Date: October 4, 1998 COST257TD(98)49
Issue: 1

Importance Sampling for Monte Carlo Simulation of Loss Systems

Pasi Lassila [1] and Jorma Virtamo [2]

Helsinki University of Technology

Abstract

In this paper we consider the use of traditional Monte Carlo simulation techniques for estimating very small blocking probabilities in multiservice loss systems. We derive an efficient sampling distribution, which has a composite form consisting of a weighted combination of exponentially twisted distributions. In the literature the composite distribution has been shown to be asymptotically efficient in systems such as the multiservice loss system. However, the asymptotic theory leaves open the question of how to choose the weights. For that we give heuristics to minimize the variance of the estimator by making the likelihood ratio as constant as possible in the set of the blocking states.

1 Introduction

Modern broadband networks have been designed to integrate several service types into the same network. On the call scale, the process describing the number of calls present in the network can be modeled by a loss system, see e.g. [5]. We are interested in calculating the steady state blocking probability for each traffic class in the system. Specifically, we are studying systems where the blocking probabilities are very small, i.e. we are dealing with probabilities of so called rare events. The steady state distribution of the system has the well known product form, from which it is easy to write down analytic expressions also for the blocking probabilities. A problem with the exact solution, however, is that it requires the calculation of a so called normalization constant, which entails the calculation of a sum over the complete allowed state space of the system. In a network of realistic size the state space very rapidly becomes astronomical.

[1] Pasi.Lassila@hut.fi

[2] Jorma.Virtamo@hut.fi

In such a situation we have two alternatives: to use analytical approximations or to simulate the problem to a desired level of accuracy. In this paper we will be dealing with the latter. Then, as the form of the stationary distribution is known, the traditional Monte Carlo (MC) method is an efficient way to perform the simulation. Alternatively, one can utilize the fact that the system can also be represented by a Markov chain and simulate that.

Since we are assuming that the blocking probabilities are very small, using direct simulation becomes inefficient because the required number of samples to be drawn increases exponentially as the estimated probability becomes smaller. To overcome this, it is possible to use importance sampling (IS), whereby one uses an alternative sampling distribution, which essentially makes the interesting samples more likely than under the original distribution. The bias in the samples caused by this is then corrected with the so called likelihood ratio.

Furthermore, we limit ourselves to studying IS distributions, which belong in the family of so called exponentially twisted distributions. Previously this has been studied by Ross in [5, chap. 6] and Mandjes in [4]. When using the traditional MC, the simulation consists of estimating two independent probabilities: the probability of hitting the blocking states and the probability of being in the allowed state space. Ross proposes heuristics based on an observation that when using the traditional MC method and the same IS distribution for both probabilities, then with the optimal, i.e. minimum variance, IS distribution half of the samples will fall within the allowed state space and half of the samples will fall within the blocking states. Based on this observation Ross has presented heuristics which attempt to increase the likelihood of the blocking states, while, at the same, it tries to limit the likelihood of generating misses from the allowed state space. However, Mandjes has noted that it is not advantageous in the rare event context to estimate the two probabilities at the same time using the same sampling distribution. Instead, it is only the probability of hitting the blocking states that requires importance sampling. Mandjes' method is motivated by large deviation results for rare events of sums of independent Poisson variables and he proposes to use an importance sampling distribution which shifts the mean of the sampling distribution to match the most probable blocking state. Essentially the method assumes that to estimate the blocking probability it is enough to identify the most probable blocking state, which also identifies the link where the blocking will occur, and to twist the sampling distribution such that its probability mass will be concentrated around the most probable blocking state.

Our approach is based on using a similar technique as proposed by Mandjes, but we extend the ideas based on the large deviation results obtained by Sadowsky et al. in [6]. There it is shown that sometimes, depending on the shape of the "interesting" set, it is not sufficient to use one twisted distribution to satisfy the conditions of asymptotical optimality for the IS distribution. Instead, one needs to use a composite distribution, which is a weighted convex combination of several exponentially twisted distributions. In this paper we illustrate that the blocking probability estimation problem has the same properties as studied in [6] and we present an IS distribution for estimating the blocking probabilities, which is of the composite form. However, the results in [6] leave open the question about the choice of the weights in the composite distribution. We propose heuristics for that, which also have a sound theoretical motivation. The slight increase in computational complexity when compared with just using a single twisted distribution appears to be well justified by the gains in the variance reduction and accuracy obtained in our numerical experiments.

