Efficient Monte Carlo Simulation of Product Form Systems

P. E. Lassila and J. T. Virtamo
Laboratory of Telecommunications Technology
Helsinki University of Technology
P.O.Box 3000, FIN-02015 HUT, Finland
Email: {Pasi.Lassila, Jorma.Virtamo}@hut.fi

July 1998

Abstract

We discuss efficient Monte Carlo simulation techniques for estimating performance measures of product form systems, e.g. the multiservice loss system. First, two methods for generating the samples are discussed, the traditional Monte Carlo and the so called Gibbs sampler. Then a variance reduction method, the conditional expectation method, is presented where the idea is to utilize known analytical results to the maximum degree. This is possible for systems with a product form probability distribution for which conditional one-dimensional expectations can easily be precomputed. The method is independent of the way the samples are generated. Finally, we study the use of a heuristic importance sampling distribution in connection with the traditional Monte Carlo method and the conditional expectation method.

1 Introduction

In this paper we will be dealing with efficient methods for the Monte Carlo simulation of systems having a product form solution. Systems possessing a product form solution represent an important class of systems and appear in various problem settings such as multiservice loss networks [4] and queuing networks [5]. In principle, product form systems are mathematically simple and well understood, and one is able to write down exact expressions for such things as the blocking probability of a call belonging to a given class. However, for realistic systems, e.g. real multiservice networks with a great number of links and usually a very great number of different traffic classes (a class is defined by the route and the connection attributes of the call), such analytical expressions defy a direct evaluation because of the huge size of the state space.

In such situations, one has to resort to simulations, and when the form of the stationary distribution is known, the Monte Carlo approach is the most obvious choice. Even then getting results with small enough confidence intervals for large systems can be a difficult task. Therefore, acceleration methods for reducing the simulation time are needed in order to allow one to study larger systems.

The sample generation method is an important part of the simulation problem where the main goal is to efficiently generate samples having some desired distribution. We focus on two methods for generating them, the traditional Monte Carlo (i.i.d. samples from the stationary distribution) and the so called Gibbs sampler [1, 6]. In the latter method we construct a Markov chain which has the correct stationary distribution, but which is not derived from e.g. the jump process associated with the arrivals and departures.

We will then describe a general variance reduction technique called the conditional expectation method, which was introduced in [2], and show how it can be applied in multiservice loss systems. The method is based on conditioning on the samples hitting certain subsets of the state space for which conditional expectations can be calculated analytically. In effect it eliminates the internal variance within each subset. The conditional expectation method is independent of the method used for generating the samples.

Finally, we study the use of importance sampling to further reduce variance in the case of rare event simulation. We combine the heuristic importance sampling method by Ross [4, chap. 6] with the conditional expectation method and show by numerical examples how the variance can be made even smaller.

2 The multiservice loss system

Throughout the paper, as an example of a product form system, we consider the calculation of blocking probabilities in a multiservice loss system. Here we only describe the system briefly (see e.g. [4] for a thorough description). Consider a network consisting of J links, indexed with $j = 1, \dots, J$, link j having a capacity of C_j resource units. The network supports K classes of calls. Associated with a class-k call, $k = 1, \dots, K$, is an offered load ρ_k and a bandwidth requirement of $b_{j,k}$ units on link j. Note that $b_{j,k} = 0$ when class-k call does not use link j. Let the vector $\mathbf{b}_j = (b_{j,1}, \dots, b_{j,K})$ denote the required bandwidths of different classes on link j. Also, we assume that the calls in each class arrive according to a Poisson process, a call is always accepted if there is enough capacity left, and that the blocked calls are cleared. The state of the system is described by the vector $\mathbf{X} = (X_1, \dots, X_K)$, where element X_k is the number of class-k calls in progress.

The set of allowed states $\mathcal{S}$ can be described as

$$\mathcal{S} = \{\mathbf{x} \mid \forall j : \mathbf{b}_j \cdot \mathbf{x} \le C_j\}, \tag{1}$$

where the scalar product is defined as $\mathbf{b}_j \cdot \mathbf{x} = \sum_k b_{j,k} x_k$. This system has the well known product form stationary distribution

$$\pi(\mathbf{x}) = \frac{1}{G} \prod_{k=1}^{K} \frac{\rho_k^{x_k}}{x_k!} = \frac{1}{G} \prod_{k=1}^{K} f(x_k, \rho_k) = \frac{f(\mathbf{x})}{G}, \tag{2}$$

where $f(x, \rho) = \rho^x / x!$, and $f(\mathbf{x}) = \prod_k \rho_k^{x_k} / x_k!$ denotes the unnormalized state probability. G is the normalization constant

$$G = \sum_{\mathbf{x} \in \mathcal{S}} f(\mathbf{x}). \tag{3}$$

The set of blocking states for a class-k call, $\mathcal{B}^k$, is

$$\mathcal{B}^k = \{\mathbf{x} \in \mathcal{S} \mid \exists j : \mathbf{b}_j \cdot (\mathbf{x} + \mathbf{e}_k) > C_j\}, \tag{4}$$

where $\mathbf{e}_k$ is a K-component vector with 1 in the k^{th} component and zeros elsewhere. The blocking probability of a class-k call, B_k, is then

$$B_k = \sum_{\mathbf{x} \in \mathcal{B}^k} \pi(\mathbf{x}) = \sum_{\mathbf{x} \in \mathcal{S}} \pi(\mathbf{x}) 1_{\mathbf{x} \in \mathcal{B}^k} = \mathrm{E}[1_{\mathbf{x} \in \mathcal{B}^k}]. \tag{5}$$

