This paper presents a method for accelerating simulations to estimate the probability of occurrence of rare events. The method, called RESTART (REPetitive Simulation Trials After Reaching Thresholds), is quite general and has a straightforward application, allowing dramatic reductions of the simulation time for an equal confidence of the results. The paper proves the efficiency of the method and shows an application example.

1. INTRODUCTION

In modern and future telecommunication networks, requirements on network performance are usually very stringent. Performance indicators defined as the probability of occurrence of certain events are specified by extremely low values; e.g., in asynchronous transmission mode (ATM), the cell loss probability should be less than $10^{-7}$.

Normal simulations are impracticable for estimating such small probabilities: in the previous example, a billion cells have to be simulated per each lost cell. To estimate the probability of such rare events with acceptable confidence would require the simulation of at least one hundred billion cells; assuming a computer time of 0.1 msecs per simulated cell, a computer time of 115 days would be necessary. This huge amount of time makes normal simulations impracticable for this purpose.

Some techniques are described in literature for accelerating those rare event simulations; importance sampling methods [1, 2, 3] and reverse time models [4, 5] are the best known. Both techniques require a complex analysis for its application to particular cases.

The RESTART (REPetitive Simulation Trials After Reaching Thresholds) method, presented in this paper, has a very simple conception and a straightforward application to each particular case. It is based on the following idea: Given the rare event $A$, the probability of which must be estimated, an event $C$ is defined so that $C \Rightarrow A$ and $P(C) > P(A)$. The probability of occurrence of event $A$ can be written as:

$$P(A) = P(C) \cdot P(A/C)$$  \hspace{1cm} (1.1)

In a simulation, $P(C)$ is normally 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. Thus, if this portion of the simulation in which $C$ occurs is increased (e.g., by many repetitive simulation trials), $P(A/C)$ and, consequently, $P(A)$ will be estimated with much better confidence. This great increase of the confidence of $P(A)$ is made with a small cost in computer time since only a small portion of the simulation is repeated. As the paper will show, this simple method allows a dramatical reduction of the simulation time for an equal confidence of the results or, vice versa, a dramatical increase of the confidence for a same simulation time.

The paper is organized as follows: the RESTART method is described in section 2 and analyzed in section 3. Based on this analysis, section 4 provides rules to optimize its parameters, section 5 reports the gain obtained with its application, and section 6 gives an application example.

2. DESCRIPTION OF RESTART

Let us see in figure 1 an example of system evolution along time. In this example, the occurrence of event $A$, the probability of which must be estimated, has been associated, for a better understanding, with a system state parameter $S$ taking values equal to or higher than a certain value. An event $C \Rightarrow A$ has been defined and associated with values of $S$ equal to or higher than an intermediate threshold $T$. Examples of possible parameters $S$ are the length of a queue, the maximum, the addition or any other function of the lengths of several queues, or the waiting time of the oldest call in the queue.

For clarity, this description will assume that event $A$, as event $C$, can only occur at certain instants, (e.g., at call arrivals) which will be called instants of interest; however, RESTART can also be applied to estimate continuous time probabilities.

Let us define events $B$ and $D$ as follows:
Figure 1: Example of the system evolution along time

B: the event C occurred in an instant of interest having occurred C in the previous one;
D: the event C occurred in an instant of interest having occurred C in the previous one.

Figure 1 can represent a normal simulation, while figure 2 represents a simulation made with RESTART. It can be a time-true or a roulette simulation with these modifications:

- When an event B occurs, the system state is saved.
- When an event D occurs, the system state of last event B is restored, and the interval B-D is simulated again.
- The above process is repeated R times, as illustrated in Fig. 2. The starting event of every trial is always the same, B, while the ending event is different, D, D', ..., D'', depending on system evolution in the trial.

- When the event D₂ occurs, simulation continues in the normal way until event B₂ occurs; then the same process is applied.
- The statistics should be accordingly corrected, either in real time or once the simulation has finished.

In simulations in which the probabilities of several types of rare events, A₁, A₂, A₃, etc., have to be estimated separately, an event C satisfying A₁ ⊆ C, A₂ ⊆ C etc., has to be chosen. Among the events A₁, A₂, ..., the probability of the most rare one will be the basis to obtain, as explained in section 4, the optimum values of P(C) and R.

3. STATISTICAL ANALYSIS

The aim of this analysis is to set the basis for optimizing the values of P(C) and R and for evaluating the gain obtained with RESTART. The analysis is made for steady state simulations.

