Queueing Network Approximations for Mass Dispensing and Vaccination Clinics

Ali Pilehvar, Jeffrey W. Herrmann

The
Institute for
Systems
Research

UNIVERSITY OF MARYLAND

A. JAMES CLARK
SCHOOL OF ENGINEERING

QUEUEING NETWORK APPROXIMATIONS FOR MASS DISPENSING AND VACCINATION CLINICS

Ali Pilehvar
Robert H. Smith School of Business
University of Maryland
College Park, MD 20742

Jeffrey W. Herrmann
Department of Mechanical Engineering
and Institute for Systems Research
University of Maryland
College Park, MD 20742
Phone: (301) 405-5433
jwh2@umd.edu

ABSTRACT

To respond to bioterrorism events or to curb outbreaks of contagious diseases, county health departments must set up and operate clinics to dispense medications or vaccines. Planning these clinics in advance of such an event requires determining clinic capacity and estimating the queueing that may occur in such facilities. We construct a queueing network model for mass dispensing and vaccination clinics and estimate the time that residents will spend at each workstation in such facilities. A key contribution is the development of useful approximations for queueing systems that have batch arrival, multiple-server batch processes and self-service stations. We compared the model's estimates to those from simulation experiments of realistic clinics using data collected from emergency preparedness exercises. Although this research was motivated by this specific application, the model should be applicable to the design and analysis of other similar queueing networks, including manufacturing systems with batch processes.

1   INTRODUCTION

This paper presents a model for estimating the performance of a queueing network. This work is motivated by research into methods for planning mass dispensing and vaccination clinics. However, because queueing networks are found in many different settings, the model presented here should be applicable to design and analysis problems in other domains as well.

The threat of an outbreak of contagious disease in the United States, caused by a terrorist act or a natural occurrence, has prompted public health departments to update and enhance their plans for responding to such events. Especially in regions that are densely populated or strategically important, such as the nation's capital, public health officials must plan for potential disasters. In the worst-case scenario, terrorists could release a lethal virus, such as smallpox, into the general population. Although different responses are available, mass vaccination should be an effective policy. Kaplan et al. (2002) compare vaccination policies for responding to a smallpox attack and show that mass vaccination results in many fewer deaths than other tactics in the most likely attack scenarios. The spread of a pandemic flu could also trigger mass vaccinations.

In the case of smallpox, every person in the affected area would have to be vaccinated within a few days. For example, Montgomery County, Maryland, would need to vaccinate nearly one million people. To vaccinate so many people in a short period, officials would have to set up mass dispensing and vaccination clinics, also known as points of dispensing (PODs). Counties across the United States are creating plans for this type of response.

2

In case of an emergency, county residents will visit clinics to receive treatment. The buildings housing these clinics will be schools, recreation centers, concert halls, and other facilities that can handle a large number of people. Clinics are not located in medical facilities because those facilities will be extremely busy during an event. There are various alternatives for transporting residents to clinics. In some plans, residents will gather at staging areas and then travel on buses to the clinics. In other plans, residents will walk to the closest clinic. For example, in the case of anthrax, a county may setup clinics at every elementary school. Mass vaccination would require every resident to visit a clinic. In other cases, such as the rapid delivery of antibiotics for anthrax, each family needs to send only one representative to obtain medication for the entire family.

Figure 1 shows the patient flow in a smallpox vaccination clinic design that is based upon federal guidelines. Each box in Figure 1 represents a station where residents receive service. The arrows show the movement of residents from one station to another. Note that not all residents follow the same path through the clinic. This is only one example of the type of clinic that our model is designed to evaluate. As described below, the model is very general and can be used for a wide range of clinic designs, which is essential because each county has its own plans for setting up and operating different types of clinics for different scenarios.
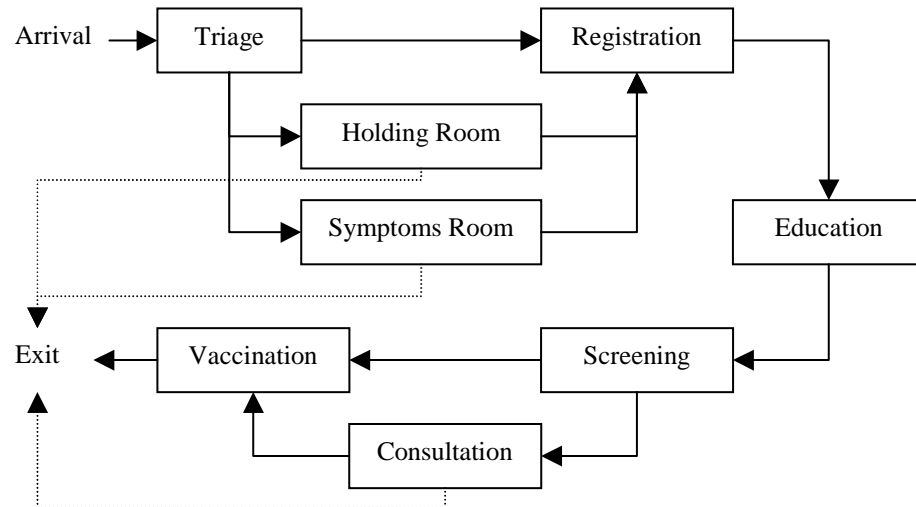

Figure 1. Flowchart of resident flow (dashed lines show residents who exit without receiving vaccinations).

Based on our experience with clinic planning, we have identified different types of stations that might be setup in a clinic. Many workstations will have multiple, parallel servers that treat one resident at a time. For example, a vaccination workstation may have a dozen nurses, and each nurse vaccinates one resident at a time. However, some

workstations will have batch processes that serve multiple residents simultaneously as a group. Moreover, there may be multiple servers so that multiple batches can be processed in parallel. For instance, at an education station, residents sit in classrooms in which they watch an informational video about the smallpox vaccine (under the direction of a staff member). Because there are multiple classrooms, different groups begin and end the process at different times.

There will also be self-service stations where residents complete paperwork (typically, medical history questionnaires) on their own. Staff may be present to answer questions, but they are not the critical resource, and modeling the process by which residents ask for and receive assistance is not essential to estimate clinic performance. One could also model the time that residents spend walking from one station to another as a self-service station.

Models of clinics are useful during the planning process. Two key clinic performance measures are the clinic capacity and the average time that a customer spends in the clinic (from arrival to departure), which we call cycle time (also known as flow time or throughput time). Clinic capacity is important for verifying that the clinic can treat the affected population in the required time. Estimating cycle time is necessary to determine how much space to allow in the clinic for queues. From the clinic planning perspective, reducing queueing is important to reduce the number of residents in the clinic, since large numbers of people increase crowding, confusion, and the chance of chaos.

In collaboration with the Montgomery County, Maryland Department of Health and Human Services (DHHS), we first developed simulation models to model resident flow through such clinics (Aaby *et al*., 2005). Then, we created software for constructing customizable spreadsheet models based on queueing network approximations, which provide a similar functionality for users without access to sophisticated simulation packages (Aaby *et al*., 2006a, b).

The fundamental problem is to evaluate the capacity and congestion of a clinic, given information about the arrival of residents to the clinic, the flow of residents through the clinic, and the processing at each workstation in the clinic. Our goal is to develop a set of approximations that require a modest amount of information about the clinic design and that can be implemented easily using common spreadsheet software. For more about the role of spreadsheets in emergency preparedness planning, see Herrmann (2008).

Good introductions to queueing theory include Gross and Harris (1974) and Hall (1991). Hopp and Spearman (2001) and Buzacott and Shanthikumar (1993) cover relevant manufacturing applications. While the study of queueing networks has resulted in numerous results, the need to develop approximations that can estimate the performance of a wide variety of clinics, including those with batch processes an self-service stations, let us to combine existing

approximations with new models for types of stations that are less common but still needed for this application (and may occur in other domains as well).

