
Simulation of Rare Events in Communications Networks

J. Keith Townsend, North Carolina State University

Zsolt Haraszti, Ericsson Infocom Systems

James A. Freebersyser, Office of Naval Research

Michael Devetsikiotis, Carleton University

ABSTRACT Computer simulation is an important tool in the analysis and design of communications networks. In spite of the advances in computational power, using simulation to obtain rare event probabilities such as cell/packet loss or delay in networks still requires prohibitively long execution times. We provide an overview of importance sampling techniques and how they can be used to provide orders of magnitude speedup for many network problems.

Many important measures of performance in communications networks are defined in terms of rare event probabilities. Two examples are cell loss probability and cell delay threshold probability in asynchronous transfer mode (ATM) networks. Obtaining accurate estimates of such rare event probabilities using computer simulation can require execution times that are prohibitively long, but simulation has long been a powerful tool in predicting performance of communication networks.

The issue of reducing execution time while simultaneously retaining the ease and flexibility of computer simulation has been a topic of investigation for a number of years. Techniques based on importance sampling (IS) have shown great promise for many different classes of communication networks (see [1] and references within).

On the spectrum of performance analysis techniques, with pure analytical solution at one end and computer simulation at the other, IS-based techniques would lie somewhere in the middle. Thus, IS-based techniques require some problem-specific time from the analyst, as well as execution time on the computer to run the simulations. The application of IS-based techniques can provide solutions to problems involving rare events that would otherwise not be possible at either end of the spectrum mentioned above.

Fundamentally, IS is based on the notion of modifying (or *biasing*, as it is commonly called) the underlying probability mass in such a way that the rare events occur much more frequently. To correct for this modification, the results are weighted in a way that yields a statistically unbiased estimator. The only fundamental restriction on how the original probability mass is biased is that the biased probability mass must not exclude the occurrence of the rare event of interest. The effort of applying IS is in determining *which* parameter(s) of the system to bias (the technique), and *how much* to bias each of them. The objective for using IS is to obtain a significant reduction in the number of trials required to obtain the same estimator precision as would be obtained in a simulation without IS. We call this factor of improvement under IS *speedup*.

In this article we examine IS and how it can be used to reduce the simulation execution time of rare events in communications networks by many orders of magnitude. Although there have been many investigations regarding the theoretical underpinnings of IS, this article is more tutorial in nature, and focuses on giving a higher-level overview of IS techniques. *Conceptually* the basic idea behind IS is to modify the stochastic behavior of the system in order to increase the sample size

of the target rare event(s). In this sense, trajectory splitting-based techniques are clearly a type of IS. Trajectory splitting achieves speedup by launching additional subtrajectories from intermediate system states. The intermediate states are visited much more often than the target states themselves and behave as gateway states to reach the target states. Trajectory splitting is discussed in more detail below.

We organize the article in terms of how IS is *used* by the community. This viewpoint yields two general classes: modification (or biasing) of *individual* stochastic elements (e.g., sources and servers), and *global* modification, which includes splitting-based techniques. We then turn to the issue of *how much* each element should be biased by discussing tuning/optimization techniques that have been used in network simulation. Applications are presented that demonstrate how IS is applied to several different communications networks.

IMPORTANCE SAMPLING

Many of the key features of IS can be explained using the following simple example. Assume that the objective is to determine the area of the two-dimensional region B in Fig 1a. Region B represents a (not so rare) probability of interest. The analytical solution would require a mathematical description of the boundary of region B, as well as a complex integration procedure. In many cases this knowledge is not obtainable; in such cases computer simulation using the Monte Carlo (MC) method is one alternative. The MC method for *estimating* this area would require generating uniformly distributed random samples over the entire space A. The estimate of the area of region B would be given by $\hat{B} = N_B/N$, where N_B is the number of hits within region B, and N is the total number of samples generated. Assuming statistically independent samples, the variance of this estimate is inversely proportional to the number of samples, N .

In general, the precision of the estimate is related to the number of hits in the important region. Thus, if our objective is to estimate the much smaller area in region C, a much larger number of samples, N , would have to be generated for an equivalent estimator variance.

Using IS, we would modify or *bias* the sampling procedure to increase the fraction of samples that result in hits. In the context of this example we represent this by arbitrarily *doubling* the probability that samples are generated in the quadrant labeled D. Thus, the average number of samples in region C is doubled, increasing the estimator precision. In this example, each sample which results in a hit in region D must

be weighted by a factor of 1/2 to yield the correct, statistically unbiased result.

One of the practical difficulties when developing an IS solution is ensuring that the regions in the space with increased sampling frequency (region D) include the important region (region C). To illustrate, assume that the region of interest is actually region E in Fig. 1b. Because of insufficient prior knowledge of system behavior, samples occur with half the probability in region E with the biasing scheme used here, thus *reducing* the number of hits and the corresponding estimator precision by a factor of two. The weight of each hit in region E is 2, rather than 1/2.

