UNE **IMP**

EHE **CWI**

QMIPS

INR **UER**

ZAR **UTO**

Q uantitative
M odeling
I n
P arallel
S ystems

Approximate Solutions for Open Networks with Breakdowns and Repairs

Ram Chakka Isi Mitrani

Computing Science Department, University of Newcastle
Newcastle upon Tyne NE1 7RU, England

Distribution Level: Unrestricted

Abstract Network models whose nodes are subject to interruptions of service are notoriously difficult to solve. There are no exact results for systems with more than two nodes, and even the approximations that have been proposed turn out to be rather unsatisfactory. We assess the existing approaches and develop new approximation methods. The latter's accuracy is evaluated by comparisons with simulations and is shown to be acceptable over a large range of parameter values.

The QMIPS Project: The QMIPS project is a collaborative research project supported by the CEC as ESPRIT-BRA project no 7269. It is being carried out by the following organisations: CWI (Amsterdam), EHEI (University of Paris V), Imperial College (London), INRIA (Sophia-Antipolis), University of Erlangen, University of Newcastle, University of Torino and University of Zaragoza.

1 Introduction

Despite the great importance and widespread use of queueing network models, there have been very few efforts to analyse their behaviour when the servers are not completely reliable. The general problem is difficult because, even under favourable assumptions, the existence of service interruptions destroys the separability of the appropriate multidimensional Markov process and renders its solution intractable.

Most of the existing studies involving breakdowns have concentrated on models with a single job queue served by one or more processors (e.g., see [1, 8, 15, 18]). Some results are available for systems where jobs from a single source are directed to one of several parallel queues, and after completion depart: in [11], the consequences of a server breakdown are the loss of all jobs in the corresponding queue and the re-direction or loss of all arrivals to that queue during the subsequent repair period; [16] treated a similar model but without job losses; [4] examined the case of two queues, only one of which is subject to breakdowns resulting in the transfer of all jobs to the other.

To our knowledge, the only 'genuine network' (in the sense that after completing service at one node jobs may go to another), which has been analysed in the presence of breakdowns and repairs, consists of only two nodes [6]. Even this apparently simple case is not treated in its full generality: a tight coupling is assumed, so that a breakdown interrupts service at both nodes simultaneously, and both nodes complete their repair simultaneously. The analysis in [6] provides ample illustration of the formidable mathematical difficulties posed by this kind of model.

In view of the above, it seems clear that approximate solutions offer the best hope of obtaining performance measures for non-trivial unreliable networks. Indeed, some steps in that direction have been taken already. A simple approach known as the *reduced work-rate approximation* was proposed in the context of both priority interruptions and breakdowns [12, 14, 17]. This is based on replacing an interruptable server with an uninterruptable but slower one, choosing the new service rate so that the overall service capacity remains the same. The evaluation of that approximation has been very limited.

Another possibility is to model the breakdowns and repairs exactly, but reduce the interactions between nodes by considering each of them in isolation. The simplest assumption would be to say that the total arrival stream into a node is Poisson. This *Poisson approximation* was suggested in [7], but its accuracy has not been evaluated.

The object of the present paper is to examine in some detail the problem of approximating the behaviour of open Jackson networks where nodes break down and are repaired independently of each other. The regime of interest is one where both the operative periods and the repair intervals are large compared to the job interarrival and service times. It turns out that, under those conditions, the reduced work-rate approximation is a very poor choice indeed. The Poisson approximation is better, but it too can be inaccurate, especially in networks with cycles.

Two new approximation methods are developed and evaluated. The first, referred to as the *MMPP-approximation*, attempts to capture the non-renewal nature of the arrival stream into a node by modelling it as a Markov-Modulated Poisson process; the departure streams are modelled as interrupted Poisson processes (the latter may of course contribute to the former). The second method takes explicit account of the operative/broken states of several (perhaps all) nodes; it will be called the *joint-state approximation*. In both approaches, if the network has cycles, then the parameters of the arrival processes are found by iterations converging to a fixed point. Thus, a considerably higher accuracy is obtained at the expense of increasing the numerical complexity of the solution.

We start by introducing the network model and the two existing approximations, reduced work-rate and Poisson, in section 2. The MMPP and joint-state approximations are described and analysed in sections 3 and 4, respectively. A number of numerical evaluation results and comparisons with simulations are presented in section 5.

