Proceedings of the 1996 Winter Simulation Conference
ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain

THE RESTART/LRE METHOD FOR RARE EVENT SIMULATION

Carmelita Görg
Friedrich Schreiber

Communication Networks
Aachen University of Technology
D-52056 Aachen, GERMANY
[cg|fschr]@comnets.rwth-aachen.de

ABSTRACT

Using the LRE-algorithm (LRE: Limited Relative Error) for the evaluation of simulated data yields the stationary distribution function of an investigated random sequence and additionally the so-called local correlation coefficient, which represents relevant correlation evidence to be included in the error measure for controlling the simulation run time. In this paper a simplified LRE-algorithm is used to evaluate discrete sequences like the occupancy of finite buffer queueing systems G/G/1/N. It is shown how this algorithm is combined with the RESTART-method for an efficient rare event simulation.

A multi-stage RESTART/LRE-algorithm has been implemented as part of a stochastic simulation system and its performance has been verified by extensive simulations of the reference system M/M/1/N, whose properties including the local correlation coefficient can be described by analytical formulas. Approximate formulas for the optimal number of stages and number of trials are given. The new algorithm has been successfully applied to several finite buffer queueing systems, including the SSMP(2)/D/1/N system with a correlated input stream, in order to gain by simulation the distribution function of the occupancy including very low loss probabilities in the order of 10^{-9} and even 10^{-21}, which are relevant for the performance of switching nodes in the ATM broadband network.

1 INTRODUCTION

The quantitative investigation of random systems by stochastic simulation is an important tool for telecommunication engineers. How to assure the *reliability* of simulated results is a problem of interest and depends on the statistical evaluation method used to control the simulation run time. Since run time is always limited, there are limitations for locating in a simulated sequence a *rare event* with sufficient accuracy.

This problem has become an issue of current interest in the performance investigation of switching nodes of the ATM broadband network, which can be modeled by a single server, finite buffer queueing system G/D/1/N with a correlated general input process "G". System design has to guarantee that the event "cell loss" is a rare event with a probability in the range of $10^{-9} \cdots 10^{-10}$ (Le Boudec, 1991). How can this requirement be investigated and verified by simulation without excessive run time?

The relevant subfield within the various methods for variance reduction, see (Kleijnen, 1974), is *importance sampling* which originated as a method for more efficient Monte Carlo integration (Kahn and Marshall, 1953) and has been applied in a quite different way, called *importance splitting* in (Shahabuddin, 1995), to the statistical investigation of Markovian sequences by (Bayes, 1970) and has been further developed and made known under the name RESTART by (Villén-Altamirano and Villén-Altamirano, 1991) ("REpetitive Simulation Trials After Reaching Threshold").

As shown in (Schreiber and Görg, 1994) this method can be combined with the so-called LRE-method, see (Schreiber and Görg, 1996), which has been designed to resolve the often debated problem, see e.g. (Bratley et al., 1987), how to evaluate correlated sequences. This combination of two methods – named the RESTART/LRE-method – provides rare event simulation with a reliable error and run length control.

In this paper the RESTART/LRE-method with multiple stages is investigated and it is shown how the optimal number of stages provide a dramatic increase of speed-up in comparison to straightforward and single stage simulations. Examples are given of simulations that would have taken several years and even millions of years that could now be reduced to give good results within one day.

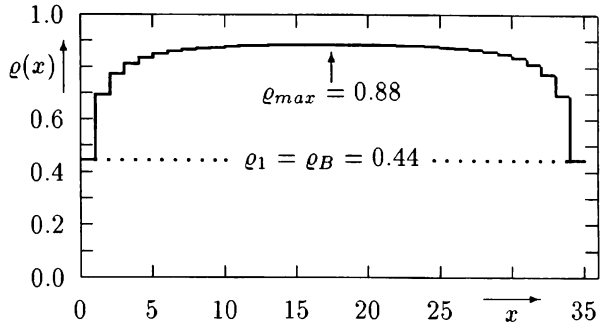

Figure 1: Queueing System M/M/1/34:
Local C.C. $\varrho(x)$ of Occupancy x ($\eta = 0.8$)

2 REFERENCE MODEL M/M/1/N

It is absolutely necessary to verify a new simulation method using reference queueing models whose relevant properties can be described by analytically derived exact formulas. A basic reference model for the rare event simulation is the single server queueing system M/M/1/N-FIFO with a finite buffer size N, whose local correlation properties have been recently investigated (Schreiber, 1994). With η denoting the traffic load and the maximum occupancy $B = N+1$, the discrete random variable "occupancy β" is described by the following equations for the stationary complementary distribution function (compl. d.f.) $G(x) = 1 - F(x)$, for the loss probability P_L, and the local correlation coefficient (local c.c.) $\varrho(x)$ for the interval $i-1 \le x < i$ and $i = 1, \cdots, B$, see (Schreiber, 1994):

$$\left.\begin{aligned} G(x) = G_i = \sum_{\beta=i}^{B} P(\beta) = \frac{\eta^i - \eta^{B+1}}{1-\eta^{B+1}}; \\ G_0 = 1; \quad G_B = P_L = \frac{\eta^B(1-\eta)}{1-\eta^{B+1}}; \end{aligned}\right\} \quad (1)$$

$$\left.\varrho(x) = \varrho_i = 1 - \frac{(1-\eta)(1-\eta^{B+1})\,\eta^i}{(1+\eta)(1-\eta^i)(\eta^i-\eta^{B+1})}.\right\} \quad (2)$$

Analytical investigations and simulations of other, less elementary single server queueing systems with finite buffer size have shown that the character of the flat maximum curve $\varrho(x)$ of system M/M/1/N is representative for several finite systems, see Figure 1.

