



# Detection of Bottleneck in Manufacturing Supply Chain using Specific KPI

Yoshiatsu Kawabata<sup>1</sup>, Yuta Hosokawa<sup>1</sup>, and Katsuhide Fujita<sup>1</sup>

Graduate School of Engineering, Tokyo University of Agriculture and Technology  
s198355z@st.go.tuat.ac.jp, hosokawa@katfuji.lab.tuat.ac.jp, katfuji@cc.tuat.ac.jp

## Abstract

With the introduction of the Digital Transformation (DX) era, It is now feasible to obtain digital data not only on the shop floor of the manufacturing facility but also across the whole supply chain (SC) network for improved management. The Bottleneck (BN) is the first tackling point to gain more throughput. By getting the accumulated Lead time (LT) data on the SC network map using simulations, we could identify the SC bottleneck regardless of the production policies such as push or pull: using a simulator that imitates the production of the entire SC network by assembling the materials; using the simple Key Performance Indicators (KPIs) that are the average lead time and the standard deviation of lead time on the simulator; identifying the BN from the remaining quantities of work in processes (WIP) between the nodes in a high-demand situation; identifying the BN based on the use of nodes at the low-demand situation virtually. In addition, our last method can depict the degree how each node is close to the BN, by sorting the rate of the utilization of the nodes.

## 1 Introduction

To avoid a loss of opportunity for the supply chain (SC) network, it must deliver the product Just In Time (JIT),[10]. However, the market demand in the real world and the capacity to produce the parts from each node fluctuate regularly. It is challenging to effectively change the production capacity for each factory or workstations (nodes) in the real world in response to demand fluctuations[12] because each node's flexibility varies. If a node is unable to match the increased demand, it would eventually become a bottleneck (BN), resulting in sales opportunity losses for all SC network members. This damages the balance sheet by creating unsold goods or overspending on capacity without generating enough revenue.

Even if the owner of the SC network deploys a special task force team of industrial engineers to save the BN, it could be challenging to find the appropriate BN; If the owner increases the capacity of the non-BN, the result brings redundant work in processes (WIP) in the push production system. Because the outcome from real BN is still insufficient, the increased WIP does not help the SC network improve the number of deliveries to clients. Thus, all the non-BN activities are a waste of time, and it would be critical for all stakeholders of the SC network to find out the right BN. Further, the emergence of the digital transformation (DX) enables

us to detect information on the shop floor inside the SC network[13]. It is feasible to increase the overall SC's efficiency by evaluating the obtained data through simulations[22]. Others, in addition to the manufacturing industries, are attempting to visualize the status of the SC network at a glance using dashboards[20].

We proposed a simulator that imitates the production of the entire SC network by assembling the materials at each node, monitored the production performance of this simulator under various conditions especially the frequency of customer order, identified the bottleneck from the remaining stocked quantities of WIP between the nodes when demand exceeds the total capacity of the SC network, and proposed the use of the nodes as the KPI for BN at the low demand situations virtually. This represented the magnitude of each node's proximity to the BN. We named this magnitude the BN degree, and the SC owners can refer to it as the priority to work for multiple Kaizen activities at multiple nodes<sup>1</sup>.

The structure of this study is as follows: first, we presented the related existing works about SC networks and multi-agent simulation; next, we proposed two strategies for detecting the BN; third, we demonstrated that the proposed methods correctly detected the BN; finally, we provided the conclusion and the future projects.

## 2 Related Works

There are various main methods in the field of Industrial Engineering for locating BNs.

- (1) The method for comparing the required production time based on the product mix, the cycle times and the average quantity, and other factors against the owner's capacity[19].
- (2) The method for predicting the BN from the node's stocked WIPs[15].
- (3) The method for predicting the BN from the rate of the node use, in other words, utilization rate[17].
- (4) Value stream analysis(VSA): The process for drawing a diagram of the material flow and the information flow, which will be used to change operations in the future by visualizing the current status, mainly in the manufacturing industry[18].

