Article

Modified Erlang Loss System for Cognitive Wireless Networks

Evsey Morozov [1,2,3,*], Stepan Rogozin [1,2,*], Hung Q. Nguyen [4] and Tuan Phung-Duc [5,*]

1   Institute of Applied Mathematical Research, Karelian Research Centre RAS, 185910 Petrozavodsk, Russia
2   Institute of Mathematics and Information Technologies, Petrozavodsk State University,
     185910 Petrozavodsk, Russia
3   Moscow Center for Fundamental and Applied Mathematics, Moscow State University,
     119991 Moscow, Russia
4   Graduate School of Science and Technology, University of Tsukuba, 1-1-1 Tennodai,
     Tsukuba 305-8573, Ibaraki, Japan; s2030113@s.tsukuba.ac.jp
5   Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai,
     Tsukuba 305-8573, Ibaraki, Japan
*   Correspondence: emorozov@karelia.ru (E.M.); ppexa@mail.ru (S.R.); tuan@sk.tsukuba.ac.jp (T.P.-D.);
     Tel.: +81-(0)29-853-5003 (T.P.-D.)

Abstract: This paper considers a modified Erlang loss system for cognitive wireless networks and related applications. A primary user has pre-emptive priority over secondary users, and the primary customer is lost if upon arrival all the channels are used by other primary users. Secondary users cognitively use idle channels, and they can stay (either in an infinite buffer or in an orbit) in cases where idle channels are not available upon arrival or they are interrupted by primary users. While the infinite buffer model represents the case with zero sensing time, the infinite orbit model represents the case with positive sensing time. We obtain an explicit stability condition for the cases where arrival processes of primary users and secondary users follow Poisson processes, and their service times follow two distinct arbitrary distributions. The stability condition is insensitive to the service time distributions and implies the maximal throughout of secondary users. Moreover, we extend the stability analysis to the system with outgoing calls. For a special case of exponential service time distributions, we analyze the buffered system in depth to show the effect of parameters on the delay performance and the mean number of interruptions of secondary users. Our simulations for distributions rather than exponential reveal that the mean number of terminations for secondary users is less sensitive to the service time distribution of primary users.

Keywords: cognitive wireless network; regenerative analysis; stability condition; priority queue; Erlang loss system; classical retrials; outgoing calls

MSC: 90B22; 60K25; 68M20

1. Introduction

In recent years, Internet traffic has increased explosively due to the increased use of smartphones, tablet computers, etc. This causes a shortage problem of the wireless spectrum. Cognitive wireless is considered a promising solution to this problem [1–5]. For the recent development of cognitive radio networks, we refer to the survey paper by Ostovar et al. [6]. In wireless networks, secondary users (unlicensed users) are allowed to cognitively use the bandwidths that are originally allocated to primary users (licensed users). Secondary users should use the bandwidths in such a way that does not interfere with the primary users. In particular, secondary users can use the bandwidths only if primary users are not present. To this end, secondary users must be aware of the presence of primary users so as to evacuate upon the arrival of primary users. From this point of view, primary users have absolute priority over secondary users, meaning that the transmission of a secondary user might be interrupted by a primary user. In the first

system, we assume that interrupted secondary users evacuate to the head of the buffer and resume their transmission as soon as a channel is available. Then, we extend the stability analysis to a *partial retrial system* when the secondary users are evacuated to a virtual orbit and then attempt to occupy the channel again. Moreover, the analysis is then extended to the system with *outgoing calls*, which is initiated during the period that a server is idle.

Motivated by the above situation, we propose analyzing a multiserver (multichannel) queueing system with an infinite buffer for secondary users while primary users have absolute priority over secondary users and are lost if all channels are already occupied by other primary users. Under this assumption (and under Poisson inputs), from the viewpoint of primary users, the system of servers behaves as an Erlang loss system, while from that of secondary users, the system is either an infinite buffer model or a retrial model where secondary users are served when some servers are not occupied by primary users. In our model, the service times of primary and secondary customers follow two distinct arbitrary distributions.

A closely related model is the paper by Mitrani and Avi-Itzhak [7]. In this paper, the author considers an $M/M/N$ system where each server is subject to random breakdowns and repairs. The author analyzes the joint distribution of the queue length and the state of the servers using a generating function approach. The model in [7] can be considered as a model with N primary customers. In our model, primary customers arrive according to a Poisson process, implying that there is an infinite number of primary customers. Akutsu and Phung-Duc [8] examine a closely related Markovian model in which secondary customers first sense the channels before occupying them. The stability condition is conjectured and verified by simulation in [8] and further proved using a Lyapunov function approach [9]. Salameh et al. [10,11] consider models with a limited number of sensing secondary users. Other queueing models of cognitive radio networks could be found in [10,12–17]. In [10,13,16,18], the service time distributions of primary and secondary customers are either restricted to Markovian distributions (exponential or phase-type distributions) and/or the assumption that the number of active secondary users are finite. In [14,15,17,19–23], models with a single channel are investigated.

In this paper, we first relax these Markovian assumptions mentioned above by considering arbitrary distributions for the service time of primary users and secondary users. Under these assumptions, we are able to obtain an explicit stability condition. To the best of our knowledge, this is the first analytical result for cognitive radio network models with multiple channels. Moreover, we extend the analysis to a model with the retrial secondary customers and with outgoing calls. Next, assuming exponential service time distributions, we study the model in depth using the *matrix analytic method* [24].

The priority queues are prevalent among models for service differentiation in communication networks and service systems, and by this reason, applications of our model are not restricted to cognitive wireless networks [25].

The rest of this paper is organized as follows. In Section 2, we describe our model in detail with a focus on the regenerative setting. In Section 3, we present the stability analysis of the basic model and, as a by-product of our analysis, obtain the well-known stability condition of a buffered multiserver multiclass system (see Theorem 2 and Remark 4). In Section 4, we prove the stability condition for the system with retrial secondary customers, while in Section 5, we study a generalization of the latter system to capture the outgoing calls as well. Section 6 presents a detailed analysis for a special case with exponential distributions for which we are able to obtain performance measures. We also present numerical and simulation experiments to show insights into the performance of our system and the sensitivity of service time distribution, while Section 7 concludes our paper.

2. Description of the System

In this section, we consider the following basic modification of the Erlang system with two classes of customers with c identical servers, Poisson inputs with rates λ_i, and general independent and identically distributed (iid) service times $\{S_n^{(i)}, n \geq 1\}$ for class-i

customers, $i = 1, 2$. Class-1 customers have *pre-emptive priority* and are lost if upon arrival all servers are busy with class-1 customers, while class-2 (non-priority) customers stay in the system regardless of the state of the system and wait in the queue, if necessary, according to a *non-idling* FCFS (first-come-first-served) service discipline. Class-1 and class-2 customers correspond to primary and secondary customers, respectively, while a server is a channel in cognitive radio networks. Denote service rates $\mu_i = 1/\mathsf{E}S^{(i)}$, where $S^{(i)}$ is the (generic) service time of class-*i* customers $i = 1, 2$. We also denote by $\{t_n, n \geq 1\}$ the arrival instants of the superposed (Poisson) input with rate $\lambda = \lambda_1 + \lambda_2$, and iid exponential interarrival times $\{\tau_n = t_{n+1} - t_n\}$ with generic interarrival time τ. By pre-emptive-resume priority, a class-1 customer occupies a server busy by a class-2 customer, provided there are no idle servers upon his arrival.

We deduce the stability conditions of this system based on the regenerative approach. More precisely, we show that the basic processes describing the dynamics of the system are regenerative and then find conditions under which these processes are *positive recurrent* [26,27]. This approach, being quite intuitive and transparent, is a powerful tool of performance and stability analysis of a wide class of queueing processes including non-Markov ones as well. The main idea of the regenerative stability analysis is based on a characterization of the limiting *remaining regeneration time*. More precisely, if this time does not go to infinity in probability, then the mean regeneration period length is finite, and thus, the process is *positive recurrent* [27,28].

