Proceedings of the 2012 Winter Simulation Conference
C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A.M. Uhrmacher, eds

OPTIMAL BATCH PROCESS ADMISSION CONTROL IN TANDEM QUEUEING SYSTEMS WITH QUEUE TIME CONSTRAINT CONSIDERATIONS

Cheng-Hung Wu
Yu-Ching Cheng
Ping-Ju Tang
Jiun-Yu Yu

Institute of Industrial Engineering, National Taiwan University
No. 1, Sec. 4, Roosevelt Road
Taipei, 10617 TAIWAN

ABSTRACT

In this paper, a dynamic control method for two-stage queueing systems with process queue time (PQT) constraints is presented. This queueing system consists of an upstream batch process machine and a downstream single process machine. The waiting time of each job in the downstream queue is constrained by an upper limit. Violation of this upper limit causes scrap of the job. A batch machine poses a problem for the two-stage system under PQT constraints. After completion of batch process, a large quantity of work-in-process (WIP) moves into the downstream queue with PQT constraints. This increases the variance of downstream queue length and the probability of scrap.

In this research, we incorporate dynamic programming algorithm in batch process admission control (BPAC) model. The performance of BPAC model is verified by simulation. Simulation results demonstrate that the proposed BPAC model outperforms other methods in every key system performance indices.

1 INTRODUCTION

This study addresses dynamic production control problems in a two-stage tandem queueing system under process queue time (PQT) constraints. The queueing system has an upstream batch process machine and a downstream single process machine. The PQT constraint is an upper bound of waiting time between two sequential processes. Process engineering sets the upper limit in waiting time to ensure production quality, and violation of PQT constraints causes scrap or rework. For example, when a work-in-process (WIP) finishes at certain process step, the subsequent operation for the job should finish within the upper bound of time. Otherwise, the WIP deteriorates and becomes scraped. Scraps reduce efficiency of capacity utilization and cause productivity loss. PQT constraint problem is a critical issue for many manufacturing systems. In semiconductor manufacturing, PQT constraints are imposed to prevent wafer surface from defects or particles after furnace tubes (Su 2003).

The two-stage queueing system model is shown in Figure 1. There are *arrival queue* and *qualified queue* ahead of the two servers, respectively. Any external job initially waits in *arrival queue* for being processed by the upstream batch process server. After finishing the upstream batch process, jobs are moved to *qualified queue* for service by the downstream single process server. Once completing the operation at the downstream server, the job leaves the system. PQT constraints are applied to the jobs staying in *qualified queue* and the downstream server.

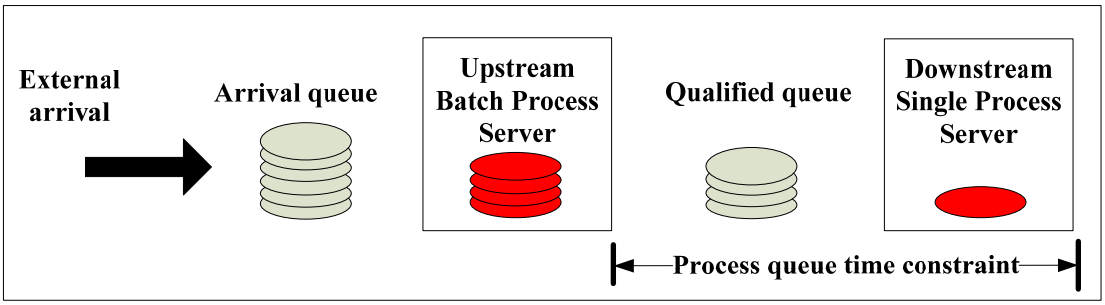


Figure 1: Batch process and process queue time constraints.

Due to the operation characteristic of the batch machine, numerous jobs would enter *qualified queue* simultaneously. But the downstream server processes one job at a time. As a result, many WIP would rest in *qualified queue*, which leads to the higher probability of violation against PQT constraints. On the other hand, if we reduce the quantity of jobs processed by the batch machine at a time, the probability of PQT constraint violation will decrease. However, it will lower the utilization of the upstream batch process machine and more inventory holding costs are incurred. Therefore, when controlling the flow rate of the upstream batch process machine, tradeoff between PQT constraints and inventory costs should be carefully considered.

The objective of this paper is to develop an efficient production control method in a two-stage queueing system under PQT constraints. The batch process admission control (BPAC) model is formulated to reduce total production costs involving scrap costs and inventory costs. Scrap costs come from violation of PQT constraints. Inventory holding costs are caused by WIP in *arrival* and *qualified queue*.

