

# Network performance optimization using Importance Sampling theory

Benjamin Feng, Changcheng Huang, Carleton University, Canada  
Mike Devetsikiotis, North Carolina State University, U.S.A  
Yanick Champoux, Alcatel, Canada

## Abstract

**In simulation runs which involve rare events, Importance Sampling (IS) is often used to speed up the simulation process to get the simulation result faster. This paper proposes a new application for the Importance Sampling theory: to predict the change in the network behavior when the network input changes. This paper will show, by modifying the way the Importance Sampling theory applied in rare event simulation, it is possible to calculate precisely the amount of adjustment required in the network input for the network performance to meet a predefined target.**

## 1. Introduction

Traffic engineering is critical for data networks such as today's Internet, or future MPLS networks. It has been widely studied by the research community [1]. Due to the complexity of Internet traffic and network topology, analytical approaches quickly become intractable. Simulation has been considered as one of the major vehicles for traffic engineering. Unfortunately simulation may introduce significant costs and require unduly long time to obtain meaningful results. This paper investigates an Importance Sampling based traffic engineering approach which is general enough to address most of the traffic engineering issues and fast enough to obtain useful results in a reasonable time.

### 1.1. Motivation

The goal of traffic engineering is to optimize a network's performance to meet predefined requirements. Thus, the capability of accurately predicting the changes in a network's behavior, when the input of that network is changed, is essential in any traffic engineering system. There are some analytical based approaches, such as queueing theory [2], effective bandwidth[3], which can be used to determine the network's performance given the network's input. However, those approaches are either too conservative, incapable of end-to-end network performance analysis, or do not work well in large complex networks (especially under self-similar traffic model[4][5]). In most cases, heuristic simulation based approaches, such as trial & error[6], golden section search[7], steepest descent[8], which can also be used to optimize the performance of a

network to meet predefined requirement(s), are adopted. These heuristic approaches accomplish this by making a series of adjustments based on the feedback from network measurements. Even though the heuristic approaches are simple, accurate, and applicable to most networks, it may take a very long time for these heuristic approaches to optimize the network performance to meet the preset requirement(s), which are typically measured in rare events.

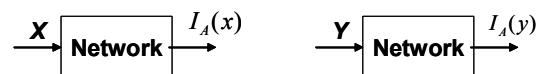
### 1.2. Contribution

This paper introduces a brand new performance optimization technique based on Importance Sampling theory. This new optimization approach is applicable to all networks as long as: 1) the network behavior, such as the probability of buffer overflow, the probability of packet loss, etc, can be described by stochastic events. 2) The network input, such as the network's input traffic flow, can be modeled as a stochastic process. 3) There is the capability to change the network input in terms of changing the attributes, such as the mean and the variance, of the input process. It is well known that most communication networks can meet these three conditions.

This Importance Sampling based approach is faster, simpler, and more universally applicable than the existing traffic engineering approaches. It can also carry out end-to-end network performance analysis under self-similar traffic model where many of the existing analytical approaches become intractable.

## 2. Importance Sampling Classical Application

We will first start with classical application of Importance Sampling theory.



**Figure 1 Network Behavior Under Different Input Processes**

In Figure 1,  $X$  and  $Y$  are two random processes modeling two different input traffic flows to the same network. Assume  $X$  and  $Y$  have the same class of distribution (in this paper, we assume the distribution is self-similar Fractal Gaussian Noise [5][4]). The only thing different between the two

process are their mean values: the mean of  $X$  is  $\nu$ , and the mean of  $Y$  is  $\nu_o$ . Twisted amount ( $m^*$ ) is the difference between  $\nu$  and  $\nu_o$ .  $I_A(t)$  is an event occurrence indicator function. In a specific time slot or a specific simulation replication run, when  $I_A=1$ , it implies the event occurred and  $I_A=0$  implies event did not occur. In this paper, the event of interest is the buffer overflow event.

Assume there are only a few overflow occurrences (rare events) in the network under  $X$ . As a result, a long period of time or many simulation replication runs are needed to accurately measure the probability of overflow. Assume  $\nu_o > \nu$ , then, there will be more overflow in the network if traffic process  $Y$  is applied. Thus, the amount of time required to measure overflow probability becomes smaller.

