Baltzer Journals Submitted: January 1996. Revised: March 1997

On the stability of retrial queues

Eitan Altman [1] and Aleksandr A. Borovkov [2]

[1] *INRIA, B.P.93*
2004 Route des Lucioles
06902 Sophia Antipolis Cedex
France
E-mail: altman@sophia.inria.fr

[2] *Institute of Mathematics*
Russian Academy of Science
Novosibirsk, 630090
Russia
E-mail: borovkov@math.nsc.su

We consider the following type of problems. Calls arrive at a queue of capacity K (which is called the primary queue), and attempt to get served by a single server. If upon arrival, the queue is full and the server is busy, the new arriving call moves into an infinite capacity orbit, from which it makes new attempts to reach the primary queue, until it finds it non-full (or it finds the server idle). If the queue is not full upon arrival, then the call (customer) waits in line, and will be served according to the FIFO order. If λ is the arrival rate (average number per time unit) of calls and μ is one over the expected service time in the facility, it is well known that $\mu > \lambda$ is not always sufficient for stability. The aim of this paper is to provide general conditions under which it is a sufficient condition. In particular, (i) we derive conditions for Harris ergodicity and obtain bounds for the rate of convergence to the steady state and large deviations results, in the case that the inter-arrival times, retrial times and service times are independent i.i.d. sequences and the retrial times are exponentially distributed; (ii) we establish conditions for strong coupling convergence to a stationary regime when either service times are general stationary ergodic (no independence assumption), and inter-arrival and retrial times are i.i.d. exponentially distributed; or when inter-arrival times are general stationary ergodic, and service and retrial times are i.i.d. exponentially distributed; (iii) we obtain conditions for the existence of uniform exponential bounds of the queue length process under some rather broad conditions on the retrial process. We finally present conditions for boundedness in distribution for the case of nonpatient (or non persistent) customers.

1 Introduction and description of the basic model

We consider in this paper the stability analysis of retrial queues. These may model both telecommunications problems, in which users that find the line busy retry to establish their call at a later time, maintenance and repair problems and many others. We specify to the single server retrial queue: customers (calls) who arrive at a queue of capacity K (which is called the primary queue), attempt to get served by a single server. If upon arrival, the queue is full, and the server is busy, the new arriving call moves into an infinite capacity orbit, from which it makes new attempts to reach the primary queue, until it finds it non-full or the server is idle. If the queue is not full upon arrival, then the call waits in line, and will be served according to the FIFO order.

Many papers have been devoted to obtaining stability conditions for the single server retrial queues. For the M/GI/1 case, i.e. Poisson arrivals, generally distributed i.i.d. service time, and exponentially distributed retrial times, it is known that the stability condition is that the arrival rate λ times the expected service time $1/\mu$ is smaller than one. This has been derived by Keilson et al. [17] and Falin [11, 12]. For general inter-arrival and retrial distribution, $\lambda < \mu$ is not always sufficient for stability, as illustrated by a counterexample, see [18]. However, Liang and Kulkarni [18] show that this condition is sufficient for the case of general independent service time, and retrial and inter-arrival times that have both finite mixture of Erlang distributions (with, however, $K = 0$). For the case of nonpersistent calls, stability conditions were obtained in [13]; in that model the expected number of retrials is finite, and after (a finite) number of retrials the call disappears even if not served. Conditions for ergodicity for the case of several servers (the multichannel case) with exponential service and retrial times and Poisson inter-arrival times, are studied in [7, 8, 9, 10, 16]. Recently, Spieksma and Tweedie [19] presented results on rate of convergence in such models, namely, establishing conditions for geometric ergodicity.

The purpose of this paper is to obtain other assumptions for which $\mu > \lambda$ is a sufficient condition for stability and to obtain boundness (in distribution) of the queue length for cases that obtaining ergodicity becomes too difficult. A source of difficulties is that in case that retrial times are not exponentially distributed, the state space in which we can construct a Markov chain becomes very complicated. We present several sets of conditions that imply different types of stability. We establish ergodicity under the above condition for the following cases: (i) when inter-arrival times, service times, and retrial times are i.i.d. independent sequences, and the retrial times are exponentially distributed (Sec. 2). Under these conditions we obtain also bounds on the rate of convergence and large deviations (Sec. 3). (ii) general stationary (in the strict sense) ergodic service times (no independence assumption), and exponentially distributed inter-arrival and retrial times (Sec. 4). We then consider the dual case: general stationary (in the strict sense) ergodic inter-arrival times, and exponentially distributed service and retrial times (Sec. 5). (iii) We obtain uniform exponential bounds of the queue length process under some rather broad conditions on the retrial process (Sec. 6).

We finally consider in Sec. 7 models in which customers are not patient (non-persistent), for which $\mu > \lambda$ turns not to be a necessary condition for boundedness. In these models, if customers are not served by some certain (random) time limit, they leave the system. If this limit has finite expectation, then no conditions on the service times are required in order to obtain boundedness in distribution. If this limit has infinite expectation, we show that the condition for boundedness in distribution is that the expected service time is finite.

Next, we present the basic model, which is analyzed in the following sections. Calls arrive to the main queue with capacity K (called the primary queue) at times $\{t_i\}, i = 1, 2, ...$; $\tau_i = t_{i+1} - t_i$ denote the inter-arrival times, $i = 1, 2, ...$ If the i-th arriving call (customer) finds the queue non-full, it joins it and waits till all customers ahead of it are served, it is then served too and leaves the system. Otherwise, it moves immediately to an orbit, it repeatedly retries to join the primary queue at times $\{R_i^j\}$, $j = 1, 2, ...$ until it finds the queue non-full; in that case it joins the queue and will be served after all customers ahead of line are served. Such a customer is called *patient.* Let $r_i^1 = R_i^1$, and $r_i^j = R_i^{j+1} - R_i^j$, $j = 2, ...$ denote the times between retrials. The n-th service duration is σ_n, and we assume that $0 < E\sigma_n < \infty$.

We shall consider in in particular $K = 0$, which means that there is no main queue in which customers can wait; if an arriving customer finds the server busy, it has to go to the retrial queue. Otherwize it starts to be served immediately.

Introduce the following processes:

- $\varphi(t)$:= the residual retrial time, i.e. the time difference between the first instant at which a customer from the orbit retries to join the primary queue after time t, and t.
- $Q(t)$:= the number of customers in the orbit at time t.
- $q(t)$:= the number of customers in the primary queue, at time t.
- $\chi(t)$:= the residual inter-arrival time at time t.
- $\mathcal{F}(t)$:= the sigma algebra generated by events describing the state of the system up to time t.

$Q(t), q(t), \varphi(t), \chi(t)$ and $\mathcal{F}(t)$ are assumed to be left continuous. $(\Omega, \mathcal{F}, \mathbf{P})$ denotes the probability space on which the above random processes are defined, such that $\mathcal{F}(t) \subset \mathcal{F}$.

2 Ergodicity under i.i.d. assumptions, and exponential retrial times

We assume throughout this section the following: The inter-arrival times, service times and retrial times are each an i.i.d. sequence, and they are mutually independent. Moreover, the retrial times in the model described in Section 1 (for each customer) are exponentially distributed with parameter θ.

In this section we define S_n to be the instant at which the n-th service time begins. We consider the Markov chain (q_n, Q_n, χ_n), where q_n denotes the number of customers in the primary queue just before S_n (i.e. $q_n = q(S_n) = q(S_n - 0)$), Q_n denotes the number of customers in the orbit an instant before S_n, and $\chi_n = \chi(S_n) = \chi(S_n - 0)$ denotes the residual inter-arrival time at that instant. (In particular, if the system is empty just before S_n, then $\chi_n = 0$). We further restrict to the embedded Markov chain $X(n) := (X_1(n), X_2(n), X_3(n)) = (q_{nm}, Q_{nm}, \chi_{nm})$, where m will be determined later. Let $\mathbf{X}$ denote the state space of this Markov chain, i.e. $\mathbf{X} = \{0, 1, ..., K\} \times \mathbb{N} \times \mathbb{R}_+$.

