TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.3, March 2014, pp. 2342 ~ 2349
DOI: http://dx.doi.org/10.11591/telkomnika.v12i3.4520

The Study of Buffer Allocation Problem in Complex Production Line

Xiaoyong Pan*, Jiang Wu, Quanwei Zhang, Dong Lai, Xin Fu, Chen Zhang
NO. 35 Mianxingdonglu, Mianyang, Sichuan, China, 0086-8162418609
*Corresponding author, e-mail: panxy@changhong.com

Abstract

This paper studies the problem of allocating buffers in stochastic production flow lines with product travel time. We build a model that decomposes the production line into the S-B-S (Station-Buffer-Station) subsystems, and use queueing theory to aggregate the subsystems. Experiments are designed for both balanced and unbalanced production lines, and with the computational results, some general rules for the buffer allocation problem are proposed.

Keywords: *buffer, queuing theory, simulated annealing*

1. Introduction

In manufacturing plants, fluctuations in production lines are common and it's necessary to put some buffers between the work stations, or else the production lines will encounter blocking and starving problems frequently. This paper studies the serial production lines where the adjacent work stations have some space in between, which could be used as buffer space. If the total length of the production line is given, the total buffer size is also determined. How to allocate the buffer space is an important issue in the design of the production lines. Koenigsberg(1958) [1] and Conway et. al. (1988) [2] point out that the buffer location and the buffer size are both important factors in the optimization of the production lines.

Most literature models the buffer location problem as a discrete Markov process and use queueing theory to solve the problem, e.g., Ovuworie(1982) [3]. A detailed review of this method is provided by Disney and Konig(1985) [4]. However, exact solutions are always difficult to get by queueing theory, especially when the problem scale grows large, which limit the application of this method. Therefore, a lot of literature applies heuristic and meta-heuristic methods to get reasonable approximate results, such as Buzacott(1967) [5], Whitt(1985) [6], Gershwin(1987) [7], Hillier and So(1991) [8], Gershwin and Schor(2000) [9], Nahas et. al.(2006) [10], Bulgak(2006) [11], and Kim et. al.(2010) [12].

There is also some literature approximates the problem as a continuous Markov process, such as Kouikoglou and Phillis(1997) [13], David et. al.(1989) [14], and da Silva Soares and Latouche(2006) [15]. Continuous model reduces the possible state space greatly, and enables the approximate methods related with gradient.

As far as we know, most research studies the allocation problem only in serial production lines. Our study proposes a model that is also suitable in complex circumstances where there are parallel and serial lines in the same time. In the same time, the research related to our study only considers the objective of maximizing the throughput, with little consideration of the cost of different layout. That is partly because in serial lines, the costs of different layout are generally the same, which is not the case in complex production lines where there are both serial and parallel lines. Also, existing literature considers little about the space limit in the production line and the travel time of work pieces in the buffer. The additional productivity of extra buffer space is decreasing because extra buffer space increases the work piece travel time. Our model absorbs the issues such as labor cost, machine cost, the work piece travel time, et. al. and thus can provide more reasonable and practical solutions.

Received August 26, 2013; Revised October 1, 2013; Accepted October 23, 2013

2. Problem Description

We study the discrete production line where K work stations $M_0, M_1,\ldots, M_K$ are processing work pieces. Between the work stations are buffer areas, and we denote them as $B_1, B_2,\ldots, B_K$. Figure 1 is an example of a complex production line. We assume that all the lines in the parallel lines have the same structure. Suppose M_i is in the stage where there are l_i parallel lines.

When the number of work stations is larger than two, we use decomposition method and study the basic unit of S-B-S (Station-Buffer-Station) sub-systems, as in Figure 2. We define the ith sub-system as T_i, and use two artificial work stations A_i and D_i to simulate the effect of upstream (M_0 to M_{i-1}) and downstream (M_i to M_K) parts of the production line on B_i. Define the capacity of T_i as the sum of the size of B_i and the size of D_i (which equals 1), and denote it as c_i. The width of a work station is W_W and the width of a buffer is W_B. The length of a work station is L_W and the width of a buffer is L_B. The distance of a work station and a buffer is L_M, and the distance of adjacent lines in the parallel lines is W_M. The available work space is limited to a square with length L and width W. The layout should not exceed the available work space. Suppose work station M_i needs W_i workers, and the total number of available workers is E. The worker number should not exceed the available number. The cost of one worker is P_W, and the cost of one machine is P_M, which combines the discounted value of the machines and the operating cost such as electric, maintenance, et. al.

