AN OPEN QUEUEING NETWORK MODEL OF MANUFACTURING SYSTEMS WITH INDEPENDENT PRODUCTION BUFFERS

Kai Furmans
Institut für Fördertechnik, Forschungsbereich Materialfluß
Universität Karlsruhe
W 7500 Karlsruhe, Germany

ABSTRACT

This paper presents a new way of modelling production buffers in manufacturing systems. The first section describes several possibilities to configure buffers in a manufacturing system. The following section contains the mathematical model and the assumptions that have to be made to get exact results. A brief explanation of the necessary operations to calculate performance measures for the network follows. The paper closes with an example that demonstrates a possible application of the presented model.

PROBLEM DESCRIPTION

The purpose of the paper is to present a modelling technique appropriate for manufacturing systems with manual or automatic material handling. Manufacturing systems usually have only limited buffers for storing unprocessed material. Especially shop floor space around machines is scarce; usually there exists a designated and limited area for storing unfinished work close to a machine or a group of machines. The machines and their allocated storage spaces can be modelled as queueing systems, the servers representing machines, the storage spaces being buffers.

Manufacturing systems consisting of many machines are often analytically treated as queueing networks with finite buffer capacity, so called blocking networks. The blocking strategy describes the behavior of jobs, which are influenced by the fact, that after completing service in one station the buffer of their destination station is full. Several papers have been published, analyzing that problem. The blocking strategies they deal with could be divided into three categories [Akyildiz 1989]: transfer blocking, service blocking and rejection blocking.

All these blocking strategies reduce the throughput of the queueing network in a way which is often practically not acceptable. If more material gathers at the machine than could be stored, other areas (walkways, fire-exits etc.) are often used to stockpile unfinished work. This leaves the manufacturing system in a state called an "unplanned state". In order to obtain a safe and smooth manufacturing operation, the probability of being in an unplanned state should be kept at a low level.

To solve this problem production buffers are added, that are not attached to a specific server but could be used to store jobs, that are not allowed to enter the local buffer of their destination system. This kind of buffers is subsequently referred to as "independent buffers". Jobs that are transferred to these buffers are called "swapped out". The buffer attached to the material handling system of a flexible manufacturing system is an example for this kind of buffers (see for instance, [Yao and Buzacott 1989]).

In more complex manufacturing systems several independent buffers can be found. This paper deals with three different types of relations between independent buffers and queueing systems: Staged buffers, shared buffers and local buffers that are used as independent buffers.

Single-staged Buffers

The most simple relation is presented in figure 3, a single-staged independent buffer. Jobs enter the manufacturing system and join the queue at A's local buffer with unlimited capacity. After completion of service at queueing system A the jobs are planned to be transferred to the buffer with limited capacity of queueing system B (straight line). If all places in B's local buffer are occupied, the job is moved to the independent buffer F (dotted line) instead. If there is space available again in B's local buffer, the jobs in F are transported to B on a first-come-first-served basis.

Shared Single-staged Buffers

Space may be saved if several queueing systems share one independent buffer (buffer F in figure 3). The swapped-out jobs stored in buffer F retain their planned destination; the workloads of the queueing systems B, D and E are not affected by the addition of an independent buffer.

The first-come-first-served order is related to the planned destination of the jobs in the buffer, i.e., a job arriving from queueing system A designated for B could leave F prior to a job already in buffer F that

