AN ANALYTICAL APPROACH FOR OPTIMISING THE NUMBER OF REPAIRMEN FOR LARGE SCALE, HOMOGENEOUS, MULTI-SERVER SYSTEMS

Orhan Gemikonakli [(a)], Hadi Sanei [(b)], Enver Ever [(c)], Altan Koçyiğit [(d)]

[(a)] School of Computing Science, Middlesex University, The Burroughs, London NW4 4BT, UK
[(b)] Halcrow Group Limited, TBG, Vineyard House, 44 Brook Green, London W6 7BY, UK
[(c)] School of Engineering, Design and Technology, University of Bradford, Richmond Road, Bradford, BD7 1DP, UK
[(d)] Informatics Institute, Middle East Technical University, 06531, ODTU, Ankara, Turkey

[(a)] o.gemikonakli@mdx.ac.uk, [(b)] saneih@halcrow.com, [(c)] eever@bradford.ac.uk, [(d)] kocyigit@metu.edu.tr

ABSTRACT

Fault tolerant, large scale multi-server systems require an optimum number of repairmen for maximising performability. However, performability evaluation of such systems is difficult due to the state space explosion problem. In this paper, a simple and flexible approximate technique capable of overcoming state space explosion problem in computing the performability of large Markov models is presented. For validation of results, simulation has been used. It is shown that, this approach can handle large state space. The proposed method allows analysing the breakdown repair behaviour of large multi server systems. An optimisation study is presented together with numerical results showing the relationship of the number of servers and the number of repairmen for optimum performance.

Keywords: Large scale multi-server systems, State Explosion problem, Markov Process, Performability and Simulation.

1. INTRODUCTION

The number of repairmen in a fault tolerant multi-server system can affect the system's overall performability significantly. When multi-server systems are considered, some of the servers in the system may be redundant because of poor repair facilities. The availability of the servers highly depends on the ratio of failure and repair rates. The probability of failures linearly increases with the number of servers employed. It is necessary to address the effects of server availability on overall performability of large scale systems. In this paper large scale, homogeneous multi-server systems are considered. The queuing model under study is a homogeneous multi server system with unreliable servers and a common queue. Homogeneous multi server systems with unreliable servers have been considered in (Chakka and Mitrani 1992; Boxma et al. 1994; Chakka et al. 2002; Mitrani 2005). However a solution method to obtain the steady state probabilities of large scale homogeneous multi server systems which is applicable to both loaded systems and systems with relatively lighter loads have not been presented.

In order to handle large scale systems, a new method is presented in (Gemikonakli et al. 2007). Numerical results are computed and validated by simulation but the effects of various numbers of repairmen have not been analysed for large scale multi server systems.

Several analytical methods have been reported for the performability evaluation of multi-server systems. In (Chakka and Mitrani 1994; Chakka et al. 2002) multi-server systems are considered with various repair disciplines (e.g. first broken first repaired, round robin etc.). Markov models are presented for heterogeneous multi-server systems with various repair strategies. Also in (Chakka and Mitrani 1992; Ever et al. 2007) numerical results presented using the spectral expansion method for various multi-server systems with breakdowns and repairs. However, the solutions used in these publications are applicable to small and medium size systems only. Larger networks give rise to state space explosion problem and most analytical solution techniques become inadequate. Mitrani's dominant eigenvalue approach (Mitrani 2005) works well especially for loaded networks due to the fact that dominant eigenvalue better represents loaded networks. This is an improvement of the Spectral Expansion method (Chakka and Mitrani 1992), however, state space is still limited with matrix sizes. Also the dominant eigen value approach may not give good approximations for systems which are not heavily loaded (Mitrani 2005). Furthermore, these works do not address the optimization of the number of repairmen for large systems.

Quasi Birth and Death processes (QBD) (Ciardo and Smirni 1999; Hung and Do 2001; Wallace 1969) have a wide range of applications in Queuing systems. A QBD process can be described as a two-dimensional Markov chain where transitions are only possible between adjacent states of a given model. The popularity of the QBD processes leads to the development of various numerical procedures for their steady-state analysis (Akar and Sohraby 1997; Bini and Meini 1996; Chakka and Mitrani 1992; Haverkort and Ost 1997; Hung and Do 2001; Krieger et al. 1998; Latouch and Ramaswami 1993; Naumov et al. 1996; Neuts 1981). These approaches have received considerable attention. In (Haverkort and Ost 1997) two such approaches are compared. An important difficulty associated with these methods is the state space explosion problem (Haverkort and van Moorsel 1995), which, in case of clusters, limits the analysis and performability evaluation of the number of processors working in parallel. Events such as break-downs, and rebooting/reconfiguration further increase this problem. In today's world, several hundred processors are likely to run in parallel.

