Applications and Applied Mathematics: An International Journal (AAM)

Volume 10 | Issue 1 Article 1

6-2015

A Note on an M/M/s Queueing System with two Reconnect and two Redial Orbits

Amina A. Bouchentouf
Djillali Liabes University of Sidi Bel Abbes

Hanane Sakhi
Sciences and Technology University of Oran Mohamed Boudiaf,

Available at
http://pvamu.edu/aam
Appl. Appl. Math.
ISSN: 1932-9466
Vol. 10, Issue 1 (June 2015), pp. 1 – 12

Applications and Applied
Mathematics:
An International Journal
(AAM)

A Note on an $M/M/s$ Queueing System with two Reconnect and two Redial Orbits

Amina Angelika Bouchentouf and Hanane Sakhi

Department of Mathematics, Djillali Liabes University of Sidi Bel Abbes
B. P. 89, Sidi Bel Abbes 22000, Algeria
bouchentouf_amina@yahoo.fr
&
Department of Mathematics, Sciences and Technology University of Oran
Mohamed Boudiaf, USTOMB BP 1505 El-M'naouar- Oran, Algeria
sakhi.hanane@yahoo.fr

Received: August 8, 2014; Accepted: January 24, 2015

Abstract

A queueing system with two reconnect orbits, two redial (retrial) orbits, s servers and two independent Poisson streams of customers is considered. An arriving customer of type i, $i = 1, 2$ is handled by an available server, if there is any; otherwise, he waits in an infinite buffer queue. A waiting customer of type i who did not get connected to a server will lose his patience and abandon after an exponentially distributed amount of time, the abandoned one may leave the system (lost customer) or move into one of the redial orbits, from which he makes a new attempt to reach the primary queue, and when a customer finishes his conversation with a server, he may comeback to the system, to one of the reconnect orbits where he will wait for another service. In this paper, a fluid model is used to derive a first order approximation for the number of customers in the redial and reconnect orbits in the heavy traffic. The fluid limit of such a model is a unique solution to a system of three differential equations.

Keywords: Abandonment; Bernoulli feedback; call center; fluid limit; queueing system; reconnect queue; redial queue; retrial

MSC 2010 No.: 60K25; 68M20; 90B22

1. Introduction

In queueing theory, such a mechanism in which ejected (or rejected) customers return at random intervals until they receive service is called a retrial queue. Retrial queues have an important application in a wide variety of fields, they are also widespread in the evaluation and design of computer networks, in telecommunications and wireless networks. A retrial queue is similar to any ordinary queueing system in that there is an arrival process and one or more servers. The elemental differences are firstly, the customers who enter during a down or busy period of the server or servers may reattempt to acquire service at some random time in the future, and secondly a waiting room, which is known as a primary queue, in the context of retrial queues is not mandatory. In place of the ordinary waiting room is a buffer called an orbit queue (*in our work we call it a redial orbit*) to which customers proceed after an unsuccessful attempt at service, and from which they retry service according to a given probabilistic or deterministic policy. Owing to the utility and interesting mathematical properties of retrial queueing models, ample literature on the subject has emerged over the past several decades. For a general survey of retrial queues and a summary of many of the results, the reader may refer to the works of (Falin, 1990; Gharbi and Ioualalen, 2006; Falin and Artalejo, 1998; Ebrahimi, 2006; Libman and Orda, 2002; Walrand, 1991; Choi and Kim, 1998) and references therein.

A queueing system with two orbits and two exogenous streams of different type serves as a model for two competing job streams in a carrier sensing multiple access system, where the jobs, after a failed attempt to network access, wait in an orbit queue (Nain, 1985; Szpankowski, 1994). An example of a carrier sensing multiple access system is a local area computer network with bus architecture. The two types of customers can be interpreted as customers with different priority requirements.

A two-class retrial system with a single- server, no waiting room, batch arrivals and classical retrial scheme was introduced and analyzed in (Kulkarni, 1986). Then, in (Falin, 1988) author extended the analysis of the model in (Kulkarni, 1986) to the multi-class setting with an arbitrary number of classes. In (Grishechkin, 1992) the author has established equivalence between the multi-class batch arrival retrial queues with classical retrial policy and branching processes with immigration. In (Moutzoukis and Langaris, 1996) a non-preemptive priority mechanism was added to the model of (Falin, 1988; Kulkarni, 1986). In (Langaris and Dimitriou, 2010) the authors have considered a multi-class retrial system where retrial classes are associated with different phases of service.

The retrial queueing model $MMAP/M_2/1$ with two retrial orbits was studied in (Avrachenkov et al., 2010), where the authors considered a retrial single-server queueing model with two types of customers. When the server is occupied, the arrival customer moves to the orbit depending on the type of the customer, one orbit is infinite while the second one is finite. Joint distribution of the number of customers in the orbits and some performance measures are computed. In (Bouchentouf and Belarbi, 2013) the authors considered two retrial queueing system with balking and feedback, the joint generating function of the number of busy server and the queue length was found by solving the Kummer differential equation by the method of series solution.

