Roser, C., Nakano, M., Tanaka, M., 2004. Holistic Manufacturing System Analysis, in: International Conference on Advanced Mechatronics. Asahikawa, Hokkaido, Japan.

Holistic Manufacturing System Analysis

Christoph Roser
Masaru Nakano
Minoru Tanaka
Toyota Central Research and Development Laboratories
Nagakute, Aichi, 480-1192, JAPAN
croser@robotics.tytlabs.co.jp

Abstract

Manufacturing systems are crucial for the Industry. The Toyota Central Research Laboratories have recently developed a number of highly successful simulation based methods to understand and predict the behavior of a manufacturing system. These methods include a reliable quantitative bottleneck detection, resulting in a machine sensitivity analysis including a prediction model of the effect of machine changes, and a blocking and starving analysis, resulting in a prediction model of the effects of buffers and a subsequent buffer optimization. The novel idea of these methods is a holistic view of the manufacturing system, i.e. understanding the system by analyzing the relations between the machines instead of analyzing machines independently. This paper provides the framework of the holistic manufacturing analysis and a brief summary of the methods already developed using this framework.

1. Introduction

The understanding and optimization of manufacturing systems is a frequently researched subject in discrete event simulation [1, 2] [3, 4]. Yet, most methods analyze the simulation results in an atomistic view, and study each machine separately. However, if the correlated manufacturing system is broken into its independent machines before analysis, it is difficult to later combine the analysis of the machines into a understanding of the entire system. However, a holistic approach studies not only the machines but also the interactions between the machines in order to understand the system, yielding a much better understanding of the manufacturing system.

Applying the holistic approach to the active times of the machines evaluates the shifting bottleneck in the system. Finding the bottleneck is no trivial task, and [5] for example simply recommends that '… the best approach is often to go to the production floor and ask knowledgeable employees …'. The presented shifting bottleneck detection method is based on a detailed analysis of the working relationships between the different machines and provides a very accurate and quantitative measure of the constraints in the machines.

Applying the holistic approach to the idle times of the machines enhances the understanding of the blocking and starving relations in the system and allows a better understanding of the effect of buffers. An excellent discussion of the effect of buffers can be found by Conway et al [6] and others [7, 8]. The presented buffer allocation method is based on a detailed analysis of the blocking and starving relationships in manufacturing systems and provides a usable prediction model of the effect of buffers in a manufacturing system, subsequently allowing the optimization of the system using only a single simulation.

2. A New Frontier: Holistic Manufacturing System Analysis

The newly developed methods described below all use a holistic approach to manufacturing system analysis compared to the atomistic view of the current simulation analysis methods. The holistic methods analyze not only independent machines, but the interaction between machines. The Merriam-Webster online dictionary defines holistic and atomistic as follows:

holistic: relating to or concerned with wholes or with complete systems rather than with the analysis of, treatment of, or dissection into parts

atomistic: characterized by or resulting from division into unconnected or antagonistic fragments

2.1. The Status Quo: Atomistic Analysis

What is called generally "a simulation" consists of many different steps, as for example building a model, verifying the model, simulating, collecting data, and finally analyzing the collected data. A manufacturing system is a network of interconnected entities, as for example machines, buffers, workers, or AGV. These entities affect each other and interact with each other. To understand the system it is crucial to analyze these interactions.

However, the status quo of current manufacturing simulation analysis is usually an atomistic analysis, completely ignoring the dependencies and interactions and achieving only a fraction of the possible conclusions which are hidden in the simulation data. For example, in a standard analysis the working times of a machine is summed up to calculate the utilization, with utter disregard of the relation of the individual working times to the other machines. Similar, the idle times are summed up to extract the percentage of the time a machine is idle, again with complete disregard of the reasons a machine is idle, which is usually caused by another entity in the system.

This approach is visualized in Figure 1, where the manufacturing system is first simulated, and then broken into its individual elements for an atomistic analysis. Naturally, based on the statistical data of the independent elements it is very difficult to determine valid conclusions for the correlated system. While these results are needed to understand a manufacturing system, they represent only a fraction of the wealth of knowledge contained in the simulation data. Furthermore, as the analysis is only based on an atomistic view, it is very difficult or almost impossible to estimate how a change of one of the entities would affect the other entities and ultimately the entire system, i.e. it is very difficult to make holistic conclusions from the atomistic analysis.

