Efficient Estimation of Loss Rates in Optical Packet Switched Networks with Wavelength Conversion

Poul E. Heegaard
Norwegian University of Science and Technology
Trondheim, N-7491, Norway
poulh@item.ntnu.no

Werner Sandmann
University of Bamberg
Bamberg, D-96045, Germany
werner.sandmann@wiai.uni-bamberg.de

Abstract

Packet losses can significantly degrade the quality of service (QoS) provided by networks. In this paper, asynchronous bufferless optical packet switched networks with multiple service classes and full wavelength conversion are considered. In case of a large number of wavelengths per fibre, packet losses become rare and thus intractable by means of conventional performance methodologies. We present a novel fast simulation technique for efficiently estimating the loss rates in this setting. Our technique applies Importance Sampling which proceeds by using an alternative probability distribution for accelerated simulation and appropriately weighting the results to obtain unbiased estimators. With our technique the necessary parameters determining the alternative distribution are adaptively obtained by ant colony optimization, a metaheuristic that is in widespread use for solving combinatorial optimization problems. Thereby, no intimate a priori knowledge of the system under consideration is required to set up the simulation. Rather, the parameters adapt to the model. The accuracy and the efficiency of the novel technique are demonstrated by numerical results.

1. Introduction

The Internet is rapidly growing and more and more new services emerged such as multimedia and realtime applications, giving rise to the need for faster transfer techniques. Optical networks featuring advanced technologies such as wavelength division multiplexing (WDM) and optical packet switching (OPS) are promising solutions capable to meet the demands of the future generation Internet [13, 15, 16, 21]. Supporting quality of service (QoS) in such networks is of major importance where packet loss rates are of special interest. These are required to be very small which means that packet losses become rare events, thereby posing great challenges on performance evaluation methodology. Exact analytical solutions are usually not available at all for complex networks. Large deviations theory (LDT) [5, 20] provides a framework for asymptotic investigations when the probability of the rare event converges to zero. However, in practice we are not interested in asymptotics but in specific probabilities in the range of $10^{-9}$ to $10^{-12}$. Moreover, LDT results can be obtained only for rather small systems. Likewise, numerical methods are typically not suitable for rare events analysis in complex networks. Hence, simulation is needed.

Direct simulation of rare events is not effective, since rare events occur too infrequently in simulations to compute reliable statistical estimates in reasonable time. Accelerated simulation is necessary in the sense that simulation time to get estimates with desired statistical accuracy, for example expressed by the relative error or the confidence interval half width of the simulation estimator, must be reduced. Since the statistical accuracy depends on the variance of the simulation estimators, such an accelerated simulation means variance reduction, and it turns out that Importance Sampling is well suited for this purpose. Importance Sampling applies a change of measure, that is the original system is simulated under an alternative probability distribution (measure), and the systematically biased results are appropriately weighted to yield unbiased estimates, cf. [8, 10, 3, 12]. In [8] the framework for stochastic processes including generalized semi-Markov processes and Markov processes has been given which thus constitutes the particular formal basis for the types of models we shall consider in this paper. Unfortunately, it is by no means guaranteed that a specific change of measure always results in reduced variances but it may even yield an infinite variance increase. Hence, the crucial issue when applying Importance Sampling is a proper choice of the alternative probability distribution which has many times shown to be an extremely difficult task. Hence, to speak of an art when choosing the change of measure seems reasonable. However, to be useful for a broad range of applications, to be used by non-experts
in Importance Sampling, and to offer the potential of integration into performance analysis tools it is necessary to come up with "automated" methods that yield Importance Sampling simulations with significantly reduced variance compared to direct simulation.

In the present paper, we are concerned with model-based analysis of asynchronous bufferless optical packet switched networks with multiple service classes and full wavelength conversion. The number of wavelengths per fibre is assumed to be large and the packet loss rate is efficiently estimated using Importance Sampling with a novel method for adaptively learning a good change of measure based on the ant colony optimization metaheuristic [6, 7, 2], a multiagent system which is inspired by the foraging behavior of ants and has been successfully applied to various combinatorial optimization problems. In [19] a similar system is designed to solve problems in telecommunication networks.