Theorem 1
Assume that $E\sigma_1 < E\tau_1$. Then $X(n)$ is Harris ergodic.

Proof. The proof is based on the use of a test function (Lyapunov function), to obtain negative drift outside some compact set, which will imply the positive recurrence of that set; we then show that the probability to reach the state $\mathbf{0} := (0,0,0)$ within some fixed number N_0 of steps from any point in that compact set is uniformly lower bounded over the set by some positive constant p. This will imply the positive recurrence of point $\mathbf{0}$ and the Harris ergodicity. Define the following independent sequences

- σ and $\{\sigma_i^n\}$, $i = 1, ..., m, n = 0, 1, 2, ... :=$ i.i.d. random variables distributed like σ_1.
- α and $\{\alpha_i^n\}$, $i = 1, ..., m, n = 0, 1, 2, ... :=$ i.i.d. exponential random variables with parameter θ.

For any $N > 0$, if $X_1(n) + X_2(n) > K + N + m$ then

$$S_{(n+1)m} - S_{mn} \leq_{st} \sum_{i=1}^{m} \left(\sigma_i^n + \frac{\alpha_i^n}{N} \right) =: v_n^m, \tag{1}$$

$$E[S_{(n+1)m} - S_{nm}] \leq E v_n^m = m \left(E\sigma_1 + \frac{1}{N\theta} \right).$$

(1) follows from the fact that the number of customers in orbit is not less that N during the whole interval $[S_{nm}, S_{(n+1)m}]$. Therefore, for any $nm \leq k < (n+1)m$, $S_{k+1} - S_k$ is stochastically upper bouned by the sum of a service time and an exponential RV with parameter $N\theta$.

Let ξ be a positive r.v., ζ_i be i.i.d. r.v.s independent of ξ and distributed like ζ, $E\zeta > 0$. Define

$$\nu(\zeta, \xi) = min\{k \geq 1 : \sum_{i=1}^{k} \zeta_i \geq \xi\} \tag{2}$$

Then $\nu(\tau, v_n^m)$ will be the number of arrivals during the period $[0, v_n^m]$, where $\chi_0 = 0$ and the customer arriving at moment 0 is included.

We use the test function $L : \mathbf{X} \to \mathbb{R}$ defined as the total number of customers in the system, i.e.

$$L(x) = x_1 + x_2. \tag{3}$$

It follows from (1) that for any $N > 0$, if $L(X(n)) > K + N + m$ then given $X(n)$,

$$L(X(n+1)) - L(X(n)) \leq_{st} \nu(\tau, v_n^m) - m \tag{4}$$

(in other words, for each real c, $\mathbf{P}(L(X(n+1)) - L(X(n)) > c | X(n)) < \mathbf{P}(\nu(\tau, v_n^m) - m > c)$). So,

$$\begin{aligned} E\left(L\left(X(n+1)\right) - L\left(X(n)\right) | X(n)\right) &\leq E\nu(\tau, v_n^m) - m \\ &= -m + \frac{E(v_1^m) + E[\chi(v_1^m)|\chi(0) = 0]}{E\tau_1} \\ &= -m + \frac{m[E\sigma_1 + (N\theta)^{-1}] + E[\chi(v_1^m)|\chi(0) = 0]}{E\tau_1} \end{aligned} \tag{5}$$

It follows from renewal arguments (see [1, 2]) that $E[\chi(t)|\chi(0) = 0]$ is $o(t)$. Hence, for any $a > 0$ there exists some constant $b > 0$ such that

$$E[\chi(v_1^m)|\chi(0) = 0] \leq aE(v_1^m) + b = ma \left(E\sigma_1 + \frac{1}{N\theta} \right) + b \tag{6}$$

Combining (5) with (6) we obtain for any $N > 0$, and $L(X(n)) > K + N + m$,

$$\begin{aligned} E\left[L\left(X(n+1)\right) - L\left(X(n)\right) \middle| X(n)\right] &\leq E\nu(\tau, v_n^m) - m \\ &\leq \frac{m\left([E\sigma_1 + (N\theta)^{-1}](1+a) - E\tau_1\right) + b}{E\tau_1} \end{aligned} \quad (7)$$

Choose N and a such that

$$[E\sigma_1 + (N\theta)^{-1}](1+a) - E\tau_1 < 0.$$

With b that corresponds to that choice of a, choose m such that

$$[E\sigma_1 + (N\theta)^{-1}](1+a) - E\tau_1 + b/m < 0. \quad (8)$$

So we have for $x \notin \mathbf{V}$, where

$$\mathbf{V} := \{x : L(x) \leq K + N + m\} \quad (9)$$

from (7),

$$E\left[L\left(X(n+1)\right) - L\left(X(n)\right) \middle| X(n) = x\right] \leq E\nu(\tau, v_n^m) - m < 0. \quad (10)$$

Next we show that $E\left[L\left(X(n+1)\right) \middle| X(n)\right]$ is bounded for any $X(n)$ in the set $\mathbf{V}$. The number of arrivals between S_{nm+k+1} and S_{nm+k} is stochastically upper bounded by $\nu(\tau, \sigma + \alpha)$. Therefore

$$L(X(n+1)) - L(X(n)) \leq_{st} \sum_{i=1}^{m} \nu_i - m, \quad (11)$$

where ν_i are independent and distributed like $\nu(\tau, \sigma + \alpha)$,

$$E\nu(\tau, \sigma + \alpha) = \int \mathbf{P}(\sigma + \alpha \in dv) E\nu(\tau, v), \quad (12)$$

$$E\nu(\tau, v) = \frac{v + E\chi(v)}{E\tau} \leq \frac{(1+a)v + b}{E\tau}.$$

So

$$E\nu(\tau, \sigma + \alpha) \leq \frac{(1+a)(E\sigma + \theta^{-1}) + b}{E\tau}$$

and for any $X(n) \in \mathbf{V}$

$$E[L(X(n+1)) - L(X(n)|X(n)] \leq m \frac{(1+a)(E\sigma + \theta^{-1}) + b}{E\tau} - m. \quad (13)$$

Together with (10) this implies that the set $\mathbf{V}$ is positive recurrent (see [4, 20]).

Next, we show that there exist some constants $N_0, p > 0$ such that

$$\inf_{x \in \mathbf{V}} \mathbf{P}(X(N_0) = \mathbf{0} | X(0) = x) \geq p. \quad (14)$$

The scheme of the proof of (14) is quite evident. Because of the exponentiality of r_i^j, the probability of the event $\bigcap_{ij\in M}\{r_i^j \in \triangle_i^j\}$ for any finite fixed set M of non-empty intervals $\triangle_i^j$ is positive. First, assume for simplicity that $\tau_n = c_1$ and $\sigma_n = c_2$ are deterministic. Since $c_1 > c_2$ then for any fixed $X_1(0) = x_1, X_2(0) = x_2, x_1 + x_2 < K + N + m$, one can always choose a finite set M of non-empty intervals $\triangle_i^j$ such that $\bigcap_{ij\in M}\{r_i^j \in \triangle_i^j\} \subset \{X(N_0) = \mathbf{0}\}$, where N_0 is defined by M. It means that

$$\inf_{x_1+x_2\leq K+N+m} \mathbf{P}(X(N_0) = \mathbf{0}|X(0) = x) > 0.$$

Further, one can preserve this inequality when allowing τ_n, σ_n to be random but in sufficiently small neighborhoods of points c_1, c_2.

In the general case, all these considerations remain true if one takes instead of points c_1, c_2 points $c_3 > c_4$ from the supports of the distributions of τ_n and σ_n, respectively. ∎

3 Rate of convergence and large deviations

We make here the same assumptions as in Section 2, i.e. the service, inter-arrival and retrial times are independent i.i.d. sequences, and the retrial times are exponentially distributed. We characterize the rate of convergence to steady state, and obtain large deviation results and existence of moments.

