Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

APPROXIMATING THE PERFORMANCE OF A STATION SUBJECT TO CHANGEOVER SETUPS

Kan Wu
Zhang Wu

School of MAE
Nanyang Technology University
50 Nanyang Avenue
Singapore, 639798, SINGAPORE

Ning Zhao

Faculty of Science
Kunming University of Science and Technology
68 Wenchang Road
Kunming, 650093, CHINA

Yijun Xu

Industrial and Operations Engineering
University of Michigan
1205 Beal Avenue
Ann Arbor, MI 481097, USA

ABSTRACT

Changeover setups are induced by switching manufacturing processes among products. They commonly exist in flexible manufacturing systems. Modeling their queue time impact correctly is of fundamental importance in evaluating the performance of production systems. In this paper, the mean queue time approximation models are proposed based on the properties of changeover setups. The models are validated by simulations and perform well in the examined cases.

1 INTRODUCTION

Due to the random effects, queueing theory plays an important role in quantifying the performance of manufacturing systems. To use the correct queueing models at the right situation, a comprehensive classification is essential. The common stochastic events in a manufacturing system can be either time-based or run-based and preemptive or non-preemptive (Wu 2014a). The run-based non-preemptive events are generally called setups and can be further classified as state-induced and product-induced setups. State-induced setups may occur when a machine change its state from idle to busy. As inferred by its name, product-induced setups are induced by products rather than state changes.

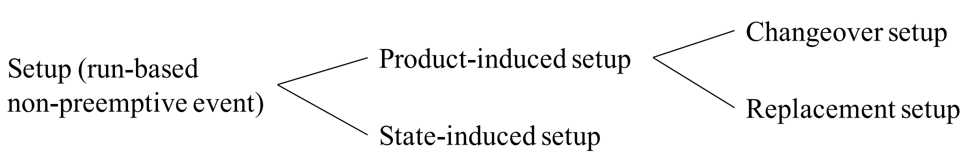

Figure 1. Classification for setups.

According to its characteristics, two types of product-induced setups commonly exist in practical manufacturing systems, i.e., changeover setup and replacement setup (see Figure 1). While a replacement setup is due to the life of a component or the usage of consumables, a changeover setup is induced by switching manufacturing processes among products. In this paper we study the mean queue time approximation of a single server queueing system with changeover setups.

When a machine produces a different product from the current one, it sometimes needs to change parts (e.g. masks of a photo lithographer) or parameter settings. If the setup process cannot be done in parallel with the processing of the previous product, it will have direct impact on job queue time and needs to be modeled explicitly. Since it switches machine settings in order to make different products, it is called a product-induced setup caused by changeover, or simply a changeover setup. Changeover setups commonly exist in the semiconductor manufacturing. For example, chemical mechanical planarization machine (CMP) machines are used to smooth wafer surfaces through both chemical and mechanical forces in a semiconductor fab. The process uses both abrasive and corrosive chemical slurry in conjunction with a polishing pad and retaining ring. The service times of a CMP machine can be around 1.5 ~ 2 minutes per wafer depending on the recipes, e.g. Trench, Oxide or Tungsten. The changeover time from Oxide to Tungsten is often less than two hours, but the changeover time from Oxide to Trench can be longer than five hours. Hence setup time distributions depend on the current product type and the product type which will be processed next.

As shown in Figure 2, changeover setups occur if there are multiple types of products. After the production process is switched to a new class, it will manufacture the same type of products consecutively until the next setup occurs. Hence, all the service times follow the same distribution between two setups. Furthermore, the setups may have different distributions when switching between different product types.

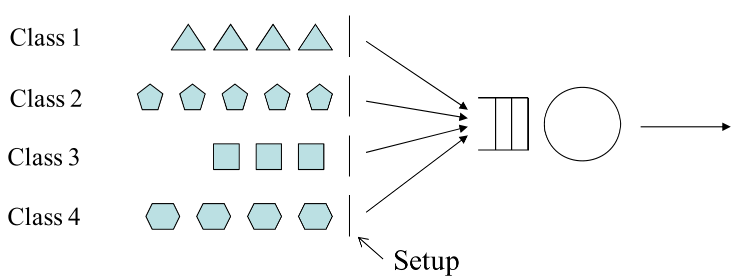

Figure 2. Changeover setups.

