Push and Pull Systems in a Dynamic Environment

Michael Zazanis
Dept. of IEOR
University of Massachusetts
Amherst, MA 01003
email: zazanis@ecs.umass.edu

Abstract

We examine Push and Pull production control systems under a make-to-order policy with safety stock. We compute the distribution of the waiting time until a demand is satisfied, as well as flow time distributions and service levels for both systems. Work-in-process levels are determined as well. The analysis is carried out under Markovian assumptions and the explicit results on flow times depend on non-overtake conditions that are satisfied for single machine stations.

1 Introduction

This paper is a first step in an attempt to model the performance of various Push and Pull production strategies and the way they respond to external demands. These demands could represent either outside orders or signals from downstream cells in the same plant. The production line is modelled as a series of single machine stations with exponential processing times, external demands are Poisson and, in this paper, two production control schemes are examined: A Push scheme with safety stock S and a Pull scheme with limited Work-in-process.

In the Push scheme each time a demand arrives it immediately authorizes the release of a new job. The demand is satisfied immediately from the stock or it is backlogged. The Pull scheme examined is similar to CONWIP (see [20, 19, 21]). The main difference is that instead of the requirement that the WIP in the system remain constant we require that the total work-in-process including the Finished Goods Inventory (FGI) remain constant and equal to N. External demands, if not immediately satisfied, are again backlogged. In this scheme the arrival of a demand authorizes the release of a new job only when the work in process is less than N or equivalently the finished goods inventory is greater than zero. Otherwise the demand is backlogged. Proper operation of this scheme, which ofcourse depends crucially upon the right choice for N, results in the same benefits in terms of increased system control as other pull systems [4, 18, 16, 20].

A number of performance criteria are considered, namely average WIP, the probability that a demand will be backlogged, and the mean time to satisfy a demand. We also examine mean flow time as well as flow time variability and obtain explicit expressions in terms of the parameters of the system.

2 Systems under a Push Policy with Safety Stock

In this section, we examine in detail the push policy described above and obtain explicit expressions for performance criteria of interest including the total WIP in the system, the probability that a demand will be backlogged, and the average amount of time for a demand to be satisfied.

Once a demand arrives, one new job will be released to join the queue at machine N (e.g., once the first demand arrives, the $(S+1)$-th job will enter the system). While there are S units of initial inventory which serve as safety stock, once a demand arrives one anticipates that a finished job will be sent out so at the same time one new job is released to the system. Therefore the work in process (WIP), which includes the unfinished jobs in process and the finished jobs in the buffer is always greater than or equal to S.

An explicit expression for $E\tau$, the mean time to fill a demand in steady state, can be obtained when service times are exponential and the demand stream is Poisson. Equations (1) and (2) show clearly the dependence of $E\tau$ on S.

2.1 Exponential servers and Poisson demands

Consider the system described earlier with the exception that the initial inventory S consists of finished jobs, in the buffer instead of raw material in the queue of machine M. In the long run one can easily show (via a coupling argument) that this modification in the initial condition makes no difference as far as $E\tau$ is considered. We assume that the demand is Poisson with rate λ, and for all $i = 1, 2, \cdots, M$, the processing time of machine i is exponential with rate μ_i. Suppose that $\rho_i = \lambda/\mu_i < 1$. Then we can compute $E\tau$ through Little's law:

Under the push policy we consider once a demand arrives, a new job is released to system. Considering only the M stations (and disregarding finished job inventory) this system behaves as an open Jackson network. Let $X_t^{(i)}$ be the WIP at machine i at time t and define

$$X_t = \sum_{i=1}^{M} X_t^{(i)},$$

the total WIP in the system. In steady state

$$P(X_t^{(1)} = n_1, \cdots, X_t^{(M)} = n_M) = \prod_{i=1}^{M} (1-\rho_i){\rho_i}^{n_i}.$$