At the present time, call centers are becoming an important way of communication with the customer. Therefore, the response-time performance of call centers is crucial for customer satisfaction. Making the right staffing decisions is also fundamental to the costs and the performances of call centers. Numerous models have been developed in order to decide on the optimal number of servers, see (Gans et al., 2002), (Halfin and Whitt, 1981), and the references therein.

So, customer redial behaviors in call centers is quite a significant aspect in queueing (Gans et al., 2002), (Aguir et al., 2008), (Sze, 1984) and the reference therein.

In addition to redials (retrial queue), reconnect is another important aspect, these two latter equivalent should be considered and modeled, see (Gans et al., 2002), (Ding et al., 2013b). Appropriately, we can say that neglecting the impact of redials and reconnects will lead to either overstaffing or understating. In case of overstaffing the performance of the call center will be good, but at needlessly high costs. and in case of understating, the performance of the call center will be degraded, and thus, it may lead to customer dissatisfaction.

When the system is heavily loaded, it would lead to low service levels. However, for large call centers, especially during the busy hours when the inbound volume is quite large, it is possible that the target service levels can be met even in heavy traffic; for more details see (Garnett et al., 2002) and (Borst et al., 2004).

Fluid models for call centers have been extensively studied, (Whitt, 2006), (Mandelbaum et al., 2002). (Whitt,2006), (Mandelbaum et al.,2002). In (Mandelbaum et al., 1999) the fluid and the diffusion approximation for time varying multiserver queue with abandonment and retrials were used, it was shown that the fluid and the diffusion approximation can both be obtained by solving sets of non-linear differential equations. In (Mandelbaum et al., 1998) more general theoretical results for the fluid and diffusion approximation for Markovian service networks was given. In (Aguir et al., 2004) the authors extended the model by allowing customer balking behavior. Fluid models have also been applied in delay announcement of customers in call centers (Ibrahim and Whitt, 2009; Ibrahim and Whitt, 2011). And recently, in (Ding et al., 2013a) the authors study call centers with one redial and one orbit; using fluid limit, they calculate the expected total arrival rate, which is then given as an input to the Erlang, a model for the purpose of calculating service levels and abandonment rates. The performance of such a procedure is validated in the case of single intervals as well as multiple intervals with changing parameters.

In the present paper, an analysis of $M/M/s$ queueing model; a call center with two reconnect, two redials orbits and two exogenous streams of different types is carried out. The goal of this work is to present a type of call center where the emphasis on reconnect and redial orbits is crucial in any telecommunication system.

Now, let us outline the structure of the paper. After the introduction in Section 2, we describe the queueing model. In Section 3, we propose a fluid model, which is a deterministic analogue of the stochastic model. We prove that the original stochastic model converges to the fluid model under proper scaling. So, roughly, we use a fluid model to derive first order approximations for the number of customers in the redial and reconnect orbits in heavy traffic, the fluid limit of such a model is the unique solution to a system of three differential equations.

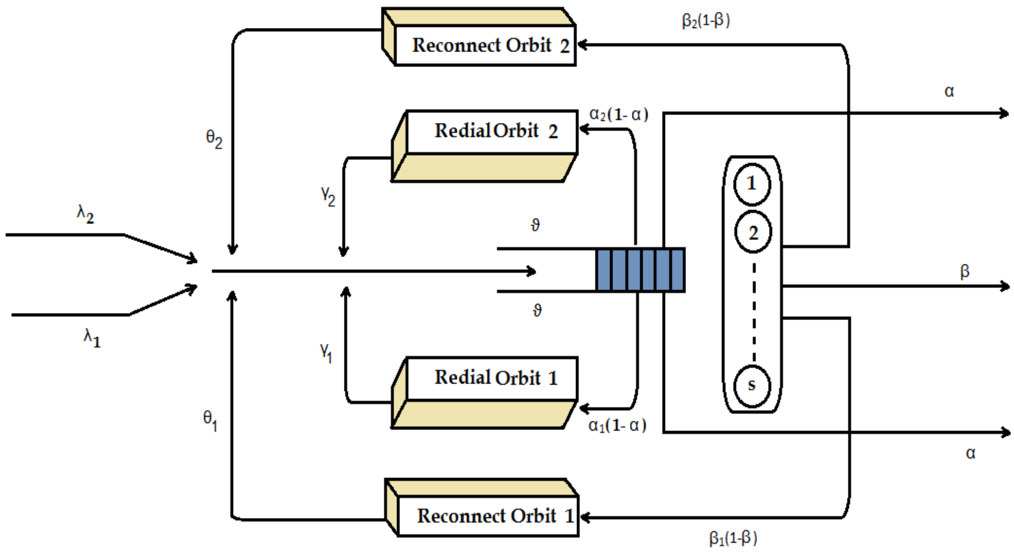

Fig. 1: A queueing model with two redial and two reconnect orbits

2. The model

Consider a queueing model with two reconnect and two redial orbits and s servers (Figure 1). Two independent Poisson streams of customers flow into a common infinite buffer queue. An arriving customer of type i, $i = 1, 2$ is handled by an available server, if there is any; otherwise, he waits in an infinite buffer queue. The customers are handled in the order of arrival (FIFO), The required service time of each customer is independent of its type.

