Proceedings of the 2016 Industrial and Systems Engineering Research Conference
H. Yang, Z. Kong, and MD Sarder, eds.

A Closed-Form Approximation for an Airport Queueing Network

Abstract ID: 196

Abhijit Gosavi and Carlos R. Chaves
Missouri University of Science and Technology
Rolla, MO 65409

Abstract

Queuing networks (QNs) arise in airports during passenger checking-in. For studying such systems, much of the literature either suggests the use of discrete-event simulation models, which are unfortunately harder to optimize, or the use of models based on the exponential distribution for inter-arrival and/or service times. Closed-form models that work for any given distribution, which are more generally applicable and are easier to optimize, are *less frequently* studied in the literature. We present a two-moment mathematical approximation, applicable *for any given distribution*, to study waiting times and queue lengths in a typical/generic airport QN. The latter usually consists of two stages of queues: a multi-server (G/G/k) queue in the first stage for ID check and parallel, single-server (G/G/1) queues in the second stage for a metal-detector/body scanner. Our main contribution lies in developing an approximation for the squared coefficient of variation of the inter-departure time in a multi-server queue in the first stage, which is necessary to compute the same for the inter-arrival time to the queue(s) in the second stage. Numerical results show that our model approximates results from discrete-event simulation well. Our model can be handily incorporated into an optimization framework to determine the optimal number of servers.

Keywords

Queueing networks, airports, G/G/1 queues, G/G/k queues

1. Introduction

Discrete-event simulation (DES) happens to be a popular tool for estimating performance measures of queueing networks (QNs) found in airports. Crook [1], Cetek [2], and Guizzi [3] discuss some classical centralized DES models for airport QNs, while Ray and Claramunt [4] present a distributed approach that can be potentially parallelized. Brunetta, *et al.* [5] develop a closed-form model named SLAM (short of Simple Landslide Aggregate Model) for answering what-if questions related to airport performance. See also Manataki and Zogafros [6] for detailed DES-based models for airport QNs. A closed-form mathematical model that evaluates QNs in airports has some advantages over DES models, e.g., it is easier to optimize and the model can also be incorporated into spreadsheet software. Humphreys and Francis [7] call for a variety of performance evaluation tools for airports, including mathematical models. To the best of our knowledge, however, airport QNs have not been studied extensively via mathematical models. Two notable exceptions are Dorton [8], who studies an approximation based on M/M/1 queues (i.e., queues whose inter-arrival and services times are exponentially distributed), and Lovell *et al.* [9], who provide a diffusion approximation and compare it to results from an M/M/1 queue.

There is empirical evidence [10] to suggest that the inter-arrival times of customers to airport queues often have the gamma distribution, for which the M/M/1 model is not satisfactory and a model based on generally distributed inter-arrival times will be more attractive. In this paper, hence, we will present a closed-form mathematical model for evaluating the performance of a multi-stage QN in which one stage has a multi-server queue with generally distributed inter-arrival and service times (G/G/k) and the other stage has single-server queues with generally distributed inter-arrival times and services times (G/G/1). Thus, our work seeks to develop a more general model in which the distributions of the inter-arrival and service times can have any given distribution. Also, for our model, only the first two moments of the inter-arrival and service times will be needed. Further, the most challenging aspect of modeling a mixed QN of this kind, where one stage has a single-server queue(s) and another has multi-server queue(s), via traditional queueing calculus (see Buzacott and Shanthikumar [11]), is approximating the squared coefficient of variation of the time between successive departures from a G/G/k queue.

Contributions of this paper: In this paper, we present a new approximation procedure for estimating the squared coefficient of variation between successive departures from a G/G/k queue, which allows us to compute the mean waiting time of (and the mean number of people in) downstream queues in the system. In particular, our approach here essentially seeks to transform the G/G/k queue in question into a fictitious G/G/1 queue, such that the variance of the service time of the latter can be estimated from the parameters (variance of the service time and k) of the G/G/k queue. The variance of the service time can in turn be used to approximate the squared coefficient of variation of the time between successive departures from the G/G/k queue. We get encouraging numerical results from our approximation, and hope that this work will lead to further improvements in these approximations for airport QNs.