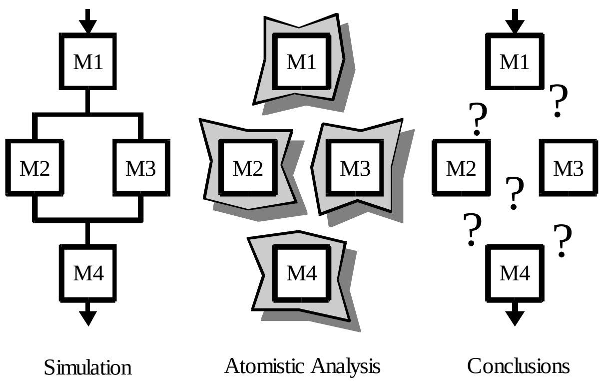

Figure 1: Atomistic Analysis Process

2.2. Holistic Analysis

The analysis presented here goes beyond the standard and is fundamentally different from the standard output analysis. The holistic analysis analyzes the interactions between the machines and statistically represents the correlations between the machines in order to understand the system. This results in a much better understanding of the system, and also allows for a prediction of the system performance based on the changes in the different entities.

This approach is visualized in Figure 2, where after the simulation of the manufacturing system, the complex interactions of the system are analyzed to obtain a holistic understanding of the system. Subsequently, valuable information about the system is obtained and it is easy to make valid conclusions for the correlated system.

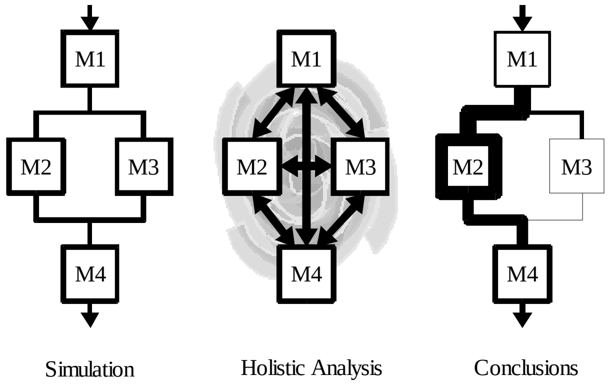

Figure 2: Holistic Analysis Process

For example the shifting bottleneck detection method as described in the next section compares the uninterrupted active (working) times of the different machines, and at any given time the machine with the longest uninterrupted active period is the bottleneck for this time. Thus the system is analyzed from a holistic point of view by comparing the different machines at different times. Subsequently, it is possible to measure the level of constrains of the machines and to make a valid prediction of the effect of a change in the machines onto the entire system.

Another example is the blocking and starving analysis as described later in this paper. The blocking and starving analysis investigates every single idle period, i.e. blocked and starved period, and tries to determine the cause of this idle period, and the path between the idle machine and the cause of the idle machine. Therefore this used a holistic view by analyzing the blocking and starving interactions between the machines. Subsequently, it was possible to make a valid prediction of the effect of a change in a buffer onto the entire system.

It can be seen that this holistic analysis offers great promise to the simulation analyst by providing valuable information about the system, allowing valid conclusions about the behavior of the system if the system changes. The current research describes only two holistic approaches based on the working times and the blocking and starving analysis, however, it is certainly possible to develop more holistic methods to answer questions about the system behavior. Of course, the holistic methods also include the results of the standard analysis, which is necessary to support the estimation of the system behavior.

While holistic methods use a slightly more complex analysis than atomistic methods, the presented methods are implemented in a analysis software and do not require an additional effort by the analyst. Overall, an holistic approach promises to obtain much more information about a system than an atomistic approach ever could, and further research in holistic analysis methods will be very useful for the industry.

3. Shifting Bottleneck detection method

The presented method will be able to detect and monitor the shifting momentary bottleneck of a production system, and also determine the average bottleneck over a selected period of time based on the duration the machines are active without interruption. More details for the different variations and uses of this method can be found in previous publications [9-13].