Therefore the negative part of $X_t - S$, $Y_t = (X_t - S)^-$, is the finished goods inventory while the positive part, $Z_t = (X_t - S)^+$ is the number of backlogged demands. Applying Little's law to the finished goods inventory buffer, we obtain

$$\lambda E\tau = E(X_t - S)^+, \qquad (1)$$

which indicates that $E\tau$ is decreasing in S. We proceed to obtain an expression for $E\tau$:

$$\begin{aligned} P(X_t = k) &= \sum_{\{\vec{n}\,:\, n_1+\cdots+n_M=k\}} \prod_{i=1}^{M} \rho_i^{n_i}(1-\rho_i) \\ &= G(M,k) \prod_{i=1}^{M}(1-\rho_i), \end{aligned}$$

with $G(M,k)$ given by:

$$G(M,k) = \sum_{\{\vec{n}\,:\, n_1+\cdots+n_M=k\}} \prod_{i=1}^{M} \rho_i^{n_i}.$$

Therefore,

$$\begin{aligned} E(X_t - S)^- &= \sum_{k=0}^{S} (S-k) P(X_t = k) \\ &= \sum_{k=0}^{S} (S-k) G(M,k) \prod_{i=1}^{M} (1-\rho_i) \\ &= \prod_{i=1}^{M} (1-\rho_i) \sum_{k=0}^{S} (S-k) G(M,k)\ . \end{aligned}$$

Since $(X_t - S)^+ = X_t - S + (X_t - S)^-$ and

$$EX_t = \sum_{i=1}^{M} \frac{\rho_i}{1-\rho_i},$$

we get

$$\begin{aligned} E\tau &= \frac{1}{\lambda} \left\{ \sum_{i=1}^{M} \frac{\rho_i}{1-\rho_i} - S \right. \\ &+ \left. \left(\prod_{i=1}^{M} (1-\rho_i) \right) \left(\sum_{k=0}^{S} (S-k) G(M,k) \right) \right\}. \qquad (2) \end{aligned}$$

Anticipating (8) which is established in the next section,

$$\begin{aligned} E\tau &= \frac{1}{\lambda} \left\{ \sum_{i=1}^{M} \frac{\rho_i}{1-\rho_i} - S \ + \ \prod_{i=1}^{M} (1-\rho_i) \times \right. \\ & \left. \sum_{m=1}^{M} \frac{\rho_m - (S+1)\rho_m^{S+1} + S\rho_m^{S+2}}{(1-\rho_m)^2 \prod_{l \neq m} (1-\rho_l/\rho_m)} \right\}. \qquad (3) \end{aligned}$$

2.2 Customer service criteria

In this framework we can address a number of related issues pertaining to the level of service:

The probability that a demand will be satisfied immediately is given by $P(S-X_t>0)=P(X_t<S)$. Since $X_t=\sum_{i=1}^{M} X_t^i$, its distribution is the convolution of N independent geometric random variables (because of the independence of the number of customers in each station). We examine two cases in detail:

a) All stations are identical. In this case the distribution of X_t is Pascal, i.e.

$$P(X_t=k)=\binom{M+k}{k}(1-\rho)^M\rho^k. \qquad (4)$$

b) If all stations have different utilizations then

$$P(X_t=k) = \sum_{i=1}^{M}\rho_i^k \frac{\prod_{l=1}^{M}(1-\rho_l)}{\prod_{l\neq i}(1-\rho_l/\rho_i)}. \qquad (5)$$

The derivation of (5) is interesting since it does not make use of convolutions directly. The partial fractions expansion through which it is obtained is shown here in the case where $\rho_i\neq\rho_j$ for $i\neq j$. The z-transform of X_t can then be written as

$$Ez^{X_t} = \prod_{i=1}^{M}\frac{1-\rho_i}{1-z\rho_i} = \sum_{i=1}^{M}\frac{A_i}{1-z\rho_i}$$

or

$$\prod_{i=1}^{M}(1-\rho_i) = \sum_{i=1}^{M}A_i\prod_{l\neq i}(1-z\rho_l)\ .$$

Letting $z=\rho_i^{-1}$ gives

$$A_i=\frac{\prod_{l=1}^{M}(1-\rho_l)}{\prod_{l\neq i}(1-\rho_l/\rho_i)}, \qquad i=1,2,\ldots,M.$$

Hence,

$$E[z^{X_t}] = \sum_{i=1}^{M}\frac{1}{1-z\rho_i}\frac{\prod_{l=1}^{M}(1-\rho_l)}{\prod_{l\neq i}(1-\rho_l/\rho_i)}, \qquad (6)$$

from which we obtain (5).

