# A Recursive Solution of a Queueing Model for a Multi-terminal System Subject to Breakdowns

J. Sztrik and T. Gál

*Department of Mathematics, University of Debrecen, Debrecen, H-4010, Hungary*

Received 12 December 1988
Revised 30 October 1989

The queueing system to be analysed is a model of a multi-terminal system subject to random breakdowns. All random variables involved here are independent and exponentially distributed. Although the stochastic process describing the system's behaviour is a Markov chain, the number of states becomes very large. The main contribution of this paper is a recursive computational approach to solve the steady-state equations concerning the problem. In equilibrium, the main performance measures of the system, such as the mean number of jobs residing at the CPU, the mean number of functional terminals, the average number of busy servers, the expected response times of jobs, and utilizations are obtained. Finally, some numerical results illustrate the problem in question.

*Keywords:* Multi-terminal System, Queueing Model, Recursive Computational Approach, Performance Measures.

## 1. Introduction

This paper deals with the analysis of a queueing system which may be used as a model of a real life system consisting of $n$ active terminals connected with a Central Processor Unit (CPU). The users at the active terminals have exponentially distributed think times with rate $\lambda$ and they generate jobs for the CPU with processing times being exponentially distributed with rate $\mu$. The service rule at the CPU may be any work-conserving discipline, that is, it does not affect the total time spent in service of any jobs, for

**J. Sztrik** is a research fellow at the Department of Mathematics, University of Debrecen, Debrecen, Hungary. He received the B.Sc. degree in 1978, the M.Sc. degree in 1980 both in probability theory and mathematical statistics from the University of Debrecen. He received the Ph.D. degree in probability theory and mathematical statistics in 1989 from the Kiev State University, Kiev, USSR. His research interests are in the field of production systems modelling and analysis, queueing theory, reliability theory, and computer science. His papers in these topics have appeared in journals such as *European Journal of Operational Research, Journal of Operational Research Society of England, Acta Cybernetica, Publicationes Mathematicae, Problems of Control and Information Theory, Serdica, Journal of Information Processing and Cybernetics, Computers and Mathematics with Applications, Performance Evaluation, Cybernetics, Probability Theory and Mathematical Statistics.* Dr. Sztrik is a member of J. Bolyai Mathematical Society, Budapest, and J. Neumann Computer Society, Budapest.

**T. Gál** is a programmer at COMPORGAN, Debrecen, Hungary. He received his B.Sc. degree in computer science in 1988 from the University of Debrecen, Debrecen, Hungary. His current research interest are computer performance evaluation, and queueing theory. He is a member of J. Bolyai Mathematical Society, Budapest, and J. Neumann Computer Society, Budapest.

North-Holland
Performance Evaluation 11 (1990) 1–7

example, FIFO or Processor Sharing. Furthermore, it is assumed that each user generates only one job at a time and he waits at the CPU before it starts thinking again. Let us suppose that the operational system is subject to random breakdowns stopping the service at the terminals and at the CPU. The failure-free operation times of the system are exponentially distributed random variables with rate $\alpha$. The restoration times of the system are assumed to be exponentially distributed random variables with rate $\beta$. The busy terminals are also subject to random breakdowns not affecting the system's operation. The failure-free operation times of the busy terminals are supposed to be exponentially distributed random variables with rate $\gamma$. The repair times of the terminals are exponentially distributed random variables with rate $\tau$. The breakdowns are serviced by $r$ repair crews providing preemptive priority to the system's failure. All random variables involved here are assumed to be independent of each other.

As can easily be seen, this model is a generalization of the classical 'machine interference problem' discussed, among others, in [1,3]. In recent years finite-source models in different forms have been effectively used, for example, for mathematical description of multiprogrammed computer systems (see [2,4,5,7,8,10]).

The purpose of the present paper is to analyse a multi-terminal system just described. The main contribution of this paper is a recursive computational approach to solve the steady-state equations concerning the problem. In equilibrium, the main performance measures of the system, such as the mean number of jobs residing at the CPU, the mean number of functional terminals, the average number of busy servers, the expected response times of jobs, and utilizations are obtained. Finally, some numerical results illustrate the problem in question.