3.1. Holistic Analysis Of The Active Duration

The presented method is based on a holistic analysis of the duration a processing machine is active without interruption. All possible machine states are grouped into two groups, being either active states or inactive states. Similar can also be done for workers, AGV, or any other entity in the system that may cause a delay. A state is active whenever the machine may cause other machines to wait. For example working on one part may cause a subsequent idle machine to wait for the completion of the part, or a machine under repair may block previous machines. A state is inactive if the associated machine is not active but instead waiting for the completion of another task, for example the arrival of a part or service, or for the removal of a part. Similar definitions can be made for any entity in a manufacturing system, as for example workers or AGV, or any entity in a discrete event system in general.

The bottleneck detection method compares the durations of the active periods of the different machines. If the analysis is based on simulation data or historical data, it is possible to determine the durations of all active periods for all machines.

However, if the analysis is used for real time monitoring, the future is unknown and the durations of the active periods are known only until the present. In this case, the active duration is measured until the present and may be updated if further information becomes available with time.

3.2. The Momentary Shifting Bottleneck

The underlying idea of the method is that at any given time the machine with the longest uninterrupted active period is the momentary bottleneck at this time. The overlap of the active period of a bottleneck with the previous or subsequent bottleneck represents the shifting of the bottleneck from one machine to another machine. In an interconnected production system, machines block and starve each other. If a machine is active, it is neither starved nor blocked. The longer a machine is active without interruption, the more likely it is that this machine blocks or starves other machines in the production system. The machine with the longest uninterrupted active period therefore has the biggest impact onto starving or blocking the other machines, therefore being the largest constraint a.k.a. the largest bottleneck. The following method describes how to determine which machine of a production system is the sole or part of a shifting bottlenecks at any time.

Figure 3 visualizes the method using a simple example consisting of only two machines. The figure shows the active periods of the machines over a short period of time. At the selected time *t*, both machines M1 and M2 are active. Yet, as M1 has the longer active period, M1 is the bottleneck machine for the time t. As there is no machine active before the current bottleneck period, there is no overlap and no shifting at the beginning of the current bottleneck period. However, at the end of the current bottleneck period, M2 is active and has the longest active period. Therefore, the subsequent bottleneck machine is M2. During the overlap between the current bottleneck period and the subsequent bottleneck period the bottleneck shifts from M1 to M2. Now, M2 is the bottleneck machine. Similarly, at the end of the bottleneck period of M2, the bottleneck shifts back to M1. Processing all available data using this method shows at what time which machine is the bottleneck machine, when the bottleneck is shifting, and when there is no bottleneck at all. Therefore, it is possible to detect and monitor the bottleneck at all times.

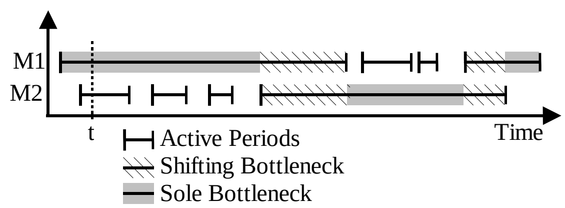

Figure 3: Shifting Bottlenecks

3.3. The Bottleneck Probability

The above method detects and monitors the momentary bottleneck at any instant of time. However, in many cases it may be of interest not to investigate an instant of time but rather a period of time. To determine the bottleneck during a period of time the available data is analyzed and the momentary bottlenecks are determined over the selected period of time. Next, the percentage of time a machine is the sole bottleneck machine and the percentage of the time a machine is part of a shifting bottleneck is measured for the selected period of time.

Figure 4 visualizes this method using the example with two machines as shown in Figure 3. M1 is the sole bottleneck more often than M2, and is also involved in a number of shifting operations. M2 is the smaller constraint, i.e., a secondary bottleneck, having being the sole bottleneck for a smaller percentage of time. Overall, an improvement of the throughput of M1 would yield a larger overall improvement of the system throughput than an improvement of M2, as M1 is the primary bottleneck during the selected period of time.

