Assembly systems with non-exponential machines: Throughput and bottlenecks

ShiNung Ching, Semyon M. Meerkov*, Liang Zhang

Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122, USA

Abstract

Manufacturing equipment on the factory floor is typically unreliable and the buffers are finite. This makes production systems stochastic and nonlinear. Numerous studies addressed the performance analysis of production systems and, in particular, assembly systems, assuming that up- and downtimes of the machines are exponentially distributed. Empirical evidence indicates, however, that the coefficients of variation of up- and downtimes are often less than 1 and, hence, the statistics are not exponential. This paper provides methods for throughput analysis and bottleneck identification in assembly systems with non-exponential machines. Since the resulting systems are non-Markovian, analytical investigation is all but impossible. Therefore, a numerical approach is pursued, based on simulations of systems at hand and a subsequent analytical approximation of the results obtained. This leads to a closed formula for the throughput and a practical method for bottleneck identification.

Keywords: Assembly systems; Unreliable machines; Finite buffers; Throughput; Bottlenecks; Simulations

1. Introduction

Assembly systems are two or more serial production lines interacting through one or more merge operations. An example of such a system is shown in Fig. 1.1 where the circles are the machines and the rectangles are the buffers; machine $m_{01}$ is the merge operation. We assume that the machines have fixed processing times but experience random breakdowns; the buffers are assumed to be finite. These assumptions, which typically take place in large volume manufacturing (e.g., automotive, electronics, etc.), make assembly systems stochastic and nonlinear.

When the breakdown and repair rates of the machines are constant, the resulting distributions of up- and downtimes are exponential, and the assembly systems can be analyzed using Markov process techniques [1–3]. Empirical evidence indicates, however, that in reality the machines are not exponential since, as it is shown in [4], the coefficients of variation of up- and downtimes ($CV_{up}$ and $CV_{down}$) are less than 1. This implies that the breakdown and the repair rates are increasing functions of time [5]. Methods for performance analysis of assembly systems with such machines are unknown. This paper is intended to contribute to this end.

Since no analytical methods for non-exponential case are available, a numerical approach is pursued. Specifically, based on the evidence obtained by simulations, we show that, if $CV_{up}$ and $CV_{down} < 1$, the throughput of an
assembly system \((TP = \text{the average number of parts produced by the last machine per unit of time in the steady state})\) is practically independent of particular distributions involved and can be approximated by a linear function of the “aggregated” coefficient of variation, with the slope defined by \(TP^0\) and \(TP^{exp}\), where \(TP^0\) and \(TP^{exp}\) are the throughputs of the system with the deterministic and exponential up- and downtimes, respectively.

Another problem addressed in this paper is that of bottleneck identification. The bottleneck (BN) has been defined in [6] as the machine with the largest effect on \(TP\), as quantified by its partial derivatives with respect to the machine parameters. For assembly systems with exponential machines, a method for BN identification has been proposed in [7]. We show in this paper that this method can be used for non-exponential machines as well.

The outline of this paper is as follows: Section 2 provides the model and the problem formulation. Sections 3 and 4 address the issues of throughput and bottlenecks, respectively. The conclusions are formulated in Section 5.

2. System model and problem formulation

2.1. Model

Let the assembly system shown in Fig. 1.1 operate according to the following assumptions:

(i) The system consists of two component lines, \(m_{ij}\), \(i = 1, 2, j = 1, \ldots, M\), a merge machine, \(m_{01}\), additional processing line, \(m_{0j}\), \(j = 2, \ldots, M_0\), and buffers, \(b_{ij}\), \(i = 1, 2, j = 1, \ldots, M_i\); \(b_{0j}\), \(j = 1, \ldots, M_0 - 1\). For convenience, we use \(I_m\) and \(I_b\) to denote the sets of all machines and buffers, respectively.

(ii) Each machine \(m_{ij}\), \(ij \in I_m\), has two states: up and down. When up, the machine is capable of producing with rate \(c_{ij}\) (parts/unit of time); when down, no production takes place.

(iii) The up- and downtimes of each machine are continuous random variables, \(t_{up,ij}\) and \(t_{down,ij}\), \(ij \in I_m\), with arbitrary unimodal probability density functions, \(f_{up,ij}(t)\) and \(f_{down,ij}(t), t \geq 0\), and the coefficients of variation denoted as \(CV_{up,ij}\) and \(CV_{down,ij}\), \(ij \in I_m\). It is assumed that these random variables are mutually independent.