The remainder of this paper is organized as follows. In Section 2 the considered optical packet switch architecture and an according model are described. The novel adaptive simulation technique is introduced in Section 3, starting with the foundations of Importance Sampling. Section 4 contains numerical results and finally, Section 5 concludes the paper and outlines further research directions.

2. Switch Architecture and Model

We consider a non-blocking asynchronous bufferless OPS with \( S \) input- and output fibres where each fibre provides \( W \) wavelengths of capacity \( C \) [bps] each. All available wavelengths at any output fibre are shared amongst \( S \) different service classes. If possible, contentions are handled by wavelength conversion such that packets are converted to idle wavelengths in case of multiple packets routed toward the same wavelength at the same time. If, due to no idle wavelengths, this is not possible then packet losses occur. We are interested in the packet loss rate at any single output fibre. We assume that packets of service class \( i \) arrive according to a Poisson process with possibly state dependent arrival rate \( \lambda_i(x) \) and that the packet length of service class \( i \) is exponentially distributed with mean packet length \( L_i \). Thus, the service rate is \( \mu_i = L_i/C \). A similar architecture was recently studied in [14] but only for two customer classes and state independent arrival rates. However, the heterogeneity of Internet services gives rise to the requirement of more than two service classes, and state dependent arrival rates can be adapted for modeling highly flexible system behavior and interconnections with the population environment. For model-based analysis we formally specify an according continuous-time Markov chain (CTMC), suitable for simulation, as follows. Since we consider \( S \) different service classes the state space is \( S \)-dimensional where each state \( x = (x_1, \ldots, x_S) \) is such that \( x_1 + \cdots + x_S \leq W \). The possible state transitions are for arrivals of class \( i \) from state \( x = (x_1, \ldots, x_S) \) to state \( (x_1, \ldots, x_{i-1}, x_i + 1, x_{i+1}, \ldots, x_S) \) with transition rate \( \lambda_i(x) \), and similarly for completions of services of class \( i \) from state \( x = (x_1, \ldots, x_S) \) to state \( (x_1, \ldots, x_{i-1}, x_i - 1, x_{i+1}, \ldots, x_S) \) with transition rate \( \mu_i(x) \). As packet losses occur when all wavelengths are occupied, we define the set \( R = \{ x \in \mathbb{N}^S : x_1 + \cdots + x_S = W \} \) as the set of target (rare) states, and we shall aim at estimating its probability. As an example for the special case of two service classes, the state transition diagram with marked target states is shown in Figure 1, where specifically the service rates for class \( i \) depend on \( x \) only in the number of entities of this class in the system. That is \( \mu_i(x) = x_i \mu_i \), where \( \mu_i \) is a constant that can be viewed as a "nominal" rate.

![Figure 1. State transition diagram for \( S = 2 \) service classes with load dependent rates where all states in the target set are marked](image)

3. Efficient Simulation Technique

In this section, we start with the foundations of Importance Sampling and the requirements for application to the OPS model, after which we proceed to the description of our novel fast simulation technique.

3.1. Importance Sampling

The most general description of Importance Sampling is in measure theoretic terms from which all applications to specific model types and domains can be obtained as special cases. Consider probability measures \( P, P^* \) on a measurable space \( (\Omega, \mathcal{A}) \) where \( P \) is absolutely continuous with respect to \( P^* \), i.e. \( \forall A \in \mathcal{A} : P^*(A) = 0 \Rightarrow P(A) = 0 \). Then
Importance Sampling, the probability measure.$c a t o r function. That means, in our case, using the indica-
an event is done by estimating the expectation of its indi-
ficult or impossible. As usual, estimating the probability of
and is therefore not available for implementation. Even if
the corresponding optimal Importance Sampling measure
zero variance but this cannot be applied in practice since
there always exists an optimal change of measure yielding
change of measure
probability measures
and jump probabilities
q

Note that with Importance Sampling all rates may be state
dependently even if they are state independent in the original
system.