2 MODEL DESCRIPTION AND FORMULATION

The definition of the variables used in BPAC model is given below.

- λ : Arrival rate of external jobs (jobs/hour)
- μ_1 : Service rate of the upstream batch process server (batch/hour)
- μ_2 : Service rate of the downstream single process server (jobs/hour)
- Q_{q_e} : Capacity of *qualified queue*
- B : Capacity of the upstream batch process server
- a : Production control for the upstream batch process server
- LT_{q_e} : Upper limit of waiting time for a job
- ω : Scrap rate, $\omega = \dfrac{1}{LT_{q_e}}$

In practical production systems, machine failures could make control policy ineffective. Therefore, BPAC model also takes the real-time machine status into consideration. Failure rates of the upstream batch server and downstream server, f_1 and f_2, are defined by the mean time between failures. Similarly, repair rates of the two servers, r_1 and r_2, are defined by mean time between repairs.

Let $S \equiv (y_1, y_2, \sigma_1, \sigma_2)$ stand for the state space.

- y_1 : Number of jobs(including WIP) in *arrival queue*
- y_2 : Number of jobs(including WIP) in *qualified queue*
- σ_1 : Upstream batch process server status, $\sigma_1 = \begin{cases} 1, & available \\ 0, & failed \end{cases}$

- σ_2 : Downstream single process server status, $\sigma_2 = \begin{cases} 1, & available \\ 0, & failed \end{cases}$

The total instantaneous cost rate, constituted by inventory holding costs and scrap costs, is defined as $c(y_1, y_2, \sigma_1, \sigma_2) = c_{hc_1} \cdot y_1 + c_{hc_2} \cdot y_2 + \omega \cdot c_{sc} \cdot y_2$, where $c_{hc_1}(c_{hc_2})$ denotes the unit holding cost rate for WIP at *arrival* (*qualified*) *queue* and c_{sc} is the unit scrap cost.

Uniformization is used to transfer a continuous time problem into a discrete time equivalent (Lippman 1975). Each decision epoch in the discrete time equivalent is the time interval between any two events. In this research, the *uniformization rate* φ is defined as

$$\varphi = \lambda + \mu_1 + \mu_2 + \gamma_1 + r_2 + f_1 + f_2 + Q_{q_e} \cdot \omega .$$

We define $V_n(y_1, y_2, \sigma_1, \sigma_2)$ as the optimal expected costs in the n^{th} iteration of the backward value iteration algorithm by Puterman (1994). $V_n(y_1, y_2, \sigma_1, \sigma_2)$ is also called the optimal value function and can be defined iteratively by the optimality equation (1). In the optimality equation (1), $a\mu_1$ refers to the production rate of the upstream server when action a is taken. In which, a is the production rate in fraction of the maximum production rate of the upstream batch server. Let initial condition $V_0(y_1, y_2, \sigma_1, \sigma_2) = 0$. In each iteration, the backward value iteration algorithm searches for an optimal action for each state such that the total cost is minimized.

Optimality Equation:

$$\begin{aligned} V_{n+1}(y_1, y_2, \sigma_1, \sigma_2) = \min_{0 \le a \le 1} & \left\{ \begin{array}{l} \dfrac{a \cdot \sigma_1 \cdot \mu_1}{\varphi} \cdot V_n\left(y_1 - \min\{y_1, B\}, y_2 + \min\{y_1, B\}, \sigma_1, \sigma_2\right) \\ + \dfrac{(1-a) \cdot \sigma_1 \cdot \mu_1}{\varphi} \cdot V_n\left(y_1, y_2, \sigma_1, \sigma_2\right) \end{array} \right\} \\ & + c(y_1, y_2, \sigma_1, \sigma_2) + \frac{\lambda}{\varphi} \cdot V_n\left(y_1 + 1, y_2, \sigma_1, \sigma_2\right) + \frac{\sigma_2 \cdot \mu_2}{\varphi} \cdot V_n\left(y_1, \max\{y_2 - 1, 0\}, \sigma_1, \sigma_2\right) \\ & + \frac{y_2 \cdot \omega}{\varphi} \cdot V_n\left(y_1, \max\{y_2 - 1, 0\}, \sigma_1, \sigma_2\right) + EH_n(y_1, y_2, \sigma_1, \sigma_2) \end{aligned}$$

, where $EH_n(y_1, y_2, \sigma_1, \sigma_2)$ depicts the expected future costs caused by other transitions types. (1)