First, we describe the regenerative structure of the system. We denote by $Q_i(t)$ the number of class-*i* customers at time instant t^- and let $Q(t) = Q_1(t) + Q_2(t)$. Obviously, $Q_1(t) \leq c$. In addition, let $W_i(t)$ be the workload (remaining work) at instant t^- of class-*i* customers, $i = 1, 2$, and let $W(t) = W_1(t) + W_2(t)$. Denote by $S_i(t)$ the remaining service time in server i ($S_i(t) = 0$ if the server is idle), and let $R(t)$ be the set of servers occupied by class-1 customers at instant t (we put $R(t) = \varnothing$ if there are no such customers). Then

$$W_1(t) = \sum_{i \in R(t)} S_i(t), \; t \geq 0. \tag{1}$$

Denote $W_n = W(t_n)$, $\quad Q_n = Q(t_n)$, so, at the arrival instant of customer n, the remaining work in all servers equals W_n and the total number of customers equals Q_n. Note that $\{W_n = 0\} = \{Q_n = 0\}$, $\; n \geq 1$. Then, the *regeneration instants* of the processes $\{Q(t)\}$ and $\{W(t)\}$ are defined as follows

$$T_{n+1} = \min\{t_k > T_n : W_k = Q_k = 0\}, \; n \geq 0, \; T_0 := 0, \tag{2}$$

with generic regeneration period length T, which is distributed as any distance between two regeneration points, i.e., $T_{n+1} - T_n$. In what follows, we assume *zero initial state*: that is, the first primary customer arrives in the idle system at instant $t_1 = T_0 = 0$. In the latter case, the first regeneration cycle length is distributed as T: that is, the *stochastic equality* $T_1 =_{st} T$ holds. We call this regenerative process (in a continuous-time case) *positive recurrent* if $\mathsf{E}T < \infty$. Define the *remaining regeneration time* in the system at instant t as

$$T(t) = \min_k (T_k - t : T_k - t > 0), \; t \geq 0. \tag{3}$$

In the regenerative stability analysis developed below, the following basic result from the *renewal theory* is used [27,29]: if there is a non-random sequence of time instances $z_i \to \infty$, $i \to \infty$ and constants $\delta > 0$ and $D < \infty$ such that

$$\inf_i \mathsf{P}(T(z_i) \leq D) \geq \delta, \tag{4}$$

then $\mathsf{E}T < \infty$.

Remark 1. *The processes $\{Q_1(t)\}$, $\{W_1(t)\}$ describing class-1 customers are regenerative with regeneration instants*

$$T_{n+1}^{(1)} = \{t_n^{(1)} > T_n^{(1)} : Q_1(t_n^{(1)}) = 0\},\ n \geq 0\ (T_0^{(1)} = 0),$$

with generic regeneration period length $T^{(1)}$, where $\{t_n^{(1)}\}$ are the arrival instants of class-1 customers. In other words, $T_n^{(1)}$ is the nth arrival instant of a class-1 customer which meets no other class-1 customers in the system. Because $Q_1(t) \leq c$ the process $\{Q_1(t)\}$ is tight, and it is known that the processes $\{S_i(t)\}$ are tight as well; see [30]. In turn, it implies the tightness and hence the positive recurrence of the processes describing class-1 customers solely [27]: that is, $E\,T^{(1)} < \infty$.

3. Stability Analysis

Because the processes describing class-1 customers are positive recurrent and the inputs are Poisson, then there exists the stationary distribution $\{P_i\}$ where P_i is the probability that exactly *i* servers are busy by the class-1 customers: that is,

$$\lim_{t \to \infty} P(Q_1(t) = i) = P_i,\ i = 0, \dots, c. \tag{5}$$

In addition, denote $\rho_2 = \lambda_2/\mu_2$. In this section, we prove the following main result.

Theorem 1. *The system is positive recurrent, that is $ET < \infty$, if and only if the following condition*

$$\rho_2 + \sum_{i=1}^{c} iP_i < c, \tag{6}$$

holds.

Proof. In the interval of time $[0, t]$, denote the following: $\hat{V}_1(t)$ is the work of class-1 customers accepted by the system; $V_2(t)$ is the arrived work of class-2 customers; and $B(t)$ is the aggregated busy time of the servers, which equals the departed work in $[0, t]$. In addition, denote the total idle time $I(t) = \sum_{i=1}^{c} I_i(t)$, where

$$I_i(t) = \int_0^t 1(S_i(u) = 0)du,\ t \geq 0,$$

is the idle time of server *i* in $[0, t]$, so $B(t) = ct - I(t)$. ($1(\cdot)$ denotes the indicator function.) Then, the following balance equation holds true:

$$\hat{V}_1(t) + V_2(t) = W_1(t) + W_2(t) + B(t) = W_1(t) + W_2(t) + ct - I(t). \tag{7}$$

By the positive recurrence of 'class-1 processes', as $t \to \infty$ (see [27]):

$$EQ_1(t) = o(t),\ \ EW_1(t) = o(t)\ \text{and}\ Q_1(t) = o(t),\ \ W_1(t) = o(t)\ \text{w.p.1.} \tag{8}$$

Now, we apply a proof by contradiction and assume that, under condition (6),

$$Q_2(t) \Rightarrow \infty,\ t \to \infty, \tag{9}$$

that is, the queue size of class-2 customers increases infinitely *in probability*, or, for each fixed *k*,

$$\lim_{t \to \infty} P(Q_2(t) > k) = 1. \tag{10}$$

It then follows that the average aggregated idle time of all servers in the interval of time $[0, t]$ is

$$EI(t) = E\left[\sum_{i=1}^{c} \int_0^t 1(S_i(u) = 0)du\right] = \sum_{i=1}^{c} \int_0^t P(S_i(u) = 0)du.$$

Because the system has an infinite capacity buffer (for the awaiting class-2 customers), then the service discipline is non-idling: that is, servers cannot be idle while there are waiting customers. Then, it is easy to see that

$$P(Q_2(t) > c) \leq P(S_i(t) > 0), \quad i = 1, \ldots, c, \tag{11}$$

and, by (10) and (11), we obtain as $t \to \infty$,

$$P(S_i(t) = 0) = 1 - P(S_i(t) > 0) \leq 1 - P(Q_2(t) > c) \to 0.$$

It now easily follows that

$$\lim_{t \to \infty} \frac{1}{t} EI(t) = 0. \tag{12}$$

Denote by $A_2(t)$ the number of class-2 customers that arrived in the interval $[0, t]$. Then,

$$V_2(t) = \sum_{k=1}^{A_2(t)} S_k^{(2)}, \quad t \geq 0,$$

is a positive recurrent process with *regenerative increments* [31] in which each arrival instant of a class-2 customer is a regeneration instant. Denote by $\{\tau_n^{(2)}\}$ the iid exponential interarrival times between class-2 customers with the generic period $\tau^{(2)}$ with parameter λ_2. Because the process $\{V_2(t)\}$ has the increment $S^{(2)}$ over each interval $\tau^{(2)}$, then it follows (see Theorem 55 in [31]) that

$$\lim_{t \to \infty} \frac{EV_2(t)}{t} = \frac{ES^{(2)}}{E\tau^{(2)}} = \rho_2. \tag{13}$$

Denote by $B_1(t)$ the total time when servers are occupied by class-1 customers in the interval $[0, t]$. Then,

$$B_1(t) = \sum_{i=1}^{c} i \int_0^t 1(Q_1(u) = i)du.$$

By (5),

$$P_i = \lim_{t \to \infty} \frac{1}{t} \int_0^t P(Q_1(u) = i)du, \quad i = 1, \ldots, c,$$

and it now follows from (8) that

$$\lim_{t \to \infty} \frac{1}{t} E\hat{V}_1(t) = \sum_{i=1}^{c} i \lim_{t \to \infty} \frac{1}{t} \int_0^t P(Q_1(u) = i)du = \sum_{i=1}^{c} iP_i. \tag{14}$$

Now, collecting together results (8) and (12)–(14), we obtain from the balance Equation (7) that

$$\rho_2 + \sum_{i=1}^{c} iP_i = \lim_{t \to \infty} \frac{1}{t} EW_2(t) + c,$$

implying

$$\lim_{t \to \infty} \frac{1}{t} EW_2(t) = \sum_{i=1}^{c} i P_i + \rho_2 - c \geq 0,$$

or $\rho_2 + \sum_{i=1}^{c} i P_i \geq c$. This contradicts (6) and shows that assumption (9) is false, and thus, $Q_2(t) \not\to \infty$. Hence, there exist constants $N < \infty$, $\delta_0 > 0$ and a (non-random) sequence $u_i \to \infty$ such that (cf. (4))

$$\inf_i P(Q_2(u_i) \leq N) \geq \delta_0. \tag{15}$$

Note that

$$W_2(t) \leq \sum_{k=1}^{Q_2(t)} S_k^{(2)} + \sum_{k=1}^{c} S_k(t), \quad t \geq 0,$$

and recall that the remaining service times processes $\{S_k(t)\}$ are tight, $k = 1, \ldots, c$. A routine but tedious calculation confirms an intuitive result: that the bound (15) implies a corresponding lower bound for the workload process $\{W_2(t)\}$, namely

$$\inf_i P(W_2(u_i) \leq D_2) \geq \delta_2,$$

for some constants $D_2 < \infty$, $\delta_2 > 0$. (In particular, see Section 2.4 in [27].) Moreover, because the positive recurrent process $\{W_1(t)\}$ is also tight, then also

$$\inf_i P(W_2(u_i) \leq D_2, W_1(u_i) \leq D_1) \geq \delta_1,$$

for some constants D_1, $\delta_1 > 0$. Denote by $\tau(t)$ the remaining (exponential) interarrival time at instant t (in the superposed Poisson input with rate λ), so $P(\tau(t) \leq x) \geq e^{-\lambda x}$ for any $x \geq 0$. Recall definition (3) and note that for arbitrary fixed u_i (satisfying (15)) and a constant $L > D_1 + D_2$, the following lower bound for the remaining regeneration time holds:

$$
\begin{aligned}
P(T(u_i) \leq L) &\geq P\Big(W_2(u_i) \leq D_2, \, W_1(u_i) \leq D_1, \, L \geq \tau(u_i) > D_1 + D_2\Big) \\
&\geq \delta_1\big(e^{-\lambda(D_1+D_2)} - e^{-\lambda L}\big) > 0.
\end{aligned}
\tag{16}
$$

