

Production Allocation Problem with Penalty by Tardiness of Delivery under Make-to-Order Environment

Yasuhiko TAKEMOTO¹, and Ikuo ARIZONO²

¹School of Business Administration, University of Hyogo
8-2-1 Gakuennishi-cho, Nishi-ku, Kobe, Hyogo 651-2197, JAPAN
e-mail:ys-take@biz.u-hyogo.ac.jp

²Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, JAPAN
e-mail:arizono@eis.osakafu-u.ac.jp

POMS 18th Annual Conference Dallas, Texas, U.S.A., May 4 to May 7, 2007

Abstract

We consider a Make-to-Order manufacturing system which can be formulated as an open queueing network composed of some GI/G/1 queueing systems. In Make-to-Order manufacturing systems, it is important to fulfill customer's requests quickly. Therefore, it has been frequently investigated how to shorten the lead time. The production allocation can be one of such the strategies. However, many researchers have focused on nothing but the expectation of lead time, *e.g.*, minimizing the expectation of lead time. In this article, we pay attention to not only the expectation but also the variance of lead time, and then evaluate a tardiness cost due to delivery behind a prescribed schedule. Especially, we assess a maximal loss among all the lead time distributions with the same expectation and variance. Then, an optimal production allocation plan is decided using production and tardiness costs. Lastly, we provide some numerical examples and perform the sensitivity analysis for the optimal production allocation.

Keywords

Distribution free approach, Due delivery interval, Lead time, Open queueing network, Tardiness cost.

1. Introduction

A large number of industries have been moving toward time-based competition [14]. Therefore, manufacturers need to obtain great ability to fulfill customer's requests quickly for the purpose of winning a competitive advantage. Especially, it is a significant issue in make-to-order (MTO) manufacturing environments. MTO manufacturers usually yield such as customized, seasonal, and short life cycle items. Therefore, the stock of final items isn't held ahead of demand and the item starts to be built up immediately after the receipt of customer's request in the MTO manufacturing systems. While the MTO manufacturing systems are able to save holding and wasting costs about surplus stock

exceeding demand and reduce the manufacturer's exposure to financial risk, it may cause long lead time, large backorders and lost sales [8]. Therefore, the MTO manufacturers have been impelled to search for ways to hold down an operating cost and to deliver quickly.

The production allocation can be one of such the strategies which are practiced for the purpose of quick delivery to customers [1, 2], where the production allocation means that customer's orders are assigned to one of some plants in which the same item can be processed. The workload can be balanced and then the waiting time can be shortened by allocating the production to some plants. On one hand, each plant usually has distinct capacity, *i.e.*, processing cost and time. Therefore, the optimal plan of production allocation to some plants should be decided by taking each plant capacity into consideration. In this article, we consider an optimal production allocation plan in the MTO manufacturing systems.

MTO manufacturing systems are in particular formulated into mathematical models using open queueing network (OQN) theory. Then, the performance of formulated systems is evaluated, such as in-process inventory, equipment utilization, and lead time [2, 7]. The lead time means time interval that a customer's order spends in the system from the moment the order is placed to the moment it is completed for delivery. The lead time is given as a random variable from the variability in the arrival of orders and processing when the system is formulated as an OQN. In the previous literature, the production allocation problem for minimizing the average lead time has been formulated and investigated [1, 5]. Also, Bitran and Dasu [1] have formulated the production allocation problem for minimizing the operating cost of the system subject to a given upper bound on the average lead time.

A regular interval from order to delivery is sometimes decided and promised between manufacturers and customers beforehand. In this article, this time interval is called "due delivery interval". The due delivery interval corresponds to the lead time that a customer might expect from the time of placing the order to the time of delivery. Then, a kind of penalties is frequently put on tardy delivery which exceeds the due delivery interval in the production management. For the purpose of evaluating the tardiness cost, the distribution of lead time should be investigated. For OQN models, Whitt [15, 16] has proposed an approximate method on evaluating the expectation and variance of lead time. Then, Vandaele *et al.* [13] have given an approximate distribution of lead time by using the expectation and variance of lead time, where the accuracy of their approximate distribution hasn't been shown. Generally speaking, it may be extremely difficult to derive the accurate distribution of lead time, *e.g.*, when the type of queues in the OQN is considered to be GI/G/1 queue with the general independent interarrival and service time distributions. Therefore, an more appropriate procedure for evaluating the tardiness cost should be developed.