This paper is organized as follows. Section 2 describes briefly the multiservice loss system and the associated simulation problem is characterized in section 3. Section 4 contains the main results of this paper, which is the derivation of the composite distribution. Numerical examples are given in section 6 and the conclusions are presented in section 7.

2 The multiservice loss system

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

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

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

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

where $f(x_k, \rho_k) = \rho_k^{x_k}/x_k!$ and G is the normalization constant

$$G = \sum_{\mathbf{x} \in \mathcal{S}} \prod_{k=1}^{K} \frac{\rho_k^{x_k}}{x_k!}.$$

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

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

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

$$B_k = \sum_{\mathbf{x} \in \mathcal{B}^k} \pi(\mathbf{x}) = \sum_{\mathbf{x} \in \mathcal{S}} \pi(\mathbf{x}) 1_{\mathbf{x} \in \mathcal{B}^k} = \mathrm{E}\left[1_{\mathbf{x} \in \mathcal{B}^k}\right]. \tag{2}$$

It should be noted that the distribution π given by (1) 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, $\mathrm{P}[\tilde{\mathbf{X}} = \tilde{\mathbf{x}}] \sim \prod_{k=1}^{K} \rho_k^{x_k}/x_k!$, we can express the blocking probability as

$$B_k = \mathrm{E}\left[1_{\tilde{\mathbf{X}} \in \mathcal{B}^k} \mid \tilde{\mathbf{X}} \in \mathcal{S}\right] = \frac{\mathrm{E}[1_{\tilde{\mathbf{X}} \in \mathcal{B}^k}]}{\mathrm{E}[1_{\tilde{\mathbf{X}} \in \mathcal{S}}]}. \tag{3}$$

3 The traditional Monte Carlo method

Let us consider a general problem of estimating the expectation

$$H = \mathrm{E}\left[h(\mathbf{X})\right]$$

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 traditional Monte Carlo method consists of drawing N independent samples $\mathbf{X}_n,\ n = 1, \ldots, N$, from the distribution P yielding an unbiased estimate

$$\hat{H} = \frac{1}{N} \sum_{n=1}^{N} h(\mathbf{X}_n). \tag{4}$$

The blocking probabilities in the multiservice loss system, given by (3), are of the above form, and we have the following estimator

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

To generate independent identically distributed samples $\tilde{\mathbf{X}}_n \in \tilde{\mathcal{S}}$, it is desirable that the components of $\tilde{\mathbf{X}}_n$ be independent to make the generation of the samples easy. Then a suitable choice for $\tilde{\mathcal{S}}$ is the Cartesian product space limited by the maximum number of allowed class-k calls $N_{\max}^k$. Formally this state space is defined as

$$\tilde{\mathcal{S}} = \{0, \ldots, N_{\max}^1\} \times \cdots \times \{0, \ldots, N_{\max}^K\}.$$

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

$$\tilde{p}(\mathbf{x}) = \frac{1}{\tilde{G}} \prod_{k=1}^{K} \frac{\rho_k^{x_k}}{x_k!}, \quad \mathbf{x} \in \tilde{\mathcal{S}}, \tag{5}$$

where

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

4 Importance sampling for multiservice loss systems

The Monte Carlo estimate (4) 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, $P^* : p^*(\mathbf{x}) = \mathrm{P}[\mathbf{X}^* = \mathbf{x}] > 0$ for $\mathbf{X}^* \in \mathcal{S}$. With respect to this distribution the expectation becomes

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

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

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

A well known result is that the optimal choice for the IS distribution in this case would be $p^*(\mathbf{x}) = p(\mathbf{x})h(\mathbf{x})/H$ for $\mathbf{x} \in \mathcal{S}$, i.e. it has the property that it makes each sample $w(\mathbf{X}_n^*)h(\mathbf{X}_n^*)$ a constant and hence the variance of the estimator zero, see e.g. [2]. This sampling distribution is, however, impractical, since it requires explicit information about the estimated quantity itself.