However, the method above is not always perfect because of the following reasons: (1) is inaccurate since it does not account for fluctuations of the product mix, even the reality is dynamic; (2) may be inaccurate by the noises of non-market-driven works in the queue stocks. There is so many WIP queue without orders from the consumers in the real SC networks, aiming for preemptive manufacturing. Some of the works might be dead-stocks if unsold. However, obscuring the real BN after the stock creates the impression that the node before the WIP stock-up works better than the node after. Further, the snapshot data of the WIP quantity can be an exceptional case of randomness from the statistical perspective. The works performed without the market's order increase the use by producing redundant WIPs, which is a waste of cash that puts the balance sheet in jeopardy. The WIP inventories can't be relied on for the judgment depending on the push or the pull production system; (3) has the same noises as (2) in the utilization rate record. This method can draw only the limited area in the same picture that flow is simple and measured in the same production policy. The policies such as "push" or "pull" of the production system should be the same among the scope of comparison. Make

---

<sup>1</sup>Even after Kaizen's actions at the BN, the development of throughput will come to halt in the real world; this is because the initial BN is no longer a BN, and another node has taken place at the top BN. Thus, SC owners must continually look for the top BN to maintain their growth[12]

to order, or preemptive push production depends on the character of the product basically. However, even among the same network producing the same product, there may be factories with different production policies. In such cases, we can't compare the utilization rates in the same manner between the factories that produce without orders and the factories that stay idle if no orders; (4) is not good at showing the BN at a glance. Unless the simple straight flow case, the widely branched river Amazon flow is not easy to draw in this diagram to find the BN and difficult to draw without skills and time. Automation of drawing is also not easy.

Both (2) and (3) have the shadowing effect at the area after the BN. If the second-worst BN is located right after the worst BN, both the queue stocks and utilization rate of the second BN might be measured less seriously than the reality. And the other nodes at the different branches of the streamline might be counted as second worst even not so serious. It is because the workload of the second BN is eased by the worst BN by receiving slower WIPs from the real pace of the throughput. Therefore the function of this simulator is very important that can lower the pace of customer demand than the one of BN without changing the specs of the production side to see the real capability of nodes.

Smart manufacturing research has become popular recently. The introduction of the Internet of Things (IoT) helps in using information about production status, analysis, and different levels of stakeholders, such as machines, factories, and Enterprise Resource Planning (ERP). However, Industry 4.0 will bring networking, visualization, and automation to monitor resources, manage industrial lines, and assist with auto-setups[22]. The inventory is a perennial issue in the manufacturing industry, and a method to prevent the bull-whip effect is being researched[5]. As the research on the AI progresses, machine learning and simulations are used in the manufacturing industry to increase efficiency[2]. In a recent study, the batch size, due dates, production capacity, WIP quantity, machine utilization rate, and other factors were considered[14].

In supply chain management (SCM), focusing on the competitive edge is crucial in business. It is an important topic to select suppliers, manage the lead time, and cope with the changing market for building and running the SC including the global logistics. Even for a single product, the SC network has spread around the world. However, companies must strengthen the relationship for day-to-day operations between components manufacturing companies all over the world and the consumer's sales network. To supply the essential items that meet the market's features depending on the regions, alignment and quick reactions are necessary, as well as strong connectivity between procurement, production, and sales[4].

SCM has characteristics in the real world that make the agent technology ideal for assisting decision making based on the simulations. Multi-agents systems can be used to model or perform tasks in SCM due to the similarities of the two systems[8]. The reasons for that are as follows:

- An SC consists of multiple parties working on multi-stage tasks, whereas a multi-agents system consists of different types of agents with varied roles and functions.
- There is no single authority: knowledge is distributed among members, decision making in the SC is accomplished through multiparty negotiation and coordination, and agents are autonomous: they are responsive to changing environment, proactive in taking self-initiated action, and social in interacting with humans and other agents.
- The structure of the SC is flexible: it can be organized differently to implement different strategies, and the agent system is flexible: agents can be organized according to various control and connection structures.

- An SC is dynamic: entities may join or leave the SC, agents can join or leave from a multi-agents system[7].

Kaihara et al. demonstrated a technique for determining the optimal SCM as a whole chain, by letting the nodes negotiate with each other in a virtual market to maximize the use of each node; three types of agent nodes were used in the virtual market: supplier agent, intermediary, and customer[6]. Supply agents employ capital to produce certain goods, and make and sell the goods to the consumers for a profit. The customer Agent purchases the goods and delivers them to downstream markets. The intermediate agent is a player that provides a trading venue such as an e-marketplace. This process uses 4-Heap algorithm[21].