In this article, we investigate an new planning of production allocation to two plants under an MTO manufacturing system in order to minimize an evaluation function composed of the production and tardiness costs. The rest of this article is organized as follows. Section 2 explains the model descriptions and notations. Section 3 formulates the MTO manufacturing system with two plants using the OQN theory and evaluates the expectation and variance of lead time. Section 4 derives an upper bound on the

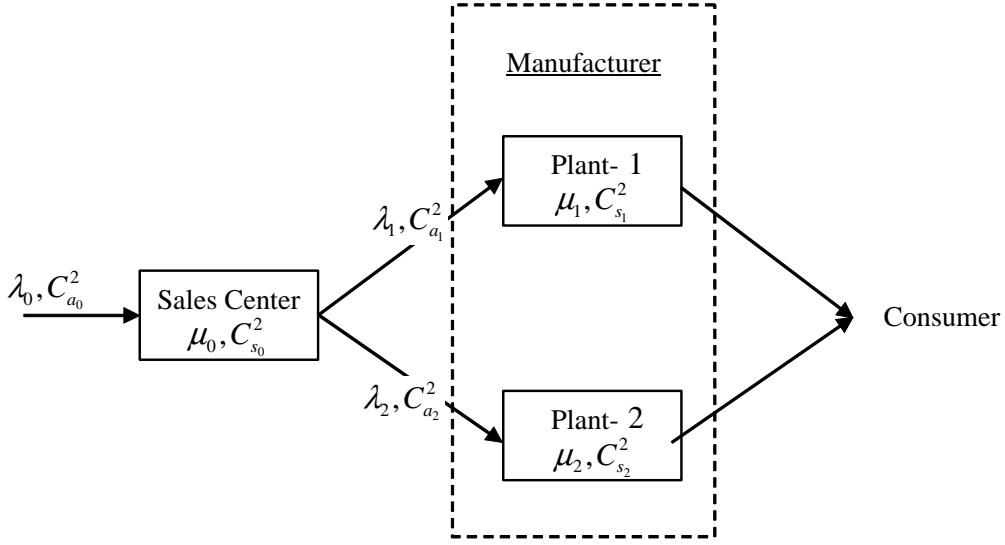


Figure 1: The outline of the MTO manufacturing system considered in this paper

tardiness cost using the idea of “distribution free approach” [10] and indicates a decision procedure with respect to the optimal plan of production allocation. Section 5 shows some numerical examples and performs the sensitivity analysis. Finally, Section 6 concludes this article.

2. Model description

In this article, we consider an MTO manufacturing system about single item that can be manufactured at each production plant. Concretely, Figure 1 illustrates the outline of the MTO manufacturing system considered in this article, where the system is composed of a sales center and two production plants. The assumptions about demand, sales center, and plants are as follows:

- Order interarrival time is independently and identically distributed, where its distribution function is unknown. Only the information about the expectation and variance of order interarrival time is given.
- The sales center and production plants are respectively defined as a GI/G/1 queueing system. The respective processing time at each plant is independently and identically distributed, where its distribution function is unknown. Only the information about the expectation and variance of respective processing time is given.
- Order arrival and production processes are independent of each other.
- The respective buffer space at each plant is unlimited.
- The discipline of “first-come, first-service (FCFS)” is adopted to sequence orders waiting for service and production.

All customer orders are accepted and then the initial service for orders is executed at the service center. After the completion of the initial service, the sales center requests the production about the order to either production plant 1 or 2. This selection of plants is based on a probabilistic allocation rule, which is defined as follows [4, 6, 9].

- A request for production is assigned to either plant 1 or 2 by Bernoulli trials.
- No information about the history and the present state of the system is used in the selection of the plant.

The production starts after the allocation of order to the plant. Then, the final item is delivered to the customer.

Further, the notations are represented as follows, where the sales center and plants are called “station”. Concretely, the station 0, 1 and 2 correspond to the sales center and the plant 1 and 2, respectively.

λ_i : demand rate per unit time at the station $i, i = 0, 1, 2$, where mean order interarrival time is given as $1/\lambda_i$.

$C_{a_i}^2$: squared coefficient of variation (SCV) of order interarrival time at the station $i, i = 0, 1, 2$, where SCV is given as (variance / mean²).

μ_i : processing rate in production per unit time at the station $i, i = 0, 1, 2$, where mean processing time is given as $1/\mu_i$.