An alternative for computing the distribution of X_t uses the normalization constants for CQNs for which a number of efficient computational algorithms exist.

$$\begin{aligned} P(X_t=k) &= \sum_{\vec{n}:n_1+\cdots+n_M=k}\prod_{i=1}^{M}(1-\rho_i)\rho_i^{n_i} \\ &= G(M,k)\prod_{i=1}^{M}(1-\rho_i)\ . \end{aligned} \qquad (7)$$

A comparison between (5) and (7) suggests the following expression for the normalization constant in a CQN

$$G(M,k) = \sum_{i=1}^{M}\frac{\rho_i^k}{\prod_{l\neq i}(1-\rho_l/\rho_i)}. \qquad (8)$$

2.3 Probability that a demand will be satisfied immediately

The probability that a demand will be satisfied immediately is given by

$$\begin{aligned} P(X_t<S) &= \sum_{k=0}^{S-1}P(X_t=k) \\ &= \prod_{i=1}^{M}(1-\rho_i)\ \sum_{k=0}^{S-1}G(M,k) \\ &= \sum_{i=1}^{M}\sum_{k=0}^{S-1}\rho_i^k\ \frac{\prod_{l=1}^{M}(1-\rho_l)}{\prod_{l\neq i}(1-\rho_l/\rho_i)} \\ &= \sum_{i=1}^{M}(1-\rho_i^S)\ \prod_{l\neq i}\frac{1-\rho_l}{1-\rho_l/\rho_i}. \end{aligned} \qquad (9)$$

2.4 Total WIP in an open system

From the above analysis we can easily obtain an expression for the distribution of the total WIP in an open system:

$$\begin{aligned} P(X_t\leq n) &= \prod_{i=1}^{M}(1-\rho_i)\sum_{k=0}^{n}G(M,k) \\ &= G(M+1,n)\prod_{i=1}^{M}(1-\rho_i), \end{aligned}$$

where $G(M+1,n)$ is the normalization constant of a CQN with n customers and $M+1$ stations with mean service times $\rho_1, \ldots, \rho_M, 1$, or equivalently $\mu_1^{-1}, \ldots, \mu_M^{-1}, \lambda^{-1}$. In view of (8)

$$G(M,n) = \sum_{m=1}^{M} \frac{\rho_m^n}{(1-1/\rho_m)\prod_{l\neq m}(1-\rho_l/\rho_m)} + \frac{1}{\prod_{m=1}^{M}(1-\rho_l)}$$

and hence

$$P(X_t \leq n) = 1 - \sum_{m=1}^{M} \frac{\rho_m^n}{(1/\rho_m - 1)\prod_{l\neq m}(1-\rho_l/\rho_m)} .$$

3 Flow time distributions for Push and Pull systems

The Pull strategy with WIP limited above by N can be modeled in the markovian case as an open queueing network with a global buffer of size N. Demands arrive according to a Poisson process and are admitted to the system only if the total number of customers present, X_t, is less than N. Otherwise they wait outside the global buffer. If at time t $X_t \leq N$ then no demands are backlogged and $N - X_t$ represents finished goods inventory. If, on the other hand, $X_t > N$ then $X_t - N$ represents the number of backlogged demands. If the probability that a demand is backlogged is small the operation of this system can be adequately approximated by a closed queueing network. (For details see [24].)

For a closed queueing network, from Boxma, Kelly, and Könheim (1984), and Daduna (1982), it follows that, if T_i is the flow time through the i'th station, the joint Laplace transform for the flow times of a tagged customer through the stations satisfies the following product form relationship

$$E[e^{-s_1T_1-\cdots-s_MT_M}] = \sum_{\vec{\jmath}\in\mathcal{S}(M,N-1)} p(j_1,\ldots,j_M)\prod_{i=1}^{M}\left(\frac{\mu_i}{\mu_i+s_i}\right)^{j_i+1}, \qquad (10)$$

where $\mathcal{S}(M,N) = \{\vec{\jmath} : j_1 + \cdots + j_M = N\}$. (The above holds provided that a non-overtake condition holds, which of course is the case for cyclic single server networks.) Setting $\rho_i = 1/\mu_i$, the rhs of the above equation can be written as

$$\sum_{\vec{\jmath}\in\mathcal{S}(M,N-1)} \frac{\rho_1^{j_1}\cdots\rho_M^{j_M}}{G(M,N-1)}\prod_{i=1}^{M}\left(\frac{\mu_i}{\mu_i+s_i}\right)^{j_i+1} = \prod_{i=1}^{M}\left(\frac{\mu_i}{\mu_i+s_i}\right)\sum_{\vec{\jmath}\in\mathcal{S}(M,N-1)}\frac{\prod_{i=1}^{M}\left(\frac{1}{\mu_i+s_i}\right)^{j_i}}{G(M,N-1)} . \qquad (11)$$

