SIMULATION OF SEQUENTIAL PRODUCTION SYSTEMS WITH IN-PROCESS INVENTORY

David R. Anderson
Brian D. Sellers
Mohammed M. Shamma
Department of Quantitative Analysis
University of Cincinnati

Abstract

This paper presents simulation results from a general sequential production system. The results are used to establish the effect of service time variability and to estimate minimum cost-in process inventory capacities.

This paper deals with the problem of finding optimal in-process inventory levels for a general production system. Specifically the system can be described as a production line with N separate stages (work stations) where an in-process inventory buffer with a fixed capacity is provided between these stages. All work units are processed through the stages in a fixed sequence. We assume an infinite supply of input at the first stage and no blocking of output from the last stage. Typically, such systems are used for high volume production, and operating costs saved by choosing an optimal size buffer will be desirable. Industrial engineers and system analysts are frequently confronted with the design and evaluation of such production line systems.

The Model

In our research we simulated 2-, 3-, and 4-stage production systems with 0, 2, 4, 6, and 8 buffer capacities and with normal service times. Coefficients of variation for the normal distributions ranged from .01 to .30. We used the normal distribution for several reasons: numerous variables seem to follow a pattern of variation that is similar to the normal distribution; and the normal distribution can be an excellent approximation to several other distributions. The chief reason for using the normal distribution, however, was its practical significance. Lind [6] and Nadler [7] found that manufacturing processes, whether machine or operator controlled, exhibit an inherent variation about their mean production rates ranging from approximately normal distributions to positively skewed distributions of the Pearson Type II curve. GPSS (General Purpose Systems Simulation) was chosen as the language for our studies because of its adaptability to manufacturing processes and especially to production lines. Not only is the structure of GPSS suited quite well to such queueing problems incapable of mathematical formulation, but it also permits the direct and complete observation of the dynamic behavior of the processes. In general simulation provided a closer fit to reality and an insight into system characteristics unobtainable through strictly analytical formulations.

In this research we investigated the behavior of systems in which the individual stages have service times with different coefficients of variation. For each system we determine an overall measure of system variation. We define the system coefficient of variation by the following formula:

$$\overline{CV} = \sqrt{\frac{(CV1)^2+(CV2)^2+\ldots+(CVN)^2}{N}} \qquad [1]$$

where $\overline{CV}$ = the system coefficient of variation

CV1 = coefficient of variation of Stage 1

CV2 = coefficient of variation of Stage 2

.
.
.

CVN = coefficient of variation of Stage N

N = number of stages in the system

The service time at each stage in the simulated production line was randomly assigned a normal distribution with a coefficient of variation of .01, .02, .03, .04, ..., .27, .28, .29, or .30. Three sets of variation patterns for the coefficient of variation of each system were calculated, and one of them was discarded on the basis of duplication or close similarity to another pattern. The example below illustrates two such patterns for a 3-stage model with a buffer capacity of 4 and a coefficient of variation for the system of 0.20.

	CV Stage 1	CV Stage 2	CV Stage 3	$\overline{CV}$ for System
System 1	.20	.05	.28	.2007
System 2	.24	.24	.08	.2013

The Simulation Process

A basic unit called a transaction travels through the simulation model with processing stations and storage areas, and statistics are gathered on its movement with respect to congestion and occupancy, total time, and delay. We used a 3000 transaction starting run for assurance that the effect of transient-state build up would not affect the steady-state statistics. We then used 15,000 transaction steady-state runs.

Two types of delay were identified in the studies: lack of work and blocking delay. Lack of work occurs at a stage when the buffer immediately preceding it is empty and the stage is available for processing; blocking occurs when the buffer immediately proceeding it is full and the stage has completed processing on its current contents. During the simulation runs we gathered statistics on blocking and lack of work delay, buffer content, facility utilization, and production times. After gathering statistics for all the systems, we used regression analysis for the formulation of functions defining various aspects of the systems according to the parameters of the number of stages in the system, the system coefficient of variation, and the buffer capacity. Our goal was to use the simulation results to find the following functional relationships:

$$\text{INVENTORY} = f\ (N, B, \overline{CV})$$
$$\text{DELAY} = f\ (N, B, \overline{CV}) \qquad [2]$$
$$\text{UTILIZATION} = f\ (N, B, \overline{CV})$$

where N = number of stages in system

B = buffer capacity

$\overline{CV}$ = system coefficient of variation

Using these equations, we developed the general cost equation for any production line system where

$$\text{TOTAL COST} = f\ (N, B, \overline{CV}) \qquad [3]$$

This equation along with the others provided the framework for determining optimal buffer capacities, optimal operating costs, and optimal utilization of production line facilities.

