Solving Multiserver Systems with Two Retrial Orbits Using Value Extrapolation: A Comparative Perspective

Mª Jose Domenech-Benlloch[1], Jose Manuel Gimenez-Guzman[2], Vicent Pla[1], Vicente Casares-Giner[1], and Jorge Martinez-Bauset[1]

[1] Dept. Comunicaciones, Universidad Politecnica Valencia
Cami de Vera, s/n 46022, València, Spain
[2] Dept. Automatica, Universidad de Alcalá
28871 Alcalá de Henares, Madrid, Spain
mdoben@doctor.upv.es, josem.gimenez@uah.es,
{vpla,vcasares}@dcom.upv.es, jmartinez@upvnet.upv.es

Abstract. In communication networks that guarantee seamless mobility of users across service areas, reattempts occur as a result of user behavior but also as automatic retries of blocked handovers. A multiserver system with two reattempt orbits is obtained when modeling these networks. However, an exact Markovian model analysis of such systems has proven to be infeasible and resorting to approximate methods is mandatory. To the best of our knowledge all the existing methods are based on computing the steady-state probabilities. We propose another approach based on the relative state values that appear in the Howard equations. We compare the proposed method with the most well-known methods appeared in the literature in a wide range of scenarios. The results of the numerical evaluation carried out show that this solution outperforms the previous approaches in terms of both accuracy and computation cost for the most common performance parameters used in retrial systems.

Keywords: Wireless and Mobile Systems and Networks (WLAN, 2G-3G-4G), Queueing Systems and Networks, Stochastic Models, Markov Models, Performance Modelling.

1 Introduction

The retrial phenomenon appears in multiple situations in telecommunications and computer networking. In this paper, we focus our attention on a generic communication network that guarantees seamless mobility to its customers by means of a cellular architecture. In this type of networks, the network coverage area is divided into cells and customers can move across different cells of the network. When a customer with an active communication moves from one cell to another, a so-called handover procedure is executed. Nowadays, perhaps the most widespread and popular example of this type of networks are the cellular telephone networks —2G and 3G— but the current perspective is that in near

K. Al-Begain, D. Fiems, and G. Horváth (Eds.): ASMTA 2009, LNCS 5513, pp. 56–70, 2009.

future a variety of technologies fitting into this category will be in place, e.g., Mobile IP, IEEE 802.16e —WiMAX— and IEEE 802.20 —Mobile Broadband Wireless Access, MBWA.

This paper deals with the case in which reattempts appear not only when a customer is blocked but also when a handover is blocked as in GSM [1]. To the best of our knowledge, the first and only paper that has considered the effect on network performance of both types of reattempts simultaneously is [2]. Now, in this paper, we refer to the former as redials and to the latter as (automatic) retrials, while we use the term reattempt to refer to any of them. Blocked handovers will be automatically retried until a reattempt succeeds or the user moves outside the handover area. In the former case the session will continue without the user noticing any disruption, while in the latter the session will be abruptly terminated. In contrast, persistence of redials depends on the user patience and an eventual abandonment results in session setup failure. 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. Therefore, both types of reattempts have different characteristics and as a consequence two separate retrial orbits have to be considered in the analysis of the system.

The modeling of repeated attempts has been the subject of numerous investigations. Two functional blocks are typically distinguished in models which consider reattempts: a block that accommodates the servers and possibly a waiting queue, and a block where users that reattempt are accommodated, usually called reattempt orbit. It is known [3] that to solve this type of systems it is necessary to resort to approximate methods. These methods are usually grouped into three categories: approximations, finite truncated methods and generalized truncated methods [3, 4]. We will direct our attention only to finite and generalized truncated methods. The finite truncated methods replace the original infinite state space by a finite one, where steady-state probabilities can be computed. On the other hand, generalized truncated methods replace the original infinite state space by another infinite but solvable state space. This last type of methods usually outperforms the other two types [4].

All the approaches presented in the literature so far rely on the numerical solution of the steady-state Kolmogorov equations of the *Continuous Time Markov Chain* (CTMC) that describes the system under consideration. Very recently, however, an alternative approach for evaluating infinite state space Markov processes has been introduced by Leino et al. [5]. The new method, named Value Extrapolation (VE), does not rely on solving the global balance equations, but considers the system in its Markov Decision Process (MDP) setting and solves the expected value from the Howard equations written for a truncated state space.

The main objective of this work is to tailor the VE method to a system with two reattempt orbits and compare its performance with the performance of other possible approximate methods. This performance evaluation is done in a cellular network scenario that guarantees seamless mobility to its users. We conclude

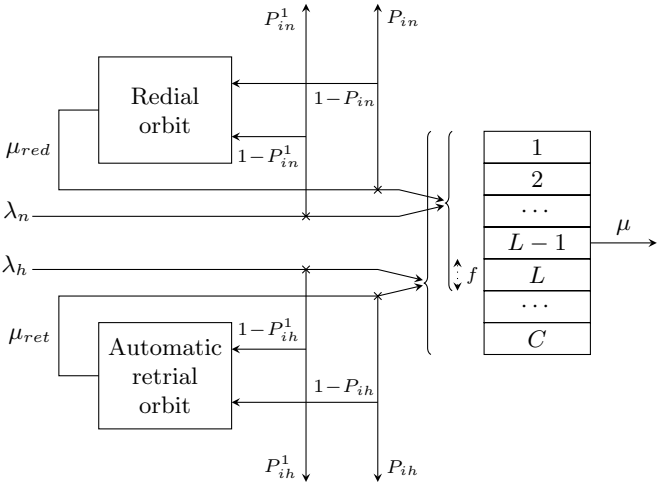

Fig. 1. System model

that VE greatly outperforms the rest of the methods throughout a wide range of scenarios not only in terms of accuracy, but also in terms of computation cost, so its use is highly recommendable.