Two-dimensional models with multiple components are effectively used for the modelling of queuing systems with multiple queues and/or multiple servers. The functional equations arising in the analysis of such processes usually present significant analytical difficulties (Boxma et al. 1994). These numerical difficulties are frequently caused by large number of steady states. In other words, difficulty caused by the rapid increase in the size of the state space of the underlying Markov process gives rise to the state space explosion problem.

When two dimensional models are considered, the cause of large number of state spaces can be either large (or infinite) number of jobs or large number of operative states (e.g. number of servers). For multi-dimensional models, where one component is finite there are good analytic-algorithmic methods, such as Matrix-geometric solution (Neuts 1981), and Spectral Expansion method (Chakka and Mitrani 1992). These methods can be used to solve the state explosion problem caused by large or infinite number of jobs (El-Rayes et al. 1999). Although systems with unbounded queuing capacities can be handled by Spectral Expansion and Matrix-geometric solution methods, as the number of operative states increases they become computationally expensive. The numerical complexity of the solutions depends on the number of operative states (Mitrani 2005). That number determines the size of the matrix R used in Matrix-geometric method, and the number of eigenvalues and eigenvectors involved in Spectral Expansion method. In both solution techniques the size of the matrices used depends on the number of operative states of the system considered. As the size of the matrices increases the computational requirements increase significantly. Because of the large size, ill-conditioned (Mitrani 2005) matrices numerical problems occur. Also, the numerical stability of the solution gets affected especially when heavily loaded systems are considered.

In this paper, a simple and flexible approximate technique capable of overcoming state space explosion problem in computing the performability of large Markov models is presented. The technique is used to analyse the effects of the number of repairmen working on the repair of failed processors, on overall performability of large scale multi server systems. For validation of the approximate solution method employed, simulation has been used. It is shown that, this approach can handle large state space. The proposed method allows analysing the breakdown repair behaviour of large multi server systems. A Poisson stream of arrivals together with exponentially distributed service times, time between failures, and repair times are assumed. An optimisation study is presented together with numerical results showing the relationship of the number of servers and number of repairmen for optimum performance.

2. APPROXIMATE SOLUTION FOR THE STEADY STATE PROBABILITIES OF LARGE MARKOV MODELS

The proposed approximate solution method uses an iterative technique to calculate the steady state probabilities of a two-dimensional Markov chain. Consider a K-server, homogeneous system with a queuing capacity L, number of repairmen R, mean arrival rate σ, mean service rate μ, mean break-down rate ξ, and mean repair rate η. On the lattice, while one-step downward transition rates (j to j-1) can be represented in terms of μ, i, and j, one step upward transition rates (j to $j+1$) depend on σ. The lateral transitions from i to i+1, and i to i-1 can be expressed in terms of $(K-i)$, R, η, and $i\xi$ respectively. Consider a continuous time, two-dimensional Markov process on a finite lattice strip. The Markov Process can be defined as $X=\{I_n, J_n;\ n=0, 1, \ldots\}$ with a state space of $(\{0, 1, 2, \ldots, K\}\text{x}\{0, 1, 2, \ldots, L\})$ where, for a multiprocessor system, K and L represent the number of processors and the queuing capacity respectively. Then, i=0, 1, 2, …, K, and j=0, 1, 2, …, L can be used to represent all possible states, (i,j), on the lattice strip. Hence, the steady state probabilities can be denoted as $p_{i,j}$. Figure 1 shows the states of such a system, where, $r_i = \min(R, K-i)$ and $k_i = \min(j, i)$.

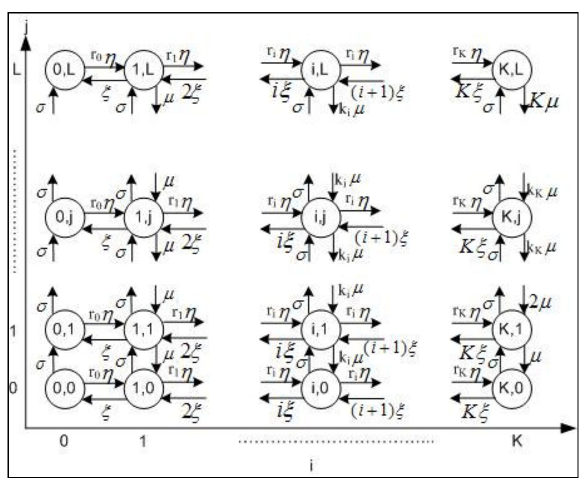

Figure 1: The states of the system under study

Let's define *column* vectors $\mathbf{v}_i$ as follows:

$$\mathbf{v}_i = \{p_{i,0}, p_{i,1}, p_{i,2}, ..., p_{i,L}\} \text{ for all } i=0,1,.., K \quad (1)$$

where i represents the number of operative servers, and $p_{i,j}$ are the steady state probabilities. For each i, $\mathbf{v}_i$ can be calculated using product form formulae. From Figure 1, for each column (i.e. i=0, 1, …, K), a set of balance equations can be obtained. An example of these balance equations is given in Eq. (2) for (i, j), where, $0 < i < K$, and $0 < j < L$.