It should be noted that the distribution π given by (2) represents the truncation of a K dimensional independent Poisson type distribution to the state space $\mathcal{S}$. Then, by defining another state space $\tilde{\mathcal{S}}$ such that $\tilde{\mathcal{S}} \supseteq \mathcal{S}$ and a random vector $\tilde{\mathbf{X}} \in \tilde{\mathcal{S}}$ with the same product form distribution, $\mathrm{P}\{\tilde{\mathbf{X}} = \tilde{\mathbf{x}}\} \sim f(\tilde{\mathbf{x}})$, we can express the blocking probability as

$$B_k = \mathrm{E}\left[1_{\tilde{\mathbf{X}} \in \mathcal{B}^k} \,|\, \tilde{\mathbf{X}} \in \mathcal{S}\right] = \frac{\mathrm{E}[1_{\tilde{\mathbf{X}} \in \mathcal{B}^k}]}{\mathrm{E}[1_{\tilde{\mathbf{X}} \in \mathcal{S}}]}. \tag{6}$$

We will see later how this formulation can be used advantageously.

To make things concrete, let us consider a system with two traffic classes, see Fig. 1. Finite capacities of different links are reflected in linear constraints on the state space $\mathcal{S}$ and the blocking states are the states on the boundary of the state space. This example will be used throughout the paper.

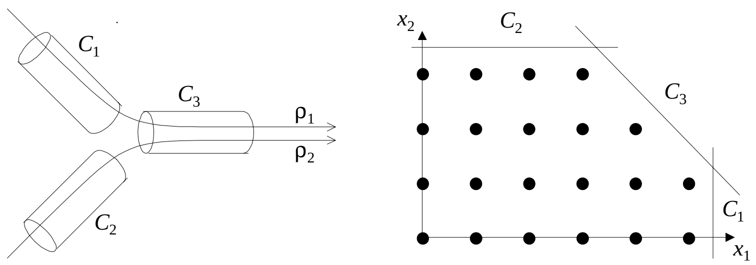

Figure 1: Simple network example and its state space.

3 Sample generation methods

The efficient generation of samples having some desired distribution is an essential part of the simulation problem. In this section, we discuss two methods for generating the samples and show how they are applied in the multiservice loss system context.

3.1 The traditional Monte Carlo method

In the traditional Monte Carlo method the idea is to generate independent identically distributed samples of $\mathbf{X}_n \in \mathcal{S}$ having some desired distribution. In the case of multiservice loss systems one often uses the rejection method which consists of generating samples of $\tilde{\mathbf{X}}$ in the larger space $\tilde{\mathcal{S}}$, where the components of $\tilde{\mathbf{X}}$ are independent, and rejecting those samples which fall outside of the allowed state space $\mathcal{S}$. The accepted samples $\mathbf{X}_n$ will then have the correct distribution π and we can estimate the blocking probabilities from (5) by $\hat{B}_k = 1/N \sum_{n=1}^{N} 1_{\mathbf{X}_n \in \mathcal{B}^k}$, where N is the number of those samples falling inside the allowed state space $\mathcal{S}$.

A particularly suitable choice for $\tilde{\mathcal{S}}$ is the Cartesian product space limited by the maximum number of allowed class-k calls $N^k_{\max}$. Formally this state space is defined as

$$\tilde{\mathcal{S}} = \{0, \ldots, N^1_{\max}\} \times \cdots \times \{0, \ldots, N^K_{\max}\}.$$

This state space has the nice property that the product form solution (2) in $\tilde{\mathcal{S}}$ means that the different components are independent and hence the samples are easy to generate.

We call this sampling distribution as the independent truncated Poisson distribution:

$$p(\mathbf{x}) = \frac{1}{\tilde{G}} \prod_{k=1}^{K} \frac{\rho_k^{x_k}}{x_k!} = \frac{f(\mathbf{x})}{\tilde{G}}, \quad \mathbf{x} \in \tilde{\mathcal{S}}, \tag{7}$$

where

$$\tilde{G} = \prod_{k=1}^{K} \sum_{n=0}^{N_{\max}^k} \frac{\rho_k^n}{n!}.$$

3.2 Gibbs sampler

The previous method was based on generating independent samples from a distribution, but we note that to estimate the blocking probabilities from (5) does not require the samples to be independent. Positive correlation between the samples just makes the estimator less efficient from the point of view of the variance. In Gibbs sampler the idea is to generate a Markov chain having the desired stationary distribution by using transition probabilities based on conditioning.

