Proceedings of the 1994 Winter Simulation Conference
ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila

RESTART: A STRAIGHTFORWARD METHOD FOR FAST SIMULATION OF RARE EVENTS

Manuel Villén-Altamirano

División de Tráfico y Prestaciones
Telefónica I+D
Emilio Vargas, 6
28043-Madrid, SPAIN

José Villén-Altamirano

Departamento de Matemática Aplicada
Universidad Politécnica de Madrid
Carretera de Valencia, km. 7
28031-Madrid, SPAIN

ABSTRACT

This paper presents a method for fast simulation of rare events, called RESTART (REpetitive Simulation Trials After Reaching Thresholds). The method obtains dramatic computer time savings for an equal confidence of the results. The paper shows the efficiency as well as the generality and the straightforward application of the method, and provides several examples of applications in different fields.

1 INTRODUCTION

The estimation of rare event probabilities is of great importance in many fields. In the two fields in which the authors work, teletraffic and reliability, performance requirements of high speed data networks, computer systems and many other electronic equipment are expressed in terms of events with very low probabilities. Probabilities of the order of 10^{-10} are often used to specify losses due to traffic congestion in ATM networks or due to failures in ultrareliable systems.

Crude simulations are impracticable for estimating such small probabilities: to estimate probabilities of 10^{-10} with acceptable confidence would require the simulation of at least one thousand billion events (which corresponds to the occurrence of one hundred rare events). A computer time of 0.1 msec. per simulated event would lead to a total time of 3 years for the crude simulation.

Simulation acceleration methods are necessary to reduce the simulation time for accurate estimation of rare events. See Frost, LaRue and Shanmugan (1988) and Kleijnen (1993) for an overview of existing methods.

The method RESTART (REpetitive Simulation Trials After Reaching Thresholds), in its first one threshold version, was presented by Villén-Altamirano, M. and J. (1991a and 1991b). Its basic idea, also used by Bayes (1970) and by Hopmans and Kleijnen (1979), is the following: given the rare event A, the probability of which must be estimated, an event C is defined so that $C \supset A$ and $p\{A\} << p\{C\} << 1$. Thus,

$$p\{A\} = p\{C\} \cdot p\{A / C\}$$

In a crude simulation, $p\{C\}$ is quite better estimated than $p\{A/C\}$, since $p\{C\}$ is estimated from the whole simulation while $p\{A/C\}$ is estimated from the small portion of the simulation in which C occurs. In RESTART simulation, the estimation of $p\{A/C\}$ is improved by means of repeated simulation trials of the portion in which C occurs. It leads to a great increase of the confidence of $p\{A\}$ with a small additional cost in computer time.

RESTART was enhanced by Villén-Altamirano, Martínez-Marrón, Gamo and Fernández-Cuesta (1994) by defining multiple thresholds (i.e., multiple events C_i satisfying $C_1 \supset C_2 \supset ... \supset A$) in which simulation retrials are performed. The great efficiency of the method with one threshold was significantly improved with multiple thresholds.

Although the main aim of present paper is to make known the method in a new audience, it also presents original contributions, as the adaptation of RESTART to transitory state simulations, criteria to choose the parameters of the method when there are constraints and the results obtained in the application of RESTART to new case studies.

The rest of the paper is organized as follows: Section 2 describes the method in its multiple thresholds version, Section 3 shows the efficiency of the method and gives criteria to choose its parameters, Section 4 presents a simulation library (ASTRO) in which RESTART is implemented and, finally, application examples to teletraffic and reliability problems are provided in Section 5.

2 DESCRIPTION OF RESTART

2.1 General Description

In Figure 1 an example of system evolution in a crude simulation is given. The occurrence of event A, the probability of which must be estimated, has been associated with a system state parameter S taking values equal to or greater than L. Figure 2 represents an example of the same simulation using RESTART.

M events C_i (satisfying $C_1 \supset C_2 \supset \ldots \supset C_M \supset A$) are defined and associated with values of S equal to or higher than intermediate thresholds T_i ($M=3$ in Fig. 1,2). Two other kinds of events are defined:

B_i: instant of the transition from $\overline{C_i}$ to C_i;

D_i: instant of the transition from C_i to $\overline{C_i}$.

