Value Extrapolation Technique to Solve Retrial Queues: A Comparative Perspective

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

ABSTRACT—While the retrial phenomenon plays an important role in communication networks and should not be ignored, retrial systems do not present an exact analytic solution, so approximate techniques are required. To the best of our knowledge, all the existing techniques are based on computing the steady states probabilities. We propose another approach based on the relative state values which appear in the Howard equations. The results of the numerical evaluation carried out show that this solution outperforms previous approaches in terms of both accuracy and computation time.

Keywords—Retrial systems, performance evaluation.

I. Introduction

A common assumption when evaluating the performance of communication systems is that users who do not obtain immediate service leave the system without retrying. However, due to the increasing number of customers and increasing network complexity, customer behavior in general and the retrial phenomenon in particular may have a non-negligible impact on the system performance [1].

Queuing theory has been developed for random walks on the semi-strip $\{0,\cdots, C\}\times\boldsymbol{Z}_+$ with infinitesimal transitions subject to conditions of space-homogeneity [2]. However, retrial systems present an infinite and non-homogeneous state space; therefore, if the number of servers is higher than 2, the equilibrium distribution can only be computed by means of approximate techniques [3].

Existing approximate techniques can be classified into two categories [3]: finite truncated and generalized truncated techniques. Examples of the finite truncated approach are the technique proposed in [4] and its generalization in [5]. Those approximations are based on the reduction of the infinite state space to a finite one by aggregating states. On the other hand, generalized truncated techniques maintain the infinite state space, but they simplify it beyond a given level. This category includes the techniques proposed in [3], [6], and [7]. In this paper, we present a novel approach and compare it with all these previous proposals in terms of accuracy and computation time.

II. Value Extrapolation Technique

All the previously mentioned approaches rely on the numerical solution of the steady-state Kolmogorov equations of the continuous time Markov chain that describes the system under consideration. Very recently, however, an alternative approach for evaluating infinite state space Markov processes was introduced in [8]. The new technique, value extrapolation (VE), does not rely on solving the global balance equations but considers the system in its Markov decision process setting.

We developed the VE technique to solve a multiserver retrial system. In our model, users arrive according to a Poisson process with rate λ contending for access to a system with C servers. Users request an exponentially distributed service time with rate μ. When a new request finds all servers occupied, it joins the retrial orbit. After an exponentially distributed time of rate μ_r, this user retries. The retrial is successful if it finds a free server. We consider an infinite capacity for the retrial orbit. This model can be represented as a bidimensional continuous time

Manuscript received Jan. 16, 2008; revised Feb. 26, 2008.

This work was supported by the Spanish Government (30% PGE) and the European Commission (70% FEDER) through projects TSI2005-07520-C03-03 and TSI2007-66869-C02-02. Ma Jose Domenech-Benlloch was supported by the Spanish Ministry of Education and Science under the contract AP-2004-3332.

Jose Manuel Gimenez-Guzman (phone: + 34 963 877 767, email: jogiguz@upvnet.upv.es), Ma Jose Domenech-Benlloch (email: mdoben@doctor.upv.es), Vicent Pla (email: vpla@dcom.upv.es), Vicente Casares-Giner (email: vcasares@dcom.upv.es), and Jorge Martinez-Bauset (email: jmartinez@upvnet.upv.es) are with the Department of Comunicaciones, Universidad Politecnica Valencia, Valencia, Spain.

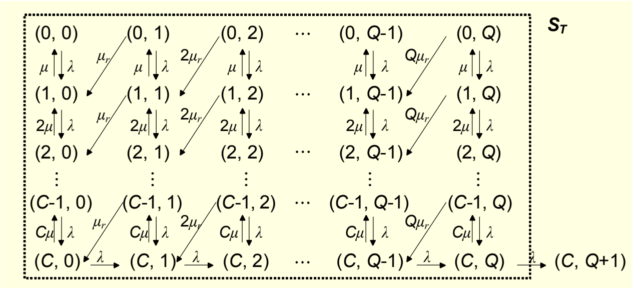

Fig. 1. Transition diagram.

Markov chain, $\boldsymbol{S}$:={s=(k, m): $k \le C$; $m \in \boldsymbol{Z}_+$}, where k is the number of users being served, and m is the number of users being served in the retrial orbit.

In a Markov decision process framework, after performing an action in state s, the system collects a revenue for that action ($r(s)$). As the number of transitions increases, the mean revenue collected converges to r, which is the mean revenue rate of the process. We choose the revenue function so that r yields the performance metric we want to compute. The relative state value ($v(s)$) tells how much greater the expected cumulative revenue over an infinite time horizon is when the system starts from the initial state s in comparison with r.

$$v(s) = E\left[\int_{t=0}^{\infty}\left(r(S(t)) - r\right)dt \,\middle|\, S(0) = s\right].$$

Revenues, relative state values, and transition rates from state s to s' ($q_{ss'}$) are related by the Howard equations [9]:

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

The number of states is infinite, as m can take any value in $\boldsymbol{Z}_+$; thus, we need to truncate the state space to $\boldsymbol{S_T}$:={s=(k, m): $k \le C$; $m \le Q$}. The transition diagram that is obtained is shown in Fig. 1. Unlike traditional truncation, that sets $q_{ss'}$=0 $\forall s' \notin \boldsymbol{S_T}$, VE performs a more efficient approximation by considering the relative state values outside $\boldsymbol{S_T}$ that appear in the Howard equations ($v(C, Q+1)$) as an extrapolation of some relative state values inside $\boldsymbol{S_T}$. It is important to choose an extrapolating function that makes the Howard equations remain a closed system of linear equations. The most common extrapolating functions that accomplish this are polynomials.