Following Borovkov [2], introduce the following classes of functions:

$\mathcal{M}^\delta$:= the class of functions $g(\cdot) : [0,\infty) \to (0,\infty)$ such that there exists a constant $\delta \geq 1$ satisfying

$$\int_0^\infty t^{\delta-1} g(t)dt < \infty.$$

For a random variable ξ, if we have $\mathbf{P}(\xi > t) \leq g(t)$ and $g(t) \in \mathcal{M}^\delta$, then $E[\max(\xi,0)^\delta] < \infty$. Hence, this class of functions will be related to statements on moments of random variables.

$\mathcal{M}_\delta$:= the class of functions $g(\cdot) : [0,\infty) \to (0,\infty)$ such that there exists a constant $\delta \geq 1$ satisfying $g(t) \leq \epsilon(t)t^{-\delta}$ where $\epsilon(t)$ is a slowly varying function (i.e., for all $a > 0$, $\lim_{t\to\infty} \epsilon(at)/\epsilon(t) = 1$).

$\mathcal{M}_C$:= the class of nonincreasing functions $g(\cdot) : [0,\infty) \to (0,\infty)$ that satisfy Cramer's condition, i.e. there exist some $c_1, c_2 > 0$ such that

$$g(t) < c_1 e^{-c_2 t} \tag{15}$$

for all $t > 0$. Condition (15) is equivalent to the existence of $c_3 > 0$ such that

$$\int_0^\infty e^{c_3 t} g(t)dt < \infty. \tag{16}$$

Note that if $g \in \mathcal{M}^\delta$ then $g \in \mathcal{M}_\delta$ with $\epsilon(t) = const$. If $g \in \mathcal{M}_\delta$, then $g \in \mathcal{M}^{\delta'}$ for all $\delta' < \delta$. Below we shall use $\mathcal{M}$ to denote any one of the classes $\mathcal{M}^\delta, \mathcal{M}_\delta, \mathcal{M}_C$. We shall say that a random variable ξ is in $\mathcal{M}$ if $\mathbf{P}(\xi > \cdot) \in \mathcal{M}$.

The classes $\mathcal{M}$ are understood as equivalent up to constants. More precisely, we consider $g(.) \in \mathcal{M}$ to represent all functions of the form

$$c_4 g(c_5 + c_6 t), \tag{17}$$

where $c_4 < \infty, c_5 < \infty, 0 < c_6 < \infty$ are constants.

Consider now an arbitrary Markov chain $Y(n)$ in $\mathbf{X}$ and assume the following standard mixing condition: for some set $\mathbf{V} \in \mathbf{X}$, there exist some constants $\hat{N}_0, \hat{p}_0 > 0$ and a measure ψ on $\mathbf{X}$ such that for all Borel sets $\mathbf{B}$,

$$\inf_{x \in \mathbf{V}} \mathbf{P}(Y(\hat{N}_0) \in \mathbf{B} | Y(0) = x) \geq \hat{p}_0 \psi(\mathbf{B}). \tag{18}$$

Note that (14) implies (18) for $X(n) = Y(n)$ and $\mathbf{V}$ given in (9). Define

$$\mathcal{T}(x) := \min_{n > 0} \{Y(n) \in \mathbf{V} \text{ given that } Y(0) = x\}. \tag{19}$$

We introduce condition (I,$\mathcal{M}$):

$$\sup_{x \in \mathbf{V}} \mathbf{P}(\mathcal{T}(x) > k) < g(k), \;\; g \in \mathcal{M} \tag{20}$$

($\mathbf{V}$ is given above (18)).

Theorem 2
([4] Chapter 3) *Assume that conditions (18) and* (I,$\mathcal{M}$) *hold, and that the test function* L *satisfies*

$$\sup_x \mathbf{P}\left(L(Y(1)) - L(Y(0)) > \cdot \;|Y(0) = x\right) \in \mathcal{M}. \tag{21}$$

Then

$$\sup_n \sup_{x \in \mathbf{V}} \mathbf{P}\left(L(Y(n))) > v | Y(0) = x\right) \leq v g(v), \quad g \in \mathcal{M}, \tag{22}$$

$$\sup_{x \in \mathbf{V}} \mathbf{P}\left(\max_{k \leq n} L(Y(k)) > v\right) \leq n g(n) + 1 - (1 - g(v))^{nc} \tag{23}$$

for some $c > 0$. *Besides,*

$$\| \mathbf{P}(Y(n) \in \cdot \;|Y(0) = x) - \pi(\cdot) \| < n g(n) \tag{24}$$

where π *is the steady state probability of the Markov chain, and the above norm refers to the total variation norm*

Note that in (24), the initial state x is fixed; $g(n)$ can be made independent of x if $|x| < hn$ for some $h > 0$.

We make use of the following:

Theorem 3
([4] Ch. 11 Sec. 41 Thm. 3) *Let* $\xi \in \mathcal{M}$ *be a non-negative random variable; Let* ζ_i *be an i.i.d. sequence with* $E\zeta_i > 0$ *and* $\max(0, -\zeta_i) \in \mathcal{M}$. *Then* $\nu(\zeta, \xi) \in \mathcal{M}$ *(see (2) for the notation).*

Corollary 4
Assume that $E\sigma_1 < E\tau_1$. If $\mathbf{P}(\sigma_1 > v) \le g(v)$ for some $g \in \mathcal{M}$ then (22), (23) and (24) hold, for process $Y(n) = X(n)$, for L defined in (3), $\mathbf{V}$ defined in (9), and $g \in \mathcal{M}$ for the same class $\mathcal{M}$ (up to some constants, as in (17)).

Proof of Corollary 4 According to considerations in the proof of Theorem 1

$$L(X(1)) - L(X(0)) \le_{st} \sum_{i=1}^{m} \nu_i - m$$

for all $X(0)$, where ν_i are independent and distributed like $\nu(\tau, \sigma + \alpha)$. But $\nu_i \in \mathcal{M}$ due to Theorem 3. This proves condition (21) of Theorem 2.

Next we establish condition (I,$\mathcal{M}$). We shall show that

$$\sup_{x \in \mathbf{V}} \mathbf{P}(\mathcal{T}(x) > \cdot) \in \mathcal{M}.$$

For $x \in \mathbf{V}$, $k > 1$ we have for $X(0) = x$

$$\mathbf{P}(\mathcal{T}(x) > k) = \mathbf{P}(\mathcal{T}(x) > k,\ X(1) \notin \mathbf{V}). \tag{25}$$

As we know, the conditional expectation of increments of $L(X(n))$ outside $\mathbf{V}$ does not exceed (in the stochastic ordering sense) $-\zeta_n$, where

$$-\zeta_n := \nu(\tau, v_n^m) - m,$$

(by (4)), $E\zeta_n > 0$ (by (10)), and $\max(0, -\zeta_n) \in \mathcal{M}$ (this follows from Theorem 3). Therefore

$$\mathbf{P}(\mathcal{T}(x) > k, X(1) \notin \mathbf{V}) \le \mathbf{P}(\nu(\zeta, L(X(1))) > k - 1,\ X(1) \notin \mathbf{V}). \tag{26}$$

Since $L(X(1)) \in \mathcal{M}$ (uniformly with respect to $x = X(0) \in \mathbf{V}$), then we obtain by Theorem 3

$$\sup_{x \in \mathbf{V}} \mathbf{P}(\mathcal{T}(x) > k) \in \mathcal{M}.$$

Thus we proved conditions (18), (20) (see the end of the proof of Theorem 1) and (21) for the Markov chain $X(n)$. By Theorem 2, this implies assertions (22)–(24). ∎

4 Ergodicity for stationary ergodic service times and i.i.d. exponential retrial and inter-arrival times

We assume throughout this section that the sequence $\{\sigma_n\}$ forms a stationary ergodic sequence, the inter-retrial and inter-arrival times are each i.i.d. exponentially distributed sequences with parameters θ and λ respectively, they are independent of each other and of the sequence $\{\sigma_n\}$. (Here and throughout, stationarity is always understood in the strict sense.) Assume $K = 0$. In this section we define S_n to be the instant when the $n-1$st service time ends (and the served customer leaves the system). We consider the process Q_n embedded immediately after times S_n, (i.e. $Q_n = Q(S_n + 0)$).