Further, previous works of simulation-based BN detection are demonstrated. Lin et al. proposed a data-driven strategy for both short and long-term throughput BN identification. This method uses the production line blockage and starvation probabilities, as well as buffer content records to identify the production bottleneck without building an analytical or simulation model. This method has been verified analytically and by simulation, and an industrial case study was used to demonstrate the implementation and validate the efficiency of the proposed bottleneck detection method[9]. The Elba project is designed to achieve an automated iterative staging to mitigate the risk of violating Service-Level Objectives (SLOs). As part of Elba, we conducted performance characterization of the system to detect BNs in their configurations. In various configuration scenarios, the proposed BN detection approach showed resilience and accuracy; It uses RUBiS (Rice University Bidding System), a well-known benchmark application, to evaluate the classifier's performance in identifying various BNs[11]. From a computer science perspective, Bodner et al. proposed high-fidelity models of manufacturing systems; such high-fidelity modeling has important benefits in prototyping system performance. However, it must be supported by a modeling discipline or structured approach to modeling factory operations. Results are implemented as generic code modules in SIMAN and are demonstrated with a case study in semiconductor manufacturing[1]. Roser et al. presented a method for detecting the BN in a discrete event system by examining the average duration of each machine's active time for all machines. The BN is the machine with the longest average uninterrupted active time. The method is widely applicable and capable of analyzing complex and sophisticated systems. The results are extremely accurate, with a high degree of confidence in distinguishing between BN and non-BN devices[16]. Further, they compared the two most used BN detection methods in terms of AGV (Automated Guided Vehicle) usage and waiting time[3].

### 3 The SC simulator to detect the BN

Figure 1 shows the flow of the SC simulator. The node signifies companies, factories, and/or processes. To produce one unit in the simulator, each node needs the data of the average lead time ( $\mu$ ) and the standard deviation of lead time ( $\sigma$ ). STC can be calculated from the lead time (LT) database. Each node's actions can order input materials upstream, produce the item with received materials, and ship downstream. The order quantities for the materials aren't above the customer's order quantity. The production starts only after the customer's order (No pre-emptive production starting: Pull flow) and the inventory of Finished goods (FG) is shipped immediately (No FG inventory).

#### The set up of node relationship

The simulator models the SC network by running on a map linked between the nodes. There are no limits on the number of nodes connected as long as the processing power is available, but

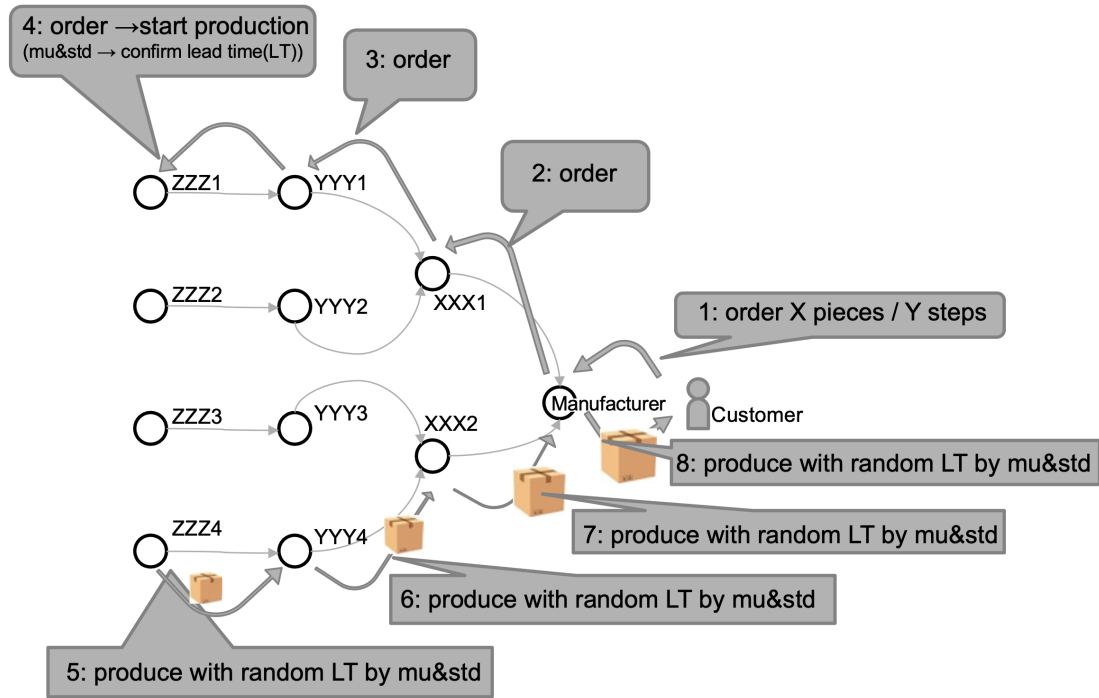


Figure 1: Flow of SC simulator

one simulation for 30 nodes up to 10,000 steps took around 3 seconds in the given environment. The input and output parts numbers, as well as quantities, are set; if all input parts have arrived, the nodes start the assembling and produce pre-determined quantities. There are no limitations on the number of input parts or their quantity. And, shipping destinations can be multiple if the output product amount in one batch is multiple. Further, all nodes are linked and the last node produces the finished goods, which are then shipped to the customer and the order is complete.

