

---

## **Cost Minimization for Unstable Concurrent Products in Multi-stage Production Line Using Queueing Analysis**

---

*Submitted 15/11/19, 1<sup>st</sup> revision 15/12/19, 2<sup>nd</sup> revision 23/01/20, accepted 21/02/20*

Chanintorn Jittawiriyankoon<sup>1</sup>, Vilasinee Srisarkun<sup>2</sup>

**Abstract:**

**Purpose:** *The paper copes with the queueing theory for evaluating a multi-stage production line process with concurrent goods. The intention of this article is to evaluate the efficiency of products assembly in the production line.*

**Design/Methodology/Approach:** *To elevate the efficiency of the assembly line it is required to control the performance of individual stations. The arrival process of concurrent products is piled up before flowing to each station. All experiments are based on queueing network analysis.*

**Findings:** *The performance analysis for unstable concurrent sub-items in the production line is discussed. The proposed analysis is based on the improvement of the total sub-production time by lessening the queue time in each station.*

**Practical implications:** *The collected data are number of workers, incoming and outgoing sub-products, throughput rate, and individual station processing time. The front loading place unpacks product items into concurrent sub-items by an operator and automatically sorts them by RFID tag or bar code identifiers. Experiments of the work based on simulation are compared and validated with results from real approximation.*

**Originality/Value:** *It is an alternative improvement to increase the efficiency of the operation in each station with minimum costs.*

**Keywords:** *Concurrent items, mathematical models, performance evaluation, production line, queueing theory, RFID, ROI.*

**JEL Codes:** *G39, K22, L22, M10, O38.*

**Paper Type:** *Research Paper.*

**Acknowledgments:**

*This research and resulting contribution are results of Assumption University of Thailand. The university partially supports financially the publication.*

---

<sup>1</sup>Graduate School of Advanced Technology Management, Assumption University, Thailand, [pct2526@yahoo.com](mailto:pct2526@yahoo.com)

<sup>2</sup>Martin de Tours School of Management and Economics, Assumption University, Thailand, [vilasineesrs@au.edu](mailto:vilasineesrs@au.edu)

## 1. Introduction

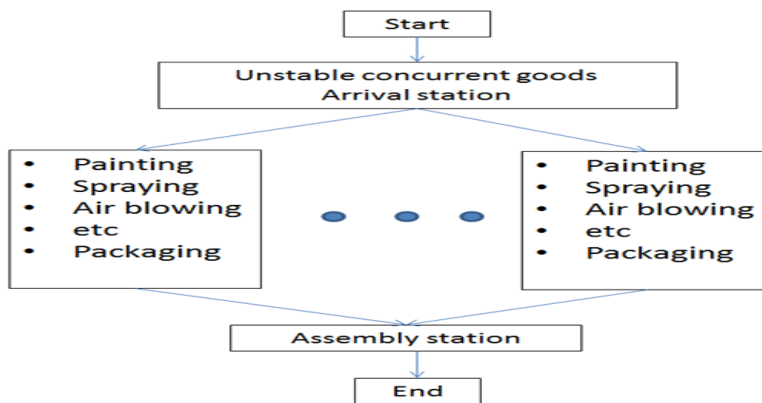
The queuing network application (Owoloko *et al.*, 2015) in production employs simulation process to investigate activities of the course to elevate productivity of work processes by lowering the mean queue time. As an engineer we can relate the queuing network to point out the impact of turn-around time which affects product's arrival time management, the efficiency of machines (resources) usage, and operation's productivity. The performance investigation involves statistical analyses, turn-around time estimation, mean waiting time at the station, and state space fust (Xuan *et al.*, 2016). Not to mention the queuing concept can help analyze the multi-stage assembly line based on representative network models. On the other hand, the approximation method is applied for the same network but with a reducible complication of steady-state space. The validation of our simulation results thus is asserted by associating with values based on estimation. These values are leading to refine both proficiency and productivity at individual station in the assembly line. This paper focuses on a study of queuing application related to unstable concurrent goods in the multi-stage production line. Concurrent process is an operational model where various stations in the production can manufacture items all together for higher performance and quality. Concurrent operation reflects processes which are being executed simultaneously.

