Yugoslav Journal of Operations Research
8 (1998), Number 2, 315-321

ANALYSIS OF A TWO-LEVEL QUEUING PRODUCTION SYSTEM WITH FINITE DEDICATED BUFFERS

George TSIOTRAS

Department of Business Administration
University of Macedonia, Greece

Abstract: This paper analyzes a two-level queuing production system with finite dedicated buffers under exponential assumptions. We consider n types of arriving jobs which upon completion of service at the first-level workstation proceed to the second level and join their dedicated workstations that are laid out in parallel. Blocking occurs when the flow of job i through the first-level queue is momentarily stopped owing to the capacity limitation of its second-level dedicated queue having been reached. Approximate results are derived for the blocking probabilities and the mean effective service time in the first-level workstation.

Keywords: Queuing networks, finite buffers, blocking, production systems.

1. INTRODUCTION

Production systems consist of a number of interconnected stations at which operations are performed on workpieces in order to convert the inputs to outputs. Queuing networks have enjoyed great popularity as models of production systems [2].

A queuing network is a multiple-node system in which a job requires service at more than one workstation (node). When limitations are imposed on the length of the queue that can be formed in front of the nodes then the interesting physical phenomenon of blocking is encountered. Blocking occurs when the flow of jobs through the queue is momentarily stopped owing to the capacity limitation of another queue having been reached.

While there is extensive literature dealing with the study of queuing networks with blocking [1, 3, 4, 6, 7, 8], a common characteristic of all studies is the assumption of single-class jobs. Limited results have been reported for queuing networks with blocking and multiple job classes [5]. It is the purpose of this paper to analyze such a two-level queuing production network with finite buffers.

2. THE MODEL

Consider the situation where there are n types of jobs, which arrive according to independent Poisson processes with respective rates λ_i $(i = 1,2,\ldots,n)$, to a two-level production system (Fig. 1). The workstation at level one has an unlimited waiting space and a single machine. The service times are exponentially distributed with parameter μ_{oi} for type i jobs.

Level two consists of n workstations working in parallel. Each workstation has a finite buffer of size M_i (i.e. the capacity of each station is $M_i + 1$) and a single machine.