When the occurrence of a setup is memoryless and service times are independent and identically distributed (i.i.d.), the queueing model for changeover setups has been proposed by Hopp and Spearman (2008). Specifically, their model assumes the number of jobs processed between two consecutive changeover setups (i.e., serial batch size) follows the same geometric distributions. This assumption simplifies the model derivation but also limits its application.

Changeover setups can be viewed as an extension of polling systems. A typical polling system consists of a number of queues, attended by a single server who visits the queues in some orders to render service to customers waiting at the queues, typically incurring a changeover time while moving from one queue to another. Federgruen and Katalan (1996) characterized the impact from the mean or distribution of changeover setup times. They proved that all moments of waiting times and queue length are reduced if the higher order moment of changeover setup times is reduced. Takagi (1988, 2000) summarized the applications of polling models under the assumptions of cyclic order and Poisson arrivals, and discussed the queueing models under four types of dispatching rules: exhaustive, gated, limited and decrementing service. The cyclic order was generalized and replaced by a periodic service order table, or a polling table (Eisenberg 1972, Baker and Rubin 1987). However, those prior models didn’t consider the case that the distribution of changeover setups in manufacturing systems may depend on the current and future product types.

While polling systems are widely studied, existing models are not general enough to satisfy the needs of our cases with changeover setups in practical manufacturing systems. To evaluate the impact on system performance from the changeover setups, corresponding queueing models have to be derived. Specifically, we propose closed form queue time approximations for the changeover setup under practical settings.

The paper is organized as follows. Queueing models for changeover setups are proposed in Sections 2. Simulation validation is given in Section 3. The importance of behavior analysis is given in Section 4. Conclusion is given in Section 5.

2 THE MODEL

In practical manufacturing systems, changeovers can be attributed to the following reasons: (1) the move target of the current product has been satisfied, (2) a job of different product type has an urgent due date, (3) the downstream stations are starving for the other product type, or (4) relative queue lengths among products. For example, when the queue length of the current product is zero, the next product could be chosen based on their queue length or due dates. However, even if the queue length is not zero, a changeover may still occur if the queue length of the other product is too long, or the due date is urgent.

The key properties of changeover setups are: (1) there are multiple products with different service time distributions, and (2) setup time distributions depend on the current product type and the product type which will be processed next. In the following we first analyze the general cases with distinct changeover setups.

Assume the inter-arrival times between consecutive jobs are i.i.d. The job arrival rate is λ and the squared coefficient of variations (SCV) of inter-arrival time is c_a^2. N classes of products are served by a single server. The service rule is FCFS (first come first served). The service time of class i is S_i (i = 1,⋯, N). A random changeover setup $T_{i,j} > 0$ is incurred when the server leaves class i to class j ($i \neq j$). If the sever serves class i jobs continuously, no changeover setup is needed, i.e. $T_{i,i} = 0$. Define the service state by $C(t)$, and $C(t) = (i-1)N + j$ if the server is performing class i service at t after the completion of class j service ($1 \leq i,\ j \leq N$). Let $C(0) = (i-1)N + i$ if the server performs class i service at the initial epoch.

