Research Journal of Applied Sciences, Engineering and Technology 4(22): 4678-4684, 2012
ISSN: 2040-7467

Submitted: March 26, 2012 Accepted: April 30, 2012 Published: November 15, 2012

Queuing Network Analysis on Hybrid Flow Shop Scheduling

[1, 2]Fuqing Zhao and [1]Qin Zhao
[1]School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China
[2]Key Laboratory of Contemporary Design and Integrated Manufacturing Technology, Ministry of Education, Northwestern Polytechnical University, 710072, China

Abstract: In this study, we consider a queuing model extension for a production system composed of several parallel machines and the same number of transporters. To obtain the minimum waiting time of the jobs in the queue, we present an exact solution for the proposed queuing model. The solution integrates M/M/C system with M/M/1 system. We obtain explicit expressions for its steady-state behavior under M/M/C and M/M/1 assumptions. Further, in order to illustrate the usefulness of the proposed methods, numerical examples are solved. On the basis of the results of these examples, some important conclusions are drawn.

Keywords: Lead time, manufacture scheduling, performance analysis, queuing theory

INTRODUCTION

Traditionally, organizations have concentrated on the reduction of product cost through mass production. Up until the 1980s, the lead time of a product in the system was not a primary concern of managers and planners in a manufacturing system. The 1990s saw the emergence of time as a strategic factor in enterprise competitiveness, resulting in a shift of focus to identifying ways to reduce production lead time. Analysis of the time spent by parts in the production system has led naturally to an increased interest in the dynamics of queuing networks, which are used to model such systems (Govil and Fu, 1999).

The manufacturing lead time is thus the sum of the set-up and processing times at each of the work stations in the job's routing sequence plus all of the time spent waiting in queues in front of the workstations needed. Clearly, the production process can be viewed as a network of queues. Queuing theory indicates that these queuing times depend upon the relative arrival and processing rates of jobs at various stages in the manufacturing process. It is reported that manufacturing lead times are often long and unreliable almost entirely due to the large proportion of time spent in the queues. Stalk and Hout reported that 95-99% of the production time is spent in queues (Stalk and Hout, 1992). Therefore, it is important to reduce the queuing times for the manufacturing companies.

The queuing theory result most widely used in analyzing and planning manufacturing systems is Little's law (Little, 1961). Little's law states that the time a new job just arriving spends in the system is the average number of jobs in the system divided by the arrival rate. The law can be applied at all levels of the system; individual work station, department and entire system. The intuitive explanation of this law can be as follows. Suppose the steady-state production rate be P and there are N jobs in the system. Every 1/P time units a new job arrives on the system and each job in the system advance one place. Spending1/Ptime units at each of N spots, the time in the system will be $T = N(1/P)$, or equivalently $N = PT$ in accord with Little law. Further more, Kobayashi has shown that Little's law is true not only for First-Come-First-Serve (FCFS) and Last-Come-First-Serve (LCFS), but also for any queue discipline (Kobayashi, 1978).

The application of queuing network theory to gain insights into the behavior of manufacturing systems dates back to the 1950s. The pioneer study on queuing networks is there search of Jackson (1957) and Jackson (1963). Since the n, a large number of analytical as well as simulation models have been used to study the behavior of queues at the different resources in a manufacturing system. Manish and Michael surveyed the contributions and applications of queuing theory in the field of discrete part manufacturing (Manish and Michael, 1999). Azaron developed an acyclic network of queues for the design of a dynamic flow shop, where each service station represents a machine. They proposed a method for approximating the distribution function of the longest path length in the network of queues by constructing a proper continuous-time Markov chain (Azaron *et al*., 2006). Chen analyzed queuing models of certain manufacturing cells under product-mix sequencing rules based on a decomposition technique (Chen, 2008). Down and

Corresponding Author : Fuqing Zhaos School of Computer and Communication Lanzhou university of technology, Lanzhou 730050, China