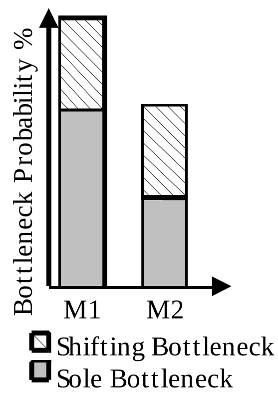

Figure 4: Average Bottleneck over Period of Time

4. Buffer Allocation Model

Buffers improve the system throughput by reducing the idle time (blocking and starving) of the machines. Therefore, to understand the buffers it is crucial to understand the blocking and starving of the machines, the causes thereof, and, most important the path to the causes and the buffer locations in between. This method analyzes every starving or blocking occurrence of every machine in the simulation, and finds the cause of the starving and blocking, and, more important, the buffer locations on the path between the idle machine and the cause thereof. The time a possible buffer location is part of a path is determined for each machine.

4.1. Holistic Starving And Blocking Analysis

To find the cause of an idleness of a machine, an algorithm has been developed that follows the cause from machine to machine or buffer until the cause of the idle period has been found.

A machine is assumed to be always either active (A), blocked (B) or starved (S). The cause of a block can always be found by following from the blocked machine downstream. If the downstream machine is also blocked or the buffer is full, continue following the block downstream. If the downstream machine is active, or the downstream buffer is not full, the cause of the block has been found. If the downstream machine is starved, it is necessary to turn around and find the cause of the starving period. An overview of the 5 possible situations is given in Figure 5.

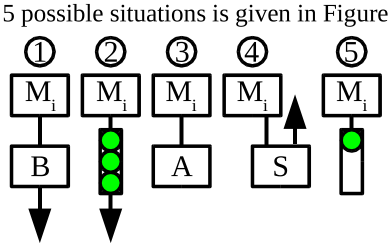

Figure 5: Situations with Blocking of Machine M_i

For starved machines, there are also a total of 5 possible situations, determining the next machine/buffer in the search for the cause of the starved machine. The cause of a starving situation is always sought upstream. If the upstream machine is also starved, or the upstream buffer is empty, continue searching for the cause upstream. If the upstream machine is active or the upstream buffer is not empty, the cause of the starving situation has been found. If the upstream machine is blocked, it is necessary to turn around and find the cause of the block for the upstream machine. An overview of the situations is given in Figure 6.

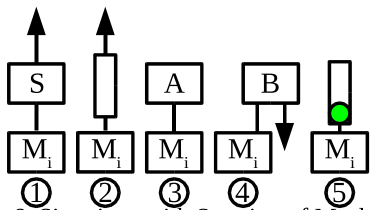

Figure 6: Situations with Starving of Machine M_i

Using this set of rules, it is possible to find the cause of an idle period for all idle periods of all machines, and to determine the period of time a buffer location was part of the path to the cause of an idle machine. This allows the conclusion of the effect of a buffer onto the different machines.

An example system with 7 machines has been analyzed, and the causes of the blocking and starving of the machines has been established. Figure 7 presents the results for machines M3 and M5 in graphical form, showing the path of the starves (cross-hatched) and blocks (diagonal-hatched) from machine M3 and machine M5 to the machine causing the starve or block. The width of the path represents the fraction of the starves/blocks following this path.