(iv) Each buffer \(b_{ij}\), \(ij \in I_b\), is characterized by its capacity \(0 < N_{ij} < \infty\).

(v) Machine \(m_{ij}\), \(i = 0, 1, 2, j > 1\), is starved at time \(t\) if it is up at time \(t\), buffer \(b_{i(j-1)}\) is empty at time \(t\). Machine \(m_{01}\) is starved if at least one of the buffers \(b_{1M_1}\) or \(b_{2M_2}\) is empty at time \(t\). Machines \(m_{11}\) and \(m_{21}\) are never starved.

(vi) Machine \(m_{ij}\), \(ij \notin \{0M_0, 1M_1, 2M_2\}\), is blocked at time \(t\) if it is up at time \(t\), buffer \(b_{ij}\) is full at time \(t\) and machine \(m_{i(j+1)}\) fails to take any work from this buffer at time \(t\). Machine \(m_{iM_i}, i = 1, 2\), is blocked if it is up, buffer \(b_{1M_i}\) is full and merge machine \(m_{01}\) fails to take any work from the buffers at time \(t\). Machine \(m_{0M_0}\) is never blocked.

Let \(T_{up,ij}\) and \(T_{down,ij}\) be the average up- and downtimes of the machine \(m_{ij}\), \(ij \in I_m\). Then its efficiency and throughput in isolation are given, respectively, by

\[
e_{ij} = \frac{T_{up,ij}}{T_{up,ij} + T_{down,ij}}. \tag{2.1}
\]

\[
TP_{iso,ij} = c_{ij}e_{ij}. \tag{2.2}
\]

For the purposes of this paper, the performance of the system defined by assumptions (i)–(vi) is characterized in terms of the following steady state characteristics:

- \(TP\) — average number of parts produced by \(m_{0M_0}\) per unit of time,
- \(ST_{ij}\) — probability of starvation of the machine \(m_{ij}, ij \neq 01\),
ST_{01} — probability of starvation of the machine \( m_{01} \) due to empty \( b_{1M_1} \),

ST_{01} — probability of starvation of the machine \( m_{01} \) due to empty \( b_{2M_2} \),

BL_{ij} — probability of blockage of the machine \( m_{ij} \), \( i, j \in I_m \).

Analysis of these characteristics leads to solutions of the problems formulated below.

2.2. Problems addressed

Clearly, the throughput and the probabilities of blockages and starvations are functionals of up- and downtime distributions parameterized by buffer capacities. Since exact evaluation of these functionals is all but impossible, the first problem addressed in this paper is to provide a closed-form analytical expression for an estimate of \( TP \) and evaluate its accuracy.

To formulate the second problem, introduce

**Definition 2.1.** Machine \( m_{ij} \) is the \( c \)-bottleneck (\( c \)-BN) of an assembly system if

\[
\frac{\partial TP}{\partial c_{ij}} > \frac{\partial TP}{\partial c_{kl}}, \quad \forall \ kl \neq ij.
\]  

(2.3)

Unfortunately, this definition can hardly be used in practice, since the partial derivatives involved in (2.3) can be neither measured on the factory floor nor calculated analytically. Therefore, the second problem of this work is to provide a method for \( c \)-BN identification, which avoids the differentiation of \( TP \) and can be applied based on factory floor measurements.

3. Throughput evaluation

Let \( TP^0 \) and \( TP^{exp} \) be the throughputs of an assembly system with up- and downtimes being deterministic (i.e., \( CV_{up,ij} = CV_{down,ij} = 0 \)) and exponentially distributed, respectively. Motivated by the results of [8] for serial lines, introduce an estimate, \( \hat{TP} \), of the throughput for arbitrary unimodal distributions of up- and downtimes as follows:

\[
\hat{TP} = TP^0 - (TP^0 - TP^{exp})CV_{ave},
\]  

(3.1)

where

\[
CV_{ave} = \frac{\sum_{ij \in I_m} (CV_{up,ij} + CV_{down,ij})}{2(M_0 + M_1 + M_2)}.
\]

Define the accuracy of this estimate as

\[
\Delta_{TP} = \frac{\vert \hat{TP} - TP \vert}{TP}.
\]  

(3.2)

**Proposition 3.1.** Assume that in the assembly system defined by assumptions (i)–(vi), \( CV_{up,ij} \leq 1 \) and \( CV_{down,ij} \leq 1 \). Then, \( \Delta_{TP} \ll 1 \).

