STABILITY ANALYSIS AND SIMULATION OF *N*-CLASS RETRIAL SYSTEM WITH CONSTANT RETRIAL RATES AND POISSON INPUTS

K. AVRACHENKOV*

Inria Sophia Antipolis 2004 Route des Lucioles Sophia Antipolis, 06902, France K.Avrachenkov@sophia.inria.fr

E. MOROZOV

Institute of Applied Mathematical Research Karelian Research Centre RAS, and Petrozavodsk State University, Russia emorozov@karelia.ru

R. NEKRASOVA

Institute of Applied Mathematical Research Karelian Research Centre RAS, and Petrozavodsk State University, Russia ruslana.nekrasova@mail.ru

B. STEYAERT

SMACS Research Group, Ghent University St-Pietersnieuwstraat 41, B-9000 Gent, Belgium bart.steyaert@ugent.be

> Received 30 September 2012 Revised 7 March 2013 Accepted 5 April 2013 Published 6 May 2014

In this paper, we study a new retrial queueing system with N classes of customers, where a class-*i* blocked customer joins orbit *i*. Orbit *i* works like a single-server queueing system with (exponential) constant retrial time (with rate $\mu_0^{(i)}$) regardless of the orbit size. Such a system is motivated by multiple telecommunication applications, for instance wireless multi-access systems, and transmission control protocols. First, we present a review of some corresponding recent results related to a single-orbit retrial system. Then, using a regenerative approach, we deduce a set of necessary stability conditions for such a system. We will show that these conditions have a very clear probabilistic interpretation. We also performed a number of simulations to show that the obtained conditions delimit the stability domain with a remarkable accuracy, being in fact the (necessary and sufficient)

*Corresponding author

stability criteria, at the very least for the 2-orbit M/M/1/1-type and M/Pareto/1/1-type retrial systems that we focus on.

Keywords: Retrial system; constant retrial rate; stability condition; regenerative approach; busy probability; multi-class system.

1. Introduction

In this paper, we first consider the single-class GI/G/m/K-type retrial queueing system with a constant retrial rate μ_0 . The external (primary) arrivals follow a renewal input with arrival epochs $\{t_n\}$ and rate λ . The system has m identical servers, and customers have i.i.d. service times $\{S_n\}$, with a generic element Sand rate $\mu := 1/ES$. If a customer finds all m servers busy and the buffer (of size $K - m < \infty$) full, it joins an infinite-capacity orbit, where only one orbit customer attempts to rejoin the primary queue after exponentially distributed time intervals with rate μ_0 (if the orbit is not empty). Such a model is referred to as a retrial model with constant retrial rate. It then follows that the orbit can be interpreted as a single-server $\cdot/M/1$ -type queue with service rate μ_0 , and with the jobs rejected from the primary queue as input. We stress that the merged stream to the orbit is in general not a GI-type arrival stream, since it is a combination of the rejected part of the primary customers and the secondary customers returning to the orbit after unsuccessful attempts to enter the primary queue. Note that the only possible source of instability of such system is an infinite growth of the orbit size.

Second, we consider a multi-class extension of the single-class model where each class has its own orbit. As is typically the case, the stability analysis of a multi-class system is much more challenging than that of the single-class variant. In the present work, we conjecture a necessary and sufficient stability condition for the multi-class multi-orbit retrial system with one server and constant retrial rates. We prove that the condition is necessary, and we provide strong indications by simulation that the condition is also sufficient.

Single-class retrial systems with constant retrial rate have been investigated in a large number of contributions. Not pretending to be exhaustive, let us mention some relevant articles. Fayolle (1986) introduced a retrial system with constant retrial rate, and derived stability conditions for the case of an M/G/1/1primary queue. Artalejo (1996) has obtained stability conditions for the Markovian M/M/2/2 case. Ramalhoto and Gómez-Corral (1998) have deduced stability conditions for the M/M/1/2 case. For the general Markovian M/M/c/K case, Ramalhoto and Gómez-Corral (1998) have obtained decomposition results assuming ergodicity (stability). The ergodicity conditions for the multiserver Markovian M/M/c/c case with recovery probability have been derived by Artalejo *et al.* (2001). Finally, a sufficient stability condition of the general single-class retrial system with constant retrial rate described above is obtained by Avrachenkov and Morozov (2010), which also turns out to be a necessary one for Markovian systems.

Retrial systems with constant retrial rate can be adopted to model a range of telecommunication systems, such as a telephone exchange system (Fayolle, 1986),

multiple access systems (Choi *et al.*, 1992, 1993), short TCP transfers (Avrachenkov and Yechiali, 2008, 2010), as well as logistic systems (Lillo, 1996). Multi-class multiorbit retrial systems are natural extensions of these applications for customers with different quality of service requirements.

The paper is organized as follows: in Sec. 2 we give a review of some related previous results (obtained mainly by the authors) concerning the analysis of a oneclass (one-orbit) system. In Sec. 3, we develop the stability analysis of a multi-orbit system, and deduce necessary stability conditions for such a system. Finally, in Sec. 4, we present simulation results which demonstrate a remarkable consistency with the theoretical results, and show that the observed conditions are very likely to be necessary and sufficient stability criteria (at least for the considered systems).