### The flow of orders and the product

The parts are handled on the steps between the different layers of the nodes in the simulator. In this simulator, the production system is “pull” without preemptive production to eliminate the tasks that were not ordered by the customer. The customer places the orders as specific intervals ( $Y$  in the Figure 1) with random dispersion within the pre-set standard deviation. In this research, the yield: the rate of the good product among all production, is ignored and all production is considered as a quality product. Therefore manufacturer node will only send orders to the upstream tiers in the quantity required to fulfill the current order. The sub-tier nodes will send the orders to the upstream tiers once they have received the orders from the downstream. All nodes will start production only if all required input parts have arrived; otherwise, the nodes will wait for the unreceived parts to arrive.

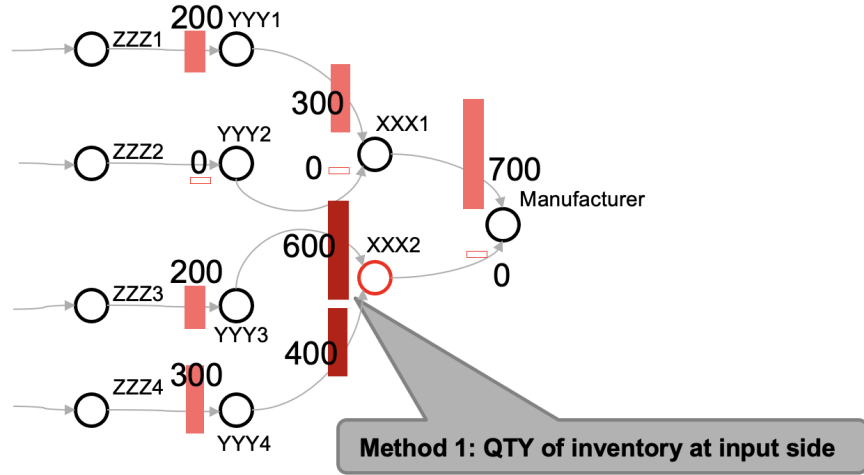


Figure 2: The quantity of WIP at the input side of the node (Method1)

### Step procedures

Once ordered, the node will calculate the needed lead time randomly with Gaussian distribution from the pre-set average LT ( $\mu$ : the number of needed steps to finish the production at that node) and the standard deviation of the LT ( $\sigma$ ). After the calculated steps, that node's manufacturing will be complete, and the product will be delivered downstream instantly; the next nodes follow after receiving all needed parts from all upstream. After completing all the pre-set processes, the simulator will produce all the data of each step such as the received number of final finished goods by the customer, and all WIP traces. A step includes the sub-steps of renewing the latest order, producing and shipping the finished goods, and determining whether the next production should begin or not.

The simulator's inputs are as follows:

- the relationship connections between all nodes
- the quantity of both each input and output in a single batch on the node
- the average LT for production on each node ( $\mu$ )
- the standard deviation of LT on each node ( $\sigma$ )

## 4 BN detection method

This paper proposes two methods to detect the BN that are simulated from the LT database on the SC map information.

### Method1: Quantity of WIP QUEUE at the input side of the node

We proposed this method first to identify BN from the node that has the highest quantity of the stocked WIP at the end of the steps (Figure 2). The following are the conditions.

**Method 2: Rate of node utilization**

Step	0	1	2	3	...	9998	9999	Total
NodeX	Off	Work	Off	Off	...	Work	Work	4500 / 1000

Figure 3: The rate of the utilization of the nodes at the low demand situation (Method2)

- The node that has the biggest quantity of the WIP at the input side including the manufacturer.
- All quantities of the materials at the input side of this node are not zero. If the reason for the production delay is due to shortages of the input material, then the delay was not caused by the low capacity, but the slow arrival of the materials.
- The top upstream tiers should be excluded because the inventories at the input side of this tier were set extremely high for the simulation.

**Method2: Rate of the utilization of the nodes at the low demand situation**

This is the experiment that uses the simulator. Therefore it is feasible to change only the demand virtually without affecting the KPIs of the *mu* and *std*. There is the prerequisite that the supply side does not run to produce the redundant WIPs and the supply side remains idle if there is no demand (Figure 3). We compared each node’s usage by simulating a recession or poor sales situation that the pace of the customer orders is slower than the pace of BN. This is the ratio of the active time of the node to the total steps. We assumed that the rate of usage indicates the nodes’ capability, and acknowledged that the higher rate of usage corresponds to higher BN degrees.