The rest of the paper is structured as follows. First, we describe the cellular network under study and its associated model. In Section 3, we enumerate and explain the main features of the methods we compare VE with. Section 4 is devoted to the description of VE and how it has been applied to the model under study. A numerical study is performed in Section 5 and finally, a summary of the paper and some concluding remarks are given in Section 6.

2 System Description and Model

We consider a cellular mobile network with a fixed channel allocation scheme and where each cell is served by a different base station, where C is the number of resources in the cell. As shown in Fig. 1 there are two arrival streams: the first one represents new sessions and the second one handovers from adjacent cells. Both arrival processes are considered to be Poisson with rates λ_n and λ_h respectively. This leads to an overall arrival rate of $\lambda = \lambda_n + \lambda_h$. For the sake of mathematical tractability, the channel holding time is assumed to be exponentially distributed with rate μ [6].

In general, blocking a new session setup is considered to be less harmful than blocking a handover attempt. Therefore, we must include an admission control policy to guarantee the prioritization of handovers —and retrials— over new sessions —and their associated redials— and therefore, assure a certain degree of Quality of Service (QoS). The most widespread technique is to reserve some resources to highest priority flows, being in our case handovers and their associated

automatic retrials. This technique can be generalized to a fractional reservation, the so-called Fractional Guard Channel (FGC) admission control policy [7]. The FGC policy 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 $L = \lfloor t \rfloor$ 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, new sessions and redials are no longer accepted. Handovers and automatic retrials are only rejected when the system is completely occupied.

When an incoming new session is blocked, according to Fig. 1, it joins the redial orbit with probability $(1 - P_{in}^1)$ or leaves the system with probability P_{in}^1. If a redial is not successful, the session returns to the redial orbit with probability $(1-P_{in})$, redialing after an exponentially distributed time with rate μ_{red}. Redials are able to access to the same resources as the new sessions. Note that P_{in}^1 and P_{in} model the impatience phenomenon of leaving the system without having been served. Similarly, P_{ih}^1, P_{ih} and μ_{ret} are the analogous parameters for automatic retrials. There are several performance parameters that are generally used to describe the behavior of this type of cellular systems with retrials and redials. On the one hand, the widely used blocking probabilities for both new sessions (P_b^n) and handovers (P_b^h). On the other hand, the mean number of users redialing (N_{red}) and handovers retrying (N_{ret}) can describe more accurately the reattempt phenomenon.

The model considered can be represented as a tridimensional (k, m, o) CTMC, where k denotes the number of sessions being served, m specifies the number of sessions in the redial orbit and o represents the number of sessions in the retrial orbit. The state space can be represented by:

$$\mathcal{S} := \{(k, m, o) : k \leq C; m \in \mathbb{Z}_+; o \in \mathbb{Z}_+\}.$$

The transition rates of this model are represented in Table 1. The main mathematical features of this queueing model consist of having two infinite dimensions —the state space of the model is $\{0, \ldots, C\} \times \mathbb{Z}_+ \times \mathbb{Z}_+$— and the space-heterogeneity along them. This heterogeneity is produced by the retrial and redial rates, which respectively depend on the number of customers in the retrial and the redial orbits.

3 Solving Methods

It is known that the classical theory, see, e.g., [8], is developed for random walks on the semi-strip $\{0, \ldots, C\} \times \mathbb{Z}_+$ with infinitesimal transitions subject to conditions of space-homogeneity. When the space-homogeneity condition does not hold the problem of calculating the equilibrium distribution has not been addressed beyond approximate methods [9]. Indeed, if we focus on the simpler case of multiserver retrial queues with only one retrial orbit, the absence of closed form solutions for the main performance characteristics when $C > 2$ can be enphasized [3].