To explain this inequality, we note that in the event

$$\Big\{ W_2(u_i) \leq D_2, \, W_1(u_i) \leq D_1, \, L \geq \tau(u_i) > D_1 + D_2 \Big\},$$

the first customer arriving after instant u_i (at the instant $u_i + \tau(u_i) \leq u_i + L$) meets a completely idle system, and thus, a regeneration occurs. It explains why then $T(u_i) \leq L$. Because the bound (16) is uniform in u_i and i, we obtain $ET < \infty$ by (4). Thus, we conclude that (6) is the *sufficient stability condition*. $\square$

Now, we show that (6) is the *necessary stability condition* as well. Namely, we assume that the (initially empty) system is positive recurrent, $ET < \infty$. Note that

$$I(t) \geq \int_0^t 1(Q(u) = 0) du =: I_0(t),$$

where $I_0(t)$ is the aggregated time when all servers are *simultaneously free* in the interval $[0, t]$. Then, it follows from the theory of regenerative processes that with probability 1,

$$\lim_{t \to \infty} \frac{I(t)}{t} \geq \lim_{t \to \infty} \frac{I_0(t)}{t} = \frac{EI_0}{ET}, \tag{17}$$

where I_0 is the idle time, during a regeneration period, when all servers are simultaneously free [26,27,31]. Because τ is exponential and $ES^{(i)} < \infty$ then, for an arbitrary (fixed) $\delta > 0$ and some $\varepsilon_1 > 0$,

$$P(\tau > \delta + S^{(i)}) = \varepsilon_1, \quad i = 1, 2,$$

implying

$$EI_0 \geq E(I_0; \tau - S^{(i)} > \delta) \geq \delta\varepsilon_1 > 0,$$

and thus, the limit in (17) is strictly positive. It then immediately follows from (13), (14), (17) and from the balance equation (7), by dividing both sides by t and letting $t \to \infty$, that the inequality (6) holds. This implies that (6) is indeed the necessary stability condition as well. Thus, the proof is completed.

Remark 2. *Following [27], one can prove that under condition (6), the system is positive recurrent under arbitrary fixed values $W(0)$ and $Q(0)$. Under such a non-zero initial condition, the 1st regeneration period length T_1 in general has another distribution than the length T (obtained under zero initial state), and according to the definition of the positive recurrent system, the finiteness of T_1 with a probability of 1 (w.p.1) is additionally required. Such a generalization can be done as in [27] (Chapter 2); however, it is rather complicated, and we omit this analysis in this paper.*

Remark 3. *It is worth mentioning that to analyze the class-1 customers solely, we can treat the system as a loss $M/G/c/0$ system in which the stationary distribution $\{P_i\}$ can be found from the celebrated Erlang formula; see for instance [26]. Denote by ζ the stationary number of servers available for class-2 customers, that is (stochastically),*

$$\zeta =_{st} c - \sum_{i=1}^{c} i\mathbf{1}(Q_1 = i).$$

Then, $E\zeta = c - \sum_{i=1}^{c} iP_i$, and condition (6) rewritten as $\rho_2 < E\zeta$ has a clear probabilistic interpretation: the traffic intensity of class-2 customers must be less than the mean number of the available servers. This condition being intuitive, however, differs with the conventional stability conditions because the number of available servers ζ is now random.

As a by-product of our analysis, we obtain the following well-known stability criterion of a multiclass system with the *infinite capacity* buffer for all awaiting customers.

Theorem 2. *The initially empty system without losses is positive recurrent if and only if the condition $\rho_1 + \rho_2 < c$ holds.*

The proof of this statement is quite similar to that given in the previous section. The only difference is that now, we replace $\hat{V}_1(t)$ by the full work $V_1(t)$ generated by all class-1 customers which arrive in the interval of time $[0, t]$ and satisfies $\lim_{t \to \infty} EV_1(t)/t = \rho_1$.

Remark 4. *Using previous arguments and evident notation, it is straightforward to prove the following well-known result: condition $\sum_{i=1}^{N} \rho_i < c$ is the stability criterion of the N-class system with no losses and with arbitrary work-conserving discipline including various priority policies, where the number of classes N is arbitrary.*

4. A System with Class-2 Retrial Customers

Now, we consider a more general cognitive radio system with primary and secondary users, and c frequency channels. A new feature of this system is that the secondary users now must sense the channels before they start using a channel. As a result, the system turns out to be a combination of a priority system and a retrial system. Under exponential assumptions, this system has been considered in the paper [9]. In this section, using previous notation and results, we extend the stability analysis to general service times distributions. We assume that the 2nd class customers join an orbit and then attempt to

occupy the server after a random time. Thus, the model becomes a (partially) retrial system. It is intuitive that under the classic retrial policy, the stability condition remains the same as in the original system considered in Section 2. Recall that T is a regeneration cycle length in continuous time, and a regeneration happens when a new customer meets the system idle. However, unlike the approach developed in Section 3 for continuous-time processes, now, we will use the processes embedded at the *departure instances* of the customers.

Theorem 3. *If condition (6) holds, then the (initially idle) system under consideration is positive recurrent, $ET < \infty$.*

Proof. Let $\{d_n, n \geq 1\}$ be the departure instances of the served customers leaving the system. As in Section 3, denote, for the time interval $[0, d_n]$, the received work by $\hat{V}_1(d_n) = \hat{V}_{1,n}$, $V_2(d_n) = V_{2,n}$, the aggregated busy time of all servers by $B(d_n) = B_n$, the total idle time of servers by $I(d_n) = I_n$, and denote by $W_{i,n}$ the remaining work of class-i customers at instant d_n, $n \geq 1; i = 1, 2$. Then, the balance of Equation (7) transforms to

$$\hat{V}_{1,n} + V_{2,n} = W_{1,n} + W_{2,n} + B_n = W_{1,n} + W_{2,n} + cd_n - I_n, \quad n \geq 1. \tag{18}$$

For the further analysis, we redefine regeneration instances of the system as follows

$$\theta_{n+1} = \inf(k > \theta_n : W_{1,k} + W_{2,k} = 0), \quad n \geq 0; \quad \theta_0 := 0, \tag{19}$$

which *count the customers meeting system idle* and starting new regeneration cycles, and we denote by θ the generic regeneration cycle length. The quantity θ is connected with the continuous-time length T by the stochastic equality $T =_{st} \sum_{i=1}^{\theta} \tau_i$ where τ_i is the ith interarrival time within a cycle. Note that by the positive recurrence of 'class-1 processes', $W_{1,n} = o(d_n)$, $W_{2,n} = o(d_n)$, $d_n \to \infty$, cf. (8). Denote also $Q_2(d_n) = Q_{2,n}$ and assume that (cf. (9))

$$Q_{2,n} \Rightarrow \infty, \quad n \to \infty. \tag{20}$$

Denote by $\Delta_k = I(d_{k+1}) - I(d_k)$ the idle time of all servers between the kth and $(k+1)$th departures. An important observation is that provided $Q_{2,k} \geq n$, the mean idle time of all servers after the kth departure is upper bounded by the constant

$$C_n := \frac{c}{\lambda + n\gamma_0},$$

and $C_n \to 0$ as $n \to \infty$. (Recall that γ_0 is the minimal retrial rate.) This shows that if $Q_{2,k} \geq n$, then $E\Delta_k \leq C_n$ can be done arbitrarily small for n large enough. It is evident that

$$I_n = \sum_{k=1}^{n-1} \Delta_k, \quad n \geq 1,$$

where by construction, $I_1 = 0$, and the (uniform) upper bound $\sup_k E\Delta_k \leq c/\lambda$ holds regardless of the states of the orbits. Now, we prove that, under assumption (20),

$$\lim_{n \to \infty} \frac{EI_n}{n} = 0. \tag{21}$$

For an arbitrary $\varepsilon > 0$, take n_1 such that for all $n \geq n_1$,

$$C_n \leq \varepsilon, \tag{22}$$

and, by (20), for an arbitrary $D > 0$, take n_2 such that, for all $n \geq n_2$:

$$P(Q_2(d_n) \leq D) \leq \varepsilon. \tag{23}$$

Now, take $n_0 = \max(n_1, n_2)$ and let $n > n_0$. Then, it follows that

$$\mathrm{E}I_n = \mathrm{E}\sum_{k=1}^{n_0}\Delta_k + \mathrm{E}\Big[\sum_{k=n_0+1}^{n}\Delta_k\Big]. \tag{24}$$

Using (20), (22) and (23), we write the 2nd summand in (24) as

$$
\begin{aligned}
\mathrm{E}\Big[\sum_{k=n_0+1}^{n}\Delta_k\Big] &= \Big[\sum_{k=n_0+1}^{n}\mathrm{E}(\Delta_k|Q_{2,k} > D)\mathrm{P}(Q_{2,k} > D)\Big] \\
&+ \Big[\sum_{k=n_0+1}^{n}\mathrm{E}(\Delta_k|Q_{2,k} \leq D)\mathrm{P}(Q_{2,k} \leq D)\Big] \\
&\leq \varepsilon(n - n_0)(1 + c/\lambda).
\end{aligned}
\tag{25}
$$

Because ε is arbitrary, then (21) follows.