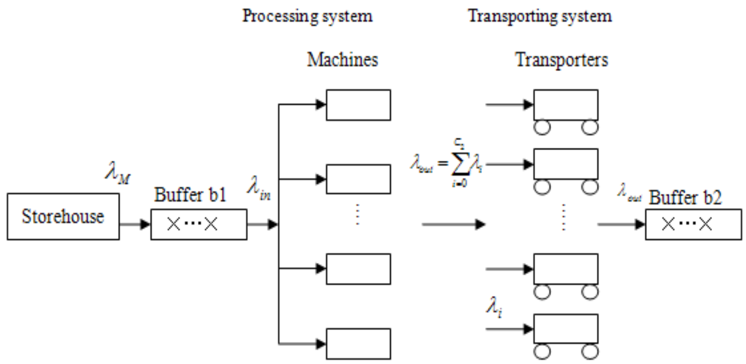

Fig. 1: Single-stage several parallel machines with transporters production system

Karakostas (2008) studied a queuing network where customers go through several stages of processing, with the class of a customer used to indicate the stage of processing (Down and Karakostas, 2008). Maglaras and Mieghem (2005) proposed an approach based on fluid-model analysis that translates the lead time specifications into deterministic constraints on the queue length vector (Maglaras and Mieghem, 2005). Stolletz (2007) proposed a new approach for the time-dependent analysis of stochastic and non-stationary M(t)M(t)c(t) queue systems (Stolletz, 2007). Ye used a modified N-policy M/G/1 queuing system with a non-reliable server to analyze the dynamic lot streaming (Ye, 2009). Missbauer gave a queuing-theoretical analysis of the clearing function concept and derive a procedure for its parameterization (Missbauer, 2009). Do considered a queuing model extension for a manufacturing cell composed of a machining center and several parallel down stream production stations under a rotation rule (Do, 2011). Zhao and Zheng developed a queuing network model of re-entrant lines and converted it to the canonical forms that are solvable by no n-linear matrix equations (Zhao and Zheng, 2010). The purpose of this study is to present a solution to the hybrid flow shop scheduling with transporters problems. We first provide details of the model under study. To obtain the minimum waiting time of the jobs in the queue, we integrate M/M/C system with M/M/1 system and derive the steady-state behavior of the proposed model. Different from the above mentioned researcher, we not only analyze the effects of the jobs arrival rate on the waiting time but also derive the probability that normal production is not affected.

METHODOLOGY

Parallel machines scheduling model:

Modeling framework: Here, for simplicity, we assume a two-stage production system, as illustrated in Fig. 1. In this system, the first stage (machine) is called processing system and the second stage is called transporting system. Now, we give the queuing model of this production system, as follow:

- We assume that there are C_1 machines in the processing system and the processing time at each machine is independent of other machines with processing rate λ_M. Jobs enter the warehouse into the buffer b_1 of the processing system with arriving rate λ_M. After a job arrives in the buffer, it goes directly to a machine for its first operation with arriving rate λ_{in} . If there are job(s) waiting for being processed, it queues up.
- The transporting system is composed of C_2 transporters. The rocessing time at each transporter is independent of other transporters and the processing rate is λ_M . After completion of processing at the first stage, jobs go to the transporting system and queue up for transporting.
- We assume the buffer of the machines which in the second processing operations is.b_2 In this study, we consider that weather the jobs reach the buffer b_2 in the shortest time after the completion of the first operations. Therefore, we need not describe the detail information of the second operations.

From the queuing theory, we know that the processing system is analyzed as M/M/C_1 system and the transporting system is analyzed as individual M/M/1system.

Assumptions: For simplicity, the following assumptions are made in this research:

- Arrival of jobs is characterized as a Poisson process.
- Service discipline is based on FIFO.
- Each job has characteristics which are statistically independent of other jobs and each machine is independent of other machines.
- A production system contains two stages and produces only one product, the first stage has several machines; the second stage is composed of several transporters. In this system, we assume the machines and transporters are not subject to break down.

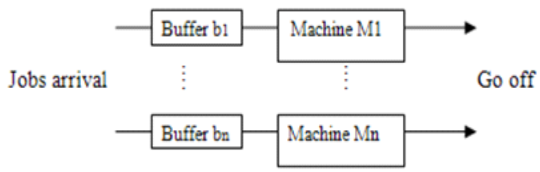

Fig. 2: Parallel processing systems with multiple buffers

- After completion of processing at a machine, jobs go to the next stage.
- Each stage has a unlimited buffer capacity, blocking is never happened.

In this model, we just set one buffer. We will give the reason in the next section.