## 2. The model and a computational approach

Let us introduce the following random variables:

$$X(t) = \begin{cases} 1 & \text{if the operational system is failed at time } t, \\ 0 & \text{otherwise,} \end{cases}$$

$Y(t)$ = the number of jobs residing at the CPU at time $t$,

$Z(t)$ = the number of failed terminals at time $t$.

Clearly, the process

$$M(t) = (X(t),\ Y(t),\ Z(t))$$

is a three-dimensional Markov chain with state space

$$((i,\ k,\ s);\ 0 \leqslant i \leqslant 1,\ 0 \leqslant k \leqslant n,\ 0 \leqslant s \leqslant n-k)$$

of dimension $(n+1)(n+2)$. Let us denote the steady-state distribution by

$$p(i,\ k,\ s) = \lim_{t \to \infty} P(X(t) = i,\ Y(t) = k,\ Z(t) = s),$$

which exists and is unique (see [1,3]). As usual, using the notion of probability flow, the global balance equations for $p(i,\ k,\ s)$ are

$$(\alpha + n\lambda + n\gamma)p(0,\ 0,\ 0) = \beta p(1,\ 0,\ 0) + \tau p(0,\ 0,\ 1,) + \mu p(0,\ 1,\ 0), \tag{1}$$

$$(\alpha + (n-s)\lambda + s\tau + (n-s)\gamma)p(0,\ 0,\ s) = \beta p(1,\ 0,\ s) + (n-s+1)\gamma p(0,\ 0,\ s-1) + (s+1)\tau p(0,\ 0,\ s+1) + \mu p(0,\ 1,\ s), \tag{2}$$

for $1 \leqslant s \leqslant r-1$,

$$(\alpha + (n-s)\lambda + r\tau + (n-s)\gamma)p(0,\ 0,\ s) = \beta p(1,\ 0,\ s) + (n-s+1)\gamma p(0,\ 0,\ s-1) + r\tau p(0,\ 0,\ s+1) + \mu p(0,\ 1,\ s), \tag{3}$$

for $r \leqslant s \leqslant n-1$,

$$(\alpha + r\tau)p(0, 0, n) = \beta p(1, 0, n) + \gamma p(0, 0, n-1), \tag{4}$$

$$(\alpha + \mu + (n-k)\lambda + (n-k)\gamma)p(0, k, 0)$$
$$= \beta p(1, k, 0) + \tau p(0, k, 1) + (n-k+1)\lambda p(0, k-1, 0) + \mu p(0, k+1, 0), \tag{5}$$

for $1 \leqslant k \leqslant n-1$,

$$(\alpha + \mu)p(0, n, 0) = \beta p(1, n, 0) + \lambda p(0, n-1, 0), \tag{6}$$

$$(\alpha + \mu + (n-k-s)\lambda + r\tau + (n-k-s)\gamma)p(0, k, s)$$
$$= \beta p(1, k, s) + \mu p(0, k+1, s) + (n-k-s+1)\gamma p(0, k, s-1)$$
$$+ r\tau p(0, k, s+1) + (n-k-s+1)\lambda p(0, k-1, s), \tag{7}$$

for $1 \leqslant k \leqslant n-r$, $r \leqslant s \leqslant n-k$,

$$(\alpha + \mu + (n-k-s)\lambda + (n-k-s)\gamma + s\tau)p(0, k, s)$$
$$= \beta p(1, k, s) + \mu p(0, k+1, s) + (n-k-s+1)\gamma p(0, k, s-1)$$
$$+(s+1)\tau p(0, k, s+1) + (n-k-s+1)\lambda p(0, k-1, s), \tag{8}$$

for $n-r+1 \leqslant k \leqslant n-1$, $1 \leqslant s \leqslant n-k$,

$$(\beta + (r-1)\tau)p(1, k, s) = \alpha p(0, k, s) + (r-1)\tau p(1, k, s+1), \tag{9}$$

for $0 \leqslant k \leqslant n-r+1$, $r-1 \leqslant s \leqslant n-k$,

$$(\beta + s\tau)p(1, k, s) = \alpha p(0, k, s) + (s+1)\tau p(1, k, s+1), \tag{10}$$

for $n-r+2 \leqslant k \leqslant n-1$, $0 \leqslant s \leqslant r-2$,

$$\beta p(1, n, 0) = \alpha p(0, n, 0). \tag{11}$$

In principle, this system of linear equations can easily be solved by standard computational methods. However, we must take into consideration that the unknowns are probabilities and therefore in the case of a large state space the round-off errors may have considerable effect on them (see [3,5,6,11]).

