Research Article

On the Convergence of Truncated Processes of Multiserver Retrial Queues

M. Jose Domenech-Benlloch, 1 Jose Manuel Gimenez-Guzman, 2 Vicent Pla, 1 Jorge Martinez-Bauset, 1 and Vicente Casares-Giner 1

1 Departamento de Comunicaciones, Universidad Politécnica de Valencia, 46022 Valencia, Spain
2 Departamento de Automática, Universidad de Alcalá, Alcalá de Henares, 28871 Madrid, Spain

Correspondence should be addressed to M. Jose Domenech-Benlloch, mdoben@doctor.upv.es

Received 1 March 2010; Accepted 28 July 2010

Academic Editor: J. Jiang

Copyright © 2010 M. Jose Domenech-Benlloch et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Retrial queues can only be solved in a closed form in very few and simple cases, so researchers must resort to approximate models. However, most of the papers that propose approximate models assume the convergence of the proposed models to their exact counterparts, without providing a rigorous mathematical proof. In this paper we demonstrate the convergence of finite truncated models with two reattempt orbits.

1. Introduction

Multiserver retrial queues (MRQs) are used to model multiple situations in our daily life activities, from queues in a supermarket to communication networks. MRQs model the activities in which a user that cannot get service immediately does not leave the system but returns after a random time to try again. The general structure of an MRQ contains two blocks: a block that accommodates the servers and a delay block for the repeated attempts. The general purpose of modelling MRQs is to obtain the steady-state probabilities for later computing some system descriptors. Unfortunately, it is only possible to derive closed-form expressions in a few and simple retrial models, so it is necessary to resort to approximate models and algorithmic methods to solve most of the MRQs [1, 2]. There is a wide literature that proposes new approximate models that are able to solve—approximately—MRQs [3–5]. Most of these approximations are based on some sort of truncation or generalized truncation [4, 6]. Intuitively, one would expect that the higher the level of truncation—that is, the more similar is the approximate model to the original one—the more accurate is the computed approximation and that as the level of truncation grows to infinity—that is, the approximate model tends to the original one—the computed approximation converges to the exact solution. The latter would be a desirable property for any approximation, not only
theoretically but also from a practical perspective since it basically states that one can get as close as wanted to the exact solution by increasing the computational effort. Unfortunately, beyond intuition there is no guarantee that such property holds in the general case. On the other hand, it is worth noting that the fulfillment of the aforementioned property cannot be verified numerically unless a previous approximation exists for which such a guarantee has been established.

As mentioned, there are an important number of published approximations to otherwise unsolvable retrial models. However, the convergence issue is only addressed in a few of them [7] and, to the best of our knowledge, all of them deal with MRQs with a single reattempt orbit [2, 7–10].

The objective of this paper is to establish the convergence of some finite truncated models with two reattempt orbits to their original and exact counterparts. More concretely, the original model for which we demonstrate the convergence is based on a cellular network system which is described in Section 2. Different approximate solution methods for that system have been proposed in [5]. However, it is important to note that it is just a possible scenario of a two-reattempt-orbit MRQ and that a similar procedure could be suitable to other systems in which there are different reattempt types.

The remainder of this paper is organized as follows. Section 2 describes the model of the system under study, while in Section 3 we mathematically verify the convergence of a family of finite truncated models to their original counterparts in a two-reattempt-orbit multiserver retrial queue. Final remarks and a summary of the paper are provided in Section 4.

2. System Model and Description

The system under study is a mobile cellular network. In this type of network, the coverage area is divided into services areas, named cells, and customers can move across different cells. When a customer with an active communication moves from one cell to another, a so-called handover procedure is executed to allocate the necessary resources in the new cell and release the unused resources in the former cell. Obviously, both new sessions and handovers will have their own reattempts (redials and retrials, resp.) but with very different characteristics. As in GSM [11], blocked handovers will be automatically retried until a reattempt succeeds or the user moves outside the handover area. In contrast, persistence of redials depends on the user patience. Another difference is that the maximum number of unsuccessful automatic retrials is set by the network operator while redials are affected by the randomness of human behavior. On the other hand, both types of reattempts may have a different importance for the network and consequently go through different admission procedures. Therefore, both types of reattempts have different characteristics and consequently two separate reattempt blocks have to be considered.