The analysis of the buffer:

Single-stage parallel processing systems with multiple buffers: In this system, there are *n* parallel machines. Before each machine has an unlimited buffer and each job has only one operation at each machine, as illustrated in Fig. 2. The system is analyzed as *n* individual M/M/1 system. We assume jobs arrival rate follows a Poisson process with the parameter λ. After a job arrives in the system, it goes directly to a machine for its first operation. If there are job (s) waiting for being processed, it queues up in the buffer. Service discipline is based on FIFO. The processing time of the machines are exponentially distributed with the parameter μ. In this system, we do not consider the breakdown of the machine.

When the system reached steady-state, the formulation of the system as follows (Lu, 2002):
Service intensity:

$$\rho = \lambda/\mu \qquad (1)$$

Average queue length of the jobs in the buffer (the number of jobs):

$$L_q = \rho^2/1\text{-}\rho \qquad (2)$$

By the Little law, the average waiting time of the jobs:

$$W_q = \frac{L_q}{\lambda} = \frac{\lambda}{\mu(\mu-\lambda)} \qquad (3)$$

The probability of jobs sojourn time ω_s do not exceed t in the system is given by follow expression:

$$P\{\omega_s > t\} = \int_t^{\infty} \mu(1-\rho)e^{-\mu(1-\rho)h}\,dh = e^{-(\mu-\lambda)t} \qquad (4)$$

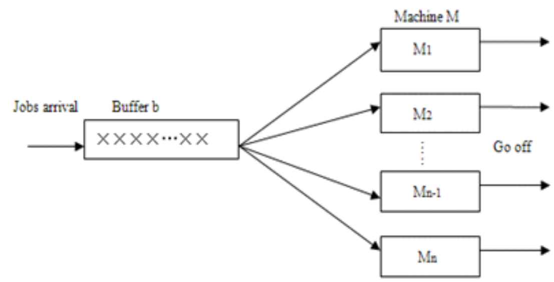

Fig. 3: Parallel processing systems with one buffer

Single-stage parallel processing systems with one buffer: Now, we consider the queuing modeling of single-stage parallel processing systems with one buffer as in Fig. 3.

The system is analyzed as M/M/C system. In this system, the job has only one operation and each machine has the same operation. We still assume jobs arrival rate follows a Poisson process with the parameter λ. After a job arrives in the system, it goes directly to a machine for its first operation. If there are job (s) waiting for being processed, it queues up in the buffer. Service discipline is based on FIFO. The processing time of the machines is also exponentially distributed with the parameter μ. In this system, we do not consider the breakdown of the machine.

When the system reached steady-state, we could deduce its stationary distribution as follows (Lu, 2002):
Service intensity:

$$\rho_1 = \frac{\lambda}{\mu}, \rho = \frac{\lambda}{n\mu} \qquad (5)$$

If we assume there are k jobs in this system that its probability is:

$$p_k = \begin{cases} \dfrac{\rho_1^k}{k!}p_0 = \dfrac{n^k}{k!}\rho^k p_0, 0 \le k < n \\ \dfrac{\rho_1^k}{n!n^{k-n}}p_0 = \dfrac{n^n}{n!}\rho^k p_0, k \ge n \end{cases} \qquad (6)$$

The probability of idle machine:

$$p_0 = \left(\sum_{k=0}^{n-1}\frac{\rho_1^k}{k!} + \frac{\rho_1^n}{n!}\frac{1}{1-\rho}\right)^{-1} \qquad (7)$$

From (5) (6), the performance of the system can be obtained as follows:

The average queue length of the jobs in the buffer, (the number of jobs):

Table 1: The comparison of performance estimates

	MM/3	M/M/1
P_0	0.075	0.25 (each subsystem)
the probability of jobs waiting	0.57	0.75 (the entire system)
L_c	1.70	2.25 (each subsystem)
W_c	1.89	7.50 (the entire system)

$$L_q = \frac{\rho_1^{n+1}}{(n-1)!(n-\rho_1)^2} p_0 \tag{8}$$

The average number of the busying machines:

$$L_q = \bar{k} = \sum_{k=0}^{n} kp_k + n\sum_{k=n+1}^{\infty} p_k = n\rho = \rho_1 \tag{9}$$

The average waiting time of the jobs:

$$W_q = \frac{L_q}{\lambda} = \frac{\rho_1^n p_0}{\mu n.n!\left(1-\rho^2\right)} \tag{10}$$

The probability of jobs queuing:

$$C(n,\rho_1) = \sum_{k=n}^{\infty} p_k = \frac{np_n}{n-\rho_1} \tag{11}$$

RESULTS ANALYSIS

Numerical examples and analysis: In the first production queuing model, we set up multiple buffers which equivalent to multiple queues. In other words, before each machine has a queue. In the second production queuing model, we only set one buffer and only one queue. In the case of processing the same jobs, we present numerical examples to illustrate the solution procedures discussed above. Relevant input parameters are as follows:

Fixed parameters:

$n = 3$, $\lambda = 0.3$, $\mu = 0.4$

Results from the two models developed above and shown in Table 1 demonstrate the comparison of performance estimates.