In the following an efficient recursive computational approach is given for determining the stationary probabilities $p(i, k, s)$. After dividing both sides by coefficient of the unknowns $p(i, k, s)$ on the left-hand side and introducing vectors

$$\boldsymbol{Y}^{(k)} = (p(0, k, 0), \ldots, p(0, k, n-k))^{\mathrm{T}},$$
$$\boldsymbol{Z}^{(k)} = (p(1, k, 0), \ldots, p(1, k, n-k))^{\mathrm{T}}$$

of dimension $n-k+1$, $k = 0, 1, \ldots, n$, the above equations can be written in matrix form as

$$\boldsymbol{Y}^{(0)} = B_0\boldsymbol{Y}^{(0)} + C_0\boldsymbol{Y}^{(1)} + D_0\boldsymbol{Z}^{(0)}, \tag{12}$$

$$\boldsymbol{Y}^{(k)} = A_k\boldsymbol{Y}^{(k-1)} + B_k\boldsymbol{Y}^{(k)} + C_k\boldsymbol{Y}^{(k+1)} + D_k\boldsymbol{Z}^{(k)}, \quad 1 \leqslant k \leqslant n-1, \tag{13}$$

$$\boldsymbol{Y}^{(n)} = A_n\boldsymbol{Y}^{(n-1)} + D_n\boldsymbol{Z}^{(n)}, \tag{14}$$

$$\boldsymbol{Z}^{(k)} = F_k\boldsymbol{Y}^{(k)} + H_k\boldsymbol{Z}^{(k)}, \quad 0 \leqslant k \leqslant n-1, \tag{15}$$

$$\boldsymbol{Z}^{(n)} = F_n\boldsymbol{Y}^{(n)}. \tag{16}$$

It is quite easy to see that the involved matrices are sparse (band diagonal) and are of the following dimensions:

- $A_k$: $(n-k+1) \times (n-k+2)$, $1 \leqslant k \leqslant n$,
- $B_k$, $H_k$: $(n-k+1) \times (n-k+1)$, $0 \leqslant k \leqslant n-1$,
- $C_k$: $(n-k+1) \times (n-k)$, $1 \leqslant k \leqslant n$,
- $D_k$, $F_k$: $(n-k+1) \times (n-k+1)$, $0 \leqslant k < n$,

The system of eqs. (12)–(16) can be considered as the steady-state global balance equations for a multi-dimensional birth-death process in which vectors $\boldsymbol{Y}^{(k)}$, $\boldsymbol{Z}^{(k)}$ play the role of probabilities and matrices $A_k$, $B_k$, $C_k$, $D_k$, $F_k$ represent the birth and death intensities.

To obtain the solution of this system we have:

**Theorem.** *The solution of eqs.* (12)–(16) *can be determined recursively by*

$$\boldsymbol{Y}^{(k)}=L_k\boldsymbol{Y}^{(k-1)},\quad 1\leqslant k\leqslant n,\qquad \boldsymbol{Z}^{(k)}=R_k\boldsymbol{Y}^{(k)},\quad 0\leqslant k\leqslant n,$$

*where*

$$R_n=F_n,\qquad R_k=(I-H_k)^{-1}F_k,\quad 0\leqslant k\leqslant n, \tag{17}$$

$$L_n=(I-D_nR_n)^{-1}A_n,\qquad L_k=(I-B_k-C_kL_{k+1}-D_kR_k)^{-1}A_k,\quad 1\leqslant k\leqslant n-1. \tag{18}$$

*The initial vector* $\boldsymbol{Y}^{(0)}$ *is calculated from*

$$(I-B_0-D_0R_0-C_0L_1)\boldsymbol{Y}^{(0)}=\mathbf{0}$$

*up to a multiplicative constant to be found from the norming condition*

$$\sum_{i=0}^{1}\sum_{k=0}^{n}\sum_{s=0}^{n-k}p(i,k,s)=1.$$

**Proof.** By the help of eqs. (15)–(16) we get $\boldsymbol{Z}^{(n)}=F_n\boldsymbol{Y}^{(n)}$ and $(I-H_k)\boldsymbol{Z}^{(k)}=F_k\boldsymbol{Y}^{(k)}$, so (17) holds.