Since $K = 0$, the time between the end of the $(n-1)$th service till the beginning of the n-th service (we call this the n-th idle period) is exponentially distributed, with parameter $\lambda + \theta Q_n$. This idle period ends when either there is an external arrival, or a call from the orbit retries to connect. The conditional probability, given $\mathcal{F}(S_n)$, that the first event to occur after the $(n-1)$th service ends (and after the served customer has left the system) is a retrial, i.e. an arrival from the orbit to the primary queue, is $\theta Q_n/(\lambda + \theta Q_n)$. The conditional probability, given $\mathcal{F}(S_n)$, that the first event to occur after the $(n-1)$th service ends (and after the served customer has left the system) is an external arrival, is $\lambda/(\lambda + \theta Q_n)$.

The process Q_n is a *Recursive Chain* (this is a generalization of a Markov chain, due to Borovkov and Foss [4, 6]). However, we shall instead express the sequence Q_n in a simple way as a Stochastic Recursive Sequence (SRS), see [2, 4, 6]. A process Q_n is said to be a SRS with a stationary ergodic driving sequence κ_n if it satisfies for all $n \geq 0$

$$Q_{n+1} = f(Q_n, \kappa_n) \tag{27}$$

where κ_n, which is called a driving sequence, is a stationary ergodic stochastic process, and f is some function.

Let u_n^1, u_n^2 be two i.i.d. sequences of random variables distributed uniformly on [0,1], mutually independent, and independent of the sequence σ_n. $u^1 = \{u_n^1\}$ will generate the arrival process, and $u^2 = \{u_n^2\}$ will generate the type of arrival (external or from the orbit) at the end the service period. More precisely, let $\Pi : R_+ \times [0,1] \to \mathbb{N}$ denote the inverse of the Poisson distribution

$$\Pi(t,x) = \inf \left\{ n \in N : \sum_{k=0}^{n} \frac{t^k e^{-t}}{k!} \geq x \right\},$$

so that $\Pi(t, u_n^1)$ is a Poisson random variable with parameter t. We have the following representation of the process Q_n:

$$Q_{n+1} = \max\left(0, Q_n + h(Q_n, \sigma_n, u_n^1, u_n^2)\right) \tag{28}$$

where

$$h(Q_n, \sigma_n, u_n^1, u_n^2) = \Pi(\lambda\sigma_n, u_n^1) - 1\left\{ u_n^2 \leq \frac{Q_n\theta}{\lambda + Q_n\theta} \right\} \tag{29}$$

The process $\kappa_n := (\sigma_n, u_n^1, u_n^2)$ is the driving sequence for the SRS, taking values in $\mathbf{Z} = \mathbb{R}_+ \times [0,1] \times [0,1]$, and is stationary ergodic. For such sequences, Borovkov [2], and later Borovkov and Foss [6] (see also [4]) obtained conditions for the coupling, and for strong coupling (defined below) to a stationary ergodic process. They show that under suitable conditions, there exists on the same probability space a stationary ergodic process $\overline{Q}(n)$ such that

$$\lim_{n\to\infty} \mathbf{P}(Q_k = \overline{Q}(k) \text{ for all } k \geq n) = 1, \tag{30}$$

the latter is known as coupling convergence of Q_n to $\overline{Q}(n)$.

Define the σ-algebras $\mathcal{F}_n^{\sigma,u} := \sigma\{\sigma_k, u_k^1, u_k^2; k \leq n\}$, $\mathcal{F}^{\sigma,u} := \sigma\{\sigma_k, u_k^1, u_k^2; k < \infty\}$. Let U^k be the k-step (measure-preserving) shift of $\mathcal{F}^{\sigma,\nu}$-measurable random variables generated by $\{\sigma_n, u_n^1, u_n^2\}$, (so that $U^k\sigma_n = \sigma_{n+k}$, etc...). Let T^k be the k-step shift transformation of sets from the σ-algebra $\mathcal{F}^{\sigma,\nu}$, so that for any Borel sets $\mathbf{B}_j \subset \mathbf{Z}$,

$$T\{\omega : (\sigma_j, u_j^1, u_j^2) \in \mathbf{B}_j; j = 1, ..., k\} = \{\omega : (\sigma_{j+1}, u_{j+1}^1, u_{j+1}^2) \in \mathbf{B}_j; j = 1, ..., k\} \quad (31)$$

and T^k is the k-th iteration of T. Define

$$\mu_k = \min\left\{n \geq -k : U^{-k}Q_{m+k} = \overline{Q}(m) \text{ for all } m \geq n\right\}. \quad (32)$$

We say that Q_n strongly couple-converges to $\overline{Q}(n)$ if $\sup_{k\geq 0}\mu_k < \infty$ *a.s.* (this implies (30), and is thus stronger than coupling convergence).

Theorem 5
Assume that $E\tau_1 > E\sigma_1$. Then the process Q_n strongly couple-converges to a unique stationary ergodic regime.

Proof. The proof is based on constructing stationary renovation events $\mathbf{A}_n$, then showing that they have strictly positive probability, and then applying an ergodic theorem introduced by Borovkov [2]. The sets $\mathbf{A}_n$ will be constructed in two steps. We shall first introduce a majorizing SRS $\{Q_n^*\}$ on the same probability space as the original one. It will have a simpler structure than the original process, which will enable us to obtain simple stationary renovating events $\mathbf{B}_n$ with positive probability, for the new process. $\mathbf{A}_n$ will be obtained as some subsets of $\mathbf{B}_n$.

We begin by constructing the majorizing SRS $\{Q_n^*\}$ on the same probability space:

$$Q_0^* = Q_0, \qquad Q_{n+1}^* = \max(C, Q_n^* + h_n^*) \quad (33)$$

where

$$h_n^* = h^*(\sigma_n, u_n^1, u_n^2) = \Pi(\lambda\sigma_n, u_n^1) - 1\left\{u_n^2 \leq \frac{C\theta}{\lambda + C\theta}\right\}. \quad (34)$$

C above is an arbitrary integer constant chosen such that $Eh_n^* < 0$, where

$$Eh_n^* = \frac{E\sigma_1}{E\tau_1} - \frac{\theta C}{\lambda + \theta C}. \quad (35)$$

It follows (see Example 1 in Borovkov and Foss [6]) that there exists some integer n_0 and a set $\mathbf{B}_0 \in \mathcal{F}_0^{\sigma,u}$ with $\mathbf{P}(\mathbf{B}_0) > 0$, such that $Q_n^* = C$ on the set $\mathbf{B}_n = T^n\mathbf{B}_0$ for all $n \geq n_0$. For the case that the initial state is $Q_0^* = v$, $\mathbf{B}_0$ is given [6] by

$$\begin{aligned}\mathbf{B}_0 = \{&h_{-1}^* \leq 0, h_{-1}^* + h_{-2}^* \leq 0, ..., \\ &h_{-1}^* + ... + h_{-n_0}^* \leq 0; h_{-1}^* + ... + h_{-n_0}^* + ... + h_{-n_0-l}^* \leq -v \text{ for all } l \geq 1\}\end{aligned} \quad (36)$$

Define the sets

$$D_0 = \left\{\Pi(\lambda\sigma_{-k}, u_{-k}^1) = 0, u_{-k}^2 \leq \frac{k\theta}{\lambda + k\theta}, k = 1, ..., C\right\}, \qquad D_n = T^nD_0. \quad (37)$$

