Journal of Manufacturing Systems 47 (2018) 43–52

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

A systematic-theoretic analysis of data-driven throughput bottleneck detection of production systems

Lin Li

Department of Mechanical and Industrial Engineering, University of Illinois at Chicago, Chicago, IL 60607, USA

ARTICLE INFO

Article history:
Received 6 September 2017
Received in revised form
16 November 2017
Accepted 1 March 2018

Keywords:
Systematic-theoretic analysis
Data-driven
Throughput bottleneck detection
Production system

ABSTRACT

Throughput is one of the most critical performance indices for design, control, and operation management of production systems. Throughput bottleneck greatly impedes the overall performance of modern production systems. However, detecting throughput bottlenecks of production systems is a complicated task due to the complexity of production system dynamics. In this paper, a new data-driven bottleneck detection method is proposed based on rigorous mathematical proof for general serial production systems, which uses the routinely available industrial data on the plant floor to identify the throughput bottleneck location within production systems in both the short-term (transient) and long-term (steady-state) periods. Case studies are conducted to illustrate the effectiveness of the proposed method. The research outcomes will enhance the intelligent decision-making and real-time operation management capabilities in the modern manufacturing enterprises.

1. Introduction

Throughput plays an important role in the design, control, and management of modern production systems. It characterizes the average number of parts produced by a production system within a specific time period. Due to many random events in production (e.g., random machine failures and varied processing time), the throughput of a production system is also random. In the literature, both analytical and simulation-based methods have been developed to estimate the expected value of the throughput (e.g., [1–8]), which is an important metric for optimal design, continuous improvement, and excellent management of production systems [2,3]. Since the overall system throughput can be increased more significantly by improving the bottleneck machine's performance than other machines', bottleneck analysis and detection are of high interest in both academia and industry. In fact, many implementations of bottleneck detection methods that can improve system performance in real-world manufacturing facilities have been reported. For instance, General Motors has increased revenue and saved over $2.1 billion by developing algorithms to identify bottlenecks, improve throughput performance, and optimize buffer allocation [9]. Trane U.S. Inc., a brand of Ingersoll Rand, has saved more than $0.7 million per year by using operations research tools to detect the bottleneck and increase throughput [10].

In academia, a substantial amount of research has been devoted to the area of bottleneck analysis for system performance improvement. One of the earliest theoretical analyses was performed by Goldratt [11], who developed the Theory of Constraints to provide a systematic tool for bottleneck analysis and system performance improvement. This theory has been widely developed and deeply rooted in the manufacturing industry [12]. Many different bottleneck detection methods have been proposed and developed based on this theory and its subsequent analysis. In general, these methods can be divided into two main categories: analytical methods and simulation-based methods.

In the analytical category, a mathematical model was developed by Meerkov and his colleagues [13], and three different bottlenecks (up-time, down-time, and cycle-time bottlenecks) were defined and an arrow-based identification method derived from the mathematical model was proposed [14–16]. Based on these results, Biller et al. [17] proposed a method to identify the bottlenecks in production lines with rework layout. In addition, Yan et al. [18] developed a method to identify the bottlenecks in the Knowledgeable Manufacturing System (KMS). In the simulation-based category, a general bottleneck identification method applied to discrete event systems was proposed by Toyota Research Laboratory [19,20]. The average duration of a machine being active or inactive was recorded. The bottlenecks and

E-mail address: linli@uic.edu

https://doi.org/10.1016/j.jmsy.2018.03.001

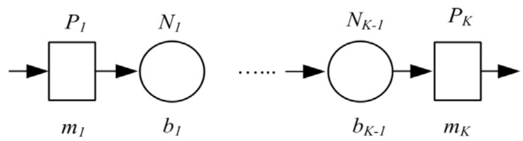

Fig. 1. A typical serial production line with *K* machines and *K*-1 buffers.

bottleneck shifting are determined by comparing the duration of a machine being active without interruption. Muthiah & Huang [21] also introduced a factory-level throughput effectiveness metric for productivity bottleneck detection and system performance improvement. Many efforts have also been devoted to other related research areas such as bottleneck-based heuristics for job scheduling [22–24] and real-time energy management [25].