General System Behavior

In viewing the data, it appears that buffer capacity and the system coefficient of variation have the most significant effect on average delay. As buffer capacity increases, delay rapidly approaches zero. There is a very large drop in delay with an increase in buffer capacity from zero to two units. This is in agreement with Hatcher's analytical results (1969) that only a small amount of storage capacity is required to reach near optimum production rates. For example, a 3-stage system with a system coefficient of variation of .20 displayed 13.3% delay for no buffer capacity, 2.11% for a buffer capacity of two, and .73% for a buffer capacity of eight. Similarly as the system coefficient of variation increases, the system displays greater delay. A 3-stage system with a buffer capacity of 4 displayed .12% average delay for a system coefficient of variation of .05 and 1.9% delay for a coefficient of .25. Increasing the number of stages in the system also increased delay but not as significantly as the other two variables. For example, for two systems with buffer capacities of 4 and a system coefficient of variation of .20, a 2-stage system displayed 1.0% delay and a 4-stage system displayed 1.7% delay.

Average system content appears to be affected by both buffer capacity and number of stages in the system. Naturally system content increases as buffer capacity is increased up to a point where blocking delay approaches zero. When buffers are large and blocking delays near zero, further increases in buffer capacity remain unused. Obviously an increase in the number of stages increases average system content by more than one unit since not only is another stage added but also another buffer. In viewing the data, however, the system coefficient of variation appears to have no significant effect on average content. For the case of no buffer, systems with higher system variation display slightly lower average content, but this does not hold as soon as buffer capacity is added.

In analyzing internal system behavior, the results for lack of work and blocking delay were in agreement with Anderson's earlier results where the coefficient of variation was held constant at each stage throughout the system. Basically, blocking delay is highest for the first stage and decreases at each stage up to the last where it is zero. Lack of work is zero for the first stage and increases up to the highest amount at the last stage. This general rule held for all systems with few exceptions. Average buffer content was, with only one exception, highest in the first buffer and lowest in the last buffer. The percentage of time the buffers were empty, a good indication of relative lack of work delay at proceedings stages, was in every case lowest for the first buffer and highest for the last buffer. In some cases the gradations from lowest lack of work and blocking delay were gradual and in other cases quite steep. However, this could not be traced to particular variation patterns. Also, evidence seems to indicate that contents of the individual buffers is independent of the system coefficient of variation.

Estimating System Delays and Contents

First we used stepwise regression to develop a formula for content. Using the following terminology

C = system coefficient of variation

N = number of stages

B = buffer capacity

we regressed the variables C, N, B, C^2, N^2, B^2, CN, CB, and NB. We obtained a correlation coefficient of .938. With the exception of B^2, all of the variables containing C were the last to enter the regression equation and did not increase the correlation coefficient significantly. Using the variables N, B, and NB, we obtained the following equation:

$$\text{Content} = .08-.27B+.93N+.41NB \qquad [4]$$

with an R^2 of .928. Using nonlinear regression did not result in any significant improvement.

After this we developed a delay equation from stepwise regression of the variables B, N, C, B^2, C^2, N^2, BN, BC, and NC. But we obtained an R^2 of only .881 and the fit was very poor. Next we resorted to nonlinear regression. As a theoretical basis for the delay equation, we considered Hunt's analytical derivation of delay in the 2-stage exponential service time system. He obtained analytically the following equation:

$$\text{Delay} = \frac{1}{B+3} \qquad [5]$$

Using this starting point, Anderson had previously developed the delay equation

$$\text{Delay} = \frac{a_1}{B+a_2} \qquad [6]$$

where $a_1 = f_1(B,N,C)$

$a_2 = f_2(B,N,C)$

After testing several formulations, we came up with the following function which is relatively simple

$$\text{Delay} = \frac{1}{B+.453}(-.134 +.131N^{.028}+.111C^{.870}+.052CN) \qquad [7]$$

with an R^2 of .985. Further complexity did not improve the fit significantly.

The Cost Model and Optimal Inventory Level

In order to evaluate systems on a cost basis and derive optimal buffer capacity for various cost structures, we use the same cost model as presented in Anderson's earlier paper. The general model for a sequential queue is shown below.

Let N = number of stages

B = buffer capacities

D = average delay/unit time

I = average contents of the system

S = total number of storage spaces

K = cost of delay/unit time

L_1 = inventory cost/unit/unit time

L_2 = storage space cost/unit/unit time

T = total cost/unit time

the

$$T = DK + IL_1 + SL_2 \quad [8]$$

But since S is known to be:

$$S = (N-1)B \quad [9]$$

we have

$$T = DK + IL_1 + (N-1)BL_2 \quad [10]$$

Using the equations we obtained from regression analysis for average delay and content, we can formulate the total cost equation where a_1 and a_2 are as defined before

$$T = \frac{a_1 K}{B+a_2} + (.08-.27B+.93N + .41NB)L_1+(N-1)BL_2 \quad [11]$$

For optimal buffer capacity with respect to cost we have

$$\frac{\partial T}{\partial B} = \frac{-a_1 K}{(B+a_2)^2} -.27L_1+.41NL_1 +(N-1)L_2 = 0 \quad [12]$$

$$B^* = \sqrt{\frac{a_1 K}{-.27L_1+.41NL_1+(N-1)L_2}} - a_2 \quad [13]$$

where $a_1 = -.134+.131N^{.028}+.111C^{.870} +.052CN$

$a_2 = .453$

The second derivative shows this to be a minimum.