## 5 Experiments

### 5.1 Scenario of the experiments

In this experiment, the end of one sequence of the procedure is to produce only one kind of end product. All the nodes of the production side are finally connected to the manufacturer, and the quantities of inputs are assembled into one selling to the final buyer in this SC without redundant leftover WIP. The final commodities are not kept in each node’s inventory, and they are promptly dispatched downstream. In the real world, production activities can occasionally fail. The rate of successful product among all production is called yield. However, we set the defect rates to zero and the Yield is 100 percent at all nodes in this simulation. Further, all WIPs in this SC network are linked to the orders. Because there are no stand-by finished goods items in the warehouse, the buyer does not receive the goods immediately after making the purchase order and must wait for the components to flow down the stream from the source.

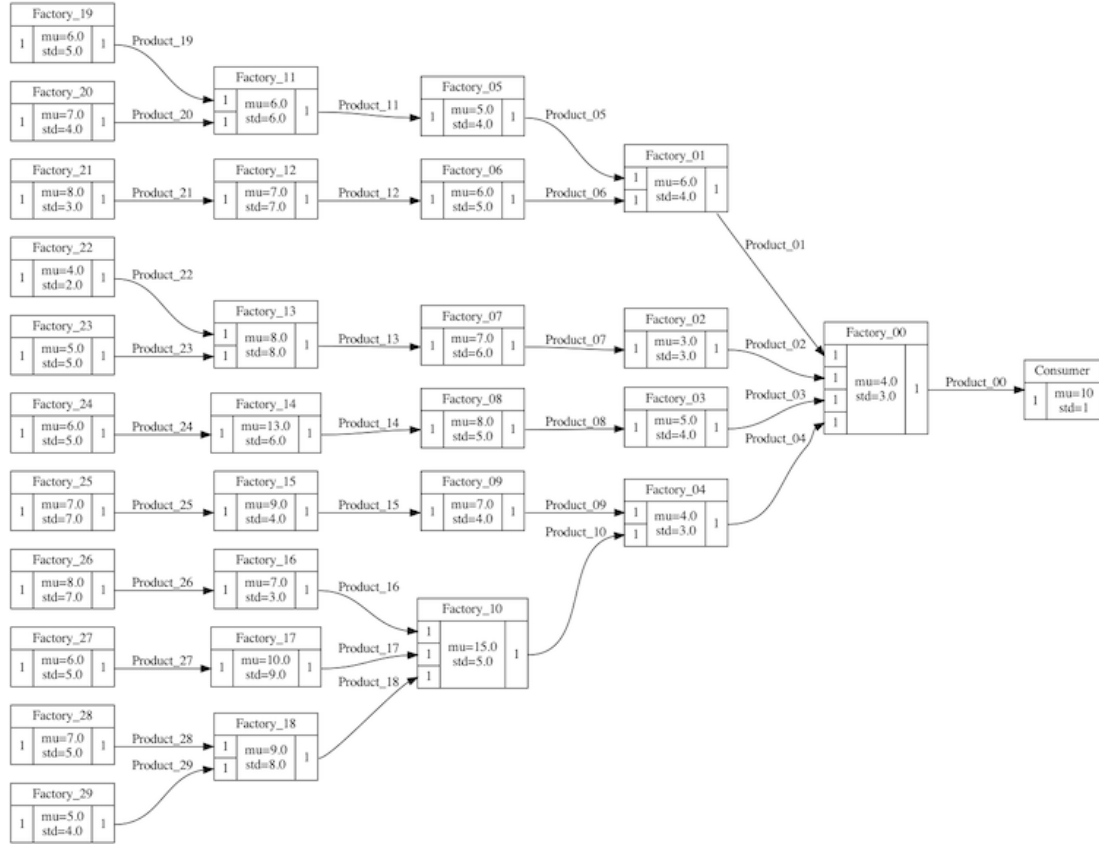


Figure 4: The structure of the SC network for this experiment

Thus, none of the nodes will run without orders, that is the perfect pull production system. This is how to eliminate the noises of the dead stocks to focus on only the signals.

Figure 4 is the diagram of this scenario with the node relationships and the KPI figures. Factory\_00 is the manufacturer. Factory\_01~04 are the Tier1 suppliers. Factory\_05~10 are Tier2. Factory\_11~18 are the Tier3. The most upstream Tier4 is Factory\_19~29 and there are enough inventories at the input side of the Tier4. We put the names of the nodes at the top center of the node boxes.  $\mu$  and  $\sigma$  are the KPIs for completing one batch's production. The finished product names can be found on the right side of the boxes, aside from the arrow. On one batch, the quantity of finished items is at the bottom right, and each input's amount is listed at the bottom left. To reduce statistical fluctuations, we repeated the experiment 100 times, with one simulation consisting of 10,000 times steps.