2. Preliminary Results

First, we introduce a more precise notion of stability, by means of the regenerative property of the system. For the GI/G/m/K-type retrial system mentioned above, we consider the (right-continuous) process $X = \{X(t) := N(t) + \nu(t), t \ge 0\}$, where at instant t, N(t) is the number of customers in the orbit (the orbit size), and $\nu(t)$ represents the total number of customers in the servers/buffer. Of course, $\nu(t) \le K$, while the orbit size N(t) can be unlimited. Denote by $X_n = X(t_n^-)$ the total number of customers just before the *n*th primary arrival, $n \ge 1$. Then regeneration epochs $\{T_n\}$ of the processes $\{X(t)\}, \{X_k\}$ (and other processes describing the system) occur when a primary arrival meets an empty system. Therefore, regeneration epochs can be defined recursively by the following standard way:

$$T_{n+1} = \inf_k (t_k > T_n : X_k = 0), \quad n \ge 0 \text{ (inf } \emptyset = \infty).$$

$$\tag{1}$$

Let T be a generic regeneration period under a zero initial state, i.e., $T = \inf(t_k > 0 : X_k = 0 | X_1 = 0)$ provided $t_1 = 0$. The process X is called *positive recurrent* if (Morozov, 2004; Sigman, 1990)

$$\mathsf{E}T < \infty.$$
 (2)

If (2) holds, then under mild regularity assumptions (say, if the input is Poisson) the stationary distribution of $\{X(t)\}$ (and other related processes) exists as $t \to \infty$ (Asmussen, 2003).

Let us now consider the basic bufferless M/M/1/1-type retrial system (denoted as Σ). In this case the pair $Y(t) := (N(t), \nu(t))$ is an irreducible Markov process with state space $\mathbf{S} = \mathbb{Z}_+ \times \{0, 1\}$. We denote by

$$\mathsf{P}_{ij} = \lim_{t \to \infty} \mathsf{P}\{N(t) = i; \nu(t) = j\}, \quad (i, j) \in \mathbf{S},\tag{3}$$

its stationary distribution, when it exists. The total number of unsuccessful attempts to enter the server is a superposition of the blocked (primary) λ -customers and the unsuccessful attempts of the secondary (orbit) customers. Denote by $\mathsf{P}_{\text{busy}} = \lim_{t\to\infty} \mathsf{P}(\nu(t) = 1)$ the stationary busy probability of the server. It follows from the explicit form of the distribution (3) by Artalejo *et al.* (2001) that the stability criterion of this Markovian system is given by

$$(\lambda + \mu_0)\mathsf{P}_{\text{busy}} < \mu_0,\tag{4}$$

where $\mathsf{P}_{\mathrm{busy}} := \sum_i \mathsf{P}_{i1}$. To give an intuitive explanation of this result, we consider a more general M/G/1/1-type retrial system (again denoted as Σ), and also an associated auxiliary system (denoted as $\hat{\Sigma}$) which consists of a first M/G/1/1 system where the input is a superposition of the same λ -input and a Poisson input with rate μ_0 , and a second $\cdot/M/1$ system (the *virtual orbit*) whose input is the stream of jobs blocked at the first system. Because the server in the first system of $\hat{\Sigma}$ is more loaded, one expects that the rate of blocked jobs in $\hat{\Sigma}$ is larger than the rate of unsuccessful attempts in the original system Σ . Then it is easy to see that the following condition must imply the stability of the virtual orbit, and hence, as shown by Avrachenkov and Morozov (2010), the stability of the real orbit in the original system Σ

$$(\lambda + \mu_0)\mathsf{P}_{\rm loss} < \mu_0,\tag{5}$$

where P_{loss} is the stationary loss probability in the first system of $\hat{\Sigma}$. It is proved by Avrachenkov and Morozov (2010) that even a general GI/G/m/K-type retrial system is indeed stable under condition (5). Moreover, condition (5) is in particular also the stability criterion for a system with Poisson λ -input. On the other hand, it is proved by Morozov and Nekrasova (2012) for a bufferless M/G/1/1-type retrial system (also see Theorem 1 below) that

$$\mathsf{P}_{\mathrm{busy}} = \lambda \mathsf{E}S := \rho. \tag{6}$$

(Note that by the PASTA property, P_{busy} is also the stationary blocking probability P_{orb} of primary customers.) Finally, using Erlang's formula for the first loss system of $\hat{\Sigma}$, we have

$$\mathsf{P}_{\rm loss} = \frac{(\lambda + \mu_0)/\mu}{1 + (\lambda + \mu_0)/\mu} = \frac{\lambda + \mu_0}{\mu + \lambda + \mu_0}.$$
(7)

A bit surprising but easy to check is that for the M/G/1/1-type retrial system, conditions (4) and (5) are in fact equivalent and can be written in the form (setting $\lambda = 1$ without loss of generality)

$$\frac{1}{\mu_0} + 1 < \mu,$$
 (8)

which is easily adopted to numerical investigation of the stability region, by varying the two parameters μ_0 and μ (see Fig. 1). Again, using Erlang's formula and stability condition (5), it is shown by Morozov and Nekrasova (2012) that the M/G/2/2-type retrial system (with rate $\lambda = 1$) is stable if $\mu > 1/2$ and

$$\mu_0 > \frac{\mu\sqrt{\mu^2 + 2\mu - 1} + 1 - \mu^2 - \mu}{2\mu - 1}.$$
(9)

Figure 2 illustrates the stability/instability region for such a 2-server system.