2 Jackson networks with breakdowns and repairs

Consider an open Jackson queueing network (e.g., see [5]), with N nodes. The external job arrival streams are Poisson, with arrival rates given by the row vector $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_N)$. The service times at all nodes are exponentially distributed, with parameters specified by the row vector $\mu = (\mu_1, \mu_2, \ldots, \mu_N)$. The routing decisions are Bernoulli, governed by the matrix $R = (r_{i,j})$, $(i, j = 1, 2, \ldots, N)$, whose elements are the probabilities that after completing service at node i, a job goes to node j. Every node contains a single server and an unbounded FIFO queue.

The possibility of relaxing some of the above assumptions will be discussed at the end of the paper.

The server at node i goes through alternating operative and inoperative periods, distributed exponentially with parameters ξ_i and η_i, respectively $(i = 1, 2, \ldots, N)$. These periods are independent of the states of other nodes and of the number of jobs in the queue. Transitions from the operative to the inoperative state are called *breakdowns*, while those from the inoperative to the operative state are *repairs*. If a breakdown occurs during a service, the latter is resumed from the point of interruption after the repair. Incoming jobs continue to join the queue during inoperative periods.

Because of the breakdowns, the average number of jobs that server i can complete per unit time, ν_i, is given by

$$\nu_i = \mu_i \frac{\eta_i}{\xi_i + \eta_i} \; ; \; i = 1, 2, \ldots, N \; . \tag{1}$$

This is the *service capacity* at node i.

Let $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$ be the vector of total job arrival rates at the N nodes. These rates are obtained as the unique solution of the set of linear equations

$$\boldsymbol{\sigma} = \gamma + \boldsymbol{\sigma} R \; . \tag{2}$$

The ergodicity condition for this network is simple: the total arrival rates should be smaller than the corresponding service capacities, i.e. $\sigma_i < \nu_i$ for all i. This will be assumed to hold. However, finding the joint or marginal distributions of queue sizes, and hence any performance measures associated with them, is an open problem for $N > 1$.

The reduced work-rate approximation consists of replacing the unreliable servers with reliable ones that have the same service capacity. In other words, the new server at node i has service rate ν_i and breakdown rate 0 $(i = 1, 2, \ldots, N)$. The arrival rates and routing probabilities remain the same. The resultant network has a product-form solution and its performance measures are easily calculable. In particular, the reduced work-rate approximation for the average queue size at node i, L_i, is

$$L_i = \frac{\sigma_i}{\nu_i - \sigma_i} \; ; \; i = 1, 2, \ldots, N \; . \tag{3}$$

The major drawback with this approach is that, since ν_i depends only on the ratio of ξ_i and η_i, and not on their separate values, the reduced work-rate approximation depends only on the long-term fraction of time that the server is operative, and is insensitive to the frequency of breakdowns and repairs. Yet the network performance measures depend very strongly on that frequency (see section 5). For example, if the average lengths of the operative and inoperative periods at node i increase, keeping their ratio and the other parameters fixed, then the average queue size at that node grows without bound while the value given by (3) remains constant. An unreliable node may be very lightly loaded, in terms of its service capacity (i.e., it may be operative and idle most of the time), and yet its average queue size may be very large. Thus, the reduced work-rate approximation can be arbitrarily bad when servers break down rarely and take a long time to repair, which is precisely the case of practical interest.

Better results are obtained by treating node i as an isolated, independent M/M/1 queue and modelling its breakdowns and repairs exactly. This approximation consists of assuming that the arrivals into node i form a Poisson stream with rate σ_i; the other assumptions are left intact. The known solution of the isolated queue (e.g., [1, 8]), yields the following Poisson approximation for the average queue sizes:

$$L_i = \frac{\sigma_i}{\nu_i - \sigma_i}[1 + \nu_i \frac{\xi_i}{\eta_i(\xi_i + \eta_i)}] \; ; \; i = 1, 2, \ldots, N \; . \tag{4}$$

Whenever breakdowns and repairs occur, the values provided by (4) are larger than those in (3). Moreover, the difference between the two becomes arbitrarily large when the average operative and inoperative periods grow in a fixed ratio.

The experiments in section 5 show that the Poisson approximation is often quite reasonable. Moreover, it is particularly good in large, strongly connected networks. This is of course due to the fact that the superposition of many input streams is approximately Poisson, almost regardless of their individual characteristics. However, the quality of the approximation can be unsatisfactory in small to moderately sized networks. There the dependency between nodes and the non-Poisson, non-renewal nature of the traffic ought to be taken into account. This is the object of the following sections.