Table 1. Transition rates of the exact model

Transition	Condition	Rate
$(k,m,o) \rightarrow (k+1,m,o)$	$0 \leq k \leq L-1$	λ
	$k = L$	$\lambda_h + f\lambda_n$
	$L < k < C$	λ_h
$(k,m,o) \rightarrow (k+1,m,o-1)$	$0 \leq k \leq C-1$	$o\mu_{ret}$
$(k,m,o) \rightarrow (k,m,o-1)$	$k = C$	$o\mu_{ret}P_{ih}$
$(k,m,o) \rightarrow (k+1,m-1,o)$	$0 \leq k \leq L-1$	$m\mu_{red}$
	$k = L$	$m\mu_{red}f$
$(k,m,o) \rightarrow (k,m-1,o)$	$k = L$	$m\mu_{red}(1-f)P_{in}$
	$L < k \leq C$	$m\mu_{red}P_{in}$
$(k,m,o) \rightarrow (k-1,m,o)$	$1 \leq k \leq C$	$k\mu$
$(k,m,o) \rightarrow (k,m,o+1)$	$k = C$	$\lambda_h(1-P_{ih}^1)$
$(k,m,o) \rightarrow (k,m+1,o)$	$k = L$	$\lambda_n(1-P_{in}^1)(1-f)$
	$L < k \leq C$	$\lambda_n(1-P_{in}^1)$

Obviously, to solve the system under study, it will also be necessary to resort to approximate models and numerical methods of solution. Although other approaches exist, for the comparison against VE we have chosen the three most well-known methods that are able to solve the problem under study. These methods are explained in the next subsections.

3.1 Double Truncation (DT)

The easiest and more intuitive method to solve the proposed model lies in the truncation of the infinite dimensions of the state space [10]. In our case, it must be applied to both the redial and retrial orbits, truncating them beyond levels Q_n and Q_h respectively and obtaining the state space:

$$\mathcal{S} := \{(k,m,o) : k \leq C; m \leq Q_n; o \leq Q_h\}.$$

Obviously, by increasing the values of Q_n and/or Q_h the considered state space in the approximation is enlarged and the accuracy of the solution is expected to improve at the expense of a higher computational cost.

The stationary probability distribution can be obtained by solving $\boldsymbol{\pi}\mathbf{Q} = \mathbf{0}$ along with the normalization condition. As $\mathbf{Q}$ is a finite matrix this system can be solved by any of the standard methods defined in classical linear algebra [11].

3.2 Double FM (DFM)

As DT, DFM belongs to the family of finite truncated methods [3]. These methods consist of replacing the original infinite state space by a finite one. However,

DFM is more sophisticated than DT as it introduces in some sense the effect of the truncated states.

In [12] we developed FM, a generalization of the approximation method proposed in [13]. Although developed initially for a single orbit scenario, FM was applied to a system like the one under study in [14]. In this case FM has been applied to both retrial and redial orbits —resulting in DFM—, reducing the state space to a finite set by aggregating all states beyond a given occupancy of the orbits, producing the same approximate state space as DT:

$$\mathcal{S} := \{(k, m, o) : k \leq C; m \leq Q_n; o \leq Q_h\}.$$

where Q_n (Q_h) defines the occupancy from which the states in the redial (retrial) orbit are aggregated. In this case states of the form $(\cdot, Q_n, \cdot)$ represent the situation where at least Q_n users are in the redial orbit. Likewise the states of the form $(\cdot, \cdot, Q_h)$ represent the situation where there are Q_h or more users in the retrial orbit. Due to that aggregation two new parameters for each orbit are introduced. The parameter M_n denotes the mean number of users in the redial orbit conditioned to those states where there are at least Q_n users in the orbit, i.e., $M_n = E(m|m \geq Q_n)$. The probability that after a successful redial the number of users in the redial orbit does not drop below Q_n is represented by p_n. For the retrial orbit the parameters M_h and p_h are defined analogously.

The global balance equations, the normalization equation and equations for parameters M_n, p_n, M_h, p_h form a system of simultaneous non-linear equations, which can be solved using, for instance, the iterative procedure shown in [14].

3.3 Truncation and Generalization (TNR)

While the two previous approximations consider a finite truncated method for each retrial orbit, this method considers the use of a generalized truncated method in one of the two orbits. Obviously, we cannot use a generalized method for both orbits as the resulting model would not be solvable. For this reason, we have applied a generalized truncated method for the automatic retrial orbit and a Truncation (T) for the redial orbit. The method chosen for the retrial orbit is the method proposed by Neuts and Rao, denoted as NR, in [15]. This method is based on the homogenization of the model beyond a given level Q_h, which supposes to restrict the maximum automatic retrial rate, i.e.,

$$\mu_{ret}(o) = \begin{cases} o\mu_{ret} & \text{if } o < Q_h \\ Q_h\mu_{ret} & \text{if } o \geq Q_h \end{cases}$$

Therefore, the resulting state space is defined by

$$\mathcal{S} := \{(k, m, o) : k \leq C; m \leq Q_n; o \in \mathbb{Z}_+\}$$

With these two approximations we have to solve a system whose state space presents two finite dimensions and an infinite one, being the infinite dimension homogeneous beyond a given level Q_h. So, we can solve the resulting system and obtain the steady-state probabilities making use of the matrix-geometric solutions for stochastic models proposed by Neuts in [8].