To extrapolate v_{Q+1}, we have used an (n–1)th degree polynomial that interpolates the n points in {(i, v_i) | v_i=$v(C, i)$, $Q-n<i \le Q$}. For example, when n=2, to extrapolate v_{Q+1}, we use points (Q, v_Q) and (Q–1, v_{Q-1}) to obtain the next interpolation polynomial, $v_x=(v_Q-v_{Q-1})x+v_Q(1-Q)+Qv_{Q-1}$. Making x=Q+1 to extrapolate the desired value, we obtain $v_{Q+1}=2v_Q-v_{Q-1}$. In general, using the Lagrange basis to reduce the complexity, and after some algebra, we obtain a simple closed-form expression for the extrapolated value:

$$v_{Q+1}^{(n)} = \sum_{k=0}^{n-1}(-1)^k \binom{n}{k+1} v_{Q-k} .$$

After several tests, we have observed that taking higher values of n improves the result, but the rate of improvement diminishes with n. In this sense, we have chosen n=9.

As we are interested in the blocking probability (P_b), the revenue function has been chosen from those states in which an attempt is blocked, namely, $r(C, m)$=1 $\forall m$, and zero elsewhere. Then we can compute P_b (r=P_b) solving the linear system of equations consisting of the $|\boldsymbol{S_T}|$ Howard equations.

III. Results and Discussion

We compare the proposed technique with those methods that we think have had more impact in the literature, namely, the proposals in [3], [5]-[7] referred to as AP, FM, Fal, and NR, respectively. In all cases the truncation/homogenization is done beyond level Q. We have studied all those techniques in a wide range of scenarios. Letting ρ=λ/($C\mu$), we have studied different system loads (ρ) by modifying λ and keeping C=10 and μ^{-1}=180 s. The retrial phenomenon has been configured with various μ_r. We have used the relative error of P_b, defined by ε_{Pb} =| $P_b^{approx}-P_b^{exact}$| / P_b^{exact} as a benchmark for comparing various techniques. To obtain an accurate estimate of P_b^{exact}, we ran all techniques, increasing the value of Q until the difference in P_b obtained by two consecutives values of Q was below 10^{-12}. As expected, all techniques converged to the same value in the performance parameter under study.

The minimum values of Q needed to obtain $\varepsilon_{Pb} < 10^{-8}$ for the different techniques are shown in Table 1, while Fig. 2 depicts the relative error as a function of Q. We conclude that AP is the best technique proposed so far (without taking into account VE). If we compare AP and VE, we conclude that VE greatly outperforms AP for low values of μ_r. For high values of μ_r, the results are not so clear: AP outperforms VE for low values of Q, but for high values, VE is the best choice.

From a practical perspective, it is perhaps more interesting to

Table 1. Minimum value of Q to achieve $\varepsilon_{Pb} < 10^{-8}$.

μ_r	0.01		0.1		1.0	
ρ	0.5	0.7	0.5	0.7	0.5	0.7
FM	24	43	19	44	21	43
Fal	25	48	21	39	17	32
NR	21	42	18	35	16	34
AP	22	40	**15**	27	**10**	**16**
VE	**15**	**22**	**15**	**26**	14	23
P_b	0.02523	0.15414	0.03245	0.20017	0.03558	0.21863

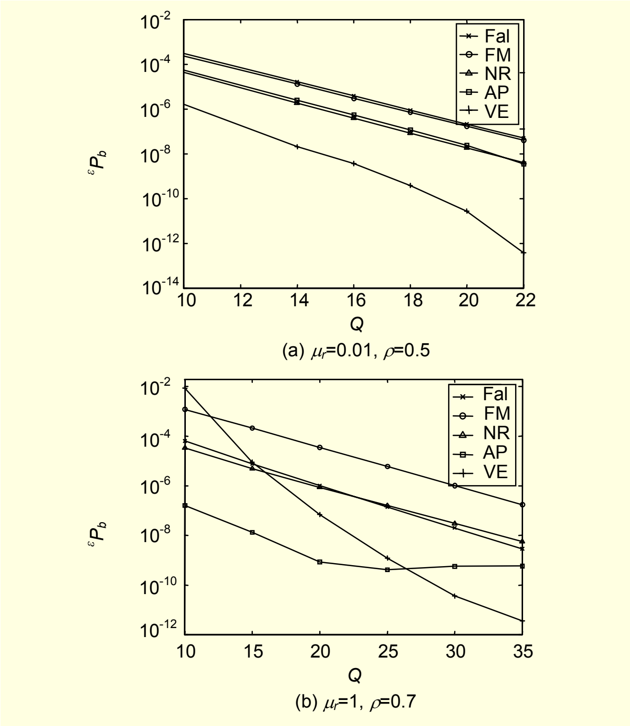

Fig. 2. Accuracy as a function of *Q*.