The rest of this paper is organized as follows. Section 2 provides the mathematical model for measuring the performance of QN. Section 3 presents numerical results from the model. Section 4 concludes the paper with remarks on future research.

2. Mathematical Model

The underlying problem here can be modeled as a 2-stage QN (see Ross [12]), where one has the data for the inter-arrival time to the first queue and knowledge of the following: (i) the number of servers in each stage, (ii) the probability of an entity leaving the first stage to join a queue in the second, and (iii) data for the service times in each queue in the system (as well as the queueing disciplines). We note that the QN we study is of the open, generalized Jackson network class [12].

In the above paragraph, by "data," we mean either the distributions of the underlying random variable (i.e., for inter-arrival times and the service times) or the values of the first two moments of the underlying random variables. Further, we will assume that each queue works on a first in first out (FIFO) discipline; note, however, that to the queue in the first stage, in many real-world systems, there may be business class travelers who will get priority. In this model, we assume that the influence of such non-FIFO travelers will be minimal. We will also assume that the travel time from exiting the first stage to joining the queue in the second stage is negligibly small. Usually, for simulation models, the underlying distributions are required, but our mathematical approximation here will rest on knowledge of only the first two moments. In other words, even if the distribution is available, only the first two moments will be needed for our approximation procedure.

Figure 1 represents the QN that we study here. Customers arrive to the first queue (the ID Check Queue, where identification documents are checked) in Figure 1, which is the first stage in the security processing. This queue is a multiple-server, single channel queue with generally distributed inter-arrival times and service times (G/G/k to use standard queueing notation). When customers complete their ID check, they are sent to one of the several parallel queues in the second stage (the Metal-Detector/Body Scanner and Carryon-Luggage Scanner Queue), shown in Figure 1. Each queue in the second stage is a single-server queue with generally distributed inter-arrival times and service times (G/G/1 to us standard queueing notation). The inputs to our model are hence (i) the first two moments of the inter-arrival time and service time to the first queue (first stage), (ii) the number of servers in the first stage, (iii) the number of queues in the second stage, and (iv) the first two moments of the service time for each queue in the second stage. The outputs from the model will be the mean waiting time and number in each queue in the system. Since our proposed mathematical model requires evaluation of a few formulas, a computer program will generate the numerical values almost instantaneously, and hence can be used to optimize the number of servers in the first stage and the number of queues in the second stage.

The basic methodology that we adopt is to apply (i) Marchal's approximation of a G/G/k queue [13] to obtain the performance measures from the first queue in the network, (ii) classical queuing calculus principles (see [11]) to obtain the first two moments of the inter-arrival time to each queue in the second stage and (iii) Marchal's approximation of a G/G/1 queue [14] to obtain the performance measures in the second queue in the network.

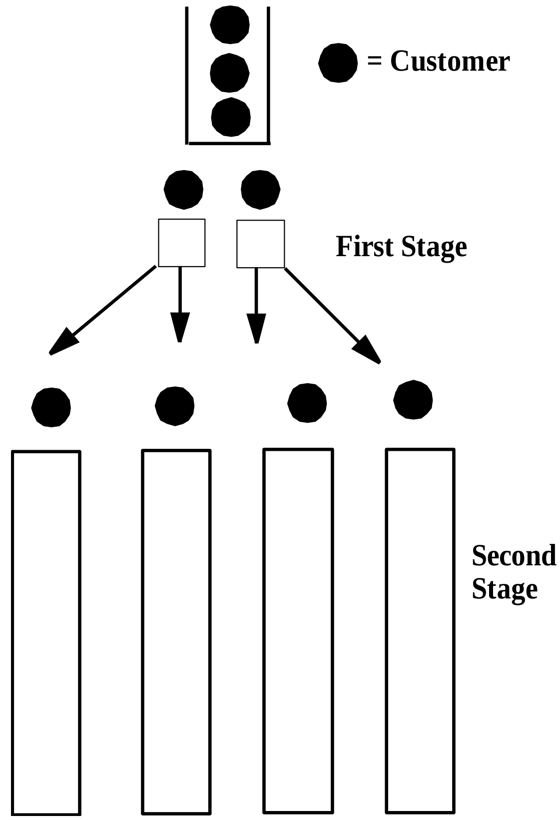

Figure 1. A QN in an Airport Security Line with 2 servers in the first stage and 4 queues in the second

