Proceedings of the 2004 Winter Simulation Conference
R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, eds.

AN ADAPTIVE APPROACH TO FAST SIMULATION OF TRAFFIC GROOMED OPTICAL NETWORKS

Chih-Chieh Hsu
Michael Devetsikiotis

ECE Department
North Carolina State University
Raleigh, NC 27695, U.S.A.

ABSTRACT

In this paper, we consider the fast simulation of traffic groomed optical networks, in which multiple sub-rate traffic streams may be carried on the same wavelength. For real-sized systems, call blocking probabilities may be rare events and require a large amount of CPU power to simulate. We present two importance sampling methods for fast simulation. For a light-load case, we prove that static IS using the Standard Clock (S-ISSC) method does indeed have bounded relative error (BRE) even in multi-class case. For a non-light-load case, we suggest, alternatively, *adaptive ISSC* (A-ISSC) which calculates the relative possibility of reaching each target in every step of simulation. Using A-ISSC, biasing methods which are proven to be optimal or have bounded error can be extended to multi-dimensional cases while still retaining a very favorable performance.

1 INTRODUCTION

Modern optical techniques such as wavelength division multiplexing (WDM) enable the capability of carrying several Terabits per second by using multiple wavelengths, each of which can carry traffic streams at the order of Gigabits per second, in each fiber. This satisfies modern applications which have huge bandwidth demands; however, in many cases, a traffic stream may only need a small fraction of the wavelength. To prevent the wasting of resources, the bandwidth of a wavelength can be divided into smaller sub-rate capacities called sub-wavelength units. A customer can require one or more sub-wavelengths to a maximum of the bandwidth of a wavelength. This technique is known as *traffic grooming* (see Dutta and Rouskas 2002). In this paper, we consider an optical network with *single-hop grooming*. In single-hop grooming, a traffic stream will not be switched to another wavelength along its source-destination path. The nodes that connect the links are add/drop multiplexers (ADMs). An ADM is the place where some of the traffic goes through while some other is dropped, which means that the traffic stream is directed to local traffic. Meanwhile, new traffic may be added from local sources, if there is sufficient capacity remaining.

A new connection, usually referred as a new *call*, may arrive at any node and have a destination of any other node. Along its path, it will require some amount of the sub-wavelength units according to its bandwidth requirement. If the requirement of a call cannot be fulfilled when it arrives, the call will be blocked. The probability that a new call will be blocked is, thus, an important indicator of the Quality of Service provided by this network. For Poisson arrivals and exponential holding times, this type of system is called a *multi-service loss system* (see Ross 1995). Closed form solutions may exist for some problems of this type; however, in real-sized networks, solutions are difficult to calculate due to the enormous state space. Simulation is, therefore, a favored and rather indispensable method to estimate the blocking probability, as well as other QoS parameters.

In large optical networks, call blocking probabilities may be *rare events* due to the large capacity of the network or a very low call arrival rate. In such cases, standard simulation may require an extremely long runtime, and usually incurs large relative errors. *Importance Sampling*, (see Heidelberger 1995) has been known as a technique to improve the accuracy of estimates of stochastic events, which permits large speed-ups of estimation of extremely low blocking probabilities. The system under study is simulated in a way that the “important" events occur more frequently by “biasing" the underlying probability distribution.

Methods for accelerating the simulation of packet-based networks of queues have been known for a while (see Heidelberger 1995; Smith, Shafi, and Gao 1997; Falkner, Devetsikiotis, and Lambadaris 1999, and references therein). Closer to the problem at hand, several methods have been proposed to use Importance Sampling to estimate the blocking probabilities in a *multi-service loss network*, e.g., Ross (1995) and Mandjes (1997) focused on the estimation of the most likely blocking link; and Lassila and Virtamo (1999 and 2000) provided methods for the cases in which more than one link may have contribution to the blocking probability. These methods are based on the product form solution and may need to calculate the very large table for the probability distribution before the simulation begins.

In this paper, we provide methods based on Importance Sampling applied to the Standard Clock method (ISSC) (Vazquez-Abad, Andrew, and Everitt 2002). In that paper, the authors showed that in the single-class case, when the arrival rates approach zero, a method called static ISSC has bounded relative error in the estimation. Here, we extend their result to a *multi-class* scenario. Furthermore, in the non-light-load case, we propose using *Adaptive ISSC*, which is based on Adaptive Importance Sampling (Heegaard 1996), that tunes the probability distribution toward the most possible target in each step.

The remainder of the paper is organized as follows. In Section 2, we provide an introduction to the model of traffic groomed optical network. In Section 3, we describe briefly the technique of Importance Sampling, as well as the two ISSC methods in question. Finally, we show validating simulation results in Section 4, and conclude in Section 5.

An example of a single-hop grooming optical network can be found in Washington and Perros (2004). For simplicity, consider a tandem network of multi-rate loss models with simultaneous resource possession. In our simulation method, this topology can be extended to a more generalized mesh network without any effort. For example, Figure 2 illustrates an optical network with seven nodes and six links. Each link is assumed to be a single fiber with W wavelengths. For single-hop grooming, this network can be seen as W identical networks, each having one wavelength in every link.

