Computers and Operations Research 88 (2017) 161–174

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

Approximating class-departure variability in tandem queues with downtime events: Regression-based variability function

Ruth Sagron*, Gad Rabinowitz, Israel Tirkel

Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, P.O.B. 653, Beer Sheva, 84105 Israel

ARTICLE INFO

Article history:
Received 9 March 2015
Revised 2 July 2017
Accepted 3 July 2017
Available online 6 July 2017

Keywords:
Regression-based variability function (RBVF)
Tandem queues
Class-departure variability
Queue performance
Decomposition approximation methods

ABSTRACT

Predicting queue performance by approximating class-departure variability in tandem queues with downtime events via existing decomposition methods is neither accurate enough nor efficient enough. Analytic approximations, if conducted alone, lack accuracy but attempting to increase accuracy by incorporating simulation to analytic approximation has proved to require significant computation efforts. The aim of this paper is to reduce the latter inefficiency by modeling the Regression-Based Variability Function (RBVF) designed to approximate the between-class effect by exploiting the departure process from a single queue. The new approach predicts performance of *n*-tandem queues by reducing the focus to two-tandem queues for each traffic intensity level, as well as by modeling different policies of downtimes (e.g. first-come-first-served or priority). Numerical experiments demonstrate that the proposed RBVF delivers both accuracy and efficiency improvements: the relative errors associated with RBVF are about three times smaller than the best existing analytic procedures and the computation efforts associated with RBVF are about five times smaller than existing analytic procedure combined with simulation.

1. Introduction

This paper is concerned with the impact on the downstream bottleneck queue of downtime events at the first queue in open tandem queuing network. Inter-arrival times of customers, as well as downtime events follow general and independent distributions. Moreover, customers pace on along the entire network and handling the downtime events halting the server, which occur only in the first queue, follow service policies of various types. Wu and McGinnis (2011) and Wu (2014) classify downtime events as either run-based or time-based arrival. Run-based arrival is associated with the number of jobs processed while time-based arrival is associated with the time passed. One can distinguish as well between scheduled operations of preventive maintenance (PM) and random breakdown events that trigger corrective maintenance (CM), both of which may be run-based or time-based. The preplanned PM is typically non-preemptive and is handled first-come-first-served (FCFS) while, on the other hand, CM is typically handled with high priority with or without preemption.

The frequent occurrence of downtime events in manufacturing presents one of the difficulties in developing proper queuing approximations for manufacturing systems, raising interest in their impact on downstream bottleneck performance. These difficulties involve significant interference among classes and generation of non-renewal and highly variable flow along the network due to long downtime events, causing starvation of downstream bottleneck queues with non-renewable flow. Closed-form solutions to predict queue performance are nearly intractable due to this inherent complexity and, hence, Buzacott and Shanthikumar (1993) and many others approximated performance analysis.

Since it is possible to model a system with downtimes as a multi-class system, our problem can be approached via decomposition without aggregation (DWOA) as done by Bitran and Tirupati (1988) and Whitt (1994). Oftentimes these procedures employ two-moment approximations to predict performance of multi-class queueing networks with deterministic routing. Yet, especially in bottleneck queues, two-moment approximations used for renewal processes are unsatisfactory for use for non-renewal processes with high variability that exist in systems with downtimes. For these systems, a better approximation is needed to represent the non-renewal variability of class-departure and thus address the challenge of interference among classes, which includes not only multiple dependent queues but also with the non-renewal flow created by the interference among classes.

Analytic decomposition methods characterize non-renewal processes within a network by moments of equivalent renewal

* Corresponding author.
E-mail addresses: ruthishilman@gmail.com, ruthshil@post.bgu.ac.il (R. Sagron), rgadi@bgu.ac.il (G. Rabinowitz), tirkel@bgu.ac.il (I. Tirkel).

http://dx.doi.org/10.1016/j.cor.2017.07.003
0305-0548/

processes. Approximating class-departure variability from the first queue and then successively from subsequent queues, thus enables predicting the performance of downstream queues by common two-moment approximations. However, the predictions obtained by using these methods lack accuracy since the approximated variability parameter of the equivalent class-departure renewal process at each queue is not representative enough of the non-renewable process.

Methods that cope with the non-renewal challenge in single-class systems, such as Whitt (1995) or Wu and McGinnis (2013), are inapplicable in systems with downtimes even when capturing downtimes by service rate and variability. It is simple to reduce systems with different types of downtimes to equivalent single-class system, by capturing the downtimes with service rate and variability, but each multi-class system exhibits a different performance. However, there have been several attempts to cope with this challenge in multi-class systems with deterministic routing. Kim (2005) was the first to reach a viable approximation for such systems by merging (for FCFS policy) the DWOA procedure with the variability function (VF) method proposed by Whitt (1995). Sagron et al. (2015) integrated Kim's method with a simulation-based VF to approximate the between-class effect in tandem queues with downtimes and reached a more accurate approximation while enabling various service policies (not only FCFS) of handling downtimes. Both methods (and VF in general), approximate the adjusted departure variability as reflected at downstream queues, rather than the actual departure variability (the common approach in the literature) and, thus, dramatically improve the predictions. However, the increased accuracy obtained by Sagron et al.'s incorporation of simulation into Kim's analytic approximation is associated with a significant increase in computation efforts.

