Информационные процессы, Том 2, № 2, 2002, стр. 198–200.

KALASHNIKOV MEMORIAL SEMINAR

Efficient Simulation of Buffer Overflow in Queueing Networks

S.K. Juneja* and V.F. Nicola**

**Indian Institute of Technology,*
Hauz Khas, New Delhi - 110016, India
email: sandeepj@cse.iitd.ernet.in
*** University of Twente,*
P.O. Box 217, 7500 AE Enschede, The Netherlands
email: nicola@cs.utwente.nl

Received October 14, 2002

Consider a continuous-time queueing network with probabilistic routing, including feedback. External batch arrival and batch service processes are characterized by independent Markov additive processes (MAPs). An effective change of measure is proposed for use in importance sampling procedure to estimate the probability of buffer overflow in an arbitrary ('target') queue during its busy period. The change of measure is state-independent and is obtained by exponentially twisting the original arrival and service processes as well as the routing probabilities. Assuming that, under the change of measure, all queues (except the target queue) are either stable or 'critical,' we determine the twisting parameters by solving a non-linear optimization problem subject to appropriate inequality constraints. Under some assumptions, it can be shown that a large amount of variance reduction is guaranteed. Experimental results using the proposed change of measure indicate that the relative error of the obtained estimates grows sub-linearly with the overflow level.

1. BACKGROUND

A significant literature exists on the efficient estimation (via simulation) of extremely small probabilities of rare events, such as buffer overflows in special classes of queueing networks. Parekh and Walrand (1989) were the first to present a heuristic for efficiently simulating networks with single server at each queue and i.i.d. inter-arrival and service time distributions using the importance sampling technique. They considered the problem of estimating the probability that starting from an empty network, the total network population exceeds a given threshold before the network re-empties. (We refer to their proposed change of measure as the $P\&W$ change of measure.) For Jackson networks, Frater et al. (1991) obtain an explicit form for the $P\&W$ change of measure. However, for the simple two-node tandem Jackson network, Glasserman and Kou (1995) show that this change of measure does not give good simulation results for a certain region of the traffic parameters. In a two-node Jackson network, Randhawa and Juneja (2001) show that if feedback is allowed, then for certain traffic parameters the suggested change of measure leads to an estimator with infinite variance. Chang et al. (1994) consider an intree network in discrete-time and propose a change of measure that gives a large amount of variance reduction compared to naive simulation. It is noteworthy that under their proposed change of measure the feeder queues are either stable or 'critical.' Kroese and Nicola (2002) consider a two-node tandem Jackson network. They derive an asymptotically optimal state-dependent change of measure (depending on the queue length at the first buffer) to estimate the probability that the second buffer overflows during its busy period. The proposed change of measure is asymptotically state-independent (i.e., for large first buffer contents). Again, under their proposed change of measure, the first queue is either stable or 'critical.' Generally, under the optimal change of measure, the likelihood ratio on paths to the rare set is 'nearly constant.' Juneja (2001) notes that a necessary condition for this 'sample

path independence' (SPI property) is that the likelihood ratio along paths joining any two states (on a path to the rare set) also be sample path independent. This is equivalent to the condition that the likelihood ratio across 'cycles' (state repetitions) equals 1.

2. THE MODEL

In this paper, we first consider a queuing network in discrete-time. The number of external arrivals to any queue during subsequent discrete-time intervals is a sequence of i.i.d. random variables independent of other arrival and service processes. The size of the batch of services completed during these intervals is again i.i.d. and independent of other processes. Probabilistic routing that includes feedback is allowed. As in Chang et al. (1994), we focus on the efficient estimation of the probability that a target queue overflows during its busy period. To this end, we identify a change of measure under which the feeder queues are either stable or critical and that satisfies the SPI property along 'most likely paths' to the rare set. We then generalize some of the results to networks in continuous-time with i.i.d. batch arrival and service sizes and i.i.d. batch inter-arrival times and batch service times. For the case where the service rates at the feeder queues are 'sufficiently large' we develop conditions for existence of the proposed change of measure. The distribution of queue lengths at the feeder queues at the initiation of the target queue busy periods plays an important role in our analysis. We focus primarily on two types of initial condition:

1. When all the feeder queues are assumed to be bounded at the start of the target queue busy period. We refer to this as the type-1 initial condition.
2. When all the feeder queues have a distribution corresponding to steady-state at the instants of busy period initiations. We refer to this as the type-2 initial condition. This, for example, is the case when estimating the steady-state loss probability, i.e., the fraction of customers lost due to buffer overflow at the target queue in steady-state (see Chang et al. 1994).

3. CONTRIBUTIONS

Our main contributions include:

1. We develop a change of measure that gives nearly constant likelihood ratios along the paths to the rare event. We show that the proposed change of measure is asymptotically optimal for type-1 initial conditions. A useful by-product of our analysis is the derivation of large deviations logarithmic limits for the probability of interest.
2. We generalize the above model to allow general i.i.d. external inter-arrival times and general i.i.d. service times at each queue and identify the change of measure that gives nearly constant likelihood ratios along paths to the rare event. We do not prove its asymptotic optimality, although we believe it to be true (for type-1 initial conditions) under appropriate regularity conditions.
3. We develop a simple heuristic non-linear program to help determine the proposed change of measure.
4. We note that the proposed change of measure may also be asymptotically optimal under type-2 initial condition, but only under certain restrictive conditions.
5. We analyze the general model when the feeder service rates are sufficiently large and show that only a unique change of measure may have nearly constant likelihood ratios amongst the class of state-independent changes of measure. We identify conditions for existence of such a change of measure. In the Jackson network settings we show that when the target queue is the original bottleneck in the network, this change of measure always exists and is identical to that of $P\&W$. Recall that $P\&W$ estimate the probability that the total network population exceeds a common large buffer in a regenerative cycle (with system empty as the regenerative state). This result may be anticipated since the total population growth is largely due to the growth in the bottleneck queue.

REFERENCES

1. Chang, C.S., Heidelberger, P., Juneja, S. and Shahabuddin, P. (1994) Effective Bandwidth and Fast Simulation of ATM Intree Networks, *Performance Evaluation,* **20**, pp. 45–65.
2. Frater, M.R., Lennon, T.M. and Anderson, B.D.O. (1991) Optimally Efficient Estimation of the Statistics of Rare Events in Queueing Networks, *IEEE Transactions on Automatic Control,* **36, 12**, pp. 1395–1404.
3. Glasserman, P. and Kou, S. (1995) Analysis of an Importance Sampling Estimator for Tandem Queues, *ACM Trans. on Modeling and Computer Simulation,* **5, 1**, pp. 22–42.
4. Juneja, S. (2001) Importance Sampling and the Cyclic Approach, *Operations Research,* **49, 6**, pp. 900–912.
5. Kroese, D.P. and Nicola, V.F. (2002) Efficient Simulation of a Tandem Jackson Network, *ACM Trans. on Modeling and Computer Simulation*, to appear.
6. Parekh, S. and Walrand, J. (1989) A Quick Simulation Method for Excessive Backlogs in Networks of Queues, *IEEE Transactions on Automatic Control,* **35, 1**, pp. 54–66.