The use of RESTART involves the following procedure:

- When an event B_i occurs, the system state is saved;
- When an event D_i occurs the system state at last event B_i is restored and the interval $[B_i, D_i)$ is simulated again;
- The process mentioned above is repeated R_i times (number of retrials for threshold i), as illustrated in Figure 2. The starting event of each trial is always the same B_i, while the ending events are different,

  $D_{i1}, D_{i2}, \ldots, D_{iR_i}$;
- During one trial of level i, an event B_{i+1} may occur and R_{i+1} retrials of level $i+1$ would be made before the trial of level i finishes;
- When event D_{iR_i} occurs simulation continues in the usual way;
- The statistics should be modified accordingly.

The way to modify the statistics is a follows: If the estimator of the probability of the rare event in crude simulation is:

$$\hat{P} = \frac{N_A}{N}$$

where N_A is the number of events A occurred in the simulation and N is, for example, the total number of simulated events, the estimator with RESTART should be:

$$\hat{P} = \frac{N_A}{N \prod_{i=1}^{M} R_i}$$

where N_A includes all the events A occurred in all retrials, while N only includes the events simulated in the first trial of each set of retrials.

2.2 Application to Transitory State Simulations

In some studies, the objective is not to evaluate the steady state probability of the rare event A, but the probability of occurrence of event A (or mean number of occurrences of event A) in a period $[t_o, t_e)$ characterized by a certain systen state in the instant t_o. With crude simulation, the system is repeatedly simulated in the interval $[t_o, t_e)$.

RESTART may also be used in this transitory state simulations. Events C_i and B_i are defined as in the steady state case. The definition of event D_i is different: D_i occurs either when there is a transition from C_i to

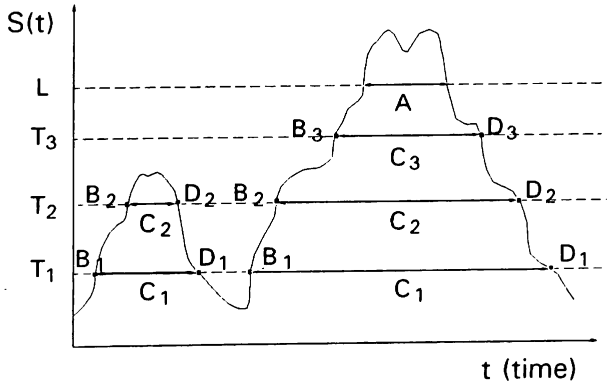

Figure 1. Crude Simulation

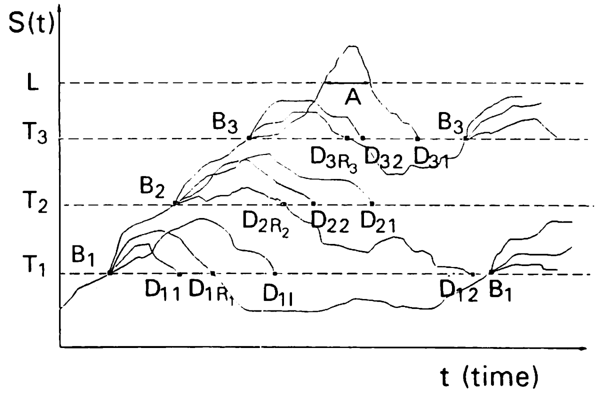

Figure 2. Simulation with RESTART

$\overline{C_i}$ or when time is equal to t_e. Observe that the simulation of a period $[t_o, t_e)$ is completed when time is equal to t_e and there is not any pending $[B_i, D_i)$ retrial.

In some applications, the state variable S is monotonously increasing, e.g., as in non-reparable reliability systems in which S is defined as the number of failed components. In these cases, D_i only occurs at time t_e. Figure 3 shows an example of this case.