This model considers two arrival streams: new sessions and handovers from adjacent cells. Both arrivals are considered to be Poisson processes with rates $\lambda_n$ and $\lambda_h$, respectively, with $\lambda = \lambda_n + \lambda_h$. Those arrivals try to access a finite number of servers $C$. When a user is served, he/she holds the resource during an exponentially distributed time with rate $\mu$ (channel holding time). When an incoming new session is blocked, it joins the redial orbit with probability $1 - P^1_{in}$ or leaves the system with probability $P^1_{in}$. If a redial is not successful, then the session returns to the redial orbit with probability $1 - P_{th}$, redialing after an exponentially distributed time with rate $\mu_{red}$. Likewise, $P^1_{ih}, P_{ih},$ and $\mu_{red}$ are the analogous parameters for automatic retrials.
In general, blocking a new session setup is considered to be less harmful than blocking a handover attempt. Thus we include an admission control policy to guarantee the prioritization of handovers—and retrials—over new sessions—and their associated redials. The technique used is known as the Fractional Guard Channel (FGC) admission control policy that is characterized by only one parameter $t (0 \leq t \leq C)$. New sessions and redials are accepted with probability 1 when there are less than $L = [t]$ resources being used and with probability $f = t - L$, when there are exactly $L$ resources in use. If there are more than $L$ busy resources, then new sessions and redials are no longer accepted. Handovers and automatic retrials are only rejected when the system is completely occupied.

Let $X(t) = (X_1(t), X_2(t), X_3(t))$ be the process associated with the original model whose state space is given by $S := \{(x_1, x_2, x_3) : x_1 = 0, 1, \ldots, C; x_2 \in \mathbb{Z}_+; x_3 \in \mathbb{Z}_+\}$, where $X_1(t)$ is the number of sessions being served, $X_2(t)$ the number of new sessions redialing, and $X_3(t)$ the number of handovers in the retrial orbit. Let $x = (x_1, x_2, x_3) \in S$ and let $e_i, i = 1, 2, 3$, be a 3-dimensional vector whose $i$th component is 1 and the rest are 0. The transitions (and their rates) from $x$ are as follows.

Arrivals correspond to transitions of the form $x \rightarrow x + e_i$ whose transition rate $(\lambda_i(x))$ is given as

$$
\begin{align*}
\lambda_1(x) &= \lambda_1(x_1) = \\
&= \begin{cases}
\lambda & \text{if } x_1 \leq L - 1, \\
\lambda_h + f \lambda_n & \text{if } x_1 = L, \\
\lambda_h & \text{if } L < x_1 < C, \\
0 & \text{if } x_1 = C,
\end{cases} \\
\lambda_2(x) &= \lambda_2(x_1) = \\
&= \begin{cases}
0 & \text{if } x_1 < L, \\
\lambda_n (1 - P_{in}^1) (1 - f) & \text{if } x_1 = L, \\
\lambda_n (1 - P_{in}^1) & \text{if } x_1 > L,
\end{cases} \\
\lambda_3(x) &= \lambda_3(x_1) = \\
&= \begin{cases}
0 & \text{if } x_1 < C, \\
\lambda_h (1 - P_{th}^1) & \text{if } x_1 = C.
\end{cases}
\end{align*}
$$

Departures correspond to transitions of the form $x \rightarrow x - e_i$ whose transition rate $(\mu_i(x))$ is given as

$$
\begin{align*}
\mu_1(x) &= x_1 \mu, \\
\mu_2(x) &= x_2 \beta_2(x_1) = x_2 \cdot \\
&= \begin{cases}
0 & \text{if } x_1 < L, \\
\mu_{red} (1 - f) P_{in} & \text{if } x_1 = L, \\
\mu_{red} P_{in} & \text{if } x_1 > L,
\end{cases} \\
\mu_3(x) &= x_3 \beta_3(x_1) = x_3 \cdot \\
&= \begin{cases}
0 & \text{if } x_1 < C, \\
\mu_{rel} P_{th} & \text{if } x_1 = C.
\end{cases}
\end{align*}
$$
The retrial orbit is bounded by $X$. Let us denote by

$$
a_{2,1}(x) = x_2 a_{2,1}(x_1) = x_2 \cdot \begin{cases} 
\mu_{\text{red}} & \text{if } x_1 \leq L - 1, \\
\mu_{\text{red}} f & \text{if } x_1 = L, \\
0 & \text{if } x_1 > L,
\end{cases}
$$

$$
a_{3,1}(x) = x_3 a_{3,1}(x_1) = x_3 \cdot \begin{cases} 
\mu_{\text{ret}} & \text{if } x_1 < C, \\
0 & \text{if } x_1 = C.
\end{cases}
$$

The process $X(t)$ is clearly a migration process since the only possible transitions are of the form $x \to x + e_i$ or $x \to x - e_i + e_j$.