For calculating the blocking probabilities, we can use the form (3) where $\tilde{\mathbf{X}}$ is defined in $\tilde{\mathcal{S}}$. We can now apply the importance sampling both for the numerator and denominator by defining a new distribution $\tilde{p}^*(\mathbf{x})$ in the space $\tilde{\mathcal{S}}$. This results in

$$B_k = \frac{\mathrm{E}_{\tilde{p}^*}\left[1_{\tilde{\mathbf{X}}^* \in \mathcal{B}^k} w(\tilde{\mathbf{X}}^*)\right]}{\mathrm{E}_{\tilde{p}^*}\left[1_{\tilde{\mathbf{X}}^* \in \mathcal{S}} w(\tilde{\mathbf{X}}^*)\right]}, \tag{6}$$

and the corresponding estimator

$$\hat{B}_k = \frac{\sum_{n=1}^{N} w(\tilde{\mathbf{X}}_n^*) 1_{\tilde{\mathbf{X}}_n^* \in \mathcal{B}^k}}{\sum_{n=1}^{N} w(\tilde{\mathbf{X}}_n^*) 1_{\tilde{\mathbf{X}}_n^* \in \mathcal{S}}}. \tag{7}$$

However, the blocking probability is a ratio of two independent expectations and thus they can be can be estimated separately in (7). By observing that as we are interested in estimating potentially very small blocking probabilities, the denominator, i.e. the probability of hitting the allowed state space $\mathcal{S}$, does not involve rare event estimation under the original distribution $\tilde{p}(\cdot)$. Hence to estimate $\mathrm{E}[1_{\tilde{\mathbf{X}} \in \mathcal{S}}]$ we do not need to apply importance sampling. In fact, under the original distribution $\tilde{p}(\cdot)$, when the network is very lightly loaded as is the case when the blocking probabilities are very small, the expectation $\mathrm{E}[1_{\tilde{\mathbf{X}} \in \mathcal{S}}] = \mathrm{P}[\tilde{\mathbf{X}} \in \mathcal{S}] \approx 1$. However, for the numerator, we can apply importance sampling and thus we are trying to find an efficient importance sampling distribution for estimating the numerator in (7)

$$\hat{H} = \sum_{n=1}^{N} w(\tilde{\mathbf{X}}_n^*) 1_{\tilde{\mathbf{X}}_n^* \in \mathcal{B}^k},$$

where the samples $\tilde{\mathbf{X}}_n^* \in \tilde{\mathcal{S}}$ have distribution $\tilde{p}^*(\tilde{\mathbf{x}})$.

What is a "good" IS distribution in this case ? When using the distribution $\tilde{p}^*(\cdot)$ the only parameters to be defined are the load parameters in the twisted distribution. When a single link is the bottleneck for each traffic class, then the idea would be basically to choose the parameters such that the probability mass of the sampling distribution is shifted to the most likely blocking state, as suggested e.g. in [4]. We consider here a more realistic setting, where for each traffic class the probability of blocking is not dominated by a single link in the network but it can be roughly equal on many of the links the traffic class uses in the network. The described situation would in reality correspond to a well engineered system, where each link in the network would be roughly equally loaded. This situation is dealt with in [4] only for the case when there exists several maximizers for the optimization problem (10), which give the same maximum value for the objective function. This can happen when for example the network has some symmetric structure. This is handled by approximating the blocking probabilities to be independent on each link and then using twisted distributions to estimate the blocking probability on each link separately. Then, by the assumed independence, the total blocking probability is the sum of individual link blocking probabilities. However, this is only an approximation, which is somewhat in contradiction with what we want in the first place when doing simulations. What we actually want is an accurate estimate for the real blocking probability B_k and not an estimate for an approximation of B_k.

4.1 Motivational Arguments

Let us consider a general problem of estimating the probability of event $\mathcal{B}$

$$\beta = \mathrm{P}[\mathbf{X} \in \mathcal{B}] = \mathrm{E}[1_{\mathbf{X} \in \mathcal{B}}],$$

where $\mathbf{X} \in \mathcal{S}$ has some distribution $p(\mathbf{x})$ and $\mathcal{B} \subseteq \mathcal{S}$. Now we wish to get insight into how the IS distribution should be chosen. By denoting with $p^*(\mathbf{x})$ the IS distribution, we can express the expectation as

$$\beta = \mathrm{E}_{p^*}[w(\mathbf{X})], \tag{8}$$

where

$$w(\mathbf{X}) = \frac{p(\mathbf{X})}{p^*(\mathbf{X})} 1_{\mathbf{X} \in \mathcal{B}}.$$