3 The MMPP approximation

A generalisation of the Poisson approximation is to treat node i as an isolated, independent queue subject to breakdowns and repairs, with jobs arriving according to a Markov-Modulated Poisson process (or MMPP, see [3]). Suppose that the arrival rate is controlled by an irreducible Markov chain, $Y(t)$, with T states numbered $1, 2, \ldots, T$ (here and whenever there is no danger of misunderstanding, we omit the node index, i). When $Y(t) = j$, the instantaneous arrival rate into the node is σ_j ($j = 1, 2, \ldots, T$). Denote the generator matrix of Y by M, and its steady-state distribution by $\mathbf{a} = (a_1, a_2, \ldots, a_T)$. The latter is obtained by solving the equations

$$\mathbf{a}M = 0 \;\; ; \;\; \sum_{j=1}^{T} a_j = 1 \; .$$

The overall arrival rate, σ, is equal to

$$\sigma = \sum_{j=1}^{T} a_j \sigma_j \ . \tag{5}$$

The queue is stable iff $\sigma < \mu\eta/(\xi + \eta)$.

Let the binary variable $\delta(t)$ indicate the state of the server at time t: 0 if broken, 1 if operative. Then the joint state of the server and the MMPP can be represented by a single random integer, $U(t) = T\delta(t) + Y(t)$, taking values in the range $1, 2, \ldots, 2T$. When $U(t) \leq T$, the server is broken; when $U(t) > T$, the server is operative. The state of the queue at time t is described by the pair of integers $[U(t), V(t)]$, where $V(t)$ is the number of jobs present. That pait is a Markov process on the state space $\{1, 2, \ldots, 2T\} \times \{0, 1, \ldots\}$. The corresponding steady-state probabilities, $p_{k,n} = P(U = k, V = n)$, can be determined exactly, e.g. by spectral expansion, [10, 9], or by the matrix-geometric method, [13]. The solution yields various performance measures, such as the average queue size:

$$L = \sum_{n=1}^{\infty} \sum_{k=1}^{2T} n p_{k,n} \ . \tag{6}$$

However, we still have to address the question of how to find the parameters T, M and σ_j, describing the arrival MMPP at each node. This is done by an iterative procedure converging to a fixed point. The first step is to determine the Laplace transform of the interval between consecutive departures from a node, when the arrival MMPP into that node is given and the probabilities $p_{k,n}$ have been calculated.

Let t be an arbitrary point of time in the steady-state, and τ be the interval between t and the next departure. Define the following conditional Laplace transforms:

$$\Upsilon(s) = E[e^{-s\tau} \,|\, U(t) \leq T, V(t) > 0]$$

(at time t, the server is broken and the queue is non-empty);

$$\Omega(s) = E[e^{-s\tau} \,|\, U(t) > T, V(t) > 0]$$

(the server is operative and the queue is non-empty);

$$\Gamma_k(s) = E[e^{-s\tau} \,|\, U(t) = k, V(t) > 0] \ , \quad k = 1, 2, \ldots, 2T$$

(the joint state of the server and the MMPP is k, and the queue is empty).

By conditioning upon the number of breakdowns that may occur during a service (geometrically distributed with parameter $\xi/(\mu + \xi)$), we obtain

$$\Omega(s) = \frac{\mu(\eta + s)}{(\eta + s)(\mu + \xi + s) - \xi\eta} \ . \tag{7}$$

Also, since a broken server has to be repaired before it resumes service,

$$\Upsilon(s) = \frac{\eta}{\eta + s}\Omega(s) = \frac{\mu\eta}{(\eta + s)(\mu + \xi + s) - \xi\eta} \ . \tag{8}$$

If the server is broken and the queue is empty, then the next event to occur may be an arrival, a repair, or a change of state in the controlling process Y. Hence, for $j = 1, 2, \ldots, T$,

$$\Gamma_j(s) = \frac{\sigma_j \Upsilon(s)}{\sigma_j + \eta + m_j + s} + \frac{\eta \Gamma_{T+j}(s)}{\sigma_j + \eta + m_j + s} + \sum_{k=1,\neq j}^{T} \frac{m_{j,k} \Gamma_k(s)}{\sigma_j + \eta + m_j + s} , \tag{9}$$

where $m_{j,k}$ is the rate at which the process Y moves from state j to state k, and m_j is the rate at which it leaves state j.