3. On the Convergence of the Doubly Truncated Process

Let us denote by $X^{(N)}(t)$ a truncated version of $X(t)$ where the number of customers in the retrial orbit is bounded by $N$. Likewise, let $X^{(M,N)}(t)$ be a doubly truncated version of $X(t)$ where the number of customers in the retrial orbit is bounded by $M$ and the number of customers in the retrial orbit is bounded by $N$. In other words, $X^{(M,N)}(t)$ is a truncation of $X^{(N)}(t)$. The state spaces of $X^{(N)}(t)$ and $X^{(M,N)}(t)$ are, respectively, given by

$$
S^{(N)} := \{(x_1, x_2, x_3) : 0 \leq x_1 \leq C; x_2, x_3 \in \mathbb{Z}_+; 0 \leq x_3 \leq N\},
$$

$$
S^{(M,N)} := \{(x_1, x_2, x_3) : 0 \leq x_1 \leq C; 0 \leq x_2 \leq M; 0 \leq x_3 \leq N\}.
$$

Note that $S^{(M,N)} \subset S^{(N)} \subset S$.

Obviously, $X^{(N)}(t)$ and $X^{(M,N)}(t)$ are also migration processes and their transition rates are as given below. For the sake of conciseness, in the following definitions we used the indicator function $\mathbb{1}_{\{\cdot\}}$, and it must be considered that $x \in S^{(N)}$ or $x \in S^{(M,N)}$ as appropriate.

Arrivals’ transition rates are given as

$$
\lambda_{1}^{(M,N)}(x) = \lambda_{1}^{(N)}(x) = \lambda_{1}(x), \\
\lambda_{2}^{(M,N)}(x) = \mathbb{1}_{\{x_2<M\}} \lambda_{2}^{(N)}(x), \\
\lambda_{3}^{(M,N)}(x) = \mathbb{1}_{\{x_3<N\}} \lambda_{3}(x).
$$

Departures’ transition rates are given as

$$
\mu_{1}^{(M,N)}(x) = \mu_{1}^{(N)}(x) = \mu_{1}(x), \\
\mu_{2}^{(M,N)}(x) = \mu_{2}^{(N)}(x) = \mu_{2}(x), \\
\mu_{3}^{(M,N)}(x) = \mu_{3}^{(N)}(x) = \mu_{3}(x).
$$
Successful reattempts’ transition rates are given as
\[
\begin{align*}
\alpha_{2,1}^{(M,N)}(x) &= \alpha_{2,1}^{(N)}(x) = \alpha_{2,1}(x), \\
\alpha_{3,1}^{(M,N)}(x) &= \alpha_{3,1}^{(N)}(x) = \alpha_{3,1}(x).
\end{align*}
\] (3.4)

The two pairs of processes \((X^{(M,N)}(t), X^{(M+1,N)}(t))\) and \((X^{(M,N)}(t), X^{(N)}(t))\) fulfill the hypotheses of [2, Statement 12; page 114]—to be more precise, it should be noted that we are using the writing of the hypotheses given in [12]. This condition is verified as shown below.

Firstly, the hypotheses of Statement 12 are rewritten particularized to the processes of our interest: \(X^{(M,N)}(t)\) and \(X^{(M+1,N)}(t)\).

Let \(x \in S^{(M,N)}\) and \(x' \in S^{(M+1,N)}\) such that \(x \leq x'\) (coordinatewise). Then we define the following:
\[
\begin{align*}
I(x, x') &:= \{i \in \{1, 2, 3\} \mid x_i = x'_i\}, \\
J_i^f(x, x') &:= \{j \in I(x, x') \setminus \{i\} \mid a_{ij}^{(M,N)}(x) > a_{ij}^{(M+1,N)}(x)\}, \\
J_i^p(x, x') &:= \{j \in I(x, x') \setminus \{i\} \mid a_{ij}^{(M,N)}(x) < a_{ij}^{(M+1,N)}(x)\},
\end{align*}
\] (3.5)