From Table 1, we see that the queuing waiting time and sojourn time of the jobs in the M/M/C queuing model is less than in M/M/1 queuing model. Therefore, we could set up one buffer in above production system. By doing this, it both reduces the lead time and saves the production cost.

The formulation of the system: For the entire system, when the system reaches steady-state, the task input is a Poisson process at rate λ_M and $\lambda_M = \lambda_{in} + \lambda_{out}$ For the processing system, theoretically, it is equal for inflow and outflow of the workload. Thus, $\lambda_{in} = \lambda_{out}$. But in the actual production, in order to ensure the production is not interrupted, the rate of the jobs arrival is greater than the rate of the jobs go off. That is, $\lambda_{in} > \lambda_{out}$. We assume the processing time of the machines is exponentially distributed with the parameter μ_M, that is, $f_M(t) = \mu_M e^{-\mu_M t}$. From (5), (8) and (10), the relevant performance parameters of the processing system can be obtained as follow:

Service intensity:

$$\rho_M = \frac{\lambda_{in}}{C_1\mu_M} \tag{12}$$

Average queue length of the jobs in the buffer (the number of jobs):

$$L_q^M = \frac{\left(C_1\rho_M\right)^{C_{1+1}}}{(C_1-1)!\left(C_1 - C_1\rho_M\right)^2} p_0 \tag{13}$$

The average waiting time of the jobs:

$$W_q^M = \frac{L_q}{\lambda_m} = \frac{\left(C_1\rho_M\right)^{C_1} p_0}{\mu_M C_1 \times C_1!\left(1-\rho_M\right)^2} \tag{14}$$

where,

$$p_0 = \left[\sum_{n=0}^{C_1-1}\frac{\left(C_1\rho_M\right)^n}{n!} + \frac{\left(C_1\rho_M\right)^{C_1}}{C_1!\left(1-\rho_M\right)}\right]^{-1}$$

After the jobs complete their first operation, they come to the transporting system with the external arrivals at rate λ_{out}. Further, the external arrival of each transporter is at rate $\lambda_T = \lambda_{out} / C_2$. From (1), (2) and (3), we have

Service intensity:

$$\rho_T = \frac{\lambda_T}{\mu_T} \tag{15}$$

The average queue length of the jobs:

$$L_q^T = \frac{C_2\rho_T}{1-\rho_T} \tag{16}$$

The average waiting time of the jobs:

$$W_q^T = \frac{L_q^T}{\lambda_T} = \frac{\lambda_T}{\mu_T\left(\mu_T - \lambda_T\right)} \tag{17}$$

For the processing system and the transporting system, the average queue length of the system is equal to the average number of scheduling jobs. In the above analysis, we see the processing systems and transporting system as two mutual independent subsystems. We described their models and analyzed their performance measures respectively, but considering the entire system, it needs the two subsystems to coordinate mutually. For example, let the inputting workload of the entire system remains a constant λ_M. For the transporting system, from (17) we know if fixed the service rate μ_T , the waiting time is an increasing function of the arrival rate.

The sojourn time of the jobs in the entire system is equal to the waiting time plus the processing time. In the processing system, the density function of the sojourn time of the scheduling task can be obtained as following (Sun and Li, 2002):

$$\omega_M(t) = \begin{cases} 0,\ t \le 0, \\ \left(C1\mu_M^2 tp_n + \mu_M - C_1\mu_M p_n\right)e^{-\mu_M t}, \\ C_1\mu_M = \lambda_{in} + \mu_{M,}\ t \ge 0 \\ \left[\mu_M + \dfrac{\mu_M^2 p_n}{(1-\rho_M)(C_1\mu_M - \lambda_{in} - \mu_M)}\right]e^{-\mu_M t} \\ \quad - \dfrac{C_1\mu^2 p_n}{C_1\mu_M - \lambda_{in} - \mu_M} \\ \times e^{-(C1\mu_M - \lambda_{in})t},\ C_1\mu_M \ne \lambda_{in} + \mu_M,\ t > 0 \end{cases} \tag{18}$$