Similarly, If the server is operative and the queue is empty, then the next event to occur may be an arrival, a breakdown, or a change of state in the controlling process Y. This yields,

$$\Gamma_{T+j}(s) = \frac{\sigma_j \Omega(s)}{\sigma_j + \xi + m_j + s} + \frac{\xi \Gamma_j(s)}{\sigma_j + \xi + m_j + s} + \sum_{k=1,\neq j}^{T} \frac{m_{j,k} \Gamma_{T+j}(s)}{\sigma_j + \xi + m_j + s} . \tag{10}$$

The linear equations (9) and (10), together with (7) and (8), determine all the conditional Laplace transforms. Now consider the queue immediately folowing a departure instant. Let $\pi_{j,0}$ be the probability that *at that epoch*, the process Y is in state j and the queue is empty. That probability is equal to the fraction of all departures which occur when the process Y is in state j, the server is operative and the queue size is 1. Remembering that the overall average number of departures per unit time is equal to σ, given by (5), we can write

$$\pi_{j,0} = \frac{p_{T+j,1} \mu}{\sigma} \; ; \;\; j = 1, 2, \ldots, T . \tag{11}$$

The probability that the queue is empty at a departure epoch is equal to $\pi_0 = \pi_{1,0} + \pi_{2,0} + \ldots + \pi_{T,0}$, while the probability that it is not empty is equal to $1 - \pi_0$. Hence, the unconditional Laplace transform of the inter-departure interval, $\Delta(s)$, is given by

$$\Delta(s) = (1 - \pi_0) \Omega(s) + \sum_{j=1}^{T} \pi_{j,0} \Gamma_{T+j}(s) . \tag{12}$$

Having obtained $\Delta(s)$, one can determine successive moments of the inter-departure interval by taking derivatives in (12), (9), (10), (7) and (8), at $s = 0$ (the first moment is of course equal to $1/\sigma$). Those moments can be used to fit an MMPP to the departure process, although the computational task is complex if large number of phases are involved. If one decides to approximate the departure process by an MMPP with D phases, there would be D^2 parameters to determine (the off-diagonal elements of the matrix M and the phase-dependent departure rates), and so D^2 moments would be needed.

It is desirable, therefore, to approximate the departure process by an MMPP with a small number of phases. Bearing in mind that there are no departures when the server is broken, we decided to choose a 2-phase approximation with departure rate of 0 during one of them. That is, we approximate the departure process from each node by an IPP (Interrupted Poisson Process). Since the latter is determined by just 3 parameters, only the first 3 moments of the inter-departure interval have to be calculated.

Markov-modulated Poisson processes can be split or merged. Bernoulli splitting according to fixed probabilities is simple: the number of phases and the generator matrix remain the same, while the phase-dependent rates are multiplied by those probabilities. Merging is more complicated: if K independent MMPP's with $T_1, T_2, \ldots, T_K$ phases respectively are merged, the resulting process has $T_1 T_2 \ldots T_K$ phases. Expressions for the parameters of the merged process are available [3].

Now, the overall arrival stream into node i is formed by merging its external arrivals (if any), with those departures from node j which are routed to node i, for all j different from i (feedbacks from i to i can be eliminated by suitably changing μ_i). If the arrival processes into those other nodes are known, then solving the corresponding isolated models, fitting an IPP to each of their departure processes and performing the appropriate splitting and merging operations, produces an MMPP arrival process for node i.

If the network is acyclic, i.e. if no job can ever return to a node it has visited already, then one can start with the nodes whose arrivals are entirely external (there must be at least one such node), and by applying the above procedure, construct MMPP arrival streams at all nodes and solve the network in one sweep.

If the network has cycles, then it is necessary to iterate. The initial step is to assume Poisson arrivals into all nodes, with rates σ_i obtained from (2), and solve the isolated models. From then on, the departure processes computed in the last iteration are used to construct the arrival processes for the next one. The procedure stops when some convergence criterion is satisfied, e.g. the change in the average queue sizes (6) is less than some ϵ.

Numerical experiments show that the MMPP approximation is very reasonable in acyclic networks, but may over-estimate the queue sizes considerably when there are cycles (see section 5). To improve the accuracy in the latter case, one should take into account the fact that the arrival rate into a node which is part of a cycle is influenced by whether it, as well as the other nodes, are broken or operative. The following approximation attemps to remedy this defficiency.