Now, it has to be shown that, for any \(x \in S^{(M,N)}\) and \(x' \in S^{(M+1,N)}\) such that \(x \leq x'\) and for all \(i \in I(x, x')\), the following inequalities hold:
\[
\begin{align*}
\lambda_i^{(M,N)}(x) + \sum_{j \in J_i^f(x, x')} a_{ij}^{(M,N)}(x) &\leq \lambda_i^{(M+1,N)}(x') + \sum_{j \in J_i^f(x, x')} a_{ij}^{(M+1,N)}(x'), \\
\mu_i^{(M,N)}(x) + \sum_{j \in J_i^p(x, x')} a_{ij}^{(M,N)}(x) &\geq \mu_i^{(M+1,N)}(x') + \sum_{j \in J_i^p(x, x')} a_{ij}^{(M+1,N)}(x').
\end{align*}
\] (3.6) (3.7)

Note that \(a_{1,2}^{(M,N)}(x) = a_{3,2}^{(M,N)}(x) = a_{1,3}^{(M,N)}(x) = a_{2,3}^{(M,N)}(x) = 0\), and therefore \(J_1^f(x, x') = J_1^p(x, x') = \emptyset\). Moreover, if \(x_1 = x'_1\) (i.e., \(1 \in I(x, x')\)), then since \(x_1 \leq x'_1\) we have that \(a_{j,1}^{(M,N)}(x) = x_1 \alpha_{j,1}(x_1) \leq x'_1 \alpha_{j,1}(x'_1) = a_{j,1}^{(M+1,N)}(x')\) for \(j = 2, 3\), and hence \(J_1^f(x, x') = \emptyset\). Likewise, since \(a_{j,1}(x_1) \geq a_{j,1}(x'_1)\) if \(x_1 \leq x'_1\), when \(x_j = x'_j\) (\(j = 2\) or \(j = 3\)) we have that
\[
\begin{align*}
a_{j,1}^{(M,N)}(x) &= x_j \alpha_{j,1}(x_1) \geq x'_j \alpha_{j,1}(x'_1) = a_{j,1}^{(M+1,N)}(x'),
\end{align*}
\] (3.8)

and hence \(J_2^p(x, x') = J_3^p(x, x') = \emptyset\).
By the above, for any $x \in S^{(M,N)}$ and $x' \in S^{(M+1,N)}$ such that $x \leq x'$, (3.6) and (3.7) become as follows.

(i) If $x_1 = x'_1$, then

$$
\lambda_1^{(M,N)}(x) + \sum_{j=2,3} a_{j,1}^{(M,N)}(x) \leq \lambda_1^{(M+1,N)}(x') + \sum_{j=2,3} a_{j,1}^{(M+1,N)}(x'),
$$

$$
\mu_1^{(M,N)}(x) \geq \mu_1^{(M+1,N)}(x').
$$

(ii) If $x_2 = x'_2$, then

$$
\lambda_2^{(M,N)}(x) \leq \lambda_2^{(M+1,N)}(x'),
$$

$$
\mu_2^{(M,N)}(x) + \|x_1 < x'_1\| a_{2,1}^{(M,N)}(x) \geq \mu_2^{(M+1,N)}(x') + \|x_1 < x'_1\| a_{2,1}^{(M+1,N)}(x').
$$

(iii) If $x_3 = x'_3$, then

$$
\lambda_3^{(M,N)}(x) \leq \lambda_3^{(M+1,N)}(x'),
$$

$$
\mu_3^{(M,N)}(x) + \|x_1 < x'_1\| a_{3,1}^{(M,N)}(x) \geq \mu_3^{(M+1,N)}(x') + \|x_1 < x'_1\| a_{3,1}^{(M+1,N)}(x').
$$

Note that for each pair $(x, x')$, depending on how many coordinates are equal, the number of inequalities to be fulfilled ranges from none to six. Below we check that (3.9)–(3.14) are indeed fulfilled as follows

(i) $x_1 = x'_1$, (3.9): from $\lambda_1^{(M,N)}(x) = \lambda_1(x_1) = \lambda_1(x'_1) = \lambda_1^{(M+1,N)}(x')$ and recalling (3.8), the inequality is easily established.

(ii) $x_1 = x'_1$, (3.10): $\mu_1^{(M,N)}(x) = x_1 \mu = x'_1 \mu = \mu_1^{(M+1,N)}(x')$.