$C_{s_i}^2$: SCV of processing time at the station $i, i = 0, 1, 2$.

ρ_i : traffic rate at the station $i, i = 0, 1, 2$, where $\rho_i = \lambda_i/\mu_i$.

c_i : processing cost per service or production at the station $i, i = 0, 1, 2$.

L : due delivery interval.

b : penalty per unit time due to tardiness of delivery.

p : probability of allocating production from the station 0 to the station 1 (decision variable), where $1 - p$ implies the probability of allocating production from the station 0 to the station 2.

$[x]^+$: $\max\{0, x\}$

$E[x]$: expectation of a random variable x .

$V[x]$: variance of a random variable x .

$\text{Cov}(x, y)$: covariance of random variables x and y .

3. Analysis of lead time property

We investigate the lead time property of the MTO manufacturing system in Figure 1 by using the theory of OQN. Whitt [15, 16] has developed the approximate evaluation of some performance for various types of OQNs and his result has been applied to a lot of researches [2, 5, 7, 11, 13]. In his analysis, the arrival and departure processes at each station are described as renewal processes. Then, every station is individually analyzed, and the decomposition method is applied. The decomposition method is described by three steps as follows [2, 15]: (i) analysis of interaction between stations in the network,

(ii) evaluation of performance at each station, and (iii) evaluation of performance for the whole network. We also apply his procedure to the MTO manufacturing system in Figure 1.

At first, note that $\lambda_0, C_{a_0}^2, \mu_i$ and $C_{s_i}^2, i = 0, 1, 2$, are given, respectively. Then, the order arrival rate λ_i at the station $i, i = 1, 2$, is respectively obtained as

$$\lambda_1 = p\lambda_0, \quad (1)$$

$$\lambda_2 = (1-p)\lambda_0, \quad (2)$$

where the traffic rate at the station i needs to satisfy $\rho_i < 1$. Further, $C_{a_i}^2, i = 1, 2$, is given as follows:

$$C_{a_1}^2 = pC_d^2 + 1 - p, \quad (3)$$

$$C_{a_2}^2 = (1-p)C_d^2 + p, \quad (4)$$

where C_d expresses the SCV of interdeparture time from the station 0 and is obtained as

$$C_d^2 = \rho_0^2 C_{s_0}^2 + (1 - \rho_0^2) C_{a_0}^2.$$

All the interaction between stations in Figure 1 can be represented by (1)–(4).

The, denote the waiting time of orders at the station $i, i = 0, 1, 2$, by w_i . By using (1)–(4), we obtain the following expectation and variance of w_i :

$$E[w_i] = \frac{\rho_i(C_{a_i}^2 + C_{s_i}^2)}{2\mu_i(1 - \rho_i)} \Phi_i, \quad (5)$$

$$V[w_i] = (E[w_i])^2 \left(\frac{\Delta_i + 1}{\Theta_i} - 1 \right), \quad (6)$$

where

$$\Phi_i = \begin{cases} \exp \left[-\frac{2(1 - \rho_i)(1 - C_{a_i}^2)^2}{3\rho_i(C_{a_i}^2 + C_{s_i}^2)} \right], & C_{a_i}^2 \leq 1 \\ 1, & C_{a_i}^2 > 1 \end{cases}$$

$$\Delta_i = \begin{cases} 2\rho_i - 1 + \frac{4(1 - \rho_i)(2C_{s_i}^2 + 1)}{3(C_{s_i}^2 + 1)}, & C_{s_i}^2 \leq 1 \\ 2\rho_i - 1 + \frac{4C_{s_i}^2(1 - \rho_i)}{(C_{s_i}^2 + 1)}, & C_{s_i}^2 > 1 \end{cases}$$

$$\Theta_i = \begin{cases} \rho_i + \frac{\rho_i(1 - \rho_i)(C_{a_i}^2 - 1)(1 + C_{a_i}^2 + \rho_i C_{s_i}^2)}{1 + \rho_i(C_{s_i}^2 - 1) + \rho_i^2(4C_{a_i}^2 + C_{s_i}^2)}, & C_{a_i}^2 \leq 1 \\ \rho_i + \frac{4\rho_i^2(1 - \rho_i)(C_{a_i}^2 - 1)}{C_{a_i}^2 + \rho_i^2(4C_{a_i}^2 + C_{s_i}^2)}, & C_{a_i}^2 > 1 \end{cases}$$