4 Value Extrapolation

All the approximate methods described in the previous sections compute the steady-state probabilities using the balance equations. Very recently, however, an alternative approach for evaluating infinite state space Markov processes has been introduced by Leino et al. [5]. This approach, named Value Extrapolation (VE), does not rely on the probability of being in a certain state, but on a new metric called relative state values, that appear when we consider the system in its MDP setting. Formally, an MDP can be defined as a tuple $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}\}$, where $\mathcal{S}$ is a set of states, $\mathcal{A}$ is a set of actions, $\mathcal{P}$ is a state transition function and $\mathcal{R}$ is a revenue function. The state of the system can be controlled by choosing actions a from $\mathcal{A}$, influencing in this way the state transitions. The transition function $\mathcal{P} : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}_+$ specifies the transition rate to other states when a certain action is taken at a given state. The first characteristic of VE is the necessity of the definition of a revenue function that must be a function of the system state, i.e., $r(s)$. Following the definition of the revenue function for every state, we will also have a mean revenue rate of the entire process (r), which will be the performance metric we want to compute.

Once the MDP framework as well as the revenue function are specified, we are able to define the relative state values. It is obvious that after performing an action in state s the system will collect a revenue for that action ($r(s)$), but, as the number of transitions increases, the average revenue collected converges to r. The relative state value ($v(s)$) indicates the difference between the total revenue incurred when the system starts at state s and the total revenue incurred in a system for which the cost rate at all states is r. If we denote by t_n the time instants in which there is a change in the system state, then

$$v(s) = E\left[\sum_{n=0}^{\infty} (r(S(t_n)) - r) \Big| S(t_0) = s\right].$$

The equations that relate revenues, relative state values, and transition rates are the Howard equations defined by:

$$r(s) - r + \sum_{s'} q_{ss'}(v(s') - v(s)) = 0 \quad \forall s.$$

There will be as many Howard equations as number of states, $|\mathcal{S}|$. The number of unknowns will be the $|\mathcal{S}|$ relative state values plus the expected revenue r, i.e, $|\mathcal{S}| + 1$ unknowns. As only the differences in the relative values appear in the Howard equations, we can set $v(\mathbf{0}) = 0$, so we will have a solvable linear system of equations with the same number of equations as unknowns.

However, a finite number of Howard equations are needed to solve the system and, therefore, we need to truncate the state space to $\hat{\mathcal{S}}$. Whereas the traditional truncation consists of doing $q_{ss'} = 0 \quad \forall s' \notin \hat{\mathcal{S}}$, VE performs a more efficient truncation. Basically, VE considers the relative state values outside $\hat{\mathcal{S}}$

that appear in the Howard equations as an extrapolation of some relative state values inside $\hat{\mathcal{S}}$. The objective of VE is to find a function $f(s)$ that interpolates some points $(s, v(s))$ for $s \in \hat{\mathcal{S}}$ so that it approximates also $(s, v(s))$ for $s \notin \hat{\mathcal{S}}$. It is important to choose a fitting function, $f(s)$, that makes the Howard equations remain a closed system of linear equations. The most common fitting functions that acomplish that fact are the polynomials. We can use all $(s, v(s))$-pairs of the state space into the fitting procedure —global fitting— or only a subset $(\mathcal{S}_f)$ of them —local fitting. The choice of $\mathcal{S}_f$ will highly depend on the relative state value we want to extrapolate. Note also that function $f(s)$ and set $\mathcal{S}_f$ need to be chosen so that parameters have unambiguous values, i.e., in the case of choosing a polynomial as the fitting function, the number of different $(s, v(s))$-pairs in $\mathcal{S}_f$ has to be equal or greater than the number of coefficients in the polynomial. Note that if the relative values outside $\hat{\mathcal{S}}$ were correctly extrapolated, the results obtained by solving the truncated model would be exact.

4.1 Howard Equations of the System

To obtain the Howard equations for a certain state of the system under study, we can classify these states into four different cases depending on the number of sessions being served (k). We next describe such cases and their corresponding Howard equations.

1. $\mathbf{k < L}$: states in which both new sessions and handovers are accepted. The transition rates that go out from these states are represented in Fig. 2. Therefore, the Howard equations related to these states are:

$$r(k,m,o) - r + \lambda[v(k+1,m,o) - v(k,m,o)] + k\mu[v(k-1,m,o) - v(k,m,o)] +$$
$$+ m\mu_{red}[v(k+1,m-1,o) - v(k,m,o)] + o\mu_{ret}[v(k+1,m,o-1) - v(k,m,o)] = 0.$$

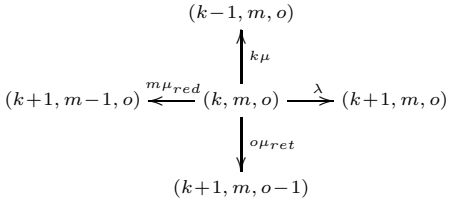

Fig. 2. Transition rates when $k < L$

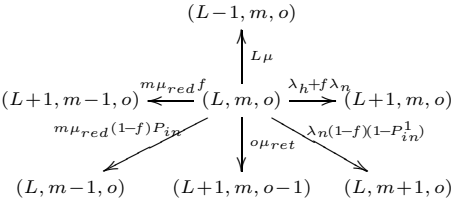

Fig. 3. Transition rates when $k = L$

2. $\mathbf{k = L}$: states in which handovers are accepted but new sessions are only accepted with probability $f = t - L$, where t is the parameter that characterizes the FGC admission control policy. Figure 3 represents the transition rates going out from these states, obtaining the next Howard equations:

$$r(L,m,o) - r + (\lambda_h + \lambda_n f)[v(L+1,m,o) - v(L,m,o)] +$$
$$+L\mu[v(L-1,m,o) - v(L,m,o)] + m\mu_{red} f[v(L+1,m-1,o) - v(L,m,o)] +$$
$$+m\mu_{red}(1-f)P_{in}[v(L,m-1,o) - v(L,m,o)] +$$
$$+o\mu_{ret}[v(L+1,m,o-1) - v(L,m,o)] +$$
$$+\lambda_n(1-f)(1-P_{in}^1)[v(L,m+1,o) - v(L,m,o)] = 0.$$