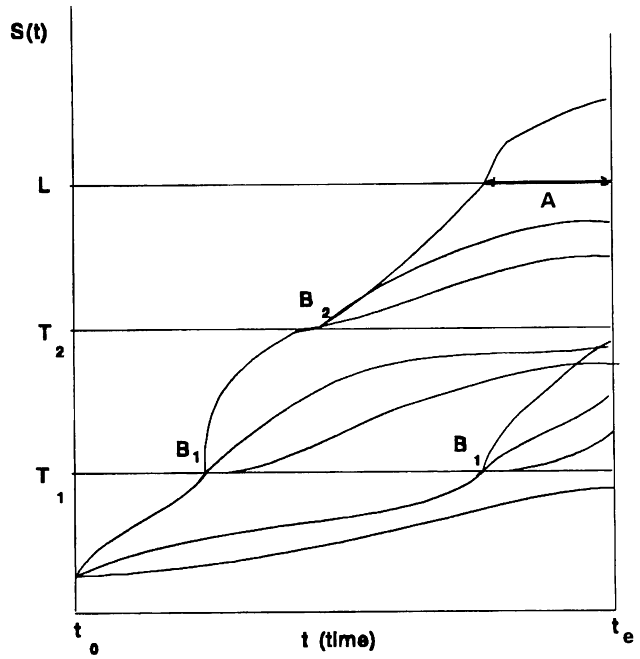

Figure 3. Transitory State Simulation with RESTART

3 EFFICIENCY OF THE METHOD

3.1 General Expression of the Gain

The gain obtained with the use of RESTART with respect to a crude simulation can be defined as:

$$G = \frac{V(\hat{P})_{(C)} \cdot C_{(C)}}{V(\hat{P})_{(R)} \cdot C_{(R)}}$$

where:

- $V(\hat{P})$ is the variance of the rare event A probability estimator $\hat{P}$;
- C is the cost of the simulation in terms of CPU time;
- Subindices $_{(C)}$ and $_{(R)}$ indicate Crude and RESTART simulations respectively.

Expresions for $V(\hat{P})_{(C)}$, $C_{(C)}$, $V(\hat{P})_{(R)}$ and $C_{(R)}$ are derived by Villén-Altamirano et al. (1994), leading to the following general expression of the gain:

$$G = \frac{1}{\dfrac{1}{\prod_{i=1}^{M} R_i} + P.\sum_{i=1}^{M} \dfrac{s_i R_i}{\prod_{j=1}^{i} P_j R_j}} \cdot \frac{1}{1 + \sum_{i=1}^{M}\left(y_i \prod_{j=1}^{i} P_j R_j \right)} \quad (1)$$

where:

- M: number of thresholds;
- $P = p\{A\}$;
- $P_1 = p\{C_1\}$;
- $P_i = p\{C_i / C_{i-1}\} \qquad (1 < i \leq M)$;
- $P_{M+1} = p\{A / C_M\}$;
- $R_i (1 \leq i \leq M)$: number of retrials from threshold i;
- $s_i (1 \leq i \leq M)$: inefficiency factor affecting $V(\hat{P})_{(R)}$;
- $y_i (1 \leq i \leq M)$: inefficiency factor affecting $C_{(R)}$.

As described by Villén-Altamirano et al. (1994), the factor s_i is mainly due to the correlation between the retrials made from the same state B_i. It mainly depends on the characteristics of the simulated system; nevertheless, it may be reduced by rescheduling, in each retrial, the events which had been scheduled when event B_i ocurred. In most applications, s_i varies between 1 and 10.

The factor y_i is due to the cost of restoring the system state of event B_i after each retrial $[B_i, D_i)$. It is equal to the ratio of the time required to simulate an interval $[B_i, D_i)$ including the restoration of B_i to the time required without including the restoration. In the simulation of complex systems, it may be worth to use hysteresis to reduce this cost. As explained in Villén-Altamirano, M. and J. (1991b), an appropriate choice of the hysteresis threshold can reduce the value of y_i to the square root of its value without hysteresis.

For extension, s_{M+1} and y_o are defined, being $s_{M+1} = y_o = 1$.

3.2 Optimal Gain

Given the difficulty of estimating the values of the inefficiency factors s_i and y_i, they will be ignored to obtain the optimal values of P_i and R_i which maximizes the gain. Thus, assuming in (1) that $s_i=y_i=1$, the optimal values obtained are:

$$P_i = e^{-2} \; ; \; R_i = e^2$$
$$M \; such \; that \; C_M \supset A \supset C_{M+1} \qquad (2)$$

This solution presents the following remarkable properties:

- The values of P_i and R_i are absolutely independent of characteristics of the system leading to straighforward application to the simulation of any system.
- It is also independent of P: RESTART can be very useful even in simulations in which the order of magnitude of the rare event probability is unknown as well as in the estimation of the whole distribution of a random variable .