## 5.2 The pre-experiment to find the BN in this scenario

We improved the KPIs of  $\mu$  and  $\sigma$  by 20% at only one node to find the right BN on this scenario and repeated the process 30 times with each node improving. If the improvement was on the non-BN, the total throughput should not change. However, it improves the total



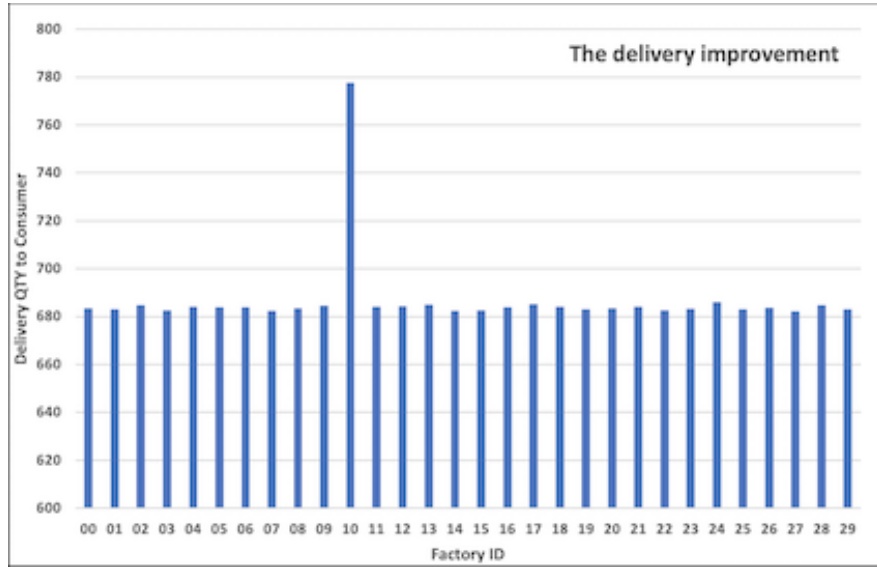


Figure 5: The number of the total delivery when the nodes are improved 20%

Table 1: The rate of the right answers on method 1 and 2

Method	rate
Method1: WIP qty before the node	99%
Method2: Utilization rate	100%

throughput if improving the right BN correctly.

The result of each 30 experiments of the quantity of the final product that customer received showed the improvement at only the Factory\_10 as Figure 5. Thus, the BN in this scenario is Factory\_10. With this as the proper BN, we deployed evaluations on two methods of the experiments. These calculations to find correct BN require massive calculations that are time-consuming and are impossible to perform manually.

### 5.3 The result of simulation

Table 1 shows the rate of the correct answer for each method, and the details are shown in Figure 6. Method 1 shows 99%, and method 2 shows 100%.

Method 1 shows Factory\_10 that is the correct answer of BN, 99 times out of 100 experiments. The last one showed BN was Factory\_00 and we examined the experiment log. One of the Factory\_00's four input parts was always 0, hence this node should not be considered as the BN. The statistical fluctuation was a rare circumstance that led to this incorrect answer. Thus, we should improve these criteria of BN in this experiment. This method was possible because the WIP is the result of the imbalance between nodes and finds the BN<sup>2</sup>.

<sup>2</sup>However, it is incorrect to evaluate the BN degree directly from the WIP quantities because the needed numbers of the parts are different for one finished items. Even if you had three wheels and one engine as the WIP at the automobile factory, the degree of the impact is not equal. The quantities of WIP must be converted

