A Quantitative Comparison of Bottleneck Detection Methods in Manufacturing Systems with Particular Consideration for Shifting Bottlenecks

Christoph Roser, Masaru Nakano

A Quantitative Comparison of Bottleneck Detection Methods in Manufacturing Systems with Particular Consideration for Shifting Bottlenecks

Christoph Roser[1], Masaru Nakano[2]

[1]Karlsruhe University of Applied Sciences, Karlsruhe, Germany
christoph.roser@hochschule-karlsruhe.de
[2]Graduate School of System Design and Management, Keio University, Yokohama, Japan
m.nakano@sdm.keio.ac.jp

Abstract. There are numerous different bottleneck detection methods both in academia and in industry. However, most do not take the shifting of bottlenecks into account. This paper compares a number of methods, namely methods based on cycle times and utilizations, waiting times and queue lengths, the arrow method, the turning point method, the active period method, and the bottleneck walk. All methods are tested against two different manufacturing systems, a pseudo static system consisting of a static system that changes once halfway through the production run and is designed to have shifting bottlenecks, and a dynamic system that includes variations for a multitude of different processes.

Keywords: Bottleneck detection · shifting bottlenecks · theory of constraints · capacity improvement

1 Introduction

Bottlenecks are processes that limit the overall system capacity of manufacturing processes. Hence, as part of our never-ending quest to produce more with less, there is frequently the need to improve system throughput. For this, it is necessary to improve the bottleneck, since improving the speed of a non-bottleneck will have no influence on the system capacity. This gives rise to the need to find the bottlenecks.

1.1 Bottleneck Definition with Respect to Shifting Bottlenecks

The need to find the bottleneck is complicated by the dynamics of real-life systems. In real systems, processes are not static but may change. There are numerous definitions of bottlenecks in literature. [1] describes a bottleneck as processes that limits output. [2] defines the bottleneck as the process whose isolated production rate has the highest sensitivity of the system's performance compared to all other processes. [3],[4] defines the bottleneck as the stage in a system that has the largest effect on slowing down or stopping the entire system. We expand these definitions to include both multiple bottlenecks and a measure of influence on the system:

> **Bottlenecks are processes that influence the throughput of the entire system. The larger the influence, the more significant is the bottleneck.**

1.2 Degree of Influence of the Bottlenecks on the Entire System

Since more than one process is likely to be a bottleneck using the definition above, it is of interest to compare the relevance of the bottlenecks. The larger the bottleneck, the larger its influence on the system throughput. While this sensitivity is difficult to obtain analytically, it can be obtained experimentally by comparing the system behavior for different cycle times. In our simulations we change the speed of the process and observe the change in the speed of the entire system. The gradient of this relation in percent represents the degree of influence of the process on the entire system. Four examples are shown in Fig. 1 with the time between parts for the process on the x-axis and the time between parts for the system on the y-axis. The horizontal and vertical dashed lines indicate the point under observation for which the gradient was measured.
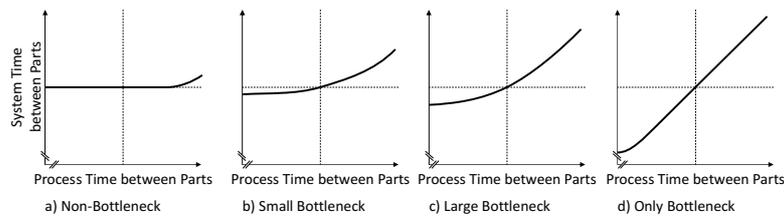
**Fig. 1.** Gradient between Process Time between Parts and System Time between Parts

If a process is no bottleneck at all, the gradient is 0% as shown in a). The other extreme is having only a single bottleneck as shown in d). In both cases major increases or decreases of the process speed will eventually change the gradient. Graph b) and c) in Fig. 1 show intermediate stages. Hence the degree of influence of a process onto the system can be between 0% and 100%.