General purpose queueing software packages are available. Two examples are Rapid Analysis of Queueing Systems (RAQS), a Windows application developed at Oklahoma State's Center for Computer Integrated Manufacturing (Kamath et al, 1995), and Queueing Theory Software Plus (QTS-Plus), an Excel-based package authored by James Thompson, Carl Harris, and Donald Gross. RAQS uses the parametric decomposition approach to solving queueing networks (Segal and Whitt, 1989 and Whitt, 1983), while QTS-Plus is based around equations from Gross and Harris (1974). Unfortunately these packages have limited usefulness to the clinic planning problem for two reasons: the assumptions regarding batch processes are too restrictive for modeling clinics, and public health emergency preparedness planners do not have the necessary background to use them without substantial training.

Certain types of stations, motivated by clinic designs, make approximating the queueing network an interesting problem. We were unable to find previously proposed models that apply to these situations, including batch processes performed by multiple parallel servers and batch arrivals with batch size variability and a batch service process whose batch size is bigger than the arrival batches. Moreover, although there has been extensive research on queueing systems with an infinite number of servers (to represent a self service station), we are not aware of any work on equations that can be used to estimate interdeparture time variability .

The queueing network model presented here builds on existing models and includes novel contributions that were developed from analysis, intuition, and experimental results. Treadwell (2006) presented an early version of this model. The results in this paper are largely based on work that is described in detail in Pilehvar (2007). As described later in this paper, we compared the model's estimates to those from simulation experiments of realistic clinics using data collected from emergency preparedness exercises.

The key contributions of this model are the new approximations for workstations with batch arrivals having batch size variability, for workstations with batch service processes and multiple parallel servers, and for self service workstations. Another important contribution is the integration of a variety of approximations into a complete queueing network model.

2 MODEL

Due to the nature of these facilities, we model a clinic as an open queueing network. In the clinic queueing model, county residents are the customers, and the servers are the clinic staff, who are the critical resource. Residents arrive according to an external (not necessarily Poisson) arrival process. When buses are used to transport residents to a clinic, arrivals will be batches of residents. Motivated by the setup of typical clinics, we assume that there is no re-entrant flow.

We decompose the queueing network by estimating the performance of each workstation using a combination of exact and approximate models. We characterize each workstation as a *G/G/m* queueing system, but some workstations have batch arrivals, some have batch processing, and some are self-service stations, which can be modeled as *G/G/∞* queueing systems.

We use "i" throughout to denote a station, with 0 referring to the arrival process, 1 through "I" referring to the stations in the clinic, and "I+1" referring to the exit. The abbreviation "SCV" refers to the squared coefficient of variation. The SCV of a random variable equals its variance divided by the square of its mean.

Inputs

P = Number of residents to be treated at the clinic (residents)

H = Length of time interval that clinic will be providing treatment (hours)

m_i = Number of staff at station i

k_i = Processing batch size at station i

t_i = Mean process time at station i (minutes) for processing k_i entities

σ_i^2 = Process time variance at station i (minutes^2)

d_{ij} = Distance from station i to station j (feet)

v = Average walking speed (feet per second)

p_{ij} = Routing probability from station i to station j

k_0 = Initial arrival batch size

c_{B01}^2 = Batch interarrival time SCV at station 1

$\overline{C}_{B01}^2$ = SCV of the batch size of batches arriving to station 1

Calculated quantities

S_i = Set of stations that send residents to station i

r_i = Arrival rate at station i (residents per minute)

λ_{Bji} = Batch flow rate from station j to station i (batches per minute)

λ_{Ai} = Batch arrival rate at station i (batches per minute)

$\overline{K}_{Bji}$ = Average batch size of batches that come to station i from station j

$\overline{C}^2_{Bji}$ = SCV of the batch size of batches that come to station i from station j

$\overline{K}_{Ai}$ = Average batch size of all batches that come to station i

$\overline{C}^2_{Ai}$ = SCV of the batch size of all batches that come to station i

c^2_{ai} = Aggregate batch interarrival time SCV at station i

c^2_{bi} = interarrival time SCV for process batches at station i (after being formed)

c^2_{ei} = Process time SCV at station i for k_i entities

c^2_{di} = Interdeparture time SCV at station i for process batches

c^2_{Bji} = Interarrival times SCV for batches that come to station i from station j

m_i' = Minimum number of staff at station i to meet required throughput.

u_i = Utilization at station i

$WTBT_i$ = Wait time to form a batch size of k_i at station i (minutes)

$WIBT_i$ = Wait in batch time at station i (minutes)

CTq_i = Average queue time at station i (minutes)

W_i = Average time spent traveling to the next station after station i (minutes)

Outputs

TH' = Required throughput (residents per minute)

CT_i = Cycle time at station i (minutes)

TCT = Total cycle time in clinic (minutes)

WIP = Average number of residents in clinic

R = Clinic capacity (residents per minute)

The throughput required to treat the population in the given time is $TH' = \dfrac{P}{60H}$.

A key concept in the queueing network model is the flow of batches from one workstation to another. An external arrival process and the departure of process batches from workstations may create move batches. The flow of batches from one workstation to another is characterized by the following: the rate at which batches flow, the variability of that flow (specifically, the SCV of the inter-arrival times), the mean batch size, and the SCV of the batch size.

S_i is the set of stations that send residents to station i: $S_i = \{ j : j < i, P_{ji} > 0 \}$.

For the first station, $S_1 = \{0\}$, representing the source from residents arrive. We assume that all arriving customers go to the first station. Therefore, $P_{01} = 1$, while $P_{0i} = 0$ and $\lambda_{B0i} = 0$ for all $i > 1$.

$$r_1 = TH'$$

$$\lambda_{A1} = \lambda_{B01} = TH' / k_0$$

$$\bar{K}_{A1} = k_0$$

$$\bar{C}^2_{A1} = \bar{C}^2_{B01}$$

We calculate the arrival rates for the other stations based on the routing probabilities:

$$r_i = \sum_{j=1}^{i-1} \lambda_{Bji} \bar{K}_{Bji} = \sum_{j=1}^{i-1} r_j P_{ji} \qquad (1)$$

$$\lambda_{Ai} = \sum_{j \in S_i} \lambda_{Bji} \qquad (2)$$

8

$$\overline{K}_{Ai} = \frac{1}{\lambda_{Ai}} \sum_{j \in S_i} \lambda_{Bji} \overline{K}_{Bji} \qquad (3)$$

Following Whitt (1983) and Fowler et al. (2002), we estimate the SCV of the batch interarrival times at each station as follows:

$$c_{ai}^2 = 1 - w_i + \frac{w_i}{\lambda_{Ai}} \sum_{j \in S_i} \lambda_{Bji} c_{Bji}^2$$

$$w_i = \frac{1}{1 + 4(1 - u_i)^2 (v_i - 1)} \qquad (4)$$

$$v_i = \left(\sum_{j \in S_i} \left(\frac{\lambda_{Bji}}{\lambda_{Ai}} \right)^2 \right)^{-1} = \frac{\lambda_{Ai}^2}{\sum_{j \in S_i} \lambda_{Bji}^2}$$

We use station arrival rates to determine the minimum staff at each station: $m_i' = \dfrac{r_i \cdot t_i}{k_i}$

We then use user-selected staff levels m_i to calculate station utilization: $u_i = \dfrac{r_i \cdot t_i}{m_i \cdot k_i}$.

The process SCV at each workstation can determined immediately: $c_{ei}^2 = \dfrac{\sigma_i^2}{t_i^2}$

The average time spent traveling to the next station after station i depends upon the routing probabilities and the average walking speed:

$$W_i = \frac{1}{60v} \sum_{n=i+1}^{I+1} P_{in} d_{in}$$

To present the remainder of the model, we will discuss six cases that are distinguished by the arrival process and the service type (individual processing, batch processing or self service). For more information about these cases, see Pilehvar (2007).