3 THE LRE-ALGORITHM

The LRE-algorithm (LRE: Limited Relative Error) has been designed to establish the distribution function (d.f.) $F(x)$ and the local correlation coefficient (local c.c.) $\varrho(x)$ of a stationary x-sequence with *a priori* unknown properties. The LRE-algorithm II given in (Schreiber, 1988) is an extension of the first LRE-algorithm I, which was defined for the evaluation of *independent* sequences. The algorithm requires, among other things, the finding of the location and size of the x-intervals and the detection of discrete points. The procedure for performing automatically this general, rather complicated evaluation task can be substantially simplified by applying the LRE-algorithm III (Schreiber and Görg, 1996), whose evaluation task has been restricted to the measurement of discrete random sequences. When, for instance, the occupancy and loss probabilities of a queueing system G/G/1/N are to be investigated, the discrete character of the random variable "occupancy β" with buffer size N, maximum occupancy $B = N+1$, and the exact range $\beta = 0, \cdots, B$ is known.

For deriving the complementary distribution function $G(x) = 1 - F(x)$ of the arrival occupancy β of a system G/G/1/N, the values of β are registered exclusively *prior* to the end of each interarrival time τ_a, that is at all instants "immediately before a new item arrives". By applying this rule the random state behavior of the system is associated to an embedded Semi-Markov chain with $B+1$ nodes.

3.1 The local correlation coefficient $\varrho(x)$

We now assume that β is generated by the random state sequence of a recurrent $(k+1)$-node Markov chain with $k = B = N+1$, whose complementary d.f. $G(x) = 1 - F(x)$ exists as shown by Figure 2. Then we can split at any point x on the real axis this chain into two parts and define an "$F(x)$-equivalent" 2-node Markov chain (Schreiber, 1987a), whose transition probabilities $p_0(x)$ and $p_1(x)$ determine its correlation coefficient $\varrho(x) = 1 - [p_0(x) + p_1(x)]$ for $i - 1 \le x < i$.

According to Equation (6a,b) in (Schreiber, 1987a) the function $\varrho(x)$ represents a well defined (first order) correlation measure for the $(k+1)$-node chain and is called the local c.c. $\varrho(x)$ due to its dependence on the location x (Ding and Schreiber, 1990). After the chain has performed n state changes (trials) the posterior mean $\tilde{\varrho}(x)$ can be obtained by relating the number of transitions a_i and $c_i \approx a_i$, respectively, across the separation line at x to the total number of observed events r_i in the left range $\beta = 0, 1, \cdots, i-1$ resp. $v_i = n - r_i$ in the right range $\beta = i, i+1, \cdots, B$, see (3). If the large sample conditions (Schreiber, 1987b):

$$n \ge 10^3; \; (r_i, v_i) \ge 10^2; \; (a_i, c_i, r_i - a_i, v_i - c_i) \ge 10$$

are fulfilled, the following posterior formulas express the *posterior compl. d.f.* $\tilde{G}(x)$, the *posterior mean occupancy* $\tilde{\beta}$, the *posterior local c.c.* $\tilde{\varrho}(x)$ with *correla-*

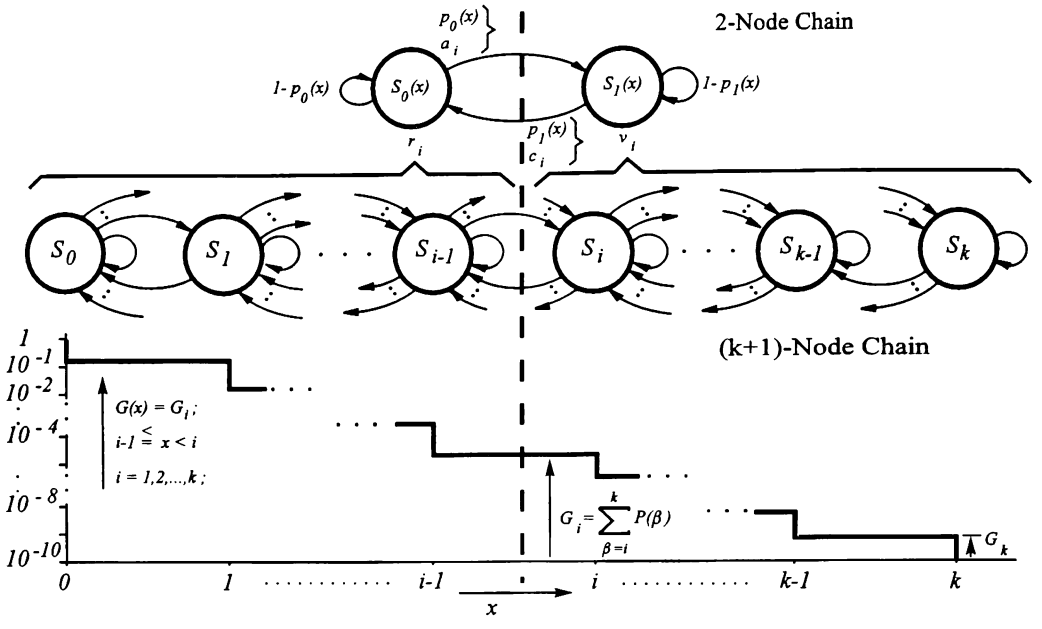

Figure 2: Discrete $(k+1)$-Node Markov Chain and the "$G(x)$-equivalent" 2-Node Markov Chain