The company provides products for distribution using multi-stage assembly flow station (Kasilingam and Seshadhri, 2015). The company claims to inspect the resourcefulness and other constraints in production line. The intention of this article is that we propose queuing analytical model to the multi-stage assembly line to evaluate the efficiency of process flow. The system where the stage comprises parallel robots and the subsequent stations (such as colorizing, assembly, wrapping, etc.) is studied. Figure 1 depicts flow in the assembly production line to complete concurrent items.

## 2. Queueing Methodology

The competitive race between companies versus their rivalries is an essential key to either launch innovative goods or to move on with the intense enhancement of existing ones. The former is costly somehow, yet fundamentally requires the replacement of several machines, and not to mention switching time while the latter eases or analyzes the employment of existing powerful resources. The paper concentrates on the cost-effective latter mechanism based on the optimal efficiency of each station in the parallel production line. Computations based on queuing concept and simulations are taken to extract performance metrics which can be meaningful to the productivity and overall improvement. Queueing network model is statistical measurement of queue length, or waiting time. It involves with the iteration and occurrences with the foundation of service time and disciplines which are in random fashion. In such a model, performance metrics can be investigated (Camaj *et al.*, 2015).

**Figure 1.** The process flow in the multi-stage production line with concurrent items



Queuing concept coincides largely with the operational research as outputs can support the decision-making in quantifying compulsory machines in the production. Besides, queuing concept helps cater analytical models to measure system behavior, for example, an investigation of network traffic congestion (Aman *et al.*, 2017). The analysis centers on how to deliver multicast traffic over mobile network. Multi-type queuing models contribute to implement resources more efficiently concerning restricted resources but severe requests, and help manage outpatients in the hospital (Zhu and Qi, 2016). In production, turnaround time is a key to identify work in progress, and can be a decision-making based on queuing analyses. The simulation tool has a great impact in determining turnaround time compared to the approximation method (Amaran *et al.*, 2014). The utilization of the simulation tool is somehow uncharted because simulation per se is not simple to familiarize and target problems as well as not to mention it is a time-consumption modeling. On the other hand, an approximation approach based on queuing theory (Khalili *et al.*, 2016) is overutilized to determine the efficiency of the production system.

Several steps are applied to analyze the arriving function at the station. As a result, arrival traffic observes Poisson distribution function. A parallel based  $c$  servers queuing model  $M/M/c$  (Shanmugasundaram and Banumathi, 2017; Yue, Yue, and Zhao, 2014) has been proposed. Let us examine the  $c$ -server-parallel queuing system where our concurrent products arrive in a Poisson fashion with rate  $\lambda$ . These arrival products are recognized as primary items. Each primary item comes into the arrival station for identifying a number of split secondary items. We assume that the individual primary items take no latency based on RFID technology (Chen *et al.*, 2013) for splitting at the station. The secondary items from the associated primary product are called “*folks*”. All folks move to parallel production line at once. It is independently manufactured folks, except the impact of queue while waiting for production line processes. The production time of all secondary items incorporates the exponential distribution with rate  $\mu$ . If a secondary item concludes the production process, it delays at the assembly station until the completion of all folks.

After all folks' integration, they are assembled into the primary product, are ready to send out. The period from coming into the arrival station to the reunion of all folks at the assembly station refers to a turnaround time. The primary product ( $A_i$ ) is split into  $A_{ij}$  folks, and it evolves the concurrency ( $a_{ij}$ ) throughout the production service. The concurrency  $a_{ij}$  is unstable and varies. Representative matrix of any primary items ( $A_i$ ) is a finite set and contains  $a_{ij}$  where  $\{a_{ij} \geq 0\}$ . An example of two primary product matrices ( $A_1$  and  $A_2$ ), where  $A_1 = [3 \ 6 \ 11 \ 1]$  and  $A_2 = [8 \ 0 \ 6 \ 10]$ . Then we compute performance metrics based on  $M/M/c$  model as follows. The model discerns exponential service ( $\mu$ ) and poisson arrival ( $\lambda$ ) with first-in first-out discipline. The key metrics of according model are summarized by:

1. Utilization factor  $\rho = \frac{\lambda}{c\mu}$ , where  $\lambda \leq c\mu$ .
2. Number of secondary items waiting in queue

$$L_q = \left[ \frac{c\mu\lambda}{c! c^c (c\mu - \lambda)^2} \rho^c \right] \left[ \frac{1}{\sum_{i=0}^{c-1} \frac{c\mu\lambda^c}{c! \mu^c (c\mu - \lambda)} + \frac{\lambda^i}{i! \mu^i}} \right]$$

3. Number of secondary items in the production line

$$L_s = c\rho + L_q$$

4. Waiting time spent in queue

$$W_q = \frac{L_s - c\rho}{\lambda}$$

5. Waiting time spent in production line

$$W_s = \frac{L_q + c\rho}{\lambda}$$

6. Turnaround time = *production time* +  $W_q$

7. Throughput = the amount of primary items passing through a production system (assembly station) in an observation period of time.