    1. Individual arrivals, individual processing

    2. Individual arrivals, batch processing

    3. Individual arrivals, self-service

    4. Mixed arrivals, individual processing

    5. Mixed arrivals, batch processing

9

6. Mixed arrivals, self-service

2.1 Individual arrivals, individual processing

In this case, residents arrive individually to the workstation. The workstation has multiple, parallel servers that serve residents individually. Thus, the workstation can be modeled as a *G/G/m* queueing system and we can use well-known results for this case.

The arrival rate r_i and arrival variability c_{ai}^2 can be determined as discussed above. In this case, $\overline{K}_{Bji} = 1$ and $C_{Bji}^2 = 0$ for all the upstream workstations *j* that send residents to workstation *i*. Moreover, it's obvious that $\lambda_{Ai} = r_i$.

The following approximation estimates the time that residents spend waiting for service:

$$CT_{qi} = \left(\frac{c_{ai}^2 + c_{ei}^2}{2} \right) \left(\frac{u_i^{\sqrt{2m_i+2}-1}}{m_i\left(1-u_i\right)} t_i \right) \cdot g_i \tag{5}$$

Where the parameter g_i, suggested by Whitt (1984) and Bitran et al. (1989), equals 1 if the batch interarrival time variability $c_{bi}^2 \geq 1$. However, if $c_{bi}^2 < 1$, then

$$g_i = e^{-2\left(1-u_i\right)\left(1-c_{ai}^2\right)^2 / \left(3u_i\left(c_{ai}^2+c_{ei}^2\right)\right)}$$

As discussed in Pilehvar (2007), when the utilization is higher than 90%, the waiting time for service can be estimated by following term:

$$CT_{qi} = \frac{c_{ai}^2 / \lambda_{Ai} + c_{ei}^2 t_i^2 \lambda_{Ai} / m_i}{2(1-u_i)} \tag{6}$$

The cycle time at station *i* is $CT_i = CT_{qi} + t_i + W_i$.

For the interdeparture variability, we use the interdeparture time SCV estimate from Hopp and Spearman (2001) and adopt results from Whitt (1983a). The batch flow from workstation *i* to a downstream workstation *n* is characterized as follows:

$$\lambda_{Bin} = r_i P_{in} \tag{7}$$

$$c_{di}^2 = 1 + \left(1 - u_i^2\right)\left(c_{ai}^2 - 1\right) + \frac{u_i^2}{\sqrt{m_i}}\left(Max(0.2, c_{ei}^2) - 1\right) \tag{8}$$

$$c_{Bin}^2 = P_{in}c_{di}^2 + \left(1 - P_{in}\right)P_{in}c_{ai}^2 + \frac{\left(1 - P_{in}\right)^2}{\lambda_{Ai}}\sum_{j=1}^{i-1}\lambda_{Bji}c_{Bji}^2 \tag{9}$$

$$\overline{K}_{Bin} = 1$$
$$\overline{C}_{Bin}^2 = 0$$

2.2 Individual arrivals, batch processing

In this case, residents arrive individually to the workstation. The workstation has multiple, parallel servers. Each server processes a group of residents simultaneously (thus, it is a parallel process batch). We assume that the server process only full batches. The most common example in this domain is an education station in a mass smallpox vaccination clinic. Each server is a staff person running video equipment in a classroom where residents watch a video about the smallpox vaccine.

In this case, arriving residents form process batches of a fixed size, and then each batch waits for a server to process it. After processing, the batch leaves the workstation.

As before, the arrival rate r_i and interarrival variability c_{ai}^2 can be determined as discussed above. In this case, $\overline{K}_{Bji} = 1$ and $\overline{C}_{Bji}^2 = 0$ for all upstream workstations j that send residents to workstation i.

$WTBT_i$ is the average time that residents spend waiting for to form a process batch:

$$WTBT_i = \frac{k_i - 1}{2r_i} \tag{10}$$

The "arrival" of process batches (after they are formed) has less variabilty than the arrival of individual residents due to variability pooling.

As shown in Pilehvar (2007), experimental results show that the best formula to calculate c_{bi}^2 in which $\overline{K}_{Bji} = 1$, is:

$$c_{bi}^2 = \frac{c_{ai}^2}{k_i} \tag{11}$$

11

Once batches are formed, the queueing system is essentially a *G/G/m* system, so we use the same approximations as before. The following approximation estimates the time that process batches spend waiting for service:

$$CT_{qi} = \left(\frac{c_{bi}^2 + c_{ei}^2}{2} \right) \left(\frac{u_i^{\sqrt{2m_i+2}-1} t_i}{m_i (1-u_i)} \right) \cdot g_i \tag{12}$$

Where the parameter g_i, suggested by Whitt (1984) and Bitran et al. (1989), equals 1 if the batch interarrival time variability $c_{bi}^2 \geq 1$. However, if $c_{bi}^2 < 1$, then

$$g_i = e^{-2(1-u_i)(1-c_{bi}^2)^2 / (3u_i(c_{bi}^2 + c_{ei}^2))}$$

Similar to individual arrival-individual service process case, for the cases with utilization higher than 90% , waiting time for service can be yielded by the following equation (In this type of queueing system the arrival batch rate to the station after being formed is $\dfrac{\lambda_{Ai}}{k_i}$):

$$CT_{qi} = \frac{k_i c_{bi}^2 / \lambda_{Ai} + c_{ei}^2 t_i^2 \lambda_{Ai} / (k_i m_i)}{2(1-u_i)} \tag{13}$$

The cycle time at station *i* includes the wait-to-batch time, the queue time, the process time, and the walking time:

$$CT_i = WTBT_i + CT_{qi} + t_i + W_i .$$

Residents leave this type of station in batches. The move batch size varies due to the stochastic routing. Not all of the residents in a particular process batch go to the same station. As shown in Pilehvar (2007), the batch flow from workstation *i* to a downstream workstation *n* is characterized as follows:

$$\lambda_{Bin} = \frac{r_i}{k_i} \left(1 - (1 - P_{in})^{k_i} \right) \tag{14}$$

$$c_{di}^2 = 1 + (1 - u_i^2)(c_{bi}^2 - 1) + \frac{u_i^2}{\sqrt{m_i}} \left(Max(0.2, c_{ei}^2) - 1 \right) \tag{15}$$

$$c_{Bin}^2 = (1 - (1 - P_{in})^{k_i}) c_{di}^2 + (1 - P_{in})^{k_i} \tag{16}$$

$$\bar{K}_{Bin} = \frac{k_i P_{in}}{1 - (1 - P_{in})^{k_i}} \tag{17}$$

12

$$\overline{C}^2_{Bin} = \frac{1 - P_{in} - \left(k_i P_{in} + 1 - P_{in}\right)\left(1 - P_{in}\right)^{k_i}}{k_i P_{in}} \tag{18}$$

The last two equations follow from the definition of a binomial distribution that is conditional upon a non-zero batch size.

2.3 Individual arrivals, self-service in which $c_{ei}^2 \leq 1$

In this case, residents arrive individually to the workstation. The residents perform the process themselves without any external resources (servers). In this domain, an example would be a workstation where each resident must complete a form. Thus, the workstation can be modeled as a $G/G/\infty$ queueing system.

The only point is that to calculate c_{ai}^2 for self service station, w_i in Equation 4 should be 1. Moreover, $\overline{K}_{Bji} = 1$ and $C_{Bji}^2 = 0$ for all the upstream workstations j that send residents to workstation i. The cycle time at station i is

$$CT_i = t_i + W_i.$$

To estimate the interdeparture variability we first take into account the following facts. For a $G/D/\infty$ system, the interdeparture variability equals the interarrival variability because the departure process is simply the arrival process shifted by a constant equal to the processing time. For a $M/G/\infty$ system, the departure process is a Poisson process; thus the interdeparture variability equals 1. For a $G/G/\infty$ system, Whitt (1983) suggests that the interdeparture variability approaches 1 as the load (the arrival rate divided by the service rate) goes to infinity. On the other hand, if the load is near 0, the service rate is relatively fast, implying that customers spend very little time in the system. Thus, we would expect the interdeparture variability to equal the interarrival variability. These imply that, in the general case (a $G/G/\infty$ system with moderate load), the interdeparture variability will be somewhere between the interarrival variability and one.