2 Reference Systems Used for Comparison

2.1 Pseudo-Dynamic System

The first system for comparing the bottleneck detection methods is a pseudo-dynamic system. This system is composed of two static systems, a) and b) as shown in Fig. 2 with unlimited demand and supply. The three processes are separated by FIFO with a capacity of 3. Process P3 always has the same cycle time of 90 seconds. However, process P1 and P2 change between system a) and b). During the first half of the observed period, the entire system behaves like static system a). During the second half of the observed period, the system behaves like static system b). For each subsystem, the bottleneck is determined easily, being P1 and P2 in the subsystems a) and b) respectively. Hence, by constructing such a pseudo-dynamic system, *we have*

forced a bottleneck shift from process P1 to process P2. P3 is never the bottleneck. Hence the degree of influence of the processes on the entire system (as per 1.2 above) should be 50% for P1 and P2. The simulation results of the pseudo-dynamic system are shown in Table 1 below.
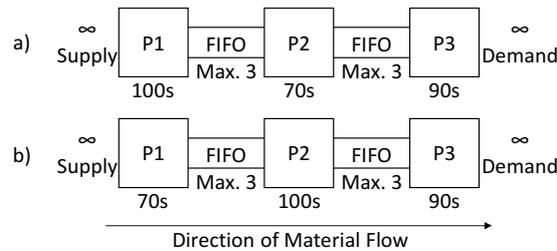
Fig. 2. Pseudo-Dynamic Manufacturing System

**Table 1.** Results of the Pseudo-Dynamic Manufacturing System

Process	First Half			Second Half			Average		
	Utilization	Blocked	Starved	Utilization	Blocked	Starved	Utilization	Blocked	Starved
P1	100%	0%	0%	70%	30%	0%	85%	15%	0%
P2	70%	0%	30%	100%	0%	0%	85%	0%	15%
P3	90%	0%	10%	90%	0%	10%	90%	0%	10%

Queue	First Half			Second Half			Average		
	Average Inventory	Average Inventory (%)	Average Waiting Time(s)	Average Inventory	Average Inventory (%)	Average Waiting Time(s)	Average Inventory	Average Inventory (%)	Average Waiting Time(s)
P1 to P2	0	0%	0	3	100%	300	1,5	50%	150
P2 to P3	0	0%	0	0	0%	0	0	0%	0

2.2 Dynamic System

The dynamic system is a more complex system, shown in Fig. 3 below. It has four processes, separated by FIFO's of capacity 3, and also with unlimited demand and supply. The cycle times of the processes are exponentially distributed with means of 90s, 95s, 100s, and 90s for the processes P1, P2, P3, and P4 respectively.
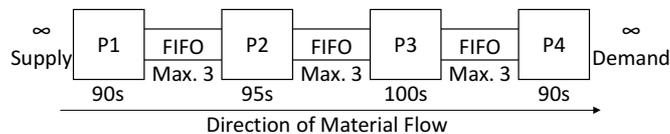
Fig. 3. Dynamic Manufacturing System

The overall performance results are shown in Table 2, based on the average of 100 simulations with 10,000,000 seconds each and including the 95% confidence interval. To determine the degree of influence this system has been simulated both for the original cycle times and for each process being 1 second faster and slower. The results of these simulations are shown in Fig. 4, Please note that the lines are not straight but have a minor convex shape. The process P3 h the slowest cycle time of

100s had the largest gradient of 47.2%, making it the largest bottleneck. The second slowest process P2 followed closely behind with a gradient of 40.3%. The fastest processes P4 and P1 had the smallest gradient with 23.7% and 21.7% each.

Table 2. Results of the Dynamic Manufacturing System

Process	Utilization	Blocked	Starved	Queue	Average Inventory	Average Inventory (%)	Average Waiting Time(s)
P1	71,03±0,051%	28,97±0,051%	0,00±0,000%				
P2	74,97±0,048%	18,36±0,055%	6,68±0,032%	P1 to P2	2,08±0,002	69,44±0,001%	264,00±0,310
P3	78,89±0,051%	7,43±0,035%	13,68±0,050%	P2 to P3	1,63±0,003	54,17±0,001%	205,94±0,362
P4	71,04±0,052%	0,00±0,000%	28,96±0,052%	P3 to P4	0,96±0,002	32,10±0,001%	122,06±0,255
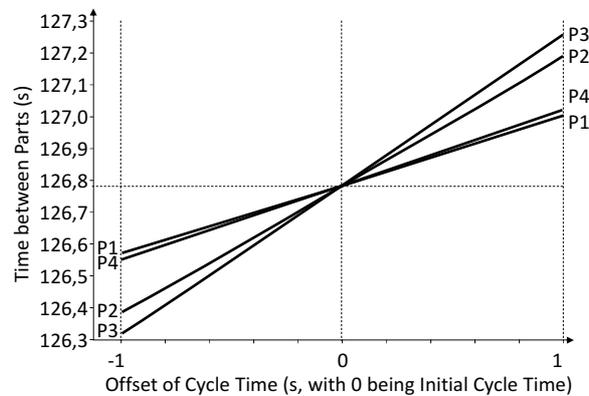
Fig. 4. Graphic Representation of the Gradients for All Four Processes of the Dynamic System