From (14) we have

$$(I-D_nF_n)\boldsymbol{Y}^{(n)}=A_n\boldsymbol{Y}^{(n-1)},$$

and assuming $\boldsymbol{Y}^{(k+1)}=L_{k+1}\boldsymbol{Y}^{(k)}$ from (13) we get

$$(I-B_k-C_kL_{k+1}-D_kR_k)\boldsymbol{Y}^{(k)}=A_k\boldsymbol{Y}^{(k-1)},\quad 1\leqslant k\leqslant n-1,$$

which yields (18).

Finally, (12) gives

$$(I-B_0-C_0L_1-D_0R_1)\boldsymbol{Y}^{(0)}=\mathbf{0}.\qquad\square$$

This system of linear equations possesses a unique solution for any $\hat{p}(0, 0, 0)$. After calculating vectors $\boldsymbol{Y}^{(k)}$, $\boldsymbol{Z}_k^{(k)}$ and summarizing their all components the steady-state probabilities $p(i, k, s)$ can be obtained after dividing each component by the sum.

The advantage of the algorithm is two-fold. On the one hand, although the number of the involved matrices are large enough, the procedure needs a relatively small storage requirement, since the matrices are sparse. On the other hand it reduces the round-off errors, hence in a given cycle only $n-k+1$ unknowns are involved instead of $(n+1)(n+2)$ ones.

It should be noted that this kind of recursive computational approach can be applied to derive the main operational characteristics of a 'machine interference problem with server's breakdowns', see e.g. [9].

## 3. The main performance measures and numerical results

(i) Mean number of jobs residing at the CPU

$$\bar{n}_j=\sum_{k=1}^{n}k\left(\sum_{i=0}^{1}\sum_{s=0}^{n-k}p(i,k,s)\right),$$

(ii) Mean number of functional terminals

$$\bar{n}_f = n - \sum_{s=1}^{n} s\left(\sum_{i=0}^{1}\sum_{k=0}^{n-s} p(i, k, s)\right),$$

(iii) Mean number of busy terminals

$$\bar{n}_b = \sum_{k=0}^{n}\sum_{s=0}^{n-k} (n-k-s)\, p(0, k, s),$$

(iv) Mean number of busy repairmen

$$\bar{n}_r = \sum_{s=1}^{r} s\left(\sum_{k=0}^{n-s} p(0, k, s)\right) + r\sum_{s=r+1}^{n}\sum_{k=0}^{n-s} p(0, k, s)$$
$$+ \sum_{s=1}^{r} s\left(\sum_{k=0}^{n-s+1} p(1, k, s-1)\right) + r\sum_{s=r+1}^{n}\sum_{k=0}^{n-s+1} p(1, k, s-1),$$

(v) Expected response times of jobs

$$\bar{T} = \bar{n}_j/(\lambda \bar{n}_b),$$

(vi) Utilization of CPU

$$U_{CPU} = \sum_{k=1}^{n}\sum_{s=0}^{n-k} p(0, k, s),$$

(vii) Utilization of terminals

$$U_t = \bar{n}_b/n,$$

(viii) Utilization of repairmen

$$U_r = \bar{n}_r/r.$$

The algorithm generating these characteristics was implemented in Turbo Pascal on an IBM PC/XT at the Institute of Mathematics, University of Debrecen, Hungary.

In the case of $\alpha = \gamma \simeq 0$ we have the M/M/1 'machine interference problem'. Using the example in [1, p. 433],

$$\lambda = 0.025, \qquad \mu = 0.25, \qquad n = 6,$$

supplemented by parameters

$$\gamma = 10^{-12}, \qquad \tau = 999, \qquad \alpha = 10^{-12}, \qquad \beta = 999, \qquad r = 1,$$

we get the following test result:

| | by formula | by the algorithm |
|---|---|---|
| $\bar{n}$ | 0.84 | 0.84 |
| $\bar{T}$ | 6.51 | 6.55 |
| $U_{CPU}$ | 52% | 52% |
| $U_t$ | 86% | 86% |

In the following, some sample results are shown to illustrate the effect of different parameters on the basic performance measures.