A waiting customer of type i who did not get connected to a server may lose his patience (impatient customer) and abandon after an exponentially distributed amount of time ψ. We assume that $\mathbb{E}(\psi) = \frac{1}{\vartheta} < \infty$, where ϑ is the abandonment rate. The impatient customer either definitively leaves the system with probability α where he is considered as a lost customer or decides to stay requesting service with probability $(1 - \alpha)$. So, if the customer decides not to leave the system, he will route to one of a separate type-i redial queue that attempts to re-dispatch its jobs with probability $(1 - \alpha)\alpha_i$, $(i = 1, 2)$ $(\alpha_1 + \alpha_2 = 1)$, which means that the customer has to choose between two redial orbits (for instance, customer of type 2 may be routed to redial orbit 1 with probability $(1 - \alpha)\alpha_1$, or to redial orbit 2 with probability $(1 - \alpha)\alpha_2$,); the choice is random and does not depend on the threshold of the redial orbits or on the type of the customer. Let's note that a redial orbit is an ordinary waiting room where a customer coming from the primary queue will be waiting until he redials after an exponentially distributed amount of time ω_i, (i=1,2) with $\mathbb{E}(\omega_i) = \gamma_i < \infty$. We assume that the service time of type-i customer has an exponential distribution with mean $\frac{1}{\mu}$. We assume that service times are independent, and the required service time of each customer is independent of its type. After the customer has been served, he may later leave the system with probability β or decide to comeback to the system for another service with probability $(1 - \beta)$. The type-i customer will enter one of the type-i reconnect queue with

probability $(1-\beta)\beta_i$, with $(\beta_1+\beta_2=1)$. As was been mentioned above (in the case of redial orbits), let's note that the customer after getting service may comeback to the system, choosing between two reconnect orbits (for instance, customer of type 2 may be routed to reconnect orbit 1 with probability $(1-\beta)\beta_1$, or to reconnect orbit 2 with probability $(1-\beta)\beta_2$,), the choice is random and does not depend on the threshold of the retrial orbits or on the type of the customer. The reconnect orbit is an ordinary waiting room where a customer coming back from the last service and requiring another service will be waiting and subsequently will reconnect after an exponentially distributed time φ_i, with $\mathbb{E}(\varphi_i)=\theta_i<\infty$.
So an abandoned customer of type i have to choose between one of the two retrial orbits, and the one who decide to comeback for another service have a choice between one of the two reconnect orbits; this means that a customer of type 1 (resp. of type 2) is not directed automatically to the redial or reconnect orbits of type 1 (resp. of type 2).

We assume that α and β do not depend on the customers' experiences in the system. These experiences include holding time, waiting time and the number of times that customers have already tried to get a service. This creates a system with five dependent queues. Such a queueing system serves as a model for two competing job streams in a carrier sensing multiple access system.

3. Fluid approximation

In this section, we calculate $\mathbb{E}(Z_Q(t))$; $\mathbb{E}(Z_R(t))$ and $\mathbb{E}(Z_O)(t)$, where $Z_Q(t)$ is the number of customers in the queue plus the number of customers in service at time t, $\mathbb{E}(Z_R)(t)=\mathbb{E}(Z_{R_1})(t)+\mathbb{E}(Z_{R_2})(t)$ is the number of customers in the redial queues 1 and 2 at time t, and $\mathbb{E}(Z_O)(t)=\mathbb{E}(Z_{O_1})(t)+\mathbb{E}(Z_{O_2})(t)$ is the number of customers in reconnect orbits 1 and 2 at time t.

So, let

$$\mathbb{E}\Delta(t) \quad = \quad \lambda_1(t)+\lambda_2(t)+(\gamma_1+\gamma_1)\mathbb{E}(Z_R(t))+(\theta_1+\theta_2)\mathbb{E}(Z_O(t)); \tag{1}$$

where $\Delta(t)$ stands for the expected number of arrivals up to time t, which is a stochastic process, $\lambda_1(t)+\lambda_2(t)$ stands for the external arrival rate at time t, $(\gamma_1+\gamma_1)\mathbb{E}(Z_R(t))$ and $(\theta_1+\theta_2)\mathbb{E}(Z_O(t))$ stand for the mean arrival rate due to redials and reconnects at time t, respectively. Therefore, once $\mathbb{E}(Z_Q(t))$; $\mathbb{E}(Z_R(t))$ and $\mathbb{E}(Z_O)(t)$, are known, $\mathbb{E}\Delta(t)$ can be obtained by equation (1). Note that $Z_Q(t)$ does not appear in Equation (1), although we will see later that $Z_R(t)$ and $Z_O(t)$ would depend on $Z_Q(t)$.