1440002-4



Fig. 2. Stability/instability region for the M/M/2/2-type retrial system with $\lambda = 1$.

Condition (5), in turn, can be rewritten as

$$\rho + \rho \frac{\lambda}{\mu_0} < 1. \tag{10}$$

One can check using (3) that in the M/M/1/1-type retrial system the term

$$\rho \frac{\lambda}{\mu_0} = \sum_{i>0} \mathsf{P}_{i0} > 0$$

is the stationary probability that the server is free while the orbit is not (we believe that this term in (10) has the same sense also in the M/G/1/1-type retrial system). It means that service discipline in this system is not *work-conserving*.

1440002-5

Note that as $\mu_0 \to \infty$, condition (10) approaches the stability criteria $\rho < 1$ of the classical $M/G/1/\infty$ system. If $\mu_0 \to 0$, then $\rho(1 + \lambda/\mu_0) \to \infty$, and thus the system becomes unstable for any $\rho > 0$. Both these results are intuitively clear.

As it has been mentioned, retrial systems with renewal input have a regenerative property. Again, consider the general GI/G/m/K-type retrial system, and let $\{z_n\}_{n\geq 0}$ be the instants when all attempts (successful or not) happen (note that the external renewal input epochs $\{t_n\} \subseteq \{z_n\}$). Recall that the process X describes the total number of customers in the system (in servers, buffer and orbit), and denote $X(z_n^-) := \hat{X}_n, n \geq 1$. It is evident that the process $\hat{X} := \{\hat{X}_n\}$ regenerates at the instants when a λ -customer meets an empty system. That is, the regeneration instants $\{\beta_n\}$ are defined by the standard recursion

$$\beta_{n+1} = \inf\{k > \beta_n : \hat{X}_k = 0\}, \quad n \ge 0 \ (\beta_0 := 0).$$
(11)

These regenerations (as well as regenerations (1) in continuous time) allow us to develop both the stability and performance analysis of the retrial system. Thus, if the stability condition (5) holds, then the regenerative process \hat{X} is positive recurrent; i.e., the mean cycle length $\mathsf{E}\beta < \infty$.

When the opposite condition $(\lambda + \mu_0)\mathsf{P}_{\text{loss}} > \mu_0$ is satisfied, the orbit increases with no limit as time increases. Thus, after an (finite w.p.1) instant t_0 , the orbit is always nonempty and the stream of attempts going from orbit to server is coupled with the Poisson input with rate μ_0 . In this case, the process \hat{X} is not positive recurrent (in fact, it is *strongly unstable*), and it is impossible to use the regenerative methodology. Hopefully, in this case we can use the so-called *quasi-regenerations* to evaluate/estimate the blocking probability P_{orb} . Denote $\nu_k = \nu(t_k^-)$. The quasiregenerations are defined as instants when an arriving λ -customer meets an empty buffer and server(s), while the state of orbit is arbitrary. More precisely, if $\alpha_0 := 0$, then the quasi-regeneration instants are defined as follows:

$$\alpha_{n+1} = \inf(k > \alpha_n : \nu_k = 0), \quad n \ge 0.$$
(12)

Applying such quasi-regenerations for a reliable estimation of $\mathsf{P}_{\rm orb}$ in the nonstationary system Σ has been discussed, by Avrachenkov and Morozov (2010) and Avrachenkov *et al.* (2011). It was shown that the distribution of the state of server in the original M/G/1/1-type retrial system approaches that in the associated auxiliary system $\hat{\Sigma}$ (with two independent Poisson inputs with rates λ and μ_0). In other words, starting from instant t_0 , quasi-regenerations become classical regenerations of the process { $\nu(t), t \geq 0$ }, solely describing the state of the server. As shown by Avrachenkov *et al.* (2011) and Morozov and Nekrasova (2011), quasi-regenerative simulation demonstrates a remarkable accuracy of the estimation of blocking probabilities in M/M/1/1-type and M/Pareto/1/1-type retrial systems. We stress that in the nonstationary regime, $\mathsf{P}_{\rm orb} = \mathsf{P}_{\rm loss}$, and hence we can compare simulation results with the explicit value of $\mathsf{P}_{\rm orb}$ satisfying (7).

We now deduce an expression for the stationary busy probability P_{busy} in the multiserver retrial system in an explicit form. Since the servers are equivalent, we

use a uniform distribution to assign a server (among the free ones, if any) for an arriving customer (primary or retrial). Denote by $B_i(t)$ the busy time of server *i* in interval (0, t], then from the previous assumptions it follows that the processes $B_i(t)$ are stochastically equivalent, $i = 1, \ldots, m$. Thus, if the limit exists, we can write

$$\mathsf{P}_{\text{busy}} = \lim_{t \to \infty} \frac{B_i(t)}{t}, \quad i = 1, \dots, m.$$
(13)

Note that in all statements below, stability means positive recurrence of the process X and other related regenerative processes in the system (i.e., positive recurrence of the system, for short). Because the (short) proof of the following statement obtained by Morozov and Nekrasova (2012) is very instructive, we reproduce it here.