tion factor $cf(x)$, and the *relative error* $d_G(x) \equiv d(x)$ for $i-1 \le x < i, i = 1, \cdots, B$, see (Schreiber, 1988):

$$\left.\begin{aligned} \tilde{G}(x) &= \tilde{G}_i = v_i/n\,;\ \tilde{\beta} = \tfrac{1}{n}\sum_{i=1}^{B} v_i\,; \\ \tilde{\varrho}(x) &= \tilde{\varrho}_i = 1 - \frac{c_i/v_i}{1 - v_i/n}\,; \\ cf(x) &= cf_i = (1+\tilde{\varrho}_i)/(1-\tilde{\varrho}_i)\,; \\ d(x) &= d_i = \left[\frac{1-v_i/n}{v_i}\cdot cf_i\right]^{1/2}. \end{aligned}\right\} \qquad (3)$$

Compared to conventional batch means and confidence interval methods, the relative error $d(x)$ has the great advantage that it contains correlation evidence, namely the measured values $\tilde{\varrho}_i$, and that it can be continuously reduced by evaluating the evidence of as many more trials as are needed to meet the error condition $d_i \le d_{max}$, with d_{max} being the prescribed error limit. The total needed number of trials n is determined by d_{max} and by the smallest G-value to be established, namely $G_{min} = \tilde{G}_B$ in the present case; with $d_B = d_{max}$ we obtain from (3):

$$n = \frac{(1-\tilde{G}_B)\,cf_B}{\tilde{G}_B\,d_{max}^2} \approx \frac{cf_B}{\tilde{G}_B\,d_{max}^2}; \quad cf_B = \frac{1+\tilde{\varrho}_B}{1-\tilde{\varrho}_B}.$$

4 THE RESTART/LRE-ALGORITHM

Verifying the loss probability $P_L = G_B = 8.1 \times 10^{-5}$ of the reference system M/M/1/34 for $\eta = 0.8$ by simulation would afford $n = 3.2 \times 10^6$ trials for an error limit of $d_{max} = 0.1$ and, assuming a production rate of 10^4 trials per second, a simulation run time of $T = 5.3$ minutes. By raising the buffer size from $N = 34$ to $N = 85$ we achieve for $\eta = 0.8$ a loss probability $P_L = 9.26 \times 10^{-10}$ and a simulation with again $d_{max} = 0.1$ would now afford $n = 2.8 \times 10^{11}$ trials and the (prohibitive) run time $T = 324$ days. The demand of a reduced 5 % error limit $d_{max} = 0.05$ would increase the simulation run time for $N = 34$ to $T = 21.4$ minutes and for $N = 85$ to $T = 3.56$ years.

This is the typical situation where straightforward simulation is no longer possible and must be replaced by a simulation based on the so-called RESTART-principle (Villén-Altamirano and Villén-Altamirano, 1991). In the following this principle will be combined with the evaluation and error control methods of LRE-algorithm III.

The main object of simulation will be the distribution function of the occupancy β of system G/G/1/N with the "rare event" being given by the loss case $\beta = N+1 = B$ and the discrete $(B+1)$-node Markov chain Figure 2 representing a model for generating the random β-sequence to be investigated. The main idea is to identify intermediate states or stages of the simulation and to *restart* the simulation in these states thus obtaining conditional probabilities that can be converted to the original distribution function using results of the previous stage.

The following three subsections describe the RESTART/LRE-algorithm with one stage, the general-

ization to m stages or intermediate points is given in section 5.

4.1 LRE-Evaluation for one stage

In its simplest version the RESTART/LRE-algorithm for one stage is performed by two sequential LRE-runs.

First LRE-Run: The first LRE-run establishes the left part of $G(x)$ up to a certain "Intermediate Point" $i = I$. The level $G_I = P(\beta \geq I)$ of this interval belongs to a "less rare event" $\beta \geq I$ compared to the level $G_B = P(\beta = B) = P(\beta > N) = P_L$ of the "rare event" $\beta = B$: $1 \gg G_I \gg G_B$. To perform the restart-feature of the second LRE-run correctly information about the intermediate state has to be collected. For instance, when a transition $I-1 \rightarrow I$ is registered, the value of the *residual service time* τ_R of the job or message in service is stored and used as a τ_R-generator for producing a random variable τ_R according to the distribution expressed by $\tilde{F}(\tau_R)$ in the second LRE-run.

To execute the first LRE-run an error limit d_{1max} has to be prescribed and a level G'_I of the intermediate point I resp. interval $I-1 \leq x < I$ has to be determined in accordance with section 4.3. G'_I can be given as a parameter or it can be approximated by the program. Alternatively the intermediate point I can also be given as a parameter.