Further, denote the flow time of orders at the station $i, i = 0, 1, 2$, by t_i . By using (5) and (6), the expectation and variance of flow time t_i at the station i are obtained as

$$E[t_i] = E[w_i] + \frac{1}{\mu_i}, \quad (7)$$

$$V[t_i] = V[w_i] + \frac{C_{s_i}^2}{\mu_i^2}. \quad (8)$$

At last, the expectation and variance of lead time t are given by

$$E[t] = E[t_0] + pE[t_1] + (1-p)E[t_2], \quad (9)$$

$$V[t] = V[t_0] + pV[t_1] + (1-p)V[t_2] + p(1-p)(E[t_1] - E[t_2])^2. \quad (10)$$

For the details, see appendix A.

4. Decision of optimal production allocation

In this section, we construct an evaluation function composed of production and tardiness costs and show a procedure for deciding a new planning of production allocation using the constructed evaluation function. We define the evaluation function per unit item as follows:

$$C(p) = c_0 + pc_1 + (1-p)c_2 + b \cdot E[t - L]^+. \quad (11)$$

Then, the respective term from first to third expresses the production cost at the plant and the last term means the penalty due to tardiness of delivery. In (11), $E[t - L]^+$ implies the average tardiness period of delivery behind the prescribed schedule and is given as follows:

$$E[t - L]^+ = \int_L^\infty (t - L)f(t; p)dt, \quad (12)$$

where $f(t; p)$ means a probability density function of lead time t for a given p . Then, $f(t; p)$ is unknown.

On decision making based on the evaluation function in (11), (12) must be evaluated for all p . In other words, $f(t; p)$ must be investigated in order to calculate (12). If the distribution of lead time were known, we could evaluate $C(p)$ and derive the optimal production allocation plan. Although we can evaluate the expectation and variance of lead time as mentioned in previous section, it is strictly difficult to derive the distribution function of lead time.

Vandaele *et al.* [13] have proposed a procedure for approximating the distribution of lead time using its expectation and variance. Concretely, they have given the approximate distribution of lead time based on a log-normal distribution, by letting the expectation and variance of lead time correspond to those of log-normal distribution. However, it can not be always guaranteed that the log-normal distribution by their approximation is the best approximate distribution of lead time. Actually, they haven't shown the adequacy of their approximation. We consider other procedures for evaluating

(12) than the procedure based on the log-normal approximation by Vandaele *et al.* [13]. So, we consult an investigation for the inventory policy presented by Moon and Gallego [10]. For $[t - L]^+ = \max\{0, t - L\}$, we obtain the following relation:

$$[t - L]^+ = \frac{|t - L| + (t - L)}{2}.$$

Then, the following inequality is derived from the Cauchy-Schwarz's inequality based on the expectation and variance of lead time:

$$\begin{aligned} E[|t - L|] &\leq \left\{E[(t - L)^2]\right\}^{1/2} \\ &= \left\{V[t] + (E[t] - L)^2\right\}^{1/2}. \end{aligned}$$

Also, it is obvious that $E[t - L] = E[t] - L$. Finally, we obtain the following relation:

$$E[t - L]^+ \leq \frac{1}{2} \left\{V[t] + (E[t] - L)^2\right\}^{1/2} + \frac{1}{2} (E[t] - L). \quad (13)$$

From (11) and (13), the following inequality is obtained:

$$\begin{aligned} C(p) &\leq c_0 + pc_1 + (1 - p)c_2 + \frac{b}{2} \left\{V[t] + (E[t] - L)^2\right\}^{1/2} + \frac{b}{2} (E[t] - L) \\ &= C_S + C_D \equiv C_{\text{DFA}}(p), \end{aligned} \quad (14)$$

where let $C_S \equiv c_0 + pc_1 + (1 - p)c_2$ as the production cost per unit item and $C_D \equiv \frac{b}{2} \left\{V[t] + (E[t] - L)^2\right\}^{1/2} + \frac{b}{2} (E[t] - L)$ as the tardiness cost per unit item. $C_{\text{DFA}}(p)$ in (13) means an upper bound of $C(p)$ in (11) among all the lead time distributions with the same expectation $E[t]$ and variance $V[t]$. Accordingly, we propose to employ $C_{\text{DFA}}(p)$ in (13) as an alternative evaluation function instead of $C(p)$. The evaluation technique mentioned above is called "distribution free approach" [10]. Then, our optimization procedure can be interpreted as a solution based on the Min-Max criterion.

