

Stochastic Hybrid Simulation with Applications to Queuing Networks

invited talk

Ben Lauwens
Royal Military Academy
30 Renaissanceaan
Brussels, Belgium
ben.lauwens@rma.ac.be

Bart Scheers
Royal Military Academy
30 Renaissanceaan
Brussels, Belgium
ben.lauwens@rma.ac.be

ABSTRACT

This paper deals with an extension to the hybrid simulation paradigm, i.e. the combination of event-driven simulation and analytical modelling, applied to packet telecommunication networks. In order to speed up the simulation only a small part of all packets, the foreground traffic, is processed in an event-driven way. On each arrival of a foreground packet, the waiting time of the packet is sampled from the virtual waiting time distribution function of the combined foreground and background traffic. This distribution function is stochastically modelled by the exact large deviations asymptotic of the virtual waiting time in a many sources regime. This novel methodology is not only valid for wired point-to-point queueing networks having a fixed transmission capacity, but it can also be applied to queueing networks for which the transmission capacity varies with the traffic load of all the elements in the network. The results obtained by the stochastic hybrid simulator are compared to full-blown event-driven simulations. An important reduction in simulation run-time is gained without sacrificing accuracy.

Categories and Subject Descriptors

C.4 [Performance of Systems]: Modelling techniques; G.3 [Probability and Statistics]: Queueing theory; I.6 [Simulation and Modelling]: Simulation Theory

General Terms

performance

Keywords

Hybrid simulation, Large deviations, Queueing network, Stochastic modelling

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIMUTools 2010 March 15–19, Torremolinos, Malaga, Spain.
Copyright 2010 ICST, ISBN 78-963-9799-87-5.

1. INTRODUCTION

Evaluating the performance characteristics of telecommunication networks is a challenging task. Measurements of relevant data on an actual running system are often impossible. Therefore, analytical models are proposed in literature. However, even after simplifying assumptions and decompositions, the resulting analytical model is usually mathematically intractable. The only alternative for predicting the performance of complex telecommunication networks is a simulation.

A packet-switched communication device can be modelled by a queueing model. In queueing theory, a queue consists of a buffer, i.e. a waiting area, and a service zone. The queue itself corresponds to a switching node or a router having both a packet buffer and an associated outgoing interface with a limited transmission capacity. A telecommunication network can be represented by interconnecting the queues. An event simulator, generally speaking, mimics the behaviour of the queueing network by tracing the packets using events.

In a simulation experiment, samples of many thousands of observations are often required to estimate some performance measure with sufficient confidence. In view of computer run-time, it is clearly important that one should attempt to reduce the volume of sampling in order to reduce the run-time of the simulation. Unfortunately, it has been a common practice to apply simulation techniques uncritically with little thought to the efficiency of simulation runs.

Two general ways of realizing a faster simulation are variance reduction and hybrid simulation. The former uses some knowledge about the system to speed up the convergence of the simulation result by reducing the number of events to be simulated. The latter combines analytical solutions of the simulation experiment with event-driven simulations having a reduced number of events.

2. HYBRID MODELS

In [4] an overview is presented of hybrid simulation/analytic models and modelling. Analytic models and simulation models can be considered as two end-points of a spectrum of possible mathematical models that can be used in modelling. An analytical model is a set of equations that can characterize a system or a problem entity. Its solution procedure usually uses either an analytical equation or a numerical algorithm that has been developed for the set of equations to obtain the desired results. A simulation model

is a dynamic or an operating model of a system that mimics the operating behaviour of the system or problem entity and contains its functional relationships. Its solution procedure consists of running a computerized model of the conceptual model, the simulator, collecting data on its behaviour, and analysing the data to obtain the desired results. It is desirable to combine analytic models and simulation models into a hybrid model and to use this combined model if it is cost efficient.

The application of hybrid models to simulate telecommunication queueing networks is a rather recent evolution. Starting around 2000, several papers were published giving different implementations of the hybrid simulation paradigm for the simulation of queueing networks representing a point-to-point packet switching network. Two general ways are considered to split the model in an analytical part and a simulation part.