In the context of a communication network, region A represents the state space, sampling represents the evolution of the sample path, and IS in a network is a modification of the stochastic behavior in a way which increases the probability that the system will visit the preferred region(s).

TECHNIQUES

IS-based techniques can be broadly grouped into two categories: techniques where the individual stochastic elements are modified or biased, and those where the global evolution of the system is manipulated.

MODIFICATION OF INDIVIDUAL STOCHASTIC ELEMENTS

The process of MC simulation involves developing models of the various key functions of the system, interconnecting these models to mimic behavior of the actual system, and finally, performing the actual simulation trials. One or more of the phenomena modeled include random number generation. Examples include models of sources, traffic routing, and packet lengths.

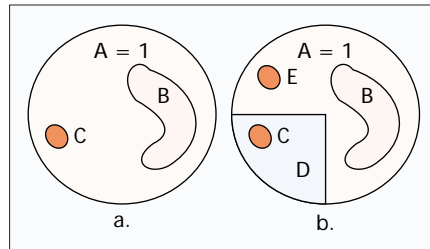
An important category of IS techniques involves modifying the underlying probability distributions of one or more of these random number generators in the simulation model. Applying IS in this fashion typically requires considerable prior knowledge about the system. This prior knowledge relates how the modified random number distributions affect the distribution of the target event(s) of interest.

A chief issue when using this technique is determining how much the distributions of the random number generators should be modified or biased. This issue of tuning/optimization is covered below.

GLOBAL MODIFICATION VIA TRAJECTORY SPLITTING

An alternative way to increase the relative number of visits to the important rare event region is to use (*trajectory*) *splitting*. As stated above, the fundamental idea of splitting is based on the assumption that there exist some well identifiable intermediate system states that are visited much more often than the target states themselves and behave as gateway states to reach the target states. For example, if the target states represent a full queue in a queuing system, the states that correspond to the case when the queue is *at least* half full can be regarded as intermediate states.

A very important feature of splitting is that the step-by-step evolution of the system follows the original probability measure. Entering the intermediate states during the simulation — which is usually characterized by crossing a *threshold* by a *control* parameter — triggers the splitting of the trajectory. The current system state is saved and a number of independent *subtrajectories* are simulated from that state. Achieving considerable variance improvement usually requires



■ **Figure 1.** Example showing sample space A and regions of interest: a) conventional Monte Carlo; b) an importance sampling biasing scheme where the probability of samples in region D is increased by a factor of two.

a hierarchical system of splitting conditions (a set of thresholds).

Existing splitting techniques assume different restrictions on how the splitting conditions are defined, apply different rules to terminate subtrajectories, and use different approaches to obtain the unbiased parameter estimates. To explain the above key features we use the splitting technique based on direct probability redistribution (DPR) as an example [2]. (To learn more about other splitting techniques the reader is referred to [3, 4] and references therein.) DPR partitions the

state-space, S , of a Markovian system into m subsets, S_1, S_2, \dots, S_m , by a mapping function, $\Gamma(s) \in [1, m]$, and assigns *oversampling factors*, $\mu = \{\mu_1, \dots, \mu_m\}$ ($\mu_1 \leq \mu_2 \leq \dots \leq \mu_m$), to each subset. $\Gamma(s)$ is the subset indicator, that is, it assigns each state its subset index.

Splitting occurs whenever the system enters a given subset from a lower-indexed subset. The number of subtrajectories and their terminating conditions are defined [2] such that every state in subset S_j is visited (relatively, and in the steady state sense) μ_j times more often than in a MC simulation. By assigning high oversampling factors to low probability subsets, we can ensure that they are visited much more often during the DPR simulation.

If the partitioning is chosen properly, this should also result in more samples in the important region. Since each state s_j is oversampled by a factor of $\mu\Gamma(s_j)$, unbiased estimates can be obtained by weighting a subset-dependent factor $1/\mu\Gamma(s_j)$ (Fig. 2). The upper left of the figure shows a sample path of the original single queue system where queue length is plotted versus discrete time; the upper right shows queue length partitioned into two subsets. Three independent subtrajectories are generated when subset S_2 is entered (since $\mu = 2$, as determined from an independent exploration process). Each important event counted while the system is in subset S_2 must be weighted by a factor of 1/3 to yield a correct estimate. The lower left of Fig. 2 shows an example illustrating subtrajectories for the same system with three subsets; in the lower right the same example as in the lower left plot is shown, except here the system evolution skips a subset.

In [2] we pointed out that DPR simulation has an equivalent Markov chain representation, and that the resulting Markov chain can be derived from the original Markov chain by manipulating the transition probabilities in a systematic way. Thus, splitting can be regarded as an indirect IS technique which, instead of modifying individual stochastic sources, changes the entire transition probability structure of the system.