We shall show that the sets $\mathbf{A}_n = \mathbf{B}_{n-C} \cap D_n$ form a stationary renovating sequence for Q_n (on the set $\mathbf{A}_n$, the values $Q_k, k > n$ do not depend (in an algebraic, not probabilistic sense) on $Q_k, k \le n$, and can be expressed as a function of the driving sequence at times $k \ge n$, see [2] for the exact definition of renovating sequences). For all $n \ge n_0 + C$ (n_0 defined above (36)), we have on $\mathbf{A}_n$, the values $Q_{n-k} \le k$, $k = 0, 1, ..., C$, and in particular, $Q_n = 0$. Moreover, $\mathbf{P}(\mathbf{A}_0) = \mathbf{P}(\mathbf{B}_{-C})\mathbf{P}\{D_0 | \mathbf{B}_{-C}\} > 0$ if we show that $\mathbf{P}(D_0|\mathbf{B}_{-C}) > 0$. Indeed, since $E(\sigma_{-C} + ... + \sigma_{-1}|\mathbf{B}_{-C}) \le (CE\sigma_1)/\mathbf{P}(\mathbf{B}_{-C})$ is finite, we have

$$\begin{aligned}
\mathbf{P}(D_0|\mathbf{B}_{-C}) &= E\left[e^{-\lambda(\sigma_{-C}+...+\sigma_{-1})}\middle| \mathbf{B}_{-C}\right] \prod_{k=1}^{C} \frac{k\theta}{\lambda + k\theta} \\
&\ge e^{-\lambda E[\sigma_{-C}+...+\sigma_{-1}|\mathbf{B}_{-C}]} \prod_{k=1}^{C} \frac{k\theta}{\lambda + k\theta} > 0.
\end{aligned}$$

The proof now follows from Theorem 1.4 in [6]. ∎

5 Ergodicity for stationary ergodic inter-arrival times and i.i.d. exponential retrial and service times

In this section we assume again $K = 0$. We further assume that the sequence $\{\tau_n\}$ forms a stationary ergodic sequence, the inter-retrial and service times are each i.i.d. exponentially distributed sequences with parameters θ and μ respectively, they are independent of each other and of the sequence $\{\tau_n\}$. Let $\mathcal{Q}_n = \mathcal{Q}(t_n) = \mathcal{Q}(t_n - 0), n = 1, 2, ...$ denote the process of total number of customers in the system (in service plus in orbit), embedded just prior to the arrival epochs t_n. We shall set $\mathcal{F}_n = \mathcal{F}(t_n)$. For each n, we define $\vec{\sigma}^n := \{\sigma_k^n, k = 1, 2, ...\}$ to be a sequence of i.i.d. exponential r.v.s with finite mean, and $\vec{\alpha}^n := \{\alpha_k^n, k = 1, 2, ...\}$ to be a sequence of i.i.d. exponential r.v.s with parameter θ. The sequences $\vec{\sigma}^n, \vec{\alpha}^n$ and τ_n are assumed to be independent of each other. Define $\Sigma_{0,n} := 0$, and for $k > 0$:

$$\Sigma_{k,n} := \sigma_1^n + \left(\frac{\alpha_1^n}{\mathcal{Q}_n} + \sigma_2^n\right) + \left(\frac{\alpha_2^n}{\mathcal{Q}_n - 1} + \sigma_3^n\right) + ... + \left(\frac{\alpha_{k-1}^n}{\mathcal{Q}_n - k + 2} + \sigma_k^n\right) \tag{38}$$

if $\mathcal{Q}_n \ge k - 1$. $\Sigma_{k,n}$ can be considered as the time it takes to serve k customers after time t_n, if there are no arrivals after time t_n. Note that just after the arrival occurs at t_n, there are $\mathcal{Q}_n + 1$ customers in the system.

Let $Z(n) := Z(\mathcal{Q}_n, \vec{\sigma}^n, \vec{\alpha}^n, \tau_n)$ be the number of customers that leave the system during the interval $[t_n, t_{n+1})$.

$$Z(n) = k \text{ if } \mathcal{Q}_n \ge k,\ \Sigma_{k,n} \le \tau_n \text{ and } \Sigma_{k+1,n} > \tau_n \text{ or if } \mathcal{Q}_n = k - 1 \text{ and } \Sigma_{k,n} \le \tau_n. \tag{39}$$

The process $\mathcal{Q}_n$ can now be represented as an SRS with driving sequence $\kappa_n = (\vec{\sigma}^n, \vec{\alpha}^n, \tau_n)$, that evolves according to

$$\mathcal{Q}_{n+1} = \mathcal{Q}_n + 1 - Z(n).$$

Define the σ-algebras $\mathcal{F}_n^{\sigma,\alpha,\tau} := \sigma\{\vec{\sigma}^n, \vec{\alpha}^n, \tau_n; k \le n\}$, $\mathcal{F}^{\sigma,\alpha,\tau} := \sigma\{\vec{\sigma}^n, \vec{\alpha}^n, \tau_n; k < \infty\}$. Let T^k be the k-step shift transformation of sets from the σ-algebra $\mathcal{F}^{\sigma,\alpha,\tau}$, so that

$$T\{\omega : (\sigma^j, \alpha^j, \tau_j) \in \mathbf{B}_j; j = 1, ..., k\} = \{\omega : (\sigma^{j+1}, \alpha^{j+1}, \tau_{j+1}) \in \mathbf{B}_j; j = 1, ..., k\} \tag{40}$$

and T^k is the k-th iteration of T.

Theorem 6
Assume that $E\tau_1 > E\sigma_1$. Then the process $\mathcal{Q}_n$ strongly couple-converges to a unique stationary ergodic regime.

Proof. The proof is based on exactly the same approach taken in the proof of Theorem 5. We begin by constructing a majorizing SRS $\overline{\mathcal{Q}}_n$ for $\mathcal{Q}_n$ on the same probability space. Roughly speaking, a majorant will look here like a process $\mathcal{Q}_n$ for the following auxiliary model: (i) the inter-retrial times occur at rate $C\theta = C/E\alpha_1^1$ when there are at least C customers in the orbit, and (ii) the retrials are stopped if there are less than C customers in the orbit. C is an arbitrary constant chosen such that $E\sigma_1 + E\alpha_1^1/C < E\tau_1$. We now define the process $\overline{\mathcal{Q}}_n$ precisely. Define first

$$\overline{Z}(n) = \min\left\{k : \sigma_1^n + \sum_{l=2}^{k}\left(\sigma_l^n + \frac{\alpha_l^n}{C}\right) > \tau_n\right\}. \tag{41}$$

Then we set

$$\begin{aligned} \overline{\mathcal{Q}}_0 &:= \mathcal{Q}_0, \\ \overline{\mathcal{Q}}_{n+1} &:= \max\{\overline{\mathcal{Q}}_n + 1 - \overline{Z}(n), C\}. \end{aligned} \tag{42}$$

Clearly $\mathcal{Q}_n \le \overline{\mathcal{Q}}_n$ for all n. Now, it easily follows from renewal arguments that

$$E\overline{Z}(n) \ge \frac{E\tau_1}{E\sigma_1 + E\alpha_1^1/C},$$

so we have $1 - E\overline{Z}(n) < 0$. It follows (see Example 1 in [6]) that there exists some integer n_0 and a set $\mathbf{B}_0 \in \mathcal{F}_0^{\sigma,\alpha,\tau}$ with $\mathbf{P}(\mathbf{B}_0) > 0$, such that $\overline{\mathcal{Q}}_n = C$ on the set $\mathbf{B}_n = T^n\mathbf{B}_0$ for all $n \ge n_0$ (see also Section 4, (36)).