Figure 2 also shows the possible routes in the network; that is, traffic of a call may arrive at any node a, and leave at any other node b, where $b > a$. All calls, whatever route they use, are grouped into K classes. Each class $k, k = 1, 2, \ldots, K$ has a bandwidth demand of d_k sub-wavelength units. An arriving class k call at node a with destination b will be accepted only when every link between a and b has at least d_k sub-wavelength units available. All of these sub-wavelength units are assigned to the call simultaneously at the time it is accepted. If a call is not accepted, it is considered *blocked*. When a call departs, all d_k sub-wavelength units on all of the links along its route are simultaneously released.

2 A MODEL FOR TRAFFIC GROOMED OPTICAL NETWORKS

In general, we may assume that there are L links in an optical network, each having a capacity of C sub-wavelength units.

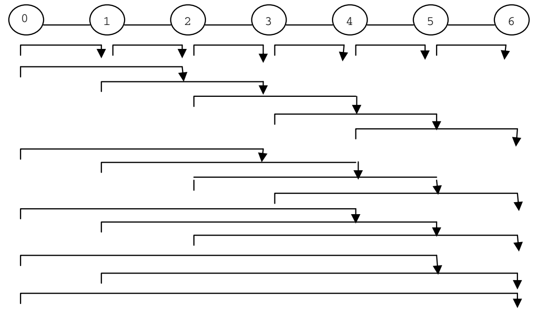

Figure 1: Seven-node Network with All Routes

To one of the R routes, say i, class k calls arrive at rate $\lambda_{i,k}$ and have mean service duration of $\frac{1}{\mu_{i,k}}$. As noted, such calls will be accepted only if there are at least d_k sub-wavelength units available in all of the links along route i.

3 IMPORTANCE SAMPLING TECHNIQUES

3.1 Basic Idea of Importance Sampling

Importance Sampling (IS) is a Monte Carlo (MC) estimation technique which aims to reduce the variance or other cost function of a given simulation estimator. Consider a case in which we want to estimate the probability $P(X \in A)$ for a random variable X with probability density function (pdf) $f(\cdot)$. Traditional MC generates samples of X and counts the number in A, that is, by calculating $\mathrm{E}[\mathbf{1}_{\{X \in A\}}]$. If $P(X \in A)$ is very small, this would require a large number of samples. Using IS, we generate samples with a pdf $f^*(\cdot)$ with $P^*(X \in A) > P(X \in A)$. If each time we observe an x within A, we increment our count by $\frac{f(x)}{f^*(x)}$ instead of 1, then, effectively, we are constructing a new "weighted" random variable, the expectation of which is also equal to $P(X \in A)$:

$$\int_{x \in A} \frac{f(x)}{f^*(x)} f^*(x) dx = \int_{x \in A} f(x) dx = P(X \in A).$$

The function $L(x) = \frac{f(x)}{f^*(x)}$ is called the Radon-Nikodym derivative or likelihood ratio. For a Markovian type system, if we have a sample path s with t steps, the Radon-Nikodym derivative would be the multiplication of each step along s, that is,

$$L(s) = \prod_{k=1}^{t} \frac{f_k(x_{k+1})}{f_k^*(x_{k+1})}$$

where $f_k(x_{k+1})$ and $f_k^*(x_{k+1})$ denote the original and twisted transition probability from state x_k to x_{k+1}.

As long as $L(x)$ can be calculated exactly [which also implies that $f^*(x)$ must be absolutely continuous with respect to $f(x)$, that is, whenever $f(x) > 0$, $f^*(x)$ must

> 0], the IS estimator is *statistically unbiased*. However, deciding how to choose an appropriate $L(x)$ so that the simulation is most efficient and results in small variance is far from trivial, depending on the system of application. A widely used criterion for evaluating bias efficiency is determining if it has *bounded relative error (BRE)* (see Shahabuddin 1994 and Heidelberger 1995 among others). The definition is as follows:

Definition 1 *The unbiased IS estimator for the rare event probability $P^\epsilon(X \in A)$, $E[L(x)\mathbf{1}_{\{x\in A\}}]$, has bounded relative error (BRE) under $f^*(x)$ if there are constants $\beta < \infty, \epsilon_0 > 0$ such that*

$$\sup_{\epsilon \le \epsilon_0} \frac{\sqrt{\text{Var}^*[L(X)\mathbf{1}_{\{X\in A\}}]}}{P^\epsilon(X \in A)} \le \beta.$$

The following lemma is a direct consequence of the above definition (Chang, Heodelberger, and Shahabuddin 1995).

Lemma 2 *If there are constants l, u and b such that $P^\epsilon(X \in A) \ge l\epsilon^b$ and $L(X)\mathbf{1}_{\{X\in A\}} \le u\epsilon^b$ a.s., then the IS estimator for $P^\epsilon(X \in A)$ has BRE.*

3.2 Static ISSC for Light-Load Case

The static ISSC method is introduced in Vazquez-Abad, Andrew, and Everitt (2002) to estimate the call blocking probability in a *cellular* telephone network. In Andrew (2004), a dynamic method is used to estimate the call blocking probability in wavelength continuous WDM networks with one class of traffic. Using a similar notation as in Andrew (2004), we define $c_{i,j}$ as the union of routes that use the j_{th} link of route i, and C_i, which is called cluster i, to be the union of all routes that intersect route i. For route i, valid states can be written as {Aggregate occupancy $\widetilde{n}$ s.t. $n^{(c_{i,L(m)})} \le W$, $\forall m = 1, \dots, M$} , where $n^{(c_{i,j})}$ is the total number of sub-wavelength units used in $c_{i,j}$; i.e., capacity occupancy of link j, $L(m) \in \{1, \dots, L\}$ is the m_{th} link that route i uses, and M is the total number of links used by route i. In other words, an aggregate occupancy $\widetilde{n}$ is said to be valid for route i if every link that the route uses does not exceed its capacity.