To achieve better computation efficiency than demonstrated by Sagron et al. (2015), this paper proposes the Regression-Based Variability Function (RBVF) to approximate the between-class effect in a queue with downtimes. Adopting Sagron et al.'s basic principles, including the within-class and between-class structure for class-departure variability, we refine the procedure to approximate the VF for the between-class effect by exploiting the theory of departure process from a single queue and the roughly linear relationship in the system. By extending the VF method in this manner, we can predict the performance of n-tandem queues by reducing the focus to two-tandem queues for each traffic intensity level. Moreover, we show that this extension makes it possible to reduce the computation efforts and increase flexibility. By relying on the within-class and between-class structure, we also show the feasibility of modeling different policies of downtimes (e.g., FCFS or priority). Finally, the RBVF approximation is dependent neither on the number of queues in the tandem, nor on the number of queues between the queue that suffers downtimes and the predicted downstream queue.

The rest of the paper proceeds as follows. The next section presents the problem description and notations. Section 3 details the RBVF model, exemplifying the roughly linear relationship stimulating our approach and explaining how only one sub-simulation suffices to determine the VF, leading to the results, including sensitivity analyses, in Section 4. Finally, Section 5 summarizes and concludes this paper.

2. Model description and notations

To convey the fundamental insights and the methodological steps of the RBVF model, we focus on a basic network of n independent tandem queues with unlimited waiting space in order to explore the impact of the downtime events occurring only at the first queue and causing high traffic variability on a bottleneck queue. Moreover, we convert here a single-class tandem queues into a dual-class one for the sake of incorporating downtimes into the model. Class-1 represents the class of interest, arriving at the first queue and then serially routed through the remaining n-1 queues in the network. Class-2 represents the downtime events arriving exogenously at the first queue, served with non-preemptive policy and then immediately leaving the system. Fig. 1 demonstrates a basic network of 10 independent tandem queues.

We demonstrate our approach for two service policies - FCFS and priority. These policies may respectively represent, for example, preventive and corrective maintenance. For both policies, we focus on time-based downtime cases and assume non-preemption for the sake of simplicity. Yet, it is possible to modify the proposed method for other downtime cases and other assumed policies.

All processes in the tandem queues are generally distributed and assumed independent. The first queue at which a Class-2 arrives is defined by $G,G/G,G/1$, where the first element represents Class-1 and the second represents Class-2 in the notations for the arrival and service characteristics. Assuming that $\rho_j < 1$, $\forall j \in 1, ..., N$, where $\rho_j = \rho_{1,j} + \rho_{2,j}$ ($\rho_{2,j} = 0$ where only Class-1 continues to queue j), and in order to examine the impact of downtime events on performances of bottleneck queues, we require that $\rho_N >> \rho_j \forall j \notin N$. Hereinafter, the following notations are used:

- A – Inter-arrival time
- $\lambda = 1/E(A)$ – Arrival rate
- S – Service duration
- $\rho = \lambda E(S)$ – Traffic intensity
- Ca^2 – Squared coefficient of variation (SCV) of the inter-arrival times
- Cd^2 – SCV of the inter-departure times
- Cs^2 – SCV of the service durations
- CB^2 – Approximation of the contribution of the between-class effect on Cd^2
- CW^2 – Approximation of the contribution of the within-class effect on Cd^2
- Wq – Average waiting time

When reference is to Class-i and Queue j indices i,j appear in the notation; when reference is to Class-i at a given queue only i appears; and when reference is to all classes at Queue j only j appears.

To differentiate between actual SCV versus approximated SCV, we denote the actual one by lower case letters and the approximated one by capital letters. Ca^2 represents an actual SCV, for example, while CA^2 represents an approximated SCV. To differentiate between the approximation for the actual Cd^2 and the approximation for its reflection in the downstream queues, the latter is denoted - adjusted CD^2. In the simulation employed later in this paper, we represent the external arrival and service processes by their rate and SCV. We use the following distributions: for $Cx^2 = 0$ - the degenerate deterministic distribution; for $Cx^2 = k^{-1}$ - the Erlang distribution (E_k); for $Cx^2 = 1$ - the exponential distribution; and for $Cx^2 > 1$ - the hyper-exponential distribution with two phases (H_2).

3. The RBVF model

This section introduces the RBVF model to approximate the between-class effect of Class-1 departure variability in tandem queues with downtimes. To reduce the computation efforts to approximate this effect by the VF procedure, RBVF is a calibrated regression model for each traffic intensity level in the downstream queues. Initially, we define the proposed approximation model. Then, in Section 3.1, we focus on the departure process in a single queue, justify the nearly linear shape of the parameters' impact on class-departure variability, and employ previous results to derive