Choose some integer m such that

$$E[\tau_m|\mathbf{B}_0] > 0. \tag{43}$$

The existence of such m follows since $\lim_{n\to\infty}\sum_{i=1}^n \tau_i = \infty$ a.s., so that $\lim_{n\to\infty}\sum_{i=1}^n E[\tau_i|\mathbf{B}_0] = \infty$; consequently, $\sum_{i=1}^n E[\tau_i|\mathbf{B}_0] > 0$ for some n. On the other hand, $E[\tau_m|\mathbf{B}_0] \le E\tau_1/\mathbf{P}(\mathbf{B}_0)$ is finite. From this and (43), we conclude that there exist some K_1 and K_2 such that

$$p_1 := \mathbf{P}(K_1 \le \tau_m \le K_2|\mathbf{B}_0) > 0. \tag{44}$$

Denote

$$D^m = \left\{\omega : \sum_{l=1}^{C+m+1}(\sigma_l^m(\omega) + \alpha_l^m(\omega)) < \tau_m(\omega)\right\}. \tag{45}$$

$A_n^m = T^n(D^m \cap \mathbf{B}_0) \in \mathcal{F}_{n+m}$ are renovating events. Indeed, for all $n > n_0$, $\mathcal{Q}_{n+m} = 0$ on A_n^m. We show that $\mathbf{P}(A_n^m) > 0$ by showing that $\mathbf{P}(D^m|\mathbf{B}_0) > 0$ (since we have already shown that $\mathbf{P}(\mathbf{B}_0) > 0$). Set $v := \min(1/E\alpha_1, 1/E\sigma_1)$. Then

$$\mathbf{P}(D^m|\mathbf{B}_0) \ge E\left[\left.\frac{(\tau_m v)^{2(C+m+1)}e^{-\tau_m v}}{[2(C+m+1)]!}\right|\mathbf{B}_0\right] \ge p_1\frac{(vK_1)^{2(C+m+1)}\,e^{-vK_2}}{[2(C+m+1)]!} > 0. \tag{46}$$

The proof is now concluded by applying Theorem 1.4 from [6]. ∎

6 Exponential boundness under i.i.d. assumptions for inter-arrivals and service times

We assume throughout this section:

- (i) The inter-arrival times are i.i.d.
- (ii) The service times $\{\sigma_n\}$ are i.i.d. and

$$\sup_v \mathbf{P}(\sigma_n > x + v | \sigma_n > v) \leq p_1(x) \quad \forall x, \tag{47}$$

  where $p_1(x_0) < 1$ for some x_0. (47) implies that $\sigma_n \leq_{st} \overline{\sigma}$, where

$$\mathbf{P}(\overline{\sigma} > v) < c_3 e^{-c_4 v} \tag{48}$$

  for all $v \geq 0$, and some positive constants c_3, c_4.

- (iii) The residual retrial times $\varphi(t)$ satisfy the following: For any ϵ there exists some probability distribution $G_\epsilon(\cdot)$ on $\mathbb{R}_+$ and some $N_1(\epsilon)$, such that on the set $Q(t) > N_1(\epsilon)$,

$$\mathbf{P}(\varphi(t) > v | \mathcal{F}(t)) \leq G_\epsilon(v) \quad \text{and} \quad \int_0^\infty G_\epsilon(v) dv < \epsilon \tag{49}$$

  and on the set $Q(t) \geq 1$

$$\mathbf{P}(\varphi(t) > v | \mathcal{F}(t)) \leq G(v) \in \mathcal{M}_C,$$

  for some probability distribution G, where the class $\mathcal{M}_C$ is defined in Section 3 (see (15)). Let β_ϵ and β be random variables with distributions G_ϵ and G respectively. Then we can take

$$\beta_\epsilon \leq_{st} \beta, \qquad E\beta_\epsilon < \epsilon. \tag{50}$$

- (iv) The sequences of retrial times, arrival times and service times are independent.

Remark 1
Note that condition (iii) is fairly general. Indeed, a sufficient condition for that is that the retrial times r_i^j of individual customers are *i.i.d.* and satisfy a condition of the type (47).

We focus on the processes Q_n and q_n, embedded at times S_n, which are defined in this section to be the instants at which the n-th service time begins. Let $X(n) = (X_1(n), X_2(n)) = (q_{nm}, Q_{nm})$. For some m, which we shall specify later, denote $\Delta(n,m) = L(X(n+1)) - L(X(n))$, where L is defined in (3).

Theorem 7
[5, 15] *Assume that there exists some $\vartheta > 0$ and some integer N_0, such that for all $v > N_0$, (i) $E\left(e^{\vartheta\Delta(n,m)} \middle| L(X(n)) > v\right) \leq \phi(\vartheta) < 1$, and (ii) $E\left(e^{\vartheta\Delta(n,m)} I\{L(X(n)) \leq N_0\}\right) \leq \psi(\vartheta) < \infty$. Then*

$$E\left(e^{\vartheta L(X(n))}\right) < \phi^n(\vartheta) E\left(e^{\vartheta L(X(0))}\right) + \frac{e^{\vartheta N_0}\psi(\vartheta)}{1 - \phi(\vartheta)} \tag{51}$$

Corollary 8
Assume that $E\sigma_1 < E\tau_1$. Then there exist some $N_0 < \infty$, some $\vartheta > 0$ for which $0 < \phi(\vartheta) < 1$ and $\psi(\vartheta) < \infty$, such that (51) holds.

Proof. The proof uses similar ideas to the one of Theorem 1. Choose some ϵ such that $0 < \epsilon < E\tau_1 - E\sigma_1$, and consider an arbitrary $N \geq N_1(\epsilon)$ where $N_1(\epsilon)$ satisfies (49). Define

- $\{\sigma_i^n\}$, $i = 1, ..., m, n = 0, 1, 2, ... :=$ i.i.d. random variables distributed like σ_1.
- $\{\beta_i^n\}$, $i = 1, ..., m, n = 0, 1, 2, ... :=$ i.i.d. random variables distributed like β_ϵ (defined in assumption (iii) above).

The sequences $\sigma_n, \sigma_i^n, \beta_i^n, \tau_n$ are assumed to be mutually independent. By the same arguments as in Section 2, we have for any $N > 0$, and for $L(X(n)) > K + N + m$

$$S_{n(m+1)} - S_{nm} \leq_{st} \sum_{i=1}^{m} (\sigma_i^n + \beta_i^n) =: \overline{v}_n^m, \qquad E[S_{n(m+1)} - S_{nm}] \leq E\overline{v}_n^m \leq m(E\sigma_1 + \epsilon). \tag{52}$$

As in Section 2, it follows from (52) that if $L\left(X(n)\right) > K + N + m$ then given $X(n)$,

$$\Delta(n, m) = L(X(n+1)) - L(X(n)) \leq_{st} \nu(\tau, \overline{v}_n^m) - m, \tag{53}$$

and everywhere

$$\Delta(n, m) = L(X(n+1)) - L(X(n)) \leq_{st} \sum_{i=1}^{m} \nu_i - m. \tag{54}$$

where ν_i are independent and distributed like $\nu(\tau, \sigma + \beta)$ (ν is defined in Sec. 2). Hence, equations (5)-(13) still hold, where in eqs. (5)-(8) ϵ replaces $(\theta N)^{-1}$, $\overline{v}_n^m$ replaces v_n^m; a, b, N and m are chosen appropriately so that (10) holds.