We assume that if the nth customer entering server j belongs to class i, then we assign the service time $S_{i,n}^{(j)}$ from the corresponding iid sequence $\{S_{i,n}^{(j)}\}$ initially intended for this class of customers, $i = 1, 2$; $j = 1, \dots, c$. (In other words, we omit the 'intermediate', not used, elements of this sequence.) Now, we consider the 'minimal' service times realized by all servers,

$$S_n^{(0)} = \min\{S_{1,n}^{(1)}, S_{2,n}^{(1)}, S_{1,n}^{(2)}, S_{2,n}^{(2)}, \dots, S_{1,n}^{(c)}, S_{2,n}^{(c)}\}, \ n \geq 1. \tag{26}$$

These times constitute an iid sequence $\{S_n^{(0)}\}$ with the generic element $S^{(0)}$. Now, we define the following instances

$$\widehat{d}_0 = 0, \ \widehat{d}_n = \sum_{k=1}^{n}S_k^{(0)}, \ n \geq 1.$$

An important observation is that $\widehat{d}_{\lfloor n/c\rfloor} \leq d_n$, $n \geq 1$. (See also Figure 1; $\lfloor x\rfloor$ denotes the maximal integer $\leq x$.) It is well-known that the convergence in mean (21) implies convergence in probability $I_n/n \Rightarrow 0$. In turn, then there exists a subsequence $n_k \to \infty$, $k \to \infty$, such that w.p.1 [27]

$$\lim_{k\to\infty}\frac{I_{n_k}}{n_k} = 0. \tag{27}$$

On the other hand, because $\widehat{d}_{\lfloor n_k/c\rfloor} \to \infty$, $k \to \infty$, it follows from the renewal theory that w.p.1

$$\lim_{k\to\infty}\frac{n_k}{\widehat{d}_{\lfloor n_k/c\rfloor}} = \frac{c}{\mathrm{E}S^{(0)}}. \tag{28}$$

Now, we return to the balance Equation (18) written for the subsequence $\{n_k\}$:

$$W_{1,n_k} + W_{2,n_k} = \hat{V}_{1,n_k} + V_{2,n_k} - cd_{n_k} + I_{n_k}, \ n \geq 1. \tag{29}$$

Thus, by (27), (28), as $k \to \infty$,

$$\frac{I_{n_k}}{d_{n_k}} = \frac{I_{n_k}}{n_k}\frac{n_k}{d_{n_k}} \leq \frac{I_{n_k}}{n_k}\frac{n_k}{\widehat{d}_{\lfloor n_k/c\rfloor}} \to 0.$$
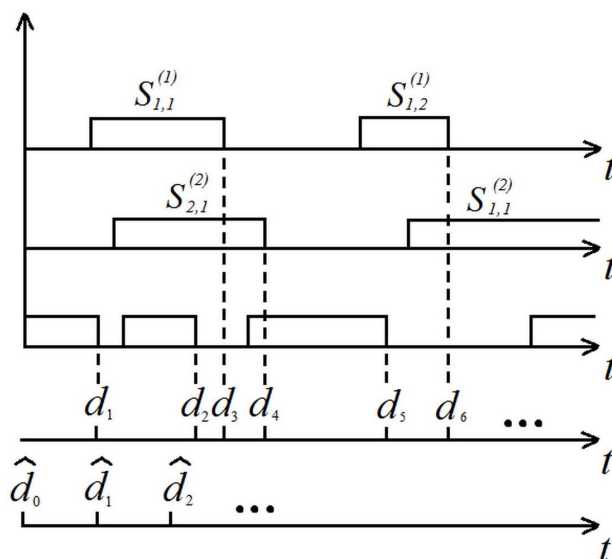
Figure 1. Construction $\widehat{d}_n$ in the system with 3 servers showing that $\widehat{d}_{\lfloor n/c \rfloor} \leq d_n$.

Now, we divide both sides of Equation (29) by d_{n_k} and let $k \to \infty$. Note that

$$\lim_{k \to \infty} W_{1,n_k}/d_{n_k} = 0, \quad \lim_{k \to \infty} W_{2,n_k}/d_{n_k} \geq 0.$$

(The latter limit exists because all other limits in (29) exist.) Recall (14), and then, it follows from (29) that

$$\rho_2 + \sum_{i=1}^{c} i P_i \geq c, \tag{30}$$

implying a contradiction with the assumption (20), that is $Q_{2,n} \not\to \infty$. We note that a 'regeneration' condition

$$\min_{1 \leq i \leq c} P(\tau > S^{(i)}) > 0$$

in this system holds automatically because the input process is Poisson. Define the remaining regeneration time,

$$\theta(n) = \inf_{i}(\theta_i - n : \theta_i - n > 0)$$

counting, at the departure instant d_n, the number of the *remaining departures* until the regeneration cycle ends; see (19). Then, using the regeneration condition and the arguments following inequality (15), we find that $\theta(n_k) \not\to \infty$, $k \to \infty$, which in turn implies the mean number of arrivals/departures within a regeneration cycle $E\theta < \infty$ [27]. Finally, by Wald's identity, $ET = E\theta \, E\tau < \infty$ (τ is exponential with parameter λ) and the proof of the Theorem 3 hereby is completed.　□

5. A System with the Outgoing Calls

Now, we consider a generalization of the partial retrial system considered in Section 4 in which the idle servers generate the so-called 'outgoing calls'. More precisely, we assume that server i generates class-i outgoing calls with rate ν_i when the server is idle. Denote the total rate by $\nu = \sum_{i=1}^{c} \nu_i$. The service times $\{Z_n^{(i)}, n \geq 1\}$ of class-i calls are assumed to be iid with mean $EZ^{(i)} < \infty$, $i = 1, \ldots, c$. It is assumed that service of these customers can not be interrupted however, as in the original system, class-1 customers can still interrupt class-2 customers. It is easy to check that (19) are the regeneration instants of the new system, and let θ be the corresponding regeneration cycle length. We now prove an intuitive result that the stability condition of the system with outgoing calls remains the same as in the original system. Namely, the following statement takes place.

Theorem 4. *If condition (6) holds in the system with outgoing calls, then $E\theta < \infty$.*

Proof. Keeping the main notation used in Equation (18) and denoting by Z_n the total work generated by the outgoing calls in the interval $[0, d_n]$, we obtain the following equation:

$$\hat{V}_{1,n} + V_{2,n} + Z_n = W_{1,n} + W_{2,n} + cd_n - I_n, \quad n \geq 1. \tag{31}$$

The outgoing calls of type i (in server i) follow a Poisson input (with rate ν_i) when the server is not occupied by the incoming customers. Let $N_n = \sum_{i=1}^{c} N_n^{(i)}$ where $N_n^{(i)}$ is the number of the outgoing calls in server i during the idle time $I_n^{(i)}$ of server i in the interval $[0, d_n]$. By construction of the total idle time process $I_n = \sum_i I_n^{(i)}$, the process N_n is dominated by a Poisson process $\hat{N}_n$ with rate ν, which counts the number of renewals in the time interval $[0, I_n]$, that is $N_n \leq_{st} \hat{N}_n$, $n \geq 1$. It is because the original process N_n, over the time period $[0, I_n]$ (obtained by a 'coupling together' the sub-intervals in which servers are idle) in general has a rate less than the 'maximal' possible rate ν. Now, we introduce the iid sequence $\hat{Z}_j = \sum_{i=1}^{c} Z_j^{(i)}$, $j \geq 1$ (with the generic element $\hat{Z}$) and consider the corresponding renewal process

$$\hat{Z}_n = \sum_{j=1}^{\hat{N}_n} \hat{Z}_j, \quad n \geq 1.$$

The new process 'dominates' the workload process generated by the outgoing calls in the original system, and the following (stochastic) inequality holds:

$$\hat{Z}_n \geq_{st} Z_n, \quad n \geq 1. \tag{32}$$

To see this, we couple together all idle times of servers in $[0, d_n]$ and realize the renewal process $\hat{Z}_n$ in the new time scale counting only the idle periods. Moreover, unlike the original process Z_n, this new renewal process does not have 'gaps' and each renewal interval is not less than the corresponding service time of an outgoing call in the original system. Now, we consider the iid 'shortest' service times realized by all servers (including the service times of the outgoing calls):

$$S_n^{(0)} = \min\{S_{1,n}^{(1)}, S_{2,n}^{(1)}, Z_n^{(1)}, S_{1,n}^{(2)}, S_{2,n}^{(2)}, Z_n^{(2)}, \ldots, S_{1,n}^{(c)}, S_{2,n}^{(c)}, Z_n^{(c)}\},$$

and define a renewal process generated by these times:

$$\widehat{d}_n = \sum_{k=1}^{n} S_k^{(0)}, \quad n \geq 1.$$

Using the same procedure as in Section 4, we obtain that $\widehat{d}_{\lfloor n/c \rfloor} \leq d_n$. (See a comment, preceding formula (26), and Figure 1). Assume that relation (20) holds, which implies relations (21) and (27). Now, we rewrite the balance Equation (31) for the corresponding subsequence $\{n_k\}$:

$$W_{1,n_k} + W_{2,n_k} = Z_{n_k} + \hat{V}_{1,n_k} + V_{2,n_k} - cd_{n_k} + I_{n_k}, \quad k \geq 1. \tag{33}$$