The first LRE-run needs n_1 number of trials:

$$n_1 = \frac{(1-\tilde{G}_I)\, cf_I}{\tilde{G}_I\, d_{1max}^2} \approx \frac{cf_I}{\tilde{G}_I\, d_{1max}^2}; \quad cf_I = \frac{1+\tilde{\varrho}_I}{1-\tilde{\varrho}_I}. \quad (4)$$

The intermediate point I is related to the following posterior quantities for $I-1 \leq x < I$ and the stopping criterion $d_I \leq d_{1max}$, compare (3):

$$\left.\begin{aligned} \tilde{G}_I &= v_I/n_1\,; \quad \tilde{G}_I \approx G'_I\,; \\ \tilde{\varrho}_I &= 1 - \frac{c_I/v_I}{1-v_I/n_1}\,. \end{aligned}\right\} \quad (5)$$

Second LRE-Run: The second LRE-run establishes the right part of $G(x)$ for $i \geq I$ "under condition $\beta \geq I$" to be enforced by the RESTART-mechanism. This means that we now have to deal with the *conditional* compl. d.f. $G(x|\beta \geq I)$, which is related to the unconditional compl. d.f. $G(x)$ for $i-1 \leq x < i\,; i = I+1, \cdots, B$ as follows:

$$G(x) = G_I \cdot G(x|\beta \geq I) = G_i = G_I \cdot G_{i|\beta \geq I}\,. \quad (6)$$

The advantage of this expression is that the conditional levels $G_{i|\beta \geq I}$ are raised by the factor G_I^{-1} (e.g. $G_I^{-1} = 10^4$) compared to the unconditional levels G_i of the original function $G(x)$: $G_{i|\beta \geq I} = G_i \cdot G_I^{-1}$ and it is this effect, which enables us to measure successfully the rare event tail of the compl. d.f. $G(x)$.

The application of the RESTART-mechanism is quite simple: whenever the "next β-value" β_n represents a state $\beta_n < I$, a reset $\beta_n := I$ and after that a restart in state I is enforced. Unavoidably, these restarts cause deviations from the original random process involved, but any detrimental effects on the simulation result can be assumed as negligible due to the applied "service time correction": for each restart the current service time τ_b is replaced by a randomly chosen value of the residual service τ_R associated to transitions $I-1 \rightarrow I$, whose empirical d.f. $\tilde{F}(\tau_R)$ has been derived during the first LRE-run.

With the simplified notation $\tilde{\varrho}_B := \tilde{\varrho}_{B|\beta \geq I}$ we obtain in correspondence to (5) statements for the posterior conditional *loss probability* for $B-1 \leq x < B$ and the stopping criterion $d_{B|\beta \geq I} \leq d_{2max}$:

$$\left.\begin{aligned} \tilde{P}_{L|\beta \geq I} &= \tilde{G}_{B|\beta \geq I} = v_B/n_2\,; \\ \tilde{\varrho}_B &= 1 - \frac{c_B/v_B}{1-v_B/n_2}\,. \end{aligned}\right\} \quad (7)$$

This simplification is justified by the fact that the posterior value $\tilde{\varrho}_B$ of the local c.c. $\varrho(x)$ of interval $B-1 \leq x < B$ at the right end is practically the same for simulation with or without RESTART. The number of restarts need no special consideration, because it represents a random quantity being controlled by the second LRE-run. Assuming $\tilde{G}_{B|\beta \geq I} \ll 1$ the final number of trials n_2 for the second LRE-run is, compare (4):

$$n_2 \approx \frac{cf_B}{\tilde{G}_{B|\beta \geq I}\, d_{2max}^2}; \quad cf_B = \frac{1+\tilde{\varrho}_B}{1-\tilde{\varrho}_B}. \quad (8)$$

4.2 Error Considerations

After both LRE-runs have been performed the levels $\tilde{G}_i$ of the *unconditional* compl.d.f. $\tilde{G}(x)$, see (6), for $i = I+1, \cdots, B$ are:

$$\tilde{G}_i = \tilde{G}_I\, \tilde{G}_{i|\beta \geq I}\,; \quad \tilde{P}_L = \tilde{G}_B = \tilde{G}_I\, \tilde{G}_{B|\beta \geq I}\,. \quad (9)$$

A careful investigation of the correct error expression for the posterior statements $\tilde{G}_i$ must take into account that all levels $\tilde{G}_i$ of the first LRE-run and $\tilde{G}_{i|\beta \geq I}$ of the second LRE-run represent mean values of the posterior random variables G_i and $G_{i|\beta \geq I}$, respectively, which – under the present large sample conditions (Schreiber, 1987b) – are normally distributed. A straightforward analysis proves that the relative error d_i of the unconditional level $\tilde{G}_i$, see (9),

is given by:

$$d_i^2 = d_I^2 + d_{i|\beta\geq I}^2 = d_{1max}^2 + d_{i|\beta\geq I}^2 \,.$$

Due to the stopping criterion (5) we have set here with good approximation $d_I = d_{1max}$. In the same way we can set $d_{B|\beta\geq I} = d_{2max}$ due to the stopping criterion (7) and find for $i = B$ the relative error d_B of the unconditional *loss probability* $\tilde{P}_L = \tilde{G}_B$, which must be equal to a prescribed error limit d_{max} for the result $\tilde{G}_B$:

$$d_B^2 = d_{max}^2 = d_I^2 + d_{B|\beta\geq I}^2 = d_{1max}^2 + d_{2max}^2 \,.$$