*Case* 1.

Input parameters:

$$\lambda = 5, \quad \mu = 50, \quad \gamma = 0.5, \quad \tau = 1.0, \quad \alpha = 0.01, \quad \beta = 1.5, \quad n = 8.$$

We should like to investigate the effect of the number of repairmen on the characteristics $\bar{n}_f$, $\bar{n}_j$, $\bar{T}$, $U_{CPU}$, $U_t$. The results are collected in Table 1.

Table 1

| $r$ | $\bar{n}_f$ | $\bar{n}_j$ | $\bar{T}$ | $U_{CPU}$ | $U_t$ |
|---|---|---|---|---|---|
| 2 | 3.8895 | 0.4901 | 0.0291 | 0.34 | 0.42 |
| 3 | 4.8936 | 0.6577 | 0.0313 | 0.42 | 0.53 |
| 4 | 5.2466 | 0.7178 | 0.0320 | 0.45 | 0.56 |
| 5 | 5.3371 | 0.7326 | 0.0321 | 0.46 | 0.57 |
| 6 | 5.3539 | 0.7352 | 0.0321 | 0.46 | 0.57 |
| 7 | 5.3558 | 0.7355 | 0.0321 | 0.46 | 0.57 |
| 8 | 5.3559 | 0.7355 | 0.0321 | 0.46 | 0.58 |

Table 2

| $\lambda$ | $\bar{n}_f$ | $\bar{n}_j$ | $\bar{T}$ | $U_{CPU}$ | $U_t$ |
|---|---|---|---|---|---|
| 6 | 5.5807 | 1.0393 | 0.0385 | 0.54 | 0.45 |
| 7 | 5.6099 | 1.2720 | 0.0422 | 0.60 | 0.43 |
| 7.5 | 5.6280 | 1.3912 | 0.0422 | 0.63 | 0.42 |
| 8 | 5.6486 | 1.5117 | 0.0461 | 0.65 | 0.41 |

As can be observed, all measures are increasing functions of $r$, and we have almost the same values for $6 \leqslant r \leqslant 8$. This behaviour of the system can be explained as follows. The greater the number of repairmen, the faster the repair of the failed terminals. Consequently, there are more functional terminals and they generate more jobs. Furthermore, if certain repairmen can service the failed terminals fast enough, then we do not need more repairmen since they have no significant effect on the characteristics.

*Case* 2.

Input parameters:

$$\mu = 50, \quad \gamma = 0.5, \quad \tau = 1.0, \quad \alpha = 0.01, \quad \beta = 1.5, \quad n = 10, \quad r = 3.$$

The performance measures are considered as function of the think time rate $\lambda$. The measures are shown in Table 2.

We can see that the measures $\bar{n}_f$, $\bar{n}_j$, $\bar{T}$, $U_{CPU}$ are increasing functions of $\lambda$, but $U_t$ is a decreasing function of $\lambda$. This is quite understandable since the greater the rate $\lambda$ the greater the number of jobs residing at the CPU and so the resulting characteristics follow from this fact.

*Case* 3.

Input parameters:

$$\lambda = 5, \quad \gamma = 0.5, \quad \tau = 1.0, \quad \alpha = 0.01, \quad \beta = 1.5, \quad n = 5, \quad r = 3.$$

The characteristics are studied as the function of the service rate $\mu$. The results can be seen in Table 3.

We can observe that $U_t$ is almost the same, while the other characteristics are decreasing function of $\mu$. Since the service rate $\mu$ is relatively great comparing to the other rates, its small changes do not influence $U_t$ at least within two digits. Furthermore, the greater the rate $\mu$ the less the number of jobs residing at the CPU and this fact explains the behaviour of the other characteristics.