The first factor in the above expression corresponds to the joint Laplace transform of the processing times for a job. The second factor, corresponding to the joint Laplace transform of waiting times can be written, taking into account (8), as

$$\left(\sum_{m=1}^{M}\frac{1}{\mu_m^{N-1}}\frac{1}{\prod_{l\neq m}(1-\mu_m/\mu_l)}\right)^{-1} \times \sum_{m=1}^{M}\frac{1}{(\mu_m+s_m)^{N-1}}\prod_{l\neq m}\left(1-\frac{\mu_m+s_m}{\mu_l+s_l}\right)^{-1} \qquad (12)$$

To obtain the Laplace transform of the flow time set $s = s_1 = \cdots = s_M$:

$$\left(\prod_{m=1}^{M}\frac{\mu_m}{\mu_m+s}\right)\frac{\sum_{m=1}^{M}\frac{1}{(\mu_m+s)^{N-1}}\prod_{l\neq m}\frac{\mu_l+s}{\mu_l-\mu_m}}{\sum_{m=1}^{M}\frac{1}{\mu_m^{N-1}}\prod_{l\neq m}\frac{\mu_l}{\mu_l-\mu_m}} . \qquad (13)$$

The throughput of the closed system is then given by

$$\lambda = \frac{G(M,N-1)}{G(M,N)} = \frac{\sum_{m=1}^{M}\frac{1}{\mu_m^{N-1}}\prod_{l\neq m}\frac{\mu_l}{\mu_l-\mu_m}}{\sum_{m=1}^{M}\frac{1}{\mu_m^{N}}\prod_{l\neq m}\frac{\mu_l}{\mu_l-\mu_m}} = \frac{\sum_{m=1}^{M}\left(\mu_m^N\prod_{l\neq m}(\mu_l-\mu_m)\right)^{-1}}{\sum_{m=1}^{M}\left(\mu_m^{N+1}\prod_{l\neq m}(\mu_l-\mu_m)\right)^{-1}} . \qquad (14)$$

For the corresponding open system the Laplace transform of the flow time is ofcourse

$$E[e^{-sT_O}] \;=\; \prod_{m=1}^{M} \frac{\mu_m - \lambda}{\mu_m - \lambda + s}\,, \tag{15}$$

where λ is given by (14).

To simplify the comparison we will confine ourselves here to the balanced case. The Push system is equivalent to an open Jackson network with M exponential stations with rate μ. Let T_O, T_C, denote the flow times for parts in the open and closed systems respectively. The closed system has N parts and a corresponding throughput

$$\lambda \;=\; \mu\frac{N}{M+N-1}. \tag{16}$$

The Laplace transform of the flow time in the open system is then given by

$$E[e^{-sT_O}] \;=\; \left(\frac{\mu-\lambda}{\mu-\lambda+s}\right)^M . \tag{17}$$

In particular the mean and variance of T_0 are given by

$$E[T_O] \;=\; \frac{M}{\mu-\lambda}, \tag{18}$$

$$\mathrm{Var}[T_O] \;=\; \frac{M}{(\mu-\lambda)^2}.$$

For the balanced case the Laplace transform of the flow time is given by,

$$E[e^{-sT_C}] \;=\; \left(\frac{\mu}{\mu+s}\right)^{M+N-1} . \tag{19}$$

In particular the mean and variance of the flow time for the closed system is

$$E[T_C] \;=\; \frac{1}{\mu}(M+N-1), \tag{20}$$

$$\mathrm{Var}[T_C] \;=\; \frac{1}{\mu^2}(M+N-1).$$

Denoting by c_O, c_C the corresponding coefficients of variation we obtain the following expression for their ratio

$$\frac{c_C}{c_O} \;=\; \sqrt{\frac{M}{M+N-1}}\,. \tag{21}$$

(21) suggests that the variability in total processing time for pull systems is always less than that of open systems, the effect becoming less pronounced under heavy utilization.