Therefore, we conducted experiments to characterize this relationship and to examine various weights for interpolating between the interarrival variability and one. Pilehvar and Herrmann (2008) found that the interdeparture variability should be a function of load, process time SCV and interarrival time variability. Therefore, from all of our experiments and analyses, we decided to use the following approximation:

$$\rho_i = r_i t_i \tag{19}$$

13

$$c_{di}^2 = \left(\frac{1 + c_{ai}^2}{1 + c_{ai}^2 + \sqrt{c_{ei}^2}\, \rho_i^{\sqrt{2}}} \right) c_{ai}^2 + \frac{\sqrt{c_{ei}^2}\, \rho_i^{\sqrt{2}}}{1 + c_{ai}^2 + \sqrt{c_{ei}^2}\, \rho_i^{\sqrt{2}}} \tag{20}$$

We have to mention to this fact that our final extracted Equation 20 can be a good estimate to calculate c_{di}^2 for any

self-service stations ($G/G/\infty$) in which $c_{ei}^2 \le 1$.

We should point out that that since we have individual departure from self service stations; it behaves as if we had individual service process. From this, the batch flow from workstation i to a downstream workstation n is characterized as follows (the first equation was given above as Equation 7):

$$\lambda_{Bin} = r_i P_{in}$$

$$c_{Bin}^2 = P_{in} c_{di}^2 + 1 - P_{in} \tag{21}$$

$$\bar{K}_{Bin} = 1$$

$$C_{Bin}^2 = 0$$

2.4 Mixed arrivals, individual processing

This case has a more general arrival process. Residents arrive to the workstation in batches and individually. The arrival batches may come from different batch processing workstations, and the batch sizes from each workstation may vary due to the routing probabilities. There are also individual arrivals from individual processing workstations. The workstation has multiple, parallel servers that serve residents individually. To analyze this case we model all of the arrivals as batches. Each batch must wait to get to the head of the queue, at which point it "opens" and at least one of the residents in the batch begins service. The other residents must wait in the batch for a server.

We calculate the variability of the arriving batch size by adapting a formula from Fowler et al. (2002), who calculate the SCV of a process time for different products that arrive at different rates. If the different products represent batches from different stations and we assume that the service time per resident is a constant, then the process time SCV is exactly the SCV of the arriving batches:

$$\bar{C}_{Ai}^2 = -1 + \frac{1}{\lambda_{Ai}\bar{K}_{Ai}^2} \sum_{j \in S_i} \lambda_{Bji} \left(1 + \bar{C}_{Bji}^2\right) \bar{K}_{Bji}^2 \tag{22}$$

14

The arrival rate r_i and arrival variability c_{ai}^2 can be determined as discussed above. To estimate the time that batches spend in the queue, we model the workstation as a $G^{[X]}/G/1$ system by combining the multiple parallel servers into one fast server that can process residents with a modified process time distribution that has a mean of T_i (calculated below) and a SCV of $\overline{C}_{Ai}^2 + c_{ei}^2 / \overline{K}_{Ai}$ from Buzacott and Shanthikumar (1993).

To estimate the time that batches spend in the queue, we model the workstation as a $G/G/1$ system by combining the multiple parallel servers into one fast server that can process residents with a modified process time distribution that has a mean of T_i (calculated below) and a SCV of $\overline{C}_{Ai} + \dfrac{c_{ei}^2}{\overline{K}_{Ai}}$.

$$T_i = \frac{\overline{K}_{Ai} t_i}{m_i} u_i^{(1-\frac{1}{m_i})} \tag{23}$$

So as we had in the first case for individual arrival and individual processing, from Equation 5 when we have only a single server:

$$CT_{qi} = \frac{1}{2}\left(c_{ai}^2 + \overline{C}_{Ai}^2 + \frac{c_{ei}^2}{\overline{K}_{Ai}} \right)\left(\frac{u_i}{1-u_i} \right) T_i g_i \tag{24}$$

The queue time estimate, as suggested by Whitt (1984) and Bitran et al. (1989), includes the parameter g_i, which equals 1 if the interarrival time SCV variability $c_{ai}^2 \geq 1$. Otherwise,

$$g_i = e^{-2(1-u_i)\left(1-c_{bi}^2\right)^2 / \left(3u_i\left(c_{ai}^2 + \overline{C}_{Ai} + c_{ei}^2/\overline{K}_{Ai}\right)\right)}$$

Similar to other cases, for the individual process stations with mixed arrival and the utilization higher than 90%, we can estimate the waiting time for batches spend in the queue by modeling the workstation as a $G/G/1$ system by combining the multiple parallel servers into one fast server that has a mean of T_i mentioned earlier in this section. In this way, this waiting time can be yielded by:

$$CT_{qi} = \frac{c_{ai}^2 / \lambda_{Ai} + (\overline{C}_{Ai}^2 + \dfrac{c_{ei}^2}{\overline{K}_{Ai}})T_i^2 \lambda_{Ai}}{2(1-u_i)} \tag{25}$$

15

When the batch reaches the front of the queue, it is opened. Some residents will go first, while others will wait. The average time spent waiting is the wait-in-batch-time.

The experiment results and analysis in Pilehvar (2007) showed that the following equations were good approximations for WIBT:

$$WIBT = \frac{(\bar{K}_{Ai} - 1)t_i}{2} \quad \text{if } m_i = 1 \tag{26}$$

$$WIBT = \frac{(\bar{K}_{Ai}^{2-u_i} - 1)t_i u_i}{(3 - u_i)m_i^{2-u_i}} \quad \text{if } m_i > 1 \tag{27}$$

The cycle time at station i is $CT_i = CT_{qi} + WIBT_i + t_i + W_i$. Using a interdeparture variability approximation from Curry and Deuermeyer (2002) for multiple servers, the batch flow from workstation i to a downstream workstation n is characterized as follows:

$$\lambda_{Bin} = r_i P_{in}$$

$$c_{di}^2 = \bar{K}_{Ai} c_{ai}^2 \left(1 - u_i^2\right) + (1 - u_i)^2 (\bar{K}_{Ai} - 1) + u_i^2 c_{ei}^2 \tag{28}$$

$$c_{Bin}^2 = P_{in} c_{di}^2 + 1 - P_{in}$$

$$\bar{K}_{Bin} = 1$$

$$C_{Bin}^2 = 0$$

Following figure demonstrates how we model a system of mixed arrivals and individual processing queueing network.
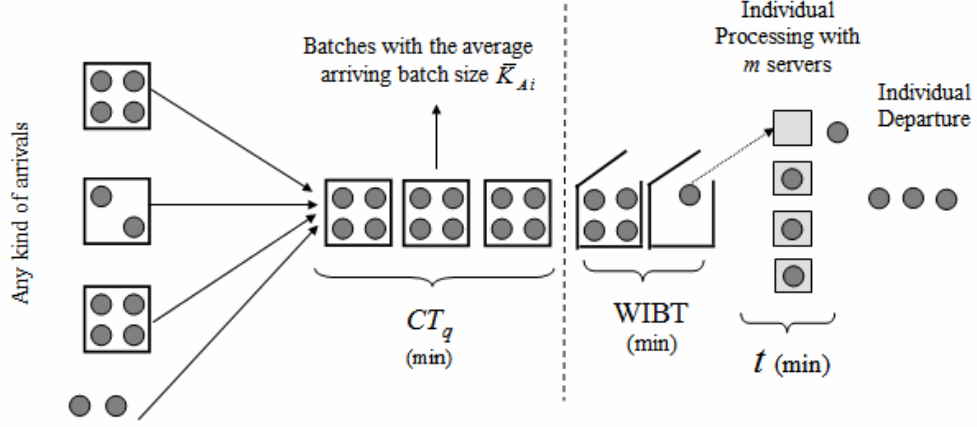
Figure 2. A system of mixed arrivals-individual processing

2.5 Mixed arrivals, batch processing

This case has the more general arrival process and process batches. Residents arrive to the workstation in batches and individually. The arrival batches may come from different batch processing workstations, and the batch sizes from each workstation may vary due to the routing probabilities. There are also individual arrivals from individual processing workstations. The workstation has multiple, parallel servers. Each server processes a group of residents simultaneously (thus, it is a parallel process batch).