The gain obtained using the parameter values given in (2) is:

$$G = \frac{4}{syPe^2\left[\ln\frac{1}{P} - 2(1-\frac{1}{s})\right]\left[\ln\frac{1}{P} - 2(1-\frac{1}{y})\right]} \approx$$
$$\approx \frac{4}{syPe^2(\ln P)^2} \qquad (3)$$

where:

$$s = \frac{1}{M}\sum_{i=1}^{M} s_i \; ; \quad y = \frac{1}{M}\sum_{i=1}^{M} y_i$$

The dramatical gain obtained with RESTART can be appreciated in Table 1, where simulation times with crude and RESTART simulations, for the same confidence of the results, are shown. (3 years for crude simulation with $P=10^{-10}$, as in the example of section 1, has been taken as reference).

Table 1: Simulation Times Using Crude and RESTART Simulations.

P	Crude Simulation	RESTART Simulation	
		sy=1	sy=10
10^{-4}	100 sec.	2 sec.	12 sec.
10^{-7}	28 hours	5 sec.	40 sec.
10^{-10}	3 years	10 sec.	90 sec.
10^{-100}	$3\text{x}10^{90}$years	16 min.	160 min.

An important remark is that, while simulation cost, for the same confidence of the results, grows linearly with $1/P$ in crude simulation, it grows linearly with $(\ln P)^2$ if RESTART is used. This property extends the applicability of simulation to any practical problem that can be envisaged regardless the value of the probability to be estimated. As can be seen in Table 1, probability values of 10^{-100} or even lower can be estimated within a reasonable time.

3.3 Choice of Parameters when There Are Constraints

In some applications, the system state variable S chosen to define the thresholds T_i can only take few discrete values. It occurrs, for example, in realibility problems in which the state variable S is defined as the number of devices in failure. In this case, the thresholds T_i cannot be adjusted to obtain the values of P_i given by (2).
For these applications, we must know the optimal number of retrials R_i for predetermined values of P_i. From (1), assuming again $s_i=y_i=1$, the optimal values obtained are:

$$R_i = \frac{1}{\sqrt{P_i \cdot P_{i+1}}} \qquad (4)$$

The gain achieved in this case is:

$$G = \frac{1}{P\left(\sum_{i=1}^{M+1}\frac{s_i}{\sqrt{P_i}}\right)\left(\sum_{i=1}^{M+1}\frac{y_i}{\sqrt{P_i}}\right)} \approx \frac{1}{syP\left(\sum_{i=1}^{M+1}\frac{1}{\sqrt{P_i}}\right)^2} \qquad (5)$$

Formula (5) provides criteria to choose the best values of P_i in cases in which, although there are some constraints in the choice, they are not totally predetermined. The best values of P_i are those which

$$Minimize \sum_{i=1}^{M+1} \frac{1}{\sqrt{P_i}} \quad , being \prod_{i=1}^{M+1} P_i = P \qquad (6)$$

If there are not any constraints for choosing P_i, the optimal choice is $P_i=e^{-2}$, which leads to the gain given by (3). In case of constraints it is derived from formula (6) that, for each i, P_i should be chosen as close as possible to e^{-2}, regardless how close to e^{-2} could be chosen for other values of i.

The above criteria are useful to maximize the gain. Nevertheless, they have not to be strictly followed, since huge gains are obtained even for values of P_i far from e^{-2}. For example, for $P=10^{-10}$ and $sy=10$, the optimal gain, corresponding to $P_i=e^{-2}$, is 1.02×10^6, and it is only reduced at most to 0.45×10^6 if the values of P_i are between 1/2 and 1/80. It shows that the choice of the parameter values is not critical, property which makes the use of RESTART simple and robust.

4 ASTRO: A GENERAL PURPOSE IMPLEMENTATION OF RESTART

An important feature of RESTART is that it has a straighforward application to each particular simulation in a way quite independent of the characteristics of the simulated system. This feature has allowed to implement RESTART in a general simulation library, called ASTRO (Advanced Simulation Tool with Restart Optimization), which has been developed in Telefónica I+D.

ASTRO provides an environment that allows a transparent and efficient use of RESTART. Functions such as monitorization of the state variable, retrial control, memory management, event rescheduling and weighting of statistics have been implemented in the library in a way which is general for any application. Only some few special actions, such as specifying the state variable to be monitored and assigning values to the thresholds and to the number of retrials, have to be taken by a user of ASTRO to use RESTART in a particular application. See Villén-Altamirano et al. (1994) for more details.