### 3. Application Model

The situation for the clarification to conquer a state space explosion is that the arrival rate of primary products must be lower than the processing rate of secondary

items in production line. Besides, other significant features essential for the simulation are number of parallel service stations, distribution function, arrival rate, service rate, and unstable concurrent matrices of primary products at the arrival station. Queuing concept models to take more variables and iterate for calculating results (Vilaplana *et al.*, 2014).

The arrival pattern of concurrent primary items is one in which we can complete arrivals by considering persistent interval of time between consecutive arrival. A Poisson fluctuation of arrival relates to random arrival (Barabesi and Pratelli, 2014). In Poisson process items attain after interval which exclusively is exponential stream. The Poisson process is significant as it is an analytical model of many engineering systems, and is defined by a variable - the average arrival rate ( $\lambda$ ) (Tse, 2014). Other essential arrival patterns (Chydzinski and Mrozowski, 2016) include batch arrival, unstable concurrent arrival, and planned arrival (i.e., an arrival rate depends on the set hour in a day). Arrival stream is how arrivals are dispersed according to time (what is the probability function of interval time between individual arrival – interarrival time). Besides, the arrival is an unstable (changeable) or a finite (fixed) set of items.

In the paper we assume that service time ( $\mu$ ) for secondary items at production line is liberated, and differs from the arrival stream but follows exponential function (Marin and Buló, 2011; Ismail *et al.*, 2018). Service time is an explanation of how long processes are needed, and whether servers are in parallel (common single queue for servers) or serial (individual server with a distinct queue) fashion. Occasionally, a pre-emptive mechanism is considered (server halts servicing a task at once for an “urgent” item).

The queuing network lets unstable concurrent items come into the arrival station and evacuate at the assembly station as shown in Figure 1. After splitting at no time then secondary items wait for production services with FIFO discipline. Three lines of production hold identical branching probability. The unstable primary products arrive with rate of  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$ , and  $\lambda_4$  items in an hour (Poisson function). Concurrent items from four sources ( $A_1$ ,  $A_2$ ,  $A_3$ , and  $A_4$ ) are determined as  $A_1 = [3 \ 4 \ 2 \ 1]$ ,  $A_2 = [6 \ 0 \ 6 \ 0]$ ,  $A_3 = [4 \ 6 \ 1 \ 2]$ , and  $A_4 = [8 \ 0 \ 6 \ 0]$ . The multi-stage production processing times for three lines observe the exponential function with mean of 2, 3, 4, and 5 minutes (service dependent) for four concurrent items, respectively. The event-driven simulation tool based on (EZSim, 2019) demonstrates performance parameters, for instance, utilization (UTL), throughput (THR), mean queue length (MQL), mean waiting time in queue (MWT), etc.

We simulate three cases associated with regular operation in the factory. These cases are assumed to be  $\lambda_1=\lambda_2=\lambda_3=\lambda_4=1$  for case one,  $\lambda_1=1$ ,  $\lambda_2=2$ ,  $\lambda_3=3$ , and  $\lambda_4=4$  for case two, and  $\lambda_1=4$ ,  $\lambda_2=3$ ,  $\lambda_3=2$ , and  $\lambda_4=1$  for case three, correspondingly. In the approximation calculation, we apply an average concurrency approach described in

(Jittawiriyankoon, 2019) and compute upon mathematic equations listed in section 2. Simulation and approximation results from experimental three cases are summarized in the following Tables 1-6. As a load balancer is put into focus, incoming requests (loads) from users traverse through the network before reaching the service units. That points out all the load traffic is hitting, where the load balancer is considered. The difference between the maximum load and minimum load in the network environment is defined as load range. Load range is also calculated for the future improvement in the specific area of production load balancing account. In case one, the probability of queue is lower than the utilization. It evidently points out that upon primary products come into the arrival station, they are being processed. They do not wait at the arrival station for too long. Multi-stage production line facility serves the secondary products effectively as shown in Table 1.