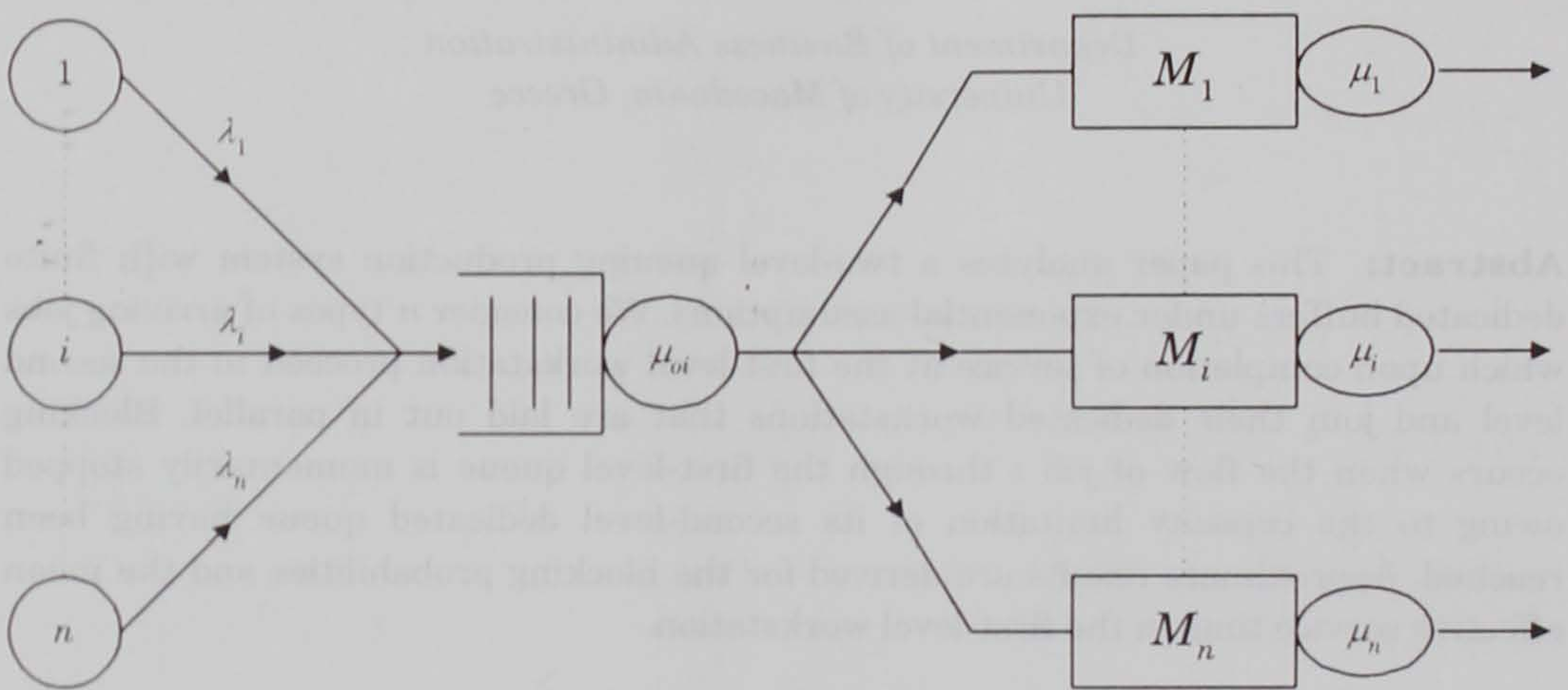

Figure 1.

Upon completion of service at level one workstation, a type i job joins its dedicated destination workstation i to continue its processing, provided that there is at least one available space there at that time. At the ith workstation of level two, the job of type i receives exponential service with rate μ_i and departs from the production system.

However, if upon completion of service at level one workstation, a type i job finds its destination workstation in full capacity, then it becomes blocked. During this time, the level one machine remains idle and it cannot provide service to any other jobs of any type that might be waiting in the queue.

The machine becomes unblocked if the blocked job's dedicated workstation at level two gains a free space (a. departure occurs from that finite queue).

3. APPROXIMATE ANALYSIS

The input process is Poisson with rate $\lambda = \sum_{i=1}^{n} \lambda_i$, which follows since the superposition of n independent Poisson processes is itself a Poisson process whose rate is the sum of the rates of the component processes.

Let S be the time elapsed from the beginning of service at the level one workstation until the machine becomes again available to serve the next job. This is called the effective service time at the level one workstation and is important in the assessment of the delay inflicted by blocking. In this section we shall derive expressions for the mean and the variance of S in terms of the blocking probabilities.

For $i = 1,2,\ldots,n$ define:

S_{oi} = service time of a type i job at level one workstation

S_i^e = effective service time of a type i job at level one station; and

S_i = service time of a type i job at its dedicated destination workstation i.

β_i = steady-state probability that a type i arrival becomes blocked upon the completion of its service at level one workstation.

The effective service time of a type i job at the level one workstation can be expressed as follows (Fig. 2).

$$S_i^e = \begin{cases} S_{oi} + S_i & \text{with probability } \beta_i \\ S_{oi} & \text{with probability } 1-\beta_i \end{cases}$$