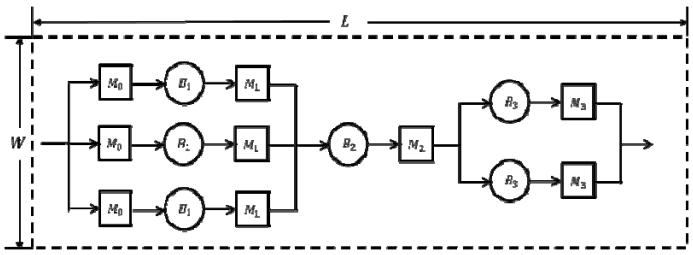


Figure 1. An Example of a Complex Production Line

Each work piece enters the production line from M_0, flows through work stations and buffer areas in sequence, and finally leaves the production line from M_K. Due to the fluctuation of the production processing time, 'blocking' or 'starving' may happen. If any A_i finishes its work and B_{i-1} has no available work pieces, then T_i is 'blocking'; if any D_i finishes its work and B_{i+1} is full, then T_i is 'blocking'.

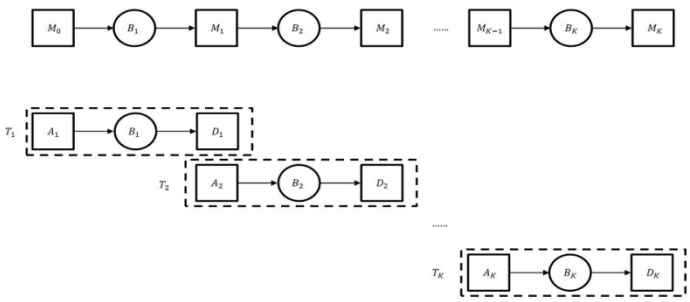


Figure 2. The Decomposition Technique of a Serial Production Line

The sub-systems are modeled as M/M/1/N queueing systems. Assume the processing time of each work station is exponentially distributed (the service rate of the ith work station is defined as μ_i). The work pieces pass through the buffer areas for certain time, which is

proportional to c_i. Define μ_i as the service rate of ith work station, and T as the time consumed to move the distance of the one work piece length. We have:

$$\bar{\mu}_i = \frac{1}{1/\mu_i + c_i T}, i = 1, 2, ..., K.$$

Let a_i be the arrival rate of work pieces, and b_i be the departure rate. Define:

$$\rho_i = \frac{a_i l_{i+1}}{b_i l_i}, i = 1, 2, ..., K.$$

Where l_{K+1} equals l_K. By queueing theory, the probability that there are k work pieces in T_i is:

$$P_{k,i} = \frac{(1-\rho_i)\rho_i^k}{1-\rho_i^{c_i+1}}, i = 1, 2, ..., K.$$

Consider the situations of 'blocking' and 'starving', we have:

$$\frac{1}{a_i} = \frac{1}{\bar{\mu}_{i-1}} + \frac{P_{0,i-1}}{a_{i-1}}, i = 2, 3, ..., K-1.$$

$$\frac{1}{b_i} = \frac{1}{\bar{\mu}_i} + \frac{P_{c_{i+1},i+1}}{a_{i+1}}, i = 1, 2, ..., K-1.$$

The first work station will never be 'waiting', and the last work station will never be 'blocking'. We have:

$$a_i = \bar{\mu}_0.$$

$$b_K = \bar{\mu}_K.$$

The throughput of the production line is:

$$X = a_K \left(1 - \frac{(1-\rho_i)\rho_i^{c_i}}{1-\rho_i^{c_i+1}}\right).$$

The labor cost is:

$$C_W = P_W \sum_{i=0}^{K} W_i l_i.$$

The capital cost is:

$$C_M = P_M \sum_{i=0}^{K} l_i.$$

The total cost is:

$$C = C_M + C_W.$$

The buffer allocation problem can be formulated as:

$$\max \frac{X(c_1,...,c_K,l_1,...,l_K)}{C(l_1,...,l_K)}.$$

s.t.