Now, it follows from the assumptions at the beginning of the section that both σ_1^i and β_1^i are in $\mathcal{M}_C$, defined in Section 3 (see (15)). This implies that both $\overline{v}_1^m$ and $\sigma + \beta$ are in $\mathcal{M}_C$. Applying Theorem 3, where $\zeta_i = \tau_i$ and $\xi = \overline{v}_1^m$, or $\xi = \sigma + \beta$, we conclude that $\Delta(n, m)$ are in $\mathcal{M}_C$, and hence there exists some $\tilde{\vartheta}$ such that $E\left(e^{\vartheta\Delta(n,m)}\right) =: \psi(\vartheta) < \infty$ holds for all $0 \leq \vartheta \leq \tilde{\vartheta}$. Hence (ii) of Theorem 7 holds. That (i) holds, follows from the fact that for $L(X(n)) > K + N + m$,

$$\mathbf{P}\left(e^{\vartheta\Delta(n,m)} \,\middle|\, X(n)\right) \leq E\left(e^{\vartheta[\nu(\tau, \overline{v}_n^m - m)]}\right) =: \mathcal{Z}(\vartheta),$$

where $\mathcal{Z}(0) = 0$ and

$$\left.\frac{d\mathcal{Z}(\vartheta)}{d\vartheta}\right|_{\vartheta=0} = E\nu(\tau, \overline{v}_n^m) - m < 0.$$

■

7 On Boundedness conditions for impatient customers

In this Section we consider the following generalization of the basic model introduced in Section 1, allowing for impatient customers. An arriving customer, say the i-th, is characterized not only by the sequence of retrial times r_i but also by a number γ_i; we consider the possibility that the customer leaves the system before it got served, if its cumulative time both in the primary queue and in orbit reaches the limit γ_i. The time γ_i can be defined in particular by a random number n_i of customer's retrials.

In all previous sections, we assumed that customers are patient, i.e. $\gamma_n = \infty$, and the stability condition was $E\sigma_1 < E\tau_1$. This condition is however not necessary in the case of impatient customers, i.e. $\gamma_n < \infty$, for which we present positive recurrence and boundedness conditions below.

First we consider the case $E\gamma_n < \infty$ assuming that $\{\tau_n, \gamma_n\}$ is a stationary ergodic sequence.

Theorem 9
The sequence $Q(t_k)$ ($t_k = \sum_{j=1}^{k-1} \tau_j$) has (independently of the behaviour of the server) a proper stationary majorant $\hat{Q}_k$: $Q(t_k) \leq \hat{Q}_k$ for all k.
Assume that $\{\tau_n\}, \{\gamma_n\}$ are independent sequences of i.i.d. r.v., and assume, for simplicity, that $Q(t_1) = 0, t_1 = 0$. Then

$$EQ(t_k) \leq E\hat{Q}_k \leq 2\left(1 + \frac{E\gamma_n}{\mathrm{med}\tau_{\mathrm{n}}}\right) \tag{55}$$

($\mathrm{med}\tau$ is the median of τ). The same inequality is true for $EQ(t)$

$$EQ(t) \leq 2\left(1 + \frac{E\gamma_n}{\mathrm{med}\tau_{\mathrm{n}}}\right) \tag{56}$$

Existence of a stationary majorant implies of course boundedness of $Q(t_k)$ in distribution, i.e. the property

$$\sup_k P(Q(t_k) > N) \to 0 \quad \text{as} \quad N \to \infty.$$

Proof. The first assersion of the Theorem follows from inequalities

$$Q(t_k) \leq \sum_{i=1}^{k} I(t_i + \gamma_i > t_k) \leq \sum_{i=1}^{\infty} I(\gamma_{k-i} > \tau_{k-i} + ... + \tau_{k-1}) =: \hat{Q}_k.$$

It was proved in [1] that if $E\gamma_n < \infty$ then $\hat{Q}_k$ is a proper r.v. (see also [14]).

Inequality (55) follows from relations

$$E\hat{Q}_k = \int P(\gamma_1 \in dv) \sum_{i=1}^{\infty} P(t_i < v) = \int P(\gamma_1 \in dv) H(v)$$

where $H(v)$ is a renewal function, satisfying inequality $H(v) \leq 2(1 + v/\mathrm{med}\tau_{\mathrm{n}})$ (see [1]). In order to prove (56) note that

$$Q(t) \leq \sum_{i=1}^{\infty} I(t_i \leq t < t_i + \gamma_i) =: \hat{Q}(t)$$

$$E\hat{Q}(t) = \int P(\gamma_1 \in dv) \sum_{i=1}^{\infty} P(t - v < t_i \leq t)$$

where

$$\sum_{i=1}^{\infty} P(t - v < t_i \leq t) = H(t) - H(t - v) \leq H(v).$$

■

Consider now the case $\gamma_n < \infty$ a.s. without any assumptions about $E\gamma_n$.

Define as before S_n to be the instant at which the n-th service begins, and $Q_n = Q(S_n - 0)$. Assume that the inter-arrival, service, retrial and patience times (γ_n) are independent i.i.d. sequences and that the retrial times are exponentially distributed with parameter θ.

Besides, we assume for simplicity $K = 0$ (this last assumption is not essential). Denote $\mathcal{T}(k) := \min\{j \geq 1 : X(k+j) \leq N\}$ (where $X(k) := Q_{km}$, and m and N are integers that will be determined later). According to the general definition in [5] we shall call the sequence $X(k)$ *positive recurrent* if for some $N \geq 0$

$$\sup_k E(\mathcal{T}(k) | X(k) \leq N) < \infty.$$

Let $\Delta(n) = X(n+1) - X(n)$. It was shown in [5] that if for some $\epsilon > 0$ and for all k

$$E(\Delta(k+n) | \mathcal{T}(k) > n) \leq -\epsilon, \tag{57}$$

$$E(X(k+1) | X(k) \leq N) < c < \infty, \tag{58}$$

then $X(k)$ is positive recurrent. The property of positive recurrence is closely related to the boundedness in distribution (see [5]).

Theorem 10
Assume that the renewal function $H(t)$ corresponding to $\tau_1, \tau_2, ...$ has a bounded density $h(t) \leq h$. If $E\sigma_1 < \infty$ then the process $X(n) = Q_{nm}$ for some fixed $m \geq 1$ satisfies conditions (57)-(58) and hence is positive recurrent.

Remark 2
The condition concerning the boundedness of the density $h(t)$ is purely technical, assumed so as to simplify the proof, and can be chipped, if $P(\tau_i = \sigma_i) < 1$.

Proof. Obviously $\Delta(k)$ does not exceed the number $e(k)$ of arrivals between S_{km} and $S_{(k+1)m}$. Since according to (13)

$$\sup_k Ee(k) < cm, \qquad c = const,$$

$\sup_k E(X(k+1)|X(k) \leq N) \leq N + Ee(k) < \infty$ (for any $N < \infty$), then the condition (58) is fulfilled. It remains to verify (57). Put for simplicity $k = 0$ (admitting that $X(k)$ are defined for negative k as well). Instead of (57) we shall estimate in fact $E(\Delta(n)|\mathcal{F}_n)$ on the set $\mathcal{T} > n$, where $\mathcal{F}_n$ is the σ-algebra generated by events, related to the prehistory of the process $X(k)$ until the moment S_{nm}; $\{\mathcal{T} > n\} \in \mathcal{F}_n$ ($\mathcal{T}$ stands for $\mathcal{T}(0)$). Note that $\{\mathcal{T} > n\} \subset \{X(n) > N\}$. In the sequel we don't use the symbol of conditional (with respect to $\mathcal{F}_n$) expectation, but shall estimate $E\Delta(n)$ for any preceeding trajectory $\{X(l), l \le n\}$(it will be the initial condition) with the property $X(n) > N$.

Denote:

$Q_l^{\rm old}, Q_0^{\rm old} > N, l \le m$, the number of customers who arrived until time S_{nm} and did not leave the system by "aging" and were not served until time S_{nm+l}.

$Q^{\rm new}$ the number of customers who arrived at the system during $(S_{nm}, S_{(n+1)m}]$ and did not leave the system by "aging" until time $S_{(n+1)m}$.

A- the number of customers who arrived at the system during $(S_{nm}, S_{(n+1)m}]$.