3 IMPLEMENTATION OF BPAC MODEL

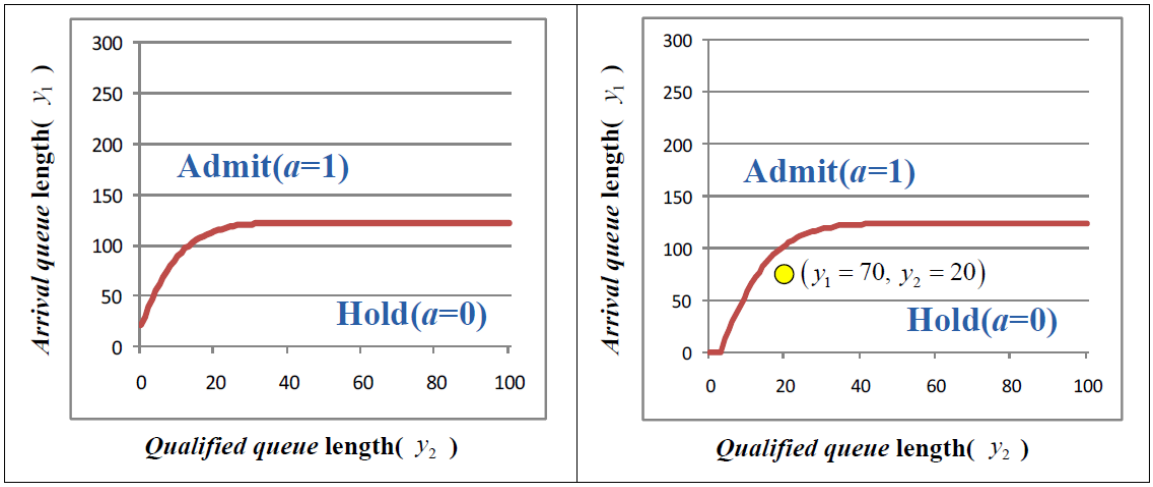
In this section, the implementation procedure of BPAC model is revealed. To illustrate how BPAC works, we provide a numerical example to explain the implementation procedure. In this example, we assume that the holding cost rates in both queues are 1 per hour and the unit scrap cost is 60. Arrival processes follow Poisson distribution with $\lambda = 6.4$ (jobs/hour). Times between server failures at the upstream and downstream stations are exponentially distributed with mean 30 hours. Server repair times at both stations are exponentially distributed with mean 0.5 hours. And the service rate at the upstream batch process server is exponentially distributed with $\mu_1 = 2$ (batch/hour). The service rate at the downstream single process server is exponentially distributed with $\mu_2 = 8$ (jobs/hour). The PQT constraint for *qualified queue* is 4.5 hours. The batch capacity of the upstream batch process server is 12 jobs. The implementation procedure of BPAC model is as follows.

Step1: Derive the dynamic admission control policy

We employ backward value iteration algorithm to solve the optimality equation (1) and find the optimal production control policy. The optimal policy will be stored in database for real-time production control. Figure 2 shows an example of the optimal control policy.

Step2: Implement the optimal policy in simulation experiments

Simulation experiments are conducted on *eM-plant* software. During simulation, the system state changes constantly. The control of the upstream batch process depends on the optimal control policy derived in step1. For instance, if the current state is $S=(70,20,1,1)$, the optimal control action is hold according to Figure 2 (b). Therefore, the upstream batch server will close and all jobs for the upstream process will be kept in *arrival queue*.



(a) $(\sigma_1,\sigma_2)=(1,0)$ (b) $(\sigma_1,\sigma_2)=(1,1)$

Figure 2: Optimal admission control policy for the batch process machine

4 SIMULATION STUDY

We carry out simulation experiments to evaluate the performance of BPAC model. In simulation study, three indices are adopted to measure the performance of different methods over simulation horizons. The performance indices are system throughput, scrap count and total cost. To analyze the performance of BPAC model, first come first serve (FCFS) control at the upstream workstation is selected for comparison. FCFS is widely used in many manufacturing systems (Akcali et al. 2001, Bernier and Frein 2004). Under FCFS, earliest arrival jobs before the upstream stage are processed first whenever the upstream machine is available.

In the following section, we compare BPAC policy with FCFS through simulation on *eM-plant*. For each dispatching rule, 5 replications are carried out. The simulation periods are 50 days with an additional warm-up period of 5 days. The experimental parameter set is the same as in section 3. Simulation results for BPAC and FCFS are shown in Table 1. The results show that the proposed BPAC method outperforms FCFS in important system criteria. First, the total throughput is raised by 4% and the total scrap count is reduced by 52%. Accordingly, BPAC model improves total throughput by cutting down scrap count. The raise on throughput will make a contribution to an enterprise's capability to fulfill demands.