TUNING/OPTIMIZATION OF PARAMETERS

LARGE DEVIATIONS, EFFECTIVE AND DECOUPLING BANDWIDTHS

Fast simulation methods based on large deviation theory (LDT) [5] have also been used successfully. LDT-based techniques, when applicable, allow analytical or numerical calculation of the amount of biasing required. Furthermore, they provide a framework that leads to arguments on *asymptotic efficiency* (i.e., asymptotic lower bounds on the acceleration offered by IS), or even *asymptotic optimality* of the IS estimator. A very thorough survey of LDT-based techniques is provided in [1].

The first step when using LDT is to specify the biased (modified) distributions as θ -conjugate *exponentially twisted*

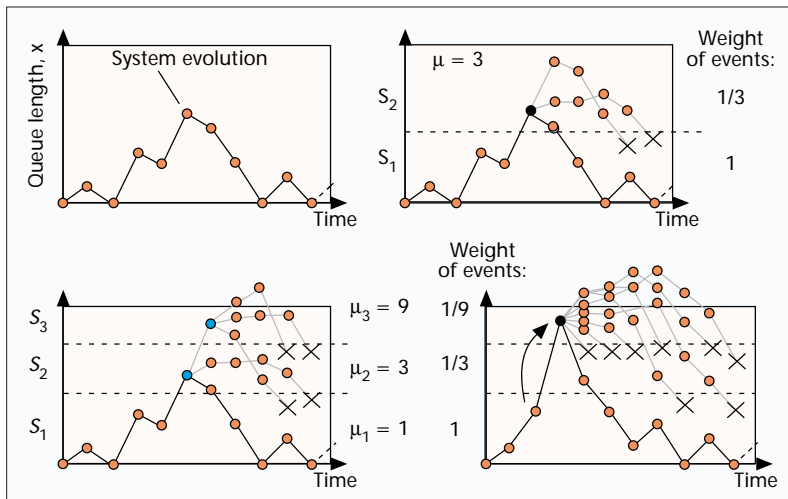


Figure 2. Splitting increases the probability of reaching rare states by splitting the system trajectory upon entering well-defined subsets of the state-space.

versions of the original, unbiased distribution. Exponentially twisted means that the biased distribution is the original distribution multiplied by an exponential with rate θ , appropriately normalized. For example, in the case of an exponential or Gaussian distribution, the exponentially twisted distribution is the original distribution shifted to a new mean.

The question now arises: what value of θ should be used? To answer this question for exponentially twisted arrival streams in queues, the theory of effective bandwidths can be invoked. Assuming that for a discrete time traffic stream $\{A_n, n = 1, 2, \dots\}$ the limit $\Lambda(\theta) = \lim_{n \rightarrow \infty} (1/n) \log E \exp[\theta \sum_{i=1}^n A_i]$ exists, then the quantity $a(\theta) = \Lambda(\theta)/\theta$ is called the effective bandwidth of the stream.

For a single queue with deterministic service, the value of θ is chosen by setting the effective bandwidth of the arrival stream equal to the service rate, and solving the equation for θ . The problem of multiple streams arriving at the same queue is addressed by using the additive property of effective bandwidths. Furthermore, a description of the aggregate effective bandwidth at the output of a queue is required. Armed with the additive property and an I/O relationship,

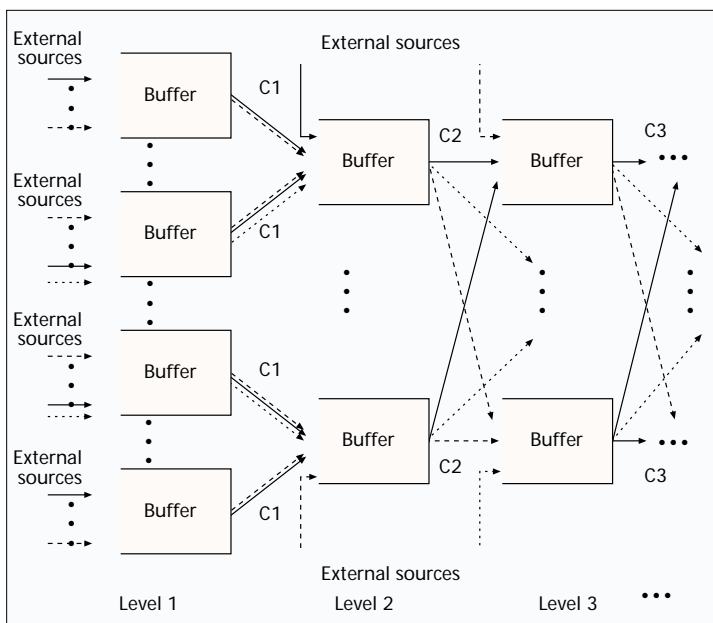


Figure 3. Diagram of a feedforward network of FIFO queues.

we can then find appropriate values for θ as a solution to a set of recursive equations, as long as the network is feedforward and there is no splitting of traffic at the output. (The splitting used in this context should not be confused with trajectory splitting discussed above.)

As an example, consider a feedforward network of FIFO queues with finite buffers and deterministic service rates, as illustrated in Fig. 3. The first two levels constitute an *intree* configuration, while the rest of the levels may incorporate splitting of departing traffic streams for subsequent transmission. External sources may exist at all levels.