5 APPLICATION EXAMPLES

RESTART can be applied to a wide variety of problems in which the probability of a rare event has to be evaluated by simulation. The authors have applied it successfully into the two fields of their work, teletraffic and reliability. This section shows some examples of application. In three of the examples analytically tractable models have been chosen in order to compare simulation with analytical results.

Example 1

It is an ATM multiplexer for which we want to estimate the queue length distribution. The input process is binomial: superposition of 80 Bernouilli sources (each one with p=0.01) representing a total load of 0.8 Erl.

For RESTART simulation, 11 thresholds were defined on the multiplexer queue length. The threshold values and the number of retrials were assigned according to formula (2). Last threshold corresponds to a queue length q=51 and a probability $p=2.4 \times 10^{-4}$. Thus, accurate results can be expected down to probabilities about $p=3 \times 10^{-11}$. The arrival of 10 million cells were simulated both for crude and RESTART simulations. Figure 4 and 5 show analytical, crude and RESTART

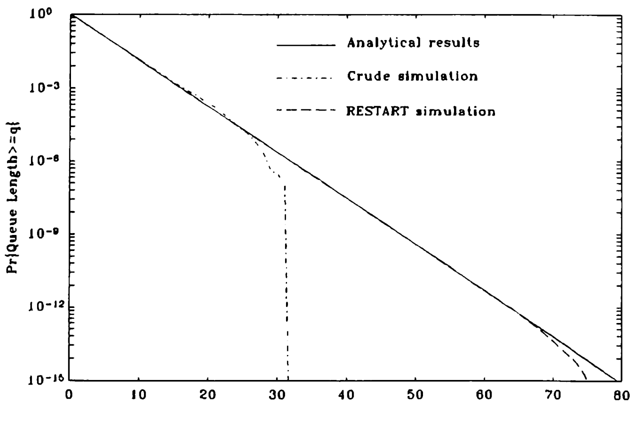

Figure 4. Queue Length Distribution in the ATM Multiplexer

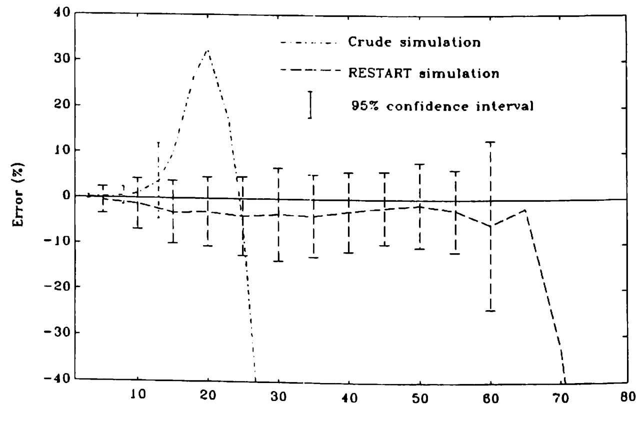

Figure 5. Relative Errors of the Simulations of the ATM Multiplexer

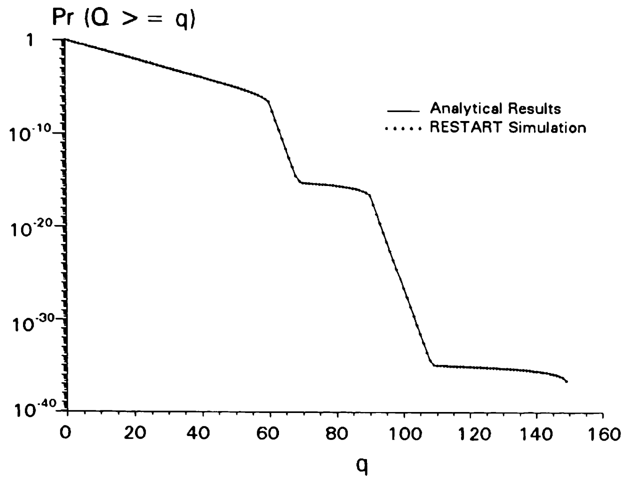

Figure 6. Queue Length Distribution of the M(s)/M/1/N System

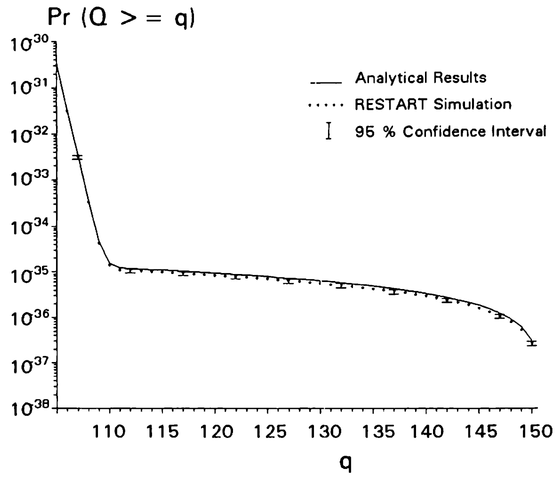

Figure 7. Queue Length Distribution of the M(s)/M/1/N System (Lower Part)