3 Analyzed Bottleneck Detection Methods

3.1 Methods Based on Cycle Times or Utilizations

One of the most common approaches in industry is to determine the bottleneck based on the largest average cycle time or the utilization. Variations of these methods are described for example, in [5]. This approach fails for the pseudo-dynamic system, erroneously determining P3 as the bottleneck. For the dynamic system, P3 has the longest average cycle time and the largest utilization of 78.89%. Therefore, the method correctly considers this process to be the main bottleneck. However, the significant influence of P2 and the smaller influences of P1 and P4 are completely ignored. Hence, utilization give a very incomplete picture of the situation.

3.2 Methods Based on Waiting Times or Queue Lengths

There are a number of different methods described in literature that determine the bottleneck based on the inventories between the processes. These use for example total waiting time [5], average waiting time [6], length of the queue [7], or combinations thereof [8]. For the pseudo-dynamic system, the first FIFO has an aver-

age inventory level of 50%. The second FIFO is always empty. The bottleneck should be at the process with the largest drop in waiting time or queue length. As such, we have two drops of equal magnitude from 100% to 50% around P1 and from 50% to 0% around P2, giving an "unclear result." In the dynamic system, the largest drop would be from 32.10% to 0% around P4. The second largest drop would be from 100% to 69.44% around P1. Hence, this approach would incorrectly consider P4 to be the bottleneck.

3.3 The Arrow Method Based on Starving and Blocking

The arrow method presented by [2] is based on the frequencies of processes being starved and blocked. "*If the frequency of manufacturing blockage of machine m_i is larger than the frequency of manufacturing starvation of machine m_{i+1}, the bottleneck is downstream of machine m_i. If the frequency of the manufacturing starvation of machine m_i is larger than the frequency of the manufacturing blockage of m_{i-1}, the bottleneck is upstream of machine m_i.*"

In the pseudo-dynamic system, process P1 is not blocked at all in the first observed half, but blocked 30s out of 100s for the second half, giving an average blocked probability of 15%. Similarly, P2 has a starving probability of also 15%. P3 has to wait for parts 10s out of 100s no matter what, and is hence 10% starved. Adding the arrows as shown in Fig. 5 clearly identifies P3 as a non-bottleneck, but fails to offer a direction between P1 and P2. The results of the dynamic system are shown in Fig. 6. The method clearly identifies the primary bottleneck P3. However, the arrow method considers P1, P2, and P4 to be non-bottlenecks.
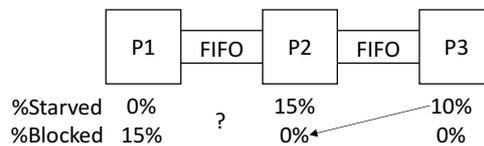

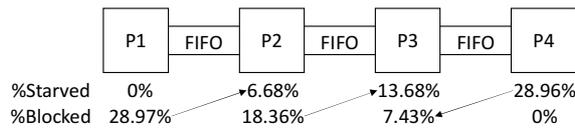**Fig. 5.** Arrow Method for the Pseudo-Static System

Fig. 6. Arrow Method for the Dynamic System

3.4 The Turning Point Method

The turning point method developed by [9] is also based on blockages and starvation similar to the Arrow method, although the calculation is more complex. In short, the bottleneck is the process where the difference between blocking and starving turns from positive to negative AND the sum of both blocked and starved must be lower than the two adjacent processes. The method can detect more than one bottleneck and even includes a ranking of multiple bottlenecks. The turning point method

fails for the pseudo-static system. According to the turning point method, there is no bottleneck in the pseudo-static system. While in the dynamic system the turning point correctly identifies P3 as the main bottleneck it misses all other bottlenecks in the dynamic system.