As can be seen from this example topology (levels 2 and higher), not all networks are of the intree type. The additional information required for general feedforward topologies concerns the effective bandwidth of individual departing streams, as opposed to the aggregate output. De

Veciana *et al.* [6] introduced the notion of decoupling bandwidths, which, in addition to the standard effective bandwidths, can be used to provide sufficient conditions so that the effective bandwidth of a specific tagged stream at the output remains the same as at the input of a buffer.

A recent algorithm for the determination of appropriate values of θ in the θ -conjugate (twisted) process for more general non-intree network configurations can be found in [7]. The algorithm makes use of decoupling bandwidths in addition to standard effective bandwidths and their additive property, and yields a value of θ as a solution to a recursive set of equations.

STOCHASTIC OPTIMIZATION

An exact closed-form representation of the estimator variance is rarely available. Instead, statistical measures of performance have been proposed, which are statistical estimates of the variability (scatter) of the MC observations involved and asymptotically consistent estimates of the estimator variance, $\sigma_{IS}^2(P, P^*)$, with respect to the bias parameter values [8].

The mean field annealing (MFA) algorithm is a variation of simulated annealing (SA) that retains the ability of SA to avoid local minima and arrive at optimal or near-optimal solutions while demonstrating more rapid convergence. MFA combines the effectiveness of SA with reduced runtimes; therefore, it can be used to minimize the noisy cost function, $C(\mathbf{a}) = \hat{\sigma}_{IS}^2(P, P^*)$ with respect to the IS-parameters, \mathbf{a} . An MFA-based algorithm that estimates near-optimal IS parameter settings $a_{1,opt}, \dots, a_{d,opt}$ is given in [8].

The same problem of choosing IS parameter values to minimize the simulation variance can also be tackled using a stochastic gradient descent (SGD) algorithm. By exploiting more information about the problem at hand (i.e., derivative information), SGD can potentially zero in on favorable bias parameter settings faster than global search techniques like MFA and similar annealing methods [9]. During the search the simulation sampling distribution is continuously changing while approaching the optimal IS distribution. Thus, the algorithm tends to constantly improve the estimator variance until near-optimal bias parameters are found.

CONDITIONAL BIASING

Parametric IS methods are not very effective when the randomness in the system is described by a uniform probability distributions. Conditional biasing is an impor-

tance sampling technique that has been shown to be effective in such cases [10, 11]. However, conditional biasing does require some prior knowledge of the system behavior that causes the important event of interest. This prior knowledge is used to partition the uniform probability distribution into intervals which will result in the important event and intervals which will not result in the important event. By drawing random variables from the intervals that cause the important event, rather than the entire uniform distribution, more important events will result. A requirement when using this technique is that the occurrence of any sequence of random variables resulting in an important event not be excluded from the biased random variable selection process. An unbiased estimate is achieved by weighting the occurrence of each important event by the ratio of the unbiased uniform probability density function (pdf), $f(\cdot)$, to the biased pdf, $f^*(\cdot)$.

Consider the case where the location of a sample in space A in Fig. 1 is described by the polar coordinates (r, θ) where r is the radius of the location from the center of A and θ is the phase angle where 0 radians is at 12 o'clock. Conventional MC simulation would select uniformly distributed random variables from $f(r) = U[0, 1]$ and $f(\theta) = U[0, 2\pi)$. Assume that region C in Fig. 1 is the important region of interest. Consider a region C^* such that $C \subseteq C^*$ where $C^* = ((r, \theta) : [0.6, 0.8], [5\pi/4, 11\pi/8])$. Conditional biasing is used by restricting the uniform probability distribution for r and θ to the respective intervals in C^* . Since $C \subseteq C^*$, no fundamental tenet of IS has been violated. The occurrence of each important event is weighted by a factor

$$w(r, \theta) = \left(\frac{f(r)}{f^*(r)} \right) \left(\frac{f(\theta)}{f^*(\theta)} \right) = \left(\frac{0.2}{1} \right) \left(\frac{\pi/8}{2\pi} \right) = \frac{1}{40}.$$

Now if $C = C^*$, then the improvement obtained would be $1/w(r, \theta) = 40$, but since this is not the case (i.e., there is room for samples to be located at the corners of C^* which are not in C) the improvement will be somewhat less. See [11] for a more complete development of conditional biasing.

ITERATIVE BALANCING FOR TRAJECTORY SPLITTING

Key issues when applying trajectory splitting to a rare event problem are:

- To find appropriate partitioning
- To choose the correct amount of splitting

For example, these are determined by $\Gamma(\cdot)$ and μ , respectively, in DPR. Although the selection of $\Gamma(\cdot)$ is a problem-specific step which requires some insight to the rare event mechanism in hand, numerous application examples demonstrate that huge simulation gains can be achieved even by very simple intuitive approaches.