Although remarkable progress about bottleneck detection and analysis has been achieved, there are still many problems that are not fully explored. First, most existing bottleneck detection methods are based on heuristics and lack the analytical proof. Second, most of these previously mentioned bottleneck detection methods only focus on the long-term analysis. With the development of new informatics technologies such as Internet of Things and Cloud Computing, large number of sensors are increasingly mounted in modern factories to collect the data of real-time performance of manufacturing systems [1]. Recently, Li et al. [2,3,26] have presented a novel approach based on the measurable online data for throughput bottleneck detection that can be used in both the short term (transient) and long term (steady state) period. However, the analytical proof is based on a three-machine-no-buffer system rather than a general N-machine-N-1-buffer system, which is oversimplified to describe the system's complex behaviors.

Based on the previous work, the main objective of this paper is to propose a new data-driven method for throughput bottleneck detection of production systems in both the short-term and long-term period, and prove the new method rigorously and analytically. Case studies will be carried out to illustrate the effectiveness of the proposed method. The research outcome has the potential to be integrated into the manufacturing execution systems [27] to improve the production and operation management of modern manufacturing enterprises.

The rest of this paper is organized as follows. Section 2 introduces the new data-driven bottleneck detection method. Section 3 presents the mathematical derivation of key performance indices of general serial production systems. The deduction of the bottleneck detection method based on the derived performance indices is shown in Section 4. Section 5 presents case studies for illustration. Finally, conclusions and future works are summarized in Section 6.

2. New bottleneck detection method

Intuitively, the bottleneck of a production system is recognized as the machine that has the strongest impact on the overall system performance. Considering both local and global properties of the production system, the throughput bottleneck can be defined based on sensitivity. Many variants of bottleneck definition exist in the literature. One of the most commonly cited definitions is as follows [14]:

The throughput bottleneck is defined as the machine if the partial derivative of the system production rate with respect to the machine's up time is the highest. That is, machine k would be the bottleneck in a serial production line with K machines and K−1 buffers if

$$k = \arg\max_i \left(\frac{\partial PR_{sys}}{\partial TUP_i}\right),\ i = 1, 2, \ldots, K$$

where PR_{sys} is the system production rate and TUP_i is the up time of machine i over a specific time period.

Following this definition of throughput bottleneck to address the shortcomings in the literature [2,3], the proposed new data-driven bottleneck detection method is given as follows.

Definition 1. *For a typical serial production system with K machines and K-1 buffers as shown in Fig. 1, machine j is the bottleneck during a time period if the following relations hold:*

for the general case $1 < j < K$:

(a) $TB_i - TS_i > 0$ *for* $i < j$ *and* $TB_i - TS_i < 0$ *for* $i > j$;
(b) $TUP_j - TB_j < TUP_{j-1} - TS_{j-1}$ or $TUP_j - TS_j < TUP_{j+1} - TB_{j+1}$;

for the special case $j = 1$:

(c) $TB_1 - TS_1 > 0$ and $TB_2 - TS_2 < 0$ and $TUP_1 < TUP_2 - TB_2$;

for the special case $j = K$:

(d) $TB_{K-1} - TS_{K-1} > 0$ and $TB_K - TS_K < 0$ and $TUP_{K-1} - TS_{K-1} > TUP_K$;

where TUP_j, TB_j, and TS_j are the up, blockage, and starvation time, respectively, for machine j over a specific time period, either in the short term or long term.

There are two conditions for the general case ($1 < j < K$) in this new method. In the first condition (a), two adjacent machines are defined as possible bottleneck candidates if their upstream machines have higher blockage than starvation and their downstream machines have lower blockage than starvation. In the second condition (b), the up time minus starvation time of the upstream machine and the up time minus the blockage time of the downstream machine of the two candidates are estimated and compared. The machine with the smaller estimated value between the two possible candidates is the bottleneck.