The idea is based on the following very general property (see [1] or [6] for more details). Again let $\mathbf{X}$ be a vector random variable with state space $\mathcal{S}$ and stationary distribution P. Let $\mathcal{A}_1, \ldots, \mathcal{A}_I$ form a partition of the state space and let $\iota(\mathbf{X})$ denote the unique index of the set to which the state $\mathbf{X}$ belongs to. Then the Markov chain $\mathbf{X}_n$ with the transition probability

$$\mathrm{P}\{\mathbf{X}_{n+1} = \mathbf{y} \,|\, \mathbf{X}_n = \mathbf{x}\} = \mathrm{P}\{\mathbf{X} = \mathbf{y} \,|\, \mathbf{X} \in \mathcal{A}_{\iota(\mathbf{x})}\} \tag{8}$$

has the invariant distribution P.

Let $\mathbf{P}$ denote the transition probability matrix with the components given by (8). The Markov chain generated by this transition matrix is reducible, because there are no transitions between different sets. However, by defining several partitions $1, \ldots, M$ with transition matrices $\mathbf{P}^{(m)}$ the Markov chain $\mathbf{X}_n$ corresponding to the compound transition matrix $\mathbf{P} = \mathbf{P}^{(1)} \cdots \mathbf{P}^{(M)}$ may be irreducible with a suitable choice of the partitions. Then P will be the unique stationary distribution of the chain.

For the purpose of estimating the blocking probability of class-k calls in the multiservice loss system we define the partition to consist of sets in the coordinate direction of traffic class k, which leads to the so called Gibbs sampler. Considering all the traffic classes we have altogether K partitions. Let us denote by $\mathcal{A}_i^k$ the i^{th} set in partition k. Using the two traffic class example, shown in Fig. 2 (left figure) for traffic class 2, the partition consists of the vertical columns. Each set $\mathcal{A}_i^k$ consists of states where the number of calls of all other classes are fixed, but the k^{th} component varies. In general, we refer to sets $\mathcal{A}_i^k$ as k-columns. The set of blocking states $\mathcal{B}^k$ for the class-k calls consists of the end points of the k-columns. The Markov chain $\mathbf{X}_n$ generated by the compound transition matrix $\mathbf{P}$ is irreducible since it is possible to move from any state $\mathbf{x}$ in the coordinate convex state space $\mathcal{S}$ to any other state $\mathbf{y}$ with at most K steps in alternating directions.

The simulation of the Markov chain $\mathbf{X}_n$ consists of making transitions with the transition matrices $\mathbf{P}^{(k)}$ in cyclical order. This is illustrated for the two traffic class example in Fig. 2 (right figure). In transitions generated with $\mathbf{P}^{(k)}$ only the component x_k changes. Starting from state $\mathbf{X}_n$ the value of x_k of the next state is obtained by drawing it from the one- dimensional truncated Poisson distribution $f(x_k, \rho_k)/g(L^k(\mathbf{X}_n), \rho_k)$ with

$x_k \in (0,\ldots,L^k(\mathbf{X}_n))$, $L^k(\mathbf{X}_n)$ denoting the length of the k-column to which the state $\mathbf{X}_n$ belongs, and $g(L,\rho)$ denoting the normalization sum

$$g(L,\rho) = \sum_{l=0}^{L} \frac{\rho^l}{l!}.$$

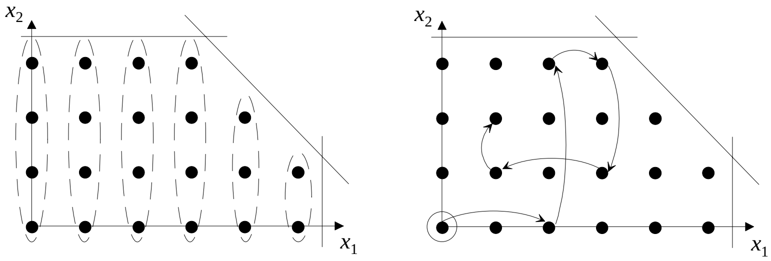

Figure 2: State space partitioning and Gibbs sampler example.

The Gibbs sampler provides a simple way of generating samples with the distribution P without any misses, which occur with the traditional Monte Carlo rejection method. All the samples $\mathbf{X}_n$ have directly the correct distribution P. This property is particularly useful when the traffic loads are high, which would cause a lot of misses to be generated in the Monte Carlo rejection method. Such a situation can occur naturally if the system is very heavily loaded or it can occur when we are using importance sampling for increasing the efficiency of the simulation, as we will see later. Of course, the Gibbs sampler can also be used to generate samples from the independent truncated Poisson distribution, but it does not necessarily give any advantage over generating them in the traditional Monte Carlo method.