Table 3

| $\mu$ | $\bar{n}_f$ | $\bar{n}_j$ | $\bar{T}$ | $U_{CPU}$ | $U_t$ |
|---|---|---|---|---|---|
| 50 | 3.3144 | 0.3761 | 0.0258 | 0.29 | 0.58 |
| 51 | 3.3135 | 0.3680 | 0.0252 | 0.29 | 0.58 |
| 52 | 3.3136 | 0.3602 | 0.0246 | 0.28 | 0.59 |
| 53 | 3.3118 | 0.3528 | 0.0240 | 0.28 | 0.59 |
| 54 | 3.3110 | 0.3456 | 0.0235 | 0.27 | 0.59 |
| 55 | 3.3103 | 0.3387 | 0.0230 | 0.27 | 0.59 |

Table 4

| $\gamma$ | $\bar{n}_f$ | $\bar{n}_j$ | $\bar{T}$ | $U_{CPU}$ | $U_t$ |
|---|---|---|---|---|---|
| 0.5 | 3.3144 | 0.3761 | 0.0258 | 0.29 | 0.58 |
| 0.6 | 3.0930 | 0.3457 | 0.0254 | 0.27 | 0.54 |
| 0.7 | 2.8954 | 0.3193 | 0.0250 | 0.25 | 0.51 |
| 0.8 | 2.7183 | 0.2961 | 0.0247 | 0.24 | 0.48 |
| 0.9 | 2.5591 | 0.2758 | 0.0244 | 0.23 | 0.45 |
| 1.0 | 2.4156 | 0.2578 | 0.0241 | 0.21 | 0.43 |

*Case* 4.
Input parameters:

$$\lambda=5,\quad \mu=50,\quad \gamma=0.5,\quad \tau=1.0,\quad \alpha=0.01,\quad \beta=1.5,\quad n=5,\quad r=3.$$

We illustrate the effect of the failure rate $\gamma$ on the performance measures collected in Table 4.

We illustrate that all measures are decreasing functions of $\gamma$, as was expected. Since the greater the failure rate $\gamma$ the less the number of busy terminals, the behaviour of the characteristics follows.

## 4. Conclusion

This paper is concerned with a queueing model of a multiterminal system subject to breakdowns. The main contribution is a recursive computational approach to solve the steady-state balance equations related to the problem. In equilibrium, the main performance measures of the system, such as the mean number of jobs residing at the CPU, the mean number of functional terminals, the average number of busy servers, the expected response times of jobs, and utilizations are obtained. Finally, some numerical results illustrate the effect of different parameters on the performance measures.

## Acknowledgment

We are very grateful to the referees for their useful comments and suggestions which improved the presentation of the paper. Special thanks are due to Professor Hideaki Takagi for helpful discussions.

## References

[1] W. Feller, An Introduction to Probability Theory and its Applications, Vol. I. (Wiley-Chapman, New York, 1950).

[2] E. Gelenbe and I. Mitrani, Analysis and Synthesis of Computer Systems (Academic Press, London, 1980).

[3] D. Gross and C.M. Harris, Fundamentals of Queueing Theory (Wiley and Sons, New York, 2nd ed., 1985).

[4] L. Kleinrock, Queueing Systems, Vol. II., Computer Applications (Wiley-Interscience, New York, 1976).

[5] S.S. Lavenberg, Ed., Computer Performance Modeling Handbook (Academic Press, New York, 1983).

[6] P.M. Snyder and W.J. Stewart, Explicit and iterative approaches to solving queueing models, *Oper. Res.* **33** (1985) 183–202.

[7] J. Sztrik, A queueing model for multiprogrammed computer systems with different I/O times, *Acta Cybernet.* **7** (1984) 127–135.

[8] J. Sztrik, A probability model for priority processor-shared multiprogrammed computer systems, *Acta Cybernet.* **7** (1986) 329–340.

[9] J. Sztrik and A. Pósafalvi, On the heterogeneous machine interference with limited server's availability, *Eur. J. Oper. Res.* **28** (1987) 321–328.

[10] H. Takagi, Queueing Analysis of Polling Models, *ACM Comput. Surveys* **20** (1988) 5–28.

[11] H.C. Tijms, Stochastical Modelling and Analysis, A Computational Approach (Wiley and Sons, New York, 1986).