At a first glance, the proper selection of μ might appear to be an $(m - 1)$ -dimensional optimization problem. Fortunately, this is not true. In fact, it has been found [2-4] that the near-optimal μ setting is when the subset probability masses are equalized (i.e., each subset is visited with approximately the same probability during splitting). In DPR, for example, this can be achieved by setting factors μ_i to be inversely proportional to the true subset probabilities. Although true subset probabilities are not known a priori, a simple iterative procedure can be used to explore subset probabilities in a step-by-step fashion; in this way, near-optimal oversampling factors can easily be obtained. The simulation overhead of such an initial exploration procedure

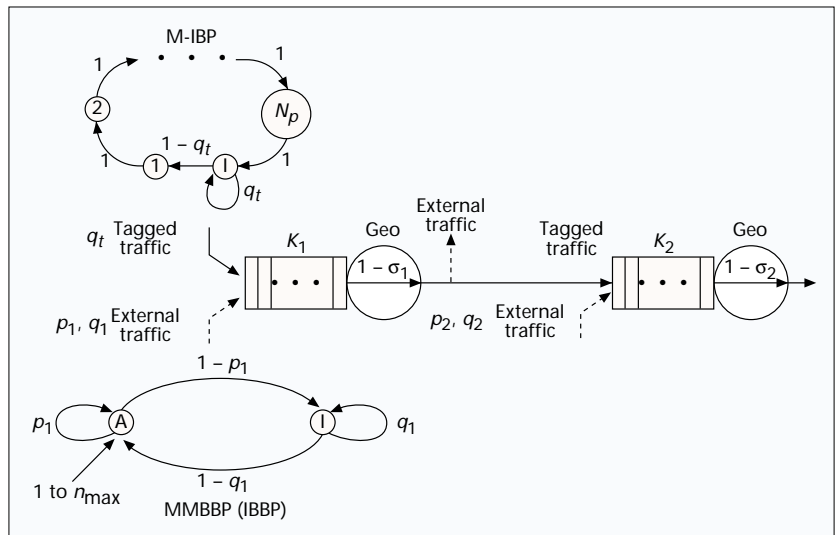


Figure 4. M -IBP + MMBBP/Geo/1/ K tandem queues.

is negligible compared to the total simulation effort needed to obtain the final (accurate) probability estimates.

APPLICATION EXAMPLES

The purpose of these examples is to illustrate the IS techniques for several different network configurations. Although the results obtained using IS in these examples include such performance indices as cell loss and probability mass for queue length, the metric of interest for each technique is speedup. Additional computation overhead is also discussed for each example. Not included in the examples is the analyst's time required to develop the IS solution. In these cases, the analyst's time is justified by the fact that conventional MC simulation would require prohibitively long execution times for many of the points on each plot. Conservatively, computer execution times would be on the order of years in these cases.

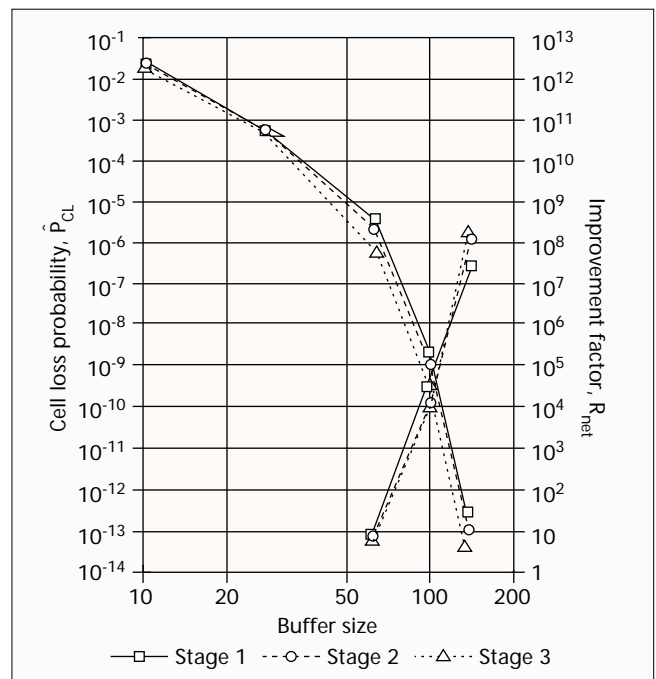


Figure 5. Estimated cell loss probabilities (decreasing curves) and speedup factors (increasing curves) for $q_t = 0.97778$, $N_p = 5$, $K = 150$, $\sigma = 0.01$, $p = 0.3$, $q = 0.825$, and $n_{max} = 5$.