First, assume that $I_{n_k} \to \infty$ as $k \to \infty$. Then, because by the renewal theory, w.p.1,

$$\lim_{k \to \infty} \frac{\hat{Z}_{n_k}}{\hat{N}_{n_k}} = E\hat{Z} \leq c \max_i EZ^{(i)} < \infty, \quad \lim_{k \to \infty} \frac{\hat{N}_{n_k}}{I_{n_k}} = \nu < \infty,$$

and from (27) and (32), we obtain

$$\frac{Z_{n_k}}{d_{n_k}} \le \frac{\widehat{Z}_{n_k}}{\widehat{d}_{\lfloor n_k/c \rfloor}} = \frac{\widehat{Z}_{n_k}}{\widehat{N}_{n_k}} \frac{\widehat{N}_{n_k}}{I_{n_k}} \frac{I_{n_k}}{\widehat{d}_{\lfloor n_k/c \rfloor}} \to 0, \ \ k \to \infty.$$

Because there can be only a finite number of the outgoing calls in any finite interval then, on the event $\{\lim_{k \to \infty} I_{n_k} < \infty\}$, we obtain $Z_{n_k}/d_{n_k} \to 0$ as well. Now, we divide both sides of (33) by d_{n_k} and let $k \to \infty$. As above, because $W_{1,n_k} + W_{2,n_k} \ge 0$, this leads to the inequality (30) contradicting (20). The rest of the proof is now follows as in Section 3 (see the proof of Theorem 1 following relation (15)). The only new element in this proof is that, realizing the unloading of the system, we must take into account the probability of the absence of the outgoing calls; for details, see [32]. $\square$

6. Performance Analysis of the System with Exponential Service Time Distributions

6.1. Quasi-Birth-and-Death (QBD) Process

In this section, we examine more closely a special case of the system described in Section 2, in which service times of both customer classes follow exponential distributions, with parameter μ_i for class-i customers.

In this pure Markovian case, for stability analysis, it is possible by applying the Matrix analytic method [24], and this alternative proof of stability is instructive. Moreover, in this setting, we can calculate the stationary distribution of the basic Markov process and as a result obtain the corresponding stationary performance indexes. In this setting, the process $\{(Q_1(t), Q_2(t)), t \ge 0\}$ is a continuous-time Markov Chain with the state space $\mathbb{S} = \{(i,j) \in \{0, 1, ..., c\} \times \mathbb{N}\}$. Grouping states into levels according to their values of $Q_2(t)$, the system can be formulated as a Quasi-Birth-and-Death process (QBD) whose infinitesimal generator Q is expressed as follows

$$Q = \begin{pmatrix} B_0 & C & O & O & O & ... & ... & ... & ... \\ A_1 & B_1 & C & O & O & ... & \vdots & \vdots & \vdots \\ O & A_2 & B_2 & C & O & ... & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ O & O & ... & A_c & B_c & C & O & ... & \vdots \\ O & O & ... & O & A_c & B_c & C & O & ... \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \end{pmatrix},$$

where O is a $(c+1) \times (c+1)$ zero matrix, and A_n, B_n, C are $(c+1) \times (c+1)$ block matrices given by $C = \text{diag}(\lambda_2)$,

$$A_n = \text{diag}(a_{n,0}, a_{n,1}, \ldots, a_{n,c}),$$

where $a_{n,i} = \min(c - i, n)\mu_2$, for $n \le c$, and $A_n = A_c$, for $n > c$;

$$B_n = \begin{pmatrix} b_{n,0} & \lambda_1 & 0 & ... & 0 & 0 & ... & 0 & 0 \\ \mu_1 & b_{n,1} & \lambda_1 & ... & 0 & 0 & ... & 0 & 0 \\ 0 & 2\mu_1 & b_{n,2} & \ddots & 0 & 0 & ... & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & ... & b_{n,c-n} & \lambda_1 & ... & 0 & 0 \\ 0 & 0 & 0 & ... & (n-1)\mu_1 & b_{n,c-(n-1)} & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & ... & 0 & 0 & ... & b_{n,c-1} & \lambda_1 \\ 0 & 0 & 0 & ... & 0 & 0 & ... & c\mu_1 & -(\lambda_2 + c\mu_1) \end{pmatrix},$$

for $n \leq c$, and $B_n = B_c$ for $n > c$. In case $n \leq c$, the diagonal elements of B_n are given by

$$b_{n,i} = -(\lambda_1 + \lambda_2 + i\mu_1 + a_{n,i}).$$

Denote by $\boldsymbol{\eta} = (\eta_0, \eta_1, ..., \eta_c)$ the row vector representing the stationary distribution of the infinitesimal generator $Q^* = A_c + B_c + C$. It is easy to see that

$$\boldsymbol{\eta} = \left(\eta_0, \rho_1 \eta_0, \frac{\rho_1^2}{2!} \eta_0, ..., \frac{\rho_1^i}{i!} \eta_0, ..., \frac{\rho_1^c}{c!} \eta_0 \right),$$

where

$$\eta_0 = \Big[\sum_{i=0}^{c} \frac{\rho_1^i}{i!} \Big]^{-1} \text{ and } \rho_1 = \frac{\lambda_1}{\mu_1}.$$

The QBD process is ergodic if and only if the following condition holds [24]

$$\boldsymbol{\eta} C \boldsymbol{e} < \boldsymbol{\eta} A_c \boldsymbol{e},$$

where $\boldsymbol{e}$ denotes the $(c+1)$-dimension column vector of ones. This stability condition can be written as

$$\lambda_2 < \frac{\sum_{i=0}^{c-1} (c-i) \frac{\rho_1^i}{i!}}{\eta_0} \mu_2, \tag{34}$$

and is identical to the stability condition (6) in the general case.

In what follows, we consider the system under the stability condition. Now, let $\pi(i,j) = P(Q_1(t) = i, Q_2(t) = j)$ denote the stationary probability of the Markov chain, and let

$$\boldsymbol{\pi}_j = (\pi(0,j), \pi(1,j), ..., \pi(c,j)) \text{ for } j \in \mathbb{N}.$$

According to matrix analytic method [24,33], we have

$$\boldsymbol{\pi}_j = \boldsymbol{\pi}_c R^{j-c}, \quad j \geq c, \qquad \boldsymbol{\pi}_j = \boldsymbol{\pi}_{j-1} R^{(j)}, \quad j = c, c-1, \dots, 1,$$

where R is the minimal non-negative solution of $C + RB_c + R^2 A_c = \boldsymbol{0}$, and

$$R^{(j)} = -C(B_j + R^{(j+1)} A_{j+1})^{-1}, \quad j = c-1, c-2, \dots, 1,$$

given that $R^{(c)} = R$, and R is numerically computed using algorithms in [24]. Finally, $\boldsymbol{\pi}_0$ is the unique solution of the following equations

$$\boldsymbol{\pi}_0 (B_0 + R^{(1)} A_1) = \boldsymbol{0},$$

$$\boldsymbol{\pi}_0 \left(I + \sum_{i=1}^{c-1} \prod_{j=1}^{i} R^{(j)} + \left(\prod_{j=1}^{c} R^{(j)} \right) (I - R)^{-1} \right) \boldsymbol{e} = 1,$$

where I denotes the $(c+1) \times (c+1)$ identity matrix.

Let EW_q denote the average waiting time of class-2 customers; then, we have

$$
\begin{aligned}
EW_q &= \frac{\sum_{j=0}^{\infty} \sum_{i=0}^{c} \max\{0, i+j-c\} \pi(i,j)}{\lambda_2} \\
&= \frac{\sum_{j=1}^{c-1} \sum_{i=c-j+1}^{c} (i+j-c) \pi(i,j) + \boldsymbol{\pi}_c (I-R)^{-2} R \boldsymbol{e} + \boldsymbol{\pi}_c (I-R)^{-1} \boldsymbol{f}}{\lambda_2},
\end{aligned}
$$

where $\boldsymbol{f} = (0, 1, 2, ..., c-1, c)^T$, due to Little's law.