5. Numerical examples and sensitivity analysis

In this section, we illustrate some numerical examples. Firstly, set the parameters for demand as follows:

$$\lambda_0 = 0.75, \quad C_{a_0}^2 = 1.00.$$

Then, the parameters for production at each station are given as follows:

$$\begin{aligned} \mu_0 &= 2.00, \quad \mu_1 = 1.25, \quad \mu_2 = 1.00, \\ C_{s_0}^2 &= 1.00, \quad C_{s_1}^2 = 1.00, \quad C_{s_2}^2 = 1.00. \end{aligned}$$

Further, specify the coefficients for each cost as:

$$c_0 = 2.00, \quad c_1 = 6.50, \quad c_2 = 5.00, \quad b = 2.00, \quad L = 4.00.$$

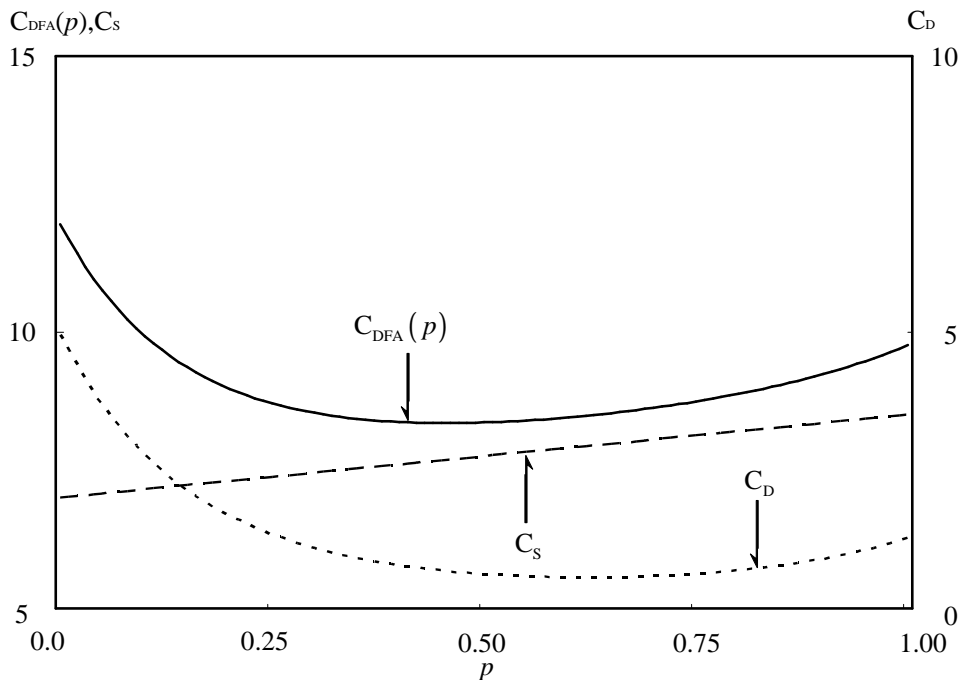


Figure 2: The behavior of C_S , C_D and $C_{DFA}(p)$ in p

We show the behavior of $C_{DFA}(p)$ in p in Figure 2. From Figure 2, it is confirmed that C_S is monotonically increased as p increases, since the production cost c_1 at the plant 1 is more expensive than one at the plant 2. On one hand, the behavior of C_D is convex in p . This is the reason why a big (or small) p brings the long queue of orders in the plant 1 (or 2). After all, $C_{DFA}(p)$ is convex in p . In addition to Figure 2, we show the numerical result for some properties such as $E[t]$, $V[t]$ and $C_{DFA}(p)$ in Table 1. From Table 1, $C_{DFA}(p)$ is minimized as $C_{DFA}(p^{*DFA}) = 8.35$ when $p^{*DFA} = 0.46$.

On one hand, Vandaele *et al.* [13] have assumed that the distribution of lead time is given the log-normal distribution as mentioned in previous section. Also, there may be some cases in which the distribution of lead time is assumed to be a plausible distribution. In this viewpoint, the distribution free approach is a somewhat pessimistic approach because the distribution of lead time is assumed to be unknown. However, it is extremely difficult to derive exactly the distribution function of lead time. The distribution free approach is considered a kind of robust solution.