The error limits d_{1max} and d_{2max} of the first and second LRE-run must be chosen in accordance with this equation. The error control in both LRE-runs safeguards that the condition $d_i \leq d_{max}$ is fulfilled for all levels $\tilde{G}_i$, $i = 1, \cdots, B$ including the unconditional levels $\tilde{G}_i$, see (9). This means that the prescribed value d_{max} is the effective error limit for the whole RESTART/LRE-algorithm.

4.3 Minimal Simulation Run Time

We replace in (4) and (8) the measured posterior quantities $\tilde{G}_I$, $\tilde{\varrho}_I$ and $\tilde{G}_{B|\beta\geq I}$, $\tilde{\varrho}_B$ by the corresponding variables G_I, ϱ_I and $G_{B|\beta\geq I}$, ϱ_B. Applying the relation $G_{B|\beta\geq I} = G_B/G_I$, $G_B \ll G_I$ which follows from (6) for $i = B$, we may therefore express the *total number of trials* $n = n_1 + n_2$ of both LRE-runs as a function of level G_I associated to the intermediate point I.

Using identical error limits for the first and second LRE-run $d_{1max} = d_{2max} = d_{Smax}$ we obtain the number of trials from $d_B = d_{max} = \sqrt{2}\, d_{Smax}$:

$$n(G_I) = n_1 + n_2 = \frac{2}{d_{max}^2}\left[\frac{cf_I}{G_I} + \frac{cf_B}{(G_B/G_I)}\right];$$

$$cf_I = \frac{1+\varrho_I}{1-\varrho_I}\,; \quad cf_B = \frac{1+\varrho_B}{1-\varrho_B}\,.$$

This function assumes its minimum n_{min} at the "optimal value" $G_I^\star$ of the variable G_I

$$\left.\begin{aligned} G_I^\star &= \sqrt{\frac{cf_I}{cf_B}}\cdot\sqrt{G_B}\,; \\ n_{min} &= n(G_I^\star) = \frac{4\sqrt{cf_I\, cf_B}}{\sqrt{G_B}\, d_{max}^2}\,. \end{aligned}\right\} \qquad (10)$$

For system M/M/1/N with $N = 85$ and $\eta = 0.8$ we have $G_B = 9.26 \times 10^{-10}$, $\varrho_B = 0.44$, and $cf_B = 2.6$. Due to the flat maximum of the function $\varrho(x)$ in Figure 1 the local c.c. is nearly constant $\varrho_I \approx \varrho_{max} = 0.89$ and therefore also the correlation factor is nearly constant $cf_I = 17.2$. For $d_{max} = 0.1$ we compute from (10) and (1): $G_I^\star = 2.56\sqrt{G_B} = 7.78 \times 10^{-5}$, $n_{min} = 8.7 \times 10^7$ trials, and the optimal intermediate point $I_{opt} = \ln[G_I^\star + \eta^{B+1}(1 - G_I^\star)]/\ln(\eta) \approx 42$.

Assuming a production rate of 10^4 trials per second the minimal simulation run time would be $T_{min} \approx 2.4$ hours. By comparing these values of n_{min} and T_{min} to $n = 2.8 \times 10^{11}$ trials and $T = 324$ days, which have been computed for a straightforward simulation of the same problem at the beginning of this section, we recognize the enormous potential of simulation speed-up by the RESTART/LRE-method.

The level G_I' of the intermediate point I to be prescribed in advance of the first LRE-run should be chosen as close as possible to the value of $G_I^\star$. But, under the conditions of normal simulation tasks, the optimal value $G_I^\star$ in (10) is *a priori* usually unknown because it depends on the values of the correlation factors cf_I, cf_B and of the loss probability $P_L = G_B$ itself, which become available only *a posteriori* as results of the statistical evaluation. For a series of RESTART/LRE-runs concerning the same simulation object the value of G_I' for each subsequent run can often be chosen closer to the optimum, because the values of cf_I and cf_B become known with increasing accuracy by the evaluation results of the previous runs. Therefore level G_I' must be chosen as a more or less good estimation of $G_I^\star$ and the needed number of trials will usually be $n > n_{min}$. In case that nothing is *a priori* known, it is recommended to use in (10) the values $cf_I = (1+\varrho_{max})/(1-\varrho_{max})$ and $cf_B = (1+\varrho_B)/(1-\varrho_B)$ of queueing system M/M/1/N, which can be computed for a given traffic load η from (2). Nevertheless, even in case of a relatively unfavorable choice of G_I' the essential simulation speed-up advantage of the RESTART-method is maintained; see also section 5.2 in (Villén-Altamirano and Villén-Altamirano, 1991).