(iii) $x_2 = x'_2$, (3.11): $\lambda_2^{(M,N)}(x) = \lambda_2(x_2) \|x_2 < x'_1\| \leq \lambda_2(x_2) \|x_2 < x'_1\| = \lambda_2^{(M+1,N)}(x')$

(iv) $x_2 = x'_2$, (3.12): dividing both sides by $x_2 (= x_2 > 0)$—if $x_2 = 0$, then both sides are zero and (3.12) is fulfilled as an equality—we obtain

$$
\beta_2(x_1) + \|x_1 < x'_1\| a_{2,1}^{(M,N)}(x_1) \geq \beta_2(x'_1) + \|x_1 < x'_1\| a_{2,1}^{(M+1,N)}(x'_1)
$$

If $x_1 = x'_1$, then the above inequality holds trivially. On the other hand, if $x_1 < x'_1$, then it follows from the fact that $\beta_2(x_1) + a_{2,1}(x_1)$ is nonincreasing in $x_1$ as $1 \geq (1 - f)P_{\infty} + f \geq P_{\infty}$.

(i) $x_3 = x'_3$, (3.13): $\lambda_3^{(M,N)}(x) = \lambda_3(x_1) \leq \lambda_3(x'_1) = \lambda_3^{(M+1,N)}(x')$.

(ii) $x_3 = x'_3$, (3.14): the same argument used to establish (3.12), with subscript 2 replaced by 3, can be applied here.

If the pair of processes under consideration are $X^{(M,N)}(t)$ and $X^{(N)}(t)$ (instead of $X^{(M,N)}(t)$ and $X^{(M+1,N)}(t)$), then the verification of the hypotheses runs as before, except for
checking the counterpart of (3.11), which has to be slightly modified as follows (recall that \( x_2 = x'_2 \) applies):

\[
\lambda_2^{(M,N)}(x) = \lambda_2(x_2)I_{|x_2| \leq M} \leq \lambda_2(x'_2) = \lambda_2^{(N)}(x').
\]  

(3.16)

On account of the abovementioned Statement, we have that

\[
X^{(M,N)}(t) \preceq_{\text{st}} X^{(M+1,N)}(t) \preceq_{\text{st}} X^{(N)}(t),
\]

where \( \preceq_{\text{st}} \) denotes the strong stochastic ordering (see the monographs in [13, 14] for more details).

Let \( \pi^{(M,N)}(x) \), \( \pi^{(N)}(x) \), and \( \pi(x) \) denote the stationary probabilities of \( X^{(M,N)}(t) \), \( X^{(N)}(t) \), and \( X(t) \). It is easy to check that

\[
\pi^{(M,N)}(x) = \overline{\pi}(x) + \overline{\pi}(x + e_1 + e_2) + \overline{\pi}(x + e_1 + e_3) + \overline{\pi}(x + e_2 + e_3)
-
\overline{\pi}(x + e_1) - \overline{\pi}(x + e_2) - \overline{\pi}(x + e_3) - \overline{\pi}(x + e_1 + e_2 + e_3),
\]

(3.18)

where \( \overline{\pi}(x) = \sum_{x_1=x_1}^{C} \sum_{x_2=x_2}^{M} \sum_{x_3=x_3}^{N} \pi(x') \).

By applying (3.17) and using (3.18) we can now proceed along the same lines that are in [2, Section 2.6.2] to conclude that \( \lim_{M \to \infty} \pi^{(M,N)}(x) = \pi^{(N)}(x) \).

Applying the same procedure as above (mutatis mutandis) to \((X^{(N)}(t), X^{(N+1)}(t))\) and \((X^{(N)}(t), X(t))\) would yield that \( \lim_{N \to \infty} \pi^{(N)}(x) = \pi(x) \). Indeed, the latter is in essence the case addressed in [2, Section 2.6.2] with a single truncation. Moreover, a repetition of the same arguments (with modifications affecting only minor details) would lead to identical results if the order in which the orbits are truncated is inverted or if we first truncate one of the orbits and then apply a generalized truncation to the other. Thus, we finally conclude that \( \lim_{M,N \to \infty} \pi^{(M,N)}(x) = \pi(x) \).

4. Conclusions

The retrial phenomenon has been thoroughly studied in the last decades. It has been shown that, in general, to solve this type of systems we must resort to approximate models. However, almost all the papers that present some approximate models make the assumption that the approximation results converge to the exact ones of the original model. In this paper, we demonstrate the convergence of a family of finite truncated models to their original counterparts in a two-reattempt-orbit multiserver retrial queue. Moreover, the followed methodology can be used as a framework to show the convergence of other approximate models, like the generalized truncated models.

Acknowledgment

This work was supported by the Spanish Government (30% PGE) and the European Commission (70% FEDER) through projects TSI2007-66869-C02-02 and TIN2008-06739-C04-02 and through the Network of Excellence “EuroNE, Anticipating the Network of the Future—From Theory to Design” (project reference 216366).
References