3.1. Notation

Apart from events A, B, C and D and number of retrials R, previously defined, let define:

\[ P = P(A) \]
\[ P_1 = P(C) \]
\[ P_2 = P(A/C) \]
\[ P'_1 = P(B) \]
\[ P'_2 = \text{expected value of the number of events } A \text{ in an interval } [B,D) \]
\[ a = \text{expected value of the number of instants of interest in an interval } [B,D) \]
\[ N = \text{number of instants of interest simulated without counting those of the retrials} \]
\[ N'_1 = \text{number of events } B \text{ which occur in the simulation without counting the retrials} \]
\[ N'_2 = \text{number of events } A \text{ which occur in the simulation counting the retrials} \]

The following relations are straightforward:

\[ P = P_1 \cdot P_2 = P'_1 \cdot P'_2 \]  \hspace{1cm} (3.1.1)
\[ P_1 = a \cdot P'_1 \]  \hspace{1cm} (3.1.2)
\[ P_2 = P'_2 / a \]  \hspace{1cm} (3.1.3)

3.2. Analysis of simulations without RESTART

In a simulation without RESTART, the following estimators are used:

\[ \hat{P}_1 = N / N \; ; \; \hat{P}_2 = N'_2 / N'_1 \; ; \; \hat{P} = N'_2 / N \]  \hspace{1cm} (3.2.1)

Thus:

\[ \hat{P} = \hat{P}_1 \cdot \hat{P}_2 \]  \hspace{1cm} (3.2.2)

Assuming that \( \hat{P}_1 \) and \( \hat{P}_2 \) are independent random variables, the variance of \( P \), \( V(P) \) is:
\[ V(\tilde{P}) = V(\tilde{P}') + V(\tilde{P}^2) + V(\tilde{P}) \cdot [E(\tilde{P}^2)]^2 + V(\tilde{P}^2) \cdot [E(\tilde{P}')]^2 \]  
(3.2.3)

### 3.2.1. Evaluation of \( V(P_1') \)

If \( X \) is defined as a random variable which takes the values 1 or 0 depending on whether the event \( B \) occurs or not in the instant of interest \( i \), \( V(P_1') \), can be written as:

\[ V(P_1') = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{V(\sum_{i=1}^{N} X_i)}{N^2} = \frac{K'_1 \cdot P'_1 (1 - P'_1)}{N} \]  
(3.2.4)

where \( K' \) is a factor which would be equal to 1 if the system states \( X \) in different instants were mutually independent, but that, due to the usual positive correlation between near instants, is normally greater than 1. \( K'_1 \), derived in [6], is given by:

\[ K'_1 = 1 + \sum_{j=1}^{N} \frac{P(\eta_j = 1 | \eta_j = -1)}{P(\eta_j = 1 | \eta_j = -1)} \]  
(3.2.5)

\( K'_1 \) can be defined as the average value of the number of events \( B \) which occur in or around a given event \( B \), minus the number of them which would occur if there were independence.

By defining: \( K = a \cdot K'_1 \)  
(3.2.6)

formula (3.2.4) can be written as:

\[ V(P'_1) = \frac{K_1 \cdot P'_1}{a \cdot N} \]  
(3.2.7)

Considering that \( a \) is the average number of events \( C \) per event \( B \), \( K \) can be defined as the average value of the number of events \( C \) which occur in or around a given event \( B \), minus the number of them which would occur if there were independence.

### 3.2.2. Evaluation of \( \hat{V}(P'_2) \)

If \( Y \) is defined as a random variable which indicates the number of events \( A \) which occur in the 1th interval \([B, D]\), \( V(P'_2) \) can be written:

\[ V(\hat{P}'_2) = E\left( \frac{1}{N} \right) \cdot V(Y_1) = \frac{V(Y_1)}{E[N_1]} = \frac{K_2 \cdot P'_2}{N \cdot P'_1} \]  
(3.2.9)

where \( K_2 \) is a factor which would be equal to 1-\( P'_1 \) if \( Y \) could only take the values 0 or 1, but that, since \( Y \) can take greater values, is greater than 1. \( K_2 \) derived in [6], is given by:

\[ K_2 = 1 + \sum_{j=1}^{N} \frac{j \cdot P(Y_1 = j)}{\sum_{k=1}^{N} k \cdot P(Y_1 = k)} \]  
(3.2.10)

From the above formula, \( K_2 \) can be defined as the average value of the number of events \( A \) which occur in or around a given event \( A \), minus the number of them which would occur if there were independence between near instants.