Moreover, the average number of class-2 customers in the system is given by

$$
\begin{aligned}
EQ_2 &= \sum_{j=0}^{\infty} j \pi_j \boldsymbol{e} \\
&= \sum_{j=1}^{c-1} \pi_j \boldsymbol{e} + c \pi_c (I - R)^{-1} \boldsymbol{e} + \pi_c (I - R)^{-2} \boldsymbol{e}.
\end{aligned}
$$

There are other performance measures which describe the QoS of such a system. For instance, if a class-2 customer is interrupted by a class-1 customer, we call it a *termination event*. Denote the set

$$
\mathbb{S}^* = \{(i, j) : i + j \geq c, \, i \leq c - 1\},
$$

which contains the states when all servers are occupied and there is at least one class-2 customer occupying a server. Then, the average number of termination events per class-2 customer EN_T can be calculated by the regenerative method as follows. Denote by $A_i(t)$ the number of class-i customers arriving in the interval $[0, t]$, $i = 1, 2$, and recall that, as $\to \infty$, the weak convergence

$$
Q(t) = Q_1(t) + Q_2(t) \Rightarrow Q := Q_1 + Q_2,
$$

holds. Then, using the basic asymptotic results for positive recurrent regenerative processes [26,27], we obtain EN_T as the following (w.p.1) limit of the fraction of class-1 arrivals interrupting class-2 customers:

$$
\begin{aligned}
EN_T &= \lim_{t \to \infty} \frac{1}{A_2(t)} \sum_{n=1}^{A_1(t)} \mathbf{1}(Q(t_n^{(1)}) \geq c, \, Q_1(t_n^{(1)}) \leq c - 1) \\
&= \lim_{t \to \infty} \frac{A_1(t)}{t} \frac{1}{A_1(t)} \sum_{n=1}^{A_1(t)} \mathbf{1}(Q(t_n^{(1)}) \geq c, \, Q_1(t_n^{(1)}) \leq c - 1) \cdot \lim_{t \to \infty} \frac{t}{A_2(t)} \\
&= \lambda_1 P(Q \geq c, \, Q_1 \leq c - 1) \frac{1}{\lambda_2} = \frac{\lambda_1}{\lambda_2} \sum_{(i,j) \in \mathbb{S}^*} P(Q_1 + Q_2 = i + j, \, Q_2 = j) \\
&= \frac{\lambda_1}{\lambda_2} \sum_{(i,j) \in \mathbb{S}^*} \pi(i, j),
\end{aligned} \tag{35}
$$

where we also use the property PASTA [26] to apply the equality

$$
\pi(i, j) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \sum_{(i,j) \in \mathbb{S}^*} \mathbf{1}(Q(t_k^{(1)}) = i + j, \, Q_2(t_k^{(1)}) = j).
$$

The equality (35) written as

$$
\lambda_2 EN_T = \lambda_1 \sum_{(i,j) \in \mathbb{S}^*} \pi(i, j)
$$

is intuitive and establishes a balance between the rate of the *interrupted* class-2 customers and the rate of *interrupting* class-1 customers.

Remark 5. *It is worth mentioning that relation (35) holds also for general service time distribution of any class of customers; however, in this case, the stationary distribution $\{\pi(i, j)\}$ of the (non-Markovian) process $\{(Q_1(t), Q_2(t)), \, t \geq 0\}$ is not analytically available.*

6.2. Simulations and Numerical Insights

In this section, we present some numerical examples of the results obtained by the matrix analytic method presented in Section 6. In our experiment, for fixed $\mu_1 = 4, \mu_2 = 20$

and values $c = 2, 5$, we show the changes in the values of some performance measures against λ_1 and λ_2. Under the same settings, we also carry out simulations and obtain the same results as those obtained by the matrix analytic method. Furthermore, to show the sensitivity of the service time distribution of primary users, we also compare the results by matrix analytic methods with those by simulations where the service time of class-1 customers follows Erlang distributions with the shape parameter $r = 5, 10$.

The scale parameter is chosen such that the mean value remains the same as in the case of exponential distributions. The duration for all experimental simulations is set at 10^6 time units, which is adequate for the simulation results to converge to their corresponding numerical results. The simulation results in all the figures are represented by the points marked with notation sim, without which the results are understood to be obtained from numerical calculations. The simulation results with service time of class-1 customers following Erlang distributions are marked with the abbreviation Erl. in the legend.

Figures 2 and 3 compare the average waiting times of class-2 customers as λ_1, λ_2 and c vary. It can be seen that the average amount of time that class-2 customers spend in the system ($\mathrm{E}W_q$) is larger when customers of either class arrive more frequently or when there are fewer servers. We observe that $\mathrm{E}W_q$ with exponential service time distribution for class-1 is largest while $\mathrm{E}W_q$ with $r = 5$ is larger than that with $r = 10$. This indicates that $\mathrm{E}W_q$ increases with the increase in the variance of service time of class-1 customers.
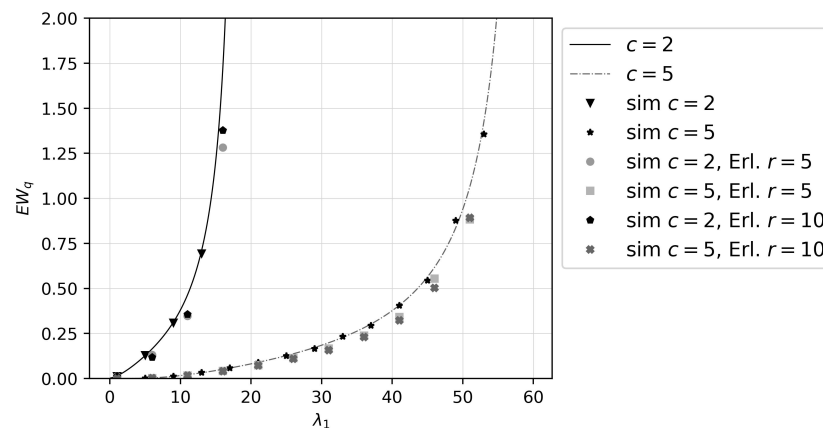

Figure 2. Average waiting time of class-2 customers against arrival rate of class-1 customers ($\lambda_2 = 8$).
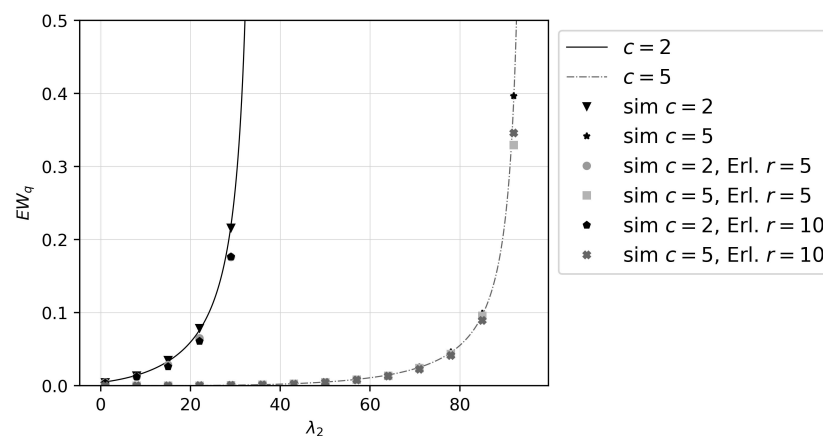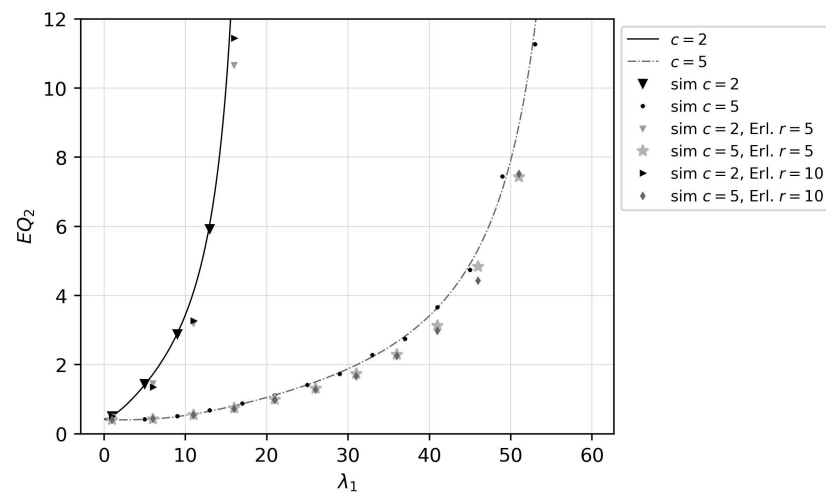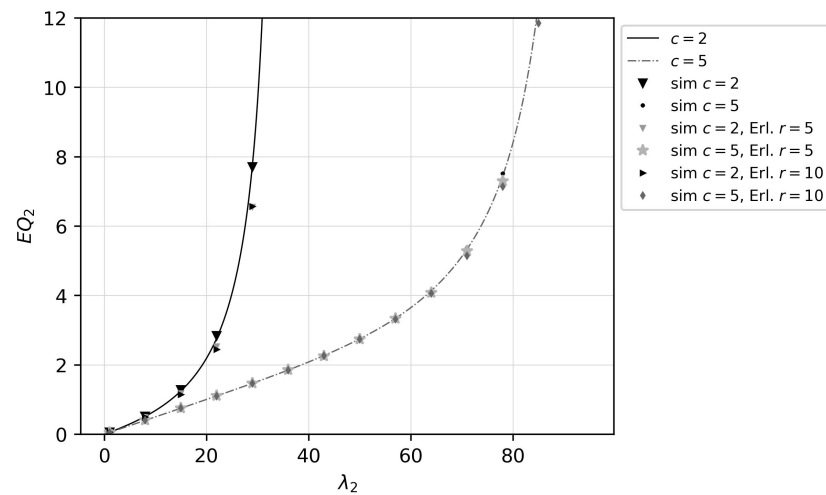

Figure 3. Average waiting time of class-2 customers against arrival rate of class-2 customers ($\lambda_1 = 1$).