As in the previous case, we model all of the arrivals as batches. Here, however, the arriving batches are combined into a process batch. Each arriving batch must wait to form the process batch. After it is formed, a process batch must wait to get to the head of the queue, at which point it begins service. We assume that the arriving batches are smaller than the process batch.

As before, we calculate the variability of the arriving batch size by adapting a formula from Fowler et al. (2002), who calculate the SCV of a process time that for different products that arrive at different rates. If the different products represent batches from different stations and we assume that the service time per resident is a constant, then the process time SCV is exactly the SCV of the arriving batches:

$$C_{Ai}^2 = \frac{1}{\lambda_{Ai} \bar{K}_{Ai}^2} \sum_{j \in S_i} \lambda_{Bji} \left(1 + C_{Bji}^2\right) \bar{k}_{ji}^2 - 1$$

Because the arrival batches are not the same as the process batches, residents must wait to form the process batches:

$$WTBT_i = \frac{k_i - 1}{2r_i} \tag{29}$$

17

For the interarrival time SCV of process batches c_{bi}^2 (after they are formed), we use the following result from Pilehvar (2007):

$$c_{bi}^2 = \frac{\overline{K}_{Ai}}{k_i} \left(\frac{\sum\limits_{j \in S_i} \overline{K}_{Bji}^2 \lambda_{Bji} c_{Bji}^2}{\sum\limits_{j \in S_i} \overline{K}_{Bji}^2 \lambda_{Bji}} + \overline{C}_{Ai}^2 \right) \tag{30}$$

The arrival rate r_i and $\overline{K}_{Ai}$ can be determined as discussed in output section. Then, we estimate the queueing in the resulting $G/G/m$ system:

$$CT_{qi} = \left(\frac{c_{bi}^2 + c_{ei}^2}{2} \right) \left(\frac{u_i^{\sqrt{2m_i+2}-1}}{m_i(1-u_i)} \right) t_i g_i$$

The queue time estimate, as suggested by Whitt (1984) and Bitran et al. (1989), includes the parameter g_i, which equals 1 if the interarrival time SCV $c_{bi}^2 \geq 1$. However, if $c_{bi}^2 < 1$, then

$$g_i = e^{-2(1-u_i)\left(1-c_{bi}^2\right)^2 / \left(3u_i\left(c_{bi}^2+c_{ei}^2\right)\right)}$$

Similar to individual arrival/batch service process, for the cases with the utilization higher than 90% from, waiting time for service can be yielded by the following equation (In this type of queueing system the arrival batch rate to the station after being formed is $\dfrac{\overline{K}_{Ai}\lambda_{Ai}}{k_i}$):

$$CT_{qi} = \frac{k_i c_{bi}^2 / (\overline{K}_{Ai}\lambda_{Ai}) + c_{ei}^2 t_i^2 \overline{K}_{Ai}\lambda_{Ai} / (k_i m_i)}{2(1-u_i)} \tag{31}$$

The cycle time at station i is $CT_i = WTBT_i + CT_{qi} + t_i + W_i$.

The batch flow from workstation i to a downstream workstation n is characterized as follows (some of which is similar to that in Section 2.2):

$$\lambda_{Bin} = \frac{r_i}{k_i}\left(1-\left(1-P_{in}\right)^{k_i}\right)$$

18

$$c_{di}^2 = 1 + \left(1 - u_i^2\right)\left(c_{bi}^2 - 1\right) + \frac{u_i^2}{\sqrt{m_i}}\left(Max(0.2, c_{ei}^2) - 1\right)$$

$$c_{Bin}^2 = (1 - (1 - P_{in})^{k_i})c_{di}^2 + (1 - P_{in})^{k_i} c_{bi}^2 \tag{32}$$

$$\overline{K}_{Bin} = \frac{k_i P_{in}}{1 - \left(1 - P_{in}\right)^{k_i}}$$

$$C_{Bin}^2 = \frac{1 - P_{in} - \left(k_i P_{in} + 1 - P_{in}\right)\left(1 - P_{in}\right)^{k_i}}{k_i P_{in}}$$

The following figure demonstrates how we model a system of mixed arrivals and batch processing queueing network where arriving batches are smaller than the process batch size.
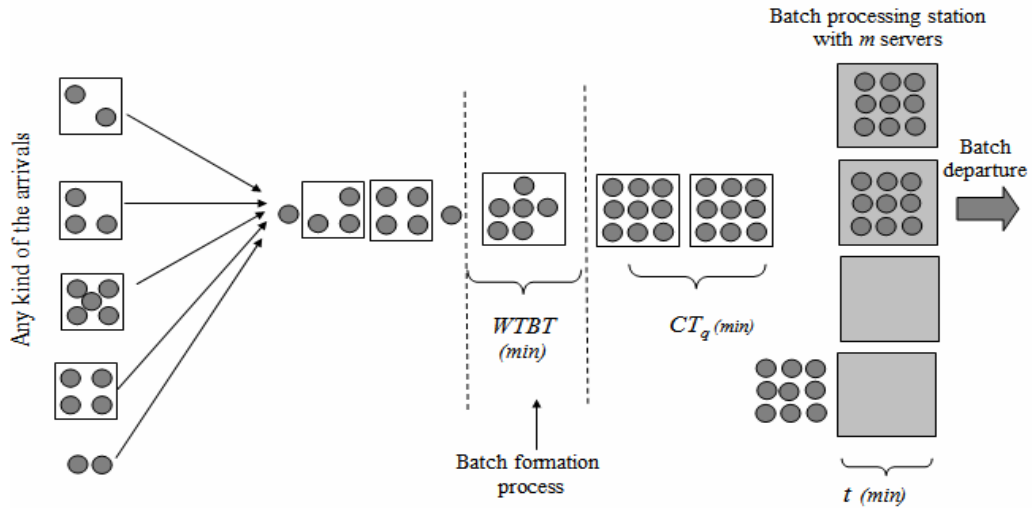

Figure 3. A system of mixed arrivals-batch processing

2.6 Mixed arrivals, self-service

In this case, residents arrive to the workstation in batches and individually. The residents perform the process themselves without any external resources. The cycle time at station i is $CT_i = t_i + W_i$. Moreover, to calculate c_{ai}^2 for self service station, w_i in Equation 4 should be 1.

To estimate the interdeparture variability we adapt the estimate used in the individual arrival, self-service case (Section 2.3). The key change is that the interarrival variability of individuals depends upon the batch size and the batch interarrival variability. From Curry (2002) we have the following:

$$\text{Interarrival variability of individuals } = \left(\overline{K}_{Ai} c_{ai}^2 + \overline{K}_{Ai} - 1 \right) \tag{33}$$

The only big assumption in the following formula is that, we assume the SCV of batch size arriving to self service station is so small, so it's ignorable in our calculation. That is why; we do not have any effect of C_{Ai}^2 in the formula.

$$\rho_i = r_i t_i$$

From Equation 33 and Equation 20, we can obtain the following:

$$c_{di}^2 = \left(\frac{\overline{K}_{Ai} c_{ai}^2 + \overline{K}_{Ai}}{\overline{K}_{Ai} c_{ai}^2 + \overline{K}_{Ai} + \sqrt{c_{ei}^2} \rho_i^{\sqrt{2}}} \right) \left(\overline{K}_{Ai} c_{ai}^2 + \overline{K}_{Ai} - 1 \right) + \frac{\sqrt{c_{ei}^2} \rho_i^{\sqrt{2}}}{\overline{K}_{Ai} c_{ai}^2 + \overline{K}_{Ai} + \sqrt{c_{ei}^2} \rho_i^{\sqrt{2}}} \tag{34}$$

We should point out that that since we have individual departure from self service stations; it behaves as if we had individual service process.