### 3.2.3. Comparison between \( K_1 \) and \( K_2 \)

First consider that the system behaves an analogous way when the parameter \( S \) is near \( L \) and when it is near \( T \), i.e., that:

\[ P(S = L + \eta_1 | S_1 = L = \eta_1) = P(S_1 = L + \eta_1 | S_1 = L = \eta_1) \]

where \( S \), \( S_1 \), ... are random variables indicating the value of the parameter \( S \) in the instants of interest \( i \), \( i-1 \), and \( \eta_1 \), \( \eta_1-1 \), ... are arbitrary values. In this case, \( K \) is a factor which would be equal to 1 if \( Y_1 \) could only take the values 0 or 1, but that, since \( Y_1 \) can take greater values, is greater than 1. \( K_2 \) derived in [6], is given by:

\[ K'_2 = 1 + \sum_{j=1}^{N} \frac{j \cdot P(Y_1 = j)}{\sum_{k=1}^{N} k \cdot P(Y_1 = k)} \]  
(3.2.11)

In general, \( K_2 \) can be greater, equal or smaller than \( K_1 \) depending on the behavior of the system when \( S \) is near \( L \) compared with the behavior when \( S \) is near \( T \) but, in most cases, \( K_1 \) and \( K_2 \) will have the same order of magnitude.

### 3.2.4. Evaluation of \( V(P) \)

From formulas (3.2.3), (3.2.7) and (3.2.9), and considering formulas (3.1.1) to (3.1.3), the following formula is derived for \( V(P) \):

\[ V(\tilde{P}) = \frac{P}{N} \left( K_1 \cdot P_2 \cdot P + K_1 \cdot P_2 \cdot P + K_2 \right) \]  
(3.2.12)

### 3.3. Analysis of simulations with RESTART

In a simulation with RESTART, \( P' \), \( P_2 \) and \( P \) are estimated by:

\[ \hat{P}'_1 = \frac{N_1}{N} ; \quad \hat{P}'_2 = \frac{N_2}{N \cdot N_1} ; \quad \hat{P} = \frac{N}{N \cdot N_1} \]  
(3.3.1)

Formulas (3.2.2) and (3.2.3) also apply when RESTART is used. \( P' \) is estimated in the same
manner as when RESTART is not used, thus, formula (3.2.7), giving \( V(P'_{1}) \) applies in this case. However, \( P'_{2} \) is estimated in a different manner and, consequently, \( V(P'_{2}) \) is different.

### 3.3.1. Evaluation of \( V(P'_{2}) \)

Define \( Y_{1j} \) and \( \bar{Y}_{1} \) as follows:

- \( Y_{1j} \) is the random variable which indicates the number of events A which occur in the jth retrial made starting in the event B, i.e., in the ith event B appearing in the simulation;
- \( \bar{Y}_{1} \) is the expected value of \( Y_{1j} \) conditioned to the system state \( B_{1} \).

From the point of view of Sampling Theory, \( P'_{2} \) is estimated by means of sampling in two stages: In the first stage, a sampling of \( N_{1} \) primary units is made, corresponding to the \( N_{1} \) events B occurring in the simulation, and, in the second stage, \( R \) secondary units are sampled for each primary unit. Both sampling stages are made with renewal of the sampled unit. Thus, considering the \( N_{1} \) primary samples as mutually independent, and making the same assumption for the \( R \) secondary samples within a primary one, it can be proved [7] that:

\[
V(Y_{1j}) = V_{1} + V_{2} \tag{3.3.2}
\]

\[
V(P'_{2}) = E \left[ \frac{1}{N_{1}} \cdot \frac{V_{1}}{P'_{1} \cdot N} + \frac{V_{2}}{P'_{2} \cdot N} \right] \tag{3.3.3}
\]

\[
V_{1} = E(\bar{Y}_{1} - P'_{2})^{2} \tag{3.3.4}
\]

\[
V_{2} = E(V_{21}) \quad \text{and} \quad V_{21} = E(Y_{1j} - \bar{Y}_{1})^{2} \tag{3.3.5}
\]

Let us define:

\[
b = \frac{V_{1}}{P'_{2} \cdot V(Y_{1j})} \tag{3.3.6}
\]

In the same manner as in formula (3.2.9), \( V(Y_{1j}) \) can be written as:

\[
V(Y_{1j}) = K_{2} \cdot P'_{2} \tag{3.3.7}
\]

Thus, \( V(P'_{2}) \) is:

\[
V(P'_{2}) = \frac{K_{2} \cdot P'_{2}}{N \cdot P'_{2} \cdot R} \left( \frac{1}{R} + b \cdot P'_{2} \cdot R^{-1} \right) = \frac{K_{2} \cdot P'_{2}}{N \cdot P'_{2} \cdot R} \left( \frac{1}{R} + P_{P}'_{2} \right) \tag{3.3.8}
\]

### 3.3.2. Bounds of \( b \)

The lower bound of \( b \) is 0, which corresponds to a case in which the definition of event B is such that all the past and present system state variables which can influence the future system evolution have the same value for all events \( B_{1} \). In this case, \( V_{1} \) and thus \( b \) will be zero.