References

[1] S. Asmussen (1987). *Applied probability and queues.* Wiley.

[2] Boxma, O.J., F.P. Kelly, and A.G. Konheim (1984). "The Product Form for Sojourn Time Distributions in Cyclic Exponential Queues," *J. ACM,* **31,** 1, 128-133.

[3] P. Brémaud (1981). *Point Processes and Queues: Martingale Dynamics.* Springer–Verlag, New York.

[4] Buzacott, J.A. and J.G. Shanthikumar (1993). *Stochastic Models of Manufacturing Systems*, Prentice Hall, 490-515.

[5] Buzacott, J.A. and D.D. Yao (1986). "On Queueing Network Models of Flexible Manufacturing Systems", *Queueing Systems,* **1,** 5-27.

[6] Buzacott, J.A. and D.D. Yao (1986). "Flexible Manufacturing Systems: A Review of Analytical Models", *Mgmt. Sci.,* **32,** 7, 890-905.

[7] Daduna, H., (1982). "Passage times for overtake-free paths in Gordon-Newell networks," *Adv. Appl. Probab.* **14,** 672-686.

[8] Delersnyder, J.-L., T.J. Dodgson, H. Muller, and P.J. O'Grady (1989). "Kanban controlled pull systems: an analytic approach", *Management Science,* **35,** 1079-1091.

[9] Graves, S.C., H.C. Meal, S. Dasu, and Y.F. Choong (1986). "Short-term production scheduling in a dynamic environment", In S. Axsater, Ch. Schneeweiss, and E. Silver, eds. *Multi-stage Production Planning and Inventory Control,* pp. 9-43, Springer Verlag.

[10] Groenevelt, H., and U. S. Karmarkar. (1988). "Dynamic Kanban Systems Case Study", *Prod. Inv. Mgmt.* **29**, 46-51.

[11] Hall, R. W. (1983). *Zero Inventories*, Dow Jones-Irwin, Homewood, Ill.

[12] Hopp, W.J., and M.L. Spearman (1991). "Throughput of a Constant Work in Process Manufacturing Line Subject to Failures", *Int. J. Prod. Res.,* **17,** 6, 557-580.

[13] Karmarkar, U. S. (1986). "Push, Pull, and Hybrid Control Schemes", Working Paper Series No. QM8614, Graduate School of Management, University of Rochester, Rochester, New York.

[14] Mitra, D., and I. Mitrani (1990). "Analysis of a Kanban discipline for cell coordination in production lines, I", *Management Science,* **36,** 1548-1566.

[15] Mitra, D., and I. Mitrani (1991). "Analysis of a Kanban Discipline for Cell Coordination in Production Lines, II: Stochastic Demands", *Operations research*, Vol. 39, No. 5, 807-823.

[16] Monden, Y. (1983). *Toyota Production System: Practical Approach to Management*, Industrial Engineering and Management Science Press, Norcross, GA.

[17] Seidman, A. (1988). "Regenerative pull (Kanban) production control policies", *European J. Operational Research,* **35,** 401-413.

[18] Schonberger, R.J. (1986). *World Class Manufacturing: The Lessons of Simplicity Applied*, The Free Press, New York.

[19] Spearman, M. (1992). "Customer Services in Pull Production Systems", *Operations research*, Vol. 40, No. 5, 948-958.

[20] Spearman, M., D. Woodruff, and W. Hopp (1990). "CONWIP: a Pull Alternative to Kanban ", *Int. Production Researches* , Vol. 28, No. 5, 879-894.

[21] Spearman, M., and M. Zazanis (1992). "Push and Pull Production Systems: Issues and Comparisons", *Operations research*, Vol. 40, No. 3, 521-531.

[22] Stoyan D. (1983). *Comparison Methods for Queues and Other Stochastic Models*, John Wiley.

[23] Yao, D.D. and J.A. Buzacott (1987). "Modeling a Class of FMSs with Reversible Routing", *Opns. Res.,* **35,** 1, 87-92.

[24] Zazanis, M.A. (1994). "Pull Systems with Limited WIP", Technical Report, IEOR Dept. Univ. of Massachusetts, Amherst, MA 01003.