изд. Рос. ун-та дружбы народов, Москва, 1995. (Bocharov P.P. and Pechinkin A.V., *Queueing theory,* Russian University of Peoples' Friendship, Moscow)

[9] **Cox D.R.**, *Renewal theory,* Methuen, London, 1962.

(Received July 24, 2003)

K.V. Mykhalevych
Dept. of Math. Reliability
V.M. Glushkov Inst. of Cybernetics
National Acad. of Sciences of Ukraine
pr. Akad. Glushkova 40
03680 Kiev-187, GSP, Ukraine

STORM Research Centre
London Metropolitan University
166-200 Holloway Road
London N78DB, UK