Figures 4 and 5 compare the average number of class-2 customers in the system ($\mathrm{E}Q_2$) as λ_1, λ_2 and c change. The observations for $\mathrm{E}Q_2$ in Figures 4 and 5 are the same as those for $\mathrm{E}W_q$ in Figures 2 and 3.

Figure 4. Average number of waiting class-2 customers against arrival rate of class-1 customers ($\lambda_2 = 8$).

Figure 5. Average number of class-2 customers against arrival rate of class-2 customers ($\lambda_1 = 1$).

Figure 6 illustrates how the average number of termination events per class-2 customer (EN_T) changes according to the arrival rates of class-1 customers and the number of servers. Obviously, when class-1 customers arrive more frequently, more class-2 customers' sessions are terminated. In addition, as the number of servers is larger, there are more spaces for all customers, and thus, class-2 customers are less likely to be kicked out of the servers. The same patterns can be observed in Figure 7. Furthermore, simulations results for the Erlang service time distributions of class-1 customers are almost the same as the corresponding ones with exponential distributions. This shows that EN_T is (almost) insensitive to the service time distribution of class-1 customers.

Figure 8 shows the distributions of the number of times that each class-2 customer is terminated. In this experiment, we set $\lambda_1 = 20, \lambda_2 = 30, c = 10$, and let the service time of class-1 customers follow an exponential distribution (denoted by Exp. in the figure) and Erlang (Erl) distributions. It can be seen that higher numbers of termination times occur with smaller probabilities. In addition, there is no significant difference in the results when we modify the distribution of service time for class-1 customers under the three settings.
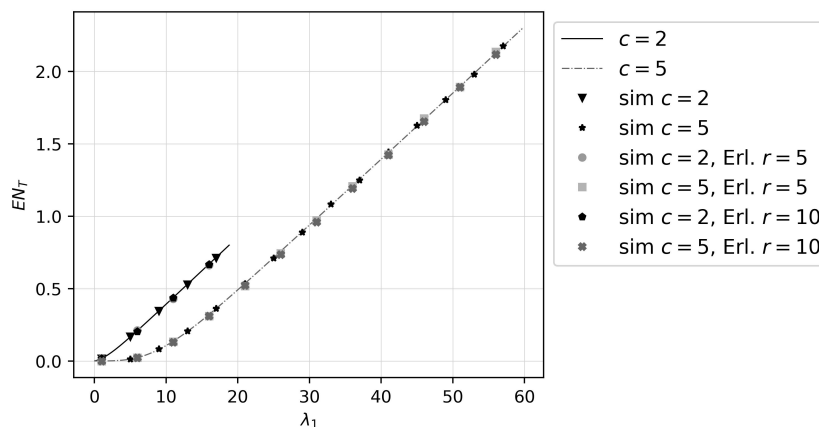
Figure 6. Average number of termination events per class-2 customers against arrival rate of class-1 customers ($\lambda_2 = 8$).
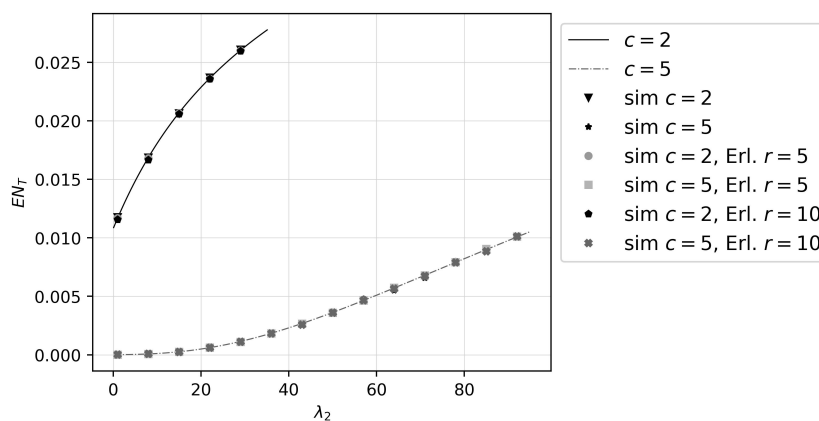

Figure 7. Average number of termination events per class-2 customers against arrival rate of class-2 customers ($\lambda_1 = 1$).
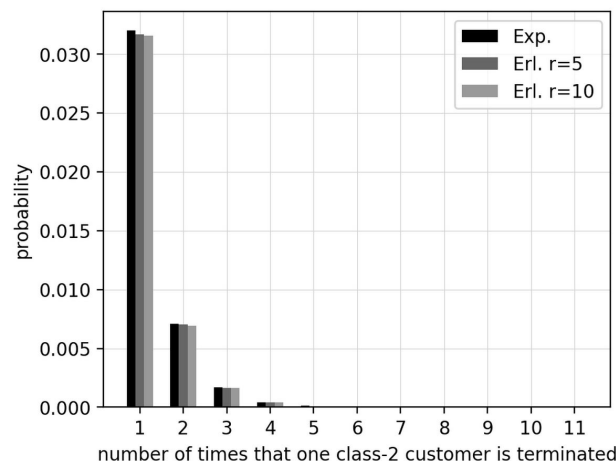

Figure 8. Distribution of number of termination events ($\lambda_1 = 20, \lambda_2 = 30, c = 10$).

Figure 9 reflects changes in the throughput of class-2 customers against λ_1 when λ_2 is fixed. It can be seen that the throughput values remain unchanged at λ_2 when λ_1 goes up to certain thresholds; then, they drop as λ_1 continues to increase. Meanwhile, Figure 10 indicates that the throughput is equal to λ_2 up to a certain threshold of λ_2; then, it remains unchanged at a value $\lambda_{\max}$ when class-2 customers arrive at very high rates. At that, $\lambda_{\max}$ equals the right-hand side of the stability condition (34).
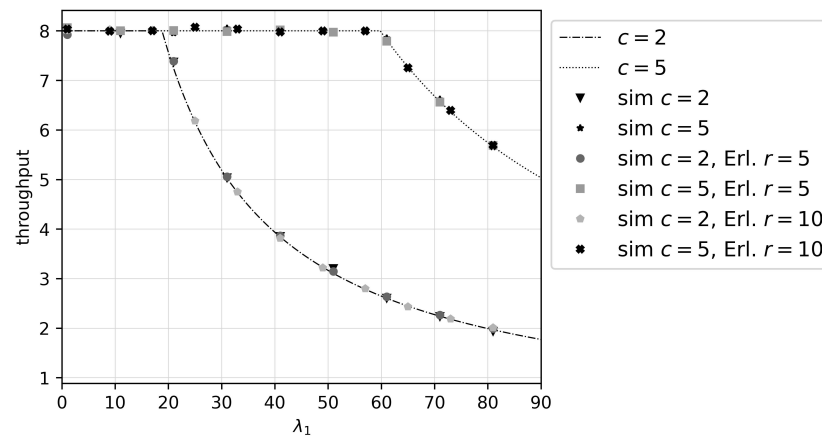
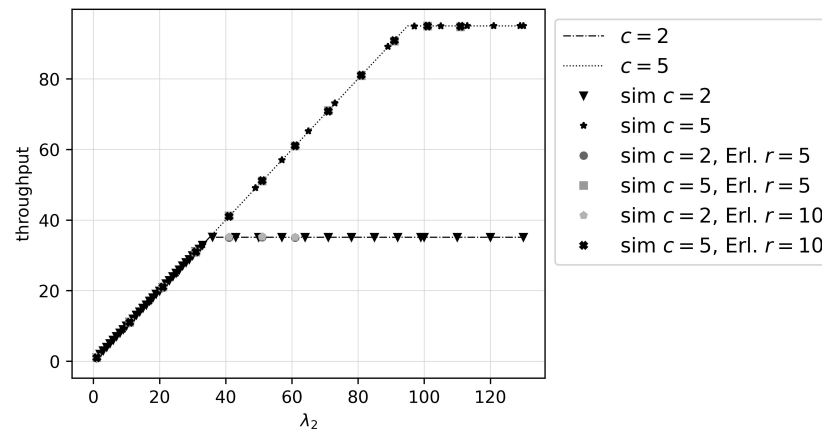**Figure 9.** Throughput against arrival rate of class-1 customers ($\lambda_2 = 8$).