5 THE MULTI-STAGE APPROACH

The extension of the RESTART/LRE-algorithm to m intermediate points $I_0, I_1, \cdots, I_{m-1}$ with G-levels: $1 \gg G_{I_0} \gg G_{I_1} \gg \cdots G_{I_{m-1}} \gg G_B$ is a natural extension of the above described algorithm with a single stage. The m intermediate points are also called stages leading to $m+1$ LRE-runs. This type of extension has also been envisaged in (Villén-Altamirano et al., 1994) and (Villén-Altamirano and Villén-Altamirano, 1994). The decisive advantage of the combination with the LRE-algorithm is that analytical formulas can be derived.

This leads under very general assumptions to the following optimal G-levels $G_{I_i}^\star$ $(i = 0, \cdots, m-1)$ and

minimal number of trials n_{min} :

$$G^{*}_{I_i}(G_B) = \left(\frac{\prod_{j=0}^{i} cf_{I_j}^{\,m-i}}{\prod_{j=i+1}^{m} cf_{I_j}^{\,i+1}} G_B^{i+1} \right)^{1/(m+1)} ;$$

$$n_{min}(G_B) = \frac{m+1}{d^2_{Smax}} \left(\frac{1}{G_B} \prod_{i=0}^{m} cf_{I_i} \right)^{1/(m+1)} . \quad (11)$$

To derive an approximation for the optimal number of stages m_{opt}, the correlation factor cf_{I_j} in (11) for $j = 0, \cdots, m-1$ is approximated by a fixed correlation factor named cf_M. The function $cf(x) = (1+\varrho(x))/(1-\varrho(x))$ is also a flat maximum curve with the maximum cf_{max}, if this is true for $\varrho(x)$ as shown in Figure 1, so that choosing the maximum for cf_M ($cf_M := cf_{max}$) is a good approximation and upper limit at the same time. Similar considerations as given in section 4.2 lead to an approximation of the error limit d_{Smax} in each run as a function of the overall error limit d_{max}:

$$d^2_{Smax} = (d^2_{max} + 1)^{1/(m+1)} - 1 \approx \frac{d^2_{max}}{m+1} .$$

Using these approximations the optimal number of stages m_{opt} follows from the necessary condition ($\partial n/\partial m = 0$) for an optimum of n_{min} of which the nearest integer is taken for simulation runs:

$$m_{opt} \approx \frac{1}{2} \ln \left(\frac{cf_B}{cf_M G_B} \right) - 1 . \quad (12)$$

Figure 3 shows the approximation of the optimal number of trials $n(m)$ on a logarithmic scale as a function of the number of stages m for the reference system M/M/1/N. The approximation is also an upper limit. Additionally a lower limit is shown.

It should be noted from (12) that the optimal number of stages does not depend on the maximum error level d_{max}. For smaller values of d_{max} the number of trials increases, but the optimum remains fixed.

A significant gain in terms of the total number of trials needed is already achieved for one intermediate point ($m = 1$), which can be seen from the comparison of values with a straightforward simulation ($m = 0$). Further improvements can be achieved when using the optimal number of stages. In the example in Figure 3 the relative speed-up $n(0)/n(1)$ is about 11 when comparing the straightforward simulation to the simulation with one intermediate point. In comparison with the optimal number of intermediate points ($m_{opt} = 3$ in Figure 3) a relative speed-up $n(0)/n(3)$ of about 18 is achieved. For this example it was possible to calculate the absolute minimum by exhaustive search, which is also given in Figure 3. For systems with very low loss probabilities the speed-up is much higher, so that now systems with, e.g. a loss probability of 10^{-20} can be simulated in affordable time, see section 6.

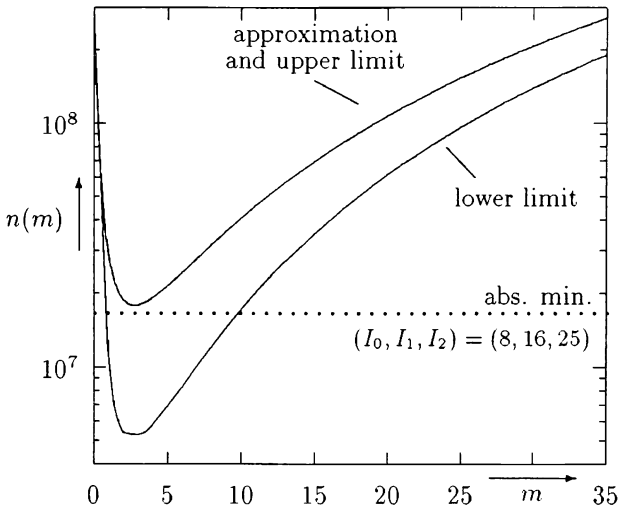

Figure 3: Queueing System M/M/1/34:
Total Number of Trials $n(m)$
versus Number of Intermediate Points m
(traffic load $\eta = 0.8$; max. error $d_{max} = 0.01$)

6 SIMULATION RESULTS

The LRE III and the RESTART/LRE-algorithm were implemented as part of a simulation system (Görg et al., 1991). Comprehensive RESTART/LRE simulation runs were performed and checked against theoretical and straightforward simulation results. Single server queueing systems of the type M/G/1/N-FIFO and G/D/1/N-FIFO were investigated.