Assume inter-arrival times and call durations are exponentially distributed. Under the *Standard Clock* simulation approach (Vakili 1991), we consider applying Importance Sampling. We would like to change the arrival rates and durations of some connections so that the blocking probability for the traffic that we are interested in, say class k at route i, becomes a non-rare event. The *event rate* for the system under the system occupancy $\widetilde{n}(t)$ is then

$$\Lambda_{\widetilde{n}(t)} = \sum_{k=1}^{K}\sum_{j=1}^{R} \lambda_{j,k} + \mu \sum_{k=1}^{K}\sum_{j=1}^{R} \widetilde{n}_{j,k}(t)$$

where $\widetilde{n}_{j,k}(t)$ is the number of class k, route j calls in the system at step t.

Consider the change of measure that swaps the aggregate arrival rates to cluster i, $\lambda = \sum_{k=1}^{K} \sum_{j\in C_i} \lambda_{i,k}$ with inverse expected call duration μ. That is, in IS, arrivals at the cluster C_i have a total rate $\lambda^* = \mu$, and the service rate for the connection in the cluster is $\mu^* = \lambda$. Inter-arrival and holding times outside the cluster have the original exponential distribution. The new event rate becomes

$$\begin{aligned}\Lambda^*_{\widetilde{n}(t)} &= \mu + \sum_{k=1}^{K}\sum_{j\notin C_i} \lambda_{j,k} + \lambda \sum_{k=1}^{K}\sum_{j\in C_i} \widetilde{n}_{j,k}(t) \\ &\quad + \mu \sum_{k=1}^{K}\sum_{j\notin C_i} \widetilde{n}_{j,k}(t),\end{aligned}$$

and the corresponding Radon-Nikodym derivative after T steps is

$$\begin{aligned}L_T &= \prod_{t=1}^{T-1} \frac{\Lambda_{\widetilde{n}(t)} e^{-\Lambda_{\widetilde{n}(t)} T_{t+1}}}{\Lambda^*_{\widetilde{n}(t)} e^{-\Lambda^*_{\widetilde{n}(t)} T_{t+1}}} \\ &\quad \times \prod_{t=1}^{T-1} \frac{\Lambda^*_{\widetilde{n}(t)}}{\Lambda_{\widetilde{n}(t)}} (H(t))\end{aligned}$$

where T_{t+1} means the time from step t to step $t+1$, and

$$H(t) = \begin{cases} \dfrac{\lambda}{\mu}, & \text{event } t+1 = \text{ arrival to cluster i} \\ \dfrac{\mu}{\lambda}, & \text{event } t+1 = \text{ departure from cluster i} \\ 1, & \text{o.w.} \end{cases}$$

Let a be the total arrivals to cluster i in T steps, and d be the total departures to cluster i in T steps,

$$L_T = e^{-\sum_{t=1}^{T-1} (\Lambda_{N(t)} - \Lambda^*_{N(t)}) T_{t+1}} \left(\frac{\lambda}{\mu}\right)^a \left(\frac{\mu}{\lambda}\right)^d$$

$$
\begin{aligned}
&\Lambda^*_{N(t)} - \Lambda_{N(t)} \\
= &\left(\mu + \sum_{k=1}^{K}\sum_{j\notin C_i}\lambda_{j,k} + \lambda\sum_{k=1}^{K}\sum_{j\in C_i}\widetilde{n}_{j,k}(t)\right. \\
&\left. +\mu\sum_{k=1}^{K}\sum_{j\notin C_i}\widetilde{n}_{j,k}(t)\right) \\
&-\left(\sum_{k=1}^{K}\sum_{j=1}^{R}\lambda_{j,k} + \sum_{k=1}^{K}\sum_{j=1}^{R}\widetilde{n}_{j,k}(t)\mu\right) \\
= &-(\mu-\lambda)(\sum_{k=1}^{K}\sum_{j\in C_i}\widetilde{n}_{j,k}(t)-1).
\end{aligned}
$$

Therefore,

$$
L_T = e^{-(\mu-\lambda)\left[\sum_{t=1}^{T-1}(\sum_{k=1}^{K}\sum_{j\in C_i}\widetilde{n}_{j,k}(t)-1)T_{t+1}\right]}\left(\frac{\lambda}{\mu}\right)^{a-d}. \quad (1)
$$

Finally, assume that $B_{i,k}$ to be the event that the system reaches the blocking rate at route i for class k

$$
\begin{aligned}
P(B_{i,k}) &= E\left[1(B_{i,k})\right] \\
&= E^*\left[L_T\cdot 1(B_{i,k})\right]
\end{aligned}
$$