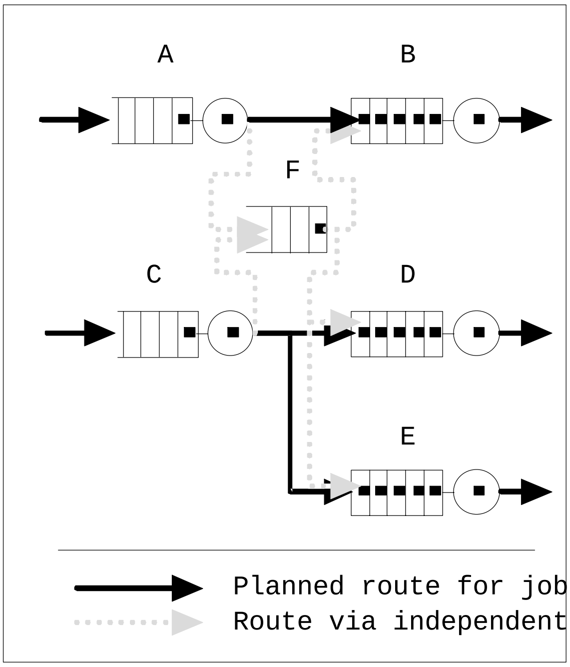

Figure 3: Shared, single-staged independent buffer

came from C bound for D, if space in the local buffer of B is the first to be available.

Multiple-staged Buffers

If independent buffers are used to model production buffers as part of a job shop, the capacity of the buffers could be exhausted as well. Therefore the necessity may arise to introduce a second or more stages of independent buffers (see figure 1). These buffers could be shared with other queueing systems as well, even on a different stage.Local Buffers As Independent Buffers

All previously mentioned configurations may cause

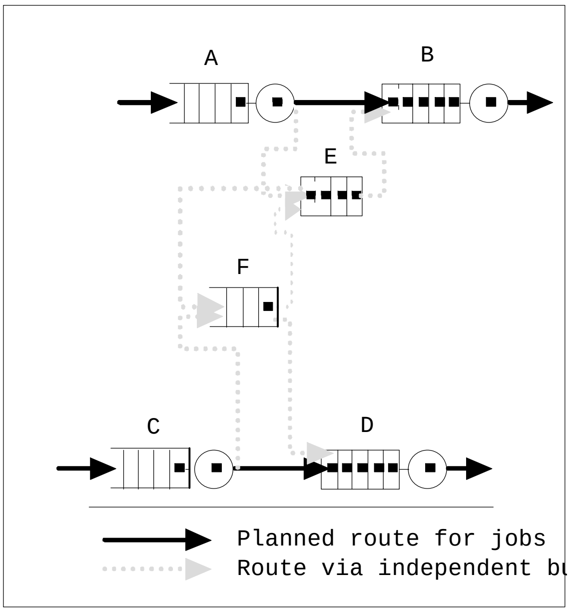

Figure 1: Shared second-stage (F) and first-stage (E) independent buffers

additional handling and transportation effort. This could be reduced by using the local buffer of one station as an independent buffer for other stations (figure 2). If a job, after completing its service at A cannot proceed to B it will be temporarily stored in the local buffer of A (path 1). If the capacity of this local buffer is used up by other jobs, the job will be conveyed to C, the independent buffer assigned to A (path 2).

Using independent buffers results in two effects: on one hand assuring flexibility and safe operation of the manufacturing system with less space used for buffers, on the other hand additional transports are generated for all jobs that are transferred to an independent buffer.

Therefore it is necessary to calculate performance measures for such a configuration. The probability of

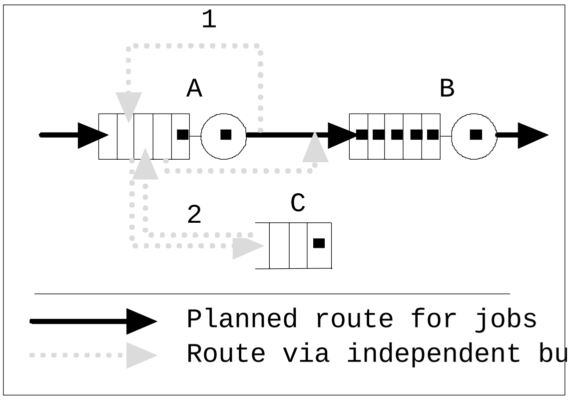

Figure 2: Local buffer and second-staged buffer

the system to be in an unplanned state and the amount of additional transportation effort are necessary data to evaluate the effectiveness of a planned configuration.

In the next section a technique is presented to calculate these data using queueing theory.

MODEL DESCRIPTION

In contrast to the work done on flexible manufacturing systems, this paper deals with open queueing networks.