In fact, it is easy to prove that our stochastic process $\{Z(t);t\geq 0\}$, which is defined by

$$Z(t):=(Z_Q(t),Z_{R_1}(t),Z_{R_2}(t),Z_{O_1}(t),Z_{O_2}(t))^T; \tag{2}$$

is a 5-dimensional Markov process, The state space of this Markov process is $\mathbb{Z}_+^5$. This stochastic process can be reduced to 3-dimensional Markov process

$$(Z_Q(t);Z_R(t);Z_O(t))^T,$$

with $Z_R(t) = Z_{R_1}(t) + Z_{R_2}(t)$ and $Z_O(t) = Z_{O_1}(t) + Z_{O_2}(t)$. Throughout the paper we will focus on the study of 3-dimensional Markov process, since we do not think that is necessary to study each of the reconnect and redial orbits separately.

And since it is a Markov process, we can truncate the system at certain large states, and numerically obtain the steady state distribution of $Z(t)$ by solving the global balance equations. But, for the model we consider, it is difficult to formulate and solve the global balance equations, since their solution offers no insight about the system. Therefore, for the convenience of practical usage, we will not be solving this Markov process, but other approximation methods instead.

A. *Fluid limit*

In this subsection, we present the fluid model, which we will show arises from the limit under a proper scaling of the stochastic model presented in Figure 1. Consider a single interval with the external arrival rates remaining constant during this interval ($\lambda_1(t) + \lambda_2(t) = \lambda_1 + \lambda_2$, $t \geq 0$). The following flow conservation equations hold for this stochastic model

$$Z_Q(t) = Z_Q(0) + \Delta_{\lambda_1}(t) + \Delta_{\lambda_2}(t) + D_R(t) + D_O(t) - D_s(t) - D_a(t). \tag{3}$$

With

$$D_{s_1}(t) + D_{s_2}(t) = D_s(t) \quad \text{and} \qquad D_{a_1}(t) + D_{a_2}(t) = D_a(t).$$

$$Z_R(t) = Z_R(0) + \sum_{j=1}^{D_a(t)} B_j(1-\alpha) - D_R(t). \tag{4}$$

$$Z_O(t) = Z_O(0) + \sum_{j=1}^{D_s(t)} B_j(1-\beta) - D_O(t). \tag{5}$$

Where $\Delta_{\lambda_1}(t)+\Delta_{\lambda_2}(t)$ is the number of external arrivals of type 1 and 2 during time interval $[0;t)$. $\Delta_{\lambda_1}(\cdot)$, and $\Delta_{\lambda_2}(\cdot)$ and Poisson processes of rates λ_1, and λ_2 respectively. $D_R(t) = D_{R_1}(t) + D_{R_2}(t)$ is the number of redials during $[0;t)$, $D_O(t) = D_{O_1}(t)+D_{O_2}(t)$; the number of reconnects during $[0;t)$, $D_s(t) = D_{s_1}(t)+D_{s_2}(t)$; number of served customers of types 1 and 2, during $[0;t)$, and $D_a(t) = D_{a_1}(t) + D_{a_2}(t)$ the number of abandoned customers of types 1 and 2 during $[0;t)$. Finally $B_j(1-\alpha)$ is a Bernoulli random variable with success probability $1-\alpha$, $j = 1, ..., D_{a_i}(t)$, such that

$$B_j(1-\alpha) = \begin{cases} 1, & \text{if the } jth \text{ abandoned customer decides to stay in the system and then} \\ & \text{enters one of the redial orbits;} \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, for given $D_a(t)$, we have $\sum_{j=1}^{D_a(t)} B_j(1-\alpha) \rightsquigarrow Bin(D_a(t), 1-\alpha)$, and for a given $D_s(t)$, we have $\sum_{j=1}^{D_s(t)} B_j(1-\beta) \rightsquigarrow Bin(D_s(t), 1-\beta)$, ($Bin(\cdot,\cdot)$ is a Binomial distribution)

Let Δ_i be independent Poisson processes of rate 1.
So,

$$D_s(t) = \Delta_1\left(\int_0^t \mu \min\{s, Z_Q(u)\}du\right), \tag{6}$$

$$D_a(t) = \Delta_2\left(\int_0^t \vartheta(Z_Q(u) - s)^+ du\right), \tag{7}$$

$$D_R(t) = \Delta_3\left(\int_0^t (\gamma_1 + \gamma_2)Z_R(u)du\right), \tag{8}$$

$$D_O(t) = \Delta_4\left(\int_0^t (\theta_1 + \theta_2)Z_O(u)du\right). \tag{9}$$