Theorem 1. Assume that the GI/G/m/K-type retrial system is stable. Then the following relation holds:

$$\rho := \lambda \mathsf{E}S = m\mathsf{P}_{\mathrm{busy}},\tag{14}$$

where P_{busy} is given in (13).

Proof. Denote by V(t) the total workload that has arrived in the system in interval (0, t], and let W(t) be the remaining workload (in orbit and in servers/buffer) at instant t. The busy time of all servers, in interval (0, t] is $B(t) = \sum_{i=1}^{m} B_i(t)$. Then we obtain the following balance equation:

$$V(t) = W(t) + B(t) = W(t) + \sum_{i=1}^{m} B_i(t), \quad t \ge 0.$$
 (15)

Let A(t) be the number of arrivals in interval (0, t], and note that

$$V(t) = \sum_{k=1}^{A(t)} S_k,$$
(16)

where S_k is the service time of customer k. By the strong law of large numbers, we have $V(t)/t \to \lambda \mathsf{E}S := \rho$. By positive recurrence, $W(t)/t \to 0$ as $t \to \infty$ w.p.1 (or W(t) = o(t)) (Smith, 1955), and also the following limit $\lim_{t\to\infty} B_i(t)/t$ exists and is equal to the stationary busy probability of (arbitrary) server $i = 1, \ldots, m$. That is, formula (13) holds. It then immediately follows from (15) that $\rho = m\mathsf{P}_{\text{busy}}$, and the proof is completed.

It is interesting to note that the same result (14) holds in stationary regime for the classical *m*-server system with no losses. However, its stability condition $\rho < m$ differs from the necessary stability condition $\rho/m < 1$ for the retrial system, which comes from the requirement $\mathsf{P}_{\text{busy}} < 1$. We stress that our result is independent of the retrial rate μ_0 . (But of course, the orbit service rate μ_0 plays an important role to guarantee positive recurrence.)

3. Multi-Orbit System

This section is devoted to a multi-class multi-orbit retrial system with constant retrial rates. Extending the analysis of the previous section, we consider now the *m*-server retrial system with *N* classes of customers and *N* orbits. In general, class*i* customers follow a renewal arrival process with rate λ_i , have i.i.d. service times $\{S_n^{(i)}, n \ge 1\}$ with rate $\mu_i = 1/\text{E}S_i$ and, provided the buffer is full upon arrival, join an *i*-type orbit, $i = 1, \ldots, N$. Thus, there are *N* different orbits with retrial customers, and the *i*-orbit behaves like the orbit of the single-class system discussed above, with service rate $\mu_0^{(i)}$. Let $\rho_i = \lambda_i/\mu_i$, $i = 1, \ldots, N$. In order to apply the previous analysis, we assume that the model has regeneration instants (1) and (11). In general, there is a difficulty to construct regenerations for a superposition of renewal processes in continuous time, therefore we will assume that it is nonetheless possible in our setting, and again denote by *T* a typical regenerations (Asmussen, 2003). For instance, note that this is a valid assumption in case of Poisson inputs.

To establish a necessary stability condition for the multi-class system, we first need the following result.

Theorem 2. If the N-orbit retrial system is stable (positive recurrent) then

$$\sum_{i=1}^{N} \rho_i = m \mathsf{P}_{\text{busy}},\tag{17}$$

where P_{busy} is given by (13).

Proof. Denote by $A_i(t)$ the number of *i*-class customers that arrived in the interval (0, t], i = 1, ..., N. Then, adopting a similar notation as before, we find that

$$V(t) = \sum_{i=1}^{N} \sum_{n=1}^{A_i(t)} S_n^{(i)},$$

and it follows that w.p.1

$$\lim_{t \to \infty} \frac{V(t)}{t} = \sum_{i=1}^{N} \rho_i.$$
(18)

Then the balance equation (15) implies (17).

Thus, the basic necessary stability condition of this system is

$$\sum_{i=1}^{N} \rho_i < m, \tag{19}$$

which is the stability criterion for a classical infinite buffer system with N (Poisson) inputs and m servers.

1440002-8

We note that for Poisson arrivals (invoking the PASTA property), the probability P_{busy} is also the stationary blocking probability for arriving primary customers.

We start from the observation that stability condition (4) for a single-class M/G/1/1-type retrial system can be written as

$$\lambda \mathsf{P}_{\mathrm{busy}} < \mu_0 (1 - \mathsf{P}_{\mathrm{busy}}), \tag{20}$$

which has a nice intuitive interpretation. Indeed, in a stable regime, the left-hand side is the rate of the (primary) blocked customers, while the right-hand side is the rate of the orbit customers which successfully enter the server.

As the following Theorem 3 shows, this result is generalized (as necessary stability conditions) to a retrial system with N orbits and N-classes of primary customers, where the *i*th class customers are generated by a Poisson arrival stream with rate λ_i and have retrial rate $\mu_0^{(i)}$, i = 1, ..., N, in the orbit.