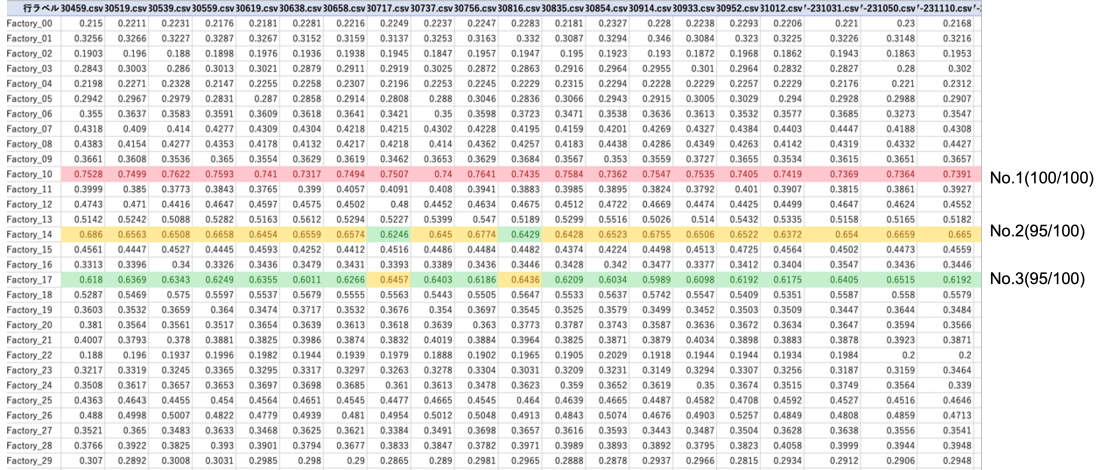


Figure 6: A part of the results of BN degree from utilization rate

Table 2: Worst three nodes under utilization rate

Position	Name	appearance
Top bottleneck	Factory_10	100%
2nd bottleneck	Factory_14	95%
3rd bottleneck	Factory_17	95%

The result of method 2 showed correctly 100 times out of 100 as Factory\_10, as the Figure 6 if we apply the rate of the usage/utilization rate under a low-demand situation. It showed Factory\_14 as the second worst BN 95 times out of 100 experiments. It also showed Factory\_17 as the third worst BN 95 times out of 100 experiments. Based on those findings, Method2 is simpler and more reliable under the current settings. Also, the Bottleneck degree was measured how each node is close to the BN as Figure 7.

## 6 Conclusion

Industrial engineers frequently find the BN from the stocked WIP quantities between nodes in the real world. However, those can be inaccurate to judge the BN. This is because those WIPs may include dead stocks that are not linked to the market orders. The node can be misunderstood as the BN based on the long queue of WIP of dead stocks even if this node produced the needed parts promptly. The utilization rate is the other famous method in the industry to find the BN. However, the policy and conditions should be the same to compare the sampled utilization rate in the same picture. It was difficult to enhance the scope of the utilization rate to compare with the outside of the factories.

This study developed the simulator from the SC map and the simple KPIs of each node that is obtainable from the accumulated LT record, enabling to estimate total production delivery. The rate of product mix, the yield, and the variation of the production lot size was not taken

into the number of the final product.

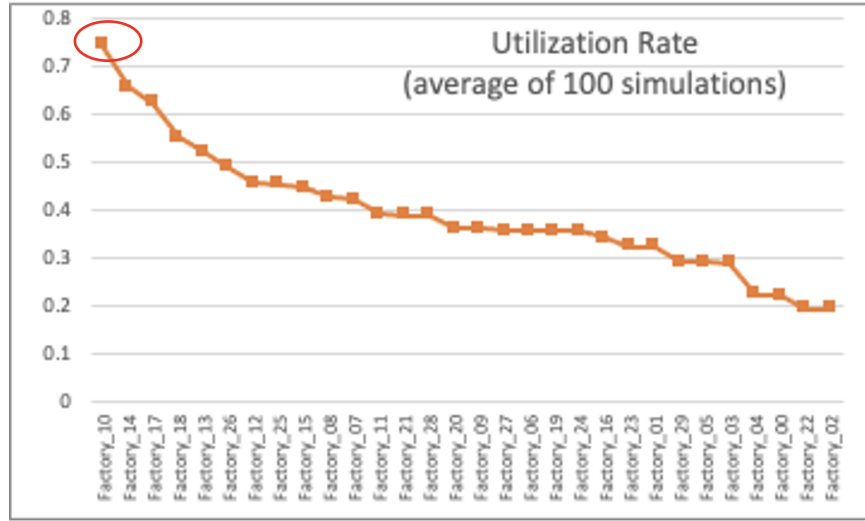


Figure 7: Bottleneck Degree

into account this time for simplicity. We proposed approaches to identify the bottleneck from the remaining stocked quantities of work in processes (WIP) between the nodes in the high-demand situation. We also proposed the rate of the usage of the nodes at the low-demand situation as the second method to detect the BN. We evaluated the ability of this simulator, and the proposed methods showed good results detecting the pre-determined correct BN. Both methods of WIP queue size and utilization rate detected the BN regardless of the differences of push or pull production philosophy of the original SC. The simulator could detect the utilization rate even if the node was located right after the BN. The second method of utilization rate also showed the rank of the magnitude of each node is close to the BN. This simulator would be practical and adaptable to a wide range of SC because the LT database is one of the most popular data obtainable from the real SC networks. And also it is because this simulator can neglect the policy of the production system of the real SC network regardless of push or pull, which is not negligible in the previous research.

The future work of this research will consider exploring varied SC maps, the yield, and the mixed model production. Further, the proposed simulator did not consider the interactions between nodes such as negotiations. The improved simulator will be able to analyze more aspects in real cases.