In ATM simulations the service time is deterministic, so that systems of type G/D/1/N are of interest. Figure 4 shows the systems M/D/1/N and M/M/1/N compared to a system SSMP(2)/D/1/N with a correlated Special Semi-Markov input Process SSMP(2) described in (Ding and Schreiber, 1990), (Ding, 1991), (Herrmann, 1993), and (Herrmann, 1994). The input process SSMP(2) distinguishes two states, in each of these states geometrically distributed interarrival times are generated with parameter q_1 and q_2. The transition probabilities between the states are indirectly given by the correlation coefficient κ and the stationary state probability P_1. In the example sys-

Table 1: Comparison of Simulated and Theoretical Values for $\eta = 0.4$ and $d_{max} = 0.03$

	M/M/1/21			M/M/1/50		
	sim.: $m = 1$	sim.: $m_{opt} = 9$	theoretical	sim.: $m = 5$	sim.: $m_{opt} = 22$	theoretical
$\tilde{G}_B$ resp. G_B	$1.091\ 10^{-9}$	$1.031\ 10^{-9}$	$1.056\ 10^{-9}$	$2.98\ 10^{-21}$	$3.02\ 10^{-21}$	$3.04\ 10^{-21}$
$\tilde{\varrho}_B$ resp. ϱ_B	0.291	0.280	0.286	0.292	0.256	0.286
$\tilde{n}$ (n straightforward)	$3.42\ 10^{8}$	$2.70\ 10^{6}$	$(18.94\ 10^{11})$	$5.14\ 10^{8}$	$1.39\ 10^{7}$	$(6.57\ 10^{23})$
Speed-Up $n/\tilde{n}$	$5.5\ 10^{3}$	$701\ 10^{3}$	-	$1.3\ 10^{15}$	$473\ 10^{15}$	-

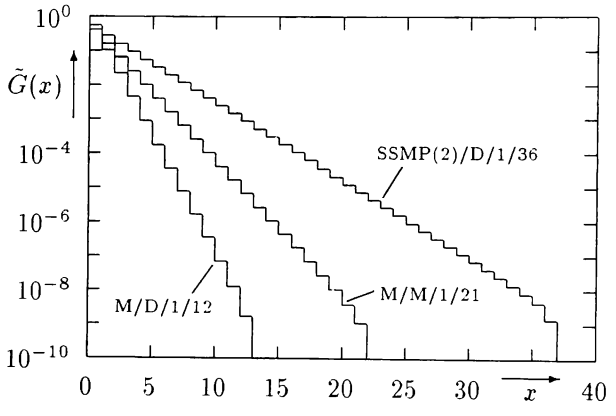

Figure 4: RESTART/LRE Simulation: Compl. Distribution Function $\tilde{G}(x)$ of Occupancy x ($m = 1$, $\eta = 0.4$, $G_B = P_L \approx 10^{-9}$, $d_{max} = 0.03$)

tem SSMP(2) the following parameters were used: $q_1 = 0.9$, $q_2 = 0.3$, $\kappa = 0.6$, $D = 3$, and P_1 was adapted according to the traffic load η.

The influence of the arrival and service time distributions on the compl. d.f. $G(x)$ of the occupancy x and the loss probability $P_L = G_B$ for a fixed traffic load of $\eta = 0.4$ is shown in Figure 4. The buffer size N for each system was chosen in such a way that the loss probability is approximately 10^{-9}. The M/D/1 system only needs a buffer size of $N = 12$ to fulfil the requirement of $P_L \approx 10^{-9}$, the M/M/1 system needs $N = 21$, whereas the SSMP(2)/D/1 system needs $N = 36$ due to the strong effect of correlated arrivals. The simulated compl. d.f. $\tilde{G}(x)$ of system M/M/1/N is in full agreement with the theoretical function $G(x)$, see (1). The SSMP(2)/D/1/N results were also checked successfully against the theoretical results given by (Herrmann, 1993). The differences are too small to be shown in the diagram.

Table 1 compares two reference systems with a loss probability of about 10^{-9} and 10^{-21}. The number of trials needed for the multi-stage simulations in comparison with the straightforward and single stage simulation are given. The simulation results are compared to the theoretical values. The main result concerning the loss probability $\tilde{P}_L = \tilde{G}_B$ resp. $P_L = G_B$ shows the conformity of simulation and theory.

If the quotient of B/m is small (e.g. $B/m<3$) the resulting values for $\tilde{\varrho}_I$, are lower than in the original process which reduces the number of trials $\tilde{n}$ in the actual simulation. n gives the theoretically needed number of trials for a straightforward simulation. The speed-up factor is $n/\tilde{n}$. As mentioned in section 4 a straightforward simulation of these cases is practically impossible. Assuming 10^4 trials/sec results in 6 and 2×10^{12} years for the examples in Table 1. With the RESTART/LRE speed-up the simulations were all performed *within one day.*

7 FINAL REMARKS

In this paper the multi-stage RESTART/LRE-algorithm was described and the speed-up in comparison to straightforward and single stage simulations was discussed for representative examples.

The RESTART/LRE-algorithm as described here can be extended and adapted to other application fields, e.g. *multi-server* systems G/G/s/N. Another extension is the investigation of rare event details of more general stationary x-sequences having a pure continuous or a mixed continuous and discrete character. For this purpose the (more complicated) LRE-algorithm II (Schreiber, 1988) must be combined with the RESTART-method. Considering more complex systems, such as queueing networks, several open issues need to be resolved. The RESTART/LRE-method needs well defined intermediate states, an algorithm to generate these states, and the statistics for evaluating the original distribution function.