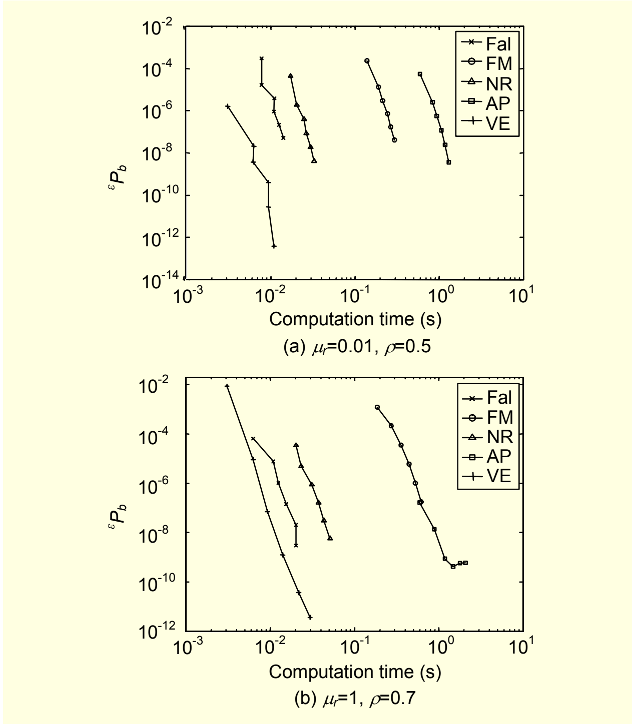

Fig. 3. Accuracy as a function of the computational time.

consider accuracy along with computation cost. We must highlight that the resolution methodology plays a fundamental role in this aspect. We have made use of the block tridiagonal structure of the techniques FM, NR, and Fal by using the algorithm proposed in [10]. To solve AP, we have used the algorithm proposed in [3], and to solve the system of equations in VE, we have used an LU factorization. Figure 3 shows a joint representation of accuracy and computation time using an Intel Pentium IV 3.2 GHz running Matlab. As the figure shows, VE yields much higher accuracy than any other technique for a given computation time. Therefore, the application of the VE approach can be strongly recommended, especially in those cases where computation time is a concern due to the size of the system and/or the number of times the basic algorithm has to be applied.

IV. Conclusion

In this paper we propose a novel technique, called VE, to solve retrial systems. It relies on the solution of the Howard equations instead of the balance equations. Results show that VE outperforms previous approaches in both accuracy and computation time, so its use is strongly recommended.

References

[1] P. Tran-Gia and M. Mandjes, “Modeling of Customer Retrial Phenomenon in Cellular Mobile Networks,” *IEEE J. on Selected Areas in Communications*, vol. 15, no. 8, Oct. 1997, pp. 1406-1414.

[2] G. Latouche and V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling*, ASA-SIAM, 1999.

[3] J.R. Artalejo and M. Pozo, “Numerical Calculation of the Stationary Distribution of the Main Multiserver Retrial Queue,” *Annals of Operations Research*, vol. 116, no. 1-4, 2002, pp. 41-56.

[4] M.A. Marsan, G.D. Carolis, E. Leonardi, R.L. Cigno, and M. Meo, “Efficient Estimation of Call Blocking Probabilities in Cellular Mobile Telephony Networks with Customer Retrials,” *IEEE J. on Selected Areas in Communications*, vol. 19, no. 2, 2001, pp. 332-346.

[5] M.J. Domenech-Benlloch, J.M. Gimenez-Guzman, J. Martinez-Bauset, and V. Casares-Giner, “Efficient and Accurate Methodology for Solving Multiserver Retrial Systems,” *IEE Electronics Letters*, vol. 41, no. 17, Aug. 2005, pp. 967-969.

[6] G.I. Falin, “Calculation of Probability Characteristics of a Multiline System with Repeat Calls,” *Moscow University Computational Mathematics and Cybernetics*, vol. 1, 1983, pp. 43-49.

[7] M.F. Neuts and B.M. Rao, “Numerical Investigation of a Multiserver Retrial Model,” *Queueing Systems*, vol. 7, no. 2, 1990, pp. 169-189.

[8] J. Leino and J. Virtamo, “An Approximative Method for Calculating Performance Measures of Markov Processes,” *Proc. VALUETOOLS*, 2006.

[9] H.C. Tijms, *Stochastic Models: An Algorithmic Approach*, John Wiley & Sons, 1994.

[10] D.P. Gaver, P.A. Jacobs, and G. Latouche, “Finite Birth-and-Death Models in Randomly Changing Environments,” *Advances in Applied Probability*, vol. 16, no. 4, 1984, pp. 715-731.