$$\begin{aligned} &\sigma p_{i,j-1} + r_{i-1}\eta p_{i-1,j} + (i+1)\xi p_{i+1,j} + k_{i,j+1}\mu p_{i,j+1} \\ &= (k_{i,j}\mu + \sigma + i\xi + r_i\eta)\, p_{i,j} \end{aligned} \quad (2)$$

By using a set of balance equations which can be derived from (2), together with $\sum_{i=0}^{K}\sum_{j=0}^{L} p_{i,j} = 1$, (3) is obtained.

$$p_{i,j} = \begin{cases} \left(\rho^j p_{i,0}\right)/\, j!, & 0 < j \le i \\ \left(\rho^j p_{i,0}\right)/\left(i!\, i^{j-i}\right), & i < j \le L \end{cases} \quad (3)$$

where, $\rho = \sigma/\mu$.

Next, let's introduce S_i representing the sum of all state probabilities for each operative state i; $S_i = \sum_{j=0}^{L} p_{i,j}$. Since each of the operative states are reached through server break-downs or repairs (horizontal transitions), the following general equation can be obtained:

$$r_{i-1}\eta S_{i-1} + (i+1)\xi S_{i+1} = (r_i\eta + i\xi) S_i .$$

Hence Eq. (4) can be obtained as following:

$$S_i = S_0 \frac{(\eta/\xi)^i}{i!} \prod_{m=0}^{i-1} r_m \quad (4)$$

where

$$S_0 = \left[1 + \sum_{k=1}^{K}\left(\frac{(\eta/\xi)^k}{k!}\prod_{m=0}^{k-1} r_m\right)\right]^{-1} \quad (5)$$

Furthermore, Eq. (5) can be used to calculate the probability that all servers are idle for a given i. For this, first using (3), it can be shown that;

$$S_i = \left(\sum_{j=0}^{i}\frac{\rho^j}{j!} + \sum_{j=i+1}^{L}\frac{\rho^j}{i!\, i^{j-i}}\right) p_{i,0} \text{, and hence,}$$

$$p_{i,0} = S_i\left(\sum_{l=0}^{i}\left(\frac{\rho^l}{l!}\right) + \frac{i^{L-i}\rho^{i+1} - \rho^{L+1}}{i!\, i^{L-i}(i-\rho)}\right)^{-1} \quad (6)$$

Remaining state probabilities can then be calculated using (3). Furthermore, $p_{0,0}$ can be obtained as follows:

$$p_{0,0} = \xi(R\eta + \sigma)^{-1} p_{1,0} \quad (7)$$

(3)-(7) define all approximate steady state probabilities, $\mathbf{v}_i$, for i operative servers. These equations can now be used in calculating approximate $p_{i,j}$. These probabilities are approximate because lateral transitions have not been taken into account yet. Once the approximate steady state probabilities are calculated, the balance equations given in (7) – (15) can be used to calculate the steady state probabilities more accurately. It is assumed that the number of repairmen is less then number of servers, and in case of breakdowns each server is repaired by a single repairman. Equations (8) – (15) can be given as follows:

$$p_{0,j} = \left(\xi p_{1,j} + \sigma p_{0,j-1}\right)(R\eta + \sigma)^{-1}, \quad (8)$$
for all j, $0<j<L$

$$p_{0,L} = (R\eta)^{-1}\left(\xi p_{1,L} + \sigma p_{0,j-1}\right) \quad (9)$$

$$p_{i,0} = \frac{\left((i+1)\xi p_{i+1,0} + \min(K-i+1,R)\eta p_{i-1,0} + \mu p_{i,1}\right)}{i\xi + \sigma + \min(K-i,R)\eta} \quad (10)$$
for all i, $0<i<K$

$$p_{i,j} = \frac{\begin{pmatrix}(i+1)\xi p_{i+1,j} + \sigma p_{i,j-1} + \\ \min(K-i+1,R)\eta p_{i-1,j} + \min(i,j+1)\mu p_{i,j+1}\end{pmatrix}}{\min(i,j)\mu + i\xi + \sigma + \min(K-i,R)\eta} \quad (11)$$
for all i, j, $0<j<L$, $0<i<K$

$$p_{i,L} = \frac{\left((i+1)\xi p_{i+1,L} + \sigma p_{i,L-1} + \min(K-i+1,R)\eta p_{i-1,L}\right)}{\min(i,L)\mu + i\xi + \min(K-i,R)\eta} \quad (12)$$
for all i, $0<i<K$

$$p_{K,0} = \left(\eta p_{K-1,0} + \mu p_{K,1}\right)(K\xi + \sigma)^{-1} \quad (13)$$

$$p_{K,j} = \frac{\left(\sigma p_{K,j-1} + \eta p_{K-1,j} + \min(K,j+1)\mu p_{K,j+1}\right)}{\min(K,j)\mu + K\xi + \sigma} \quad (14)$$
for all j, $0<j<L$

$$p_{K,L} = \frac{\left(\sigma p_{K,L-1} + \eta p_{K-1,L}\right)}{K(\mu + \xi)} \quad (15)$$