Considering load balancing in each production line, so far it is not yet a problem. In the approximation, results somehow differ from those collected upon simulation as depicted in Table 2. However, the approximation triumphs over the state space complexity as well as computation cost. Processing capacity as tabulated in Table 3 grows as the excessive waiting time at the arrival station causes a congestion due to heavy arrival process in case two. By the way this processing capacity extends higher ROI (profit upon investment) for the company and shareholders. More nation-wide production lines (channels) can be introduced with multi-stage facility.

Efficient production management brings to progress. The approximation results of case two are displayed in Table 4. Note that in this case the approximation values do not align much with simulation due to the adjustment of average concurrency in the computation. Based on this case, it can be encouraged that queuing concept is invalid to be adopted. From Table 5 and 6 representing results from case three, it is detected that once the arrival process is decreasing compared to case two, while the processing capacity in the production line remains unchanged. However, this study case brings unbalance load problems due to the load range figure, and may not embrace for the multi-stage production model. Furthermore, we spot the limitation of queueing theory application for case three as well of which results from the approximation are sensitive and not close.

**Table 1.** Simulation Results (case one)

Simulation				
Line	MQL	MWT	THR	UTL
1	0.11	1.84	30	0.22
2	0.15	2.23	34	0.27
3	0.06	0.84	37	0.26
Load range	0.09			

**Table 2.** *Approximation Results (case one)*

Approximation				
Line	MQL	MWT	THR	UTL
1	0.125	2.857	36.5	0.31
2	0.125	2.857	36.5	0.31
3	0.125	2.857	36.5	0.31
Load range	0			

**Table 3.** *Simulation Results (case two)*

Simulation				
Line	MQL	MWT	THR	UTL
1	2.24	10.42	93	0.79
2	1.85	11.42	79	0.76
3	9.3	37.95	103	0.97
Load range	7.45			

**Table 4.** *Approximation Results (case two)*

Approximation				
Line	MQL	MWT	THR	UTL
1	14.55	38.3	113.7	0.93
2	14.55	38.3	113.7	0.93
3	14.55	38.3	113.7	0.93
Load range	0			

**Table 5.** *Simulation Results (case three)*

Simulation				
Line	MQL	MWT	THR	UTL
1	0.23	1.35	85	0.44
2	0.46	2.82	78	0.51
3	0.57	3.18	87	0.52
Load range	0.34			

**Table 6.** *Approximation Results (case three).*

Approximation				
Line	MQL	MWT	THR	UTL
1	2.2	7.5	62.6	0.7
2	2.2	7.5	62.6	0.7
3	2.2	7.5	62.6	0.7
Load range	0			

#### 4. Conclusion

The article applies simulation and queuing approach to investigate the efficiency of three production lines. The study has found that queuing concept can compute production system metrics such as mean queue length, utilization, and throughput. The production includes color spray, air-blowing, assembly, packing and etc. Arrival and departure distribution is analyzed for input parameters (for both simulation and approximation) to identify whether they follow Poisson and Exponential fashion. The calculated results from queuing application are compared to simulation results. We found that accuracy between results using simulation and approximation is high only for the case that the probability of queue does not exceed the utilization factor. To recover production time by balancing and subsidizing the queue time in each line is advised. Another development is to escalate the efficiency of the processing capacity (operation) in each line at least cost, especially for the case of heavy arrival of primary products as well as the preparation for future growth in operation. The improvement with an increment of the operational efficiency in each station must follow cost-effective analysis to minimize all costs.

To lessen the queuing time and upsurge the efficiency will reflect an improvement. One potential development is to balance the load in each line to obtain maximum equal utilization. We obtain the efficiency of the line which is about 25% in case one and above 75% in the remaining two cases. We also found that the load imbalance arises in case two (load range is 7.45). Hence, the company should iron out this problem to earn a better being efficiency. The study per se is beneficial for the company regarding future improvement by accounting all insightful information relevant to the production line performance. It is at ease and more practical for the executives of company to strategize ahead the production process in near future. The study emphasizes multi-stage production line in the industry by applying queuing concept, as such the company can avoid the complex simulation modeling and save computational cost. However, we can conclude that queuing concept is usable in case of low arrival traffic analysis. It is apparent to confirm that the results may diverge due to the characteristics of input parameters and distribution functions. It is remarkable that the queuing model application helps analyze the basic measures, and the anticipated values of the behavior of the operation, and the production process.