## References

- [1] Douglas A. Bodner and Leon F. Mcginnis. A structured approach to simulation modeling of manufacturing systems. In *Proceedings of the 2002 Industrial Engineering Research Conference*, 2002.
- [2] Alok R Chaturvedi, George K Hutchinson, and Derek L Nazareth. A synergistic approach to manufacturing systems control using machine learning and simulation. *Journal of Intelligent Manufacturing*, 3(1):43–57, 1992.

- [3] Stephen E. Chick, Paul J. Sánchez, Don Ferrin, and Douglas J. Morrice. Comparison of bottleneck detection methods for agv systems. In *Winter Simulation Conference 2003*, pages 1192–1198, 2003.
- [4] Mark Goh, Joseph YS Lim, and Fanwen Meng. A stochastic model for risk management in global supply chain networks. *European Journal of Operational Research*, 182(1):164–173, 2007.
- [5] Kiyong Jeong and Jae-Dong Hong. The impact of information sharing on bullwhip effect reduction in a supply chain. *Journal of Intelligent Manufacturing*, 30(4):1739–1751, 2019.
- [6] Toshiya KAIHARA, Susumu FUJII, and Kenji OHYA. A study on artificial market based on economics of complex systems. *Transactions of the Institute of Systems, Control and Information Engineers*, 17(4):170–177, 2004.
- [7] Rasoul Karimi, Caro Lucas, and Behzad Moshiri. New multi attributes procurement auction for agent-based supply chain formation. *International Journal of Computer Science and Network Security*, 7(4):255–261, 2007.
- [8] Averill M. Law and Michael G. McComas. Simulation of manufacturing systems. In *Proceedings of the 30th Conference on Winter Simulation*, WSC '98, pages 49–52, Washington, DC, USA, 1998. IEEE Computer Society Press.
- [9] Lin Li, Qing Chang, and Jun Ni. Data driven bottleneck detection of manufacturing systems. *International Journal of Production Research*, 47(18):5019–5036, 2009.
- [10] Jeffrey K Liker. *Toyota way: 14 management principles from the world's greatest manufacturer*. McGraw-Hill Education, 2004.
- [11] Simon Malkowski, Markus Hedwig, Jason Parekh, Calton Pu, and Akhil Sahai. Bottleneck detection using statistical intervention analysis. In *Proceedings of the Distributed Systems: Operations and Management 18th IFIP/IEEE International Conference on Managing Virtualization of Networks and Services*, DSOM'07, pages 122–134. Springer-Verlag, 2007.
- [12] Kenneth N McKay and Thomas E Morton. Review of: Critical chain. *IIE TRANSACTIONS*, 30(8):759–762, 1998.
- [13] Ercan Oztemel and Samet Gursev. A taxonomy of industry 4.0 and related technologies. *Industry 4.0*, page 45, 2020.
- [14] Shwetank Parihar and Chandan Bhar. Development of framework for mitigating production bottleneck related risks: A case study on thermosetting plastic products manufacturing firm. *Management Insight*, 11(2):91–99, 2015.
- [15] Glenn C Parry and CE Turner. Application of lean visual process management tools. *Production planning & control*, 17(1):77–86, 2006.
- [16] C. Roser, M. Nakano, and M. Tanaka. A practical bottleneck detection method. In *Proceeding of the 2001 Winter Simulation Conference (Cat. No.01CH37304)*, volume 2, pages 949–953 vol.2, 2001.
- [17] Christoph Roser, Masaru Nakano, and Minoru Tanaka. Shifting bottleneck detection. In *Proceedings of the Winter Simulation Conference*, volume 2, pages 1079–1086. IEEE, 2002.
- [18] Mike Rother and John Shook. *Value-stream Mapping Workshop: Participant Guide: a Learning Solution from LEI*. Lean Enterprise Institute LEI, 2009.
- [19] Nigel Slack, Stuart Chambers, and Robert Johnston. *Operations management*. Pearson education, 2010.
- [20] Dusan Stefanovic and Nenad Stefanovic. Methodology for modeling and analysis of supply networks. *Journal of Intelligent Manufacturing*, 19(4):485–503, 2008.
- [21] Peter R Wurman, William E Walsh, and Michael P Wellman. Flexible double auctions for electronic commerce: Theory and implementation. *Decision Support Systems*, 24(1):17–27, 1998.
- [22] SooCheol Yoon, Jumyung Um, Suk-Hwan Suh, Ian Stroud, and Joo-Sung Yoon. Smart factory information service bus (sibus) for manufacturing application: requirement, architecture and implementation. *Journal of Intelligent Manufacturing*, 30(1):363–382, 2019.