Proceedings of 3rd
Australian Teletraffic
Research Seminar,
Melbourne, November 1988

Estimation of Rare Event Statistics in Data Communications Networks *

Michael R. Frater † Brian D.O. Anderson †

October 18, 2006

Abstract

The estimation of the statistics of rare events is inherently costly, simply because of their rarity. This paper describes some recently developed techniques for the efficient simulation of rare events in telecommunications networks. The work described is split into three broad themes: the optimal simulation of buffer overflows in queueing networks, reverse-time modeling of random processes and optimal control. One of the new results presented provides an interconnection between all three of these themes.

1 Introduction

From one viewpoint, data communications networks can be viewed as interconnections of queues and servers, with practical considerations limiting the size of a buffer at any one queueing site. It is not satisfactory if overfilling of buffers occurs, and it is desirable to have methods for estimating the probability of a buffer overfilling in a given time. This is a rare event, but one that is bound to occur some time.

Within this paper we formulate some conjectures and possible approaches to the development of methods for estimating such probabilities, and indeed probabilities of other rare events in communications systems. The interaction of several ideas is involved:

- Rare Events and Importance Sampling
- Optimal Control
- Reverse Time Models of Random Processes

*Work supported by Australian Telecommunications and Electronics Research Board (ATERB).

†Department of Systems Engineering, Research School of Physical Sciences, Australian National University, G.P.O. Box 4, Canberra, A.C.T., 2601, Australia.

2 Rare Events - Examples

To help fix ideas, let us give some examples of rare events.

1. Consider a ball at the bottom of a finite well with smooth interior (see Figure 1). In the absence of random forces, the ball is at a stable equilibrium. If the ball is subject to random forces (say white noise) of modest intensity, one could conceive of a particular, rare, time-trajectory of random forces that drove the ball up the side of the well and over the top. This would be a rare event.

2. In decision directed equalization, it is known [1] that an adaptive equalizer can get hung up at an incorrect equilibrium. The motion of the equalizer taps is such that they move around, as if driven by noise, in the vicinity of the incorrect equilibrium, i.e. it is to all intents and purposes a noisy equilibrium. Generally, after a very long time, the random tap motions will lead to the tap settings being sufficiently far from the incorrect equilibrium that they are captured by another equilibrium.

3. In a well defined phased-lock loop, cycle slipping occurs only occasionally, i.e. the mean time between slips is much greater than the loop bandwidth. Cycle slipping occurs when 'adverse' noise sequences or realizations cause an unusually large error in the phase estimation, and cycle slips are rare events.

4. In a properly designed message network, with queues and buffers, etc., the buffer contents will take certain average values, but exhibit fluctuations around those values. Large fluctuations are less common than small ones, and very large fluctuations, such as might cause attempted overfilling of a buffer, are rare events.

The above examples of rare events are characterised by the two properties:

1. the event is certain to happen some time; and

2. the mean time between occurrences of the event is very long (and it is in this sense that the event is rare); this mean time may be adjustable by the designer of the system in which the events occur.

3 Direct Simulation of Rare Events

Even given an accurate computer model of a process in which rare events occur, their simulation for the purpose of obtaining statistics such as mean time between occurrences is intrinsically costly. Why? In order to get estimates of mean times or probabilities with any accuracy, one has to run the experiment often, and if any one occurrence of interest is occasional, the

time to collect statistics from which a reasonably accurate estimate can be obtained is very long. The more rare is the event, the more costly is the simulation exercise.