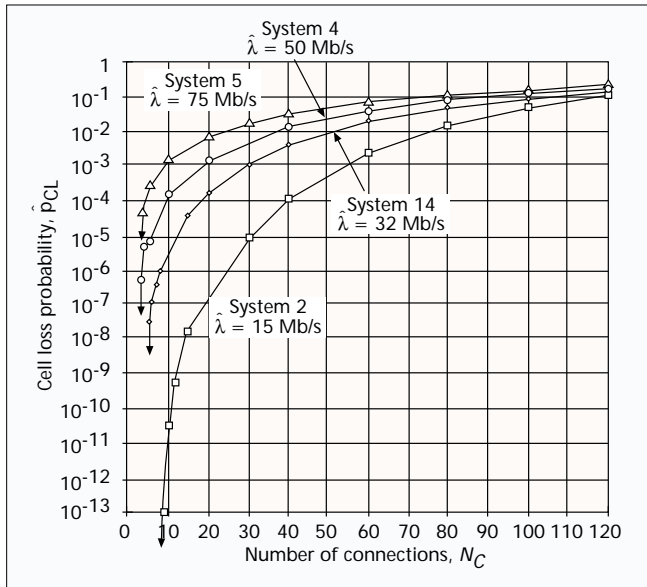


Figure 6. Cell loss probability as a function of the number of connections for $\hat{\lambda}$ varying from 15 to 32, 50, and 75 Mb/s when the remaining system parameters are fixed at $\bar{\lambda} = 1$ Mb/s, $\hat{B} = 200$ cells, $\hat{\mu} = 120$ Mb/s, and $K = 100$ cells.

STEADY-STATE SIMULATION OF CELL LOSS PROBABILITIES

To obtain steady-state estimates of cell loss rates, the *regenerative* method can be used [12]. When the regenerative method is impractical, the technique of *A-cycles* [1] or of approximate regeneration [8] can be used instead.

First, a period of *warmup* is allowed so that the system enters steady state. Then a number of cycles is simulated, both with and without IS. At the end of each nonimportance sampling cycle, the state of the system is stored and an IS cycle follows.

At the end of an IS cycle, the state is restored to that state at the end of the previous nonimportance sampling cycle, and so on. A confidence interval can then be constructed using the batch means technique.

APPLICATION OF STOCHASTIC OPTIMIZATION

We consider the application of the SGD technique to the tandem network of M-IBP+MMBBP/Geo/1/ K queues shown in Fig. 4. A single stage of this slotted time queuing model has one server with a buffer that holds K cells. There are two independent traffic streams entering the first stage. The first stream, called the *tagged* traffic, is modeled by a modified-interrupted Bernoulli process (M-IBP). The second stream, called the *external* traffic, is modeled by a Markov modulated Bernoulli batch process (MMBBP). Tandem networks of M-IBP+MMBBP/Geo/1/ K queues can make up an end-to-end model of the nodes in an ATM network. The geometric server models the link carrying the traffic to the next node in the network.

The estimate of the cell loss probability at the input of the S th stage in the tandem network is obtained by using the SGD algorithm to minimize the estimate of the variance of the average number of tagged cell losses per regenerative cycle (RC) at the S th stage with respect to the bias parameters associated with the traffic source and service processes. The search is started by using the unbiased parameter values for a short queue length, one in which the estimate is obtainable by conventional MC simulation.

The probability of cell loss estimates of the tagged traffic for a three-stage tandem network are shown in Fig. 5. The use of the SGD algorithm resulted in speedup factors over conventional MC simulation of up to eight orders of magnitude for the

rare cell loss probabilities of interest. The SGD search overhead reduces this speedup by less than an order of magnitude.

APPLICATION OF CONDITIONAL BIASING

Consider an ATM switch where, instead of the traditional stochastic models for the source traffic and service characteristics, the operational approach is used to describe the behavior of the system. The traffic for each of N_C connections entering the ATM switch is described by the triplet $(\hat{\lambda}, \bar{\lambda}, \hat{B})$ where $\hat{\lambda}$ is the peak cell rate, $\bar{\lambda}$ is the average cell rate, and \hat{B} is the maximum burst length in cells at the peak cell rate. This approach is used so that there is a match between the simulation parameters and the ATM Forum standardized connection traffic descriptors. The ATM switch has a service rate of $\hat{\mu}$ and a buffer that can hold K cells. Traffic arrives to the switch in the *greedy* pattern where an entire burst of \hat{B} cells arrives at the switch at the peak rate $\hat{\lambda}$ followed by a silence period such that the average rate $\bar{\lambda}$ is maintained. This results in a periodic behavior for the cell arrivals with period $T = \hat{B}\hat{\lambda} / \bar{\lambda}$ arrival slots, where an arrival slot is the time period occupied by one ATM cell. A connection can start in any arrival slot within the T length period, so we describe the connection starting slot position with a uniform probability density.

In the system model we consider, cell losses occur when the connections start close enough together that the server cannot keep up with the arriving cells over a time $t < T$ and a buffer overflow results in cell loss. For the minimum number of connections that causes cell loss, N_{C_0} , we can determine an interval of c_0 arrival slots in which all the connections must start their cell burst in order for cell loss to result. This prior knowledge of what causes the important event to occur is used to condition the selection of the random variables in the simulation. A curve of an example cell loss probability estimate is shown in Fig. 6. The performance curves show the increase in cell loss probability that results from increasing the peak cell rate $\hat{\lambda}$ from 15 to 32, to 50, and to 75 Mb/s, respectively. For each performance curve, the bottom points were obtained by simulation using IS, and the upper points were obtained by conventional MC simulation. The speedup obtained for the bottommost point on each curve is 2.1×10^{10} , 2.2×10^5 , 4.1×10^3 , and 1.8×10^3 , for $\hat{\lambda}$ of 15, 32, 50, and 75 Mb/s, respectively. The downward pointing arrows indicate that cell losses cannot occur without a sufficient number of connections. The simulation overhead involved with the technique is negligible.