4 The joint-state approximation

The joint state of all nodes in the network at time t is described by the vector of indicators $\delta_1(t), \delta_2(t), \ldots, \delta_N(t)$, where $\delta_i(t)$ is 0 if node i is broken, 1 if operative. Alternatively, that joint state can be described by a single integer random variable, $U(t)$, whose values have N-digit binary representations given by the indicator vector ($i.e., U(t) = 0, 1, \ldots, 2^N - 1$).

Consider the queue at node i in isolation from the other queues, but controlled by the Markovian environment $U(t)$. That environment changes phase whenever *any* operative node breaks down or a broken one is repaired; the generator matrix associated with those transitions is known, since all ξ and η parameters are gigen. The instantaneous service rate at node i when $U(t) = k$, $\mu_{i,k}$, is equal to μ_i if node i is operative in phase k and is 0 otherwise.

The joint-state approximation is based on assuming that the arrival process into node i is MMPP and is controlled by $U(t)$. The instantaneous arrival rate is $\sigma_{i,k}$ when $U(t) = k$. Then the pair $[U(t), V_i(t)]$, where $V_i(t)$ is the number of jobs present at node i, is a Markov process on the state space $\{0, 1, \ldots, 2^N - 1\} \times \{0, 1, \ldots\}$. If the rates $\sigma_{i,k}$ are known, the steady-state probabilities, $p_i(k, n) = P(U = k, V_i = n)$, and the desired performance measures, can be determined.

The departure rate from node i in phase k, $d_{i,k}$, is equal to

$$d_{i,k} = \mu_{i,k}[1 - \frac{p_i(k,0)}{p(k)}] , \tag{13}$$

where $p(k)$ is the marginal steady-state probability that the environment is in phase k. Hence, the arrival rate into node i in phase k is equal to

$$\sigma_{i,k} = \sum_{j=1}^{N} d_{j,k} r_{j,i} , \tag{14}$$

where $r_{j,i}$ is the routing probability from node j to node i.

Now we are in a similar situation to that in the last section. If the network is acyclic, then starting from the nodes whose inputs are entirely external, and using their outputs to calculate the inputs into the downstream nodes, the approximate solution for the whole network is obtained in one sweep. Moreover, a slight simplification of the computation is possible in this case: any node which does not contribute, directly or indirectly, to the traffic at node i may be excluded from the latter's Markovian environment.

If the network has cycles, then an iterative procedure is applied: initially, all inputs rates are assumed to be phase-independent and given by (2). Subsequently, the departure rates computed in the last iteration are used to find the arrival rates for the next one. The full environment $U(T)$ controls all nodes in every iteration.

5 Numerical results

The first experiment is intended merely to illustrate the inadequacy of the reduced work-rate approximation, so that it can be eliminated from further comparisons. The example involves a simple network of two identical nodes in tandem. In figure 1, the approximated and simulated values of the average queue size at node 2 are plotted against the average operative period, $1/\xi$. The ratio η/ξ is kept fixed, so that the service capacities remain constant (each server is operative about 80% of the time). The figure amply confirms the fact that the queue size increases with the lengths of the operative and inoperative periods. Moreover, that increase appears to be almost linear. The reduced work-rate approximation, on the other hand, does not change and therefore becomes progressively worse. In this example, the MMPP and the joint-state approximations are nearly identical and very accurate, while the Poisson approximation is slightly worse.

Next, we examine the quality of the Poisson approximation when the size of the network increases. A number of fully connected networks with 3, 5, 10 and 15 nodes are generated, choosing their parameters at random while making sure that all are stable and subjected to roughly similar loads. In each case, the average response time is estimated by means of the Poisson approximation, and also by simulation; the relative error of the former with respect to the latter is evaluated. The results, displayed in figure 2(a) show that both the average and the variability of the error decrease rapidly with N. In the 10-node networks, all observed errors are less than 15%, while in the 15-node ones they are less than 10%.