The results using the RESTART/LRE-algorithm are well established and will be further investigated to make them available in practical simulation tools.

ACKNOWLEDGEMENTS

The authors thank O. Fuss and W. Olzem, Aachen, for supporting these investigations by simulation work and Dr.-Ing. C. Herrman for helping with the *mutual verification* of simulation and theoretical results.

REFERENCES

Bayes, A. (1970). Statistical techniques for simulation models. *Australian Comp. Journ.*, 2:180–184.

Bratley, P., Fox, B., and Schrage, L. (1987). *A guide to simulation.* Springer, New York.

Ding, W. (1991). A unified correlated input process model for telecommunication networks. In *13th Int. Teletraffic Congress*, Copenhagen, Denmark. North-Holland.

Ding, W. and Schreiber, F. (1990). Local correlation properties of random sequences generated by queueing system M/M/1. *AEÜ*, 44:384–389.

Görg, C., Kluth, B., and Salewski, U. (1991). SIC – a system for stochastic simulation in C++. In *Modelling and Simulation 1991, Proceedings of the 1991 European Simulation Multiconference*, pages 413–414.

Herrmann, C. (1993). Analysis of the discrete-time SMP/D/1/s finite buffer queue with applications in ATM. In *IEEE INFOCOM'93, San Francisco, 1993, Session 2a.3, pp. 160-167.*

Herrmann, C. (1994). The local correlation properties of some discrete-time finite queues with semi-Markovian input. *AEÜ*, 48:244–255.

Kahn, H. and Marshall, A. (1953). Methods of reducing sample size in Monte Carlo computations. *J. Oper. Res. Society of America*, 1:263–278.

Kleijnen, J. (1974). *Statistical techniques in simulation*, volume I. Marcel Dekker, New York.

Le Boudec, J. (1991). An efficient solution method for Markov models of ATM links with loss priorities. *IEEE J. Select. Areas Commun.*, 9:408–417.

Schreiber, F. (1987a). The empirical stationary distribution function of Markovian correlated random sequences. *AEÜ*, 41:257–263.

Schreiber, F. (1987b). Measurement of the stationary state probabilities of the 2-node Markov chain. *AEÜ*, 41:117–122.

Schreiber, F. (1988). Effective control of simulation runs by a new evaluation algorithm for correlated random sequences. *AEÜ*, 42:347–354.

Schreiber, F. (1994). Local correlation properties of the finite buffer queueing system M/M/1/N. *AEÜ*, 48:63–67.

Schreiber, F. and Görg, C. (1994). Rare event simulation: a modified RESTART-method using the LRE-algorithm. In *Proceedings 14th Int. Teletraffic Congress*, pages 787–796. North-Holland.

Schreiber, F. and Görg, C. (1996). Stochastic simulation: a simplified LRE-algorithm for discrete random sequences. *AEÜ*, 50.

Shahabuddin, P. (1995). Rare event simulation in stochastic models. In *Proceedings of the 1995 Winter Simulation Conference*, pages 178–185, Arlington, Virginia, USA.

Villén-Altamirano, M., Martínez-Marrón, A., Gamo, J., and Fernández-Cuesta, F. (1994). Enhancements of the accelerated simulation method RESTART by considering multiple thresholds. In *Proceedings 14th Int. Teletraffic Congress*, pages 797–810. North-Holland.

Villén-Altamirano, M. and Villén-Altamirano, J. (1991). RESTART: a method for accelerating rare event simulations. In Cohen, J. and Pack, C., editors, *Queueing, Performance and Control in ATM*, pages 71–76. Proc. 13th Int. Teletraffic Congress, Copenhagen, North-Holland.

Villén-Altamirano, M. and Villén-Altamirano, J. (1994). RESTART: A straightforward method for fast simulation of rare events. In *Proc. of the 1994 Winter Simulation Conference*, pages 282–289.

AUTHOR BIOGRAPHIES

CARMELITA GÖRG received the Dipl.-Inform. degree in computer science from the University of Karlsruhe, Germany, in 1976, and the Dr.rer.nat. degree from Aachen University of Technology, Germany, in 1984 for her dissertation in the field of queueing strategies for communication networks. From 1985 until 1989 she worked as a consultant in the field of communication networks. Since 1989 she has been working as a group leader at the Communication Networks Institute, Aachen University of Technology. Her research interests are: performance analysis of communication networks, stochastic simulation, rare event simulation, high speed networks, and new services in telecommunication networks.

FRIEDRICH SCHREIBER studied communication engineering from 1948–1956 at the Universities of Karlsruhe, Rhode Island/USA, and Stuttgart, where he received the Dr.-Ing. degree for his dissertation in the field of magnetic materials. From 1956–1971 he worked at the Central Laboratories of Siemens AG in Munich. From 1971–1990 he was full professor and is now professor emeritus at the Aachen University of Technology, Faculty of Electrical Engineering. His research interests are: measurement of user behavior in data networks, optimal queueing strategies for communication networks, stochastic simulation techniques, quasi-ideal random generators, and new statistical evaluation methods for correlated random sequences.