It is evident that

$$\Delta(n) \le Q_m^{\rm old} - Q_0^{\rm old} + Q^{\rm new}.$$

Let us estimate $EQ^{\rm new}$. The number of arrivals during $(S_{nm+l}, S_{(n+1)m}] \le_{st} \sum_{i=l+1}^m \nu_i =: A_{l,m}$ (ν_i was defined below (11)). Denote

$$A_{0,l} := A - A_{l,m}.$$

Let $B_{0,l}$ denote the number of arrivals among $A_{0,l}$ who did not leave the system by "aging" until moment $S_{(n+1)m}$, and define similarly $B_{l,m}$. $B_{0,l}$ satisfies inequalities

$$B_{0,l} \le \sum_{i=1}^{A_{0,l}} I(\gamma_i > \sum_{k=1}^{m-l} \sigma_k),$$

$$EB_{0,l} \le P\left(\gamma_1 > \sum_{k=1}^{m-l} \sigma_k\right) EA_{0,l} \le lE\nu_1 P(\gamma_1 > \sum_{k=1}^{m-l} \sigma_k).$$

For any given $\epsilon > 0$ we can choose $m-l$ large enough to make the right-hand side smaller than ϵm. As

$$EB_{l,m} \le E\Delta_{l,m} = (m-l)E\nu_1, \qquad Q^{\rm new} \le B_{0,l} + B_{l,m}$$

then

$$EQ^{\rm new} \le \epsilon m + (m-l)E\nu_1 \le 2\epsilon m$$

for $m - l \le \epsilon m / E\nu_1$. So $EQ^{\rm new} = o(m)$ as $m \to \infty$.

Consider now two possibilities:

1. $P(Q_m^{\rm old} \le N/2) \ge 1/2$.

In this case, since $\Delta^{\rm old} := Q_m^{\rm old} - Q_0^{\rm old} \le 0$ then

$$E\left(\Delta^{\rm old}; Q_m^{\rm old} \le \frac{N}{2}\right) \le -\frac{N}{2} P\left(Q_m^{\rm old} \le \frac{N}{2}\right) \le -\frac{N}{4},$$

$$E\Delta(n) \le E\left(\Delta^{\rm old}; Q_m^{\rm old} \le \frac{N}{2}\right) + o(m) \le -\frac{N}{4} + o(m) \le -\frac{N}{5}$$

for $N >> m$.

2. It remains to consider the case $P(Q_m^{\rm old} \le N/2) < 1/2$. Since

$$E\Delta(n) \le E\left(\Delta^{\rm old}; Q_m^{\rm old} > \frac{N}{2}\right) + o(m)$$

we need to estimate

$$E\left(\Delta^{\text{old}}; Q_m^{\text{old}} > N/2\right).$$

m customers were served during the period $(S_{nm}, S_{(n+1)m}]$. Each time more than $N/2-m$ "old" customers (who arrived before S_{nm}) and less than A "new" customers (who arrived after S_{nm}) claimed to be served. Denote by δ_i the r.v. which equals 1 if the i-th served customer was "new", and equals 0 if the i-th served customer was "old", so $\sum_{i=1}^m \delta_i$ is the number of "new", customers served during $(S_{nm}, S_{(n+1)m}]$. Let $S_{nm} + \lambda_i$ be the moment when i-th service was given, and $\mathcal{F}(\lambda_i)$ - the σ-algebra generated by the process describing the system until the time $S_{nm} + \lambda_i - 0$. For $Q_m^{\text{old}} > N/2$,

$$E(\delta_i|\mathcal{F}(\lambda_i)) \leq \frac{A(\lambda_i)\theta + h}{(N/2-m)\theta},$$

where $A(\lambda_i)$ is a number of customers who arrived during $(S_{nm}, S_{nm} + \lambda_i]$. It implies

$$\begin{aligned} E\left(\sum_{i=1}^m \delta_i; Q_m^{\text{old}} > \frac{N}{2}\right) &\leq \sum_{i=1}^m E\left(E(\delta_i|\mathcal{F}(\lambda_i)); Q_m^{\text{old}} > \frac{N}{2}\right) \\ &\leq \frac{EA\theta m + h}{(N/2-m)\theta} \leq \frac{\theta m^2 E\nu_1 + h}{(N/2-m)\theta} = o(m) \end{aligned} \qquad (59)$$

for $N >> m$, and since $Q_m^{\text{old}} - Q_0^{\text{old}} \leq -m + \sum_{i=1}^m \delta_i$, then

$$E\left(\Delta^{\text{old}}; Q_m^{\text{old}} > \frac{N}{2}\right) \leq -mP\left(Q_m^{\text{old}} > \frac{N}{2}\right) + o(m) < -\frac{m}{2} + o(m).$$

Due to (59) this gives us

$$E\Delta(n) \leq -\frac{m}{4}$$

for large enough m and N. Condition (57) is proved. ∎

References

[1] A. A. Borovkov, *Stochastic Processes in Queueing Theory*, Springer-Verlag, 1976.

[2] A. A. Borovkov, *Asymptotic Methods in Queueing Theory*, John Wiley & Sons, 1984.

[3] A. A. Borovkov, "Ergodicity and stability of multidimensional Markov chains", *SIAM Theory Probab. Appl.*, **35**, No. 3, pp. 542-546, 1989.

[4] A. A. Borovkov, *Ergodicity and stability of stochastic processes*, to appear (in Russian).

[5] A. A. Borovkov, "On boundedness of stochastic sequences and estimates of their moments", *Ann. Appl. Prob.*, to appear.

[6] A. A. Borovkov and S. G. Foss, "Stochastically recursive sequences and their generalizations", *Siberian Advances in Mathematics*, **2**, No. 1, pp. 16-81, 1992.

[7] J. W. Cohen, "Basic problems of telephone traffic theory and the influence of repeated calls", *Philips Telecommun. Rev.* **18**, No. 2, pp. 49-100, 1957.

[8] N. Deul, "Stationary conditions for multi-server queueing systems with repeated calls", *Elektr. Informationsverarbeit. Kybernet.* **16**, pp. 607-613, 1980.

[9] G. I. Falin, "On sufficient conditions for ergodicity of multichannel queueing systems with repeated calls", *Adv. Appl. Prob.* **16**, pp. 447-448, 1984.

[10] G. I. Falin, "The ergodicity of multilinear queueing systems with repeated calls", *Tekhnicheskaya Kibernetika*, No. 4, pp. 63-68, 1986.

[11] G. I. Falin, "Single-line repeated orders queueing systems", *Math. Operat. forsch. Statist. Ser. Optim.*, **5**, pp. 649-667, 1986.

[12] G. I. Falin, "On ergodicity of multichannel queueing systems with repeated calls", *Sov. J. Comput. Statistik, Optimization* **5**, pp. 649-667, 1986.

[13] G. I. Falin, "Estimation of error in approximation of countable Markov chains associated with models of repeated calls", *Vestnik Moscov. Univ. Ser. 1, Math. Mech.* **2**, pp. 12-15, 1987.

[14] G. I. Falin, "A survey of retrial queues", *QUESTA*, **5**, 127-167, 1990.

[15] B. Hajek, "Hitting-time and occupation-time bounds implied by drift analysis with applications", *Adv. Appl. Prob.*, **14**, pp. 502-525, 1982.

[16] T. Hanschke, "Explicit formulas for the characteristics of the M/M/2/2 queue with repeated attempts", *J. Appl. Probab.* **24**, pp. 486-494, 1987.

[17] J. Keilson, J. Cozzolino and H. Young, "A service system with unfilled requests repeated", *Operations Research*, **16**, pp. 1126-1137, 1968.

[18] H.-L. Liang and V. G. Kulkarni, "Stability condition for a single-server retrial queue", *Adv. Appl. Prob.* **25**, No. 3, pp. 690-701, 1993.

[19] F. M. Spieksma and R. L. Tweedie, "Strengthening ergodicity to geometric ergodicity of Markov chains", Technical report no TW-92-05, Univ. of Leiden, 1992, *to appear in Stoch. Models.*

[20] R. L. Tweedie, "Criteria for classifying general Markov Chains", *Adv. Appl. Prob.* **8**, pp. 737-771, 1976.