Also, let $\beta^* = \mathrm{E}_{p^*}[1_{\mathbf{X} \in \mathcal{B}}]$ be the probability of event $\mathcal{B}$ with respect to the distribution $p^*(\mathbf{x})$. By conditioning on the value of the random variable $I = 1_{\mathbf{X} \in \mathcal{B}}$ we can express (8) as

$$\begin{aligned} \beta &= \mathrm{E}_{p^*}[w(\mathbf{X})] \\ &= \mathrm{E}_{p^*}[\mathrm{E}_{p^*}[w(\mathbf{X}) | I]] \\ &= \beta^* \mathrm{E}_{p^*}[w(\mathbf{X}) | I = 1]. \end{aligned}$$

Thus we have $\mathrm{E}_{p^*}\left[w(\mathbf{X})|I=1\right] = \beta/\beta^*$. The variance of $w(\mathbf{X})$ is then

$$\begin{aligned}
\mathrm{V}_{p^*}\left[w(\mathbf{x})\right] &= \mathrm{V}_{p^*}\left[\mathrm{E}_{p^*}\left[w(\mathbf{X})|I\right]\right] + \mathrm{E}_{p^*}\left[\mathrm{V}_{p^*}\left[w(\mathbf{X})|I\right]\right] \\
&= \mathrm{E}_{p^*}\left[\mathrm{E}_{p^*}\left[w(\mathbf{X})|I\right]^2\right] - \mathrm{E}_{p^*}\left[\mathrm{E}_{p^*}\left[w(\mathbf{X})|I\right]\right]^2 + \beta^*\sigma^{*2} \\
&= \frac{\beta^2}{\beta^*} - \beta^2 + \beta^*\sigma^{*2}, \qquad (9)
\end{aligned}$$

where $\sigma^{*2} = \mathrm{V}_{p^*}\left[w(\mathbf{X})|I=1\right]$ is the variance of $w(\mathbf{X})$ in $\mathcal{B}$ under $p^*(\mathbf{x})$. From this formulation we are able to get insight into the effect of the IS distribution.

When no twisting is used and $p^*(\mathbf{x}) = p(\mathbf{x})$ then also $\beta^* = \beta$ and $\sigma^{*2} = 0$. In this case

$$\mathrm{V}_p\left[w(\mathbf{X})\right] = \beta(1-\beta) \approx \beta.$$

By increasing the probability β^* of $\mathcal{B}$ under $p^*(\mathbf{x})$, the first and most important term in (9) can be reduced. Ideally, if one can make $\beta^* = 1$, the first and second term completely cancel. If the probability of $\mathcal{B}$ can be increased uniformly, i.e. with $w(\mathbf{x})$ constant in $\mathcal{B}$, then $\sigma^{*2} = 0$ and the estimator would have a zero variance.

In practice one is limited to a family of twisted distributions, and one has to compromise between the two factors. It is important to increase the probability of $\mathcal{B}$ but at the same time it is important to keep $w(\mathbf{x})$ as constant as possible in $\mathcal{B}$ in order to minimize σ^{*2}.

4.2 Derivation of the IS distribution

The basic idea is to search all dominating points (most probable blocking states) on all the links of the network, which gives us an importance sampling distribution for effectively sampling the blocking states on each link j. This is illustrated in Fig. 1, which shows an example with two traffic classes. Now we can choose one of these points and twist the original distribution such that the main mass of the distribution is centered around that point. In this way we can maximize β^*. However, if we use only one point then the distribution of $w(\mathbf{x})$ in $\mathcal{B}$ will be uneven giving rise to a large σ^{*2}. Therefore, it is more advantageous to use a composite distribution which is a weighted combination of the individual twisted distributions such that $w(\mathbf{x})$ will be more uniform in the whole $\mathcal{B}$.

This approach is supported by the results of [6], which show that so called asymptotically optimal twisting distribution for $\mathcal{B}$ indeed is of this composite form. A failure to include all the dominating points means that the sampling distribution can be asymptotically very inefficient. The asymptotic theory, however, leaves the weights in the composite distribution open. We will fix them by the heuristics of maximal uniformity of $w(\mathbf{x})$ in $\mathcal{B}$.