Theorem 3. Assume that the N-class M/G/1/1-type retrial system with N orbits is positive recurrent. Then the following conditions hold:

$$\lambda_i \sum_{i=1}^{N} \rho_i < \mu_0^{(i)} \left(1 - \sum_{i=1}^{N} \rho_i \right), \quad i = 1, \dots, N.$$
(21)

Proof. Recall the notation B(t) for the server busy period introduced before, and note that, in view of the Poisson inputs, the number of the blocked *i*-class customers arriving to orbit *i* in interval (0,t] is stochastically equivalent to $A_i(B(t)), i =$ $1, \ldots, N$. More precisely, for each *t*, we couple together all busy (sub)periods of type *i* (in interval (0,t]) to obtain a busy time "period" of total length B(t). Then the Poisson process which is formed by all (blocked) *i*-class customers arriving during period (0, B(t)] is equivalent to $A_i(B(t)), i = 1, \ldots, N$. Denote by $D_i(t)$ the number of renewals in the (Poisson) process with rate $\mu_0^{(i)}$ in interval (0, t], $i = 1, \ldots, N$. Note that the actual (successful) departures from orbit *i* happen only when the server is empty and orbit *i* is not empty. Hence, the number of departures from orbit *i* (in interval (0, t]) is equivalent (again by the PASTA property) to $D_i(t - B(t) - I_i(t))$, where $I_i(t)$ is the total empty time of orbit *i* within the empty periods of the server in interval $(0, t], i = 1, \ldots, N$. Moreover, we have the following balance equations:

$$A_i(B(t)) =_{st} N_i(t) + D_i(t - B(t) - I_i(t)),$$
(22)

where $=_{st}$ represents stochastic equivalence. Recall that $N_i(t)$ is the size of orbit i at instant t, i = 1, ..., N, then from the positive recurrence property (i.e., stability), $N_i(t) = o(t)$ as $t \to \infty$ w.p.1 for each i (Smith, 1955). Moreover, by the same reason, it follows from regenerative theory that for each cumulative process $\{I_i(t), t \ge 0\}$, the following limit exists w.p.1:

$$\lim_{t \to \infty} \frac{I_i(t)}{t} = \frac{\mathsf{E}I_i}{\mathsf{E}T} := \gamma_i,\tag{23}$$

where I_i is a typical empty period of orbit *i* within the idle period separating two adjacent regeneration cycles. We now show that $\min_i \gamma_i > 0$. Denote by S_j a typical service time of customer *j* (with a given distribution function F_j), and select a constant $d_j < \infty$ in such a way that

$$F_j(d_j) \equiv \mathsf{P}(S_j \le d_j) \ge \sigma_j,$$

for arbitrary fixed $\sigma_j > 0, j = 1, ..., N$. Note that this is possible since $\max_j \mathsf{E}S_j < \infty$. Denote further $\Lambda = \sum_{i=1}^N \lambda_i$ as the rate of the merged input, and let τ be a typical interarrival time of this input. Assume that a regeneration period is started by a *j*-class customer which of course enters the server. Note that this happens with probability

$$c_j = \frac{\lambda_j}{\Lambda}, \quad j = 1, \dots, N.$$

Given this event, the probability that the next customer again meets an empty system (i.e., starts a new regeneration period), provided that the whole system has been empty (before his arrival) during a time period not less than some $\delta > 0$, satisfies the inequality

$$\mathsf{P}(\tau - S_j \ge \delta) \ge \int_0^\infty e^{-(x+\delta)\Lambda} dF_j(x)$$
$$\ge e^{-(d_j+\delta)\Lambda} \sigma_j := r_j > 0.$$
(24)

It is easy to see that for any i,

$$\mathsf{E}I_i \ge \delta \mathsf{P}(I_i \ge \delta) \ge \delta \sum_i c_i r_i := r > 0,$$
(25)

and hence, $\gamma_i \ge r/\mathsf{E}T > 0$. Now we apply the strong law of large numbers to obtain w.p.1, as $t \to \infty$,

$$\frac{A_i(B(t))}{t} = \frac{A_i(B(t))}{B(t)} \frac{B(t)}{t} \to \lambda_i \mathsf{P}_{\text{busy}},$$

$$\frac{D_i(t - B(t))}{t} = \frac{D_i(t - B(t))}{t - B(t)} \frac{(t - B(t))}{t} \to (1 - \mathsf{P}_{\text{busy}})\mu_0^{(i)}, \quad i = 1, \dots, N.$$
(26)

Moreover,

$$D_i(t - B(t) - I_i(t)) =_{st} D_i(t - B(t)) - D_i(I_i(t)).$$

Since $I_i(t) \to \infty$, we then have

$$\lim_{t \to \infty} \frac{D_i(I_i(t))}{t} = \lim_{t \to \infty} \frac{D_i(I_i(t))}{I_i(t)} \frac{I_i(t)}{t} \ge \mu_0^{(i)} \gamma_i > 0, \quad i = 1, \dots, N.$$

It therefore immediately follows from (19) and (22) that (21) indeed holds.