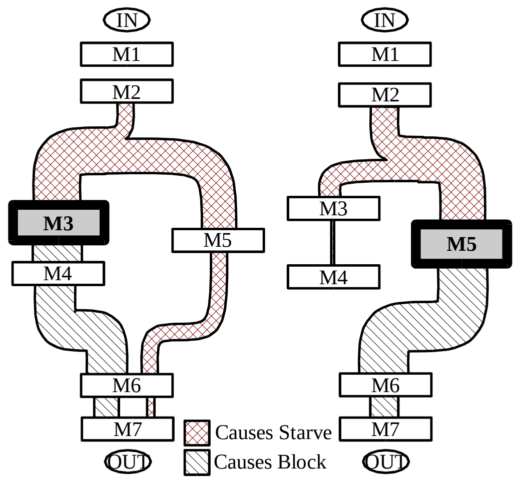

Figure 7: Causes of Blocking and Starving of Machines M3 and M5

In the simulated example, machine M3 is blocked 6.1% of the time. Whenever machine M3 is blocked, the path to the cause of the block leads to the next downstream machine M4 (100% of the blocked time). However, M4 itself is rarely the cause of the block. Most the paths continue to machine M6 (78% of the blocked time), and M7 (46% of the blocked time). Therefore, a buffer increase before machine M7 affects the blocking of machine M3 46% of the time. Machine M3 is also starved for 5.8% of the time. The path to the cause of the starve splits, with 38% of the starving periods caused by machine M2, and 62% following to machine M5. From machine M5 the paths continue to M6 (27% of the starving time), and from there to M7 (15% of the starving time). The causes of the starving and blocking of machine M5 can be traced in a similar style. The path between the idle machines and the cause thereof allows an estimation of the effect of buffers. Only buffers in these path affect the machines.

4.2. Buffer Prediction Model

The blocking and starving analysis determines which buffer affects which machine and how often. Combined with an estimation of the number of parts in the buffer, the effect on the machines can be determined. As a next step, the bottleneck probability can be used to determine the effect onto the entire system based on the effect onto the individual machines.

The prediction model for the effects of buffers has been validated extensively for a number of different systems and buffers [14]. For example, Figure 8 shows the comparison of the predicted time per part to the measured time per part for a buffer BM3 located before machine M3 in the same manufacturing system as shown in Figure 7. The continuous line shows the measured data including the 95% confidence intervals, and the dotted lines shows the predicted system performance. The predicted performance follows the measured data very nicely. The overall root mean squared error RMSE was only 0.24s.

Roser, C., Nakano, M., Tanaka, M., 2004. Holistic Manufacturing System Analysis, in: International Conference on Advanced Mechatronics. Asahikawa, Hokkaido, Japan.

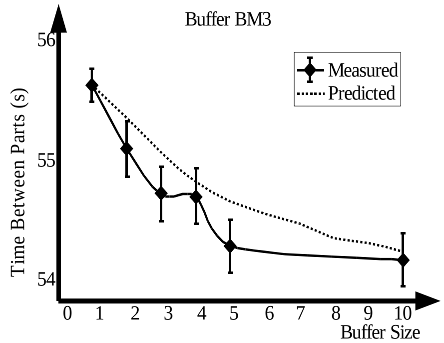

Figure 8: Performance Prediction for Buffer BM3

4.3. Buffer Optimization

The buffer prediction model can also easily be used to optimize a manufacturing system. As the prediction model allows the rapid comparison of many different buffer allocations, a wide variety of optimization methods can be used, as for example a gradient based method or a genetic algorithm. A good description of a wide range of optimization methods can be found in [15].

Two different optimization approaches are possible. One approach is to use only a single simulation to create a prediction model for the optimization. The other approach uses a multi-step optimization, where the prediction model is used for a local optimization, after which a new simulation verifies the system and is itself optimized again. Figure 9 shows a multi step optimization for a system similar to Figure 7, where the step size is limited to 15 buffer spaces per buffer. The simulation quickly reached an optimal plateau after 4 steps, and no further improvement was possible after step 13.

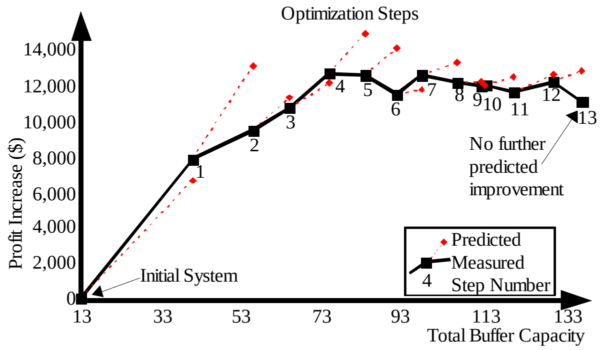

Figure 9: Multi Step Buffer Optimization

5. Conclusions

This paper summarizes a medley of methods recently developed by the Toyota Central Research and Development Laboratories related to the analysis of manufacturing system simulations. Unique to all the methods is the underlying holistic approach of analyzing not only the entities of the manufacturing system but also the interactions between these entities. All of the methods have been validated and implemented in an automated analysis software used by some companies in the Toyota group. This software creates an easy to use MS Excel file containing all the data and prediction models for optimizations.