For the proof of these statements, we refer to (Pang et al., 2007). To introduce the fluid limit, we consider a sequence of models as in Figure 1 such that, in the n-th model, the external arrival rates are $\lambda_1 n$, $\lambda_2 n$ and the number of servers is ns. We add the superscript "(n)" to all notations in the n-th model. Similarly to (3)-(5), we then have for the n$-th$ model:

$$Z_Q^{(n)}(t) = Z_Q^{(n)}(0) + \Delta_{\lambda_1 n}^{(n)}(t) + \Delta_{\lambda_2 n}^{(n)}(t) + D_R^{(n)}(t) + D_O^{(n)}(t) - D_s^{(n)}(t) - D_a^{(n)}(t), \tag{10}$$

$$Z_R^{(n)}(t) = Z_R^{(n)}(0) + \sum_{j=1}^{D_a^{(n)}(t)} B_j(1-\alpha) - D_R^{(n)}(t). \tag{11}$$

$$Z_O^{(n)}(t) = Z_O^{(n)}(0) + \sum_{j=1}^{D_s^{(n)}(t)} B_j(1-\beta) - D_O^{(n)}(t). \tag{12}$$

Now, dividing by n on both sides of equations (10)-(12),we have

$$\bar{Z}_Q^{(n)}(t) = \bar{Z}_Q^{(n)}(0) + G_Q^{(n)}\left(\bar{Z}^{(n)}\right)(t) + \int_0^t H_Q\left(\bar{Z}^{(n)}\right)(u)du, \tag{13}$$

$$\bar{Z}_R^{(n)}(t) = \bar{Z}_R^{(n)}(0) + G_R^{(n)}\left(\bar{Z}^{(n)}\right)(t) + \int_0^t H_R\left(\bar{Z}^{(n)}\right)(u)du, \tag{14}$$

$$\bar{Z}_O^{(n)}(t) = \bar{Z}_O^{(n)}(0) + G_O^{(n)}\left(\bar{Z}^{(n)}\right)(t) + \int_0^t H_O\left(\bar{Z}^{(n)}\right)(u)du, \tag{15}$$

Where

$$\begin{aligned} G_Q^{(n)}\left(\bar{Z}^{(n)}\right)(t) = \quad & \left(\frac{\Delta_{\lambda_1 n}^{(n)}(t) + \Delta_{\lambda_2 n}^{(n)}(t)}{n} - (\lambda_1 + \lambda_2)t\right) - \left(\bar{D}_s^{(n)}(t) - \int_0^t \mu \min\{s, \bar{Z}_Q^{(n)}(u)\}du\right) \\ & - \left(\bar{D}_a^{(n)}(t) - \int_0^t \vartheta(\bar{Z}_Q^{(n)}(u) - s)^+ du\right) + \left(\bar{D}_R^{(n)}(t) - \int_0^t (\gamma_1 + \gamma_2)\bar{Z}_R^{(n)}(u)du\right) \\ & + \left(\bar{D}_O^{(n)}(t) - \int_0^t (\theta_1 + \theta_2)\bar{Z}_O^{(n)}(u)du\right) \end{aligned} \tag{16}$$

$$G_R^{(n)}\left(\bar{Z}^{(n)}\right)(t) = \left(\sum_{j=1}^{n\bar{D_a}^{(n)}(t)} B_j(1-\alpha)/n - \int_0^t (1-\alpha)\vartheta\left(\bar{Z}_Q^{(n)}(u)-s\right)^+ du\right) - \left(\bar{D}_R^{(n)}(t) - \int_0^t (\gamma_1+\gamma_2)\bar{Z}_R^{(n)}(u)du\right) \tag{17}$$

$$G_O^{(n)}\left(\bar{Z}^{(n)}\right)(t) = \left(\sum_{j=1}^{n\bar{D_s}^{(n)}(t)} B_j(1-\beta)/n - \int_0^t (1-\beta)\mu\min\{s,\bar{Z}_Q^{(n)}(u)\}du\right) - \left(\bar{D}_O^{(n)}(t) - \int_0^t (\theta_1+\theta_2)\bar{Z}_O^{(n)}(u)du\right). \tag{18}$$

And

$$\bar{D_s}^{(n)}(t) = \Delta_1\left(n\int_0^t \mu\min\{s,\bar{Z}_Q^{(n)}(u)\}du\right)/n \tag{19}$$

$$\bar{D_a}^{(n)}(t) = \Delta_2\left(n\int_0^t \vartheta(\bar{Z}_Q^{(n)}(u)-s)^+du\right)/n \tag{20}$$

$$\bar{D}_R^{(n)}(t) = \Delta_3\left(n\int_0^t (\gamma_1+\gamma_2)\bar{Z}_R^{(n)}(u)du\right)/n \tag{21}$$

$$\bar{D}_O^{(n)}(t) = \Delta_4\left(n\int_0^t (\theta_1+\theta_2)\bar{Z}_O^{(n)}(u)du\right)/n \tag{22}$$