Since the steady state probabilities for state (i, j) are initially calculated independent of states $(i$-1, $j)$ and $(i$+1, $j)$, it is important to use a technique to compensate for the unaccounted effects of the latter two states on state (i, j). This can be achieved through the use of the balance equations (7) – (15) together with an iterative process. We have K+1 vectors, if $t_{x,y}$ represents transitions from vector x to vector y where $x \neq y$, then, the transitions can be summarized as follows:

$$[\mathbf{v}_0]\begin{matrix}\rightarrow t_{0,1}\rightarrow\\ \leftarrow t_{1,0}\leftarrow\end{matrix}[\mathbf{v}_1]..[\mathbf{v}_{i-1}]\begin{matrix}\rightarrow t_{i-1,i}\rightarrow\\ \leftarrow t_{i,i-1}\leftarrow\end{matrix}[\mathbf{v}_i]..[\mathbf{v}_{K-1}]\begin{matrix}\rightarrow t_{K-1,K}\rightarrow\\ \leftarrow t_{K,K-1}\leftarrow\end{matrix}[\mathbf{v}_K]$$

Then, an iterative procedure can be followed to accurately calculate $p_{i,j}$. The procedure can be given as follows:

(i). First, $\mathbf{v}_i$ are calculated for i = 0, 1, 2, …, K using (1) , (3), (4), (5), and (6).
(ii). Knowing approximate $p_{i,j}$, the balance equations given in (7) - (15) are used to calculate the correct steady state probabilities.
(iii). The sum of all probabilities is calculated for the queuing system considered.
(iv). Steps (ii) and (iii) are repeated until the sum of probabilities converges to one. Once the correct state probabilities are obtained, various performability measures can be calculated.

The iterative method presented in this study is much faster than simulation and more importantly, unlike most analytical techniques, it can handle large numbers of servers working in parallel without giving rise to state space explosion problem for most practical systems. Furthermore, the method provides accurate results for both heavily loaded networks and networks with relatively lighter loads.

3. NUMERICAL RESULTS AND DISCUSSIONS

Numerical results are presented in this section in order to show the effects of the number of repairmen on systems performability. For all computations parameters are taken as, μ=2, ξ=0.01, η=0.5, and L=1000.

Figures 2-7 show the MQL performance of K-server systems as a function of σ, R and K. It is clear from Figure 2 that when only one repairman is present, increasing the number of servers will not result in an increase in MQL performance. As the number of repairmen is increased, the effects of the number of processors become more significant. Considering that, failure rate can be expressed as $i\xi$, while mean repair rate is min(K-i, R)η, the relationship between these two parameters is important. For system efficiency, $R\eta > K\xi$ should be satisfied. As an example, for R=2, and K=64, we obtain 1 > 0.64. Results also show that there is an upper limit for R as well. For a 64-server system, increasing R from one to two significantly improves systems MQL performance and R>2 has no significant effect on system performance. Figure 8 shows the mean queue length of a 512-server system as a function of R and σ. Here, R plays a significant role, the significance of R will only diminish when the network is extremely loaded or R is large (e.g. R >10).

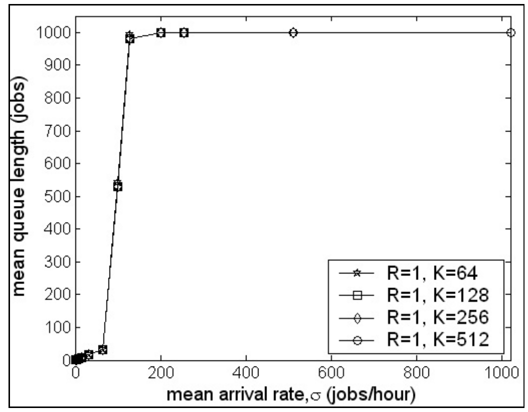

Figure 2: MQL Using a Single Repairman

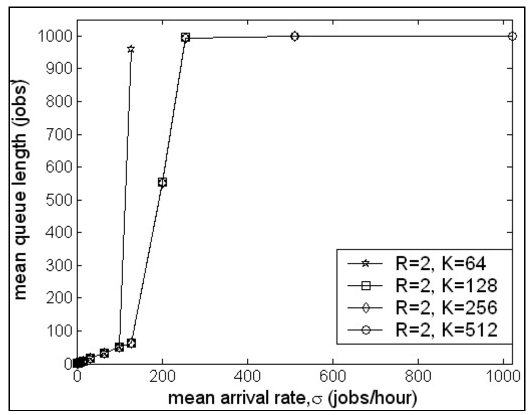

Figure 3: MQL for R=2

Further computations have been carried out to demonstrate the effects of R on system performance for K=512.