By way of (a possibly oversimplified) example, suppose A is the event that a given buffer in a certain network overflows in an interval of length 1 hour. We record the occurrence or non-occurrence of overflows over a period of n hours, not necessarily contiguous. We also suppose that the probability of overflow in hour i is independent of the probability of overflow in hour j for $i \neq j$. Suppose also that overflow is rare. Our aim is to estimate $P(A)$ from the n hours of experimental data. Thus we have a number of runs of an experiment that has a binary outcome, "success" ($X = 1$) corresponding to overflow or failure ($X = 0$), and nearly all the time, we get failures. Evidently A corresponds to a success, $X = 1$. Now the obvious unbiased estimate of $P(A)$ after n repeats of the experiment is

$$\widehat{P}(A) = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{1}$$

This has (it turns out) a standard deviation of $\sqrt{\frac{P(A)[1-P(A)]}{n}}$. To have a 20% error on $\hat{P}(A)$ with 95% confidence, we must take $\frac{\sqrt{1-P(A)}}{\sqrt{nP(A)}} \leq 0.1$, then for $P(A) \simeq 10^{-4}$, we need 10^6 samples, and if $P(A) \simeq 10^{-6}$, we need 10^8 samples.

The mean time between occurrences of A, the other statistic most of interest, is readily obtainable from $P(A)$ as $\frac{1}{P(A)}$ in this case.

4 Theory Applicable to Rare Event Calculations: Importance Sampling

There are two major approaches which have been used to try to make more practical the task of calculating probabilities of rare events.

- Importance sampling [2,3]
- Approximation by diffusions, coupled with optimal control [2,4,5,6]

The idea in importance sampling is as follows. Suppose we are interested in certain (rare) events occurring in a system S that we can simulate on a computer. Then instead of simulating S we simulate a second system $\overline{S}$, which has the property that events in S and $\overline{S}$ correspond in some way. In particular, to the rare events A in S correspond events $\overline{A}$ in $\overline{S}$. The correspondence is such that

- the events $\overline{A}$ in $\overline{S}$ are more frequent than the events A in S

- the connection between S and $\overline{S}$ allows one to infer $P(A)$ if one knows $\overline{P}(\overline{A})$.

This means that one can simulate $\overline{S}$ and obtain, by random experiment, an estimate of the probability $\overline{P}(\overline{A})$ of the $\overline{A}$ event. Then, knowing how S and $\overline{S}$ are related, one can compute the estimate of $P(A)$. Because the $\overline{A}$ events are more frequent, the determination of their statistics by simulation need not be so time consuming. So there is time saving in estimating $P(A)$. The advantages posed by speed up are equally applicable to the problem of obtaining (by simulation) the mean time between successive occurrences of a rare event.

As a simple example, see [7], consider as the system S an M/M/1 queue with arrival rate μ and service rate λ with $\mu/\lambda > 1$ (see Figure 2). Suppose we are interested in estimating for some large integer N the value of the expectation $E(T)$, where T denotes the first time the queue length hits N, assuming it is initially zero. (The mean time between successive occasions with queue length N could also be of interest; generally, it will differ very little from $E(T)$.)

This happens to be a problem which is analytically solvable. Nevertheless, it also serves to illustrate the idea in importance sampling. The second system $\overline{S}$ we consider is shown in Figure 3. It is also an M/M/1 queue, but with different arrival and service rates. The fact that S and $\overline{S}$ have the same structure with different transition probabilities between states is typical of importance sampling. Notice that in fact the queue in Figure 3 is unstable. In Figure 2, A is the event that the queue length is N starting at 0. In Figure 3, $\overline{A}$ is the event that the queue length is N, starting at 0. Obviously, the computation of the mean time for $\overline{A}$ to occur via simulation will be rapid. From it, one can deduce the mean time for A to occur as we now outline.

It is well known that one can set up an embedded discrete time Markov chain $\{X_k, k = 0, l, 2, \ldots\}$, with X_k denoting the queue length just after the k-th change of that length, i.e. just after the k-th occurrence of an arrival or departure. For convenience, one can rescale so that $\lambda + \mu = 1$.