Figure 2(b) shows the approximated and simulated average response times in a randomly chosen 17-node network under increasing load (the total external arrival rate is

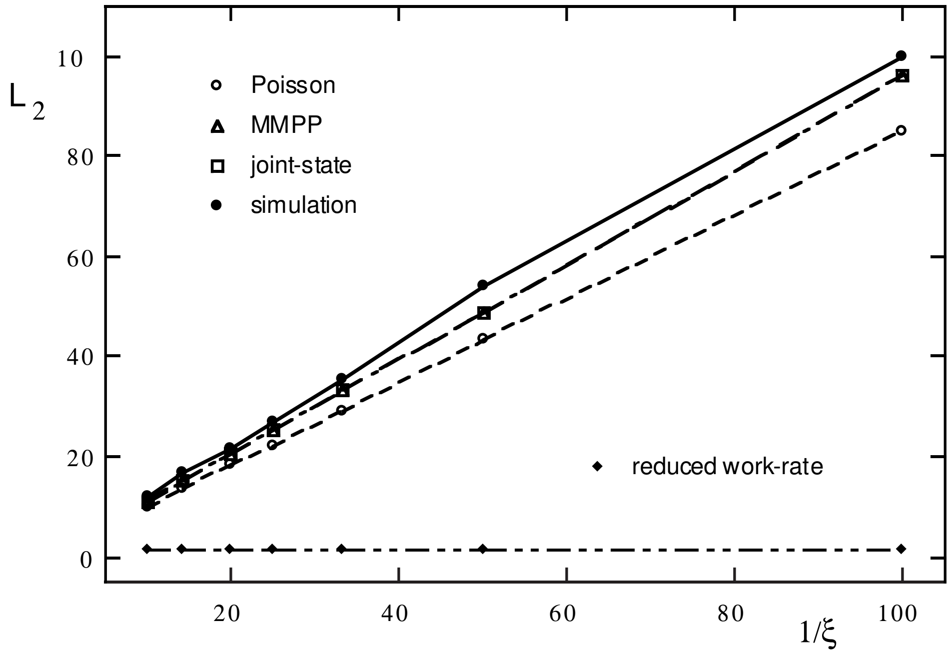

Figure 1: Approximations for a 2-node tandem network
$\sigma = 1$; $\mu_1 = \mu_2 = 2$; $\xi_1 = \xi_2 = \xi$; $\eta_1 = \eta_2 = 5\xi$

varied). All approximated values, including the heavy load ones, are well within the corresponding confidence intervals.

Thus, the Poisson approximation is adequate in networks where each node has a reasonable number of inputs (10 or more). The more complex approximations find their applications in small networks, or networks whose structure is such that the number of inputs into each node is small (e.g., nodes in tandem). Further experimentation with that type of networks has led us to conclude that when they do not have cycles, both the MMPP and the joint-state approximations are quite accurate (relative errors are usually less than 10%). However, when there are cycles, the joint-state approximation is clearly preferable.

Figure 3 provides an illustration of the last remark, in the context of a three-node network with feedbacks from node 3 to node 1. The performance measure is the average response time. It is evaluated by the Poisson, MMPP and joint-state approximations, and also by simulation, for different external arrival rates. In graph 3(a), the servers are operative about 95% of the time, while in 3(b), that fraction is between 60% and 65%. Both graphs show that the Poisson and MMPP approximations strongly overestimate the performance measure. The joint-state approximation is much better; its relative error is on the order of 15% or less.

6 Conclusion

We have made some prograss with a difficult problem. The reduced work-rate approximation can generally be dismissed from consideration. The Poisson approximation is