**Justification:** A set of 10,000 assembly systems has been generated with the parameters selected randomly and equiprobably from the sets

\[
M_0, M_1, M_2 \in \{1, 2, 3, 4, 5\},
\]  

(3.3)

\[
T_{down,ij} \in \{1, 5\},
\]  

(3.4)

\[
e_{ij} \in (0.75, 0.95],
\]  

(3.5)

\[
c_{ij} \in (0.8, 1.2],
\]  

(3.6)

\[
N_{ij} \in \{2, 3, \ldots, 10\}.
\]  

(3.7)
For each of these systems, 50 additional assembly systems have been constructed by selecting the distributions of the machines’ up- and downtimes and the coefficients of variation randomly and equiprobably from the following sets

\[
 f_{\text{up},ij}, f_{\text{down},ij} \in \{\text{exp}, W, ga, LN\},
\]

\[
 CV_{\text{up},ij}, CV_{\text{down},ij} \in [0.1, 1].
\]  

where \(\text{exp}, W, ga\) and \(LN\) stand for the exponential, Weibull, gamma, and log-normal distributions:

\[
 f_{\text{up}}(t) = \lambda e^{-\lambda t}, \quad f_{\text{down}}(t) = \mu e^{-\mu t}, \quad t \geq 0, \quad \text{for exponential},
\]

\[
 f_{\text{up}}(t) = \lambda A e^{-\lambda t} \Lambda t^{A-1}, \quad f_{\text{down}}(t) = \mu M e^{-\mu t} \Gamma M^{M-1}, \quad t \geq 0, \quad \text{for Weibull},
\]

\[
 f_{\text{up}}(t) = \lambda e^{-\lambda t} \left(\frac{\Gamma A}{\Lambda t^{A-1}}\right), \quad f_{\text{down}}(t) = \mu e^{-\mu t} \left(\frac{\Gamma M}{\mu t^{M-1}}\right), \quad t \geq 0, \quad \text{for gamma},
\]

\[
 f_{\text{up}}(t) = \frac{1}{\sqrt{2\pi M t}} e^{-\frac{(\ln t - \lambda)^2}{2\sigma^2}}, \quad f_{\text{down}}(t) = \frac{1}{\sqrt{2\pi M t}} e^{-\frac{(\ln t - \mu)^2}{2\sigma^2}}, \quad t \geq 0, \quad \text{for log-normal}
\]

and \(\lambda, \mu, A, M\) are positive constants.

For each of the 500,000 assembly systems, thus constructed, \(TP\) and \(\hat{TP}\) have been evaluated, respectively, by simulations and by calculations using (3.1). The accuracy of the approximation has been evaluated by (3.2). The simulations used a C++ code representing the system at hand and the following

**Simulation procedure 3.1:**

- Select initial state of each machine up with probability \(e_{ij}\) and down with probability \(1 - e_{ij}\), \(ij \in I_m\).
- For each line under consideration, carry out 20 runs of the simulation code.
- In each run, use the first 100,000 time slots as a warm-up period and the subsequent 1,000,000 time slots to statistically evaluate \(TP, ST_{ij}\) and \(BL_{ij}\); this results in 95% confidence intervals of less than 0.001 for all performance measures.

This approach is illustrated by a system shown in Fig. 3.1. For this system, the simulations result in \(TP^0 = 0.7066\) and \(TP^{\text{exp}} = 0.6284\). Fig. 3.2 shows the linear function representing expression (3.1). To investigate the accuracy of this expression for the system at hand, fifty additional systems are generated and their \(TPs\), obtained by simulations, are indicated in Fig. 3.2. Clearly, these \(TPs\) are close to that provided by (3.1), with the largest error being less than 0.025.

Analyzing in a similar manner all 500,000 assembly systems under consideration, we determine that the average value of \(\Delta TP\) is 0.0211 and the maximum one is 0.0676. Thus, we conclude that Proposition 3.1 is justified.