4 The conditional expectation method

In this section we first give a general description of a variance reduction method for Monte Carlo simulations, which we have named as the conditional expectation method. Then we show how it can be applied in the context of the multiservice loss system.

4.1 The general formulation

Let us consider a general problem of estimating the expectation

$$H = \mathrm{E}[h(\mathbf{X})] \tag{9}$$

of some function $h(\cdot)$ of a vector random variable $\mathbf{X} \in \mathcal{S}$ with some state space $\mathcal{S}$ and having a distribution P. The Monte Carlo method consists of drawing N independent samples $\mathbf{X}_n,\ n = 1,\ldots,N$, from the distribution P yielding an unbiased estimate

$$\hat{H} = \frac{1}{N}\sum_{n=1}^{N} h(\mathbf{X}_n). \tag{10}$$

Now, the following elementary identity holds always:

$$H = \mathrm{E}[h(\mathbf{X})] = \mathrm{E}[\mathrm{E}\,[h(\mathbf{X})\,|\,g(\mathbf{X})]],$$

where $g(\cdot)$ is another function. Assume now that the conditional expectation

$$\eta(\mathbf{x}) = \mathrm{E}\left[h(\mathbf{X}) \,|\, g(\mathbf{X}) = \mathbf{x}\right]$$

can be calculated analytically. Then the expectation of $h(\mathbf{X})$ becomes

$$H = \mathrm{E}[h(\mathbf{X})] = \mathrm{E}[\eta(g(\mathbf{X}))],$$

Correspondingly, we get a new Monte Carlo estimator for H,

$$\hat{H} = \frac{1}{N} \sum_{n=1}^{N} \eta(g(\mathbf{X}_n)). \tag{11}$$

More specifically, we consider the case where the state space $\mathcal{S}$ has a partitioning into sets $\mathcal{A}_i$, $i = 1, \ldots, I$. A state $\mathbf{X}$ belongs to one and only one of the sets $\mathcal{A}_i$. Let us again denote the unique index of this set by $\iota(\mathbf{X})$. We use this discrete valued function as the function $g(\cdot)$ in the above formulae. So, finally, our estimator is

$$\hat{H} = \frac{1}{N} \sum_{n=1}^{N} \eta(\iota(\mathbf{X}_n)) = \frac{1}{N} \sum_{i=1}^{I} \eta(i) N_i, \tag{12}$$

where N_i is the count of the samples having $\iota(\mathbf{X}_n) = i$ or, equivalently, $\mathbf{X}_n \in \mathcal{A}_i$, and $\eta(i) = \mathrm{E}\left[h(\mathbf{X}) \,|\, \mathbf{X} \in \mathcal{A}_i\right]$. In fact, the latter form could have been written directly from

$$\mathrm{E}[h(\mathbf{X})] = \sum_{i} \mathrm{E}\left[h(\mathbf{X}) \,|\, \mathbf{X} \in \mathcal{A}_i\right] \, \mathrm{P}\{\mathbf{X} \in \mathcal{A}_i\}.$$

Since $\eta(i)$ represents the conditional expectation of $h(\mathbf{X})$ over the set $\mathcal{A}_i$ it is intuitively obvious that the variance of estimator (12) is smaller than that of (10) (see [2] for a formal explanation).

The method described above is simple. However, it is very useful in cases where one is able to define a partition of the state space $\mathcal{S}$ such that the conditional expectations $\eta(i) = \mathrm{E}\left[h(\mathbf{X}) \,|\, \mathbf{X} \in \mathcal{A}_i\right]$ can be calculated analytically for all i. This requirement is nicely fulfilled by systems with product form state probabilities such as the multiservice loss system.

4.2 Application to loss systems

Specifically, in the case of the multiservice loss system, we use the same K partitions as with the Gibbs sampler (see section 3.2), i.e. partition k consists of columns in the direction k, and $\mathcal{A}_i^k$ denotes the i^{th} k-column in partition k. Now, the blocking probability, given by (5), is an expectation of the considered type with $h(\mathbf{X}) = 1_{\mathbf{X} \in \mathcal{B}^k}$. Because of the product form of the state probabilities (2) the conditional expectation $\eta^k(i) = \mathrm{E}\left[1_{\mathbf{X} \in \mathcal{B}^k} \,|\, \mathbf{X} \in \mathcal{A}_i^k\right]$ can be calculated easily. The probability distribution constrained to column i in partition k (set $\mathcal{A}_i^k$) is a truncated Poisson distribution and the conditional blocking probability is given by the Erlang loss function (B formula), $\mathrm{erl}(L_i^k, \rho_k)$, where L_i^k denotes the length of the k-column i (set $\mathcal{A}_i^k$). This leads to the estimator

$$\hat{B}_k = \frac{1}{N} \sum_{n=1}^{N} \mathrm{erl}(L^k(\mathbf{X}_n), \rho_k), \tag{13}$$

where, for clarity, we have written directly $L^k(\mathbf{X})$ (instead of $L^k_{\iota(\mathbf{X})}$) for the length of the k-column to which the state $\mathbf{X}$ belongs.