The total number of workers should not exceed the available number, that is:

$$\sum_{i=0}^{K} W_i l_i \le E.$$

The layout of the production line should not exceed the available work area, that is:

$$(K+1)L_W + \sum_{i=1}^{K} c_i L_B + 2KL_M \le L.$$

$$\mathrm{Max}\{L_0, L_1, ..., L_K\} \times (\mathrm{Max}\{W_B, W_W\} + W_M) - W_M \le W.$$

All the variables should be positive integers, that is:

$$c_i \in Z, c_i \ge 0, i = 1, ..., K.$$

$$l_j \in Z, l_j \ge 0, j = 0, ..., K.$$

3. Research Method

We use simulated annealing algorithm to solve the problem. The core idea of simulated annealing is quite similar to that of thermodynamic principles of crystallization and metal cooling and annealing. A large number of liquid molecules move freely at high temperature, but the mobility will decease as the temperature cools down. Atoms are arranged in rows and form a pure crystal, which is the state with the least energy. However, if we decease the temperate very quickly, the liquid metal is said to be 'quenched', and will stop at a high energy polycrystalline state or amorphous state. Therefore, the essence of getting a perfect crystal is to decrease the temperate slowly to allow the atoms redistribute their positions before they lost their mobility. This is called annealing in technical definitions.

Simulated annealing method is a meta-heuristic technique to solve large-scale global optimization problems, especially when there are many local optimal points, and the searching space is very large. In our problem, we have an objective which is not a simple combination of the *K* variables. The searching space is so large that it's almost impossible to enumerate all possible solutions. Therefore, we adopt simulated annealing in our problem.

According to the principle of annealing, the solid is heated to a sufficiently high temperature, and then is cooled down slowly. When heated, the solid particles become disordered in position, with very high internal energy. When cooled down slowly, they are ordered and finally reach equilibrium, with the least internal energy. The Metropolis criterion describes this process in mathematical form, i.e., the probability of particles to reach balance state is $e^{-\delta E/(kT)}$ at temperature *T*, where *E* is the internal energy, and *k* is Boltzmann constant. When we use simulated annealing to solve combinatorial optimization problems, the internal energy *E* is replaced with the objective function value *X*, and the temperature *T* is replaced with the control parameter *t*. Starting from an initial solution, we 'generate a new solution, calculate the objective difference, and accept or reject' iteratively, and gradually decay the *t* value. When the algorithm terminates, we get an approximate solution for the optimization problem. The annealing process is controlled by a cooling process table, which includes the control parameter *t*, the decay factor δt, the iteration number *L* and the stopping condition *S* for each *t*. The solution generating process is:

a. Generate a new feasible solution from the current solutions. For the ease of calculation, the new solutions are generated by simple transformation, such as edge-exchange and node-exchange methods.
b. Calculate the difference of the objective value with the older ones.
c. Determine whether the new solution *S'* is accepted. The most commonly used criterion is Metropolis principle: if $\delta t' < 0$, we accept the new solution and use it to replace the old one *S*, or else we accept *S'* only with a probability $e^{-\delta t'/T}$.
d. If the new solution is accepted or rejected, the iteration is completed.

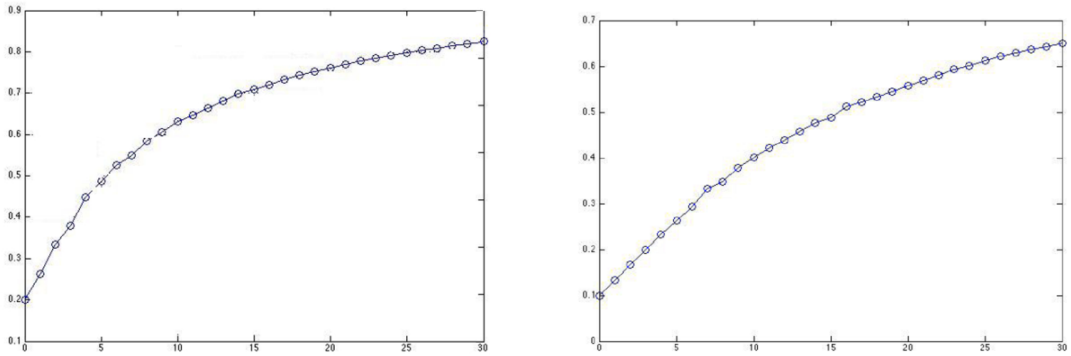
4. Results and Analysis

Similar to Conway et. al. (1988) [2], we design experiments under circumstances of both balanced and unbalanced production lines. The objective is to maximize the throughput. The production line is in a 'pushing' way, that is, the first work station will never be 'starving', and the last work station will never be 'blocking'. The model and the algorithm are coded with Matlab.