simulation results. Figure 4 shows the absolute results, while figure 5 highlights the difference between RESTART and normal simulations showing the errors of both of them with respect to analytical results. As expected, RESTART simulation provides accurate results (confidence interval smaller than 10%) up to q=56 (p=2.6×10^{-11}). This accuracy is achieved with crude simulation only for queue lengths shorter than q=14 (probabilities higher than p=2.4×10^{-3}). Since the simulation time in crude simulation, for the same confidence of the results, grows linearly wiht *1/p*, we can estimate that the crude simulation should have been 9.2×10^{7} times longer to obtain 10% accuracy for p=2.6×10^{-11}. Thus, the gain obtained is 9.2×10^{7}, similar to the theoretical one expected from (3): 3.5×10^{7} (for s=1 and y=1)

Example 2

The aim of the second example is to show how the extremely low value of the probability to be estimated is not a limitation for RESTART. Probabilities down to 10^{-37}, lower than those which can be envisaged in present practical problems, are estimated in this example. The system is an *M(s)/M/1/N* queue, where *M(s)* indicates a Markovian arrival with state-dependent rate, and N=150 places in the queue. 36 thresholds where defined on the queue length, with $P_i \approx 0.1$ and $R_i = 10$. The arrival of 36 million customers were simulated. Figure 6 shows the analytical and RESTART simulation results. The accuracy of the results (confidence interval smaller than 11% in the whole range) makes difficult to appreciate the differences in this figure. Thus, a zoom of lower part of Figure 6 is shown in Figure 7, in which we can better observe the small differences obtained for this low range of probabilities.

Example 3

This example is a reliability reparable system with 10 components, 3 cold spares and 1 reparation service. Times of component failure and of reparation are exponentially distributed with rates λ and μ respectively. The system fails when a component fails and there are not spares.

Three thresholds are defined corresponding to *S*=1, 2 or 3 failed components. The unavailability is given by the fraction of time in which *S*=4.

Crude and RESTART simulations were made for μ / λ varying from 100 to 1000, which corresponds to system unavailabilities from 9.0×10^{-5} to 9.9×10^{-9}. The CPU time of each simulation case was 30 minutes. Figure 8 shows the relative errors of both types of simulations with respect to analytical results. It can be observed that

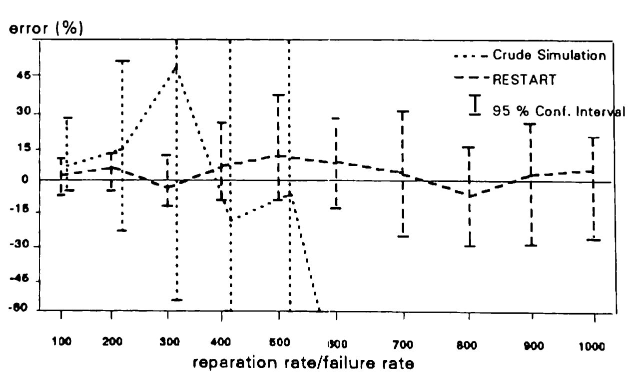

Figure 8. Relative Errors of the Simulations of the Reparable System

the accuracy of RESTART decreases very slowly for lower probabilities, being acceptable in all the cases. However, the accuracy of crude simulation is always worse, being unacceptable for μ / λ = 300 or greater, which corresponds to unavailabilities equal or greater than 10^{-6}.

Example 4

The two nearly-independent triads system is a non-reparable ultrareliable system studied by Geist (1994). There are two subsystems each with 3 components and the system fails when there are two failed components in each subsystem or when there is a nearly-coincident failure of one component of each triad. The objective is to evaluate the unreliability of the system in a period [*to, te*).

Three thresholds were defined corresponding to 1, 2 or 3 failed components.