3.2. Adaptive Change of Measure

Adaptive approaches aim at learning a good change of
measure without intimate knowledge of the model at hand.
Starting with some initial change of measure, a couple of
independent simulation runs are performed, and the change
of measure is updated according to rules that depend on and
thereby characterize the specific adaptive method. Then
the simulation is continued by making multiple independent
simulation runs with the updated measure and so on
until finally the method converges to a measure with which
the actual simulation is performed. Most of the adaptive ap-
proaches reported in the literature aim at either directly min-
imizing the (estimated) variance of the Importance Sam-
pling estimator or a related property such as for example the
cross-entropy between the actually used measure and the
optimal one [17]. These approaches typically become either
computationally expensive and/or exhaust storage when ap-
plied to complex models without making too restrictive as-
sumptions.

Our technique makes use of LDT-based reasoning on
how rare events occur. Although a purely analytical ap-
proach to rare event analysis is usually impossible for
complex networks, LDT gives some valuable insights and
guidelines. In particular, rare events typically occur via cer-
tain most likely paths and it is known that the change of
measure should mainly emphasize these paths. Indeed, sev-
eral successful applications of Importance Sampling have
been reported where LDT was utilized to obtain a suitable
change of measure. However, the drawback of this ap-
proach is that the change of measure is specifically designed
for one particular model, and seemingly obvious modificati-
ions or extensions usually fail even for only slightly dif-
ferent or larger models. While it is difficult to determine
the most likely paths analytically, they can be estimated.
For example, in [9] the transition rates are changed in accordance to the importance of a path with this transition as the first step where the target importance is determined by a "lookahead" approach inspired by the failure distance, a metric for the distance from any state to a target set of rare states that was introduced in [4]. In [4] all transition rates are changed after each simulated transition. As the major drawback the simulation speed-up strongly depends on an accurate computation of the failure distances which involves the computation of minimum cut sets, in general an NP-hard problem. Hence, the applicability of this approach is quite limited. In [9] the path likelihood is estimated by determining the most likely path from the current state to any state in the set of rare states of interest. However, the computational demand of the lookahead approach increases as the dimensionality of the state space increases.

### 3.3. Search and Update Procedure

In our new technique the target importance is instead determined by ant colony optimization. More specifically, a search and update procedure inspired by the foraging behavior of (real) ants is applied and further equipped with an explicit memory and evaluation facility for (artificial) ants. The ants search iteratively for paths in a connected graph (a network) between source nodes and destination nodes. The path quality is evaluated on arrival to a destination node and then each ant backtracks over the links along the reverse path quality is evaluated on arrival to a destination node and involves the computation of minimum cut sets, in general an NP-hard problem. Hence, the applicability of this approach is quite limited. In [9] the path likelihood is estimated by determining the most likely path from the current state to any state in the set of rare states of interest. However, the computational demand of the lookahead approach increases as the dimensionality of the state space increases.

In our new technique the target importance is instead determined by ant colony optimization. More specifically, a search and update procedure inspired by the foraging behavior of (real) ants is applied and further equipped with an explicit memory and evaluation facility for (artificial) ants. The ants search iteratively for paths in a connected graph (a network) between source nodes and destination nodes. The path quality is evaluated on arrival to a destination node and then each ant backtracks over the links along the reverse path quality is evaluated on arrival to a destination node and involves the computation of minimum cut sets, in general an NP-hard problem. Hence, the applicability of this approach is quite limited. In [9] the path likelihood is estimated by determining the most likely path from the current state to any state in the set of rare states of interest. However, the computational demand of the lookahead approach increases as the dimensionality of the state space increases.