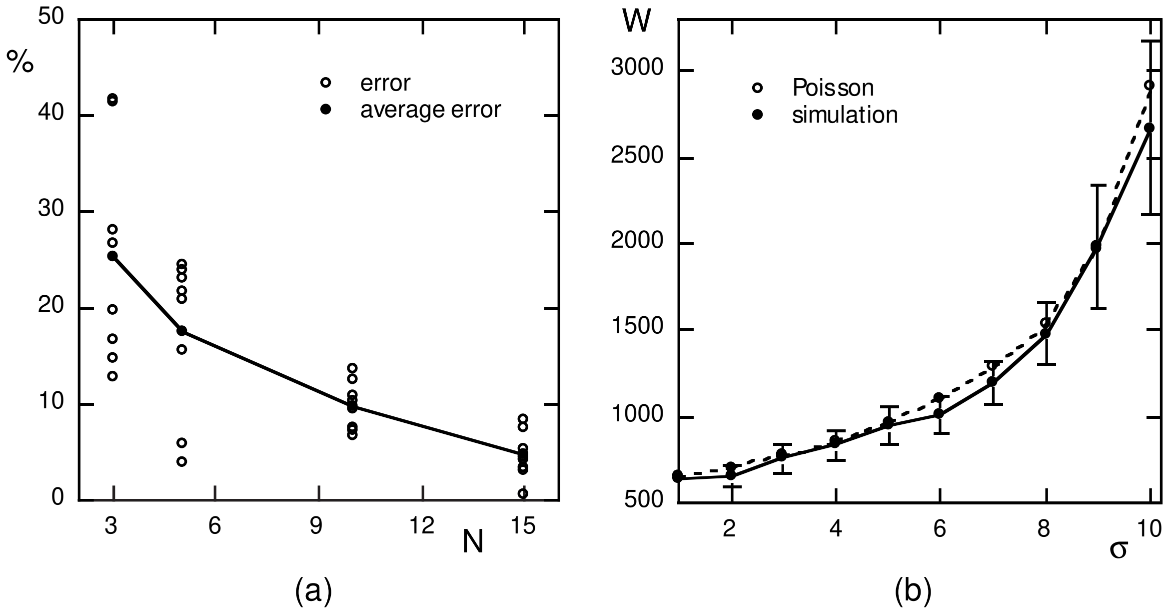

Figure 2: Poisson approximation for average response time *vs* simulation
Graph (a): random networks of different sizes; Graph (b): a 20-node network

acceptable in networks with high degree of merging of arrivals. Where that is not the case, one has to deal with the obstacles posed by the very volatile and state-dependent nature of the traffic between nodes. The latter can be overcome, at least partially, by using non-Poisson approximations for the arrival streams. Of the two suggested here, the joint-state approximation is recommended for networks containing cycles. Both the joint-state and the MMPP approximations appear to be adequate in feedforward networks.

The accuracy of the approximations can probably be improved, at the price of increasing their complexity. For example, while a node is broken, its queue tends to accumulate and there may be a large backlog of jobs at the start of the next operative period. Consequently, for an initial fraction of that period the departure rate is higher than average. That phenomenon can be included into the approximation by adding extra phases to the Markovian environment.

Acknowledgement

This work was carried out in connection with the Basic Research projects QMIPS (Quantitative Methods In Parallel Systems), and PDCS2 (Predictably Dependable Computer Systems), funded by the European Union.

References

[1] B. Avi-Itzhak and P. Naor, "Some Queueing Problems with the Service Station Subject to Breakdowns", Operations Research, 11, pp. 303-320, 1963.