To simplify the problem, we assume that the size of work stations and buffers are the same, while the length of the work station equals the distance of the adjacent buffers and work stations, and the width equals the distance of the adjacent parallel lines. That is, $L_W = L_B = L_M$, and $W_W = W_B = W_M$.

4.1. Balanced Production Line

Assume the processing speed of all work stations is one work piece per unit time. We study the situations where there are 5 and 10 work stations in a serial line respectively. The width is limited to $5W_M$. The maximum objective of the production line vs. the total length is demonstrated in Figure 3. It's clear that the benefit of the buffer space is decreasing as the total length increases. Hatcher (1969) [16] shows that 10 buffer space is adequate for production line in most circumstances. Our research demonstrated that it's not always a good idea to design 10 buffer spaces in total. As we can see from Figure 3, when there are 5 work stations, the throughput of production line with 10 buffer spaces can reach 60% of ideal throughput (when there are no fluctuation), but the number becomes 40% when there are 10 work stations. We recommend designers to calculate the different proposals first, together with the cost of increasing the production line length, before making a decision.



5 work stations 10 work stations

Figure 3. The Objective vs. the Buffer Space for Balanced Production Lines

The buffer allocation results are listed in Table 1 and Table 2. As we can see, the optimal allocations are symetrial in most cases, with a little more allocated in the center.

Table 1. The Buffer Allocation Results of the Balanced Production Lines with 5 Work Stations

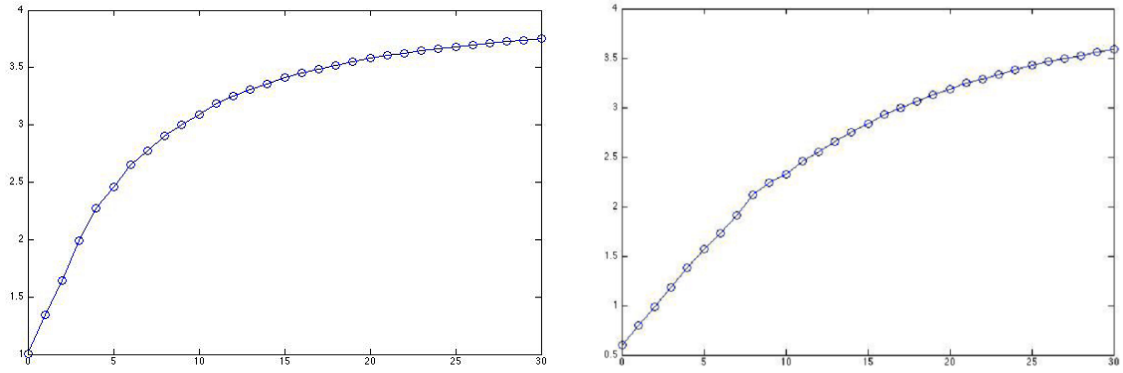
Buffer Space	c_1	c_2	c_3	c_4	l_1	l_2	l_3	l_4	l_5
15	3	5	4	3	1	1	1	1	1
16	3	5	5	3	1	1	1	1	1
17	4	5	4	4	1	1	1	1	1
18	4	5	5	4	1	1	2	1	1
19	4	6	5	4	1	1	2	1	1
20	4	6	6	4	1	1	2	1	1
21	5	6	6	4	1	2	2	2	1
22	5	6	6	5	2	2	2	2	2
23	5	7	6	5	2	2	3	2	2
24	5	7	7	5	2	2	3	2	2
25	6	7	7	5	2	2	3	2	2
26	6	7	7	6	2	2	3	2	2
27	6	8	7	6	2	3	3	3	2
28	6	8	8	6	2	3	3	3	2
29	7	8	8	6	2	3	3	3	2
30	7	8	8	7	2	3	3	3	2

Table 2. The Buffer Allocation Results of the Balanced Production Lines with 10 Work Stations