This rather obvious decomposition does not seem to have been exploited in the simulation context in spite of its significant advantages. Note that in the standard Monte Carlo simulation, a sample point $\mathbf{X}_n$ gives a contribution to the blocking probability only when it hits the set $\mathcal{B}^k$. In contrast, in the proposed method, for each sample point we collect the conditional expectation over the whole column, always containing a blocking state at the end of the column. Further, note that there is no penalty for this advantage as the values of $\mathrm{erl}(L, \rho)$ can be easily precomputed and stored into an array for all the values of L and ρ needed.

4.3 Alternative application to loss networks

In the previous application, the samples $\mathbf{X}_n$ have to be generated in the state space $\mathcal{S}$ from the distribution (2). However, as we noted at the end of Chapter 2, eq. (6), we can define the blocking probability by considering a random vector $\tilde{\mathbf{X}}$ in a larger state space $\tilde{\mathcal{S}}$. Monte Carlo method can be applied both for the numerator and the denominator leading to the estimator

$$\hat{B}_k = \frac{\sum_{n=1}^{N} 1_{\tilde{\mathbf{X}}_n \in \mathcal{B}^k}}{\sum_{n=1}^{N} 1_{\tilde{\mathbf{X}}_n \in \mathcal{S}}}. \tag{14}$$

If the same samples $\tilde{\mathbf{X}}_n$ are used in both the numerator and the denominator this estimator, in effect, reduces to the estimator $\hat{B}_k = 1/N_S \sum_n 1_{\mathbf{X}_n \in \mathcal{B}^k}$, where the samples $\mathbf{X}_n$ are obtained from $\tilde{\mathbf{X}}_n$ by including only those N_S samples which fall within $\mathcal{S}$, i.e. the $\mathbf{X}_n$ are generated by the rejection method.

Something new, however, is obtained when we notice that both the numerator and the denominator of (6) can be estimated with the conditional expectation method. To this end we define K partitions of space $\tilde{\mathcal{S}}$ into sets $\tilde{\mathcal{A}}^k_i$ where the sets in the k^{th} partition consists of k-columns in the space $\tilde{\mathcal{S}}$, as illustrated in Fig. 4. Further, we define the

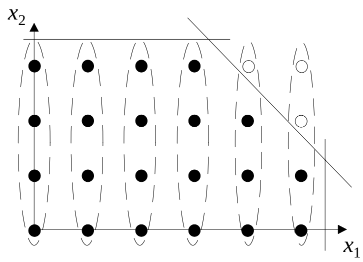

Figure 3: New partitioning of the state space.

conditional expectations

$$\begin{aligned} \eta^k(i) &= \mathrm{E}\left[1_{\tilde{\mathbf{X}} \in \mathcal{B}^k} \,|\, \tilde{\mathbf{X}} \in \tilde{\mathcal{A}}^k_i\right], \\ \vartheta^k(i) &= \mathrm{E}\left[1_{\tilde{\mathbf{X}} \in \mathcal{S}} \,|\, \tilde{\mathbf{X}} \in \tilde{\mathcal{A}}^k_i\right]. \end{aligned}$$

Note that the set $\mathcal{B}^k$ is still formed by the endpoints of the k-columns $\mathcal{A}^k_i$, not by those of the $\tilde{\mathcal{A}}^k_i$. Both of the conditional expectations can be calculated analytically,

$$\eta^k(i) = \begin{cases} 0, & \tilde{\mathcal{A}}^k_i \cap \mathcal{S} = \varnothing, \\ f(L^k_i, \rho_k)/g(N^k_{\max}, \rho_k), & \tilde{\mathcal{A}}^k_i \cap \mathcal{S} \neq \varnothing, \end{cases}$$

$$\vartheta^k(i) = \begin{cases} 0, & \tilde{\mathcal{A}}_i^k \cap \mathcal{S} = \varnothing, \\ g(L_i^k, \rho_k)/g(N_{\max}^k, \rho_k), & \tilde{\mathcal{A}}_i^k \cap \mathcal{S} \neq \varnothing, \end{cases}$$

where, again, L_i^k is the length of the column $\mathcal{A}_i^k$ ($\subseteq \tilde{\mathcal{A}}_i^k$), and, as before, $g(L,\rho) = \sum_{l=0}^{L} \rho^l/l!$.