The batch flow from workstation i to a downstream workstation n is characterized as follows:

$$\lambda_{Bin} = r_i P_{in}$$

$$c_{Bin}^2 = P_{in} c_{di}^2 + 1 - P_{in}$$

$$\overline{K}_{Bin} = 1$$
$$C_{Bin}^2 = 0$$

2.7 Clinic Performance Measures

The clinic capacity is determined by bounds set by each station's capacity and the relative arrival rates:

$$R = \min_{i=1,\dots,I} \left\{ \frac{k_i m_i r_1}{t_i r_i} \right\} \tag{35}$$

Because of the stochastic routing, the clinic's total cycle time is a weighted sum of the station cycle times:

$$TCT = \frac{1}{r_1} \sum_{i=1}^{I} r_i CT_i \qquad (36)$$

The average number of residents in the clinic follows from Little's Law:

$$WIP = r_1 \cdot TCT \qquad (37)$$

3 EXPERIMENTAL RESULTS

In order to evaluate this queueing network approximation, we compared the model's results to the results from discrete-event simulation models created and run using Rockwell Software's Arena®. To validate our formulas for our clinic modeling (proposed six cases), we will carry out four different sets of experiments.

In all of these experiments, the first station in the clinic has a batch arrival process. The travel time between stations is negligible in these experiments.

In the first experiment, the clinic has no self service stations, and arrival batch size is fixed. The remaining three experiments have two tests each. In each test, the clinic has different combinations of stations and includes self service workstations, which helps us to evaluate clinics with the different types of workstations described previously. Additionally, these clinics have the arrival batch size variability.

For all of these experiments, we use a confidence interval of 95% for the simulation results.

3.1 Experiment One

We obtained data for this experiment from a time study of a mass smallpox vaccination clinic exercise on June 21, 2004, by the Montgomery County, Maryland, DHHS. From the exercise we collected data on the processing times at each workstation as well as measuring how long residents spent in the clinic. The exercise, which lasted a few hours, had hundreds of volunteers go through the clinic as residents. No residents received actual vaccinations or medications.

We constructed a simulation model and ran 10 replications of 2000 hours, with 5000 hours of warm-up time allowed to achieve steady state. Data was recorded for mean total time and mean queueing time at each node, as well as mean time in system and mean system WIP. The model was tested at several levels of resident arrival rates, from 20% to 97.5% of clinic capacity.

21

The arrival batches were always 50 residents. The batch interarrival times were exponentially distributed. Table 1 describes the eight stations in the clinic. Table 2 shows the routing probabilities from one station to another (only station numbers are shown to save space). Table 3 lists the capacity of each station and its bound on the clinic capacity. The vaccination station is the bottleneck station, and the clinic capacity is 5.123 residents per minute. Moreover, Table 4 shows the average total cycle time (in minutes) from the simulation results and the queueing network model in 21 scenarios in which the arrival rate increases to nearly the clinic capacity.

Table 1: Parameters for Experiment One.

Workstation	Number of Staff	Mean service Time (min.)	Service Time SCV	Processing time distribution (min.)	Batch processing size k_i
1. Triage	5	0.259	0.268	0.125+EXPO(0.134)	1
2. Symptoms Room	3	1.213	0.264	0.59 + EXPO(0.623)	1
3. Holding Room	3	3.800	1.000	EXPO(3.8)	1
4. Registration	8	0.122	0.630	0.025+EXPO(0.0995)	1
5. Education	8	24.000	0.111	18+EXPO(6)	30
6. Screening	9	1.724	0.261	0.999 + GAMM(1.07, 0.678)	1
7. Consultation	6	3.770	0.308	GAMM(1.16, 3.25)	1
8. Vaccination	16	3.260	0.124	1 + GAMM(0.581, 3.89)	1

Table 2. Routing Table for Experiment One.

				To				
From	2	3	4	5	6	7	8	Exit
1	0.048	0.032	0.921	0	0	0	0	0
2		0	0.67	0	0	0	0	0.33
3			0.65	0	0	0	0	0.35
4				1.00	0	0	0	0
5					1.00	0	0	0
6						0.262	0.738	0
7							0.941	0.059
8								1.00

Table 3. Capacity for first experiment

Workstation	Station capacity (residents/min)	Relative Throughput	Bound on clinic capacity (residents/min)
1. Triage	19.293	1.000	19.293
2. Symptoms Room	2.473	0.048	51.849
3. Holding Room	0.789	0.032	24.905
4. Registration	65.844	0.973	67.659
5. Education	10.000	0.973	10.276
6. Screening	5.219	0.973	5.363
7. Consultation	1.592	0.255	6.249
8. Vaccination	4.908	0.958	5.123

Table 4. Comparison of total cycle time for first experiment

Scenario	Arrival rate to the clinic (residents/min)	Total cycle time from simulation	Total cycle time from queueing network model	Percentage error %
1	5.00	253.23	126.17	50.18%
2	4.85	126.85	96.25	24.13%
3	4.75	99.06	86.23	12.96%
4	4.60	79.95	76.65	4.12%
5	4.50	69.87	72.24	3.41%
6	4.25	60.51	59.32	1.97%
7	4.17	58.62	57.46	1.99%
8	3.57	48.87	49.19	0.64%
9	3.13	45.69	46.05	0.80%
10	2.78	44.48	44.59	0.26%
11	2.63	44.14	44.17	0.07%
12	2.50	44.01	43.88	0.31%
13	2.27	43.66	43.56	0.22%
14	2.00	43.83	43.51	0.73%
15	1.85	44.05	43.64	0.92%
16	1.67	44.50	44.00	1.12%
17	1.52	45.03	44.49	1.19%
18	1.43	45.56	44.86	1.53%
19	1.25	46.72	45.91	1.74%
20	1.11	47.88	47.07	1.69%
21	1.00	49.25	48.30	1.94%

Figure 4 shows the total cycle time estimates from the queueing network model and the discrete event simulation for a variety of arrival rates. The plot for the simulation results includes error bars showing the 95% confidence interval on each estimate.

Due to the batching at the education station, total cycle time does not always decrease as the arrival rate decreases. At low arrival rates, there is significant waiting to form the education batches, which increases total cycle time.