We use the A-cycle method (see Heidelberger 1995), in which a quasi-regenerative set A is defined. Two consecutive entries into A from A' is called an *"A-cycle"*. Under this method, the blocking probability can be expressed as

$$
\begin{aligned}
P(B) &= \frac{E[T_B]}{E[T_A]} \\
&= \frac{E[T_B|A_B]P(A_B)}{E[T_A]}
\end{aligned}
$$

where T_A, T_B and A_B are the A-cycle length, time spent in blocking states in an A-cycle, and the event that an A-cycle contains blocking state, respectively. These three are all functions of i and k, that is, the traffic in which we are interested. We can estimate $P(A_B)$ by turning on IS when the system state reaches A from A' until it reaches the blocking state. Then we turn off IS and perform ordinary simulation from this point until it leaves the set A, and $E[T_B|A_B]$ can be calculated. Moreover, for $E[T_A]$, we can simply use the ordinary simulation. This simulation can also be used to provide the *entry points* for the IS to use. Figure 2 shows the illustration of A-cycle method and also shows that an ordinary simulation can be used in more than one set of IS which may focus on the blocking state of different routes or links. In this paper, we define A to be the set of all states in which the number of calls in cluster C_i is greater than some threshold θ.

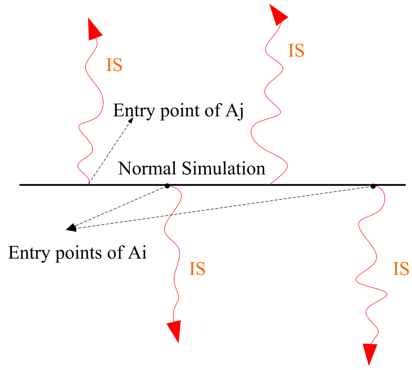

Figure 2: A-cycle Method

We prove in the following that, if the call arrival rates $\lambda_{i,k} = \varphi_{i,k}\epsilon$ and $\mu > \lambda$, when $\epsilon \longrightarrow 0$, then this method will have a Bounded Relative Error (BRE).

Proof First, we show that $L_T\mathbf{1}_{\{A_B\}} \le u\epsilon^b$.

When we reach a blocking state for route i, class k calls, there are at least $\frac{C}{Max\{d_k\}}$ active calls within cluster i at time T. Moreover, from the definition of A, the active calls within this cluster will be at least $\frac{\theta}{Max\{d_k\}} + 1$.

Therefore,

$$
(\mu-\lambda)\left[\sum_{t=1}^{T-1}(\sum_{k=1}^{K}\sum_{j\in C_i}\widetilde{n}_{j,k}(t)-1)T_{t+1}\right] \ge 0.
$$

Furthermore, since

$$
a-d \ge \frac{C-\theta}{Max\{d_k\}},
$$

we have

$$
\left(\frac{\lambda}{\mu}\right)^{a-d} < \left(\frac{\lambda}{\mu}\right)^{\frac{C-\theta}{Max\{d_k\}}}.
$$

From (1), we get

$$
\begin{aligned}
L_T\mathbf{1}_{\{A_B\}} &\le \left(\frac{\lambda}{\mu}\right)^{\frac{C-\theta}{Max\{d_k\}}} \le \left(\frac{\lambda}{\mu}\right)^{\frac{C}{Max\{d_k\}}} \\
&= u\epsilon^b
\end{aligned}
$$

, where

$$
u = \left(\sum_{j\in C_i}\sum_{k}\varphi_{j,k}/\mu\right)^{\frac{C}{Max\{d_k\}}}
$$

and

$$
b = \frac{C}{Max\{d_k\}}.
$$

Next, we need to show that $P(A_B) \geq l\epsilon^b$.

Consider a sample path in which only $\frac{C}{Max\{d_k\}}$ class k^*, route i calls arrive, where k^* is the class such that $d_{k^*} = Max\{d_k\}$.

This will bring the system into a blocking state, and the probability of this sample path is a lower bound of $P(A_B)$.

We have

$$\begin{aligned} P(A_B) &\geq \left(\frac{\lambda_{i,k^*}}{\Lambda_{\tilde{n}}}\right)^{\frac{C}{Max\{d_k\}}} \\ &\geq \left(\frac{\varphi_{i,k^*}}{C\mu + \sum_i \sum_k \lambda_{i,k}}\right)^{\frac{C}{Max\{d_k\}}} \epsilon^{\frac{C}{Max\{d_k\}}} \\ &= l\epsilon^b \end{aligned}$$

where $l = \left(\frac{\varphi_{i,k^*}}{C\mu+\sum_i \sum_k \lambda_{i,k}}\right)^{\frac{C}{Max\{d_k\}}}$.