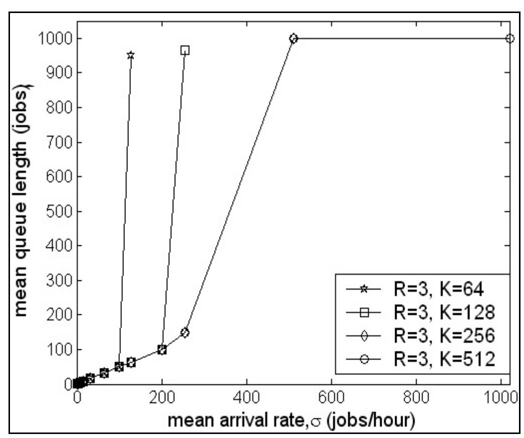

Figure 4: MQL for R=3

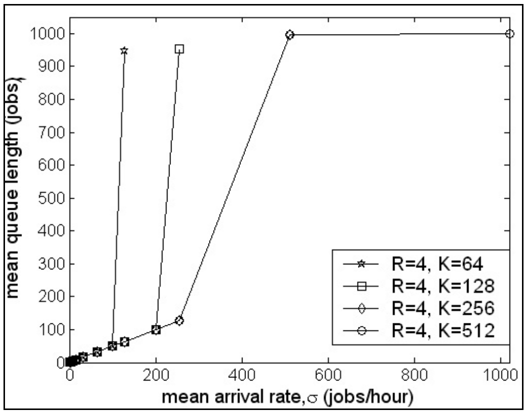

Figure 5: MQL for R=4

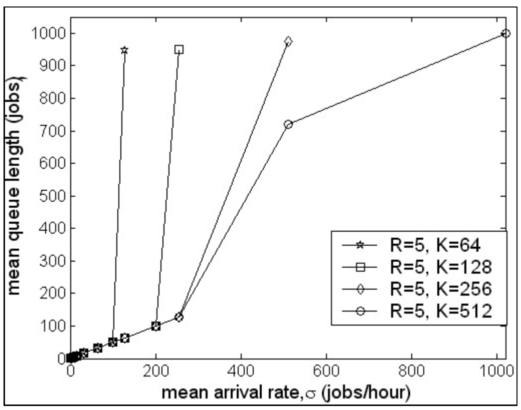

Figure 6: MQL for R=5

Figure 9 shows the effect of R on the throughput of the system, while Figure 10 shows the effect of R on the mean response time of such a system. Results clearly show that there is a threshold for R which depends on not only K but also σ. Increasing R beyond that threshold does not have much significance on mean system response time. However, when the system throughput is considered, the significance of R increases as the system's load increases. Again, large R does not seem to have much significance on system performability.