With the aid of these conditional expectations the blocking probability (6) can be written as

$$B_k = \frac{\mathrm{E}[\eta^k(\iota^k(\tilde{\mathbf{X}}))]}{\mathrm{E}[\vartheta^k(\iota^k(\tilde{\mathbf{X}}))]},$$

where $\iota^k(\tilde{\mathbf{X}})$ now denotes the unique index i of set $\tilde{\mathcal{A}}_i^k$ to which $\tilde{\mathbf{X}}$ belongs. The corresponding Monte Carlo estimator becomes

$$\hat{B}_k = \frac{\sum_{n=1}^{N} \eta^k(\iota^k(\tilde{\mathbf{X}}_n))}{\sum_{n=1}^{N} \vartheta^k(\iota^k(\tilde{\mathbf{X}}_n))} = \frac{\sum_{n=1}^{N} f(L(\tilde{\mathbf{X}}_n), \rho_k) 1_{\tilde{\mathbf{X}}_n^{(k)} \in \mathcal{S}}}{\sum_{n=1}^{N} g(L(\tilde{\mathbf{X}}_n), \rho_k) 1_{\tilde{\mathbf{X}}_n^{(k)} \in \mathcal{S}}}, \tag{15}$$

where $L^k(\tilde{\mathbf{X}})$ denotes the length of the column $\mathcal{A}_i^k$ to which $\tilde{\mathbf{X}}$ belongs, $\tilde{\mathbf{X}}_n^{(k)}$ is the K-vector obtained from $\tilde{\mathbf{X}}_n$ by setting its k^{th} component to 0 (the set $\tilde{\mathcal{A}}_i^k$ to which $\tilde{\mathbf{X}}_n$ belongs has a nonempty intersection with $\mathcal{S}$ if and only if $\tilde{\mathbf{X}}_n^{(k)} \in \mathcal{S}$), and where we have utilized the fact that $g(N_{\max}^k, \rho_k)$ is constant for traffic class k and cancels out.

In this formulation even a sample $\tilde{\mathbf{X}}_n$ falling outside $\mathcal{S}$ can give a contribution to the numerator and denominator of (15).

5 Importance sampling for multiservice loss systems

In this section we introduce the heuristic importance sampling method by Ross for the multiservice loss system and then we combine it with the conditional averaging method to obtain further variance reduction.

5.1 The importance sampling distribution

The Monte Carlo estimate (10) for the expectation $\mathrm{E}[h(\mathbf{X})]$ is not very efficient in cases where the main contribution to this expectation comes from values of $\mathbf{X}$ which are rare under the distribution $P : p(\mathbf{X}) = \mathrm{P}\{\mathbf{X} = \mathbf{x}\}$ defined in the state space $\mathcal{S}$. In such cases the variance can be reduced by making the sampling from another distribution, so called importance sampling distribution, $P^* : p^*(\mathbf{x}) = \mathrm{P}\{\mathbf{X}^* = \mathbf{x}\} > 0$. With respect to this distribution the expectation becomes

$$H = \mathrm{E}[h(\mathbf{X})] = \mathrm{E}_{p^*}\left[h(\mathbf{X}^*) w(\mathbf{X}^*)\right],$$

where $w(\mathbf{X}^*) = p(\mathbf{X}^*)/p^*(\mathbf{X}^*)$ is the likelihood ratio. This leads to the Monte Carlo estimator

$$\hat{H} = \frac{1}{N} \sum_{n=1}^{N} h(\mathbf{X}_n^*) w(\mathbf{X}_n^*).$$

For calculating the blocking probabilities, we can use the form (6) where $\tilde{\mathbf{X}}$ is defined in a larger state space $\tilde{\mathcal{S}} \supseteq \mathcal{S}$. We can now apply the importance sampling both for the numerator and denominator by defining a new distribution $p^*(\cdot)$ in the space $\tilde{\mathcal{S}}$. This results in

$$B_k = \frac{\mathrm{E}_{p^*}\left[1_{\tilde{\mathbf{X}}^* \in \mathcal{B}^k} w(\tilde{\mathbf{X}}^*)\right]}{\mathrm{E}_{p^*}\left[1_{\tilde{\mathbf{X}}^* \in \mathcal{S}} w(\tilde{\mathbf{X}}^*)\right]}, \tag{16}$$

and the corresponding estimator

$$\hat{B}_k = \frac{\sum_{n=1}^{N} w(\tilde{\mathbf{X}}_n^*) 1_{\tilde{\mathbf{X}}_n^* \in \mathcal{B}^k}}{\sum_{n=1}^{N} w(\tilde{\mathbf{X}}_n^*) 1_{\tilde{\mathbf{X}}_n^* \in \mathcal{S}}}. \tag{17}$$

The idea with the importance sampling is to increase the likelihood of more important points in the state space. In the case of estimator (17), in particular, that means that we should try to increase the probability of the event $\tilde{\mathbf{X}}_n^* \in \mathcal{B}^k$ without increasing too much the number of misses $\tilde{\mathbf{X}}_n^* \notin \mathcal{S}$. Another important consideration in choosing the distribution $p^*(\cdot)$ is that the generation of the samples from this distribution and the calculation of the likelihood ratio $w(\cdot)$ should be computationally easy. In practice, compromises have to be made between these somewhat conflicting requirements.