From all of the above and Lemma 2, it follows that this method has BRE. □

3.3 Adaptive ISSC

The assumption of light load in the last section does not seem to be always practical in real networks. In this section, we would like to consider instead the situation where the arrival rates do not go to zero. Blocking probabilities may still be rare events for all or some calls due to *high capacity* or fast service. As to the method of IS biasing on this kind of problem, several studies have been published. Among the M links a route uses, if call blocking events happen in a certain link, say l', far more frequently than in all of the other links, we can estimate the call blocking probability by biasing the distribution toward the overwhelming of link l'. This is called a *single target system* and is discussed in Ross (1995) and Mandjes (1997). However, modern design of optical networks would prevent the existence of this kind of bottleneck and try to balance the load among all links. Under such a model, biasing the system toward a certain link will cause an *overbiasing* problem (Smith, Shafi, and Gao 1997) and, thus, underestimate the blocking probability. In Lassila and Virtamo (1999) and Lassila and Virtamo (2000), the authors provide methods to distribute blocking among links, but these depend on the product form solution and may have to pre-calculate a large amount of tables before simulation.

The idea of Adaptive Importance Sampling (Heegaard 1996) is to sample a link before each step according to the relative importance of each link, then bias the distribution as if the link is the only target. After an arrival or a departure, another sampling of links is done and the biasing distribution may also be changed. These steps are repeated until the system reaches a blocking state. This idea is shown in Figure 3, where a simplified model of four links and two traffic types is shown.

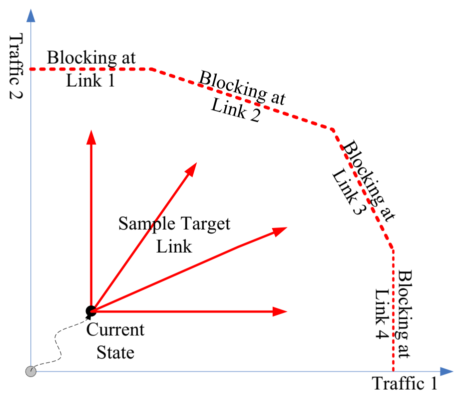

Figure 3: Adaptive Importance Sampling

Assume that we want to sample the links from a probability distribution $\widehat{f_{\tilde{n}(t)}}(j)$, where $j = 1, \ldots, L$, and $\tilde{n}(t)$ is the aggregate system state at step t. Recall that if we want to use IS to estimate the probability that $X \in A = \{A_1, \ldots, A_N\}$, the "best" biasing is to sample only the elements in A and preserve the proportion of their relative probability (Heidelberger 1995). Of course, this requires the understanding of the probability we want to estimate and, thus, is not feasible. However, biasing the probability so that the relative probabilities of importance samples are preserved, does help to prevent over-biasing problems and reduce the estimator variance. Therefore, we want $\widehat{f_{\tilde{n}(t)}}(j)$ to be an *approximation* of the relation of the true blocking probabilities among the links, starting from state $\tilde{n}(t)$. We calculate $\widehat{f_{\tilde{n}(t)}}(j)$ by estimating how easily the blocking state of link j can be achieved from the current state $\tilde{n}(t)$. We use the probability that the system will be in the blocking state of link j in the *shortest path* from $\tilde{n}(t)$ to that state as the weight of link j. $\widehat{f_{\tilde{n}(t)}}(j)$ is a probability distribution that is proportional to the weight of link j.

To calculate the weight, we consider the shortest path, which is the most likely path to reach the blocking state from the current state. For a system with arrival and call duration independent of system state, the shortest path would be successive arrival of route i, class k traffic whose $d_k \frac{\lambda_{i,k}}{\mu_{i,k}}$ is largest among the routes that use class j until the system reaches the blocking state of route j. Since $d_k \frac{\lambda_{i,k}}{\mu_{i,k}}$ is *independent* of system states, this can be calculated in advance; and in each step, we only need to calculate M weights, one for each link used by the route in which we are interested.

If the arrival and/or the call duration are dependent on the system state, we have to consider all possible sample paths, which may require an excessive amount of computing time. One possibility is to choose the largest $d_k \frac{\lambda_{i,k}(\tilde{n})}{\mu_{i,k}(\tilde{n})}$ for each step and use this resulting path as the shortest path. This method requires $O(M \cdot K \cdot N(c))$ computations, where $N(c)$ is the mean value of the number of routes that use link j over the M links that the route in which we are interested is using.

When the target link j is decided, we can then bias the system toward that link. Exponential biasing, which is suggested in the single target system, can be used; or, for calculation efficiency, change of aggregate arrivals of the routes that use link j with the aggregate service rate of the active calls that use link j, which has asymptotically optimal in single-class method, may also be used.

Using the Standard Clock method, our algorithm using Adaptive ISSC is as follows:

1. For each j link that the route in which we are interested uses, find the shortest path from current state to the blocking state of the link and then calculate the probability that the system will be in the blocking state. Call this probability w_j.
2. Sample a target link $\widehat{j}$ from the distribution that $P(j) = \frac{w_j}{\sum_k w_k}$.
3. Bias the system in favor of $\widehat{j}$, sum up the arrival and departure rates, sample the next event and determine its attributes.
4. Calculate the Radon-Nikodym derivative for this step as

$$\begin{aligned} L(t) &= \frac{\Lambda_{\widetilde{n}(t)} e^{-\Lambda_{\widetilde{n}(t)} T_{t+1}}}{\Lambda^*_{\widetilde{n}(t)} e^{-\Lambda^*_{\widetilde{n}(t)} T_{t+1}}} \\ &\quad \times \frac{\Lambda^*_{\widetilde{n}(t)}}{\Lambda_{\widetilde{n}(t)}} (H(t)) \\ &= e^{(\Lambda_{N(t)} - \Lambda^*_{N(t)}) T_{t+1}} H(t) \end{aligned}$$

   where T_{t+1} means the time to the next event, and

   $H(t)$ differs according to the biasing method used. If we switch the arrival rate to $\widehat{j}$ with the aggregate inverse average call duration at $\widehat{j}$,

$$H(t) = \begin{cases} \frac{\lambda}{\mu}, & \text{event} = \text{ arrival to } \widehat{j} \\ \frac{\mu}{\lambda}, & \text{event} = \text{ departure from } \widehat{j} \\ 1, & \text{o.w.} \end{cases}$$