And

$$\int_0^t H_Q\left(\bar{Z}^{(n)}\right)(u)du = \int_0^t \lambda_1+\lambda_2+(\gamma_1+\gamma_2)\bar{Z}_R^{(n)}(u)+(\theta_1+\theta_2)\bar{Z}_O^{(n)}(u) -\mu\min\{s,\bar{Z}_Q^{(n)}(u)\} - \vartheta\left(\bar{Z}_Q^{(n)}(u)-s\right)^+ du \tag{23}$$

$$\int_0^t H_R\left(\bar{Z}^{(n)}\right)(u)du = \int_0^t (1-\alpha)\vartheta\left(\bar{Z}_Q^{(n)}(u)-s\right)^+ - (\gamma_1+\gamma_2)\bar{Z}_R^{(n)}(u)du \tag{24}$$

$$\int_0^t H_O\left(\bar{Z}^{(n)}\right)(u)du = \int_0^t (1-\beta)\mu\min\{s,\bar{Z}_Q^{(n)}(u)\} - (\theta_1+\theta_2)\bar{Z}_O^{(n)}(u)du \tag{25}$$

For the convenience of notation, we rewrite equations (13)-(15) in the vector form

$$\bar{Z}^{(n)}(t) = \bar{Z}^{(n)}(0) + G^{(n)}\left(\bar{Z}^{(n)}\right)(t) + \int_0^t H_Q(\bar{Z}^{(n)})(u)du \tag{26}$$

where

$$G^{(n)}\left(\bar{Z}^{(n)}\right)(t) = (G_Q^{(n)}\left(\bar{Z}^{(n)}\right)(t), G_R^{(n)}\left(\bar{Z}^{(n)}\right)(t), G_O^{(n)}\left(\bar{Z}^{(n)}\right)(t))^T$$

$$H^{(n)}\left(\bar{Z}^{(n)}\right)(u) = (H_Q^{(n)}\left(\bar{Z}^{(n)}\right)(u), H_R^{(n)}\left(\bar{Z}^{(n)}\right)(u), H_O^{(n)}\left(\bar{Z}^{(n)}(u)\right)^T$$

Now, let us define the fluid scaled process

$$\bar{Z}^n(t) = (\bar{Z}_Q^{(n)}(t), \bar{Z}_R^{(n)}(t), \bar{Z}_O^{(n)}(t))^T$$

Where

$$\bar{Z}_Q^{(n)}(t) = \frac{Z_Q^{(n)}(t)}{n}, \quad \bar{Z}_R^{(n)}(t) = \frac{Z_R^{(n)}(t)}{n}, \quad \bar{Z}_O^{(n)}(t) = \frac{Z_O^{(n)}(t)}{n}.$$

Let $D([0,\infty),\mathbb{R}^3)$ be the space of right continuous functions with left limits in $\mathbb{R}^3$ having the domain $[0,\infty)$. We endow $D([0,\infty),\mathbb{R}^3)$ with the usual Skorokhod J_1 topology. Assume that $\{X^{(n)}\}_{n=1}^{\infty}$ is a sequence of stochastic processes, then $X^{(n)} \rightarrow^d x$ means that $X^{(n)}$ converge weakly to stochastic process x.

Definition. If there exists a limit in distribution for the scaled process $\{\bar{Z}^{(n)}(\cdot)\}_{n=1}^{\infty}$, i.e. $\bar{z}^{(n)}(\cdot) \rightarrow^d z(\cdot)$, then $z(\cdot)$ is called the fluid limit of the original stochastic model.

Now, let us define the fluid limit of the original stochastic model; so if

$$H_0: \ (\overline{Z}_Q^{(n)}(0), \overline{Z}_R^{(n)}(0), \overline{Z}_O^{(n)}(0)) \rightarrow^d (z_Q(0), Z_R(0), z_O(0)) \ \text{ as } \ n \rightarrow \infty,$$

holds, then the fluid limit of the original stochastic model is the unique solution to the following system of equations

$$\begin{aligned} Z_Q(t) = Z_Q(0) + (\lambda_1+\lambda_2)t + (\gamma_1+\gamma_2)\int_0^t Z_R(u)du + (\theta_1+\theta_2)\int_0^t Z_O(u)du - \\ \mu\int_0^t \min\{s, Z_Q(u)\}du - \vartheta\int_0^t (Z_Q(u)-s)^+ du \end{aligned} \tag{27}$$

$$Z_R(t) = Z_R(0) + (1-\alpha)\vartheta\int_0^t (Z_Q(u)-s)^+ du - (\gamma_1+\gamma_2)\int_0^t Z_R(u)du. \tag{28}$$

$$Z_O(t) = Z_O(0) + (1-\beta)\mu\int_0^t \min\{s, Z_Q(u)\}du - (\theta_1+\theta_2)\int_0^t Z_O(u)du. \tag{29}$$

The proof is as presented in (Ding et al., 2013b).