Buffer Space	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8	l_9	l_{10}
15	1	2	2	2	1	2	2	2	1	1	1	1	1	2	1	1	1	1	1
16	1	2	2	2	2	2	2	2	1	1	1	1	1	2	1	1	1	1	1
17	1	2	2	2	3	2	2	2	1	1	1	1	1	2	1	1	1	1	1
18	1	2	2	3	2	3	2	2	1	1	1	1	1	2	1	1	1	1	1
19	2	2	2	2	3	2	2	2	2	1	1	1	2	2	2	1	1	1	1
20	2	2	2	3	2	3	2	2	2	1	1	1	2	2	2	1	1	1	1
21	2	2	2	3	3	3	2	2	2	1	1	1	2	2	2	1	1	1	1
22	2	2	3	3	2	3	3	2	2	1	1	1	2	2	2	1	1	1	1
23	2	2	3	3	3	3	3	2	2	1	1	1	2	2	2	1	1	1	1
24	2	3	3	3	3	3	3	2	2	1	1	2	2	2	2	1	1	1	1
25	2	3	3	3	3	3	3	3	2	1	1	2	2	2	2	1	1	1	1
26	2	3	3	3	4	3	3	3	2	1	1	2	2	2	2	1	1	1	1
27	2	3	3	4	3	4	3	3	2	1	1	2	2	2	2	2	1	1	1
28	2	3	4	3	4	3	4	3	2	1	1	2	2	2	2	2	2	1	1
29	2	3	4	4	3	4	4	3	2	1	1	2	2	2	2	2	2	1	1
30	3	3	3	4	4	4	3	3	3	1	1	2	2	2	2	2	2	1	1

4.2. Unbalanced Production Line

The unbalanced production lines are much more complicated than balanced ones. We study two examples. The first example has 5 work stations, and the processing speed of each work station is 4, 7, 10, 7 and 4 work pieces per unit time. The second example has 10 work stations, and the processing speed of each work station is 8, 4, 7, 5, 10, 7, 7, 4, 7 and 5 work pieces per unit time. The maximum throughput of the production line vs. the total buffer size is demonstrated as Figure 4. As we can see, the curve is very similar to Figure 3, indicating that the effects of total buffer size on throughput are similar in both balanced and unbalanced lines.



5 work stations 10 work stations

Figure 4. The Objective vs. the Buffer Size for Unbalanced Production Lines

The buffer allocation results are listed in Table 3 and Table 4. Bulgak et. al. (1995) [17] shows that the buffer size is positively related to the congestion possibility. Our research indicates that, the situation is more compliated when we consider the complex production line, and the optimal buffer allocation schemes often take very irregular forms.

Table 3. The Buffer Allocation Results of the Unbalanced Production Lines with 5 Work Stations

Buffer Space	c_1	c_2	c_3	c_4	l_1	l_2	l_3	l_4	l_5
15	6	4	3	2	1	1	1	1	1
16	7	4	2	3	1	1	1	1	1
17	7	5	2	3	1	1	1	1	1
18	8	4	3	3	1	1	1	1	1
19	8	5	3	3	1	2	1	1	1
20	9	5	3	3	1	2	2	1	1
21	10	5	3	3	1	2	2	1	1
22	10	6	3	3	2	2	2	2	1
23	11	6	3	3	2	2	3	2	1
24	12	6	3	3	2	2	3	2	1
25	12	6	3	4	2	2	2	2	1
26	13	6	3	4	2	2	2	2	2
27	14	6	3	4	2	2	2	2	2
28	14	7	3	4	2	2	2	2	2
29	15	7	3	4	2	2	2	2	2
30	16	7	3	4	2	2	2	2	2

Table 4. The Buffer Allocation Results of the Unbalanced Production Lines with 10 Work Stations