Now let α be the probability that with $X_0 = 0$, X_k hits N before hitting zero again. Let T denote the first time X_k reaches N and let T_0 denote the time to hit 0 for the first time after leaving 0. It is not hard to check that

$$E(T) = \frac{1-\alpha}{\alpha} E(T_0) \tag{2}$$

Now $E(T_0)$ can be easily obtained by direct simulation on S, while α, which will be small, is obtainable from $\overline{S}$, as follows. Let us call a *cycle* of the system S or $\overline{S}$ a movement from 0 to the first time *either* 0 is reached again, *or* N is reached. Define $V_k = 1_{\{X_m \text{reaches } N \text{ in cycle } k\}}$. For S, we have

$$E(V_k) = \alpha \tag{3}$$

Let L_k denote the likelihood ratio $\frac{dP}{d\overline{P}}$ during cycle k. Notice that the L_k are i.i.d. and

$$\overline{E}(L_k V_k) = E(V_k) = \alpha \tag{4}$$

Now L_k is readily definable for S and $\overline{S}$. Suppose that $V_k = 1$, that there are l departures and $N + l$ arrivals. Then

$$\begin{array}{c} \overline{P}(V_k) = \mu^{N+l}\lambda^l \\ P(V_k) = \lambda^{N+l}\mu^l \end{array} \tag{5}$$

and so

$$L_k = \frac{dP}{d\overline{P}} = (\frac{\lambda}{\mu})^N \tag{6}$$

on the set $\{V_k = 1\}$.

There are frequent occurrences of the set $\{V_k = 1\}$ for the system $\overline{S}$, since it is unstable. We examine p cycles for $\overline{S}$ and estimate α by

$$\begin{aligned} \widehat{\alpha} &= \frac{L_1 V_1 + L_2 V_2 + \ldots + L_p V_p}{p} \\ &= (\frac{\lambda}{\mu})^N \frac{\text{Number of cycles for which } V_k = 1}{p} \end{aligned} \tag{7}$$

The speed up factor in simulation time obtained by using $\overline{S}$ instead of S and requiring equal accuracy in estimating a turns out to be, see [7]:

$$[N(\frac{\lambda}{\mu})^N(1 - \frac{\lambda}{\mu})]^{-1} \tag{8}$$

In the above analysis we have shown how knowing and simulating $\overline{S}$, it can be used to compute a and thus important statistics concerning S. But we have not described what led us to the particular $\overline{S}$ chosen, rather than something else. *A major issue in the use of importance sampling is how one should construct $\overline{S}$ from S.* To an extent, the problems of obtaining the probability of a rare event or the mean time between occurrences of a rare event (which are both problems of excessive computer time) are being replaced by another difficult problem ('How should we obtain $\overline{S}$ from S?') in importance sampling.

This problem can be posed as an optimization problem, in the following way. Let A be a rare event for S: $\alpha = P(A) \ll 1$. For a direct Monte Carlo simulation involving n independent experiments we could estimate α via

$$\widehat{\alpha}_n = \frac{1}{n}\sum_{i=1}^{n} 1_A(\omega_i) \tag{9}$$

where the ω_i are the i.i.d. outcomes of the experiments. The variance of $\widehat{\alpha}_n$ is easily computed as

$$E(\alpha - \widehat{a}_n)^2 = \frac{1}{n}(\alpha - \alpha^2) \tag{10}$$

Alternatively, consider a probability measure $\overline{P}$ associated with a system $\overline{S}$, with P absolutely continuous with respect to $\overline{P}$. Using $\overline{S}$ we can obtain a second estimation

$$\widehat{\overline{\alpha}}_n = \frac{1}{n}\sum_{i=1}^{n} 1_A(\overline{\omega}_i)L(\overline{\omega}_i) \tag{11}$$

where $L = \frac{dP}{d\overline{P}}$ and the $\overline{\omega}_i$ are the outcome of n experiments using $\overline{S}$. The variance of $\widehat{\overline{\alpha}}$ is obtainable as

$$\frac{1}{n}(\int_A L^2(\omega)d\overline{P}(\omega) - \alpha^2) \tag{12}$$

We want this to be as accurate as possible. So we want to adjust all the transition probabilities in S to new ones in $\overline{S}$ so that

$$(\sigma^*)^2 = \int_A L^2(\omega)d\overline{P}(\omega) \tag{13}$$

is minimized.