Let $x = (x_1, \cdots, x_S)$ be the state vector and $\pi(x, y)$ a path between two states $x$ and $y$. A path between state $x$ and a state in the rare event set is denoted $\pi(x, \mathcal{R})$ and the $r$-th sampled path is $\pi^{(r)}$. The probability of a path $\pi$ is $p(\pi)$. The maximum normalized probability of a path from state $x$ to a state in $\mathcal{R}$, given that an arrival of service class $i$ occurred in state $x$ is

$$p_{\text{max}}(\pi(x, \mathcal{R}) \mid i\text{-arrival}) = \frac{\sum_{r=1}^{R} p(\pi^{(r)}(x, \mathcal{R}) \mid i\text{-arrival})}{\max_x \{\sum_{j=1}^{R} p(\pi^{(j)}(x, \mathcal{R}) \mid k\text{-arrival})\}}$$

(1

where for any path $\pi(x, \mathcal{R})$ from $x$ to the set of rare states, $p(\pi(x, \mathcal{R}) \mid i\text{-arrival})$ denotes the probability of that path given that an $i$-arrival transition occurred in state $x$. The normalization factor in (1) is the largest accumulated target probability of all $k$-arrival transitions in state $x$. The ant colony optimization then applies the following procedure that is repeated for every simulation iteration $r = 1, \cdots, R$:

**repeat**

- Sample a path $\pi^{(r)}$ from $x$ towards a target state in $\mathcal{R}$; Stop when hitting the origin states;
- if $\pi^{(r)}$ contains states of the rare event set $\mathcal{R}$ then
  - for each state $y \in \pi^{(r)}$ update the maximum path probability $p_{\text{max}}(\pi(y, \mathcal{R}) \mid i\text{-arrival})$ according to equation (1);

**until** <end of simulation condition>.

The random search of the ants in every state is governed by a *random proportional rule* that is incrementally updated for every new path found by the ants. This proportional rule is determined by the normalized, possibly state dependent Importance Sampling transition rates $\lambda_i^*(x)$ and $\mu_i^*(x)$ in each state. Even in case of state independent rates in the original model, that is $\lambda_i(x)$ and $\mu_i(x)$ as denoted above are constant for all states $x$, the Importance Sampling transition rates may be state dependent. In fact, only this flexibility of state dependent rates renders successful applications of Importance Sampling possible in complex settings. Each sampled path that includes visits to $\mathcal{R}$ invokes a recalculation of the maximum path likelihood for all states in the path. Then the transition rates are changed for all transitions in the path according to the following updating rule. Denote by $1_i$ the $S$-dimensional vector with entry 1 at component $i$ and zero elsewhere and define $\Delta_i(x) = \max(\mu_i(x+1_i)-\lambda_i(x); 0)$. Then the arrival and service rates for Importance Sampling are updated according to

$$\lambda_i^*(x) = \lambda_i(x) + p_{\text{max}}(\pi(x, \mathcal{R}) \mid i\text{-arrival}) \cdot \Delta_i(x),$$

$$\mu_i^*(x) = \mu_i(x) - p_{\text{max}}(\pi(x, \mathcal{R}) \mid i\text{-arrival}) \cdot \Delta_i(x-1_i).$$

Note that only transition rates along the sampled path are updated which strongly reduced the computational demands compared to other approaches. Initially, when no information of the target likelihood exists, the ants search the state space by a guided random walk. After this initial phase the ants use the updating rules.