Since a changeover setup is induced by multiple independent random factors and only depends on the current and the next product types, it is reasonable to assume the occurrence of a changeover setup at each job is Markovian, and the service process can be modeled as a discrete time Markov chain $\{C(t), t \geq 0\}$ with state space $\{1,2,,\cdots,N^2\}$. Assume that the transition probability from serving class i to class k, i.e., from state $(i-1)N + j$ to state $(k-1)N + i$ is $p_{(i-1)N+j,(k-1)N+i}$ $(1 \leq i,j,k \leq N)$. Note that state $(i-1)N + j$ corresponds to the case that the server is serving class i and state $(k-1)N + i'$ denotes that the previous service class before class k is i'. $p_{(i-1)N+j,(k-1)N+i'} = 0$ if $i \neq i'$. The transition probability matrix of $\{C(t), t \geq 0\}$ is

$$P = \left[p_{m,n}\right]_{N^2 \times N^2}, \tag{1}$$

where $p_{(i-1)N+j,(k-1)N+i'}$ =0, $p_{(i-1)N+j,(k-1)N+i} = p_{(i-1)N+j',(k-1)N+i}$, and $\sum_{n=1}^{N^2} p_{m,n} = 1$ $(1 \leq i,i',j,j',k \leq N, i \neq i', 1 \leq m \leq N^2)$.

Under the existence of setups, the concept of service time has to be generalized. From the view point of capacity, generalized service time is defined as follows (Wu *et al.* 2011b):

> Generalized service time = job departure time − the time epoch when the job first claims capacity of the machine,

where job departure time is the time epoch that a job releases machine capacity. A job claims capacity of a machine if (i) the job is present at the machine, (ii) the preceding job has released machine capacity, and (iii) the machine is ready to process this job, or is ready to perform a product-induced setup.

Hence, the generalized service time at state $(i-1)N+j$ is $G_{(i-1)N+j}=S_i+T_{j,i}$. When the Markov process $\{C(t), t\geq 0\}$ is ergodic, the stationary probability vector $\pi=(\pi_1,\pi_2,\cdots,\pi_{N^2})$ exists. The mean generalized service time at the server is

$$E(G)=\sum_{i=1}^{N^2}E(G_i)\pi_i. \tag{2}$$

And the expected value of the second moment of generalized service time is

$$E(G^2)=\sum_{i=1}^{N^2}E\left(G_i^2\right)\pi_i. \tag{3}$$

For stability, we assume $\rho_G=\lambda E(G)<1$. Based on Kingman's heavy-traffic approximation for a *G*/*G*/1 queue (Medhi 2002), the mean queue time (QT) can be approximated by

$$E(QT)\approx\frac{c_a^2+c_G^2}{2}\frac{\rho_G}{1-\rho_G}E(G), \tag{4}$$

where $c_G^2={(E(G^2)-E^2(G))}\big/{E^2(G)}$.

3 MODEL VALIDATION

Simulation experiments are conducted to validate the mean queue time approximation for changeover setup models. The following results validates Eq. (4) for the model in which the changeover setups depend on the current and the following product types.

Assume jobs arrive according to a Poisson stream. Ten different mean arrival rates resulting in system utilizations ρ_G from 10% to 95% are considered. There is a single server which provides three classes of service. A random changeover setup $T_{i,j}$ is incurred when the server leaves class *i* service to class *j* service (*i*,*j*=1,2,3). The service times and changeover setup times follow gamma distributions. Let $E(S_1)$ =30 minutes, $E(S_2)$ =40 minutes, $E(S_3)$ =50 minutes. The mean changeover setup times are $E(T_{2,1})=30, E(T_{3,1})=35, E(T_{1,2})=28, E(T_{3,2})=46, E(T_{1,3})=42, E(T_{2,3})=58$. The transition probability matrix of $\{C(t), t\geq 0\}$ is

$$P=\begin{bmatrix} 2/5 & 0 & 0 & 1/5 & 0 & 0 & 2/5 & 0 & 0 \\ 2/5 & 0 & 0 & 1/5 & 0 & 0 & 2/5 & 0 & 0 \\ 2/5 & 0 & 0 & 1/5 & 0 & 0 & 2/5 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 1/2 & 0 & 0 & 1/6 & 0 \\ 0 & 1/3 & 0 & 0 & 1/2 & 0 & 0 & 1/6 & 0 \\ 0 & 1/3 & 0 & 0 & 1/2 & 0 & 0 & 1/6 & 0 \\ 0 & 0 & 1/6 & 0 & 0 & 1/6 & 0 & 0 & 2/3 \\ 0 & 0 & 1/6 & 0 & 0 & 1/6 & 0 & 0 & 2/3 \\ 0 & 0 & 1/6 & 0 & 0 & 1/6 & 0 & 0 & 2/3 \end{bmatrix}.$$