Therefore, we investigate the optimal production allocation plan when the distribution of lead time is assumed to be a particular distribution. Suppose that the lead time obeys the log-normal distribution as:

$$f_{LN}(t) = \frac{1}{t\sqrt{2\pi\delta}} \exp \left\{ -\frac{(\log t - \nu)^2}{2\delta^2} \right\},$$

where ν and δ mean the respective parameters about location and scale. By letting the expectation and variance of lead time correspond to those of the log-normal distribution,

Table 1: $E[t]$ and $V[t]$ of the lead-time and C_S, C_D and $C_{DFA}(p)$ in some p in Figure 2

p	$E[t]$	$V[t]$	C_S	C_D	$C_{DFA}(p)$
0.00	4.80	16.64	7.00	4.96	11.96
0.10	3.65	9.68	7.15	2.78	9.93
0.20	2.98	6.21	7.30	1.67	8.97
0.30	2.57	4.30	7.45	1.09	8.54
0.40	2.31	3.21	7.60	0.77	8.37
0.46	2.22	2.81	7.69	0.66	8.35
0.50	2.17	2.63	7.75	0.61	8.36
0.60	2.12	2.40	7.90	0.56	8.46
0.70	2.15	2.47	8.05	0.58	8.63
0.80	2.27	2.83	8.20	0.68	8.88
0.90	2.47	3.52	8.35	0.89	9.24
1.00	2.80	4.64	8.50	1.27	9.77

the respective parameters about location and scale are given as follows:

$$\delta^2 = \log \left(\frac{V[t]}{E[t]^2} + 1 \right), \quad \nu = \log E[t] - \frac{\delta^2}{2}.$$

When the log-normal distribution with ν and δ as mentioned the above is assumed to be the distribution of lead time, we can calculate (12). Under the same condition in Figure 2, the optimal production allocation plan and the cost are obtained as

$$p^{*LN} = 0.41, \quad C_{LN}(p^{*LN}) = 8.08,$$

where $C_{LN}(p)$ implies the evaluation function by substituting the log-normal distribution for $f(t; p)$ in (12) and then p^{*LN} means the optimal production allocation plan based on the evaluation function $C_{LN}(p)$. Naturally, $C_{LN}(p^{*LN})$ is more economical than $C_{DFA}(p^{*DFA})$ because the distribution of lead time is given. Consider the difference between $C_{LN}(p^{*DFA})$ and $C_{LN}(p^{*LN})$. The difference is called the expected value of additional information (EVAI) and expresses the largest amount that we would be willing to pay for the knowledge about the distribution [10]. When the distribution of lead time is assumed the log-normal distribution, the EVAI is defined as follows:

$$EVAI_{LN} = C_{LN}(p^{*DFA}) - C_{LN}(p^{*LN}).$$

Since $EVAI_{LN}$ is obtained as 0.01 due to $C_{LN}(p^{*DFA} = 0.46) = 8.09$, we would conclude that there is not so much benefit by specifying the distribution of lead time as the log-normal distribution in comparison with the effort to specify it.

Further, we show some numerical examples for the purpose of verifying the impact of the variability of processing time on the allocation plan. In previous literature, it has been reported that the variability of the processing time in many manufacturing operations is smaller than 1 [3]. In Table 2, the parameters $C_{s_1}^2$ and $C_{s_2}^2$ are equal to

Table 2: The behavior of the optimal production allocation rate under $C_{s_1}^2 = 0.50$ and/or $C_{s_2}^2 = 0.50$

	$C_{s_1}^2$	$C_{s_2}^2$	$E[t]$	$V[t]$	p^{*DFA}	$C_{DFA}(p^{*DFA})$
Case 2-1	1.00	1.00	2.22	2.81	0.46	8.35
Case 2-2	0.50	1.00	2.15	2.43	0.48	8.29
Case 2-3	1.00	0.50	2.27	2.38	0.35	8.11
Case 2-4	0.50	0.50	2.21	2.11	0.37	8.07

Table 3: The behavior of the optimal production allocation rate under $C_{s_1}^2 = 2.00$ and/or $C_{s_2}^2 = 2.00$