According to Table 1, even though the inventory cost for BPAC is 26% more than that of FCFS, the total cost is still improved by 39%. Since BPAC method produces a stricter control policy, jobs are more possibly kept at upstream queue to avoid scraps. Even though the average queue length is longer, BPAC

could effectively lower the probability of violation against PQT constraints. Therefore, the outstanding performance of BPAC on total cost results from the significant reduction on scrap count.

Table 1: Performance results of different control policies

Control policy	Performance index			
	Throughput	Scrap count	Inventory cost	Total cost
BPAC	7238	353	22512	43717
FCFS	6933	733	16719	60711

To show the robustness of BPAC in general systems, we compare BPAC with FCFS in another 8 randomly generated systems. Table 2 shows the average performance improvement of BPAC in those 8 systems. According to the simulation results, total costs and throughput are improved by 29.1% and 4.3% in average. Meanwhile, the violation of PQT constraints is reduced by 59.1% in average.

Table 2: Average performance improvement in eight systems

	Performance index			
	Throughput	Scrap count	Inventory cost	Total cost
Performance Improvement under BPAC	4.3%	59.1%	-28.4%	29.1%

5 CONCLUSION

This paper studies a two-stage tandem queueing system under PQT constraints. In this system, after completion of the upstream batch process, jobs enter *qualified queue* for the downstream single process with specified PQT constraints. If a job cannot be processed within the PQT limit, it becomes scraped and the scrap cost is incurred. Due to the processing characteristic of the upstream server, plenty of WIP enter *qualified queue* simultaneously. As a consequence, a long queue is formed and the probability of violation against PQT constraints increases. In addition, uncertainty factors like process time, machine breakdown and repair also complicate this problem. Thus, we develop an efficient mechanism called BPAC for solution of the optimal control policy.

Incorporating Markov decision process (MDP) in BPAC model, we develop a tool to acquire the optimal control actions under different system states. To validate our model, BPAC policy is compared with FCFS through simulation study. Simulation results indicate that BPAC not only reduces total production costs significantly, but also raises production output. The performance improvement on these critical indices proves that BPAC is an efficient control policy in the two-stage manufacturing system under PQT constraints.

ACKNOWLEDGMENTS

This research is supported in part by the National Science Council of Taiwan under grant NSC 99-2221-E-002-152-MY3.

REFERENCES

Akcali, E., K. Nemoto, and R. Uzsoy. 2001. "Cycle-Time Improvements for Photolithography Process in Semiconductor Manufacturing." *IEEE Transactions on Semiconductor Manufacturing* 14(1): 48-56.

Bernier, V., and Y. Frein. 2004. "Local Scheduling Problems Submitted to Global FIFO Processing Constraints." *International Journal of Production Research* 42(8): 1483-1504.

Lippman, S. A. 1975. "Applying a New Device in the Optimization of Exponential Queueing Systems." *Operations Research* 23(4): 687-710.

Puterman, M. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* 1st ed. New York: John Willy & Sons, Inc.

Su, L. H. 2003. "A Hybrid Two-Stage Flowshop with Limited Waiting Time Constraints." *Computer & Industrial Engineering* 44:409-424.

AUTHOR BIOGRAPHIES

CHENG-HUNG WU is an Associate Professor at the Institute of Industrial Engineering, National Taiwan University, Taipei, TAIWAN. He received his Ph.D. and M.S. degree in industrial and operations engineering from the University of Michigan – Ann Arbor in 2005 and 2006. He was a technical consultant at Ares International Corp. and Oracle from 2000 to 2002. His research interests include: (1) decisions under uncertainties, particularly in applications such as production/supply-chain system analysis and dynamic control; (2) information and decision support systems (especially operations management and supply chain management); and (3) theoretical work in Markov decision processes and stochastic programming. His email address is wuchn@ntu.edu.tw.

YU-CHING CHENG was a graduate student in the Institute of Industrial Engineering at National Taiwan University. He received his M.S. degree from National Taiwan University in 2010. His research interests are in dynamic control of production systems.

PING-JU TANG is a graduate student in the Institute of Industrial Engineering at National Taiwan University. He received his bachelor degree from National Cheng Kung University in 2010. His research interests are in capacity planning and competition in duopoly. His email address is r99546011@ntu.eud.tw.

JIUN-YU YU is an Assistant Professor at the Department of Business Administration, National Taiwan University, Taipei, TAIWAN. He received his D.Phil. and M.Sc. degrees in applied statistics from the University of Oxford, UK. His research interests include: (1) decision making under uncertainty, (2) operations management, particularly in healthcare services, and (3) system dynamics. His email address is jyyu@ntu.edu.tw.