where,

$$p_n = \begin{cases} \dfrac{(C_1\rho_M)^n p_0}{n!}, & n = 1,2,\ldots,C_1 - 1 \\ \dfrac{(C_1^{C_1}\rho_m^n)^n p_0}{C_1!}, & n = C_1, C_1+1,\ldots, \end{cases}$$

$$p_0 = \left[\sum_{n=0}^{C_1-1}\frac{(C_1\rho_M)^n}{n!} + \frac{(C_1\rho_M)^{C_1}}{C_1!(1-\rho_M)}\right]^{-1}$$

Further, the average sojourn time of the jobs in the processing system is denoted by $W^M{}_s$, thus:

$$\begin{aligned} W_s^M &= E[T] = \frac{\rho_M p_n}{\lambda_{in}(1-\rho_M)^2} + \frac{1}{\mu_M} \\ &= \frac{\rho_M^{C_1} p_0}{\mu_M C_1 \times C_1!(1-\rho_M)^2} + \frac{1}{\mu_M} \end{aligned} \tag{19}$$

The jobs through the processing system into the transporting system, because the transporting system is analyzed as C_2 M/M/1 models, for the transporting system, the density function of the sojourn time T is given by:

$$\omega_T(t) = (\mu_T - \lambda_T)e^{-(\mu_T-\lambda_T)t} \tag{20}$$

Further, the average sojourn time of the transporting system can be obtained by using (20) as follows:

$$W_S^T = E[T] = \frac{1}{(\mu_C - \lambda_T)} \tag{21}$$

Therefore, the average sojourn time of the entire system is:

$$W = W^M{}_s + W^T{}_s \tag{22}$$

The sojourn time of the jobs in the entire system is equal to the time from buffer b_1 to buffer b_2 and its density function is given by following expression (Qu and Hu, 2004):

$$\omega(t) = \omega_M(t)\otimes\omega_T(t) \tag{23}$$

Let $\tilde{\mu} = (\mu_T - \lambda_T)$

when,

$$n_M\,\mu_M = \lambda_{in} + \mu_M$$

Thus:

$$\begin{aligned} \omega(t) &= \int_0^t \tilde{\mu}e^{-\tilde{\mu}(t-u)} \times \left[C_1\mu^2 tp_n + \mu_M - C_1\mu_M p_n\right]du = \\ &\frac{\tilde{\mu}\mu_M(1-\rho_M-p_n)(\tilde{\mu}-\mu_M) - \tilde{\mu}\,\mu_M(C_1\mu_M-\lambda_T)p_n}{(1-\rho_M)(\tilde{\mu}-\mu_M)_2} \\ &\left(e^{-\mu_M t} - e^{-\tilde{\mu}t}\right) \\ &+ \frac{\tilde{\mu}\mu_M(C_1\mu_M-\lambda_T)p_n}{(1-\rho_M)(\tilde{\mu}-\mu_M)}te^{-\mu_A t} = A\left(e^{-\mu_M t} - e^{-\tilde{\mu}t}\right) + Bte^{-\mu_M t} \end{aligned}$$

where,

$$A = \frac{\tilde{\mu}\mu_M(1-\rho_M-p_n)(\tilde{\mu}-\mu_M) - \tilde{\mu}\mu_M(C_1\mu_M-\lambda_T)p_n}{(1-\rho_M)(\tilde{\mu}-\mu_M)^2}$$

$$B = \frac{\tilde{\mu}\mu_M(C_1\mu_M-\lambda_T)p_n}{(1-\rho_M)(\tilde{\mu}-\mu_M)}$$

Therefore, the probability of the jobs transported to buffer b_2 timely without affecting the production can be obtained as follows:

Table 2: The minimum waiting time and the optimal value of P

Arrival	M-wait	M-stay	T-wait	T-stay	W-wait	W-stay	P
$\lambda_m = 0.8$, $\lambda_T = 0.8$	0.19	1.85	2.32	4.32	2.51	6.17	0.61
$\lambda_m = 0.9$, $\lambda_T = 0.7$	0.26	1.93	1.70	3.70	1.96	5.63	0.65s
$\lambda_m = 1.0$, $\lambda_T = 0.6$	0.36	2.03	1.33	3.33	1.69	5.36	0.73
$\lambda_m = 1.1$, $\lambda_T = 0.5$	0.50	2.17	1.03	3.03	1.53	5.20	0.81
$\lambda_m = 1.2$, $\lambda_T = 0.4$	0.74	2.41	0.73	2.73	1.47	5.14	0.83
$\lambda_m = 1.3$, $\lambda_T = 0.3$	1.08	2.75	0.51	2.51	1.59	5.26	0.79
$\lambda_m = 0.4$, $\lambda_T = 0.2$	1.45	3.12	0.39	2.39	1.84	5.51	0.70

$$P\left(t < t_0\right) = \int_0^t \omega(t)dt = \int_0^t \left[A\left(e^{-\mu_M t} - e^{-\tilde{\mu} t} + Bte^{-\mu_A t}\right)\right]dt = \frac{A}{\tilde{\mu}}e^{-\tilde{\mu} t_0} - \frac{\left(A + Bt_0\right)\mu_M + B}{\mu_M^2}e^{-\mu_M t_0} + \frac{A}{\mu_M} - \frac{A}{\tilde{\mu}} + \frac{b}{\mu_A^2} \tag{24}$$

Similarly, when $n_M\, \mu_M \neq \lambda_{in} + \mu_M$

$$P(t < t_0) = \frac{C - D}{\tilde{\mu}}e^{-\tilde{\mu} t_0} - \frac{C}{\mu_M}e^{-\mu_M t 0} + \frac{D}{\left(1 - \rho_M\right)C_1\mu_M}e^{-(1-\rho_M)C_1\mu_M t_0} + \frac{C}{\tilde{\mu}} - \frac{C - D}{\tilde{\mu}}\frac{D}{\left(1 - \rho_M\right)C_1\mu_M} \tag{25}$$

where,

$$C = \frac{\tilde{\mu}\mu_M\left[P_c + \left(1 - \rho_M\right)\left(C_1 - C_1\rho_M - 1\right)\right]}{\left(\tilde{\mu} - \mu_M\right)\left(1 - \rho_M\right)\left(C_1 - C_1\rho_M - 1\right)}$$

$$D = \frac{C_1\tilde{\mu}\mu_M P_n}{\left(C_1 - C_1\rho_M - 1\right)\left(\tilde{\mu} - C_1\mu_M + C_1\mu_M\rho_M\right)}$$

Numerical examples and analysis: In this section, we present numerical examples to illustrate the solution procedures discussed above. Relevant input parameters are as follows:

Fixed parameters:

$C_1 = 3$, $C_2 = 3$, $\mu_M = 0.6$, $\mu_T = 0.5$, $\lambda_M = 1.6$

In this table, the queuing waiting time W_q, the average sojourn time and the optimal probability p that is computed by means of expressions (14), (17) and (24), (25) for all circumstances are shown in Table 2.

Future, Fig. 4 and 5 illustrate the influence of every varied parameter on the average waiting time and the probability P, respectively. Some important conclusions drawn from Fig. 4 and 5 are as follows:

- From Fig. 4, we see that the average waiting time of the processing system M-wait indeed increases as the arrival rate λ_M increases. The average waiting time of the transporting system T-wait decrease as the arrival rate λ_M increase. When $\lambda_M = 1.2$, $\lambda_T = 0.4$, the minimum average waiting time for the entire system is obta

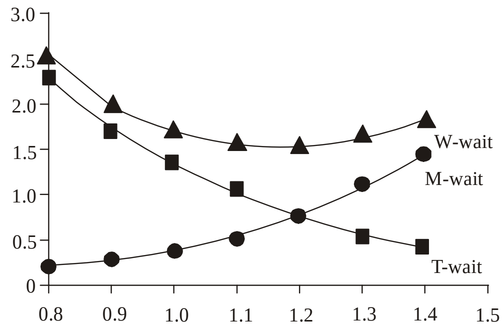

Fig. 4: Minimum waiting time

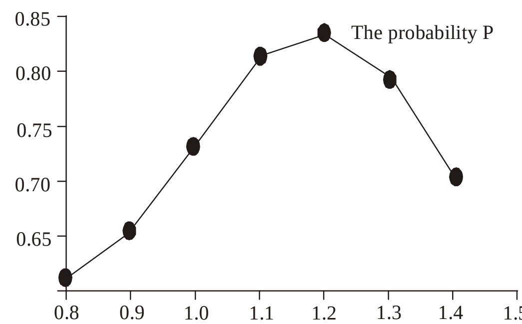

Fig. 5: Optimal probability P

- From Fig. 5, we can see that the probability P is a convex function of the arrival rate λ_M. It is also to say when the whole waiting time increases as the probability P decreases.