Based on these considerations Ross has derived a heuristic importance sampling distribution to be used for the estimation of blocking probabilities in multiservice loss systems [4, chap. 6]. The distribution is again of the same product form but the actual traffic intensities ρ_k are replaced by virtual loads γ_k. Basically the idea is to increase the offered loads if the load is too light and to decrease the load if it is too high. With this sampling distribution the likelihood ratio is, up to a constant multiplier (which cancels out in the ratio estimator),

$$w(\mathbf{x}) = \prod_{k=1}^{K} \left(\frac{\rho_k}{\gamma_k}\right)^{x_k}.$$

Specifically, the Ross heuristics include rules for selecting the values for the parameters γ_k. The heuristics are rather conservative in that they try to limit the magnitude of the shift in order to avoid a large proportion of misses $\tilde{\mathbf{X}}_n^* \notin \mathcal{S}$. This is in fact a limitation of the Ross heuristics, and e.g. in [3, chap. 4] it is shown that the numerator and denominator in (17) should have a different importance sampling distribution. Note that if Gibbs sampler is used, it is easy to generate the samples $\mathbf{X}_n^*$ directly in $\mathcal{S}$ whence the problem with the misses does not occur at all, and much stronger tilting of the distribution is possible.

5.2 Combining importance sampling and conditional expectation methods

The conditional expectation method can also be applied to the Monte Carlo evaluation of the expectations occurring in the importance sampling result (16). The only difference now is that instead of having $h(\mathbf{X}) = 1_{\mathbf{X} \in \mathcal{B}^k}$ or $h(\mathbf{X}) = 1_{\mathbf{X} \in \mathcal{S}}$ we have to multiply these by the weight factor $w(\mathbf{X})$. The samples $\tilde{\mathbf{X}}_n^*$ are to be generated in the space $\tilde{\mathcal{S}}$ with the tilted distribution $p^*(\cdot)$.

The conditional expectations can again be calculated analytically and this results in the new estimator

$$\hat{B}_k = \frac{\sum_{n=1}^{N} w^{(k)}(\tilde{\mathbf{X}}_n^*) f(L^k(\tilde{\mathbf{X}}_n^*), \rho_k) 1_{\tilde{\mathbf{X}}_n^{(k)*} \in \mathcal{S}}}{\sum_{n=1}^{N} w^{(k)}(\tilde{\mathbf{X}}_n^*) g(L^k(\tilde{\mathbf{X}}_n^*), \rho_k) 1_{\tilde{\mathbf{X}}_n^{(k)*} \in \mathcal{S}}}, \tag{18}$$

where

$$w^{(k)}(\mathbf{x}) = \prod_{j \neq k} \left(\frac{\rho_j}{\gamma_j} \right)^{x_j}.$$

In (18), the one dimensional expectations are essentially those obtained with the actual distribution, i.e. functions $f(\cdot,\cdot)$ and $g(\cdot,\cdot)$ are calculated with load ρ_k instead of γ_k. The effect of the importance sampling is corrected with the weight factor $w^{(k)}(\tilde{\mathbf{X}}_n^*)1_{\tilde{\mathbf{X}}_n^{(k)*} \in \mathcal{S}}$, which in fact is the likelihood ratio for the occurrence of the k-column to which $\tilde{\mathbf{X}}_n^*$ belongs.

Note that if the space $\tilde{\mathcal{S}}$ has a more general shape so that the sets $\tilde{\mathcal{A}}_i^k$ have different sizes, the length of the k-column to which $\tilde{\mathbf{X}}_n^*$ belongs being denoted by $\tilde{L}^k(\tilde{\mathbf{X}}_n^*)$, then the weight factor becomes $w^{(k)}(\tilde{\mathbf{X}}_n^*)1_{\tilde{\mathbf{X}}_n^{(k)*} \in \mathcal{S}}/g(\tilde{L}^k(\tilde{\mathbf{X}}_n^*), \gamma_k)$. In particular, if $\tilde{\mathcal{S}} = \mathcal{S}$ then the divisor in this weight is simply $g(L^k(\tilde{\mathbf{X}}_n^*), \gamma_k)$.

6 Numerical results

Here we compare through numerical examples the efficiency of the different methods presented in this paper. As an example we use the same four link star network as studied by Ross in [4, chap 6] with moderate and light traffic loads (cases 1 and 2 in the tables). Blocking probabilities and the 95% confidence intervals are given for two typical traffic classes (out of 12). We also experiment with a larger network (case 3) where the link capacities have been increased roughly by a factor of 20 and the traffic intensities have been increased correspondingly.