See [10, 11] for more complete descriptions of the use of conditional biasing in estimating cell loss, cell delay, and cell delay variation, respectively, in ATM networks.

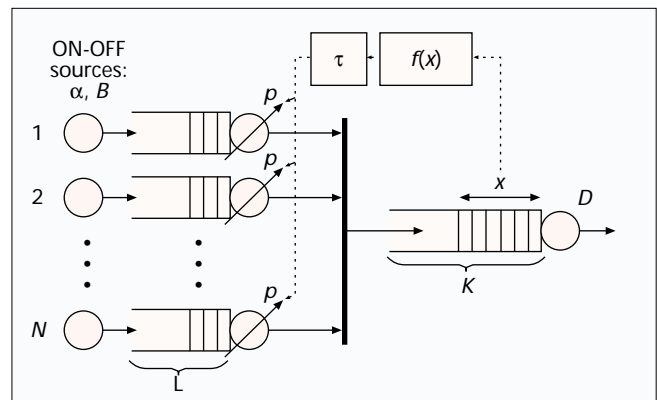


Figure 7. Model of an ATM multiplexer stage with internal traffic control.

APPLICATION OF DPR-BASED SPLITTING SIMULATION

Systems with internal control loops cannot be handled easily by input-biased IS techniques. To demonstrate one of the main advantages of DPR-based trajectory splitting, here we consider a simple system with a control loop: a queuing system with feedback (Fig. 7). The model can be regarded as a generic building block for ATM switches with internal traffic control. Incoming cells from the independent ON/OFF sources are multiplexed to the outgoing main buffer via a nonlocking shared bus. The buffer can store up to K cells. To minimize the occupancy of the main buffer and the probability of losing cells, a feedback signal is broadcast by the main buffer to the input ports. In our generic model the feedback signal is a probability value, p , with which input ports are allowed to pass cells to the bus. Cells that are held back are stored at the input ports in buffers with capacity L cells. The value of p is derived from the actual queue length, x , via the function $p = f(x)$, and the feedback latency is modeled by a constant delay factor, τ .

Figure 8 presents queue length probability mass functions (pmfs) obtained by DPR for the case $K = 84$ cells, $L = 500$ cells, and $\tau = 3$. The number of sources is $N = 16$ and the mean burst length is $B = 3$ cells. We use a two-threshold feedback function where $f(x) = 1$ if $x \leq 44$, $f(x) = 0.5$ if $44 < x \leq 74$, and $f(x) = 0$ otherwise. The mean probability estimates and their 95 percent confidence intervals (both shown in Fig. 8) were obtained from 50 independent DPR replicas per curve, each replica consisting of as few as 10^6 ATM slots. The total simulation time to generate one curve (including the iterative subset balancing procedure) took less than 40 min using a conventional desktop workstation. The corresponding speedup factors vary between 10^4 ($\alpha = 0.6$) and 10^{13} ($\alpha = 0.2$). To demonstrate the effect of feedback, pmf curves for the uncontrolled cases are also shown in Fig. 8.

CONCLUSION

Although application of any IS technique requires a problem-specific analytical phase, these techniques have been applied to a number of different networking problems that required obtaining rare event probabilities. As computer technology continues to advance, simulation will be used to evaluate more complicated network mechanisms. At the same time, more reliable networks will be characterized by even rarer events. Both trends will increase the importance of IS-based techniques for rare event simulation.

REFERENCES

- [1] P. Heidelberger, "Fast Simulation of Rare Events in Queueing and Reliability Models," *ACM Trans. Modeling and Comp. Simulation*, vol. 5, no. 1, Jan. 1995.
- [2] Z. Haraszti and J. K. Townsend, "The Theory of Direct Probability Redistribution and its Application to Rare Event Simulation," *Proc. ICC '98*, June 1998.
- [3] M. Villén-Altamirano *et al.*, "Enhancement of the Accelerated Simulation Method RESTART by Considering Multiple Thresholds," *Proc. ITC 14*, vol. 1a, France, 1994, pp. 797–810.
- [4] P. Shahabuddin, P. Glasserman, and P. Heidelberger, "Splitting for Rare Event Simulation: Analysis of Simple Cases," *1996 Winter Simulation Conf.*, Coronado, CA, Dec. 1996, pp. 302–8.
- [5] J. S. Sadowsky and J. A. Bucklew, "On Large Deviation Theory and Asymptotically Efficient Monte Carlo Estimation," *IEEE Trans. Info. Theory*, vol. 36, no. 3, May 1990, pp. 579–88.
- [6] G. de Veciana, C. Courcoubetis, and J. Walrand, "Decoupling Bandwidths for Networks: A Decomposition Approach to Resource Management," *Proc. IEEE INFOCOM '94*, Toronto, Canada, May 1994.
- [7] M. Falkner, M. Devetsikiotis, and I. Lambadaris, "Issues in Fast Simulation of Networks of Queues," presented at the 18th Ann. Mtg. Canadian Applied Math. Soc., CAMS/SCMA '97, The Fields Institute, Toronto,