Therefore, $\lambda_M = 1.2$, $\lambda_T = 0.4$ could be the optimal choice.

CONCLUSION

This study considers a queuing model extension for a production system composed several parallel machines and the same number of transporters. A queuing network analysis for determining the minimum waiting time and the probability that normal production is not affected is presented. The numerical examples provided in the final part of the study demonstrated that the method could be available and effective. Although the example in the study includes only two stages, we can easily extend it to a multistage environment. The breakdown of the machines and transporters, a limited buffer or a multiproduct system should be considered in the future search.

ACKNOWLEDGMENT

This study is financially supported by the National Natural Science Foundation of China under Grant No. 61064011. And it was also supported by Scientific research funds in Gansu Universities, Science Foundation for the Excellent Youth Scholars of Lanzhou University of Technology, Educational Commission of Gansu Province of China, Natural Science Foundation of Gansu Province and Returned Overseas Scholars Fund under Grant No. 1114ZTC139, 1014ZCX017, 1014ZTC090, 1114ZSB091 and 1014ZSB115, respectively.

REFERENCES

Azaron, A., H. Katagiri, K. Kato and M. Sakawa, 2006. Longest path analysis in networks of queues: Dynamic scheduling problems. Eur. J. Oper. Res., 174(1): 132-149.

Chen, J.T., 2008. Queueing models of certain manufacturing cells under product-mix sequencing rules. Eur. J. Oper. Res., 188: 826-837.

Do, T.V., 2011. A new solution for aqueueing model of amanufacturing cell with negative customers under arotation rule. Perform. Evaluation, 68: 330-337.

Down, D.G. and G. Karakostas, 2008. Maximizing throughput in queueing networks with limited flexibility. Eur. J. Oper. Res., 187: 98-112.

Govil, M.K. and M.C. Fu, 1999. Queueing theory in manufacturing: A survey. J. Manuf. Syst., 18(3): 214-240.

Jackson, J.H., 1957. Networks of waiting lines. Oper. Res., 5: 518-521.

Jackson, J.K., 1963. Job shop-like queuing systems. Manage. Sci., 10: 131-142.

Kobayashi, H., 1978. Modelling and Analysis: An Introduction to System Performance Evaluation Methodology. Addison Wesley, New York.

Little, J.D.C., 1961. A proof of the queuing formula.J $-\lambda$W. Oper. Res., 9: 383-389.

Lu, C.L., 2002. Queuing Theory. Retrieved from: http://www.win.tue.nl.

Maglaras, C. and J.V. Mieghem, 2005. Queueing systems with leadtime constraints: A fluid-model approach for admission and sequencing control. Eur. J. Oper. Res., 167: 179-207.

Manish, K.G. and C.F. Michael, 1999. Queuing theory in manufacturing: A survey. J. Manuf. Syst., 18(3): 214-240.

Missbauer, H., 2009. Order release planning with clearing functions: A queueing-theoretical analysis of the clearing function concept. Prod. Econ., 131: 399-406.

Qu, W.M. and Q.Y. Hu, 2004. The CIMS logistics scheduling system modeling and simulation. Comput. Integr. Manuf. Syst., 10: 1067-1072.

Stalk, Jr., G. and T.M. Hout, 1992. Competing Against Time. Free Press, New York.

Stolletz, R., 2007. Approximation of the non-stationary M(t)/M(t)/c(t)-queueusing stationary queueing models: The stationary backlog-carryover approach. Eur. J. Oper. Res., 190: 478-493.

Sun, R.H. and J.P. Li, 2002. The Basis of Queuing Theory. Retrieved from: http://cs.gmu.edu.

Ye, T.F., 2009. Queueing network analysis on dynamic lot streaming. Comput. Oper. Res., 36: 415-424.

Zhao, L.N. and Y.P. Zheng, 2010. An open queuing network model on reentrant production system and the algorithm. Cont. Decis., 15: 181-185.