The most probable blocking states are found as a result of a simple maximization problem of a Poisson-type distribution on a given linear constraint. To search the dominating point

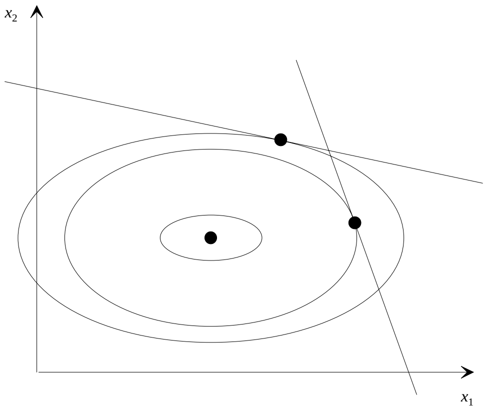

Figure 1: Most probable blocking states in a two traffic example with two link constraints.

on each link we need to solve the following optimization problem for all $j = 1, \ldots, J$:

$$\max_{\mathbf{x}} \quad \frac{1}{\tilde{G}} \prod_{k=1}^{K} \frac{\rho_k^{x_k}}{x_k!} = \prod_{k=1}^{K} f(x_k, \rho_k)$$

$$\text{subject to} \quad \mathbf{x} \cdot \mathbf{b}_j = C_j.$$

This problem has been considered elsewhere in the literature (e.g. [3] and [4]) and usually it has been solved by using first the Stirling's approximation for the factorial term and then solving the constrained optimization problem by standard techniques. However, the optimal solution can be found easily even without using the Stirling's approximation. The expression can then be solved effectively by using standard numerical techniques. For this, consider the following equivalent optimization problem (the term $1/\tilde{G}$ is just a constant and has been omitted)

$$\max_{\mathbf{x}} \quad \sum_{k=1}^{K} \log f(x_k, \rho_k) \tag{10}$$

$$\text{subject to} \quad \mathbf{x} \cdot \mathbf{b}_j = C_j.$$

Introducing the Lagrange multiplier θ for the linear constraint, we are lead to the unconstrained maximization of the following objective function

$$\max_{\mathbf{x}} \sum_{k=1}^{K} \left[\log f(x_k, \rho_k) - \theta x_k b_{j,k}\right] - \theta C_j. \tag{11}$$

Let us denote by

$$h(x_k, \rho_k) = \frac{\partial}{\partial x_k} \log f(x_k, \rho_k).$$

Maximization of (11) with respect to the x_k, leads to the solution

$$x_k = h^{-1}(\theta b_{j,k}, \rho_k), \quad k = 1, \ldots, K, \tag{12}$$

where $h^{-1}(\cdot, \rho)$ denotes the inverse function of $h(\cdot, \rho)$. It is easy to check that $h(\cdot, \cdot)$ is a monotonously decreasing function of x_k and hence the inverse function is also well defined. The Lagrange multiplier θ is determined simply by requiring that the solution must satisfy the constraint of (10), i.e.

$$\sum_{k=1}^{K} b_{j,k} h^{-1}(\theta b_{j,k}, \rho_k) = C_j. \tag{13}$$

Hence, the solution to (10) is obtained by first solving (13) and then evaluating (12) for $k = 1, \ldots, K$. Let us now denote by $\mathbf{x}_j^*$ the solution for link j. This solution represents the "most likely" blocking state on the link j.

At this point we restrict ourselves to only consider IS distributions belonging in the family of exponentially twisted distributions. In case of the Poisson distributions these are also Poisson distributions but with different load parameters $\boldsymbol{\gamma}$, instead of $\boldsymbol{\rho}$. Now the main mass of the sampling distribution can be moved to the point $\mathbf{x}_j^*$ by selecting $\boldsymbol{\gamma}_j = \mathbf{x}_j^*$. Then, let us denote by $\tilde{p}_j^*(\tilde{\mathbf{x}})$ a truncated Poisson density of the form (5), where the load parameter of each class $\gamma_{j,k} = x_{j,k}^*$. To be precise, each distribution $\tilde{p}_j^*(\tilde{\mathbf{x}})$ is given by

$$\tilde{p}_j^*(\tilde{\mathbf{x}}) = \frac{1}{\tilde{G}_j^*} \prod_{k=1}^{K} f(\tilde{x}_k, \gamma_{j,k}),$$

where $\tilde{G}_j^*$ is the normalization constant

$$\tilde{G}_j^* = \prod_{k=1}^{K} \sum_{n=0}^{N_{\max}^k} f(\tilde{x}_k, \gamma_{j,k}).$$

These distributions will effectively move the probability mass of the sampling distribution close to the capacity constraint of link j giving an efficient sampling distribution for the blocking states resulting from link j. In fact, if these parameters were used in sampling from independent infinite dimensional Poisson distributions, the twisting would correspond to moving the mean of the distribution to the most probable blocking state. This is again illustrated in Fig. 2, which shows the sampling distributions of a two traffic class example with two link constraints.