The standard open queueing networks are extended to include independent buffers. The network consists of a number of queueing systems (1,...,M) and several independent buffers (1,...,P). Buffers and queueing systems are subsequently called nodes.

The size of the local buffer of node i is s_i , the total capacity of node i is $s'_i = s_i + m_i$ (including the number of parallel servers m_i).

The relations between queueing systems and buffers are defined by an M+P-dimensional overflow-adjacency matrix R.

We assign the value of '1' to element r_{ij} of matrix *R*, if blocked jobs, designated for *i* are transferred to the buffer of *j* instead. The matrix *R* is interpreted as a graph of the relations between buffers.

Basic Assumptions

To calculate exact performance measures, it is required that

- all jobs belong to one class,
- all inter arrival and service-times are negative-exponential distributed,
- service-order is first-come-first-served,
- the adjacency-relations defined in matrix *R* are circular-free,
- the routing probability-matrix *Q* is independent of the state of the queueing system,
- all queueing systems have either a local buffer with unlimited capacity or direct or indirect access to an independent buffer with unlimited capacity, ensuring that swapping out of a blocked job is always possible.

Although the subsequent sections concentrate on exact results, it should be mentioned that the application of the presented methods yields approximate results for open queueing networks with general distributed inter arrival and service-times as well. The results are not exact because the state-probabilities of the queueing systems are not independent. The quality of the approximate results also strongly relies on the algorithm that computes steady-state-probabilities for all queueing systems under the assumption of infinite buffer capacity for all local buffers. A promising approach for the approximate calculation of state-probabilities for queueing-networks of G|G|1-systems is presented in [Haßlinger and Rieger 92].

CALCULATING PERFORMANCE DATA

The computation of performance data is done in four steps, that are briefly explained in that paper. A full explanation of all steps is given in [Furmans 1992].

Building The Equivalent Queueing Network With Unlimited Buffers

The basic idea is to determine the flow in an equivalent network, which is identical to the network with independent buffers except that all local buffers have unlimited capacity. Therefore the local buffers are not necessary and can be omitted. If all service and inter-arrival times are negative-exponential distributed the equivalent network is of the Jackson-type. Therefore a product-form solution exists and the individual queueing-systems can be analyzed as if they were independent [Jackson 1957]. The steady-state-probabilities for the individual queueing systems can be computed with simple formulas (see for instance, [Bolch 1989] pages 42-43, [Hillier and Lieberman 1986], pages 543-551).

Determining The Critical Number Of Jobs

The number of possible states that the equivalent network is able to assume is unlimited, but not all of these states are of interest. We concentrate on those states, that lead to a partial filling of the independent buffers with limited capacity. A critical number of customers *K'* is determined up to where the steady-state-probabilities of the queueing systems must be computed. It is done in the following way:

1. $\mathcal{P}$ is the set of all queueing-systems and independent buffers whose capacity is limited. For every $j \in \mathcal{P}$ determine a set $\mathcal{Y}_j$ containing all buffers and queueing systems from where *j* is reachable.
2. For every $i \in \mathcal{Y}_j$ compute the sum S_{ij} of buffer spaces s'_h of all nodes *h* on the path from *i* to *j* (including *i* itself).
3. *K'* must be chosen so that $K' \geq \max\{S_{ij}\}$

Transformation Of State-Probabilities

A state *k* of the network is represented by an *M*-dimensional vector $k = (k_1, k_2, \ldots, k_i, \ldots, k_M)^t$. Every state *k* of the equivalent network is projected onto a state *k'* in the network with limited buffers (an *M+P*-dimensional vector). The projection depends on the type of occuring adjacency relations. In the simple case (single-staged independent buffers) it is defined as follows for every component of the vector k':

$$k'_i := \begin{cases} k_i & \text{if } k_i \leq s'_i \wedge 1 \leq i \leq M \\ s'_i & \text{if } k_i > s'_i \wedge 1 \leq i \leq M \\ \sum_{j=1}^{M} (k_i - s'_i) r_{ji} & \text{if } M+1 \leq i \leq M+P \end{cases}$$

The projection is injective but not bijective since two or more states *k* may be projected on the same state *k'*. In that case, the according probabilities *p(k)* must be added. Considering the number of relevant states, the projection of every single state would be a tedious task.