#### References:

- Aman, A.H.M. et al. 2017. Simulation Analysis for Multicast Context Delivery Network Mobility Management. *Indonesian Journal of Electrical Engineering and Informatics*, 5(4), 390-394.
- Amaran, S., Sahinidis, N., Sharda, B., Bury, S. 2014. Simulation optimization: A review of algorithms and applications. *Annals of Operations Research*, 12(4), 301-333.
- Barabesi, L., Pratelli, L. 2014. A note on a universal random variate generator for integer-valued random variables. *Statistics and Computing*, 24(4), 589-596.



- Camaj, J., Dolinayova, A., Lalinska, J., Bariak, M. 2015. The Technological Problem of Simulation of the Logistics Centre. Proceeding of the 17<sup>th</sup> International Conference on Logistics and Transportation, 2613-2617.
- Chen, Y.C., Chen, R.S., Ye, C.P., Sun, H.M. 2013. RFID Application on Manufacturing Process Control in Semiconductor Industry. Proceedings of the World Congress on Engineering, 2, 941-943.
- Chydzinski, A., Mrozowski, P. 2016. Queues with Dropping Functions and General Arrival Processes. PLoS ONE, 11(3), e0150702.
- EZSIM. <http://www.bkshoshnevis.com/> (23.12.2019).
- Ismail, A. et al. 2018. Exponential MLWDF (EXP-MLWDF) Downlink Scheduling Algorithm Evaluated in LTE for High Mobility and Dense Area Scenario. International Journal of Electrical and Computer Engineering, 8(3), 1618-1628.
- Jittawiriyankoon, C. 2019. Performance Evaluation of Proposed Load Balancing Algorithm with Unstable Concurrent Programs. Indonesian Journal of Electrical Engineering and Computer Science, 14(3), 1452-1459.
- Jittawiriyankoon, C., Srisarkun, V. 2017. Performance Evaluation and Prediction of Parallel Big Data using MOA. International Journal of Engineering and Technology, 9(2), 865-870.
- Kasilingam, S., Seshadhri, V. 2015. Design of Flow Control Mechanism Model for Markovian Multi Stage Queueing Network. World Engineering and Applied Sciences Journal, 6(3), 191-197.
- Khalili, S., Mohammadzade, H., Fallahnezhad, M.S. 2016. A new approach based on queueing theory for solving the assembly line balancing problem using fuzzy prioritization techniques. Scientia Iranica, 23, 387-398.
- Marin, A., Buló, S.R. 2011. Explicit solutions for queues with hypo-exponential service time and applications to product-form analysis. Proceedings of the 5<sup>th</sup> International ICST Conference on Performance Evaluation Methodologies and Tools, 166-175.
- Owoloko, A.E., Ayoku, A.S., Adeleke, O.J., Edeki, S.O., Owoloko, I. 2015. On the application of the Open Jackson queueing network. Global Journal of Pure and Applied Mathematics, 11(4), 2299-2313.
- Shanmugasundaram, S., Banumathi, P. 2017. A Comparative Study on M/M/1 and M/M/C Queueing Models Using Monte Carlo Simulation. Global Journal of Pure and Applied Mathematics, 13(11), 7843-7853.
- Tse, K.K. 2014. Some Applications of the Poisson Process. Applied Mathematics, 5, 3011-3017.
- Vilaplana, J. et al. 2014. A queueing theory model for cloud computing. The Journal of Supercomputing, 69(1), 492-507.
- Xuan, S., Yang, W., Dong, H., Zhang, J. 2016. Performance Evaluation Model for Application Layer Firewalls. PLoS One, 11(11), 1-15.
- Yue, D., Yue, W., Zhao, G. 2014. Analysis of an M/M/c Queueing System with Impatient Customers and Synchronous Vacations. Journal of Applied Mathematics, Article ID 893094, 1-11.
- Zhu, M., Qi, E. 2016. A Multi-Type Queueing Network Analysis Method for Controlling Server Number in the Outpatient. The open Automation and Control Systems Journal, 8, 21-33.