Three cases with different settings are considered:

Case 1: The SCV of S_i is 0.2, i =1, 2, 3. And SCV's of $T_{2,1}, T_{3,1}, T_{1,2}, T_{3,2}, T_{1,3}, T_{2,3}$ are 0.2, 0.3, 0.1, 0.5, 0.1 and 0.4 respectively.
Case 2: The SCV of S_i is 0.4, i =1, 2, 3. And SCV's of $T_{2,1}, T_{3,1}, T_{1,2}, T_{3,2}, T_{1,3}, T_{2,3}$ are 0.2, 0.3, 0.1, 0.5, 0.1 and 0.4 respectively.

Case 3: The SCV of S_i is 0.2, i =1, 2, 3. And SCV's of $T_{2,1}$, $T_{3,1}$, $T_{1,2}$, $T_{3,2}$, $T_{1,3}$, $T_{2,3}$ are 0.4, 0.6, 0.2, 1, 0.2 and 0.8 respectively.

Since the SCV of service time is practical manufacturing systems is generally small, all three cases have small service time SCV (Inman 1999). The difference between Case 1 and Case 2 is that the SCV of S_i in Case 2 is two times of that in Case 1. The difference between Case 1 and Case 3 is that the SCV of $T_{i,j}$ in Case 3 is two times of that in Case 1 (*i, j*=1,2,3).

Thirty replications are conducted at each utilization ρ_G. Each of the 30 replications is composed of 1000,000 jobs after discarding the first 2000,000 jobs for warm-up. The results are shown in Table 1, where SQT is the queue time from simulation, 95% half-width of the confidence intervals of the corresponding sample mean is given after its sample mean and AQT is the approximate queue time computed based on Eq. (4). Due to the large sample size, the 95% half-width of the confidence intervals are all smaller than 1% in all test cases. The percentage difference from AQT compared with SQT is reported in "Diff %" (Diff% = AQT/SQT – 1).

Table 1. Mean queue time comparison for queueing systems with distinct distributed changeover setups.

	Case 1			Case 2			Case 3		
ρ_G	SQT	AQT	Diff%	SQT	AQT	Diff%	SQT	AQT	Diff%
10%	4.26 ± 0.01	4.25	-0.1%	4.60 ± 0.01	4.594	-0.2%	4.45 ± 0.01	4.44	-0.2%
20%	9.59 ± 0.02	9.57	-0.2%	10.35 ± 0.02	10.34	-0.1%	10.01 ± 0.02	9.99	-0.2%
30%	16.43 ± 0.03	16.40	-0.1%	17.73 ± 0.03	17.72	0.0%	17.15 ± 0.03	17.12	-0.2%
40%	25.52 ± 0.04	25.52	0.0%	27.56 ± 0.05	27.56	0.0%	26.67 ± 0.04	26.63	-0.1%
50%	38.28 ± 0.07	38.27	0.0%	41.29 ± 0.07	41.34	0.1%	40.05 ± 0.06	39.95	-0.3%
60%	57.43 ± 0.12	57.41	0.0%	62.02 ± 0.12	62.01	0.0%	59.94 ± 0.09	59.92	0.0%
70%	89.22 ± 0.20	89.31	0.1%	96.24 ± 0.22	96.46	0.2%	93.23 ± 0.18	93.21	0.0%
80%	152.66 ± 0.48	153.10	0.3%	165.19 ± 0.54	165.4	0.1%	159.23 ± 0.60	159.79	0.4%
90%	342.51 ± 2.00	344.47	0.6%	368.65 ± 2.58	372.1	0.9%	357.75 ± 2.52	359.53	0.5%
95%	723.04 ± 7.32	727.22	0.6%	786.69 ± 8.60	785.5	-0.2%	762.35 ± 8.96	759.02	-0.4%

Comparing Case 2 and Case 3 with Case 1, we find that the increase of variability of service times and changeover setups cause the mean queue time increase. The percentage difference (Diff %) is small at all utilizations in the three cases. The approximation performs well. Its effect doesn't change too much because of the variability of service times and changeover setups.