	$C_{s_1}^2$	$C_{s_2}^2$	$E[t]$	$V[t]$	p^{*DFA}	$C_{DFA}(p^{*DFA})$
Case 3-1	1.00	1.00	2.22	2.81	0.46	8.35
Case 3-2	2.00	1.00	2.35	3.73	0.41	8.50
Case 3-3	1.00	2.00	2.20	3.54	0.61	8.72
Case 3-4	2.00	2.00	2.37	5.21	0.54	8.98

and/or smaller than them in Table 1, where the others are the same as settings in Table 1. Then, Case 2-1 in Table 2 corresponds to the situation in Table 1. From Table 2, it is confirmed that p^{*DFA} in Case 2-2 is larger than one in Case 2-1. This is brought by the reason why the penalty due to tardiness is decreased by allocating more production to the plant 1 with the smaller SCV. Case 2-2 and Case 2-3 are opposite cases. In Case 2-4, both SCVs of the processing time in the plants 1 and 2 become smaller. Then, more production is allocated to the plant 2 with the smaller production cost and the evaluation function $C_{DFA}(p^{*DFA})$ is decreased in comparison with Case 2-1. The property of these results is quite reasonable.

Also, Souza *et al.* [12] have reported that electronics remanufacturers have SCV between 1.50 and 2.00 under reverse logistics environment. Then, Souza and Ketzenberg [11] have considered the production allocation to an ordinary manufacturer and a remanufacturer, where their research has considered nothing but the expectation of lead time. In Table 3, the parameters $C_{s_1}^2$ and $C_{s_2}^2$ are equal to and /or larger than them in Table 1, where the others are the same as settings in Table 1. Then, Case 3-1 in Table 3 corresponds to the situation in Table 1. From Table 3, it is confirmed that p^{*DFA} in Case 3-2 is smaller than one in Case 3-1. This is brought by the reason why the penalty due to tardiness is decreased by allocating more production to the plant 2 with the smaller SCV. Case 3-2 and Case 3-3 are opposite cases. In Case 3-4, both SCVs of the processing time in the plants 1 and 2 becomes larger. Then, more production is allocated to the plant 1 with the larger processing rate. The property of these results is also quite reasonable.

6. Concluding Remarks

In this article, we have considered the production allocation problem with the penalty by tardiness of delivery behind the prescribed schedule under the MTO manufacturing environment. We have described the MTO manufacturing system using the OQN model. In the OQN analysis, each plant has been defined as the GI/G/1 queue, and then the expectation and variance of lead time in the whole system have been evaluated. When the distribution function of lead time is unknown, we have paid the attention to the worst situation among all the lead time distributions with the same expectation and variance. Further, we have applied the idea of the distribution free approach to obtaining the quasi-optimal production allocation plan for the worst situation among all the lead time distributions. Finally, the quasi-optimal production allocation plan has been successfully derived using the evaluation function composed of production and tardiness costs.

In this article, we consider the production allocation to two different plants. Of course, it would be general to formulate the model when the number of plants is n . Also, it has been assumed that the production is allocated to plants based on the probabilistic allocation rule in the same manner in previous investigations such as [2, 5]. Perhaps, operators may pay attention to the present queue condition in each plant in the allocation of production to plants. The formulation of the adaptive production allocation may be an interesting object. However, the object like this will be our future research.

Acknowledgments

Then, this research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology, Grant-in-Aid for Young Scientists (B), No.18710136, 2006. We would like to appreciate it for our research.

Appendix A: Derivation of (9) and (10)

In this appendix, we show the derivation of the expectation and variance of lead time in (9) and (10). The number of visits to the station i for each order is expressed as $N_i, i = 0, 1, 2$. Then, the expectation and variance of lead time are given as follows [5, 15]:

$$\begin{aligned} E[t] &= \sum_{i=0}^2 E[N_i]E[t_i] \\ V[t] &= \sum_{i=0}^2 E[N_i]V[t_i] + \sum_{i=0}^2 \sum_{j=0}^2 \text{Cov}(N_i, N_j)E[t_i]E[t_j] \end{aligned}$$

where $E[N_i], i = 0, 1, 2$, means the expected number of visits to the station i for each order and is given as

$$E[N_i] = \frac{\lambda_i}{\lambda_0}$$

By the way, orders from customers always arrive at the sales center, *i.e.* the station 0. Then, the production flows from the station 0 to the station $i, i = 1, 2$, in this article. Then, $N_i, i = 1, 2$, is a random variable of the Bernoulli distribution with parameters N_0 and p . When $i = j$, $\text{Cov}(N_i, N_i)$ means the variance of N_i . Therefore, $\text{Cov}(N_i, N_i), i = 1, 2$, is obtained as

$$\begin{aligned}\text{Cov}(N_i, N_i) &= E[N_i^2] - E[N_i]^2 \\ &= \begin{cases} 0 & (i = 0) \\ p(1-p) & (i = 1) \\ (1-p)p & (i = 2) \end{cases}\end{aligned}$$