The Importance Sampling theorem states [9]:

$$E_X \{I_A(x)\} = \int I_A(t) \frac{f_X(t)}{f_Y(t)} f_Y(t) dt \quad (1)$$

where  $N_Y$  is the number of simulation replication runs executed;  $f_Y(y)$  is the pdf of the input process  $Y$ ;  $f_X(x)$  is the pdf of the input process  $X$ ;  $I_A$  is the indicator of event occurrences when  $Y$  is applied to the network. With the captured trace of event  $A$  (which will take a shorter time to collect) and the likelihood ratio  $L$  (a ratio of the two distributions:  $\frac{f_X(t)}{f_Y(t)}$ ), the expected probability of

event  $A$  occurrences in the network, with  $X$  as the network input, can be found even when we have applied  $Y$  as the network input.

$$\tilde{P}_X' = \frac{\sum_{i=1}^{N_Y} I_{A_i} L_i}{N_Y} \quad (2)$$

is used to approximate (1). Thus, by applying Importance Sampling, we have reduced the amount of simulation time required to get the simulation result.

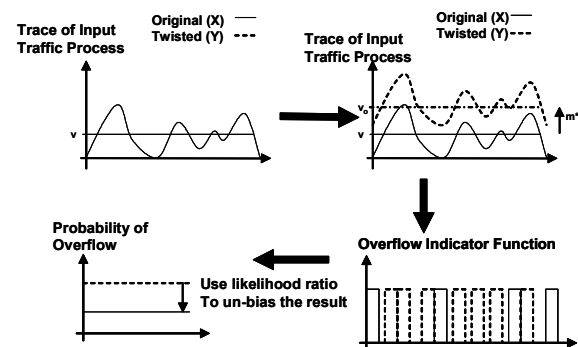


Figure 2 Importance Sampling in Rare Event Simulation Summary

Figure 2 shows a summary of the application of Importance Sampling in speeding up rare event simulation. We start with a known “original” input process  $X$  with a fixed mean value of  $\nu$ . We try to find an input process  $Y$  (twisted process) with mean of  $\nu_o$  that will make overflow events occur more often. Importance Sampling based optimization is a “reverse” of this concept.

### 3. Importance Sampling in Network Performance Optimization

Previously,  $\nu$  is known and fixed.  $\nu_o$  is changed. Now,  $\nu_o$  is known and fixed,  $\nu$  is changed. We treat the current input to the network as the twisted process. We try to find the original input process  $X$  such that when  $X$  is applied into the network, the probability of overflow in the network meets a predefined requirement.

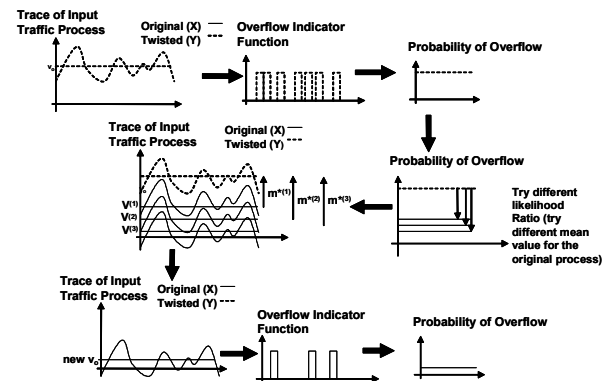


Figure 3 Importance Sampling based Traffic Engineering Process

Figure 3 is a graphical view of the IS approach. In order to calculate the likelihood ratio, the IS based approach starts with the capturing of a trace of the current network input traffic flow and a trace of buffer overflow occurrences currently in the network. The number of replication runs used in the simulation run ( $N_Y$ ) is also recorded. Equation (2) indicates that

$$\tilde{P}_X' = \frac{\sum_{i=1}^{N_Y} I_{A_i} L_i}{N_Y}$$

The “twisted”/“shifted” input process now models the traffic flow currently being applied to the network. Therefore,  $I_{A_i}$  is the overflow occurrence trace that was captured. As a result, the term  $I_{A_i}$  in equation (2) is fixed. The  $N_Y$  term in (2) is also captured and therefore it is fixed. The only term remaining in (2) that is not fixed is the likelihood ratio  $L_i$ . Since the likelihood ratio is a ratio of the two input process’ distributions, and the two

distributions are only different in their mean values (by the twisted amount  $m^*$ ):

The probability of buffer overflow, when the “original” input traffic process is applied to the network, can be made into anything desired by setting the twisted amount  $m^*$  to the appropriate value. No more simulation is required since everything needed to solve equation (2) have been collected or has been set.