3.5   The Active Period Method

The active period method was developed by [10], [11]. In this method, a process is considered active whenever the process is not waiting for parts or material. At any given time, the process with the longest active period is the momentary bottleneck. Overlap between the longest active periods are times of shifting bottlenecks. Periods with no overlaps are sole bottlenecks. The total bottleneck probability is the likelihood of a process being a sole or a shifting bottleneck. Regarding the pseudo-dynamic system, the active period correctly identifies the bottleneck likelihood of 50% for P1 and P2 each, whereas P3 is never the bottleneck. The bottleneck probabilities for the dynamic system including a 95% confidence interval were $24.1\pm3.7\%$ for P1; $36.1\pm4.8\%$ for P2; $49.8\pm2.9\%$ for P3, and $24.3\pm3.2\%$ for P4. These results match almost perfectly with the experimental results from 2.2 as shown in Fig. 7.
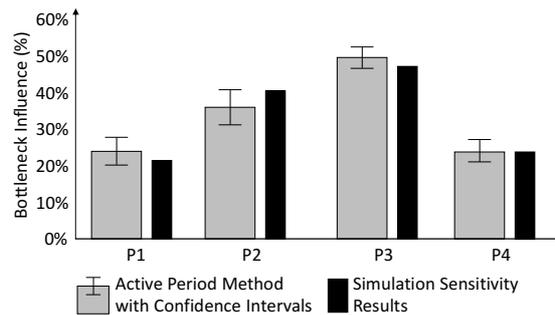

Fig. 7. Active Period Method results for the dynamic system
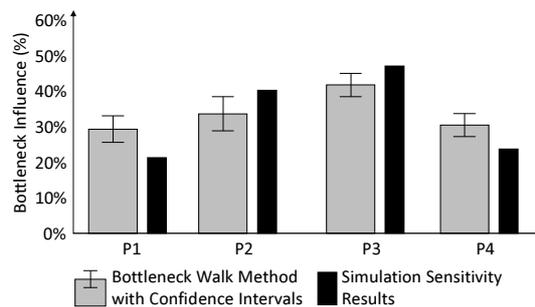
3.6   The Bottleneck Walk

Fig. 8. Bottleneck Walk Results for the Dynamic System

The bottleneck walk [12] uses observations of processes being starved and blocked and inventory levels to determine the direction of the bottleneck. The method is particularly suited for use on the shop floor, as no mathematical calculations or detailed measurements are required. For the pseudo-dynamic system, the bottleneck walk correctly determines both P1 and P2 are bottlenecks with an influence of about 5 0%.In the dynamic system, the bottleneck probabilities and their 95% confidence interval were 28.7±2.9% for P1, 31.4±2.2% for P2, 40.8±1.7% for P3 and 29.7±1.9% for P4.While these results are not as good as the active period method, they still come very close to the true sensitivity as shown in Fig. 8.

4　　Summary of Results and Conclusion

Overall, the accuracy of these bottleneck detection methods varies widely. Table 3 shows the overview of the results for all examined methods and systems. Fields in gray represent an incorrectly identified process. The last column shows the mean squared error of the bottleneck likelihoods.

Table 3. Results Overview

Examined System		Pseudo-Dynamic System			Dynamic System				Mean Squared
	Process	P1	P2	P3	P1	P2	P3	P4	Error
Baseline Simulation Results	Bottleneck	Prim.	Prim.	No BN	Sec.	Sec.	Prim.	Sec.	n/a
	Gradient	50,0%	50,0%	0,0%	21,7%	40,3%	47,2%	23,7%	
Max Cycle Time/Utilization	Bottleneck	No BN	No BN	Prim.	No BN	No BN	Prim.	No BN	29,21%
	Gradient	0,0%	0,0%	100,0%	0,0%	0,0%	100,0%	0,0%	
Waiting Time or Queue Length	Bottleneck	Unknown	Unknown	No BN	No BN	No BN	No BN	Prim.	20,29%
	Gradient	Unknown	Unknown	0,0%	0,0%	0,0%	0,0%	100,0%	
Arrow Method	Bottleneck	Unknown	Unknown	No BN	No BN	No BN	Prim.	No BN	10,89%
	Gradient	Unknown	Unknown	0,0%	0,0%	0,0%	100,0%	0,0%	
Turning Point Method	Bottleneck	No BN	No BN	No BN	No BN	No BN	Prim.	No BN	14,92%
	Gradient	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%	0,0%	
Active Period Method	Bottleneck	Prim.	Prim.	No BN	Sec.	Sec.	Prim.	Sec.	0,04%
	Gradient	50,0%	50,0%	0,0%	24,1%	36,1%	49,8%	24,3%	
Bottleneck Walk	Bottleneck	Prim.	Prim.	No BN	Sec.	Sec.	Prim.	Sec.	0,29%
	Gradient	50,0%	50,0%	0,0%	28,7%	31,4%	40,8%	29,7%	