5. Repeat steps 1-4 until the system reaches any of the blocking states. The overall Radon-Nikodym derivative of this sample path is the product of the derivatives of all steps.

In Adaptive ISSC, we still apply the A-cycle method as described in the previous section.

4 SIMULATION MODEL AND RESULTS

4.1 Simulation Model

Consider a seven-ADM tandem optical network as in Figure 2. Each link between two adjacent ADMs is modeled as a single wavelength that is groomed to carry 24 sub-wavelength units. All 21 possible routes are considered, and for each route assume there are 4 classes of traffic, that is, K=4. The demands for sub-wavelength units of the four classes are $d_1 = 2, d_2 = 6, d_3 = 10$, and $d_4 = 16$.

Assume that the routes are numbered from 1 to 21 from the top-left route to the very bottom one. For example, the route between node 0 and node 1 is route 1, the route between node 0 and node 2 is route 7, and the route that uses all of the links is route 21, etc. The arrival rates (packets per second) are assigned as Table 1.

Table 1: Call Arrival Rates of Simulation

Route Class	1	2	3	4
1	1	0.99	0.98	0.97
2	0.96	0.95	0.94	0.93
3	0.92	0.91	0.9	0.89
4	0.88	0.87	0.86	0.85
5	0.84	0.83	0.82	0.81
6	0.80	0.79	0.78	0.77
7-18	0.1	0.1	0.1	0.1
19-20	0.03	0.03	0.03	0.03
21	0.01	0.01	0.01	0.01

The service rate for all calls is set to be 500 packets per second. Inter-arrival times and call duration for all calls are assumed to be exponentially distributed with the above rates.

Ordinary Monte-Carlo simulation, Static ISSC and Adaptive ISSC are all performed for this scenario using ARENA 7.0 as the simulation tool. For Monte Carlo simulation, a 5000-day simulation is done for each of 30 independent replications. This requires more than 2 real days in a Pentium-4 1.7G machine dedicated to the simulation. For both of the ISSC methods, θ is set to be 0, that is, A-cycles are defined to be the time between the beginning of two consecutive busy periods of link j. An ordinary simulation is done for 10 days for each method. Both methods use the first 10,000 entry points of their own ordinary simulation. The method of batch means is used to calculate the variance of the blocking rate using both ISSC methods. For the A-ISSC method, we use an interchanging of the aggregate arrival rate with the aggregate inverse average duration. The simulation time of the blocking probability for *all* 84 route/classes using S-ISSC and A-ISSC is approximately 65 minutes and 2.5 hours, respectively.

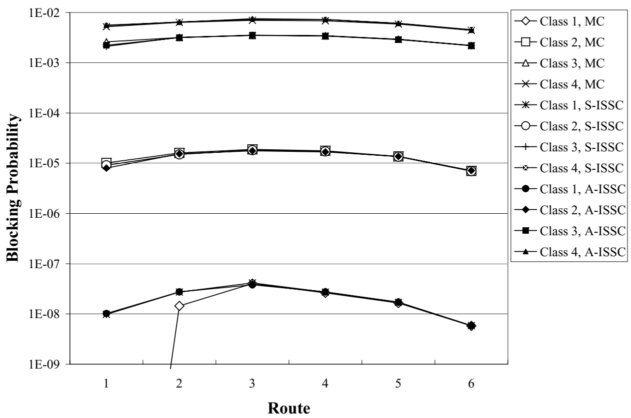

Figure 4: Routes that Use One Link

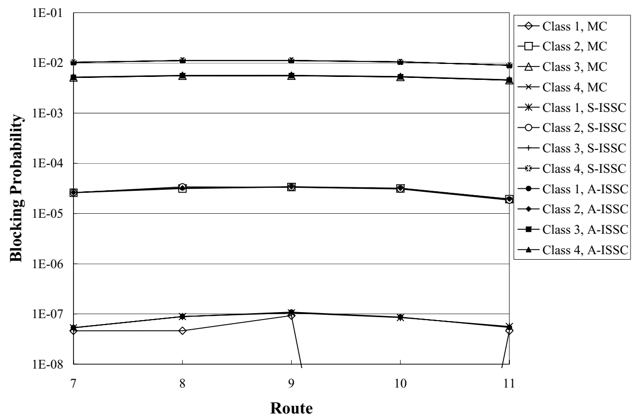

Figure 5: Routes that Use Two Links

4.2 Simulation Results

The call blocking probabilities for different routes and classes are shown in Figures 4, 5 and 6. From the figures, we see that even for such long execution time, blocking probabilities for some of the traffic streams are still too small to estimate by the ordinary Monte Carlo method. This can also be seen from the large relative error derived from the ordinary simulation of the class 1 traffic, which typically has blocking probabilities less than 10^{-7}. Table 2 shows the estimation of call probabilities and relative errors in percentage for the class 1 calls that use two links, i.e., route 7 through route 11, under all three methods. Notice that for route 10, traditional Monte Carlo simulation fails to estimate the blocking probability. Meanwhile, we can see that the relative error is much larger in ordinary simulation than in the two IS methods that we propose.