However, a traffic class uses only a subset of all the links in the network. Thus, when estimating the blocking probability of traffic class k, we only need to sample on those links, which the traffic class k actually uses and the complete sampling distribution for traffic class k is then a weighted combination of those distributions. For this let us denote by R_k the set of links the traffic class k uses. Formally we have that

$$R_k = \{j \in 1, \ldots, J \mid b_{j,k} > 0\}.$$

Also, let J_k denote the number of links in the set R_k. Then the sampling distribution for traffic class k is expressed as

$$\tilde{p}^*(\tilde{\mathbf{x}}) = \sum_{j \in R_k} P_j \, \tilde{p}_j^*(\tilde{\mathbf{x}}), \tag{14}$$

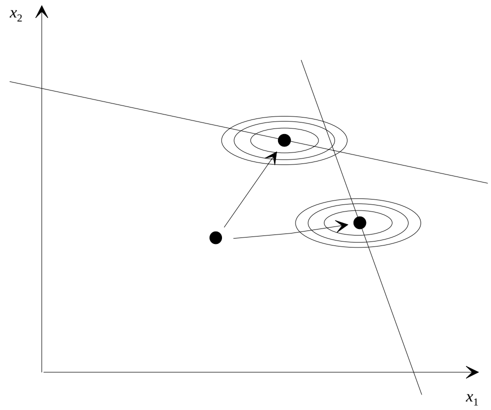

Figure 2: The twisted distributions in a two traffic example with two link constraints.

where P_j is the probability of using the distribution $\tilde{p}_j^*(\cdot)$ for generating the sample. What remains is to find a method to obtain these weights. Again, the used heuristics is based on attempting to keep the likelihood ratio as constant as possible in $\mathcal{B}^k$. With J_k parameters P_i available, we cannot make $w(\mathbf{x})$ constant in the whole $\mathcal{B}^k$. Instead, we choose to require $w(\mathbf{x})$ to be constant, η, at the dominating points $\mathbf{x}_j^*$. This requirement leads to a set of linear equations

$$\sum_{i \in R_k} P_i \ \tilde{p}_i^*(\mathbf{x}_j^*) = \eta \tilde{p}(\mathbf{x}_j^*), \quad \forall j = 1, \ldots, J, \tag{15}$$

where the constant η is chosen such that $\sum_i P_i = 1$.

5 Numerical examples

Here we illustrate the efficiency of using the presented composite distributions for evaluating blocking probabilities in a multiservice loss system by considering the numerical example from [1], where Heegaard uses an adaptive importance sampling scheme in a Markov chain simulation setting. The example corresponds to a network where usually the blocking probability of a class is not dominated by a single link constraint. The network has 11 links and 10 traffic classes. However, the link sizes are small enough to permit the calculation of the exact solution. We selected as examples three traffic classes (2, 4 and 6) to illustrate the differences in accuracy when using just one twisted distribution and the composite distribution. Using just a single twisted distribution for sampling corresponds to the simulation method given in [4]. In this case the load parameters are chosen to be the components of $\mathbf{x}_j^*$ giving the overall largest value for (10) over $j \in R_k$. Also, the method of Ross [5, chap. 6] corresponds to the use of a single twisted distribution. However, the method suffers from being too conservative and e.g. in the examples here the method is not able to produce any estimate even after 1 000 000 samples. To compare the results we compute the relative difference of the estimate, given by $(\hat{B}_k - B_k)/B_k$, and its estimated 95% confidence interval. For each example we give the results when using $N = 10\ 000$ or $N = 100\ 000$ samples in the simulation.