Therefore another method is proposed. The steady-state-probabilities $p(k_i = 0)$ up to $p(k_i = K')$ for the individual queueing systems have to be computed. Now several cases for the transformation of state-probabilities of every node have to be considered.

A) $i = 1 \ldots M$ (queueing systems)

A 1) $s'_i < \infty$ (limited local buffer)

$$p(k'_i) := \begin{cases} p(k_i) & \text{if } k < s_i \\ 1 - \sum_{i=1}^{s'_i - 1} p(k_i) & \text{if } k = s_i \end{cases} \qquad (1)$$

A 2) $s'_i = \infty$ (unlimited local buffer)

$$p(k'_i) = p(k_i) \qquad (2)$$

B) $i = M+1 \ldots M+P$ (independent buffers)

B 1) $s'_i = \infty$ (unlimited independent buffer, first stage)

The number of jobs in an independent buffer j is the sum of all jobs that could not be stored in the local buffers of the queueing systems i with r_{ij}=1.

The variable o_i denominates the number of jobs, that have to be swapped out to the related independent buffer. The vector

$$P_i^o = \left(p(o_i = 0), p(o_i = 1), \cdots, p(o_i = K')\right)^T$$

is the vector of the associated probabilities. They are derived from the equivalent network with infinite buffers as follows:

$$p(o_i) := \begin{cases} p(k_i = s'_i + o_i) & \text{if } o_i > 0 \\ \sum_{h=0}^{s'_i} p(k_i = h) & \text{if } o_i = 0 \end{cases} \qquad (3)$$

The probability of a sum of independent random variables could be computed with the convolution operator ⊗ (see [Feller 1968], pages 266 - 270). The vector of the steady-state-probabilities of the independent buffer is defined as

$$P_j = (p(k_j = 0), p(k_j = 1), \cdots, p(k_j = K'))^T$$

and computed by

$$P_j = \bigotimes_{j:r_{ij}=1} P_i^o \qquad (4)$$

B 2) $s'_i = \infty$ (limited and unlimited independent buffers, multiple-staged)

If the network contains multiple-staged buffers, the nodes have to be topologically sorted according to their relations defined by the adjacency matrix. In the sequence of their sorting the queueing systems are treated regarding cases A 1 and A 2. Buffers with limited capacities are first considered infinite and their virtual steady-state-probabilities are calculated with formula (4). If the capacity of node i is finite, in the next step the "overflow-probabilities" P_i^o are computed and serve as a computation-basis for those nodes that follow in the topological sequence.

C) Local buffers as independent buffers

As in the preceding case, a topological sorting is assumed and all nodes are treated in their sorting order. The number of jobs in the local buffer of a queueing system i is the sum of those jobs that wait for processing at i and those that are swapped out from other nodes that have an adjacency relation with node i. If the buffer-capacity is infinite the steady-state-probabilities could be calculated with

$$P'_j = P_j \otimes \bigotimes_{j:r_{ij}=1} P_i^o . \qquad (5)$$

If the capacity of the buffer is limited, the respective "overflow-probabilities" have to be computed using (3).

<u>Calculating The Additional Flow In The Network</u>

The mean number of jobs in the limited buffers of the network can be easily computed based on the modified state-probabilities of all limited buffers. By choosing a large K' the respective values for all unlimited buffers could be approximated with every desired accuracy.

The additional flow in the network is calculated as follows.