In general, direct minimization of this quantity is extemely difficult and less direct arguments using large deviations theory have been used ([2], [7]). However, for the example of the M/M/1 queue, we have found that a direct minimization of this quantity is possible.

For the M/M/1 queue, suppose that $\overline{S}$ is obtained from S by varying λ, μ to λ^* and μ^*. Then the likelihood ratio (L) takes on two values, one for an upward step ($\frac{\lambda}{\lambda^*}$), the other for a downward step ($\frac{\lambda}{\lambda^*}$). The value of the likelihood ratio for a trajectory is found by multiplying the likelihood ratios for the steps making up the trajectory. Since this system is discrete both in time and state-space, equation (4.11) can be rewritten as follows:

$$(\sigma^*)^2 = \sum_A (\frac{\lambda^2}{\lambda^*})^{N+l}(\frac{\mu^2}{\mu^*})^l \tag{14}$$

where A is the set of all trajectories that exit reach N before reaching 0. Let

$$C_n = \sum_{A_n} (\frac{\lambda^2}{\lambda^*})^{N+l}(\frac{\mu^2}{\mu^*})^l \tag{15}$$

where A_n is the set of all trajectories that start at n and reach N before reaching 0. Then we can use the first step method to find C_n, and hence $(\sigma^*)^2$. The first step equation is:

$$C_n = \frac{\lambda^2}{\lambda^*}C_{n+1} + \frac{\mu^2}{\mu^*}C_{n-1} \tag{16}$$

with boundary conditions $C0 = 0$ (since $C_1 = (\frac{\lambda^2}{\lambda^*}C_2)$ and $C_N = 1$ (since $C_{N-1} = \frac{\lambda^2}{\lambda^*} + \frac{\mu^2}{\mu^* C_{N-2}}$) Solving these equations, we find:

$$C_n = (2\frac{\lambda^2}{\lambda^*})^{N-n}[\frac{(1+\sqrt{1-4\frac{\lambda^2}{\lambda^*}\frac{\mu^2}{\mu^*}})^n - (1-\sqrt{1-4\frac{\lambda^2}{\lambda^*}\frac{\mu^2}{\mu^*}})^n}{(1+\sqrt{1-4\frac{\lambda^2}{\lambda^*}\frac{\mu^2}{\mu^*}})^N - (1-\sqrt{1-4\frac{\lambda^2}{\lambda^*}\frac{\mu^2}{\mu^*}})^N}] \tag{17}$$

In fact, all the trajectories are defined to start at 1. Hence, the quantity we desire to minimize is

$$(\sigma^*)^2 = C_1 \tag{18}$$

When C_1 is differentiated with respect to λ^*, it is found that its derivative goes to zero for large N when $\lambda^* = 1 - \lambda$. i.e. for an M/M/1 queue, the variance of the probability estimator is minimized by exchanging the arrival and service rates.

Unfortunately, it does not seem to be possible to apply the methods used here to more complicated examples, such as network of M/M/1 queues, as in these examples, analytic solutions of the first-step equations cannot be found.

5 Rare Events And Diffusion Processes

A diffusion process is formally modelled by an equation of the form

$$dx = f(x)dt + g(x)dw \tag{19}$$

Here, x is an n-vector, $\frac{dw}{dt}$ is an m-vector of white noise processes, f(x) is an n-vector function and g(x) is an $n \times m$ matrix function. Certain smoothness assumptions are fulfulled by f, g. As already mentioned, finite state Markov processes can be approximated by diffusion processes - indeed this has been done in [2], in considering the Aloha systems, which is actually a finite state process.

Now rare event theory is much more developed for diffusion processes [2], [4], [5], [6]. In this section, we review some of the results, and point out a hitherto unknown fact.