4. **Bottleneck identification**

Consider the assembly systems shown in Figs. 4.1 and 4.2 and assume that the probabilities of machine blockages and starvations, shown therein, are either measured on the factory floor or evaluated by simulations. Assign arrows directed from one machine to another according to the following rule:

- if \(BL_{ij} > ST_{i(j+1)}, ij \neq 1M_1, 2M_2, 0M_0\), direct the arrow from \(m_{ij}\) to \(m_{i(j+1)}\);
- if \(ST_{ij} > BL_{i(j-1)}, ij \neq 11, 21, 01\), direct the arrow from \(m_{ij}\) to \(m_{i(j-1)}\).
Motivated by the results of [9] for serial lines, we formulate

**c-BN Indicator**: In an assembly system defined by assumptions (i)–(vi) with $\text{CV}_{\text{up},ij} \leq 1$ and $\text{CV}_{\text{down},ij} \leq 1$, the c-BN is a machine with no emanating arrows.

**Proposition 4.1.** The c-BN Indicator holds with frequency $1 - \epsilon$, where $\epsilon \ll 1$.

**Justification:** A set of 100,000 assembly systems has been generated with parameters selected randomly and equiprobably from sets (3.3)–(3.9). For each of these systems,
• c-BN was identified using Definition 2.1, with partial derivatives evaluated using Simulation Procedure 3.1 as
\[
\frac{\partial TP}{\partial c_{ij}} \approx \frac{\Delta TP}{\Delta c_{ij}},
\]
with \(\Delta c_{ij} = 0.03\);
• machines with no emanating arrows were identified using the probabilities of machine blockages and starvations obtained by simulations.

Analyzing the results for all 100,000 assembly systems considered, we determine that a single machine with no emanating arrows is indeed the c-BN in 90.1% of cases, and the set of the multiple machines with no emanating arrows contains the c-BN in 96.6% of cases (see Figs. 4.1 and 4.2 for examples). Thus, we conclude that the c-BN Indicator holds with frequency \(1 - \epsilon\), where \(\epsilon < 0.1\).

**Remark.** In practice, it is often assumed that the bottleneck is the machine with lowest throughput in isolation, i.e., the machine defined by
\[
TP_{iso,ij} = \min_{kl \in I_m} TP_{iso,kl}.
\]

It turns out, however, that it is seldom the case: Among the 100,000 assembly systems analyzed, only 27% had the worst machine as the bottleneck in the sense of Definition 2.1. Thus, the c-BN Indicator provides a substantially more accurate tool for bottlenecks’ identification.

5. Conclusions

This paper offers practical methods for analysis and improvement of assembly systems with non-exponential machines. As far as the throughput is concerned, it offers a simple formula for \(TP\) evaluation without knowing the statistical models of machine reliability, but just based on their first two moments. As far as the bottlenecks are concerned, it offers a way for c-BN identification without knowing machine and buffer parameters, but just using the

---

**Fig. 4.2.** Multiple machines with no emanating arrows.

- **c-BN** was identified using **Definition 2.1**, with partial derivatives evaluated using Simulation Procedure 3.1 as

\[
\frac{\partial TP}{\partial c_{ij}} \approx \frac{\Delta TP}{\Delta c_{ij}},
\]

with \(\Delta c_{ij} = 0.03\);
- **machines** with no emanating arrows were identified using the probabilities of machine blockages and starvations obtained by simulations.

Analyzing the results for all 100,000 assembly systems considered, we determine that a single machine with no emanating arrows is indeed the **c-BN** in 90.1% of cases, and the set of the multiple machines with no emanating arrows contains the **c-BN** in 96.6% of cases (see **Figs. 4.1 and 4.2 for examples**). Thus, we conclude that the **c-BN Indicator** holds with frequency \(1 - \epsilon\), where \(\epsilon < 0.1\).

**Remark.** In practice, it is often assumed that the bottleneck is the machine with lowest throughput in isolation, i.e., the machine defined by

\[
TP_{iso,ij} = \min_{kl \in I_m} TP_{iso,kl}.
\]

It turns out, however, that it is seldom the case: Among the 100,000 assembly systems analyzed, only 27% had the worst machine as the bottleneck in the sense of **Definition 2.1**. Thus, the **c-BN Indicator** provides a substantially more accurate tool for bottlenecks’ identification.

**5. Conclusions**

This paper offers practical methods for analysis and improvement of assembly systems with non-exponential machines. As far as the throughput is concerned, it offers a simple formula for \(TP\) evaluation without knowing the statistical models of machine reliability, but just based on their first two moments. As far as the bottlenecks are concerned, it offers a way for c-BN identification without knowing machine and buffer parameters, but just using the
frequencies of machine blockages and starvations. Both of these results are justified numerically. Analytical proofs of these results, however challenging they may be, would be of significant interest.

References