Note that system failure due to a two-component nearly coincident failure occurs below the third threshold. Consequently, the number of occurrences of this type of system failure must not be divided by R3 for estimating its probability.

For different component failure rates, crude and RESTART simulations long enough to obtain accurate unreliability estimations (confidence intervals smaller than 15%) have been made. Table 2 shows the results obtained and the required CPU time. Observe that for estimating unreliabilities of the order of 10^{-9}, the gain achieved with RESTART is 100,000, in spite of the existing constraints to define the thresholds.

Table 2: Unreliability of Two Nearly-Independent Triads

Failure rate	Unreliability	Simulation time	
		Crude	RESTART
0.075	2.2×10^{-4}	2 hours	4 min.
0.030	7.5×10^{-6}	70 hours	30 min.
0.015	4.6×10^{-7}	50 days(*)	70 min.
0.003	7.8×10^{-10}	74 years(*)	6 hours

(*) Estimated values

6 CONCLUSIONS

RESTART is a simulation acceleration method that obtains dramatic time savings in the simulation of rare events. The gain obtained is more significant as the probability *P* of the rare event is lower: while computer time needed with crude simulation is proportional to *1/P*, with RESTART, it is proportional to $(ln\ P)^2$. This property allows to estimate probability values of 10^{-100} or even lower within a reasonable time.

The generality of the method allows its application to a large variety of problems. It has already been succesfully applied to several teletraffic and reliability problems, and it can also be applied to other fields.

Another important feature of RESTART is its straightforward application to each particular case in a way quite independent of the characteristics of the simulated system. The analysis required to choose the parameters of the method is very simple and not critical, since the method is robust against a non-optimal choice.

Its generality and its striaghtforward application has allowed to implement RESTART in a general purpose simulation library (ASTRO) which performs most RESTART procedures in a way which is general for any application.

REFERENCES

Bayes, A.J. 1970. Statistical techniques for simulation models. *The Australian Computer Journal* 2: 180-184.

Frost, V., W.W. LaRue, and K.S. Shanmugan. 1988. Efficient techniques for the simulation of computer communications network. *IEEE Journal on Selected Areas in Communications* 6: 146-157.

Geist, R. 1991. Extended behavioral decomposition for estimating ultrahigh reability. *IEEE Transactions on Reliability* 40: 22-27.

Hopmans, A.C.M., and J.P.C. Kleijnen. 1979. Importance Sampling in system simulation: A practical failure?. *Mathematics and Computing in Simulation XXI*. North-Holland Publishing Company 209-220.

Kleijnen, J. 1993. *Simulation, a statistical perspective.* New York: Wiley.

Villén-Altamirano, M., and J. Villén-Altamirano. 1991a. RESTART: A method for accelerating rare events simulations. In *Proceedings of the 13th International Teletraffic Congress, Queuing performance and control in ATM.* Ed. J.W. Cohen, and C.D. Pack, 71-76, North-Holland.

Villén-Altamirano, M., and J. Villén-Altamirano. 1991b. Accelerated simulation of rare events using RESTART method with hysteresis. In *Proceedings of the ITC Specialists Seminar, Telecommunications Services for Developing Economies,* Ed J. Filipiak, 240-251, The University of Mining and Metallurgy, Krakov.

Villén-Altamirano M., A. Martínez-Marrón, J. Gamo, and F. Fernández-Cuesta. 1994. Enhancement of the accelerated simulation method RESTART by considering multiple thresholds. In *Proceedings of the 14th International Teletraffic Congress, The fundamental role of teletraffic in the evolution of telecommunications networks,* Ed J. Labetoulle, and J.W. Roberts, 787-810, Elsevier.

AUTHOR BIOGRAPHIES

MANUEL VILLÉN-ALTAMIRANO is Head of the Traffic Performance Division of Telefónica I+D, Madrid. He received the M.S. degree in telecommunication engineering from Politechnical University of Madrid in 1970. He is Rapporteur for "Traffic engineering for broadband ISDN" in the International Telecommunication Union. His research interests are focused on performance analysis and traffic management of telecommunication networks.

JOSÉ VILLÉN-ALTAMIRANO is Professor, Head of the Department of Applied Mathematics at the Politechnical University of Madrid. He received M.S. degree in Mathematics from Complutense University of Madrid in 1977 and he received Ph. D. degree in Computer Science from Polithecnical University of Madrid in 1988. His research interests are focused on reliability, queueing theory and simulation.