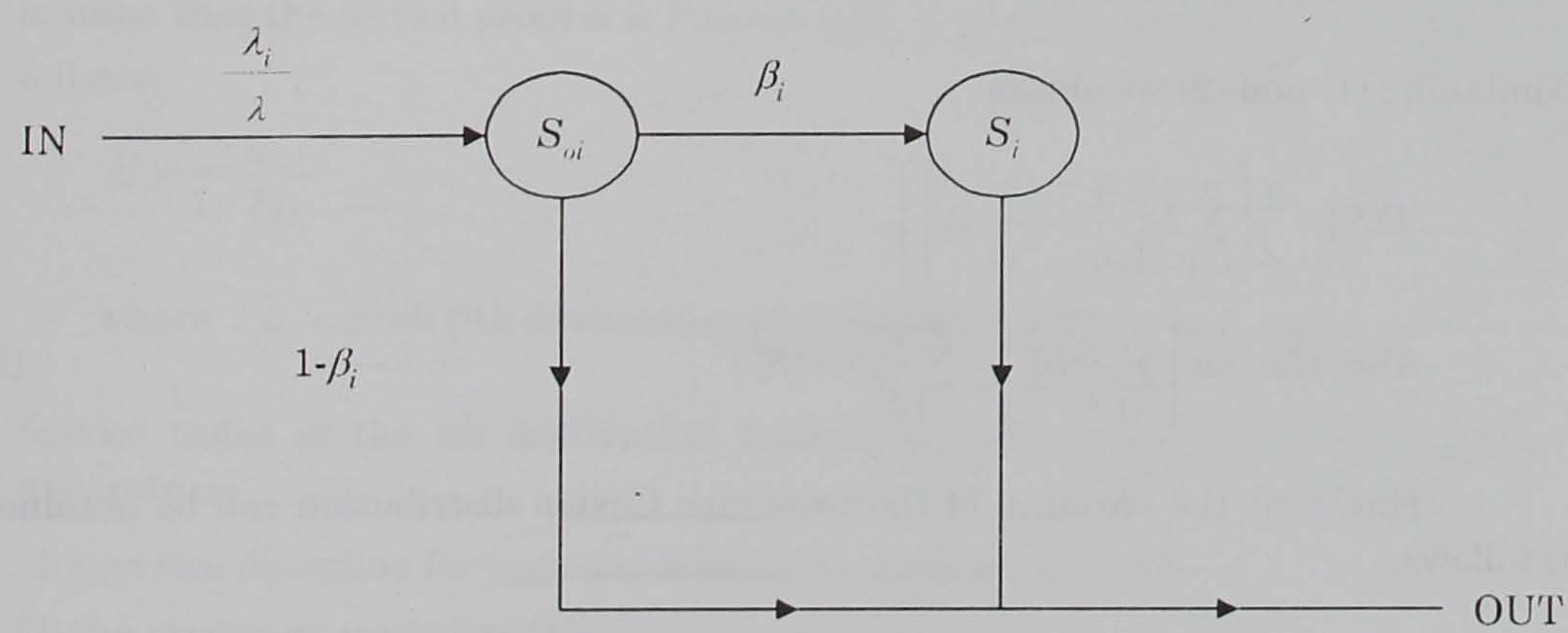

Figure 2.

The effective service time at the level one workstation is:

$$S = S_i^e \text{ with probability } \frac{\lambda_i}{\lambda},\ i = 1.2.\ldots n$$

The mean effective service time at the level one workstation is:

$$E[S] = \frac{1}{\lambda}\left\{\sum_{i=1}^{n} \lambda_i E[S_i^e]\right\} \tag{2}$$

The probability density function of S_i^e (Coxian distribution) is:

$$b_i(t) = c_{oi}\mu_{oi}e^{-\mu_{oi}t} + c_{1i}\mu_i e^{-\mu_i t} \text{ for } t > 0 \text{ and } \mu_{oi} \neq \mu_i$$

$$c_{oi} = \frac{\mu_{oi}(1-\beta_i) - \mu_i}{\mu_{oi} - \mu_i} \text{ and } c_{1i} = 1 - c_{oi}$$

The mean of the two-stage Coxian distribution is:

$$E = [S_i^e] = \int_0^\infty t b_i(t)dt = \int_0^\infty t[c_{oi}\mu_{oi}e^{-\mu_{oi}t} + c_{1i}\mu_i e^{-\mu_i t}]dt =$$

$$= c_{oi}\mu_{oi}\int_0^\infty te^{-\mu_{oi}t}dt + c_{1i}\mu_i\int_0^\infty te^{-\mu_i t}dt$$

$$= \frac{c_{oi}}{\mu_{oi}} + \frac{c_{1i}}{\mu_i} = \frac{c_{oi}}{\mu_{oi}} + \frac{1-c_{oi}}{\mu_i}$$