Further, consider $\text{Cov}(N_i, N_j)$ when $i \neq j$. $\text{Cov}(N_i, N_j)$ has the relation as:

$$\text{Cov}(N_i, N_j) = E[N_i N_j] - E[N_i]E[N_j]$$

Then, it is obvious that $\text{Cov}(N_i, N_j)$ equals to $\text{Cov}(N_j, N_i)$. When $i = 0$ and $j \neq 0$, the covariance $\text{Cov}(N_0, N_j)$ is obtained as

$$\begin{aligned}\text{Cov}(N_0, N_1) &= E[N_0 N_1] - E[N_0]E[N_1] \\ &= pE[N_0^2] - pE[N_0]^2 \\ &= 0 \\ \text{Cov}(N_0, N_2) &= E[N_0 N_2] - E[N_0]E[N_2] \\ &= (1-p)E[N_0^2] - (1-p)E[N_0]^2 \\ &= 0\end{aligned}$$

$\text{Cov}(N_i, N_j)$ is evaluated as the following equation when $i \neq 0, j \neq 0$ and $i \neq j$.

$$\begin{aligned}\text{Cov}(N_1, N_2) &= E[N_1(N_0 - N_1)] - E[N_1]E[N_2] \\ &= pE[N_0^2] - E[N_1^2] - E[N_1]E[N_2] \\ &= -p(1-p)\end{aligned}$$

Therefore, the expectation and variance of lead time are obtained as (9) and (10), respectively.

References

- [1] Bitran, G. R. and Dasu, S., 1992. A review of open queueing network models of manufacturing systems, *Queueing Systems*, **12**(1-2), 95–133.
- [2] Bitran, G. R. and Morabito, R., 1996. Open queueing networks: Optimazation and performance evaluation models for discrete manufacturing systems, *Production and Operations Management*, **5**(2), 163–193. 841–843.

- [3] Bitran, G. R. and Tirupani, D., 1988. Multiproducts queueing networks with deterministic routing: Decomposition approach and the notation of inference, *Management Science*, **34**(1), 75–100.
- [4] Bretthauer, K. M., 2000. Optimal service and arrival rate in Jackson queueing networks, *Naval Research Logistics*, **47**(1), 1–17.
- [5] Buzacott, J. and Shanthikumar, J., 1993. *Stochastic Models of Manufacturing Systems*, (Prentice Hall: Englewood Cliffs).
- [6] Combe, M. B. and Boxma, O. J., 1994. Optimization of static traffic allocation policies, *Theoretical Computer Science*, **125**(1), 17–43.
- [7] Govil, M. K. and Fu, M. C., 1999. Queueing Theory in Manufacturing: A Review, *Journal of Manufacturing Systems*, **18**(4), 214–240.
- [8] Gupta, D. and Benjaafar, S., 2004. Make-to-order, make-to-stock, or delay product differentiation? A common framework for modeling and analysis, *IIE Transactions*, **36**(6), 529–546.
- [9] Koole, G., 1999. On the static assignment to parallel servers, *IEEE Transactions on Automatic Control*, **44**(8), 1588–1592.
- [10] Moon, I. and Gallego, G., 1994. Distribution free procedures for some inventory models, *International Journal Operational Research*, **46**(6), 651–658.
- [11] Souza, G. C. and Ketzenberg, M. E., 2002. Two-stage make-to-order remanufacturing with service-level constraints, *International Journal of Production Research*, **40**(2), 477–493.
- [12] Souza, G. C., Ketzenberg, M. E., and Guide, V. D. R., 2002. Capacitated remanufacturing with service level constraints, *Production and Operations Management*, 2002, **11**(2), 231–248.
- [13] Vandaele, N., Boeck, L. D. and Callewier, D., 2002. An open queueing network for leadtime analysis, *IIE Transactions*, **34**(1), 1–9.
- [14] Weng, Z. K., 1999. Strategies for integrating lead time and customer-order decisions, *IIE Transactions*, **31**(2), 161–171.
- [15] Whitt, W., 1983a. Queueing network analyzer, *Bell System Technical Journal*, 1983, **62**(9), 2779–2815.
- [16] Whitt, W., 1983b. Performance of the queueing network analyzer, *Bell System Technical Journal*, **62**(9), 2817–2843.