Suppose that $\dot{x} = f(x)$ is asymptotically stable. Hence if there is no noise in (5.1), trajectories will decay, while if there is small noise, say (5.1) is replaced by

$$dx = f(x)dt + \epsilon g(x)dw \tag{20}$$

with ϵ small, then trajectories will tend not to get larger, but rarely may get large. Consider the following large time scale event: Let D be a bounded region containing zero, with boundary ∂D, and consider the event that a trajectory starts at $x(0) = \chi$ and reaches D. Let $\tau^{(\chi,\epsilon)}$ be the time required

for this. Then the expected value of τ can be characterized in terms of the solution of a certain optimal control problem, which we now discuss.

Consider the system

$$\dot{x} = f(x) + g(x)u \tag{21}$$

where we must minimize, via choice of $u(\cdot)$ and T the performance function

$$V(x_0, u(\cdot)) = \frac{1}{2}\int_{-T}^{0} u'udt \tag{22}$$

subject to the constraints $x(-T) = 0$, $x(0) = x_0$. Denote the optimal performance by $V^*(x_0)$; it is a fact that at any intermediate point y on an optimal trajectory encountered at some time $t < 0$, the optimal control can be expressed in feedback form as

$$u^*(t) = g'\frac{\partial V^*}{\partial y} \tag{23}$$

Optimal trajectories of (5.3) than all have the form

$$\dot{x} = f(x) + gg'\frac{\partial V^*}{\partial x} \tag{24}$$

with the constraint $x(0) = x_0$.

The main connections with the exit problem is twofold. First,

$$\lim_{\epsilon \downarrow 0} \epsilon^2 \log E(\tau^{x_0,\epsilon}) = \inf x_0 \in \partial D V^*(x_0) \tag{25}$$

i.e. $V^*(x)$ contains information about the exit time. Secondly, suppose there exist one or more points $y \in \partial D$ for which

$$V^*(y) = \inf_{x_0 \in \partial D} V^*(x_0) \tag{26}$$

Then as $\epsilon \to 0$, the exiting trajectories for (5.2) exit more and more closely with probability approaching 1 to one of the y, i.e. for arbitrary $\delta > 0$, with x(t) denoting the solution of (5.2) at time t, there holds

$$\lim_{\epsilon \downarrow 0} p(\min_{\substack{ysatisfying \\ (5.8)}} \|X(\tau^{x_0,\epsilon}) - y\| < \delta) = 1 \tag{27}$$

There is however an additional observation which can be made: there is a *reverse time* model corresponding to (5.2), i.e. a stochastic differential equation evolving *backwards* in time with trajectories indistinguishable from those of (5.2). Calculations based on [11] show that as $\epsilon \to 0$, this is given ever more accurately by

$$dx = (f(x) + g(x)g'(x)\frac{\partial V^*}{\partial x})dt + \epsilon\, g(x)d\overline{w} \tag{28}$$

where $\frac{d\overline{w}}{dt}$ is an r-vector white noise process. The trajectories of (5.10) cluster round the mean

$$\frac{dx}{dt} = f(x) + g(x)g'(x)\frac{\partial V^*}{\partial x} \tag{29}$$

which is nothing other than (5.6).

The crucial observation is that *the solution of the optimal control problem and the solutions of the problem of constructing a reverse time modal are*, as $\epsilon \downarrow 0$, *effectively the same.*

Now it is recognized in [2] that the optimal control solution can be used to define the system $\overline{S}$ for use in importance sampling. It is virtually (5.9). So here we have a connection between (a) the importance sampling construction (b) optimal control (c) reverse time models.

Is there some general principle that is operative here, and might extend beyond diffusion equations? Recall the simple example of the M/M/1 queue. The system $\overline{S}$ found there is exactly that associated with a reverse time model of S! So in this instance, the connection (a) $\Leftrightarrow$ (c) certainly exists.

6 Jackson Networks and General Queueing Networks

A Jackson network is an interconnection of M/M/1 queues, as described in previous sections, with routing probabilities independent of both time and the state of the network. If a steady-state distribution exists for the network (i.e. it is stable in the mean), then node i is subjected to traffic arriving at rate λ_i, where the λ_i are found by solving the traffic equations:

$$\lambda_i = \gamma_i + \sum_j \lambda_j q_{ji} \tag{30}$$

where γ_i is the rate of external arrivals at node i, and q_{ji} the probability that a job will be routed from node j to node i.