Since Eq. (4) assumes that service time distributions are i.i.d., the errors could be caused by the service time dependence among the jobs of the same product type. In other words, the approximate model would perform better if the probability to process the same product type consecutively is not extremely high. Otherwise, the approximate model tends to underestimate the mean queue time. Furthermore, while the model still performs well in heavy traffic, the approximation model performs the best at around 60% utilization, which gives the smallest approximate error. This phenomenon could be also caused by the i.i.d. assumption of service times.

4 BEHAVIOR ANALYSIS

The classification and models proposed in this paper is done by a systematic approach. That is, (1) systemic classification of events, (2) defining service time from the view point of capacity, (3) deriving queueing models accordingly. As shown in Figure 3, the previous two belong to behavior analysis, while the last one belongs to model derivation.

Behavior Analysis	Classification of Events
	Service Time Definition
Model Derivation	Queueing Model Derivation

Figure 3. Methodology for studying manufacturing system behavior.

The importance of behavior analysis can be seen in Wu and Hui (2008) and Wu, McGinnis, and Zwart (2011a), for example. Based on a detailed behavior analysis, Wu and Hui (2008) studied the characteristic of service time and identify the dependence between effective service times and arrival rates. This key observation directly leads to the definition of generalized service time (Wu, McGinnis, and Zwart 2011b). Through a detailed behavior analysis, Wu, McGinnis, and Zwart (2011a) identified the dependence between batching times and queueing times. This observation leads to a better queue time approximation model for a parallel batching station.

By the same token, Wu (2014a) proposed a comprehensive classification for different types of interruptions of a station in practice manufacturing systems. Queueing models of each type of interruption is also given. Behavior analysis is in vain without model derivation, but derivation without thoughtful analysis may lead to incorrect models as well. While model derivation is often appreciated in academia, behavior analysis can sometimes be overlooked but is indeed a key step to derive a proper model.

5 CONCLUSION

In order to satisfy varying customer demand by product variety, flexible manufacturing systems are prevalent in modern manufacturing systems. Since changeover setups commonly exist in flexible manufacturing systems, properly evaluating its impact on queue time plays a key role in improving system performance.

When (1) there are a single service time distribution and a single setup time distribution, and (2) the occurrence of the setup is memoryless and follows geometric distributions, a changeover setups can be simply modeled by the queueing model proposed by Hopp and Spearman (2008). Based on this approach, integrated queueing models, which consider all types of interruptions or batching behaviors are proposed by Wu, McGinnis, and Zwart (2011b) and Wu (2014b), respectively.

Changeover setups are induced by switching manufacturing processes among products. Under proper justification based on the observations from practical production lines, we assume that the occurrence of changeover only depends on the current and the next product type. With this assumption, the changeover process can be modeled as a discrete time Markov chain. While the model performs well when the service time dependence is weak, it may underestimate queue time when the dependence is high. Future research is expected to improve the model.

In addition to a changeover setup, a replacement setup also commonly exists in production lines. Model derivation for different types of product-induced setup is left for future research (Wu and Zhao 2015). In order to quantify the queue time performance of a station in a queueing network, the effect of different types of setups on the intrinsic ratio (Wu and McGinnis 2012, Wu and McGinnis 2013) is also an interesting topic and left for future research.

ACKNOWLEDGMENTS

This research is supported in part by the Singapore MOE Academic Research Funding Tier 1 Grant M4011107.

REFERENCES

Baker, J., and I. Rubin. 1987. "Polling with a general-service order table." *IEEE Transactions on Communications* 35 (3): 283-288.

Eisenberg, M. 1972. "Queues with periodic service and changeover time." *Operations Research* 20(2): 440-451.