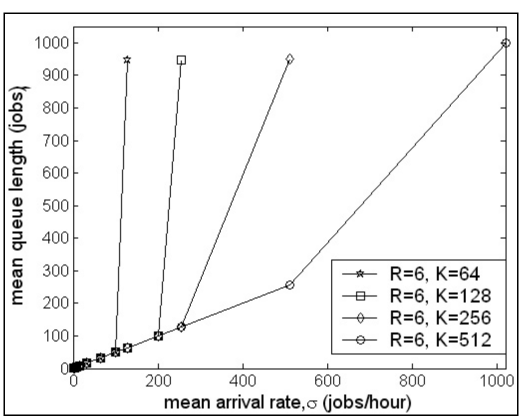

Figure 7: MQL for R=6

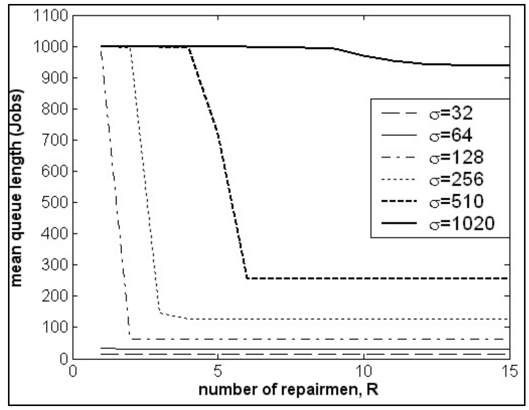

Figure 8: MQL as a Function of R and σ, for K=512

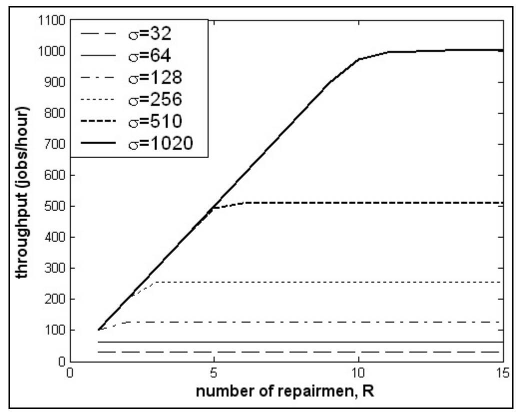

Figure 9: Throughput for K=512 and Various R

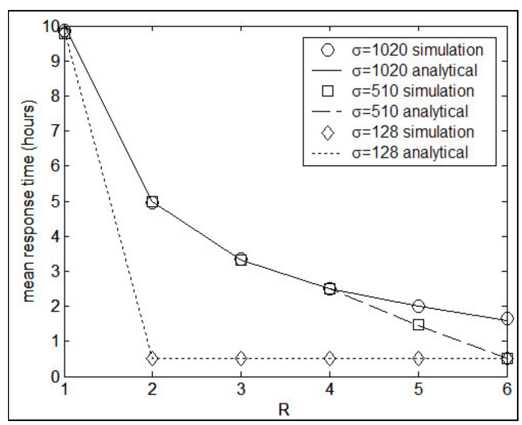

Figure 10: Mean Response Time as a function of R, for K=512

4. MODEL VALIDATION

In this section, we examine the validity of the approximate analytical model proposed in this paper by computer simulations. For this purpose, we have developed an event driven simulation system in the C++ programming language. The simulator developed simulates the actual multi-processor system under consideration. In order to ensure that the developed simulator correctly simulates the intended system, several tests have been carried out. Some of these tests are as follows:

- Unit tests and code walkthroughs have been conducted during the software development.
- Job arrival, job service, processor failure, and processor repair events and the system state for several simulation runs are traced for consistency checking purposes.
- For various multi-server system configurations with different parameters, the operative state probabilities obtained from the simulation are compared to the values obtained from closed form expressions (equations 4 and 5).
- M/M/K/L queuing systems are simulated (by disabling failures) with different K, L, arrival rate, and service rate parameters and the results obtained from the simulation are compared to the values obtained from analytical solutions (using well-known closed form expressions - see equation 3).

For the validation of the proposed method, we have carried out extensive simulation experiments and reached very close performance conclusions for various scenarios. For simulations we used a 5% confidence interval with a confidence level of 95%.

Figures 10-11 and Tables 1-4 present MQL results obtained using both the proposed analytical method and simulation.

Table 1: A Comparison of MQL Results From the Proposed Analytical Approach and Simulation for K=512

Analytical						
σ	R=1	R=2	R=3	R=4	R=5	R=6
64	32.947	32	32	32	32	32
128	982.425	64.030	64	64	64	64
256	999.345	994.127	147.429	128	128	128
510	999.755	999.384	998.537	995.915	719.123	257.359
1020	999.891	999.755	999.582	999.351	999.031	998
Simulation						
σ	R=1	R=2	R=3	R=4	R=5	R=6
64	32.234	31.971	32.007	32.017	32.035	32.0078
128	983.018	64.032	63.999	64.011	63.989	64.0056
256	997.652	991.711	150.787	128.107	127.925	128.022
510	999.011	998.529	997.493	994.611	729.035	257.7967
1020	999.851	999.702	999.55	999.3	998.986	998.4893

Table 2: A Comparison of MQL Results From the Proposed Analytical Approach and Simulation for K=256

Analytical					
σ	R=1	R=2	R=3	R=4	R=5
64	32.9465	32	32	32	32
128	982.425	64.0302	64	64	64
256	999.345	994.127	147.429	128	128
510	999.755	999.348	998.537	995.944	973.966
Simulation					
σ	R=1	R=2	R=3	R=4	R=5
64	33.0037	31.9974	32.0007	32.0014	31.9592
128	977.6245	64.0356	64.0141	63.9736	63.9754
256	998.3492	991.6139	146.1522	128.0276	128.0276
510	999.6659	999.2278	998.3227	995.7597	973.394

Table 3: A Comparison of MQL Results From the Proposed Analytical Approach and Simulation for K=128

Analytical				
σ	R=1	R=2	R=3	R=4
64	32.9465	32	32	32
128	982.425	64.0303	64	64
200	998.956	553.004	100.232	100.051
256	999.345	995.419	965.345	952.776
Simulation				
σ	R=1	R=2	R=3	R=4
64	33.3109	32.0099	32.0271	32.0043
128	978.3478	64.0106	63.9932	64.0448
200	998.3102	551.8926	100.1956	100.027
256	999.1635	995.2569	964.1577	953.292

Table 4: A Comparison of MQL Results From the Proposed Analytical Approach and Simulation for K=64

Analytical				
σ	R=1	R=2	R=3	R=4
64	32.9694	32	32	32
100	542.46	50.416	50.2514	50.2358
128	994.453	959.3	949.691	947.646
Simulation				
σ	R=1	R=2	R=3	R=4
64	32.3929	32	32.035	32.0099
100	548.9022	50.5001	50.2368	50.2445
128	993.862	958.5472	951.5519	945.927

The results presented in Tables 1-4 clearly show that the discrepancy of analytical results and simulation results are less than 5%.