If $p_i(k)$ is the invariant probability that node i contains k jobs, and $p(k_1, \ldots, k_M)$ the invariant probability that node l contains k_1 jobs etc., Jackson's theorem (see [13], p 174) tells us:

$$p(k_1, \ldots, k_M) = \prod p_i(k_i) \tag{31}$$

The following statements follow from this result:

1. The numbers of jobs at each node are independent of each other.

2. Each node behaves as if it were subjected to a stream of Poisson arrivals, with rate determined by the traffic equations.

It should be noted that Jackson's theorem applies only to the invariant probability, and cannot be used for time-dependent probabilities.

While the mean time between overflows can be calculated analytically for a single M/M/l queue (eg [7]), this is not possible for more general Jackson networks, so simulation is necessary. Just as it may not be practical to perform direct simulation of a single M/M/1 queue, other techniques may be required for the simulation of networks.

The analysis of Jackson networks presents significant difficulties. Large Deviations Theory has been used to obtain a number of approximate results [7]. Some of these, along with our extensions to these results, are presented below.

Consider the case of two queues in tandem (Figure 4). If, for example $\mu_1 > \mu_2$, then if the size of the buffers (N) is large, the overflow statistics will be dominated by the second buffer's behaviour, and μ_1 will be of little importance, i.e. from the point of view of buffer overflows, the system essentially behaves as a single buffer with arrival rate λ and service rate μ_2. Hence, it is reasonable to suspect that for large N, the optimal simulation system will behave very like a single buffer with arrival rate μ_2 and service rate λ. Let us now see how this conclusion can be rigorously established, using methods that will apply to general Jackson networks and possibly general queueing networks.

Let $\xi_1 \ldots \xi_d$ be i.i.d random variables in $\Re^d$. Let F be the distribution function of ξ_k and m its mean. Assume that the Laplace transform of F

$$M(s) = \int_{\Re^d} \exp < s, z > dF(z) \tag{32}$$

is finite in a neighbourhood of 0. Then the Cramér or Legendre transform is defined as as

$$h(y) = \sup_{\Re^d}[< s, y > -logM(s)] \tag{33}$$

For example, the Cramér transform of an exponential distribution with parameter λ is

$$h_\lambda(u) = \begin{cases} \lambda u - \log(\lambda u) - 1 & u > 0 \\ \infty & otherwise \end{cases} \tag{34}$$

The following properties of the Cramér transform are used in this paper:

1. h is convex,

2. h(m) = 0, where m is the mean of the distribution function.

For the tandem queues, call a cycle a piece of a trajectory starting at the zero state and terminating on the first occasion when either the total buffer contents equals N or again equals zero. If α is the probability that a cycle

reaches N, then for a pair of buffers in tandem, it can be argued (see [7]) that a is approximated by

$$\alpha = \exp -N \min_{\substack{\lambda'>0,\mu_1',\mu_2'\geq 0,\\ \lambda'>\mu_1' or \lambda'>\mu_2'}} \frac{1}{R}[\lambda' h_\lambda(\frac{1}{\lambda'}) + \mu_1' h_{\mu_1}(\frac{1}{\mu_1'}) + \mu_2' h_{\mu_2}(\frac{1}{\mu_2'})] \tag{35}$$

where

$$R = \begin{cases} \frac{1}{\lambda'-\mu_1'} & \mu_1' < \mu_2' \\ \frac{1}{\lambda'-\mu_1'} & otherwise \end{cases} \tag{36}$$

Analytic minimization implies that for optimal simulation (in the sense of variance), λ should be swapped with the smaller of μ_1 and μ_2. For example, if $\mu_1 > \mu_2$, we have $\lambda' = \mu_2$, $\mu_1' = \mu_1$, $\mu_2 = \lambda$. In fact, we can show that this solution generalizes to networks of M queues in tandem, i.e., we exchange the external arrival rate with the smallest of the service rates. Also, we can show that this solution is unique.