The bottleneck detection method determines the momentary shifting bottleneck based on a holistic comparison of the machine working times. This provides a quantitative measure of the constraints, allowing the subsequent use of the bottleneck probability for a sensitivity analysis and a prediction model. Overall, this method greatly enhances the understanding of the system by quantitatively determining the primary and secondary bottlenecks and the non-bottlenecks.

The buffer prediction model is based on a detailed holistic analysis of the blocking and starving situations in the manufacturing system, determining the cause of every idle period in the system. This allows for a subsequent prediction model of the effect of buffers onto the system and a optimization of the buffer allocations. This prediction model greatly reduces the time needed to understand and improve the effect of buffers in the system.

Roser, C., Nakano, M., Tanaka, M., 2004. Holistic Manufacturing System Analysis, in: International Conference on Advanced Mechatronics. Asahikawa, Hokkaido, Japan.

Further research includes the development of additional methods in the promising field of holistic simulation analysis. By statistically analyzing the interactions in a manufacturing system or any discrete event system in general, a much deeper understanding of the system can be obtained. If the relations between the system entities is understood, the effect of changes in one entity can be estimated, greatly benefiting the industry by improving their competitiveness.

6. References

[1] M. Fu, S. Andradottir, J. S. Carson, F. Glover, C. R. Harrell, Y.-C. Ho, J. P. Kelly, and S. M. Robinson, "Integrating Optimization and Simulation: Research and Practice," presented at Winter Simulation Conference, Orlando, Florida, USA, 2000.

[2] J. Boesel, R. O. B. Jr., F. Glover, J. P. Kelly, and E. Westwig, "Future of Simulation Optimization," presented at Winter Simulation Conference, Arlington, Virginia, USA, 2001.

[3] F. Azadivar, "Simulation Optimization Methodologies," presented at Winter Simulation Conference, Phoenix, AZ, USA, 1999.

[4] J. R. Swisher, P. D. Hyden, S. H. Jacobson, and L. W. Schruben, "A Survey of Simulation Optimization Techniques and Procedures," presented at Winter Simulation Conference, Orlando, Florida, USA, 2000.

[5] J. F. I. Cox and M. S. Spencer, *The Constraints Management Handbook*. Boca Raton, Florida: CRC Press - St. Lucie Press, 1997.

[6] R. W. Conway, W. Maxwell, J. O. McClain, and L. J. Thomas, "The role of work-in-process inventory in serial production lines," *Operations Research*, vol. 36, pp. 229-241, 1988.

[7] M. Caramanis, H. Pan, and O. Anli, "Is there a Trade off between Lean and Agile Manufacturing? A Supply Chain Investigation," presented at Third Aegean International Conference on "Design and Analysis of Manufacturing Systems", Tinos Island, Greece, 2001.

[8] D. Brittan, "When Bad Things Happen to Good Factories," *Technology Review*, 1996.

[9] C. Roser, M. Nakano, and M. Tanaka, "Shifting Bottleneck Detection," presented at Winter Simulation Conference, San Diego, CA, USA, 2002.

[10] C. Roser, M. Nakano, and M. Tanaka, "Detecting Shifting Bottlenecks," presented at International Symposium on Scheduling, Hamamatsu, Japan, 2002.

[11] C. Roser, M. Nakano, and M. Tanaka, "Tracking Shifting Bottlenecks," presented at Japan-USA Symposium on Flexible Automation, Hiroshima, Japan, 2002.

[12] C. Roser, M. Nakano, and M. Tanaka, "Throughput Sensitivity Analysis using a single simulation," presented at Winter Simulation Conference, San Diego, CA, USA, 2002.

[13] C. Roser, M. Nakano, and M. Tanaka, "Constraint Management in Manufacturing Systems," *International Journal of the Japan Society of Mechanical Engineering, Series C, Special Issue on Advanced Scheduling*, vol. 46, pp. 73-80, 2003.

[14] C. Roser, M. Nakano, and M. Tanaka, "Buffer Allocation Model based on a Single Simulation," presented at Winter Simulation Conference, New Orleans, Louisiana, USA, 2003.

[15] G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, *Optimization*, vol. 1. Amsterdam: Elsevier Science, 1994.