Figure 4 shows the call blocking probabilities of the routes that use only one link, i.e., route 1 through route 6. For these routes, we simply have a single target system. From the figure, we see that the IS methods do very well in estimating the probabilities. For multi-target cases, Figure 5 and Figure 6 represent the blocking probabilities for the routes that use two links and the routes that use four to six links, respectively. From the figures we can see that the estimates from the two IS methods are consistent with each other.

Moreover, besides some of the class 1 cases that cannot be successfully estimated by MC simulation, the results from MC and those from IS are almost the same. Thus, we can conclude that such methods do not suffer from the overbiasing problem and can generate very accurate, unbiased estimates. Finally, to see the efficiency of the IS methods, Figure 7 shows the relative error, which is defined as the standard deviation divided by the mean of the estimation, versus the call blocking probabilities. In this graph, the relative error for the ordinary Monte Carlo simulation explodes when the blocking probability approaches 10^{-9}, while there are only small fluctuations for both ISSC

Table 2: Blocking Probabilities and Relative Errors

Route	P(B) MC	RE of MC	P(B) S-ISSC	RE of S-ISSC	P(B) of A-ISSC	RE of A-ISSC
7	$4.63 \cdot 10^{-8}$	548.79	$5.30 \cdot 10^{-8}$	1.949	$5.35 \cdot 10^{-8}$	1.912
8	$4.63 \cdot 10^{-8}$	548.18	$8.90 \cdot 10^{-8}$	1.672	$8.86 \cdot 10^{-8}$	1.766
9	$9.30 \cdot 10^{-8}$	378.68	$1.05 \cdot 10^{-7}$	2.670	$1.09 \cdot 10^{-8}$	1.038
10	0	N/A	$8.50 \cdot 10^{-8}$	1.366	$8.66 \cdot 10^{-8}$	1.432
11	$4.64 \cdot 10^{-8}$	547.59	$5.64 \cdot 10^{-8}$	2.276	$5.40 \cdot 10^{-8}$	1.037

methods. Furthermore, we can see that the relative error of our adaptive A-ISSC is smaller than that of S-ISSC. This shows the efficiency of the A-ISSC for such cases that are indeed non-light loaded.

5 CONCLUSION

In this paper, we have described two importance sampling methods to simulate the call blocking probability in a traffic groomed optical network that have simultaneous resource possession. In the light-load case, we have proved that static ISSC has a bounded relative error (BRE) in such networks, where different classes of traffic may require different amounts of sub-wavelength units. In cases other than that of light-load, we argued that *adaptive* ISSC may instead provide an efficient and robust way to set biasing parameters. In our simulation results, adaptive ISSC provides very accurate estimates of the call blocking probability, as well as extremely low relative variances.

As mentioned in the second section, these methods can be extended to be used in more generalized, "mesh" networks without much additional effort. Moreover, unlike the methods that depend on the product form solution, adaptive ISSC may be used in *single target system* models in which

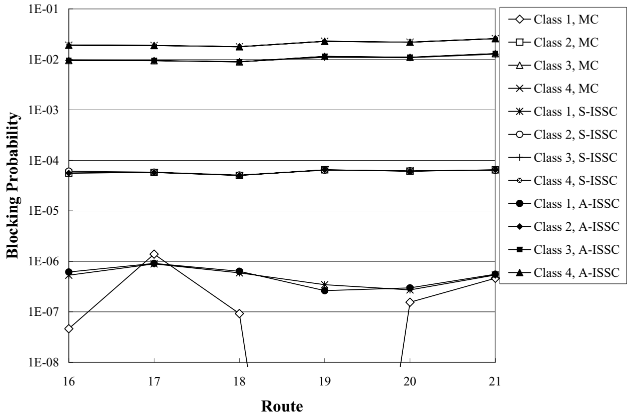

Figure 6: Routes that Use Four to Six links

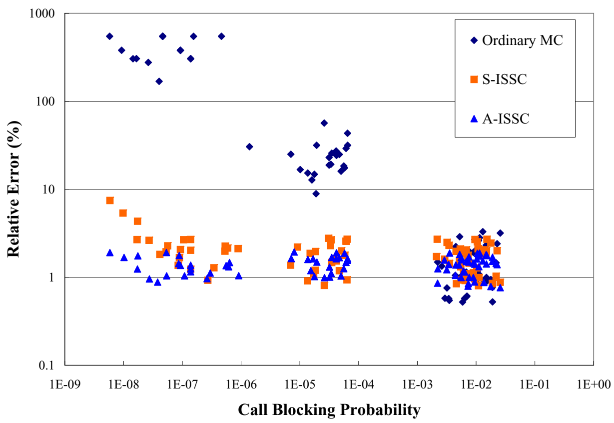

Figure 7: Relative Error vs. Call Blocking Probability

there exist optimal or good biasing methods. For example, for a network with self-similar input traffic, arrivals may be modeled as sum of fractal modulated poisson processes, which converges asymptotically to a Poisson arrival in fixed time intervals (Tsybakov and Georganas 1997). In a case like this, the asymptotically optimal bias of single target cases exists and may be extended to multi-target systems by using Adaptive ISSC.