The flow of swapped-out jobs from node i is denominated as λ_i^o. It is necessary to swap out a job, if it arrives while all buffer space is occupied. It was shown in [Sevcik and Mitrani 1981] that the probability of an arriving job $p_{\mathrm{i}}^{\text{arrival}}(k_i)$ finding exactly k_i jobs in i is the probability of finding k_i jobs at a random instant in i. Thus the flow from a queueing system to the associated independent buffer (case A 1) is computed by

$$\lambda_i^o = \lambda_i \sum_{k_i = s'_i}^{K'} p_i^{\text{arrival}}(k_i) = \lambda_i \sum_{k_i = s'_i}^{K'} p_i(k_i) \qquad (6)$$
$$= \lambda_i \, p(k'_i = s'_i)$$

The flow from an independent buffer with limited capacity h to the adjacent buffer l has to be traced back to the probability that a job arriving at node i is swapped out to h and then to l. The probability for such an event is denominated $p^{i \to h \to l}$. It is equal to the probability, that a job arrives at node i in the corresponding equivalent network with infinite buffers under a certain condition. The condition is that the local buffer of node i contains already $k_i \geq s'_i$ jobs and the sum of all jobs, that are swapped out from other nodes to h exceeds $s'_h - (k_i - s'_i)$. The set $\mathcal{J}$ contains all nodes, that have adjacency relations to h.

$$\mathcal{J} = \{ j | r_{jh} = 1\}$$

The probability $p^{i \to h \to l}$ can then be written as

$$p^{i \to h \to l} = \sum_{k_i = s'_i}^{K'} \left[p^{\text{arrival}}(k_i) \cdot p\left(\sum_{j \in \mathcal{J} \setminus i} o_j \geq s'_h - (k_i - s'_i) \right) \right] \quad (7)$$

A more efficient way to replace (7) is presented in [Furmans 1992].

The flow from the independent buffer h to the adjacent buffer l is subsequently computed by

$$\lambda_h^o = \sum_{r_{ih}=1} \lambda_i \cdot p^{i \to h \to l} \quad . \qquad (8)$$

This method can be extended to multiple-staged buffers by determining all nodes i from where an independent buffer h is reachable.

EXAMPLE

The results obtained by the presented methods will be illustrated by a small example.

A manufacturing system consists of four queueing systems (i = 1 … 4) with local buffers and an independent buffer (i = 5). The arrival rate at the first queueing system is $\lambda_{01} = 8$. The service rate is $\mu_i = 10$. A buffer with infinite capacity (i = 6) is added for all unplanned states, i.e., when jobs could not be stored in the designated buffers of the nodes i = 1 … 5. As a measure of the probability that the manufacturing system is in an unplanned state, the probability that node 6 is not empty is taken .

The effects of the buffer allocation are discussed as follows. A constant number of 20 buffer spaces is assigned to the nodes in 6 combinations. The local buffers of the queueing systems are of equal size s_i (i = 1 … 4), the independent buffer i = 5 assumes s_5 = 20 - 4 s_i (see Table 1). The adjacency matrix R is:

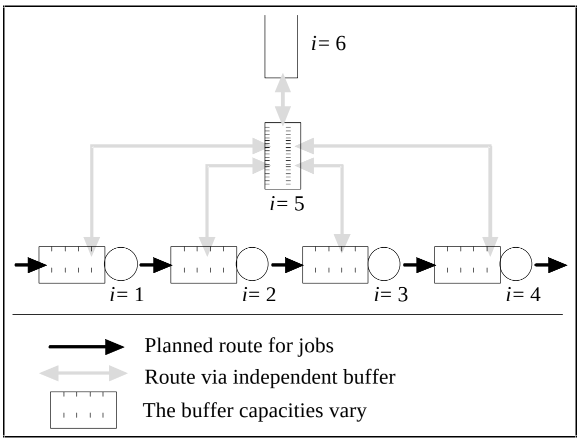

Figure 4: Example network with independent buffers

$$R = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The critical number K' of jobs depends on the configuration. Configuration 5 for example requires a critical number of jobs $K' \geq 16 + 2 = 18$.

Configuration No.	1	2	3	4	5	6
s_i if i = 1 … 4	5	4	3	2	1	0
s_5	0	4	8	12	16	20
Fraction of buffer space allocated at node 5	0	0.2	0.4	0.6	0.8	1.0

Table 1: Data of configurations with different buffer allocation

The trade-off between security (low probability of an unplanned state) and additional transportation can be seen in figure 5. The additional flow is identical for all queueing systems, node 1 is arbitrarily chosen. If the central buffer 5 has no capacity, the additional flow of 2.0972 is conveyed to the buffer with infinite capacity. The maximum additional flow is generated if no local buffer space is allocated at the queueing systems.