3. $\mathbf{L < k < C}$: states where handovers are accepted but new sessions are blocked, as shown in Fig. 4. That leads to the Howard equations:

$$r(k,m,o) - r + \lambda_h[v(k+1,m,o) - v(k,m,o)] + k\mu[v(k-1,m,o) - v(k,m,o)] +$$
$$+m\mu_{red}P_{in}[v(k,m-1,o) - v(k,m,o)] + o\mu_{ret}[v(k+1,m,o-1) - v(k,m,o)] +$$
$$+\lambda_n(1-P_{in}^1)[v(k,m+1,o) - v(k,m,o)] = 0.$$

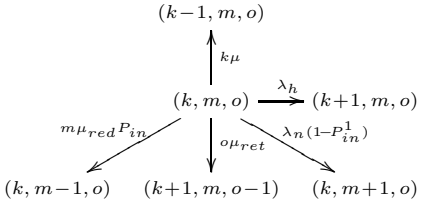

Fig. 4. Transition rates when $L < k < C$

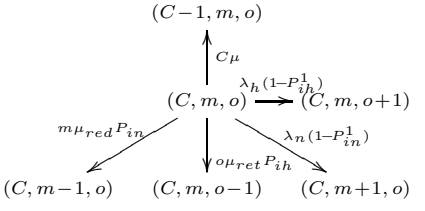

Fig. 5. Transition rates when $k = C$

4. $\mathbf{k = C}$: states where both new sessions and handovers are blocked, being the transition rates as shown in Fig. 5 and their corresponding Howard equations:

$$r(C,m,o) - r + \lambda_h(1-P_{ih}^1)[v(C,m,o+1) - v(C,m,o)] +$$
$$+C\mu[v(C-1,m,o) - v(C,m,o)] + m\mu_{red}P_{in}[v(C,m-1,o) - v(C,m,o)] +$$
$$+o\mu_{ret}P_{ih}[v(C,m,o-1) - v(C,m,o)] +$$
$$+\lambda_n(1-P_{in}^1)[v(C,m+1,o) - v(C,m,o)] = 0.$$

4.2 Revenue Function

As performance parameters are not computed from the steady-state probabilities as usual, it is important to explain more carefully how they are computed. For that purpose we must set the inputs $r(s)$ in the Howard equations properly in order to ensure that the revenue rate of the entire process r is equal to the performance parameter we want to compute. In a nutshell, r will be the parameter we want to compute if we let $r(s)$ to be the value of that parameter when the system is in state s. Table 2 gives several examples on how $r(s)$ can be set in order to obtain the performance parameters under study.

Table 2. Revenue function definition

Parameter	Value
P_b^h	$r(k,m,o) = 1$ for $k = C$, $\forall m$, $\forall o$
	$r(k,m,o) = 0$ otherwise
P_b^n	$r(k,m,o) = 1 - f$ for $k = L$, $\forall m$, $\forall o$
	$r(k,m,o) = 1$ for $k \geq L$, $\forall m$, $\forall o$
	$r(k,m) = 0$ otherwise
N_{ret}	$r(k,m,o) = o$ $\forall k$, $\forall m$, $\forall o$
N_{red}	$r(k,m,o) = m$ $\forall k$, $\forall m$, $\forall o$

4.3 Polynomial Fitting and Solution

Note that in the system under study the number of states is infinite because both m and o can take any value in $\mathbb{Z}_+$, thus some truncation is needed. We have made a truncation similar to DT and DFM, obtaining a truncated state space defined by:

$$\hat{\mathcal{S}} := \{s = (k,m,o) : k \leq C; m \leq Q_n; o \leq Q_h\}.$$

Therefore, in the system under study, we have truncated the state space beyond a value of Q_n (Q_h) for the occupancy of the redial (automatic retrial) orbit. However, in the Howard equations of the truncated state space, relative state values of some states appear that do not belong to the truncated state space, being $v(C,m,Q_h+1)$ $\forall m$ and $v(k,Q_n+1,o)$ for $k \geq L$ and $\forall o$. Therefore, we must extrapolate these two sets of states to obtain a closed system of equations. We have used a $(n-1)$-th degree polynomial that interpolates the n points in $\{(j,v_j)|v_j = v(C,m,j), \forall m, Q_h - n < j \leq Q_h\}$ to extrapolate $v(C,m,Q_h+1)$. To extrapolate $v(k,Q_n+1,o)$ for $k \geq L$ we interpolate the p points in $\{(i,v_i)|v_i = v(k,i,o), k \geq L, Q_n - p < i \leq Q_n, \forall o\}$. Note that including value extrapolation neither increase the computational cost nor the number of Howard equations, remaining in $|\hat{\mathcal{S}}| = (C+1) \times (Q_n+1) \times (Q_h+1)$.