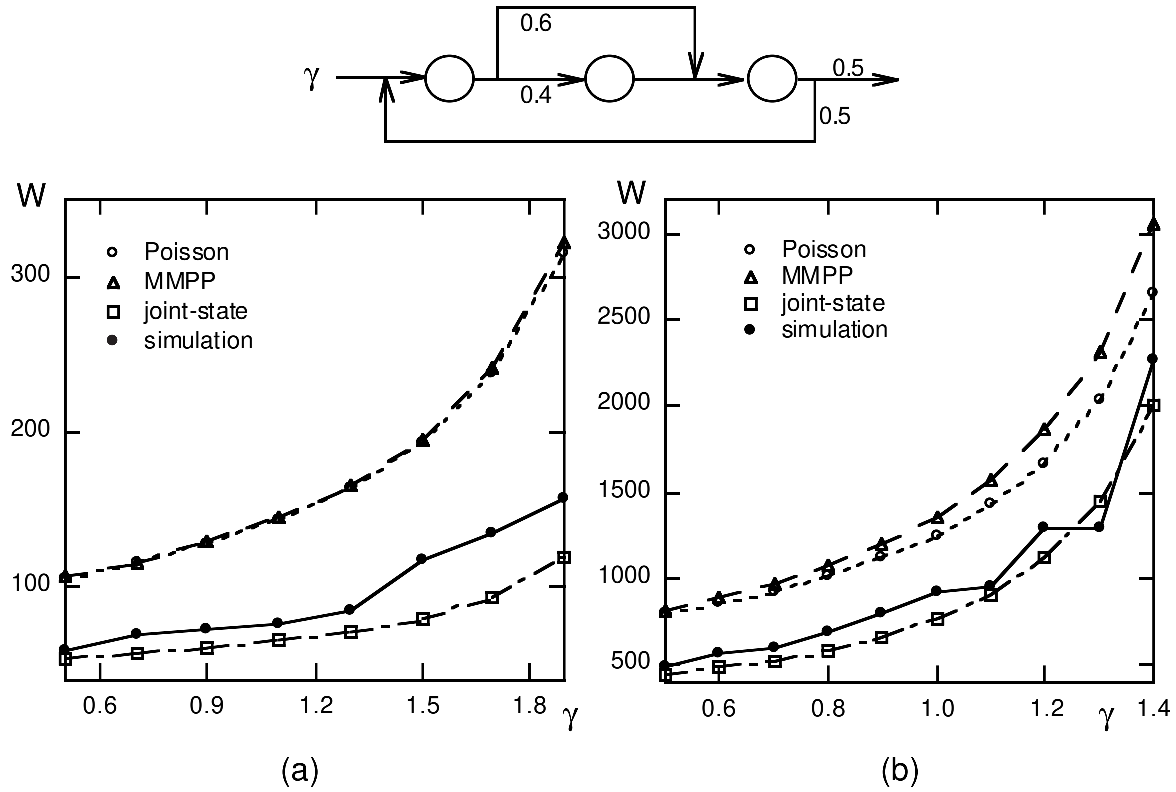

Figure 3: Average response time in a 3-node network with feedbacks
$(\mu_1, \mu_2, \mu_3) = (5, 6, 7)$; $(\eta_1, \eta_2, \eta_3) = (0.002, 0.003, 0.005)$
Graph (a): $(\xi_1, \xi_2, \xi_3) = (0.0001, 0.0002, 0.0003)$; Graph (b):
$(\xi_1, \xi_2, \xi_3) = (0.001, 0.002, 0.003)$

[2] W. Feller, *An Introduction to Probability Theory and its Applications, Vol. 1*, Wiley, 1968.

[3] W. Fischer and K. Meier-Hellstern (1993), The Markov-modulated Poisson proccess, *Performance Evaluation*, **18**, 2, 149-171

[4] O. Idrissi-Kacemi, N. Mikou and S. Saadi, "Two Processors Only Interacting During Breakdown: The Case Where the Load is Not Lost", submitted for publication.

[5] L. Kleinrock, *Queueing Systems, Vol. 1*, Wiley, 1975.

[6] N. Mikou, "A Two-Node Jackson Network Subject to Breakdowns", Stochastic Models, 4, pp. 523-552, 1988.

[7] I. Mitrani, "Networks of Unreliable Computers", in *Computer Architectures and Networks* (eds. E. Gelenbe and R. Mahl), North-Holland, 1974.

[8] I. Mitrani and B. Avi-Itzhak, "A Many-Server Queue with Service Interruptions", Operations Research, 16, pp. 628-638, 1968.

[9] I. Mitrani and R. Chakka (1995), "Spectral Expansion Solution for a Class of Markov Models: Application and Comparison with the Matrix-Geometric Method", Performance Evaluation, 23, 1995.

[10] I. Mitrani and D. Mitra (1992), A Spectral Expansion Method for Random Walks on Semi-Infinite Strips, in *Iterative Methods in Linear Algebra* (Eds R. Beauwens and P. de Groen), North-Holland

[11] I. Mitrani and P.E. Wright (1994), Routing in the Presence of Breakdowns, Performance Evaluation, 20, 151-164

[12] M. Reiser, "Interactive Modelling of Computer Systems", IBM Sys. J., 15, pp. 309-327, 1976.

[13] M.F. Neuts (1981), *Matrix Geometric Solutions in Stochastic Models*, John Hopkins Press

[14] M. Reiser, "A Queueing Network Analysis of Computer Comunication Networks with Window Flow Control", IEEE Trans. Commun., 8, pp. 1199-1209, 1979.

[15] B. Sengupta, "A Queue with Service Interruptions in an Alternating Markovian Environment", Operations Research, 38, pp. 308-318, 1990.

[16] N. Thomas and I. Mitrani, "Routing Among Different Nodes Where Servers Break Down Without Losing Jobs", Procs., IPDS'95, Erlangen, 1995.

[17] B. Vinod and T. Altiok, "Approximating Unreliable Queueing Networks Under the Assumption of Exponentiality", J. Oper. Res. Soc., 37, pp. 309-316, 1986.

[18] H.C. White and L.S. Christie, "Queueing with Preemptive Priorities or with Breakdown", Operations Research, 6, pp. 79-95, 1958.