Another method to determine the buffer configuration is to establish an acceptable probability of the system being in an unplanned state. Then determine

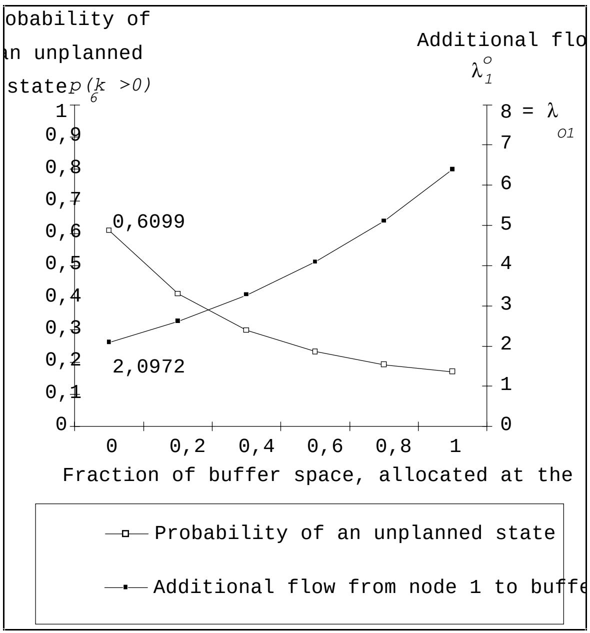

Figure 5: Probability of an unplanned state versus additional flow for the network in figure 4 with different configurations

configurations that do not exceed the given probability and compare the amount of additional transportation and the necessary investments in buffer space of the configurations. The basic data for this comparison could be supplied with the presented model.

SUMMARY

The introduction of independent buffers in queueing networks broadens the spectrum of problems that could be treated analytically. The requirements of computing time are minimal, the necessary calculations can still be performed on small computers in a very short time.

Further work should be done to integrate new algorithms that calculate steady-state-probabilities for G|G|s-networks.

REFERENCES

Akyildiz, I. F. 1989. "Product Form Approximations for Queueing Networks with Multiple Servers and Blocking." *IEEE Transactions on Computers*, Vol. 38, No. 1.

Bolch, G. 1989. *Leistungsbewertung von Rechensystemen.* B.G. Teubner, Stuttgart.

Feller, W. 1968. *An Introduction to Probability Theory and Its Applications 3rd Edition, Vol. I.* Wiley International, New York, N.Y.

Furmans, K. 1992. *Ein Beitrag zur theoretischen Behandlung von Materialflußpuffern in Bediensystemnetzwerken.* Dissertation, Fakultät für Maschinenbau, Universität Karlsruhe.

Haßlinger, G. and Rieger, E. S. 1991. "Modellierung offener Warteschlangennetzwerke durch Erneuerungsprozesse im diskreten Zeitbereich." in (Eds.:) Lehmann, Lehmann: *Messung, Modellierung und Bewertung von Rechensystemen, Informatik Fachberichte 286*. Springer-Verlag, Berlin.

Hillier, F. S. and Lieberman, G. J. 1986. *Introduction to Operations Research*. Holden Day, Oakland, CA.

Jackson, J. R. 1957. "Networks of Waiting Lines." *Operations Research* Vol. 5, No. 4.

Sevcik, K. C. and Mitrani, I. 1981. "The Distribution of Queueing Network States at Input and Output Instants." *Journal of the ACM* Vol. 28. No. 2.

Yao, D. D. and Buzacott, J. A. 1989. "Models of Flexible Manufacturing Systems with Limited Local Buffers." *International Journal of Productions Research* Vol. 24, No. 1.

ACKNOWLEDGEMENTS

This work was supported by a scholarship of the Stiftung Industrieforschung, Köln.