After some algebra, and using the Lagrange basis to reduce the complexity of the procedure, we obtain a simple closed-form expression for the extrapolated value of both sets

$$v(C,m,Q_h+1)^{(n)} = \sum_{j=0}^{n-1} (-1)^j \binom{n}{j+1} v(C,m,Q_h-j), \forall m,$$

and

$$v(k,Q_n+1,o)^{(g)} = \sum_{i=0}^{g-1} (-1)^i \binom{g}{i+1} v(k,Q_n-i,o), k \geq L, \forall o.$$

5 Results and Discussion

In this section a number of numerical examples are presented with the purpose of illustrating the capabilities and versatility of our model and the analysis

methodology. The numerical analysis is also aimed at assessing a comparison between the proposed methodology and previous approaches not only in terms of accuracy but also in terms of computation cost.

For the numerical experiments a basic configuration of the system is used and then the different parameters are varied. Thus, unless otherwise indicated, the value of the parameters will be those of the basic configuration: $C = 10$, $t = 9$, $\mu = 1$, $P_{ih}^1 = P_{in}^1 = 0$, $P_{ih} = P_{in} = 0.2$, and $\mu_{red} = \mu_{ret} = 1$. The values of λ_n and λ_h have been modified by means of the system load $\rho = \lambda/C\mu$, being $\lambda = \lambda_n + \lambda_h$ and taking $\lambda_h = 2\lambda_n$ in all cases. It must be noted that, due to the introduction of the impatience phenomenon modeled by P_{in}^1, P_{in}, P_{ih}^1, and P_{ih}, we will be able to consider values of $\rho > 1$.

5.1 VE Performance

The objective of this section is to study the performance of different extrapolation polynomials in a wide range of scenarios. Obviously, as stated in Section 3, for the system under study we are not able to compute the exact values of the most common performance parameters. For this reason, the first step is to assume that the exact value can be obtained choosing increasing and sufficiently high values of the truncation level. More specifically, we ran all methods presented in Section 3 and VE until the value of all the performance parameters under study had stabilized up to the 8th decimal digit.

In the system under study, there are two different truncation levels that must be specified, namely Q_n and Q_h. The purpose of this study will be to determine the pair (Q_n, Q_h) that makes the cardinality of the problem $((C + 1) \times (Q_n + 1) \times (Q_h + 1))$ as small as possible while a certain accuracy criterion is met. To fulfil these requirements we must define a direction of search to determine the desired (Q_n, Q_h) pair.

To avoid an exhaustive search to determine (Q_n, Q_h) we have used an algorithm similar to the one proposed in [16]. Our algorithm increase (Q_n, Q_h) along the diagonal until we obtain a system that fulfils the desired accuracy and later we decrease both parameters separately following descendent directions of the coordinate axis and finally take the best solution in terms of the cardinality of the problem. The rationale behind this last movement for only one of the two parameters (Q_n or Q_h) is the fact that, generally, $Q_n \neq Q_h$, and this cannot be accomplished only with the diagonal movement, so the solution with this last movement improves the initial diagonal movement.

In Table 3 we show the minimum complexity of the problem needed to fulfil a relative error lower than 10^{-4} for parameters P_b^n and P_b^h, for different loads (ρ) and reattempt rates ($\{\mu_{red}, \mu_{ret}\}$) and for different orders of the extrapolation polynomial.

Note that VEx denotes the use of an extrapolation polynomial of order x. Note also that the numbers shown in each cell represent the product $(Q_n+1)\times(Q_h+1)$ which defines the complexity and it is denoted by Ω, although the cardinality of the problem should also include the factor $(C+1)$. However, we have omitted this factor as it is common to all cases. Therefore, the best order for the extrapolation

Table 3. Minimum Ω to obtain relative errors lower than 10^{-4} in P_b^n/P_b^h

μ_{red}, μ_{ret}	ρ	VE1	VE2	VE3	VE4	VE5	VE6
{1,1}	0.4	25/30	**12/12**	16/16	25/25	36/36	49/49
	0.8	144/144	49/72	64/72	49/**35**	**36**/36	49/49
	1.2	484/506	342/342	240/**36**	**98**/120	121/132	99/120
{2,0.5}	0.4	20/25	**12/12**	16/16	25/25	36/36	49/49
	0.8	130/90	45/55	56/64	**36/30**	**36**/36	49/49
	1.2	-/-	432/336	280/170	**99/136**	126/144	135/168
{0.5,2}	0.4	20/25	**12/12**	16/16	25/25	36/36	49/49
	0.8	160/160	66/110	80/100	56/49	**36/42**	49/49
	1.2	-/-	-/-	400/-	154/189	**144/187**	162/198
{0.5,0.5}	0.4	25/30	**9/9**	16/16	25/25	36/36	49/49
	0.8	224/160	100/121	90/100	48/**35**	**36**/36	49/49
	1.2	-/-	-/-	-/-	**168**/280	195/**196**	441/378

polynomial will be the one that has the lowest Ω, which is in bold in the table. Moreover, we denote by "-" those cases in which the computer could not obtain a result because of lack of memory[1].