In the next section, we derive from equations (27)-(29) the fluid limit in stationarity.

4. The stationarity of the model

In this section we develop conditions under which $z(t)$ is constant. By differentiating Equations (27)-(29), we obtain

$$\lambda_1 + \lambda_2 = \mu \min\{s, z_Q(\infty)\}) + \vartheta(z_Q(\infty) - s)^+ - (\gamma_1 + \gamma_2) z_R(\infty) - (\theta_1 + \theta_2) z_O(\infty). \tag{30}$$

$$(\gamma_1 + \gamma_2) z_R(\infty) = (1 - \alpha)\vartheta(z_Q(\infty) - s)^+. \tag{31}$$

$$(\theta_1 + \theta_2) z_O(\infty) = (1 - \beta)\mu \ \min\{s, z_Q(\infty)\}. \tag{32}$$

Where $z_Q(\infty) = \lim_{t\longrightarrow\infty} z_Q(t)$, $z_R(\infty) = \lim_{t\longrightarrow\infty} z_R(t)$, $z_O(\infty) = \lim_{t\longrightarrow\infty} z_O(t)$. Equations (30)-(32) can be easily solved with respect to $z_Q(\infty)$, $z_R(\infty)$ and $z_O(\infty)$, yielding

$$Z_Q(\infty) = \begin{cases} \frac{\lambda_1+\lambda_2}{\mu\beta} & \text{if } \frac{\lambda_1+\lambda_2}{s\mu} < \beta \\ \frac{\lambda_1+\lambda_2-\beta\mu s}{\vartheta\alpha} + s & \text{otherwise} \end{cases} \tag{33}$$

$$Z_R(\infty) = \begin{cases} 0 & \text{if } \frac{\lambda_1+\lambda_2}{s\mu} < \beta \\ \frac{(1-\alpha)\vartheta(z_Q(\infty)-s)}{\gamma_1+\gamma_2} & \text{otherwise.} \end{cases} \tag{34}$$

$$Z_O(\infty) = \begin{cases} \frac{(1-\beta)\mu z_Q(\infty)}{\theta_1+\theta_2} & \text{if } \frac{\lambda_1+\lambda_2}{s\mu} < \beta \\ \frac{(1-\beta)\mu s}{\theta_1+\theta_2} & \text{otherwise} \end{cases} \tag{35}$$

B. Discussion on the results

At first, let us notice that $\frac{\lambda_1+\lambda_2}{s\mu}$ is the load of the system due to the external arrivals. $\frac{\lambda_1+\lambda_2}{\mu\beta}$ is the total load when there is no redials. Then let us see that in expressions (33), (34) and (35), the value of $\frac{\lambda_1+\lambda_2}{\mu\beta}$ determines whether the fluid model is in heavy traffic or not. So, now if $\frac{\lambda_1+\lambda_2}{\mu\beta} < 1$, since the fluid limit is deterministic, we get $z_Q(\infty) < s$, and there is no customer in the two redial orbits which means that $z_R(\infty) = 0$ would hold, which also means that there is no abandonment at all in the fluid limit. In reality, due to the variability of the service duration and patience, abandonments would not be 0, though very small. Now, if $\frac{\lambda_1+\lambda_2}{\mu\beta} > 1$, by equation (33), we have $z_Q(\infty) > s$. Consequently, the fluid model indicates that there will be $z_Q(\infty) - s$ amount of customers waiting, each with rate ϑ, and customers will be routed to the redial orbits with rate $(1 - \alpha)\vartheta(z_Q(\infty) - s)$.

5. Conclusion

In our paper, a call center with two retrial orbits, two redial orbits and two exogenous streams of different types was studied; we analyzed this system where customers can abandon, and the abandoned one may redial; also when a customer finishes service, he may reconnect. In this work, a fluid model is used to derive first order approximations for the number of customers in the redial and reconnect orbits in heavy traffic; the fluid limit of such a model is the unique solution to a system of three differential equations.

Acknowledgments

The authors are thankful to anonymous referees and the Editor-in-Chief Professor Aliakbar Montazer Haghighi for useful comments and suggestions towards the improvement of this paper.

REFERENCES

Aguir, M., Akşin, O., Karaesmen, F., and Dallery, Y. (2008). On the interaction between retrials and sizing of call centers. *European Journal of Operational Research*, 191(2):398–408.

Aguir, M., Karaesmen, F., Aksin, O., and Chauvet, F. (2004). The impact of retrials on call center performance. *OR Spectrum*, 26(3):353–376.

Avrachenkov, K., Dudin, A., and Klimenok, V. (2010). Queueing model $mmap/m_2/1$ with two orbits. *Lecture Notes in Computer Science*, 6235:107–118.