Federgruen, A., and Z. Katalan. 1996. "The impact of setup times on the performance of multiclass service and production systems." *Operations Research* 44 (6): 989-1001.

Hopp, W.J., and M.L. Spearman. 2008. *Factory physics: Foundations of manufacturing management*: Burr Ridge, Illinois: Irwin/McGraw-Hill.

Inman, R.R. 1999. "Empirical evalucation of exponential and independence assumptions in queueing models of manufacturing systems." *Production and Operations Management* 8 (4): 409-432.

Medhi, J. 2002. *Stochastic models in queueing theory*. Academic Press.

Takagi, H. 1988. "Queuing analysis of polling models." *ACM Computing Surveys* 20 (1): 5-28.

Takagi, H. 2000. "Analysis and application of polling models." In *Performance Evaluation: Origins and Directions*, edited by G. Haring, C. Lindemann, and M. Reiser, 423-442. Springer.

Wu, K. 2014a. "Classification of queueing models for a workstation with interruptions: A review." *International Journal of Production Research* 52 (3): 902-917.

Wu, K. 2014b. "Taxonomy of batch queueing models in manufacturing systems." *European Journal of Operational Research* 237 (1): 129-135.

Wu, K. and K. Hui. 2008. "The determination and indetermination of service times in manufacturing systems." *IEEE Transactions on Semiconductor Manufacturing* 21 (1): 72-82.

Wu, K. and L. McGinnis. 2012. "Performance evaluation for general queueing networks in manufacturing systems: Characterizing the trade-off between queue time and utilization." *European Journal of Operational Research* 221 (2): 328-339.

Wu, K. and L. McGinnis. 2013. "Interpolation approximations for queues in series." *IIE Transactions* 45 (3): 273-290.

Wu, K., L. McGinnis and B. Zwart. 2011a. "Approximating the performance of a batch service queue using the $m/m^k/1$ model." *IEEE Transactions on Automation Science and Engineering* 8(1):95-102.

Wu, K., L. McGinnis and B. Zwart. 2011b. "Queueing models for a single machine subject to multiple types of interruptions." *IIE Transactions* 43(10): 753-759.

Wu, K. and N. Zhao. 2015. "Queue Time Approximations for a Workstation with Product-Induced Setups." Working paper.

AUTHOR BIOGRAPHIES

KAN WU is an Assistant Professor in the Division of Systems and Engineering Management at Nanyang Technological University. He received the B.S. degree from National Tsinghua University, M.S. degree from University of California at Berkeley, and Ph.D. degree in Industrial and Systems Engineering from Georgia Institute of Technology. He has 10 years' experience in the semiconductor industry, from consultants to managers. Before joining NTU, he was the CTO and founding team member of a startup company in the US. His PhD dissertation was awarded the 3rd place for the IIE Pritsker Doctoral Dissertation Award in 2010. His research interests are primarily in the areas of queueing theory, with applications in the performance evaluation of supply chains and manufacturing systems. His email is kan626@gmail.com.

NING ZHAO is an Assistant Professor in Faculty of Science at Kunming University of Science and Technology. She received the M.S. degree in Applied Mathematics from Shanghai Jiaotong University, and received a Ph.D in Business Information Systems from University of Macau. Her research interest is queueing theory. Her email is zhaoning1102@gmail.com.

YIJUN XU is a Graduate Student of Industrial and Operations Engineering Department at University of Michigan. He got his Bachelor degree in Mechanical Engineering from Nanyang Technological University. His email is yijunx@umich.edu.

ZHANG WU is a Senior Research Fellow in the School of Mechanical and Aerospace Engineering at Nanyang Technological University, Singapore. Dr. Wu received his B.Eng degree in Mechanical Engineering from Huazhong University of Science and Technology, China, and both his M.Eng and Ph. D. degrees in Mechanical Engineering from McMaster University, Canada. His research interests include quality control, reliability, nonlinear optimization and geometrical tolerance. His email is MZWU@ntu.edu.sg.