From the results in Table 3 we can conclude that there is not a clear choice in the order of the extrapolation polynomial that is able to get the lowest Ω in all cases. Neither the lowest nor the highest orders offer the best results. When the load is not high ($\rho = 0.4$), VE2 offers the lowest complexities, due to the fact that VE3-VE6 offer the result of the minimum Ω they require to work, e.g., to extrapolate with VE4 at least $Q_n = Q_h = 4$ is needed and therefore, the minimum Ω required to use VE4 is $(4+1) \times (4+1) = 25$. When the retrial orbits are more heavily loaded, VE4 is a good choice, as it offers low values of Ω. Therefore, hereafter we will use the polynomial of order 4 (VE4) and we will simply denote it as VE.

5.2 Comparison among Different Methods

Accuracy: The objective of this section is to compare the performance of VE with DT, DFM, and TNR. In Table 4 we show the minimum values of Ω needed to obtain a relative error lower than 10^{-4} for N_{red}. The results for the rest of performance parameters have been omitted as N_{red} is usually the worst case for all methods and results are found to be qualitatively equivalent for all performance parameters. We show in bold the best results, i.e., those that offer the minimum complexity Ω. Results show that VE clearly outperforms classical methods as it needs a much lower value of Ω to achieve the desired accuracy in all the scenarios under study. Moreover, and what is probably more important, there are some scenarios where VE is the only method that is able to get a result due to the complexity of those scenarios produced by having low reattempt rates.

[1] Results have been obtained using Matlab running in an Intel Core 2 Quad Q6600 with 4GB RAM memory.

Table 4. Minimum Ω to obtain relative errors lower than 10^{-4} in N_{red}

	$\mu_{red}, \mu_{ret} = \{1, 1\}$					$\mu_{red}, \mu_{ret} = \{2, 0.5\}$					$\mu_{red}, \mu_{ret} = \{0.5, 5\}$					$\mu_{red}, \mu_{ret} = \{0.5, 0.5\}$				
ρ	0.4	0.6	0.8	1.0	1.2	0.4	0.6	0.8	1.0	1.2	0.4	0.6	0.8	1.0	1.2	0.4	0.6	0.8	1.0	1.2
DT	64	143	324	550	930	56	132	304	522	-	54	120	264	-	-	63	180	528	-	-
DFM	48	72	208	360	324	49	100	176	378	-	45	98	198	-	-	56	126	352	-	-
TNR	48	91	180	400	651	48	99	182	196	640	36	90	192	-	-	54	135	240	-	-
VE	**25**	**25**	**35**	**110**	**196**	**25**	**25**	**35**	**108**	**204**	**25**	**25**	**60**	**66**	**161**	**25**	**25**	**45**	**195**	**396**

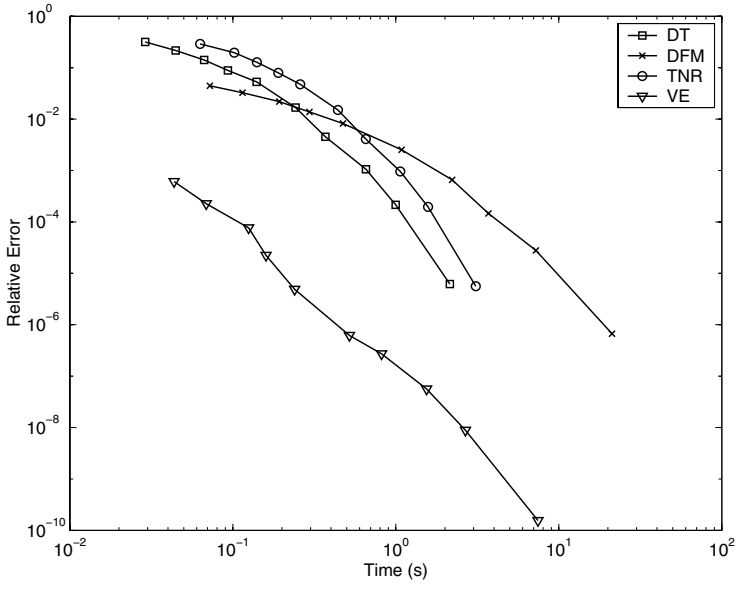

Fig. 6. Computation time for different methods

Computation cost: Although it is shown that VE clearly outperforms the other methods in terms of accuracy, it is also interesting to study their associated computation cost. From a practical perspective, it is more interesting to consider accuracy along with computation time. Figure 6 shows a joint representation of both parameters. As the figure shows, VE yields much higher accuracy than any other method for a given computation time. Results should be interpreted carefully, because computation cost highly depend on the algorithm used to solve the resulting system of equations. More concretely, in order to compute matrix $\mathbf{R}$ that appear in TNR we have used the logarithmic reduction algorithm as proposed in [17, Section 8.4], using a precision of 10^{-6} for the iterative procedure. Moreover, for solving the systems obtained with the DT, DFM, and TNR methods we have made use of the efficient algorithm described in [11] that takes advantage of the block-tridiagonal structure that presents the infinitesimal generator. Unfortunately, the linear system of equations obtained in VE has no

longer such a block-tridiagonal structure, and therefore, we must use a more general algorithm. More concretely, we have used LU factorization.