In addition to MQL, throughput and response times have also been computed using the analytical approach and compared to simulation results. Response time results are shown in Figure 10.

In (Gemikonakli et al. 2007) computational efficiencies of the proposed system and the Spectral Expansion method are compared. In that study K=49 has been considered; this is due to state space limitations of the Spectral Expansion method. Computational performances of the two methods are close for a range of utilisation values, however, for loaded networks, the proposed technique is slower than the Spectral Expansion method, but still much faster than simulation.

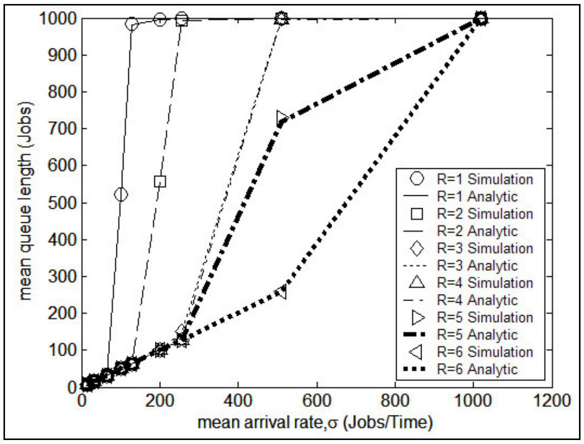

Figure 11: MQL as a function of R and σ and K=512

5. CONCLUSION AND FUTURE DEVELOPMENTS

In this paper, an analytical approach is used to calculate the optimum number of repairmen for large scale, fault tolerant, homogeneous, multi-server systems, to achieve the best performability possible. Inter-arrival times, service times, mean time between failures, and repair times have been assumed to be exponentially distributed. In order to avoid the state space explosion problem inherent to most analytical approaches, an approximate solution technique is proposed for the performance evaluation of such systems. Results obtained using this approach has been confirmed to be accurate using simulation results. The proposed technique has been used to compute the mean queue length, mean response time and throughput of various systems for different numbers of repairmen. The approach lends itself as a powerful technique in evaluating and optimising the performance of large scale networks under various scenarios. The case studies concerned show the strong relationship between number of processors and number of repairmen for operating a large scale multi-server system efficiently. Some threshold values have been obtained for specific case studies. For small and medium scale systems $R\eta/K\xi >> 1$, and hence specific relationship between R and K is not that evident. This case further highlights the importance of the approach used. The proposed method is a flexible one and can be extended to the case of heterogeneous systems, systems with rebooting and reconfiguration delays, highly available systems, Beowulf systems and many other practical, fault tolerant multi-server systems.

REFERENCES

Akar, N., Sohraby, K. 1997. Finite and Infinite QBD Chains: A Simple and Unifying Algorithmic Approach. *Proceedings of IEEE INFOCOM*, pp. 1105-1113. April 1997, (Kobe, Japan).

Bini, D., Meini, B. 1996. On the solution of a non-linear equation arising in queueing problems, SIAM J. Matrix. *Anal*. Appl. 17, 906-926.

Boxma, O.J., Koole, G.M., Liu, Z. 1994. Queueing-theoretic solution methods for models of parallel distributed systems. *Performance Evaluation of Parallel and Distributed Systems, eds*., pp. 1-24. (CWI Tract 105, Amsterdam)

Chakka, R., Mitrani, I. 1992. A numerical solution method for multiprocessor systems with general breakdowns and repairs. *Proceedings of the 6th International Conference Modelling Techniques and Tools*, pp. 289-304. September 1992, (Edinburgh, UK)

Chakka, R., Mitrani, I. 1994. Heterogeneous Multiprocessor Systems with Breakdowns: Performance and Optimal Repair Strategies. *Theoretical Computer Science 125(1)*, pp. 91-109.

Chakka, R., Gemikonakli, O., Basappa, P. 2002. Modelling multiserver systems with time or operation dependent breakdowns, alternate repair strategies, reconfiguration and rebooting delays. *Proceeding of the SPECTS 2002*, pp. 266-277, July 2002 (San Diego, CA, USA)

Ciardo, G., Smirni, E. 1999. ETAQA: An Efficient Technique for the Analysis of QBD-processes by Aggregation. *Performance Evaluation 36-37*, pp. 71-93.