Buffer Space	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8	l_9	l_{10}
15	1	1	2	2	2	1	2	2	2	1	1	1	1	1	1	1	1	1	1
16	1	2	2	2	2	1	2	2	2	1	1	1	1	1	1	1	1	1	1
17	1	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1
18	1	2	3	2	2	1	2	3	2	1	1	1	1	1	1	1	1	1	1
19	1	2	3	2	2	2	2	2	3	1	1	1	2	2	1	1	1	1	1
20	1	2	3	2	2	2	2	3	3	1	1	1	2	2	1	1	1	1	1
21	1	2	3	3	2	2	2	3	3	1	1	1	2	2	1	1	1	1	1
22	1	3	3	3	2	2	2	3	3	1	1	2	2	2	2	1	1	1	1
23	1	3	3	3	2	2	2	4	3	1	1	2	2	2	2	1	1	1	1
24	1	3	3	3	2	2	3	3	4	1	1	2	2	2	2	1	1	1	1
25	1	3	4	3	2	2	2	4	4	1	1	2	2	2	2	1	1	1	1
26	1	3	4	3	3	2	2	4	4	1	2	2	2	2	2	1	1	1	1
27	1	3	4	4	2	2	3	4	4	1	2	2	2	2	2	2	1	1	1
28	1	4	4	3	3	2	3	4	4	2	2	2	2	2	2	2	2	1	1
29	2	3	4	3	3	2	3	4	5	2	2	2	2	2	2	2	2	1	1
30	2	3	4	4	3	2	3	4	5	2	2	2	2	2	2	2	2	1	1

4. Conclusion

The problem of allocating buffers in stochastic production flow lines with product travel time was studied in this paper. The using “queueing theory” to aggregate the subsystems, a model that decomposes the production line into the S-B-S (Station-Buffer-Station) subsystems has been developed. Experiments have been designed for both balanced and unbalanced production lines, and with the computational results, some general rules for the buffer allocation problem have been proposed.

References

[1] Koenigsberg E. Cyclic queues. *Journal of the Operational Research Society*. 1958; 9(1): 22-35.
[2] Conway R, Maxwell W, McClain JO, Thomas LJ. The role of work-in-process inventory in serial production lines. *Operations Research*. 1988; 36(2): 229-241.
[3] Ovuworie GC. An unreliable series production line: keeping it running with buffer stocks. *The International Journal of Production Research*. 1982; 20(5): 607-617.
[4] Disney RL, Konig D. Queueing networks: a survey of their random processes. *SIAM review*. 1985; 27(3): 335-403.
[5] Buzacott JA. Automatic transfer lines with buffer stocks. *International Journal of Production Research*. 1967; 5(3): 183-200.
[6] Whitt W. The best order for queues in series. *Management Science*. 1985; 31(4): 475-487.

[7] Gershwin SB. An effcient decomposition method for the approximate evaluation of tandem queues with finite storage space and blocking. *Operations Research*. 1987; 35(2): 291-305.
[8] Hillier FS, So KC. The effect of machine breakdowns and interstage storage on the performance of production line systems. *International Journal of Production Research*. 1991; 29(10): 2043-2055.
[9] Gershwin SB, Schor JE. Effffient algorithms for buffer space allocation. *Annals of Operations Research*. 2000; 93(1-4): 117-144.
[10] Nahas N, Ait-Kadi D, Nourelfath M. A new approach for buffer allocation in unreliable production lines. *International Journal of Production Economics*. 2006; 103(2): 873-881.
[11] Bulgak AA. Analysis and design of split and merge unpaced assembly systems by metamodelling and stochastic search. *International Journal of Production Research*. 2006; 44(18-19): 4067-4080.
[12] Kim S, Cox JF, Mabin VJ. An exploratory study of protective inventory in a re-entrant line with protective capacity. *International Journal of Production Research*. 2010; 48(14): 4153-4178.
[13] Kouikoglou VS, Phillis YA. A continuousow model for production networks with finite buffers, unreliable machines, and multiple products. *International Journal of Production Research*. 1997; 35(2): 381-397.
[14] David R, Xie X. Properties of continous models of transfer lines with unreliable machines and finite buffers. *IMA Journal of Management Mathematics*. 1989; 2(4): 281-308.
[15] da Silva Soares A. Latouche G. Matrix-analytic methods for fluid queues with finite buffers. *Performance Evaluation*. 2006; 63(4): 295-314.
[16] Hatcher JM. The effect of internal storage on the production rate of a series of stages having exponential service times. *AiIE Transactions*. 1969; 1(2): 150-156.
[17] Bulgak AA, Diwan PD, Inozu B. Buffer size optimization in asynchronous assembly systems using genetic algorithms. *Computers and Industrial Engineering*. 1995; 28(2): 309-322.