Remark 1. Note that the basic stability condition (19) with m = 1 follows immediately from the more tight condition (21).

1440002-10

Remark 2. We have stated above a few times a commonly recognized result, being that under positive recurrence (stability) conditions, the inequality $\mathsf{P}_{\text{busy}} < 1$ holds. Indeed, using the previous proof, one can give strict regenerative arguments to verify this statement. Namely, the mean empty time of the system (before a new regeneration cycle starts) is lower bounded by a positive constant r see Eq. (25). Denote by I(t) an empty time of the system in the interval (0, t]. Then, by positive recurrence and the property of Poisson inputs, the following limits exist:

$$\lim_{t \to \infty} \frac{I(t)}{t} = \frac{\mathsf{E}I}{\mathsf{E}T} := \mathsf{P}_0 = \lim_{t \to \infty} \mathsf{P}(X(t) = 0),$$

where I is a typical empty period between two regeneration cycles. On the other hand, it is easy to see that $r \leq \mathsf{E}I$. This implies that $\mathsf{P}_0 > 0$ and hence, $\mathsf{P}_{\mathrm{busy}} < 1$.

As the simulation results in Sec. 4 will show, the stability conditions (21) are, in fact, the (necessary and sufficient) stability criteria, at least for the considered 2-class retrial systems. To support these observations, we note that $B(t) \leq V(t)$ and $V(t)/t \rightarrow \sum_i \rho_i$ regardless of the regime of the system. In particular, this holds if the system is not positive recurrent, implying $\mathbf{E}T = \infty$. (Note that in this case we cannot claim that $W(t) = o(t), N_i(t) = o(t), t \rightarrow \infty$, and the previous analysis based on balance equations such as (15) becomes useless.) Nevertheless, we may deduce from $A_i(B(t)) \leq_{st} A_i(V(t))$ that the following upper bound for the input rate to orbit *i* holds:

$$\limsup_{t \to \infty} \frac{A_i(t)}{t} \le \lim_{t \to \infty} \frac{A_i(V(t))}{V(t)} \frac{V(t)}{t} = \lambda_i \sum \rho_i, \quad i = 1, \dots, N.$$
(27)

On the other hand, the maximal long-term output from orbit i, that is, the potential rate of the (successful) attempts of orbit-i customers from a *permanently nonempty* orbit is lower bounded by

$$\liminf_{t \to \infty} \frac{D_i(t - B(t))}{t}$$

$$= \lim_{t \to \infty} \frac{D_i(t)}{t} - \limsup_{t \to \infty} \frac{D_i(B(t))}{t}$$

$$= \mu_0^{(i)} - \limsup_{t \to \infty} \frac{D_i(B(t))}{B(t)} \frac{B(t)}{t}$$

$$= \mu_0^{(i)} - \mu_0^{(i)} \limsup_{t \to \infty} \frac{B(t)}{t} \ge \mu_0^{(i)} \left(1 - \sum_i \rho_i\right), \quad i = 1, \dots, N. \quad (28)$$

To obtain the last equality, we use the convergence w.p.1 $B(t) \to \infty$. Moreover, to obtain the last inequality in (28), we rely on the following observation. While in the stable regime, we may write (see, for instance (15)),

$$\lim_{t \to \infty} \frac{V(t)}{t} = \lim_{t \to \infty} \frac{B(t)}{t} = \sum_{i} \rho_i,$$
(29)

while in the nonstable regime, we obtain the inequality

$$\limsup_{t \to \infty} \frac{B(t)}{t} \le \sum_{i} \rho_i,\tag{30}$$

because the orbit customers accumulating into an infinitely increasing orbit size actually cannot reach the server in a finite time, and in fact "disappear" (lost). Thus, under conditions (21), as follows from (27)–(30), the long-term input rate to (each) orbit is less than the potential long-term output rate from the ("saturated") orbit. This exactly corresponds to what happens in a classical system under the typical stability (negative drift) condition. It indicates that conditions (21) must also be sufficient stability conditions. Although we may deduce from this discussion that each orbit does not increase unlimitedly (in a sense), however, increasing asynchronous oscillations of different orbits are still possible, in principle. This in turn, does not allow one to uniformly upper bound all orbits sizes and, as a result, to establish stability.

4. Simulation Results

In this section, we present some numerical results which show that conditions (21) are not only necessary, but in fact are stability criteria, at least for the considered 2-class (2-orbit) M/G/1/1-type retrial systems. This is verified for exponential and Pareto service time distributions. For such a stationary system, using previous notations, we have $P_{\text{busy}} = \rho_1 + \rho_2$ and, by Theorem 3, necessary stability conditions are

$$\lambda_1 \mathsf{P}_{\text{busy}} < (1 - \mathsf{P}_{\text{busy}}) \mu_0^{(1)},\tag{31}$$

$$\lambda_2 \mathsf{P}_{\text{busy}} < (1 - \mathsf{P}_{\text{busy}}) \mu_0^{(2)}. \tag{32}$$