Borst, S., Mandelbaum, A., and Reiman, M. (2004). Dimensioning large call centers. *Operations Research*, 52(1):17–34.

Bouchentouf, A. and Belarbi, F. (2013). Performance evaluation of two markovian retrial queueing model with balking and feedback. *Acta Univ. Sapientiae, Mathematica*, 5(2):132–146.

Choi, B. and Kim, Y. (1998). The $m/m/c$ retrial queue with geometric loss and feedback. *Computers Math. Applications*, 36(6):41–52.

Ding, S., Remerova, M., van der Mei, R., and Zwart, B. (2013a). Fluid approximation of a call center model with redials and reconnects. *Cornell University Library,arXiv:1311.6248v1 [Math.PR]*.

Ding, S., van der Mei, R., and Koole, G. (2013b). A method for estimation of redial and reconnect probabilities in call centers. *Proceedings of the 2013 Winter Simulation Conference, Winter Simulation Conference*.

Ebrahimi, N. (2006). System reliability based on system wear. *Stochastic Models*, 22(1):21–36.

Falin, G. (1988). On a multiclass batch arrival retrial queue. *Adv. Appl. Probab*, 20:483–487.

Falin, G. (1990). A survey of retrial queues. *Queueing Systems: Theory and Applications*, 7(2):127–167.

Falin, G. I. and Artalejo, J. R. (1998). An infinite source retrial queue. *European Journal of Operational Research*, 108(2):409–424.

Gans, N., Koole, G., and Mandelbaum, A. (2002). Telephone calls centers: a tutorial and literature review. *European Jour. Oper. Res.*

Garnett, O., Mandelbaum, A., and Reiman, M. (2002). Designing a call center with impatient customers. *Manufacturing & Service Operations Management*, 4(3):208–227.

Gharbi, N. and Ioualalen, M. (2006). Gspn analysis of retrial systems with server breakdowns and repairs. *Applied Mathematics and Computation*, 174(2):1151–1168.

Grishechkin, S. (1992). Multiclass batch arrival retrial queues analyzed as branching processes with immigration. *Queueing Syst*, 11:395–418.

Halfin, S. and Whitt, W. (1981). Heavy-traffic limits for queues with many exponential servers. *Operations Research*, 29(3):567–588.

Ibrahim, R. and Whitt, W. (2009). Real-time delay estimation based on delay history. *Manufacturing & Service Operations Management*, 11(3):397–415.

Ibrahim, R. and Whitt, W. (2011). Wait-time predictors for customer service systems with time-varying demand and capacity. *Operations Research*, 59(5):1106–1118.

Kulkarni, V. (1986). Expected waiting times in a multiclass batch arrival retrial queue. *J. Appl. Probab*, 23:144–154.

Langaris, C. and Dimitriou, I. (2010). A queueing system with n-phases of service and $(n-1)$-types of retrial customers. *Eur. J. Oper. Res*, 205:638–649.

Libman, L. and Orda, A. (2002). Optimal retrial and timeout strategies for accessing network resources. *IEEE/ACM Transactions on Networking*, 10(4):551–564.

Mandelbaum, A., Massey, W., and Reiman, M. (1998). Strong approximations for markovian service networks. *Queueing Systems*, 30(1-2):149–201.

Mandelbaum, A., Massey, W., Reiman, M., and Rider, B. (1999). Time varying multiserver queues with abandonment and retrials. *Proceedings of the 16th International Teletraffic Conference*, 4:4–7.

Mandelbaum, A., Massey, W., Reiman, M., Stolyar, A., and Rider, B. (2002). Queue lengths and waiting times for multiserver queues with abandonment and retrials. *Telecommunication Systems*, 21(2-4):149–171.

Moutzoukis, E. and Langaris, C. (1996). Non-preemptive priorities and vacations in a multiclass retrial queueing system. *Stoch Models*, 12(3):455–472.

Nain, P. (1985). Analysis of a two-node aloha network with infinite capacity buffers. *Proc. Int. Seminar on Computer Networking and Performance Evaluation, T. Hasegawa, H. Takagi, and Y. Takahashi (eds.), Tokyo, Japan*, pages 18–20.

Pang, G., Talreja, R., and Whitt, W. (2007). Martingale proofs of many-server heavy-traffic limits for markovian queues. *Probability Surveys*, 4:193–267.

Sze, D. (1984). Or practice - a queueing model for telephone operator staffing. *Operations Research*, 32(2):229–249.

Szpankowski, W. (1994). Stability conditions for some multiqueue distributed systems: buffered random access systems. *Adv. Appl. Probab*, 26:498–515.

Walrand, J. (1991). *Communication Networks: A First Course*. The Aksen Associates Series in Electrical and Computer Engineering. Richard D. Irwin, Inc., and Aksen Associates, Inc., Homewood, IL and Boston, MA.

Whitt, W. (2006). Fluid models for multiserver queues with abandonments. *Operations Research*, 54(1):37–54.