By definition, the subscript μ of $h_\mu(\cdot)$ is the expected number of services per unit time. Hence, the mean time between services is $\frac{1}{\mu}$. Therefore, for buffers whose service rates are the same in both the original and optimal-simulation systems, (i.e. $\mu_j = \mu_j'$) we have

$$h_{u_j}(\frac{1}{\mu_j'}) = h_{u_j}(\frac{1}{\mu_j}) = 0 \tag{37}$$

since the Cramér transform has value zero at the mean of its associated distribution function.

Hence, we can conclude that for sufficiently large N, the optimal simulation system depends only on the statistics of one service rate and the arrival process, assuming that no two service rates are actually equal. It is important to note that this property does not depend in any way on the statistics of the service rates of buffers other than the one dominating the overflow statistics.

From these results, we suspect that the mean time between overflows in a general Jackson network can be found by solving the traffic equations for the network and finding the mean time between overflows for the buffers individually, using their effective arrival and service rates. From these times, the mean time between overflows for the network can be found by adding them like resistances in parallel.

7 Conclusion

There are a number of key tasks yet to be performed. These include the connection of optimal control, reverse-time modelling and importance sampling for discrete-time systems without the need for approximation by diffusion

equations, and there are many known results for reverse-time models (e.g. [8]) that are yet to be brought to bear. Also, a rigorous justification is required to show that the overflow of queues in general networks is determined only by the net arrival and service rates of individual queues.

REFERENCES

[1] J.E. Mazo, "Analysis of decision-directed equalizer", *Bell Syst. Tech. J.*, Vol. 59, No. 10, December 1980, pp. 1857-1876.

[2] M. Cottrell, J.-C. Fort and G. Malgouyres, "Large deviations and rare events in the study of stochastic algorithms", *I.E.E.E. Trans. Auto. Contr.*, Vol. AC-28, No. 9, September 1983, pp. 907-919.

[3] D. Gaver and G. Thompson, *Programming and Probability Models in Operations Research*, Monterey, CA: Brooks/Cole, 1972.

[4] M.I. Friedlin and A.D. Wentzell, *Random Perturbation of Dynamic Systems*, New York: Springer Verlag, 1984.

[5] J. Zabczyk, "Exit Problems and Control Theory", *Syst. and Contr. Letters*, Vol. 6, No. 3, August 1985, pp. 165-172.

[6] H. Kushner, *Approximation and Weak Convergence Methods for Random Processes*, Cambridge, MA: M.I.T. Press, 1984.

[7] S. Parekh and J. Walrand, "Quick simulation of excessive backlogs in networks of queues", *Proc. 25th I.E.E.E. Conf. on Dec. and Contr.*, Athens 1986, pp. 979-986.

[8] F.P. Kelly, *Reversibility and Stochastic Networks*, New York: John Wiley and Sons, 1979.

[9] B.D.O. Anderson and T. Kailath, "Forward, backward and dynamically reversible Markovian models for second order processes", *I.E.E.E. Trans. Circuits and Systems*, Vol. CAS-26, No. 11, November 1979, pp. 956-965.

[10] B.D.O. Anderson and T. Kailath, "Forwards and backwards models for finite-state Markov processes", *Advances in Applied Probability*, Vol.11, 1978, pp.118-133.

[11] B.D.O. Anderson, "Reverse-time diffusion equation models", *Stoch. Proc. and Applic.*, Vol.12, No.31, 1982, pp.313-326.

[12] R.J. Elliot and B.D.O. Anderson, "Reverse time diffusions", *Stochastic Processes and their Applications*, Vol.19, 1985, pp.327-339.

[13] D. Bertsekas and R. Gallager, *Data Networks*, New Jersey: Prentice-Hall, 1987.

8 Errata

Submitted to 3rd Australian Teletraffic Research Seminar, 3rd November 1988.