2.1 Notation

We begin with some notation:

k: Number of servers in the multi-server queue in the first stage

λ: Mean rate of arrival= $\frac{1}{\mathrm{E(inter-arrival\ time)}}$ to the queue in the first stage

μ: Mean service rate = $\frac{1}{\mathrm{E(service\ time)}}$ of the queue in the first stage

ρ: Utilization in the first stage = $\frac{\lambda}{k\mu}$

$W_q^{G/G/k}$: Mean wait time in the multi-server queue in the first stage

$W_{q,i}^{G/G/1}$: Mean wait time in the *i*th queue in the second stage

σ_a^2: Variance of the inter-arrival time to the first queue

σ_s^2: Variance of the service time of one server in the first queue

C_a^2 : Squared coefficient of variation for the inter-arrival time = $\frac{\sigma_a^2}{\left(\frac{1}{\lambda}\right)^2}$

C_s^2: Squared coefficient of variation for the service time of one server in the first stage

C_d^2: Squared coefficient of variation for the time between successive departures from the first stage

$C_{a,i}^2$: Squared coefficient of variation for the inter-arrival time to the *i*th queue in the second stage

2.2 Model

We now present details of how the performance metrics are actually calculated for each stage. For the first stage, we use an approximation suggested by Marchal [13] for G/G/k queues for the mean waiting time in the queue:

$$W_q^{G/G/k} = \frac{\pi_0\left(\frac{\lambda}{\mu}\right)^k \rho}{k!\,(1-\rho)^2}\,\frac{(1+C_s^2)\left(C_s^2+(\rho^2 C_s^2)\right)}{2(\rho^2 C_s^2)\lambda} \tag{1}$$

in which we need the steady-state probability of having no customers in the system of an M/M/k queue, which is defined as follows [12]: $\pi_0 = \sum_{m=0}^{k-1} \frac{\mathrm{k\rho^m}}{\mathrm{m!}}$.

Now, we discuss the model for the second stage. Note that the mean rate of arrival to the *i*th queue in the second stage, λ_i, is given by: $\lambda_i = P_i\lambda$ where P_i is the probability of a customer selecting the *i*th queue in the second stage. Then, one can calculate the squared coefficient of variation of the time between successive departures from the first queue, based on the approximation in [11], via:

$$C_d^2 = \rho^2 C_s^2 + (1-\rho^2)C_a^2 \tag{2}$$

in which we approximate the squared coefficient of the service time in a G/G/k queue by the following:

$$C_s^2 = \frac{\sigma_s^2 k}{\left(\frac{1}{\mu}\right)^2}. \tag{3}$$

The approximation is rooted in aggregating the variance of the *k* servers and treating the multi-server queue as one whose variance is $\sigma_s^2 k$, effectively transforming the G/G/k queue into a fictitious G/G/1 queue for the purpose of modeling. The above equation is a key contribution of this paper. The squared coefficient of variability of the inter-arrival time at the *i*th queue in the second stage will then be given by the following well-known equation of queueing network calculus [11]: $C_{a,i}^2 = 1 - P_i + (P_i * C_d^2)$.

Using μ_i to denote the mean service rate of the ith queue in the second stage, we have that the utilization in the *i*th queue is given by $\rho_i = \frac{\lambda_i}{\mu_i}$. One can now compute the mean waiting time in the *i*th queue of the second stage using Marchal's approximation for a G/G/1 queue [14], which results in the following formula for the mean waiting time in the *i*th queue of the second stage:

$$W_{q,i}^{G/G/1} \cong \frac{\rho_i^2\,(1+C_{s,i}^2)(C_{a,i}^2+\rho_i^2 C_{s,i}^2)}{2\,(1-\rho_i^2)(1+\rho_i^2 C_{s,i}^2)\lambda_i} \tag{4}$$

where $C_{s,i}^2$ denotes the squared coefficient of variability of the service time of the *i*th queue. We note that the mean number in any queue in the entire QN can be easily computed from the mean waiting time in the queue via Little's rule.