With the dynamic system, all but the waiting time or queue length were able to determine the primary bottleneck correctly. However, only the active period method and the bottleneck walk were able to quantify the secondary bottlenecks. Regarding the pseudo-dynamic system, the shifting of the bottlenecks makes the detection more difficult. Only the active period method and the bottleneck walk were able to identify the bottlenecks correctly, although the arrow method and the waiting time/queue length method were undecided between the two bottlenecks. Only the active period method and the bottleneck walk were able to measure the bottleneck likelihood. The active period method were also the methods recommended by [13].

Overall, to detect shifting bottlenecks, it is imperative to **first detect the momentary bottleneck before calculating averages of the overall effect on the system**. Any method using averages before detecting the bottlenecks is likely to fall short for shifting bottlenecks. Of the presented methods, the active period method is particularly well suited for data-rich environments like simulations, whereas the bottleneck walk is best suited for a shop-floor-based observation.

5   References

[1]   L. J. Krajewski, L. P. Ritzman, and M. K. Malhotra, *Operations Management*, 9 edition. Upper Saddle River, N.J: Prentice Hall, 2009.

[2]   C.-T. Kuo, J.-T. Lim, and S. M. Meerkov, "Bottlenecks in Serial Production Lines: A System-Theoretic Approach," *Math. Probl. Eng.*, vol. 2, pp. 233–276, 1996.

[3]   C. Roser, M. Nakano, and M. Tanaka, "Time Shifting Bottlenecks in Manufacturing," in *International Conference on Advanced Mechatronics*, Asahikawa, Hokkaido, Japan, 2004.

[4]   C. Roser, M. Nakano, and M. Tanaka, "Monitoring Bottlenecks in Dynamic Discrete Event Systems," in *European Simulation Multiconference*, Magdeburg, Germany, 2004.

[5]   A. M. Law and D. W. Kelton, *Simulation Modeling & Analysis*, 2nd ed. McGraw Hill, 1991.

[6]   P. K. Pollett, "Modelling congestion in closed queueing networks," *Int. Trans. Oper. Res.*, vol. 7, no. 4–5, pp. 319–330, Sep. 2000.

[7]   S. R. Lawrence and A. H. Buss, "Shifting Production Bottlenecks: Causes, Cures, and Conundrums," *J. Prod. Oper. Manag.*, vol. 3, no. 1, pp. 21–37, 1994.

[8]   G. F. Elmasry and C. J. McCann, "Bottleneck discovery in large-scale networks based on the expected value of per-hop delay," in *2003 IEEE Military Communications Conference, 2003. MILCOM '03*, 2003, vol. 1, pp. 405–410 Vol.1.

[9]   L. Li, Q. Chang, and J. Ni, "Data driven bottleneck detection of manufacturing systems," *Int. J. Prod. Res.*, vol. 47, no. 18, pp. 5019–5036, Jun. 2009.

[10]   C. Roser, M. Nakano, and M. Tanaka, "Shifting Bottleneck Detection," in *Winter Simulation Conference*, San Diego, CA, USA, 2002, pp. 1079–1086.

[11]   C. Roser, M. Nakano, and M. Tanaka, "Detecting Shifting Bottlenecks," in *International Symposium on Scheduling*, Hamamatsu, Japan, 2002, pp. 59–62.

[12]   C. Roser, K. Lorentzen, and J. Deuse, "Reliable Shop Floor Bottleneck Detection for Flow Lines through Process and Inventory Observations," in *Proceedings of the Robust Manufacturing Conference*, Bremen, Germany, 2014.

[13]   E. Lima, L. Chwif, and M. R. P. Barreto, "Metodology for selecting the best suitable bottleneck detection method," in *Simulation Conference, 2008. WSC 2008. Winter*, 2008, pp. 1746–1751.