Finally,

$$E[S_i^e] = \frac{1}{\mu_{oi}} + \frac{1}{\mu_i}\beta_i \tag{2}$$

Combining (1) and (2) we obtain:

$$E[S] = \frac{1}{\lambda}\left[\sum_{i=1}^{n} \lambda_i \left\{\frac{1}{\mu_{oi}} + \frac{1}{\mu_i}\beta_i\right\}\right]$$

Also, the $\operatorname{var}\left[\sum_{i=1}^{n} \frac{\lambda_i}{\lambda} S_i^e\right] = \sum_{i=1}^{n} \frac{\lambda_i^2}{\lambda^2} \operatorname{var}[S_i^e]$ (3)

However, the variance of the two-stage Coxian distribution can be obtained as follows:

$$E[(S_i^e)^2] = \int_0^\infty t^2[c_{oi}\mu_{oi}e^{-\mu_{oi}t} + c_i\mu_i e^{-\mu_i t}]dt = \frac{2c_{oi}}{\mu_{oi}^2} + \frac{2c_i}{\mu_i^2}$$

or

$$E[(S_i^e)^2] = \frac{2}{\mu_{oi}^2} + \frac{2\beta_i}{\mu_{oi}\mu_i} + \frac{2\beta_i}{\mu_i^2}$$

and

$$\mathrm{var}[S_i^e] = \frac{1}{\mu_{oi}^2} + \frac{2\beta_i}{\mu_i^2} - \frac{\beta_i^2}{\mu_i^2} \tag{4}$$

Combining (3) and (4) we obtain:

$$\mathrm{var}\left[\sum_{i=1}^{n} \frac{\lambda_i}{\lambda} S_i^e\right] = \frac{1}{\lambda^2}\left[\sum_{i=1}^{n} \lambda_i^2 \left(\frac{1}{\mu_{oi}^2} + \frac{2\beta_i}{\mu_i^2} - \frac{\beta_i^2}{\mu_i^2}\right)\right]$$

The approximate analysis performed in this section decomposes the two-level queuing network into two individual queuing systems as follows:

Level one system:

- The input process is Poisson with rate λ_i for type i jobs.
- The service time of a type i job will be its effective service time, S_i^e
- The service discipline is FCFS
- All the processes are independent.

Level two system:

- Although the arrivals to the ith destination workstation are not Poisson, we assume that the arrival process is Poisson with revised arrival rate λ_i', defined as follows:

$$\lambda_i' = \frac{\lambda_i}{1 - P_{M_i}} \tag{5}$$

where P_{M_i} = Prob (ith destination workstation is full), $i = 1, 2, \ldots, n$.

- Service times at the ith destination workstation are exponentially distributed with rate μ_i
- The service discipline for each destination workstation is FCFS
- All the processes are independent.

Then for the ith destination workstation we have:

$$P_{M_i} = \frac{\left[\frac{\lambda_i'}{\mu_i}\right]^{M_i}\left[1-\frac{\lambda_i'}{\mu_i}\right]}{1-\left[\frac{\lambda_i'}{\mu_i}\right]^{M_i+1}}, \quad i = 1,2,\dots,n \tag{6}$$

However, since $P_{M_i} = \beta_i$ from (5) and (6) we obtain:

$$\beta_i = \frac{\left[\frac{\lambda_i}{\mu_i(1-\beta_i)}\right]^{M_i}\left[1-\frac{\lambda_i}{\mu_i(1-\beta_i)}\right]}{1-\left[\frac{\lambda_i}{\mu_i(1-\beta_i)}\right]^{M_i+1}}, \quad i = 1,2,\dots,n$$

The above equation can be easily solved numerically for β_i, using for example Newton's method as follows:

Let

$$f(\beta_i) = \beta_i - \beta_i\left[\frac{\lambda_i}{\mu_i(1-\beta_i)}\right]^{M_i+1} - \left[\frac{\lambda_i}{\mu_i(1-\beta_i)}\right]^{M_i}\cdot\left[1-\frac{\lambda_i}{\mu_i(1-\beta_i)}\right]$$

Start with $\beta_i^{(0)} = \alpha$ where a small, $0 < \alpha < 1$

$$\beta_i^{(n)} = \beta_i^{(n-1)} - \frac{f(\beta_i^{(n-1)})}{f'(\beta_i^{(n-1)})}$$

with

$$f'(\beta_i) = 1-\left[\frac{\lambda_i}{\mu_i(1-\beta_i)}\right]^{M_i+1} - \beta_i\left[\frac{\lambda_i^{M_i+1}(M_i+1)}{\mu_i^{M_i+1}(1-\beta_i)^{M_i+2}}\right] -$$

$$-\left[\frac{\lambda_i}{\mu_i(1-\beta_i)}\right]^{M_i+1}\cdot\left[\frac{-\lambda_i}{\mu_i(1-\beta_i)^2}\right] - \left[1-\frac{\lambda_i}{\mu_i(1-\beta_i)}\right]\frac{\lambda_i^{M_i}M_i}{\mu_i^{M_i}(1-\beta_i)^{M_i+1}}$$

The method will converge.

4. STABILITY CONDITION

An open network is stable if long-run distribution exists for the number of jobs in the input queue. Thus, the system under consideration is stable if jobs arrive at a rate that is smaller than the rate at which they can be served at the level one workstation. Then, the condition for stability is:

$$\lambda < \frac{1}{E[S]} \Rightarrow \lambda E[S] < 1 \Rightarrow \lambda \cdot \frac{1}{\lambda} + \left[\sum_{i=1}^{n} \lambda_i \left\{ \frac{1}{\mu_{oi}} + \frac{1}{\mu_i} \beta_i \right\} \right] < 1$$

or

$$\sum_{i=1}^{n} \lambda_i \left\{ \frac{1}{\mu_{oi}} + \frac{1}{\mu_i} \beta_i \right\} < 1 .$$

5. CONCLUSION

In this paper we modeled and approximated a multiple class two-level queuing production system with finite capacity. The approximation procedure decomposes the production network into two queuing systems with revised arrival and service processes. These individual queuing systems are then analyzed in isolation. Results on the mean effective service time and blocking probabilities are obtained.

REFERENCES

[1] Altiok, T., and Perros, H.G., "Approximate analysis of arbitrary configurations of open queuing networks with blocking", *Ann. Ops. Res.*, 9 (1987) 481-509.

[2] Bucacott, J.A., and Yao, D.D., "On queuing network models of flexible manufacturing systems", *Queuing Syst.*, 1 (1986) 5-27.

[3] Perros, H.G., "Queuing networks with blocking, A bibliography", *ACM Sigmet, Perform. Eval. Rev.*, 12 (1984).

[4] Perros, H.G., and Atliok, T., "Approximate analysis of open networks of queues with blocking: tandem configurations", *IEEE Trans. Softw. Engng.*, 12 (1986).

[5] Tsiotras, G.D., "Modeling and analysis of multiple-class tandem queuing systems with finite capacities", *Math. and Comput. Modeling*, 13 (2) (1990) 1-7.

[6] Tsiotras, G.D., and Tapiero, C.S., "WIP and CSP-1 Quality Control in a tandem queuing production system", *Computers Math. Applic.*, 23 (1) (1992) 89-101.

[7] Tsiotras, G.D., "Exact analysis of the joint production quantity-quality of a two-station stable tandem manufacturing system with restricted capacity", *Computers Math. Applic.*, 21 (4) (1991) 23-31.

[8] Tsiotras, G., and Badr, H., "A recursive methodology for the derivation of the blocking probabilities of a tandem queue with finite capacity", *Computers and Operations Research*, 13 (5) (1990) 475-479.