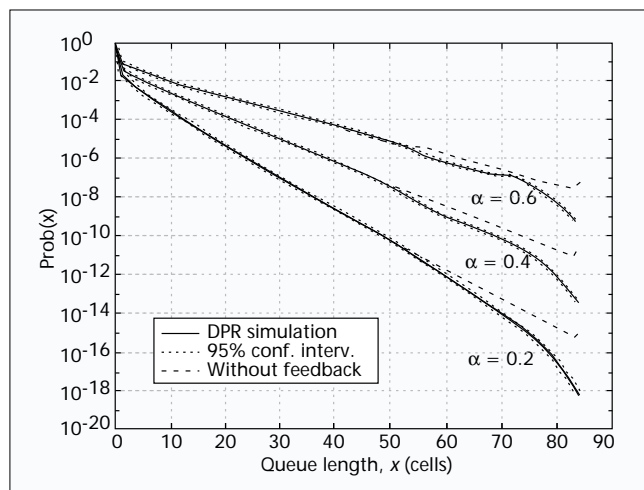


Figure 8. Queue length probability tails for the feedback queuing system, simulated using a trajectory splitting simulation based on DPR.

Canada, May 1997; also at the 4th IEEE Wksp. Architecture and Implementation of High Perf. Commun. Sys., Sani Beach, Halkidiki, Greece, June 1997.

- [8] M. Devetsikiotis and J. K. Townsend, "Statistical Optimization of Dynamic Importance Sampling Parameters for Efficient Simulation of Communication Networks," *IEEE/ACM Trans. Networking*, vol. 1, no. 3, June 1993, pp. 293–305.
- [9] J. A. Freebersyser, M. Devetsikiotis, and J. K. Townsend, "Efficient Simulation of High-Speed Networks Using Importance Sampling and Stochastic Gradient Techniques," *Proc. IEEE GLOBECOM '94*, San Francisco, CA, Nov. 1994, pp. 1095–99.
- [10] J. A. Freebersyser and J. K. Townsend, "Efficient Simulation of Cell Loss Probability in ATM Networks with Heterogeneous Connection Traffic Descriptors," *Proc. ICC '96*, Dallas, TX, June 1996, pp. 320–27.
- [11] A. A. Akyamac and J. K. Townsend, "Conditional Importance Sampling and Its Application to ATM Switch Analysis," *Proc. IEEE ICC '98*, Atlanta, GA, June 1998.
- [12] M. A. Crane and A. J. Lemoine, *An Introduction to the Regenerative Method for Simulation Analysis*, Berlin: Springer-Verlag, 1977.

BIOGRAPHIES

J. KEITH TOWNSEND (jkt@eos.ncsu.edu) received the Ph.D. degree in electrical engineering 1988 from the University of Kansas. Before graduate study he was an engineer in the Avionics Design Group at Bell Helicopter Textron, Fort Worth, Texas. He joined faculty of the Electrical and Computer Engineering Department at North Carolina State University in July 1988, where he is now an associate professor. His current research interests include wireless communication systems, and rare event simulation techniques for communication links and networks.

MICHAEL DEVETSIKIOTIS (mike@sce.carleton.ca) received the Ph.D. degree in electrical engineering from North Carolina State University, Raleigh, in 1993. He has been an assistant professor in the Department of Systems and Computer Engineering, Carleton University, since July 1996. His research has been in telecommunication systems modeling, efficient simulation, and traffic characterization. His present focus is on the performance of broadband communication networks as they become larger in size, and more complex in topology and traffic.

ZSOLT HARASZTI (Zsolt.Haraszti@ericsson.com) received his M.Sc. degree at the Technical University of Budapest, Hungary, in 1993. He joined Ericsson Telecom AB in 1995 as a research fellow of the Traffic Analysis and Simulation Laboratory. Since 1997 he has been a research project manager at Ericsson's Switching Laboratory in Stockholm, Sweden. His main research interest is in communications system modeling, simulation, and network traffic management. His Ph.D. research focus is on efficient simulation of rare events in packet-switched networks.

JAMES A. FREEBERSYSER (freebej@onr.navy.mil) is the program officer in communications and networks at the Office of Naval Research in Arlington, Virginia. He received his degrees in electrical engineering from Duke University (B.S.) in May 1988, the University of Virginia (M.S.) in January 1990, and North Carolina State University (Ph.D.) in December 1995. His research interests are computer-aided modeling, design, simulation, and analysis of communications systems, including both high-speed networks and mobile wireless networks.