Figure 10. Throughput against arrival rate of class-2 customers ($\lambda_1 = 1$).

Finally, it is noticeable that when we let the service times of class-1 customers follow Erlang distribution, the results do not change much as compared to the exponential case. This agrees with our stability condition, which depends on only the mean service time of class-1 customers.

7. Conclusions

In this paper, we have considered a modified Erlang system for cognitive radio networks and related applications. Using the regenerative methodology, we have established the stability condition, which is insensitive to the service time distributions of primary and secondary users. This analysis is sequentially considered for: (i) the system where there exists an infinite buffer for the awaiting secondary users, (ii) the system where the secondary users meeting a busy server join a virtual orbit and then make retrials and finally, (iii) the system where an idle server generates outgoing calls. The obtained results have implied that the throughput of secondary users is insensitive to the service time distributions of primary and secondary users, provided that the means are fixed. For the case of exponential service time distributions of primary and secondary customers, we have derived some stationary performance measures. Our extensive simulations have shown that the mean waiting time of secondary users increases with the increase in the variance of service time of primary users, while the mean number of terminations of secondary users is almost insensitive to the service time distributions of primary users, provided that the means are fixed. Our findings could be used in resource allocation in cognitive radio networks and related applications.

References

1. Akyildiz, I.F.; Lee, W.Y.; Vuran, M.C.; Mohanty, S. NeXt generation dynamic spectrum access cognitive radio wireless networks: A survey. *Comput. Netw.* **2006**, *50*, 2127–2159. [CrossRef]
2. Mitola, J.; Maguire, G.Q. Cognitive radio: Making software radios more personal. *IEEE Pers. Commun.* **1999**, *6*, 13–18. [CrossRef]
3. Wang, B.; Liu, K.R. Advances in cognitive radio networks: A survey. *IEEE J. Sel. Top. Signal Process.* **2010**, *5*, 5–23. [CrossRef]
4. Akyildiz, I.F.; Lee, W.Y.; Vuran, M.C.; Mohanty, S. A survey on spectrum management in cognitive radio networks. *IEEE Commun. Mag.* **2008**, *46*, 40–48. [CrossRef]
5. Letaief, K.B.; Zhang, W. Cooperative communications for cognitive radio networks. *Proc. IEEE* **2009**, *97*, 878–893. [CrossRef]
6. Ostovar, A.; Keshavarz, H.; Quan, Z. Cognitive radio networks for green wireless communications: An overview. *Telecommun. Syst.* **2021**, *76*, 129–138. [CrossRef]
7. Mitrany, I.L.; Avi-Itzhak, B. A many-server queue with service interruptions. *Oper. Res.* **1968**, *16*, 628–638. [CrossRef]
8. Akutsu, K.; Phung-Duc, T. Analysis of retrial queues for cognitive wireless networks with sensing time of secondary users. In Proceedings of the 14th International Conference, QTNA 2019, Ghent, Belgium, 27–29 August 2019; Springer: Cham, Switzerland, 2019; pp. 77–91.
9. Phung-Duc, T.; Akutsu, K.; Kawanishi, K.; Salameh, O.; Wittevrongel, S. Annals of Operations Research. *Perf. Eval.* **2022**, *310*, 641–660.
10. Salameh, O.; De Turck, K.; Bruneel, H.B.C.; Wittevrongel, S. Analysis of secondary user performance in cognitive radio networks with reactive spectrum handoff. *Telecommun. Syst.* **2017**, *65*, 539–550. [CrossRef]
11. Salameh, O.; Bruneel, H.; Wittevrongel, S. Performance evaluation of cognitive radio networks with imperfect spectrum sensing and bursty primary user traffic. *Math. Probl. Eng.* **2020**, *2020*, 4102046. [CrossRef]
12. Wong, E.W.A.; Foh, C.H. Analysis of cognitive radio spectrum access with finite user population. *IEEE Wirel. Commun. Lett.* **2009**, *13*, 294–296. [CrossRef]
13. Konishi, Y.; Masuyama, H.; Kasahara, S.; Takahashi, Y. Performance analysis of dynamic spectrum handoff scheme with variable bandwidth demand of secondary users for cognitive radio networks. *Wirel. Netw.* **2013**, *19*, 607–617. [CrossRef]
14. Shajin, D.; Dudin, A.N.; Dudina, O.; Krishnamoorthy, A. A two-priority single server retrial queue with additional items. *J. Ind. Manag. Optim.* **2020**, *16*, 2891–2912. [CrossRef]
15. Zhang, Y.; Wang, J.; Li, W.W. Optimal pricing strategies in cognitive radio networks with heterogeneous secondary users and retrials. *IEEE Access* **2019**, *7*, 30937–30950. [CrossRef]
16. Dudin, A.N.; Lee, M.H.; Dudina, O.; Lee, S.K. Analysis of priority retrial queue with many types of customers and servers reservation as a model of cognitive radio system. *IEEE Trans. Commun.* **2016**, *65*, 186–199. [CrossRef]
17. Paul, S.; Phung-Duc, T. Retrial queueing model with two-way communication, unreliable server and resume of interrupted call for cognitive radio networks. In *Information Technologies and Mathematical Modelling. Queueing Theory and Applications*; Springer: Cham, Switzerland, 2018; pp. 213–224.
18. Liu, J.; Jin, S.; Yue, W. Performance evaluation and system optimization of Green cognitive radio networks with a multiple-sleep mode. *Ann. Oper. Res.* **2019**, *277*, 371–391. [CrossRef]
19. Nazarov, A.; Phung-Duc, T.; Paul, S. Unreliable single-server queue with two-Way communication and retrials of blocked and interrupted calls for cognitive radio networks. In *Distributed Computer and Communication Networks*; Springer: Cham, Switzerland, 2018; pp. 276–287.
20. Dimitriou, I.; Phung-Duc, T. Analysis of cognitive radio networks with cooperative communication. In Proceedings of the 13th EAI International Conference on Performance Evaluation Methodologies and Tools, Tsukuba, Japan, 18–20 May 2020; pp. 192–195.
21. Dragieva, V.I.; Phung-Duc, T. Unreliable single-server queue with two-Way communication and retrials of blocked and interrupted calls for cognitive radio networks. In *Analytical and Stochastic Modelling Techniques and Applications*; Springer: Cham, Switzerland, 2019; pp. 18–32.

22. Kim, K. T-preemptive priority queue and its application to the analysis of an opportunistic spectrum access in cognitive radio networks. *Comput. Oper. Res.* **2012**, *39*, 1394–1401. [CrossRef]
23. Azarfar, A.; Frigon, J.F.; Sanso, B. Priority queueing models for cognitive radio networks with traffic differentiation. *EURASIP J Wirel. Commun. Netw.* **2014**, *2014*, 206. [CrossRef]
24. Neuts, M. *Matrix Geometric Solutions in Stochastic Models—An Algorithmic Approach*; Johns Hopkins University Press: Baltimore, MD, USA, 1981.
25. Vishnevskiy, V.M.; Samouylov, K.E.; Yarkina, N.V. Mathematical model of LTE cells with machine-type and broadband communications. *Autom. Remote Control* **2020**, *81*, 622–636. [CrossRef]
26. Asmussen, S. *Applied Probability and Queues*, 2nd ed.; Springer: New York, NY, USA, 2003.
27. Morozov, E.; Steyaert, B. *Stability Analysis of Regenerative Queueing Models*; Springer: Cham, Switzerland, 2021.
28. Morozov, E. Weak regeneration in modeling of queueing processes. *Queueing Syst.* **2004**, *46*, 295–315. [CrossRef]
29. Feller, W. *An Introduction to Probability Theory*; Wiley: New York, NY, USA, 1971.
30. Morozov, E. The tightness in the ergodic analysis of regenerative queueing processes. *Queueing Syst.* **1997**, *27*, 179–203. [CrossRef]
31. Serfozo, R. *Basics of Applied Stochastic Processes*; Springer: Berlin/Heidelberg, Germany, 2009.
32. Morozov, E.; Phung-Duc, T. Stability analysis of a multiclass retrial system with classical retrial policy. *Perf. Eval.* **2017**, *112*, 15–26. [CrossRef]
33. Phung-Duc, T.; Masuyama, H.; Kasahara, S.; Takahashi, Y. A simple algorithm for the rate matrices of level-dependent QBD processes. In Proceedings of the 5th International Conference on Queueing Theory and Network Applications, Beijing, China, 24–26 July 2010; pp. 46–52.