When the arrival rate is low to moderate, there is a small difference between the estimates from the queueing network model and the discrete event simulation. As the arrival rate approaches the clinic capacity, the difference between the two estimates is large due almost entirely to different estimates for the cycle time at the vaccination station, which has the highest utilization (since it is the clinic bottleneck).
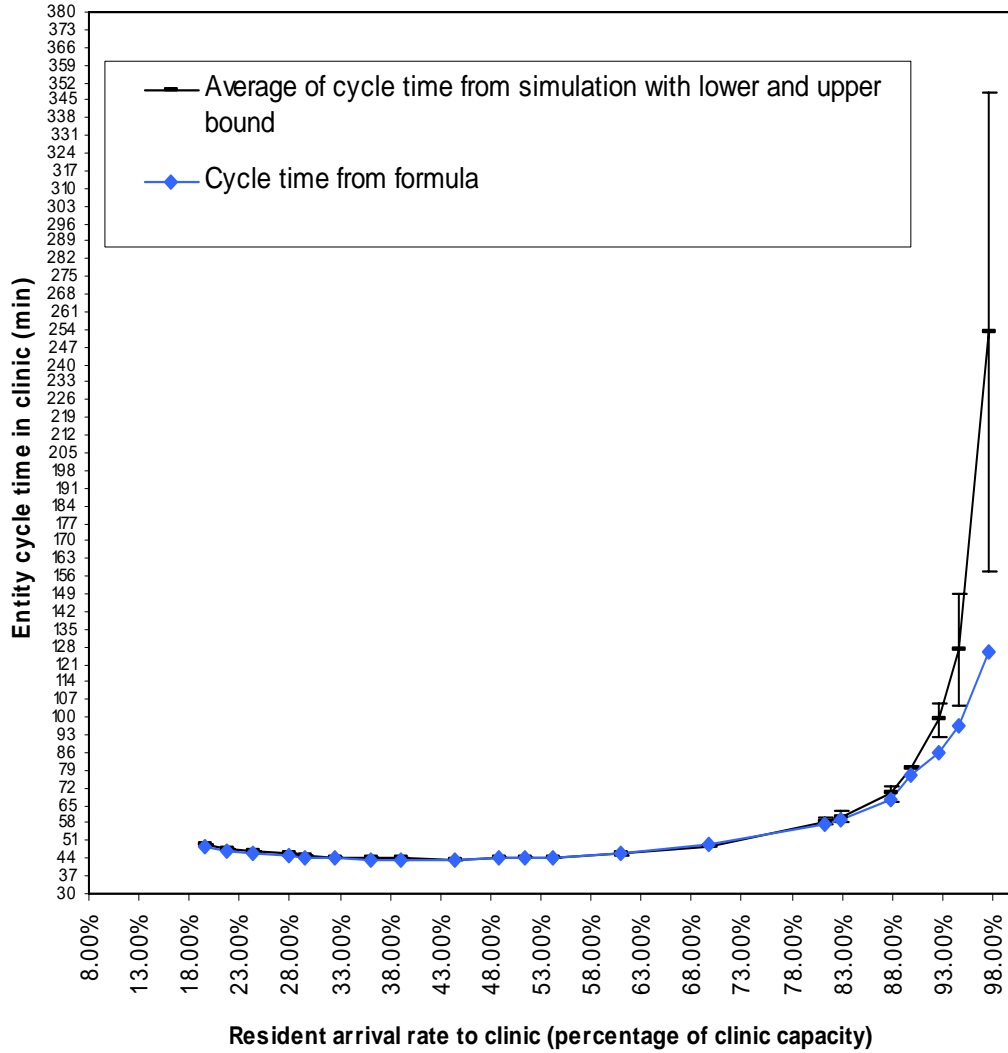

Figure 4. Comparison of average total cycle time for first experiment

3.2 Experiment Two

In this experiment, the clinic has seven workstations with different specifications. All of the workstations have individual service process and the fourth station is a self service station with individual arrival.

The model was tested at several levels of resident arrival rates, from 20% to 94.5% of clinic capacity under the different scenario. We ran 10 replications of 4000 hours, with 1000 hours of warm-up time allowed to achieve steady state for each scenario. Data was recorded for mean total time and mean queueing time at each node, as well as mean time in system and mean system WIP.

This experiment consisted of two tests: Test 2-1 and Test 2-2. In both tests, the average arrival batch size equals 20. The batch interarrival times were exponentially distributed, and all of the service process distributions had gamma distributions to allow different process time SCV. In Test 2-1, the SCV of the arrival batch size is 0.05; in Test 2-2, the SCV of the arrival batch size is 0.2.

Table 5 describes each of the seven stations in the clinic. Note that station 4 is a self-service station. Table 6 shows the routing probabilities from one station to another. Table 7 lists the capacity of each station and its bound on the clinic capacity.

In this experiment, station number 7 is the bottleneck station, and the clinic capacity is 10.704 residents per minute. We should say that in this experiment, for each test we have eight scenarios with different arrival rates.

Table 5. Parameters for clinic in Experiment Two. (Station 4 is a self-service station.)

Workstation Number	Staff	Mean service Time (min.)	Service time SCV	Processing time distribution (min.)	Batch processing size k_i
1	15	1	1.11	GAMM(0.9,1.11)	1
2	9	1.752	0.52	GAMM (1.91,0.92)	1
3	8	1.154	0.40	GAMM (2.5,0.46)	1
4	n.a.	6	0.56	GAMM (1.8,3.33)	1
5	5	2	1.00	GAMM (1,2)	1
6	7	1.5	0.44	GAMM (2.25,0.67)	1
7	9	2	0.50	GAMM (2,1)	1

Table 6. Routing table for the clinic in Experiment Two

	To						
From	2	3	4	5	6	7	Exit
1	0.200	0.300	0.500	0	0	0	0
2		0.400	0	0	0.600	0	0
3			0.700	0	0.000	0.300	0
4				0.250	0.350	0.400	0
5					0	0	1.00
6						0	1.00
7							1.00

Table 7. Table capacity for the clinic stations in Experiment Two

Workstation	Station capacity (residents/min)	Relative Throughput	Bound on clinic capacity (residents/min)
1	15	1.000	15
2	5.137	0.200	25.685
3	6.932	0.380	18.243
4	n.a.	0.766	n.a.
5	2.5	0.192	13.055
6	4.667	0.388	12.024
7	4.500	0.420	10.704

Results for Test 2-1

In Test 2-1, the arrival batch size distribution is 1+Poisson (19), and the arrival batch size SCV equals 0.05. Table 8 shows the average total cycle time (in minutes) from the simulation results (Test 2-1) and from our queueing network model. Moreover, Figure 5 shows the total cycle time estimates from the queueing network model and the discrete event simulation for a variety of arrival rates in Test 2-1. The plot for the simulation results includes error bars showing the 95% confidence interval on each estimate.

Table 8. Comparison of total cycle time for the clinic (Test 2-1)

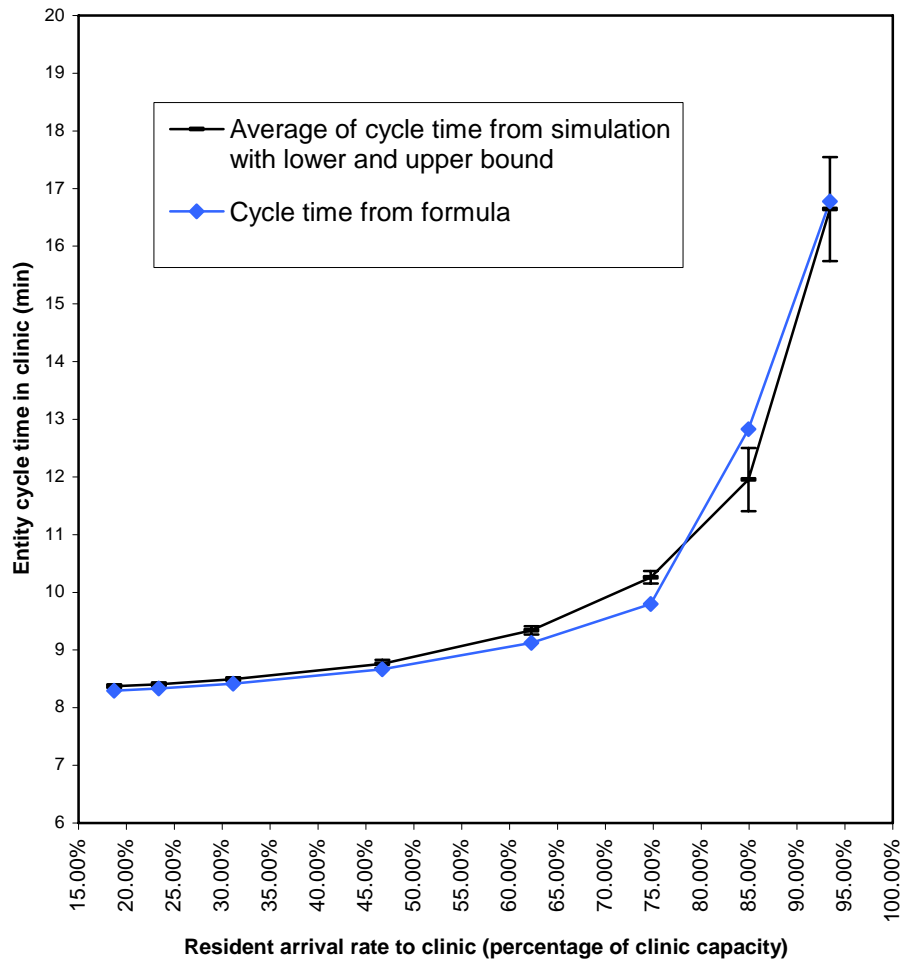| Scenario	Arrival rate to the clinic (residents/min)	Total cycle time from simulation	Total cycle time from queueing network model	Percentage error %
1	10.00	16.65	16.75	0.66%
2	9.09	11.96	12.82	7.23%
3	8.00	10.26	9.80	4.49%
4	6.67	9.34	9.13	2.30%
5	5.00	8.76	8.67	1.05%
6	3.33	8.50	8.42	0.93%
7	2.50	8.41	8.34	0.83%
8	2.00	8.37	8.30	0.90%