Since  $v_o = v + m^*$ , when using different values of twisted amount ( $m^*$ ), the mean rate ( $v$ ) of the “original” traffic process will take on different

values. As a result,  $\tilde{P}_X'$  in equation (2) represents the probability of buffer overflow when the network is under different “original” network traffic processes ( $X$ ), each with different mean values ( $v$ ). By reducing the current input traffic ( $Y$ )’s mean rate to match the mean rate of any one of those “original” input processes, the current input process becomes that specific “original” input process since the two processes only differ by their mean values. Therefore, (2) can predict the network behavior, in terms of the probability of buffer overflow, when the mean rate of the network input traffic changes. This capability makes the performance optimization extremely easy: By collecting the overflow occurrences and input traffic trace, we can calculate how the probability of overflow will change as we change the mean rate of the incoming traffic.

#### 4. Simulation Results

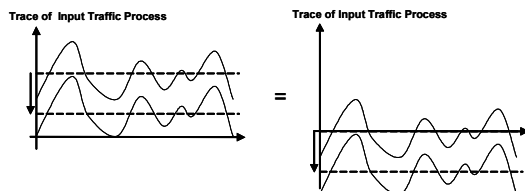


Figure 4 Input Traffic Process Shift/Twist

The random input process generated by an existing FGN traffic generator has zero mean and unity variance. Since we want to predict how the network will behave if the mean value of the current network input is reduced, as a result, the mean values of the FGN processes, after the adjustments are made, are not positive, as shown in Figure 4. Because the Lindley equation  $Q_k = (Q_{k-1} + X_k - \mu)^+$  [4] is used to model the queue behavior, the queue size depends on the difference between the instantaneous incoming traffic rate ( $X_k$ ) and the service rate ( $\mu$ ); therefore, a negative mean valued input process is valid.

Tandem Queue topologies are used in the simulation. Both buffers has the size of 50 units. Service rate for Buffer 1 is 0.8 units/sec and service rate for buffer 2 is 0.62 units/sec. The traffic source has a Hurst Parameter [5] of 0.8.

#### 4.1.1. Probability of Overflow

The new IS approach is used to predict the end-to-end probability of overflow in the network. Monte Carlo [10] Simulation is used to verify the prediction. As it can be seen in Figure 5, the prediction is fairly accurate. Most existing analytical approaches can not handle this kind of end to end performance analysis, especially under self-similar traffic model.

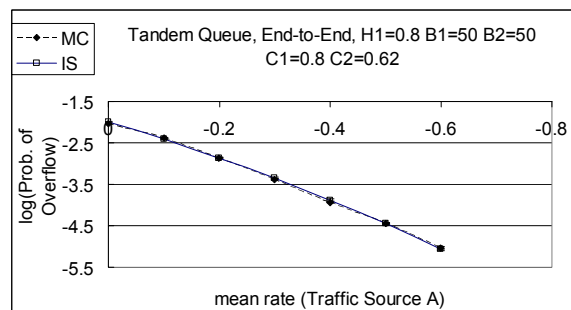


Figure 5 End-to-end Probability of Overflow

We also predict the probability of overflow in Q1 and Q2, individually. As it can be seen Figure 6, even though Q1 is congested (with overflow), the prediction on the second queue is still very accurate, as it can be seen in Figure 6.