class	B_k	N	Single twist	Composite distribution
2	$0.587 \cdot 10^{-9}$	10 000	-0.329 ± 0.020	-0.014 ± 0.032
2	$0.587 \cdot 10^{-9}$	100 000	-0.317 ± 0.026	-0.015 ± 0.020
6	$0.244 \cdot 10^{-9}$	10 000	-0.184 ± 0.013	-0.086 ± 0.051
6	$0.244 \cdot 10^{-9}$	100 000	-0.111 ± 0.054	0.025 ± 0.089
4	$0.186 \cdot 10^{-9}$	10 000	-0.051 ± 0.031	-0.043 ± 0.032
4	$0.186 \cdot 10^{-9}$	100 000	0.016 ± 0.039	-0.019 ± 0.027

Table 1: The relative difference for the estimates and the estimated 95% confidence intervals.

From the results we can clearly see the better accuracy of using the composite distribution, see especially the results of traffic class 2. However, for traffic class 4 it could be seen from the results of the likelihood maximization problem that there is basically only a single link where the main contribution to the blocking probability comes. Then, as the results in the table show, it is sufficient to use only a single twisted distribution.

For the cases covered here Heegaard obtained results with a relative accuracy of approximately $10\% - 20\%$, but the estimated confidence intervals were wide enough to include also the correct values. However, to obtain the results the simulation required 15 replicas of 300 000 so called regenerative cycles (paths starting from an empty system and ending there).

The results also reveal another interesting phenomenon which appears when using IS: the variances are heavily under estimated. When using IS and a very heavily twisted distribution, it appears that the value of the estimated variance is usually quite far from the correct value and hence the estimated confidence intervals can be very far from the truth. The estimated results appear to be very accurate judging by the estimated variance, but, as can be seen, the known correct value does not even lie within the confidence intervals. This is again a manifestation of the observation made earlier that in a "good" IS distribution also the variance of the likelihood ratio in the "interesting set" should be minimized. When using just a single twisted distribution the problem is that the likelihood ratio can have a huge value in some points in the state space, but under the twisted distribution these points have a very small probability. Then during the simulation we can never observe these significant points even though they would give a huge contribution, and hence also the estimates are not accurate. The composite method is not totally immune to this either, since by comparing the reduction in the variance of the estimator when increasing the number of samples from 10 000 to 100 000, the variance is not reduced correspondingly by the factor 10. In fact, from the results of traffic class 6, we can see that although the relative accuracy got better when increasing the number of the samples the estimated variance actually became worse.

6 Conclusions

In this paper we considered the efficient simulation of very small blocking probabilities in a multiservice loss system when using the traditional Monte Carlo method for the simulation. Previous research on efficient IS distributions for simulating loss systems has considered the use of only one exponentially twisted distribution. We derived an IS distribution for estimating the blocking probabilities, which has the form of a composite distribution. The distribution is a weighted combination of several exponentially twisted distributions, each of which corresponds to a distribution for effectively sampling the blocking states on a single link. In [6] the composite form has, in fact, been shown to be asymptotically optimal in the case of e.g. the loss system. However, the asymptotic theory leaves open the choice of the weights. For this, we have also presented heuristics, which try to minimize the variance of the samples within the set of the blocking states. This is done by choosing the weights such that the likelihood ratio in the most probable blocking states is a constant. The numerical results confirm the accuracy of this method, when compared with just using a single exponentially twisted distribution.

References

[1] P. E. Heegaard, "Efficient Simulation of Network Performance by Importance Sampling", Teletraffic Contributions for the Information Age, Proceedings of ITC-15, vol. 2a, Elsevier, Netherlands, 1997, pp. 623–632.

[2] P. Heidelberger, "Fast Simulation of Rare Events in Queuing and Reliability Models", ACM Transactions on Modeling and Computer Simulation, vol. 5, no. 1, 1995, pp. 43–85.

[3] F. P. Kelly, "Blocking Probabilities in Large Circuit Switched Networks", Advances in Applied Probability, vol. 18, 1986, pp. 473–505.

[4] M. Mandjes, "Fast simulation of blocking probabilities in loss networks", European Journal of Operations Research, Vol. 101, 1997, pp. 393-405.

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

[6] J. S. Sadowsky, J. A. Bucklew, "On Large Deviations Theory and Asymptotically Efficient Monte Carlo Estimation", IEEE Transactions on Information Theory, vol. 36, no. 3, 1990, pp. 579–588.