We consider a particular case, $\mu_1 = \mu_2 := \mu$. The stability/instability region of such an M/M/1/1-type system with parameters $\lambda_1 = 0.5$ and $\lambda_2 = 4$ is plotted in Fig. 3. The values of parameters $(\mu, \mu_0^{(1)}, \mu_0^{(2)})$, when condition (31) holds, while condition (32) is violated, is shown by the inclined hatching zone on Fig. 3 (i.e., the intermediate region between the two full curves). To obtain the opposite case, one can take, for instance, $\mu = 15, \mu_0^{(1)} = 0.01$ and $\mu_0^{(2)} = 4$. Define the *stability measure* of orbit *i* as [see (32)]

$$\Gamma^{(i)} := 1 - \mathsf{P}_{\text{busy}}\left(\frac{\lambda_i}{\mu_0^{(i)}} + 1\right), \quad i = 1, 2.$$
(33)

By analogy we can define the corresponding *instability measure* $\hat{\Gamma}^{(i)} := -\Gamma^{(i)}, i = 1, 2.$

As simulation shows, these measures allow to delimit stability/instability regions with a remarkable accuracy.

We study the dynamics of the orbits depending on the value of the above introduced measures. Figures 4–7 show the dynamics for the M/M/1/1-type system with



Fig. 3. Stability/instability region for the 2-class M/M/1/1-type system with $\lambda_1 = 0.5$ and $\lambda_2 = 4$.



Fig. 4. Orbit dynamics for the 2-class M/M/1/1-type system with $\Gamma^{(1)} = 0.493$, $\Gamma^{(2)} = 0.1$ and $\mu = 10$.



Fig. 5. Orbit dynamics for the 2-class M/M/1/1-type system with $\Gamma^{(1)} = 0.493$, $\Gamma^{(2)} = -0.05$ and $\mu = 10$.



Fig. 6. Orbit dynamics for the 2-class M/M/1/1-type system with $\Gamma^{(1)} = 0.504$, $\Gamma^{(2)} = -1.25$ and $\mu = 10$.



Fig. 7. Orbit dynamics for the 2-class M/M/1/1-type system with $\Gamma^{(1)} = -3.05$, $\Gamma^{(2)} = -8$ and $\mu = 2.5$.

 $\lambda_1 = 0.5, \lambda_2 = 4$ and different $\Gamma^{(i)}$. For instance, as Fig. 4 shows, both orbits are stable due to $\Gamma^{(1)} > 0, \Gamma^{(2)} > 0$.

Figures 5 and 6 demonstrate the situation when $\Gamma^{(1)} > 0$, $\Gamma^{(2)} < 0$, that is, when only the 1st orbit satisfies the stability condition. It is expected that the 1st orbit is stationary, while the 2nd one will be unstable, and this is confirmed by the results shown in these figures. Moreover, as Fig. 6 shows, the 2nd orbit evolves faster to infinity compared to the case presented in Fig. 5. This can be explained by the larger value of the instability measure $\hat{\Gamma}^{(2)}$. Finally, as Fig. 7 shows that both orbits become unstable when both conditions (31) and (32) are violated, that is $\Gamma^{(1)} < 0$ and $\Gamma^{(2)} < 0$.

In addition, Figs. 8 and 9 demonstrate the dynamics of the orbits for the 2-class M/Pareto/1/1-type system with Pareto service time complementary cumulative distribution function $P(S \ge x) = (x/x_0)^{-\alpha}, x \ge x_0$ (with the scale parameter



Fig. 8. Orbit dynamics for the 2-class M/Pareto/1/1-type system with $\Gamma^{(1)} = 0.04$, $\Gamma^{(2)} = -9$ and $\mu = 0.875$.



Fig. 9. Orbit dynamics for the 2-class M/Pareto/1/1-type system with $\Gamma^{(1)} = -9$, $\Gamma^{(2)} = 0.006$ and $\mu = 0.875$.

 $x_0 = 1$ and the shape parameter $\alpha = 8$) with input rates $\lambda_1 := 0.1$, $\lambda_2 := 0.7$ and service rate $\mu = (\alpha - 1)/(\alpha x_0) = 0.875$. Again these results are consistent with the observed stability conditions.

Thus, as simulations show, conditions (31) and (32) are, in fact, stability criteria for (at least) the 2-class retrial systems considered here.

5. Conclusion

In this paper, we have considered a retrial queueing system with N classes of customers in which a class-*i* blocked customer joins orbit *i*. Orbit *i* operates as a single-server queueing system with exponential service time with rate $\mu_0^{(i)}$, which is indeed the constant retrial rate of orbit *i*-class customers in our model. We have obtained the necessary stability conditions of such a system, expressed in terms of the system parameters that are given. Interestingly, these conditions have a clear probabilistic interpretation. Simulations demonstrate that these condition allow to delimit the stability region with a remarkable accuracy. These results, combined with some key observations concerning the retrial system, strongly indicate that these conditions are indeed the stability criteria, at least for the considered 2-orbit M/M/1/1 and M/Pareto/1/1 systems.

Acknowledgments

The research of EM and RN is supported by the Program of strategic development of Petrozavodsk State University for 2012–2016 and Russian Foundation for Basic research, Project no. 10-07-00017.