3. Numerical Results

We tested our mathematical model on ten representative cases, which vary in terms of: (i) squared coefficients of inter-arrival time in the system, (ii) squared coefficient of variation of service time for queues in the first and second stage in the network, (iii) the number of servers, and (iv) the service time distributions. In all the cases we studied, there were five parallel queues in the second stage of which one was significantly slower and was used to represent the server used for special/additional screening, which requires extra time; the other four had identical service rates. We ran discrete-event simulations to benchmark the performance of our model. The error was computed against results from simulations. The error in the mean wait in the queue was calculated as: $Error\ \% = \frac{|W_q^{Model} - W_q^{Simulation}|}{W_q^{Simulation}}\ X\ 100.$

The numerical results from all our experiments are presented in Tables 1 and 2 below. The computer programs for evaluating the mean waiting times and queue length using our approximation were written in MATLAB, while those for the simulation results were written in ARENA. All programs were run on an Intel Pentium Processor with a speed of 2.66 GHz on a 64-bit operating system. The computer programs for our mathematical model took about 10 milliseconds; however, the simulation programs took longer (about 1 minute), since they involve multiple replications.

Usually, queueing approximations can result in errors of about 25% (see [13-14]). Therefore, our numerical results are quite encouraging: on the low end, the error computed was 0% and on the high-end the error was 26.5%.

Table 1: Results for first queue in the network: T (min, mode, max) denotes the triangular distribution, N(mean, variance) denotes the normal distribution, and Gm (mean, variance) denotes the gamma distribution. The inter-arrival time has a gamma distribution whose mean is 5 for each case and whose C_a^2 value is specified for each case in the table.

Case	k	C_a^2	Service Dist. First Stage	μ	C_s^2	$W_q^{Simulation}$	W_q^{Model}	% Error
1	2	0.45	T(1.33, 3.33, 15.33)	0.15	0.215	1.3952	1.6134	15.64
2	2	0.50	T(1.33, 3.33, 15.33)	0.15	0.215	1.4617	1.7613	20.50
3	3	0.60	N(10,5)	0.10	0.05	1.1643	1.4203	21.99
4	3	0.65	N(10,5)	0.10	0.05	1.2134	1.5344	26.45
5	4	0.75	N(10,5)	0.075	0.05	1.4112	1.5008	6.35
6	4	0.95	N(10,5)	0.075	0.15	2.1968	2.0739	5.59
7	6	0.65	Gm(20,60)	0.05	0.15	0.89842	1.1001	22.45
8	6	0.70	Gm(20,60)	0.05	0.15	1.1029	1.1769	6.71
9	7	0.65	T(4.67,11.87, 53.67)	0.043	0.215	0.9466	1.0338	9.21
10	7	0.70	T(4.67,11.87, 53.67)	0.043	0.215	1.1915	1.1031	7.42

Table 2: Results of the second queue in the network which consists of five single-server queues in parallel where service times are normally-distributed. The first four servers have a mean service rate of 1/20, while the fifth server has a rate of $1/23$. Also, P_i = 1/5 for all values of *i*.

	i= 1:5	Servers 1 - 4			Server 5		
Case	$C_{s,i}^2$	$W_q^{Simulation}$	W_q^{Model}	% Error	$W_q^{Simulation}$	W_q^{Model}	% Error
1	0.10	37.84375	38.5873	1.96	138.59	127.9198	7.70
2	0.15	40.73775	40.7397	0.00	141.08	135.1733	4.19
3	0.10	39.60825	38.67	2.37	156.17	128.1881	17.92
4	0.15	39.635	40.8236	3.00	141.05	135.4432	3.98
5	0.10	45.796	39.3593	14.06	119.73	130.4235	8.93
6	0.15	44.05875	42.5957	3.32	186.78	141.1413	24.43
7	0.10	38.36725	39.2673	2.35	141.95	130.1254	8.33
8	0.15	51.751	41.4298	19.94	186.94	137.3925	26.50
9	0.10	38.0355	39.5063	3.87	117.52	130.9	11.39
10	0.15	48.90425	41.6723	14.79	186.09	138.1723	25.75