Finally, in certain modern optical networks, such as those using Optical Burst Switching (Battestilli and Perros 2003), calls may not hold all of the links they require at the time they arrive. On the contrary, they may obtain the link just one hop in advance of the traffic burst. Whether the importance sampling provided in this paper can be applied to such network models will be the next step of our research.

REFERENCES

Andrew, L. L. H. 2004. Fast simulation of wavelength continuous WDM networks. *IEEE/ACM Transactions on Networking, to appear*.

Battestilli, T., and H. Perros. 2003. An introduction to optical burst switching. *IEEE Communications Optical Magazine* 41:S10 – S15.

Chang, C.-S., P. Heodelberger, and P. Shahabuddin. 1995. Fast simulation of packet loss rates in a shared buffer communication switch. *ACM Transactions on Modeling and Computer Simulation* 5 (4): 306–325.

Dutta, R., and G. Rouskas. 2002. Traffic grooming in wdm networks: Past and future. *IEEE Network* 16 (6): 46–56.

Falkner, M., M. Devetsikiotis, and I. Lambadaris. 1999. Fast simulation of networks of queues with effective and decoupling bandwidths. *ACM Transactions on Modeling and Computer Simulation* 9 (1): 45–58.

Heegaard, P. E. 1996. Adaptive optimisation of importance sampling for multi-dimensional state space models with irregular resource boundaries. In *Proceedings of the 13th Nordic Teletraffic Seminar, Trondheim, Norway*, 176–189.

Heidelberger, P. 1995. Fast Simulation of Rare Events in Queueing and Reliability Models. *ACM Transactions on Modeling and Computer Simulation* 5 (1): 43–85.

Lassila, P. E., and J. T. Virtamo. 1999. Efficient importance sampling for monte carlo simulation of loss systems. In *ITC-16 Conference on Teletraffic Engineering in a Competitive World*, 787–796. Elsevier, Amsterdam, The Netherlands.

Lassila, P. E., and J. T. Virtamo. 2000. Nearly optimal importance sampling for monte carlo simulation of loss systems. *ACM Transactions on Modeling and Computer Simulation* 10 (4).

Mandjes, M. 1997. Fast simulation of blocking probabilities in loss networks. *European Journal in Operational Research* 101:393–405.

Ross, K. W. 1995. *Multiservice loss models for broadband telecommunication networks*. Springer-Verlag.

Shahabuddin, P. 1994. Importance sampling for the simulation of highly reliable markovian systems. *Management Science* 40 (3): 333–352.

Smith, P. J., M. Shafi, and H. Gao. 1997. Quick simulation: A review of importance sampling techniques in communications systems. *IEEE Journal on Selected Areas in Communications* 15 (4): 597–613.

Tsybakov, B., and N. D. Georganas. 1997. On the self-similar traffic in atm queues: definitions, overflow probability bound and cell delay distributions. *IEEE/ACM Transactions on Networking* 5 (3): 397–409.

Vakili, P. 1991. Using a standard clock technique for efficient simulation. *Operations Research Letters* 10:445–452.

Vazquez-Abad, F. J., L. L. H. Andrew, and D. Everitt. 2002. Estimation of blocking probabilities in cellular networks with dynamic channel assignment. *ACM Transactions on Modeling and Computer Simulation* 12 (1).

Washington, A. N., and H. G. Perros. 2004. Call blocking probabilities in a traffic groomed tandem optical network. *Journal of Computer Networks, Special issue dedicated to the memory of Professor Olga Casals,* 45.

AUTHOR BIOGRAPHIES

CHIH-CHIEH HSU (IEEE S 2002) received the B.S. and M.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan in 1998 and 2000. He is working toward the Ph.D. degree in the Department of Electrical and Computer Engineering at North Carolina State University. His research interests include analysis and implementation of efficient simulation, optimum network control, optical networks, and wireless ad-hoc networks. His e-mail address is <chsu@ncsu.edu>.

MICHAEL DEVETSIKIOTIS (IEEE S 1985, M 1994, SM 2003) is an Associate Professor of the Department of Electrical and Computer Engineering at North Carolina State University. His research work has been in the areas of telecommunication systems modeling, performance evaluation, and efficient simulation; traffic characterization and management; and optimization techniques applied to the analysis and design of communication systems. His present focus is on the performance and nonlinear behavior of communication networks as they become larger in size,and more complex in topology and traffic. He currently serves as the chairman of the IEEE Communication Society committee on communication systems integration and modeling (CSIM), as an Associate Editor for the ACM Transactions on Modeling and Computer Simulation (TOMACS), and as an Associate Editor for the IEEE Communications Letters. Finally, he is co-chair of the High Speed Networks symposium, IEEE ICC 2004, held in Paris, June 2004. His e-mail address is <mdevets@ncsu.edu>