The upper bound of \( b \) is determined based on:

\[
V_{1} = E(\bar{Y}_{1} - P'_{2})^{2} = E(\bar{Y}_{1}^{2} - P'_{2}^{2}) \leq E(\bar{P} - P'_{2}^{2}) = P'_{2}^{2} - P'_{2} \tag{3.3.9}
\]

\[
P'_{2} \text{ being equal to } \text{Max}(\bar{Y}) \text{ for any } i. \text{ Thus, the bounds of } b \text{ are:}
\]

\[
0 \leq b \leq \frac{a}{K_{2}} \left( \frac{P'_{2}}{P_{P}^{2}} - 1 \right) \tag{3.3.10}
\]

The value of \( b \) can be reduced by an appropriate definition of event B which includes the values of most system state variables influencing the future system evolution. A reduction of \( b \) will increase the efficiency of RESTART.

### 3.3.3. Evaluation of \( V(P) \)

On the basis of formulas (3.2.3), (3.2.7) and (3.3.8), \( V(P) \) can be written as:

\[
V(P) = \frac{P}{N} \left[ K_{1} \cdot P_{P} + \left( K_{2} \cdot P_{P} + K_{2} \right) \left( \frac{1}{R} + b \cdot P_{P} \right) \right] = \frac{P}{N} \left[ K_{1} \cdot P_{P} + K_{2} \left( \frac{1}{R} + b \cdot P_{P} \right) \right] \tag{3.3.11}
\]

The comparison of (3.3.11) with (3.2.12) clearly indicates the advantage of RESTART. This advantage, greater if \( b \) is smaller, will be evaluated in the next sections.

### 4. OPTIMIZATION OF RESTART PARAMETERS

The cost of simulation can be measured as the number of instants of interest simulated. Using RESTART, this cost is given by:

\[
C = N + N_{1} \cdot a \cdot (R - 1) = N \left( 1 + P_{1} \cdot R \right) \tag{4.1}
\]

Since, given a certain simulation method, the cost, \( C \), is inversely proportional to the variance of the estimator, \( V(P) \), and the objective of RESTART is to decrease \( C \) for a same \( V(P) \) or vice versa, the product \( C \cdot V(P) \) is the parameter which should be minimized. From (3.3.11) and (4.1), and assuming that \( K_{1}, K_{2} \) and \( b \) do not depend on the value of \( P'_{2} \), it can be derived that the minimum of \( C \cdot V(P) \) is obtained for:

\[
V(P_{P}') = \frac{K_{2} \cdot P_{P}'}{N_{P} \cdot P_{P}'} \left( \frac{1}{R} + b \cdot P_{P}' \cdot R^{-1} \right) = \frac{K_{2} \cdot P_{P}'}{N_{P} \cdot P_{P}'} \left( \frac{1}{R} + P_{P}' \right) \tag{3.3.8}
\]
The optimum value of $R$ is always given by formula (4.2), even if the optimum value of $P_1$ had not been chosen, and vice versa.

5. GAIN OBTAINED WITH RESTART

The optimum values of $P$ and $R$ derived in section 4 lead to the following value of $C\cdot V(P)$:

$$C\cdot V(P) = 4\cdot K_2 \cdot \sqrt{(K_1 / K_2 + b) \cdot P} \quad (5.1)$$

Since the product $C\cdot V(\tilde{P})$ when RESTART is not used is $K_2 \cdot P$, the gain, $G$, obtained is:

$$G = \frac{1}{4 \sqrt{(K_1 / K_2 + b) \cdot P}} \quad (5.2)$$

5.1. Bounds of the gain

Formula (5.2.) can be difficult to evaluate in practice due to the difficulty of knowing $K_1$, $K_2$ and $b$. But, based on the bounds of $b$ given by formula (3.3.10), we can write:

$$\frac{1}{4 \sqrt{(\frac{K_1}{K_2} + b) \cdot P}} \leq G \leq \frac{1}{4 \sqrt{K_1 \cdot P}} \quad (5.1.1)$$

It most applications, according to section 3.2.3, it will be $K_1 \leq K_2$, $a \leq K_2$. Thus, formula (5.1.2) will be normally a safe-side approximation of (5.1.1):

$$\frac{1}{4 \sqrt{(\frac{P'_1}{P'} + b) \cdot P}} \leq G \leq \frac{1}{4 \sqrt{P}} \quad (5.1.2)$$

The above bounds of $G$ are easy to evaluate (at least, approximately) in most applications. Based on (5.1.2), Table 1 gives bounds of $G$ for several values of $P$ and $P'_1 / P'$.