References

- Artalejo, JR (1996). Stationary analysis of the characteristics of the M/M/2 queue with constant repeated attempts. Opsearch, 33, 83–95.
- Artalejo, JR, A Gómez-Corral and MF Neuts (2001). Analysis of multiserver queues with constant retrial rate. European Journal of Operational Research, 135, 569–581.
- Asmussen, S (2003). *Applied Probability and Queues*, 2nd edn. New York: Springer-Verlag. Avrachenkov, K and U Yechiali (2008). Retrial networks with finite buffers and their
- application to internet data traffic. Probability in the Engineering and Informational Sciences, 22, 519–536.
- Avrachenkov, K and U Yechiali (2010). On tandem blocking queues with a common retrial queue. Computers and Operations Research, 37(7), 1174–1180.
- Avrachenkov, K and E Morozov (2010). Stability analysis of GI/G/c/K retrial queue with constant retrial rate. INRIA Research Report No. 7335. Available at http://hal.inria.fr/inria-00499261/en/. Last accessed on January 2014.
- Avrachenkov, K, RS Goricheva and EV Morozov (2011). Verification of stability region of a retrial queuing system by regenerative method. In Proc. Int. Conf. "Modern Probabilistic Methods for Analysis and Optimization of Information and Telecommunication Networks", Minsk, pp. 22–28.
- Bertsekas, D and R Gallager (1992). *Data Networks*, 2nd edn., New Jersey, USA: Prentice-Hall International.

- Choi, BD, YW Shin and WC Ahn (1992). Retrial queues with collision arising from unslotted CSMA/CD protocol. Queueing Systems, 11, 335–356.
- Choi, BD, KH Rhee and KK Park (1993). The M/G/1 retrial queue with retrial rate control policy. Probability in the Engineering and Informational Sciences, 7, 29–46.
- Fayolle, G (1986). A simple telephone exchange with delayed feedback. In *Teletraffic Anal-ysis and Computer Performance Evaluation*, OJ Boxma, JW Cohen and HC Tijms (Eds.), Teletrac Analysis and Computer Performance Evaluation, Vol. 7, Amsterdam: Elsevier Science Publishers, pp. 245–253.
- Lillo, RE (1996). A G/M/1-queue with exponential retrial. TOP, 4(1), 99–120.
- Morozov, E (2004). Weak regeneration in modeling of queueing processes. Queueing Systems, 46, 295–315.
- Morozov, E (2007). A multiserver retrial queue: Regenerative stability analysis. Queueing Systems, 56, 157–168.
- Morozov, E and R Nekrasova (2011). Estimation of blocking probability in retrial queuing system with constant retrial rate. In *Proc. Institute of Applied Mathematical research, Karelian Research Centre RAS*, pp. 63–74 (in Russian).
- Morozov, E and R Nekrasova (2012). On the estimation of the overflow probability in regenerative finite buffer queueing systems. *Informatics and their Applications*, 3(6), 1977–1991 (in Russian).
- Ramalhoto, MF and A Gómez-Corral (1998). Some decomposition formulae for M/M/r/r+d queues with constant retrial rate. Stochastic Models, 14, 123–145.
- Sigman, K (1990). One-dependent regenerative processes and queues in continuous time. Mathematics of Operations Research, 15, 175–189.
- Smith, WL (1955). Regenerative stochastic processes. Proceedings of Royal Society, Series A, 232, 6–31.

Konstantin Avrachenkov received the Master degree in Control Theory from St. Petersburg State Polytechnic University (1996), PhD degree in Mathematics from University of South Australia (2000) and Habilitation (Doctor of Science) from University of Nice Sophia Antipolis (2010). Currently he is a Director of Research at INRIA Sophia Antipolis, France. His main research interests are Markov processes, singular perturbation theory, queueing theory, mathematical programming, game theory and performance evaluation of data networks.

Evsey Morozov received the PhD degree in Physics and Mathematics from Kiev Institute of Cybernetics in 1979 and the degree of Doctor of Sciences from Moscow Institute of Control Systems in 1996. Since 1986 he has been working as a Professor of the Department of Applied Mathematics and Cybernetics at the Petrozavodsk State University, Russia. Since 2000, he is also a leading researcher at the Institute of Applied Mathematical Research of Karelian Research Centre of Russian Academy of Sciences. His main research interests include stability and performance analysis of queues, with applications to modern telecommunication systems and networks and also regenerative simulation.

Ruslana Nekrasova received the Master degree from Mathematics department of Petrozavodsk State University in 2009. Then she has received PhD degree from Institute of Applied Mathematical Research in Karelian Research Centre of the Russian Academy of Sciences (IAMR KRC RAS) in 2012. Currently she works as a junior researcher at IAMR KRC RAS. Her main research interests are regenerative processes, retrial queueing systems and regenerative estimation of time averages.

Bart Steyaert was born in Roeselare, Belgium, in 1964. He received the degrees of Licentiate in Physics and Licentiate in Computer Science in 1987 and 1989, respectively, and a PhD in Engineering Sciences in 2008, all from the University of Ghent, Belgium. Since January 1990, he has been working as a researcher at the SMACS Research Group, Department for Telecommunications and Information Processing (TELIN), at the same university, in the context of various research projects. His main research interests include discrete-time queueing models, traffic control, and stochastic modelling of high-speed communications networks.