### 4. Numerical Results

We have performed a series of simulation experiments for different numbers of service classes as well as different numbers of wavelengths. Here, we present numerical results for $S = 2, 3, 4, 5$ service classes with $W = 4, 8, 16$ wavelengths. The load level $\rho_i = \lambda_i/\mu_i$ of each service class $i = 1, \cdots, S$ is varied such that the loss probabilities are in the range from $10^{-7}$ to $10^{-11}$. In the simulations, without loss of generality the rates are normalized such that $\lambda_1 + \mu_1 = 1$ which gives $\lambda_i = \rho_i/(1 + \rho_i)$ and $\mu_i = 1/(1 + \rho_i)$. The specific system parameters used within the simulation experiments are listed in Table 1.
Note that these parameters are not chosen as flexible as they could be. It is possible to analyze more general settings with our simulation technique. However, in order to compare the results to exact results we need parameter settings for which exact values can be numerically computed. In all cases \( \rho_i = \rho \) are the same for all service classes, except Case VIII where the service class 5 has a higher load level than the other service classes. The objective is to estimate the loss probabilities, i.e. the probability that all wavelengths are in use. With the parameter settings given in Table 1 losses are rare events. The set of rare event states \( \mathcal{R} = \{ x \in \mathbb{N}^8 : x_1 + \cdots + x_8 = W \} \) for all simulated cases is also contained the table.

Also note that we dropped the explicit expression of state dependence in the rates for ease and readability of notation but we also considered state dependent rates. Of course, state dependence in general can be any type of dependence, any function on the state vector. However, we should not construct artificial types of dependence but choose dependencies that well reflect reality. Moreover, for the sake of comparisons to exact results we choose dependencies that fit the assumptions of available numerical methods and can thus be handled by them. State dependent rates here mean dependencies such that \( \lambda_i(x) = (W_i - x_i)\lambda_i \) and \( \mu_i(x) = x_i\mu_i \). Once given this specification of state dependence, it is indeed convenient to drop it in the notation and set \( \lambda_i = \lambda_i(x) \), \( \mu_i = \mu_i(x) \) and \( \rho_i = \rho_i(x) \). The system load in this case is then accordingly given by \( \rho_i = \lambda_i / (\lambda_i + \mu_i) \).

In the simulated cases with state independent rates the system load is given by \( \rho = (\lambda_i + \mu_i) / W \). In Table 2 the estimated loss probabilities based on 3,000,000 regenerative cycles along with the relative errors of the simulation estimates are given. The simulation results are compared with exact results obtained by a numerical method, the convolution algorithm described in [11]. Note again, that such exact results can be obtained for the cases we have chosen and that this choices were made to provide evidence on the validity and accuracy of our method but our method applies to far more general settings.

<table>
<thead>
<tr>
<th>Case</th>
<th>( \rho_i )</th>
<th>( W )</th>
<th>( S )</th>
<th>( \mathcal{R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.05</td>
<td>16</td>
<td>2</td>
<td>( \sum_{i=1}^{S=2} x_i = 16 )</td>
</tr>
<tr>
<td>II</td>
<td>0.0025</td>
<td>4</td>
<td>2</td>
<td>( \sum_{i=1}^{S=4} x_i = 4 )</td>
</tr>
<tr>
<td>III</td>
<td>0.01 (state dep. rates)</td>
<td>8</td>
<td>2</td>
<td>( \sum_{i=1}^{S=8} x_i = 8 )</td>
</tr>
<tr>
<td>IV</td>
<td>0.10 (state dep. rates)</td>
<td>16</td>
<td>2</td>
<td>( \sum_{i=1}^{S=16} x_i = 16 )</td>
</tr>
<tr>
<td>V</td>
<td>0.0375</td>
<td>16</td>
<td>3</td>
<td>( \sum_{i=1}^{S=3} x_i = 16 )</td>
</tr>
<tr>
<td>VI</td>
<td>0.0025</td>
<td>4</td>
<td>3</td>
<td>( \sum_{i=1}^{S=4} x_i = 4 )</td>
</tr>
<tr>
<td>VII</td>
<td>0.025</td>
<td>16</td>
<td>4</td>
<td>( \sum_{i=1}^{S=16} x_i = 16 )</td>
</tr>
<tr>
<td>VIII</td>
<td>0.0025</td>
<td>4</td>
<td>4</td>
<td>( \sum_{i=1}^{S=4} x_i = 4 )</td>
</tr>
<tr>
<td>IX</td>
<td>0.01 (state dep. rates)</td>
<td>8</td>
<td>3</td>
<td>( \sum_{i=1}^{S=8} x_i = 8 )</td>
</tr>
<tr>
<td>X</td>
<td>0.0025</td>
<td>4</td>
<td>5</td>
<td>( \sum_{i=1}^{S=5} x_i = 4 )</td>
</tr>
<tr>
<td>XI</td>
<td>One 0.01, rest 0.001</td>
<td>4</td>
<td>5</td>
<td>( \sum_{i=1}^{S=5} x_i = 4 )</td>
</tr>
</tbody>
</table>