El-Rayes, A., Kwiatkowska, M., Norman, G. 1999. Solving Infinite Stochastic Process Algebra Models Through Matrix-Geometric Methods. *Proceeding of the 7th Process Algebras and Performance Modelling Workshop, PAPM'99*, pp. 41-62. September 1999, (University of Zaragoza),

Gemikonakli, O., Sanei, H., Ever, E. 2007. Approximate Solution for the Performability of Markovian Queuing Networks with a Large Number of Servers. *Proceeding of the 5th International Workshop on Signal Processing for Wireless Communication, SPWC 2007*. June **2007,** (King's College, University of London).

Haverkort, B.R., Ost, A. 1997. Steady-State Analysis of Infinite Stochastic Petri Nets: Comparing the Spectral Expansion and the Matrix-Geometric Method. *Proceeding of the Seventh International Workshop on Petri Nets and Performance Models*, pp. 36-45.June 1997 (Saint Malo, France).

Haverkort, B.R., van Moorsel, A.P.A. 1995. Using the Probabilistic Evaluation Tool for the Analytical Solution of Large Markov Models. *Proceeding of the Sixth International Workshop on Petri Nets and Performance Models (PNPM '95)*, pp. 206-207, October 1995, (Durham, NC).

Hung, T.T., Do, T.V. 2001. Computational Aspects for Steady State Analysis of QBD Processes, Department of Telecommunication. *Periodica, Polytechnica Ser. El. Eng*. Vol.44, No.2, pp. 179-200.

Krieger, U.R., Naumov, V., Wagner, D. 1998. Analysis of a Finite FIFO Buffer in an Advanced Packet-Switched Network. *IEICE Trans. Commun. E81-B*, pp. 937-947.

Latouch, G., Ramaswami, V. 1993. A Logarithmic Reduction Algorithm for quasi birth-death processes. *Journal of Applied Probability, 39*, pp. 650-674.

Mitrani, I. 2005. Approximate solutions for heavily loaded Markov-modulated queues. *Perform. Eval. 62(1-4)*, pp. 117-131.

Naumov, V., Krieger, U.R., Wagner, D. 1996. Analysis of a Multi-Server Delay-Loss System with a General Markovian Arrival Process. *In S.R. Chakravarthy and A.S. Alfa, editors*, Matrix-analytic methods in stochastic models., *volume 183 of Lecture Note in Pure and Applied Mathematics, Marcel Dekker*, September 1996.

Neuts, M.F. 1981. Matrix Geometric Solutions in Stochastic Model. *Johns Hopkins University Press, Baltimore*.

Wallace, V.L. 1969. The Solution of Quasi Birth and Death Processes Arising From Multiple Access Computer Systems. Thesis PhD, University of Michigan.

AUTHORS BIOGRAPHY

ORHAN GEMIKONAKLI obtained his first degree in electrical engineering from Eastern Mediterranean University in Cyprus, in 1984. He then continued his studies at King's College, University of London where he obtained his MSc in Digital Electronics, Computers and Communications and PhD in Telecommunications in 1985 and 1990 respectively. He worked there as a post-doctoral Research Associate for a couple of years before moving in 1990 to Middlesex University, London where he is now the Head of Department, Computer Communications. He is a member of the IEE, a member of the IEEE and a Chartered Engineer. His research interest is in network security and performability modelling of complex systems. His e-mail address is: o.gemikonakli@mdx.ac.uk and his Web-page can be found at http://www.cs.mdx.ac.uk/staffpages/orhan/.

HADI SANEI obtained his BEng in Computer Hardware Engineering from his native country of Iran. Later he obtained his MSc in Computer Networks in July 2006 from Middlesex University in London. He also is a professional member of BCS (British Computer Society). During his master degree he demonstrated his research interest in performance and reliability modelling and simulation of complex systems. Currently he is a Systems Engineer / RAMS analyst at Halcrow Group Ltd (UK offices). Also he is a PhD candidate for a joint research project at UCLSE (UCL Systems Engineering Dept.), University of London and Halcrow Group Ltd. His email address is saneih@halcrow.com.

ENVER EVER obtained his BSc. degree from the Department of Computer Engineering, Eastern Mediterranean University, Cyprus in 2002, his MSc in Computer Networks from Middlesex University, London in 2003, and his PhD degree in Performance and Reliability modelling from Middlesex University, London in 2008. Currently he is a researcher in The University of Bradford's Engineering, Design and Technology Department working on development of various state-of-the-art computer systems in Markov Modelling, Risk, Reliability, and Safety analysis. His research interest is in networks, performance and reliability modelling of complex systems. His e-mail address is eever@bradford.ac.uk.

ALTAN KOCYIGIT obtained BSc, MSc, and PhD degrees from the Middle East Technical University Department of Electrical and Electronics Engineering Department in 1993, 1997, and 2001, respectively. He has been with the Middle East Technical University Informatics Institute since 2002. His research interests include computer networking, software engineering, and parallel/distributed processing. His e-mail address is kocyigit@metu.edu.tr.