Table 1 shows that the gain obtained is dramatic, mainly when $P = 10$, case in which this gain is more necessary.

5.2. Robustness for errors in the choice of the parameters.

The optimum values of $P$ and $R$ given by formula (4.2) may be difficult to calculate since $K_1$, $K_2$ and $b$ may be unknown. In this case, $K_1 = K_2$ and $b = 0$ could be assumed, leading to:

$$P = \sqrt{(K_1 / K_2 + b) \cdot P} \quad (4.2)$$

Moreover, $P$ may not be exactly known in some cases. Also, an event $C$ (or a threshold $T$) has to be chosen with probability of occurrence (or of exceeding $T$) equal to $\sqrt{P}$, and the probability of the chosen event could differ from that value. Table 2 shows the robustness of RESTART for all the above types of errors. It gives the gain obtained when the values of the parameters are:

- the optimum ones, given by formulas (4.2);
- approximate ones, given by formulas (5.1.3);
- erroneous ones, taken as the combination of values of $P'$ and $R$, each within the range of one tenth to ten times its approximated value, which gives minimum gain.

The table shows how robust RESTART is. A large gain is obtained even for erroneous values; it means that a careful analysis is not necessary to obtain a very important advantage.

Table 2: Gains obtained for different adjustments of the parameters.

<table>
<thead>
<tr>
<th>$P$</th>
<th>$P'_1 / P'$</th>
<th>LOWER BOUND</th>
<th>UPPER BOUND</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-9</td>
<td>1</td>
<td>7,900</td>
<td>7,900</td>
</tr>
<tr>
<td>10-9</td>
<td>10</td>
<td>2,500</td>
<td>2,500</td>
</tr>
<tr>
<td>10-9</td>
<td>100</td>
<td>790</td>
<td>790</td>
</tr>
<tr>
<td>10-6</td>
<td>1</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td>10-6</td>
<td>10</td>
<td>79</td>
<td>25</td>
</tr>
<tr>
<td>10-6</td>
<td>100</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 3: Computer times to which 115 days of normal simulation are reduced.
Table 3 gives the sensitivity to the choice of the parameter values in terms of computer time required with RESTART for a simulation which, without its use, would require 115 days. The table shows how the impracticable time of 115 days is reduced to reasonable times in most cases, very small in some of them. Any way, results obtained in a short simulation can be used to correct the erroneous parameters, avoiding cases as the last one of the table.

6. APPLICATION EXAMPLE

RESTART has already been successfully used to simulate an ATM multiplexer, as shown in a companion paper [8]. Its use has allowed to obtain the good confidence required in the study presented there. For a better comparison, a case which is analytically tractable has been taken to be presented here. In this case, the source input process is binomial (80 Bernouilli sources, each with \( p = 0.01 \), giving a total load of 0.8 Erl.). Events A and C were defined by thresholds reached by the multiplexer queue length. Threshold T has been chosen equal to 24 in order to have \( P = 10^{-4.8} \). The number of retrials has been \( R = 10^4 \).

Figure 3 shows results obtained by analytical formulas, and by normal and RESTART simulations, each one of 10 million cells. We can observe that normal simulation gives accurate results up to \( 10^{-5} \), while the results obtained with RESTART were accurate up to \( 10^{-9} \).

7. CONCLUSIONS

The paper presents a method to accelerate simulations for estimating the probability of rare events. The method called RESTART (REPpetitive Simulations Trials After Reachig Thresholds) has the following features:

- It allows a dramatic reduction of simulation time for an equal confidence of the results;
- It is general enough to be applied to most simulation models;
- It has a straightforward application to each particular case;
- It allows the application of simulation to evaluate future systems and networks, as, e.g. ATM, where the required time for the simulation made the application of classical methods impracticable.

Although a very important advantage is obtained from the use of RESTART in the present state, the authors consider that further developments on the same concept can lead to even greater advantages. Two topics are presently planned to be investigated:

- To introduce hysteresis by defining the ending event D of the retrials by a threshold lower than that defining the starting event B; it will prolong the retrials thus reducing the impact of the possible restoration costs;
- To define several thresholds, with a different set of retrials for each threshold reached; it can multiply the gain obtained with the use of only one threshold.

REFERENCES