In order to compare the cost properties of the two models, let L equal the effective space-holding cost at the optimal buffer size of B*. From the equation we must have

$$-.27L+.4NL+(N-1)L=.27L_1+.41NL_1 +(N-1)L_2 \quad [14]$$

giving

$$L = \frac{-.27L_1+.41NL_1+(N-1)L_2}{1.41N-1.27} \quad [15]$$

Rewriting the equation, we get

$$B^* = \sqrt{\frac{a_1 K}{(1.41N-1.27)L}} - a_2 \quad [16]$$

Let ϕ = the ratio of delay cost K to the effective holding cost

$$\phi = K/L \quad [17]$$

Then

$$B^* = \sqrt{\frac{a_1 \phi}{(1.41N-1.27)}} - a_2 \quad [18]$$

and

$$T^* = DK + LI + (N-1)B^* L$$
$$= DK + IK/\phi + (N-1)B^* K/\phi$$
$$= (D + I/\phi + (N-1)B^*/\phi) K \quad [19]$$

Letting $K=\phi$ we generate the following table for optimal buffer capacities and optimal costs for the given ratios of ϕ.

		Buffer Sizes					Total Cost				
		$\overline{CV}$					$\overline{CV}$				
ϕ	N	.05	.10	.15	.20	.25	.05	.10	.15	.20	.25
1,000	2	2	4	4	5	6	10.32	13.74	16.33	18.44	20.35
	3	2	3	3	4	5	15.75	20.92	25.01	28.11	31.12
	4	2	2	3	4	4	21.01	28.01	32.75	37.21	40.91
5,000	2	6	9	10	12	13	21.28	29.08	34.89	39.71	43.94
	3	5	7	8	10	11	33.34	44.83	53.64	61.01	67.44
	4	4	6	7	9	10	44.61	59.43	71.08	80.76	89.31
10,000	2	9	12	15	17	19	29.60	40.57	48.57	55.64	61.62
	3	7	10	12	14	15	46.51	62.76	75.17	85.59	94.72
	4	6	9	11	12	14	62.22	83.28	99.62	113.38	125.46
50,000	2	20	28	34	39	43	64.62	89.17	107.57	122.88	136.25
	3	17	23	27	31	35	102.14	138.46	166.20	189.47	209.88
	4	15	20	25	28	31	136.58	183.99	220.49	251.19	278.23

Simulation Summary

In total we simulated the production process of 18,000 units a total of 150 times. These simulations required over seven hours of computer time and cost roughly $4800.

Conclusion

A general class of simulation models of production lines have been studied to formulate theories on system behavior. All of the systems had normally distributed processing times as an approximation to conditions often encountered in the manufacturing environment. Results show that the behavior of such systems is a function of buffer capacity, the system coefficient of variation, and the number of stages in the system, and that the pattern of variation among the individual stages has no significant effect on the behavior of the system. From the results, it appears possible to obtain an excellent approximation to optimum buffer size for any system meeting the general assumptions of the model. Also we can gain insight into the cost of constraints on buffer sizes less than optimum. Simulation has provided significant results to the general in-process inventory problem where a theoretical approach would have been virtually impossible.

References

1. Anderson, David R., "Simulation Studies on Sequential Queues", Proceedings American Institute of Decision Sciences Southeastern Conference, July, 1971.
2. Barten, Kenneth, "A Queueing Simulator for Determining Optimum Inventory Levels in a Sequential Process", Journal of Industrial Engineering, Vol. VIII, No. 4, July-August, 1962.
3. Buchan, Joseph and Koenigsberg, Ernest, Scientific Inventory Management, Chapter 22, Prentice-Hall, 1963.
4. Hatcher, Jerome M., "The Effect of Internal Storage on the Production Rate of a Series of Stages Having Exponential Service Times", A.I.I.E. Transactions, Vol. I, June, 1969.
5. Hunt, Gordon C., "Sequential Arrays of Waiting Series", Operations Research, December, 1956.
6. Lind, Warren E., "A Statistical Analysis of Work-Time Distributions", MSIE Thesis, Georgia Institute of Technology, June, 1953.
7. Nadler, Gerald, Motion and Time Study, Chapter 20, New York, McGraw-Hill Book Company, 1955.
8. Young, H. H., "Optimization Models for Production Lines", Journal of Industrial Engineering, January, 1967.