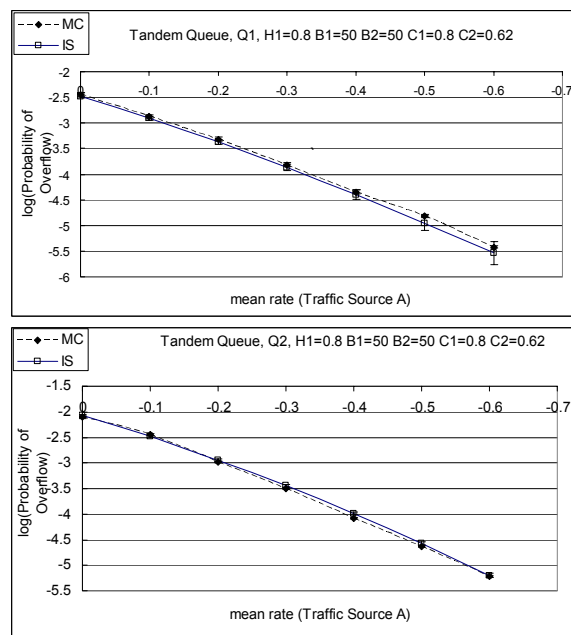
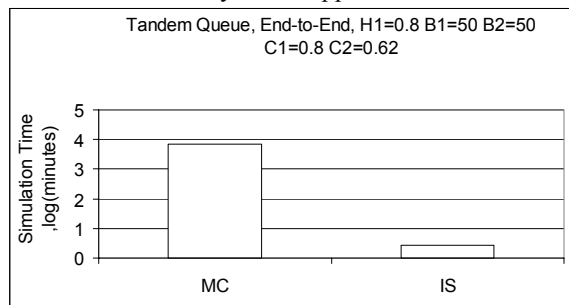


Figure 6 Tandem Queue, individual Probability of Overflow

#### 4.1.2. Time Consumption

As the network event of interest becomes more and more rare during optimization, it will take more and more time for heuristic based optimization approaches to measure it. The beauty of using Importance Sampling is that when we first collecting

the samples, the network event could still be very frequent. After the necessary traces are collected, the rest are pure off-line calculation. The time required is extremely short. Assume other optimization processes are able to achieve their target in one try (in one Monte Carlo run), and we compare this time with the time used by the IS approach.



**Figure 7 Simulation Time Comparison**

Figure 7 shows the simulation times required for traditional optimization approach (Monte Carlo) versus the IS approach. The simulation time is recorded in units of minutes. Because of the huge difference in the simulation time, the simulation times are plotted in the logarithm scale.

## 5. Conclusions

Using Importance Sampling theory in traffic engineering is a simple and effective way of predicting how the behavior of the network will change if network inputs are changed. The simulation results in this paper verified that this new approach is accurate in predicting the probability of overflow in a network with self-similar input traffic under different input traffic mean rates.

For complex networks, the Importance Sampling based performance optimization technique is more applicable because it does not require the knowledge of the internal details or internal actives (whether if there are overflow in upstream nodes) of a network. The Importance Sampling based approach can also perform end-to-end performance predictions in which some of the analytical approach can not perform under self-similar traffic model.

## Reference

- [1] D. O. Awduche, A. Chiu, A. Elwalid, I. Widjaja, and X. Xiao, "Overview and Principles of Internet Traffic Engineering", IETF Internet Draft, 2002.
- [2] J. Banks, J. S. Carson, II, B. L. Nelson, David M. Nicol, "Discrete Event System Simulation, Third Edition", Pages 204-445, Prentice Hall, 2001
- [3] F. Haciomeroglu, "On-line Measurement-based Capacity Allocation Schemes", A dissertation submitted to the Graduate Faculty of North Carolina State University, 2003

[4] C. Huang, "Long Range Dependent Traffic: Modeling, Simulation, and Congestion Control", Doctor of Philosophy thesis, Carleton University, Canada, 1997

[5] V. Paxson and S. Floyd. "Wide Area Traffic: The failure of Poisson Modeling IEEE/ACM Transactions on Networking, 3(3):226-244, 1995

[6] <http://www.m-w.com/cgi-bin/dictionary?book=Dictionary&va=trial+error>, Merriam-Webster, Incorporated, 2004

[7] W. L. Winston, M. Venkataramanan, "Introduction to Mathematical Programming", 4<sup>th</sup> edition, page 692-697, 744-745, Thomson Brooks/Cole, 2003

[8] W. L. Winston, M. Venkataramanan, "Introduction to Mathematical Programming", 4<sup>th</sup> edition, page 703-706, Thomson Brooks/Cole, 2003

[9] R. Y. Rubinstein, B. Melamed "Modern Simulation and Modeling" page 100-101, John Wiley & Sons Inc, 1998

[10] <http://mathworld.wolfram.com/MonteCarloMethod.html>, Wolfram Research, 2004