The first splits physically the network in sub-networks described by an analytical model, a fluid-flow approximation of the traffic streams, and in sub-networks implemented by a simulation model, a detailed packet-event simulator. A major problem has to be solved: how can packets realistically be generated from a fluid flow?

The second splits the traffic in two parts: the foreground packets, directly related to the performance measures of interest, and the background stream, having only an indirect relation with the performance measures of interest. The background traffic stream is in this case also a deterministic fluid flow and the foreground traffic process is modelled by an event-driven simulator. Two main problems have to be tackled for this model: how is the synchronization done between the background fluid flow and the foreground packet stream and how is the time-dependent behaviour of real traffic traces mapped to changes in the variables of the analytical equations? The former problem is solved in several papers but the latter seems to be of no concern to the hybrid modelling community.

3. NOVEL METHODOLOGY

The novel stochastic hybrid methodology is a model in which the foreground packets are simulated in an event-driven way. In this simulation scheme, first proposed in [1], the splitting of the traffic into foreground and background traffic is combined with the general idea of using a large deviations technique to get insight in the rare events of interest. The background fluid-flows are no longer described by a set of differential equations but a stochastic model is used.

At the queue level the impact of the stochastic fluid-flows on the foreground packets is estimated by exploiting the ASTA¹ property ([3]). If the lack of bias assumption is respected, the waiting time of a packet from a specific source

has the same distribution as the virtual waiting time of the aggregation of all traffic streams at the queueing system:

$$F_W(w) - F_V(w) = \frac{\text{Cov} \{ \mathbb{I}_{\{V \leq w\}}, \lambda \mid \mathbb{I}_{\{V \leq w\}} \}}{\mathbb{E} \{ \lambda \}} \quad (1)$$

where λ is the arrival rate of foreground packets, $F_V(v)$ the probability distribution of the virtual waiting time $F_W(w)$ the probability distribution of the waiting time and \mathbb{I} the indicator function.

The packets of the foreground traffic are processed using an adapted discrete-event simulation kernel. Each arrival event is delayed due to the buffer occupation by the background traffic. The delay estimate is sampled from the virtual waiting time probability distribution function of the combined foreground and background traffic at the queueing system.

The probability distribution function is modelled by a large deviations approximation for which the number of flows scales together with the transmission capacity and the buffer size ([2]). In the many sources large deviations limit a fluid-flow is characterized by its effective bandwidth which does not depend on the specific queueing system or on the other fluid-flows. The notion of effective bandwidth allows a straightforward generalization to large queueing networks.

In order to validate the novel stochastic hybrid simulation scheme, the results of the hybrid stochastic simulation runs are compared with the results of full-blown event-driven simulations for a large set of scenarios. The Kolmogorov-Smirnov test shows that the waiting time distributions of the foreground packets for the pure event-driven simulation do not differ from the waiting time distribution for the stochastic hybrid simulation scheme. The run time of the stochastic hybrid simulation is significantly less than the run time of the event-driven simulation depending on the proportion of the events generated by the foreground traffic and the events generated by the background traffic.

4. REFERENCES

- [1] B. Lauwens, J. Potemans, B. Scheers, and A. Van de Capelle. Hybrid simulation of a FIFO queueing system with trace-driven background traffic. In *ValueTools*, pages 1–10, 2007.
- [2] N. Likhanov and R. R. Mazumdar. Cell loss asymptotics for buffers fed with a large number of independent stationary sources. *Journal of Applied Probability*, 36(1):86–96, 1999.
- [3] B. Melamed and W. Whitt. On arrivals that see time averages. *Operations Research*, 38(1):156–172, 1990.
- [4] R. Sargent. A historical view of hybrid simulation/analytic models. In *Winter Simulation Conference*, pages 383–386, 1994.

¹The ASTA (Arrivals See Time Averages) property is closely related to the well known PASTA property: Poisson Arrivals See Time Averages. Due to the memoryless property of the Poisson process the distribution of the number of packets in the buffer is always equivalent to the long-run time average distribution.