The results show the good performance of our novel simulation technique. All estimates are statistically accurate with small relative errors. Simulation runtimes were all in the range from only a few seconds to a few minutes. Finally, we like to emphasize that it was not possible to compare our results to results from simulation experiments without Importance Sampling, simply because in direct simulations without Importance Sampling no packet losses were observed within reasonable time. This can be easily explained by a rough argumentation as follows. If we consider an event that has a probability of \( 10^{-11} \) to occur within a regenerative cycle then on average \( 10^{11} \) cycles must be simulated to observe only one single occurrence of this event. It becomes even worse if one wants to make statistically valid estimations. Assuming as a rule of thumb that at least one hundred observations of an event of interest are necessary to provide reasonable statistics then on average \( 10^{13} \) regenerative cycles can be expected necessary to yield feasible results. Thus, since we only needed to simulate 3,000,000 cycles, we have an improvement factor which is in the order of \( 10^7 \). Now, assume one minute as a reasonable representative average runtime for \( 10^6 \) cycles. Then to obtain results that are of the accuracy of ours, direct simulation requires approximately 7,000 days which is more than nineteen years. Of course, the above argumentation is only rough but rather tends to underestimate the effort required by direct simulation than to overrate it (since only one hundred necessary observations of the rare event is quite optimistic in order to get proper statistical accuracy). Hence, the simulation speed-up provided by Importance Sampling compared to direct simulation is indeed dramatic.
5. Conclusion

In this paper, we have simulated the packet loss in an asynchronous bufferless optical packet switched network with multiple service classes and full wavelength conversion. The packet loss should occur with a low probability, thus making it a rare event which poses great challenges to performance analysis methodologies and renders direct simulation infeasible. We have presented a novel fast simulation technique for efficiently estimating the loss rates in this setting. Our technique applies Importance Sampling which proceeds by using an alternative probability distribution for accelerated simulation and appropriately weighting the results to obtain unbiased estimators. With our technique the necessary parameters determining the alternative distribution are adaptively obtained by the ant colony optimization metaheuristic. An exceptionally strong feature of our method is that no intimate a priori knowledge of the system under consideration is required to set up the simulation. Furthermore, the parameters gradually adapt to the model as the model's state space is explored and the Importance Sampling change of measure is set accordingly. This change of measure is applied by the succeeding simulation trials. The accuracy and the efficiency of the novel technique are demonstrated by numerical results for an optical packet switched network model. These results are very accurate with small relative errors. To demonstrate how the simulation technique works for different model structures, more simulation studies are planned, for instance for optical packet switched nodes with preemptive priority levels, or with an upper bound of the number of wavelengths allowed to be allocated by a low priority service class. It should be noted that the applicability of the method is not restricted to exponentially distributed times. It is easy to incorporate phase-type distributions into the model and then to apply the ant colony optimization procedure similarly as it is done in the present paper. This will be another topic of future investigations. Further research also includes systematic studies of the properties of the Importance Sampling estimators obtained via the ant colony optimization. Although a main motivation is the applicability to rare events with probabilities in orders of magnitudes of practical interest, it is of course also interesting and useful to examine asymptotic properties such as, e.g., those recently considered in [18, 1] to gain insights how the method behaves as the probability of interest converges to zero.

References