It can be seen that in the system under study the computation times needed for any of the methods are not very high from a human point of view. For that reason, the time results should be compared qualitatively, as the time units may be different from just seconds when we solve more complex systems or when we have to solve the basic retrial system several times —for example to balance the incoming handover rate to the outgoing handover rate, as shown in [18]—.

6 Conclusions

In mobile communication systems like cellular networks, Mobile IP or the recently defined IEEE 802.16e and IEEE 802.20 networks, mobile operators must guarantee seamless mobility to its customers. In these networks, repeated attempts occur due to user redials when their session establishments are blocked and also due to automatic retries when a handover fails. The Markovian model describing such a complex network is a multiserver retrial system that presents space-heterogeneity along two infinite dimensions. To the best of our knowledge, all the methods studied in the literature to solve these systems are based on their steady-state probabilities. In this paper, we propose an alternative method based on a different metric: the relative state values and the Howard equations that relate them.

We have compared the proposed method with the most well-known approaches appeared in the literature so far. The results show that the proposed method greatly outperforms previous approaches not only in terms of accuracy, but also in terms of computation cost. Moreover, we have shown that in some scenarios the proposed method is the only one that is able to guarantee a certain accuracy. For all those reasons the proposed method is highly recommendable to solve this type of systems.

Acknowledgements

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.

References

1. Mouly, M., Pautet, M.B.: The GSM system for mobile communications. Published by the authors (1992)
2. Onur, E., Deliç, H., Ersoy, C., Çaglayan, M.U.: Measurement-based replanning of cell capacities in GSM networks. Computer Networks 39, 749–767 (2002)
3. Artalejo, J.R., Pozo, M.: Numerical calculation of the stationary distribution of the main multiserver retrial queue. Annals of Operations Research 116(1–4), 41–56 (2002)

4. Domenech-Benlloch, M.J., Gimenez-Guzman, J.M., Pla, V., Martinez-Bauset, J., Casares-Giner, V.: Generalized Truncated Methods for an Efficient Solution of Retrial Systems. Mathematical Problems in Engineering, Article ID 183089 (2008)
5. Leino, J., Penttinen, A., Virtamo, J.: Flow level performance analysis of wireless data networks: A case study. In: Proceedings of IEEE ICC, vol. 3, pp. 961–966 (2006)
6. Khan, F., Zeghlache, D.: Effect of Cell Residence Time Distribution on the Performance of Cellular Mobile Networks. In: Proceedings of IEEE VTC 1997, pp. 949–953 (1997)
7. Ramjee, R., Nagarajan, R., Towsley, D.: On optimal call admission control in cellular networks. Wireless Networks Journal (WINET) 3(1), 29–41 (1997)
8. Neuts, M.: Matrix-geometric Solutions in Stochastic Models: An Algorithmic Approach. The Johns Hopkins University Press (1981)
9. Falin, G., Templeton, J.: Retrial Queues. Chapman & Hall, Boca Raton (1997)
10. Wilkinson, R.I.: Theories for toll traffic engineering in the USA. The Bell System Technical Journal 35(2), 421–514 (1956)
11. Servi, L.D.: Algorithmic solutions to two-dimensional birth-death processes with application to capacity planning. Telecommunication Systems 21(2–4), 205–212 (2002)
12. Domenech-Benlloch, M.J., Gimenez-Guzman, J.M., Martinez-Bauset, J., Casares-Giner, V.: Efficient and accurate methodology for solving multiserver retrial systems. IEE Electronic Letters 41(17), 967–969 (2005)
13. Marsan, M.A., Carolis, G.D., Leonardi, E., Cigno, R.L., Meo, M.: Efficient estimation of call blocking probabilities in cellular mobile telephony networks with customer retrials. IEEE Journal on Selected Areas in Communications 19(2), 332–346 (2001)
14. Gimenez-Guzman, J.M., Domenech-Benlloch, M.J., Pla, V., Casares-Giner, V., Martinez-Bauset, J.: Guaranteeing Seamless Mobility with User Redials and Automatic Handover Retrials. Journal of Universal Computer Science 14(10), 1597–1624 (2008)
15. Neuts, M., Rao, B.: Numerical investigation of a multiserver retrial model. Queueing systems 7, 169–190 (1990)
16. Artalejo, J.R., Pla, V.: On the impact of customer balking, impatience and retrials in telecommunication systems. Computers & Mathematics with Applications 57(2), 217–229 (2009)
17. Latouche, G., Ramaswami, V.: Introduction to Matrix Analytic Methods in Stochastic Modeling. In: ASA-SIAM 1999 (1999)
18. Marsan, M.A., Carolis, G.D., Leonardi, E., Cigno, R.L., Meo, M.: How many cells should be considered to accurately predict the performance of cellular networks? In: Proceedings European Wireless (1999)