In Table 1 we compare the efficiency of the conditional expectation method (CE method in the tables), with samples generated in $\mathcal{S}$ and $\tilde{\mathcal{S}}$, against those obtained with standard Monte Carlo simulation. From the results one can see that by using the conditional expectation method a significant variance reduction is obtained and that the reduction factor increases as the system size increases. Note that if the standard deviation is reduced by a factor of e.g. 7 as in the case 3 for traffic class 2, this corresponds to a reduction of 49 times in the required number of samples.

Table 1: Blocking probabilities (%) with confidence intervals

Case	Class	Standard MC	CE method in $\mathcal{S}$	CE method in $\tilde{\mathcal{S}}$
1	2	0.295 ± 0.034	0.287 ± 0.010	0.301 ± 0.010
1	8	1.960 ± 0.090	1.945 ± 0.042	1.976 ± 0.031
2	2	0.052 ± 0.014	0.043 ± 0.006	0.046 ± 0.004
2	8	0.360 ± 0.040	0.343 ± 0.011	0.350 ± 0.012
3	2	0.112 ± 0.021	0.116 ± 0.004	0.114 ± 0.003
3	8	0.600 ± 0.049	0.596 ± 0.006	0.595 ± 0.008

In Table 2 we compare the efficiency of the conditional expectation method, again, both in $\mathcal{S}$ and in $\tilde{\mathcal{S}}$ combined with the importance sampling (IS) distribution introduced in Chapter 5, against the standard Monte Carlo method with the same importance sampling distribution (results obtained by Ross). Again clear variance reduction is obtained with the usage of the conditional expectation methods and the reduction factors are roughly of the same order as with the results in Table 1. Note that from the moderate traffic results in Table 2 (cf. the corresponding results in Table 1) we can see that the usage of the CE method in $\tilde{\mathcal{S}}$ gives some variance reduction over the same method in $\mathcal{S}$ when the probability of a sample falling outside $\mathcal{S}$ increases (the offered loads in the sampling distribution were highest for the moderate traffic case in Table 2), which also seems intuitively plausible. This property can prove to be very useful when considering other importance sampling distributions with much heavier tilting.

Table 2: Blocking probabilities (%) with confidence intervals

Case	Class	Standard MC + IS	CE method in $\mathcal{S}$ + IS	CE method in $\tilde{\mathcal{S}}$ + IS
1	2	0.283 ± 0.022	0.287 ± 0.008	0.294 ± 0.007
1	8	1.935 ± 0.065	1.941 ± 0.036	1.957 ± 0.027
2	2	0.041 ± 0.006	0.047 ± 0.002	0.046 ± 0.002
2	8	0.335 ± 0.018	0.348 ± 0.006	0.341 ± 0.007
3	2	0.109 ± 0.017	0.115 ± 0.003	0.113 ± 0.002
3	8	0.594 ± 0.037	0.591 ± 0.010	0.595 ± 0.008

7 Conclusions

In this paper we have considered efficient methods for the Monte Carlo simulation of product form systems such as the multiservice loss system. We have first decomposed the simulation problem into two separate problems, namely that of the method for generating the samples and the method used to collect the contribution from the generated samples. For the sample generation problem we presented two computationally easy methods, the traditional Monte Carlo method and the Gibbs sampler. For the sample contribution collection problem we have presented the conditional expectation method, which gives significant variance reduction, and is easy to apply in systems having a product form. The method is based on partitioning the state space into sets such that within each set the conditional expectation of the estimated function can be calculated analytically, which in effect eliminates the internal variance within the set from the estimator. As a new result, we presented a method where we can define a larger sampling state space and use the conditional expectation method to collect the contribution of even those samples, which fall outside the allowed state space. Finally, we studied the use of a heuristic importance sampling distribution for the sample generation. The used heuristics is not, however, very efficient and suffers from being too conservative. The future research will be directed towards defining more efficient importance sampling distributions.

References

[1] Lassila, P. E., Virtamo J. T., “Using Gibbs Sampler in Simulating Multiservice Loss Systems”, to appear in the proceedings of PICS ’98, Lund, May 25.-28., 1997.

[2] Lassila, P. E., Virtamo J. T., “Variance Reduction in Monte Carlo Simulation of Product Form Systems”, IEE Electronics Letters, vol. 34, no. 12, June 1998, p. 1204-1205.

[3] M. Mandjes, "Rare Event Analysis of Communication Networks", Ph.D. Thesis, Vrije University, Tinbergen Institute Research Series, 1996.

[4] Ross, K. W., "Multiservice Loss Models for Broadband Telecommunication Networks", Springer-Verlag, London, 1995.

[5] Ross, K. W., Wang, J., "Asymptotically Optimal Importance Sampling for Product-Form Queuing Networks", ACM Transactions on Modeling and Computer Simulation, Vol. 3, No. 3, July 1993, pp. 245-268.

[6] Tierney, L. M., "Markov Chains for Exploring Posterior Distributions", The Annals of Statistics, 1994, vol. 22, No. 4, pp. 1701-1728.