4. Conclusions

Queuing approximations such as the G/G/1 approximation presented above are now widely used in manufacturing systems for measuring lead times (see Askin and Goldberg [15]). Queueing *network* (QN) approximations, which are more complex than the same for approximating a single queue, have also been used extensively in modeling production lines [16]. In this paper, we presented a new mathematical model for approximating a (mixed) 2-stage QN in which in one stage there is a G/G/k queue and in another there is a set of parallel queues, each belonging to the G/G/1 family. Our contribution in this paper is in formulating a novel way to compute the squared coefficient of variation for the time between successive departures from the G/G/k queue in the first stage. We obtained encouraging numerical results with our approximation procedure.

There are multiple avenues for future research based on this work. First, our approximating procedure can be used to optimize the number of servers. Another potential line for further research would measure the variance of the waiting time in each queue using the third moment. Finally, an important direction for queueing approximations in general is to find ways to reduce their errors.

References

1. Crook, S., 1998, "The Use of Simulation and Virtual Reality in the Design and Operation of Airport Terminals," Proceedings of the International Conference on Simulation, Conference Publication No. 457, 8-10.
2. Cetek, C., 2012, "Training Airport Capacity Development Using Discrete-Event Simulation," International Journal of Engineering Science and Technology, 4 (8), 3750 -3757.
3. Guizzi, G., Murino, T, and Romano, E., 2009, "A Discrete Event Simulation to Model Passenger Flow in an Airport Terminal," Mathematical Methods and Applied Computing, 2, 427 – 434.
4. Ray, C., and Claramunt C., 2003, "A Distributed System for the Simulation of People Flows in an Airport Terminal," Knowledge-Based Systems, 16, 191 – 203.
5. Brunetta, L., Righi,, L., and Andreatta, G. 1999, "An Operations Research model for the Evaluation of an Airport Terminal: SLAM (Simple Landside Aggregate Model)," Journal of Air Transport Management, 5, 161 – 175.
6. Manataki, I.E., and Zografos, K.G. 2009, "A Generic System Dynamic Based Tool for Airport Terminal Performance Analysis," Transportation Research Part C: Emerging Technologies, 17 (4), 428 – 443.
7. Humphreys, I. and Francis, G., 2002, "Performance Measurement: A Review of Airports," International Journal of Transport Management. 1 (2), 79 – 85.
8. Dorton, S.L., 2011. "Analysis of Airport Security Screening Checkpoints Using Queuing Networks and Discrete Event Simulation: A Theoretical and Empirical Approach," Unpublished M.S. Thesis, Embry Riddle Aeronautical University, Florida.
9. Lovell, D.J., Vlachou, K., Rabbani, T. and Bayen, A., 2013. "A Diffusion Approximation to a Single Airport Queue." Transportation Research Part C: Emerging Technologies, 33, 227-237.
10. Khadgi, P., 2009, "Simulation Analysis of Passenger Check-In and Baggage Screening Area at Chicago-Rockford International Airport," Northern Illinois University (NIU) Engineering Review, 1(1), 29-34.
11. Buzacott, J.A. and Shanthikumar, J.G., 1993, Stochastic Models of Manufacturing Systems, Prentice Hall, NJ.
12. Ross., S. M., 1997, Introduction to Probability Models, Academic Press, San Diego, CA, USA,
13. Marchal, W.G., 1985, "Numerical Performance of Approximate Queuing Formulae with Application to Flexible Manufacturing Systems," Annals of Operations Research. 3, 141 – 152
14. Marchal, W.G., 1976, "An Approximation Formula for Waiting Times in Single-Server Queues," AIIE Transactions, 8, 473.
15. Askin, R., and Goldberg, J., 2002, Design and Analysis of Lean Production Systems. Wiley, NY.
16. Papadopolous, H.T., Heavey, C., and Browne, J., 1993, Queueing Theory in Manufacturing Systems Analysis and Design, Springer, NY.

.