Figure 5. Comparison of total cycle time for the clinic (Test 2-1)

Results for Test 2-2

In Test 2-2, the arrival batch size distribution is 100-Poisson (80), and the arrival batch size SCV equals 0.2. Table 9 shows the average total cycle time (in minutes) from the simulation results and from our queueing network model. Figure 6 shows the total cycle time estimates from the queueing network model and the discrete event simulation for a variety of arrival rates in Test 2-2. The plot for the simulation results includes error bars showing the 95% confidence interval on each estimate.

Table 9. Comparison of total cycle time for the clinic (Test 2-2)

Scenario	Arrival rate to the clinic (residents/min)	Total cycle time from simulation	Total cycle time from queueing network model	Percentage error %
1	10.00	18.80	18.08	3.78%
2	9.09	12.72	12.92	1.55%
3	8.00	10.71	9.86	7.90%
4	6.67	9.60	9.16	4.55%
5	5.00	8.95	8.69	2.98%
6	3.33	8.62	8.42	2.26%
7	2.50	8.51	8.34	2.03%
8	2.00	8.47	8.30	2.02%

Cycle time from simulation (min) with lower and upper bound
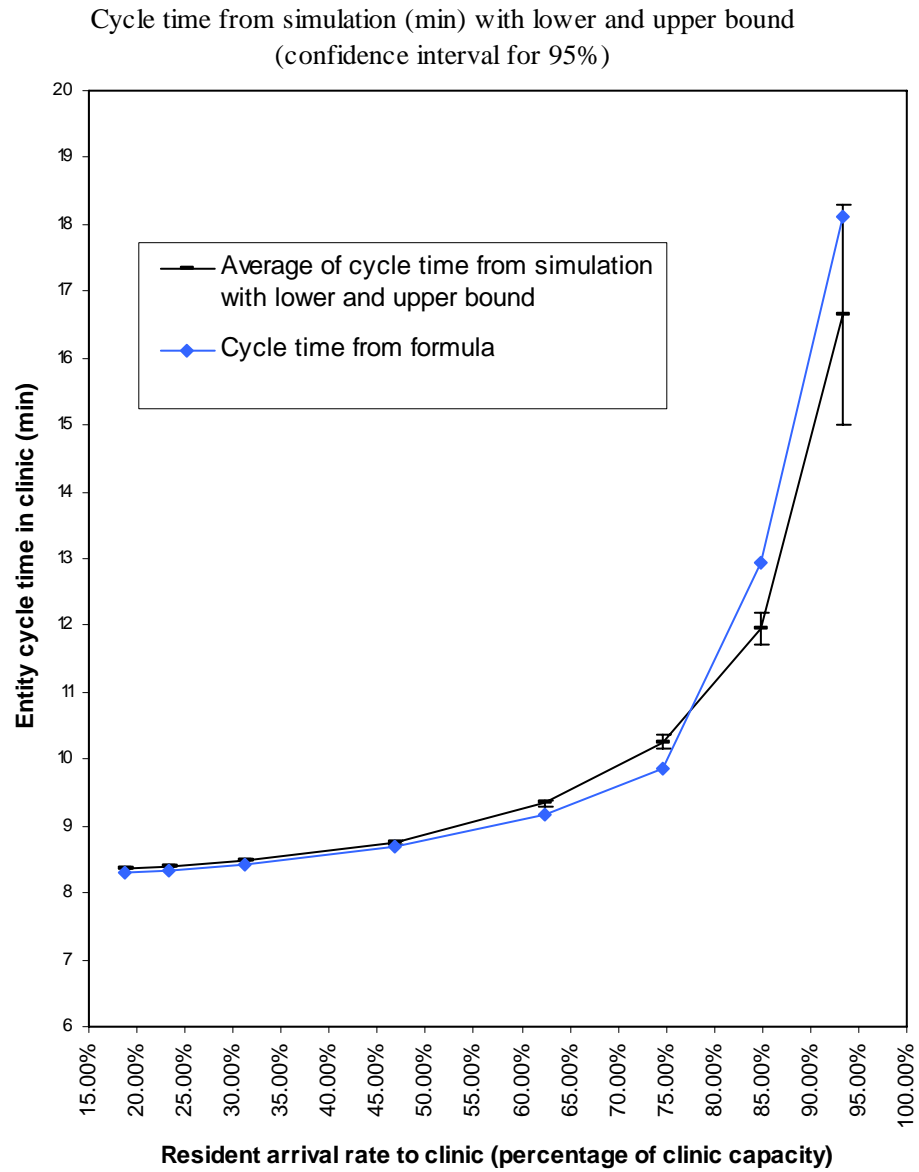(confidence interval for 95%)

Figure 6. Comparison of total cycle time for the clinic (Test 2-2)

In second test since the arrival rate is from low to moderate, there is a small difference between the estimates from the queueing network model and the discrete event simulation.

3.3 Experiment Three

In this experiment, the clinic has seven workstations with different specifications. The first three workstations have individual service processes. The fourth station is a self-service station. The last three stations are batch process workstations.

The model was tested at several levels of resident arrival rates, from 39% to 97.5% of clinic capacity under the different scenarios. We ran 10 replications of 4000 hours, with 1000 hours of warm-up time allowed to achieve steady state for each scenario. Data was recorded for mean total time and mean queueing time at each node, as well as mean time in system and mean system WIP.

This experiment consisted of two tests: Test 3-1 and Test 3-2. In both tests, the arrival batch size distribution is 1+Poisson (49), the average arrival batch size equals 50, and the arrival batch size SCV equals 0.02. The batch interarrival times were exponentially distributed, and all of the service process distributions had gamma distributions to allow different process time SCV. In Test 3-1, the process time SCV at station 4 (the self-service station) equals 0.56; in Test 3-2, this SCV equals 1.

The routing probabilities from one station to another in this experiment are the same as second test (See Table 6). Table 10 lists the capacity of each station and its bound on the clinic capacity which is similar for both Test 3-1 and 3-2. In this experiment, station six is the bottleneck station, and the clinic capacity is 10.736 residents per minute.

In this experiment, each test has eight scenarios with different arrival rates.

Table 10. Table capacity for the clinic stations in the Experiment Three.
(Station 4 is a self-service station.)

Workstation #	Station capacity (residents/min)	Relative Throughput	Bound on clinic capacity (residents/min)
1	12.126	1.000	12.126
2	2.854	0.200	14.269
3	6.066	0.380	15.963
4	n.a.	0.766	n.a.
5	12.500	0.192	65.274
6	4.167	0.388	10.736
7	5.000	0.420	11.893

Results for Test 3-1

In Test 3-1, the process time SCV of the self service station is 0.56. Table 11 describes each of the seven stations in the clinic for Test 3-1. Table 12 shows the average total cycle time (in minutes) from the simulation results and from our queueing network model. Figure 7 shows the total cycle time estimates from the queueing network model and the discrete event simulation for a variety of arrival rates in Test 3-1. The plot for the simulation results includes error bars showing the 95% confidence interval on each estimate.

Table 11. Parameters for clinic stations (Test 3-1)

Workstation	Staff	Mean service Time (min.)	Service time SCV	Processing time distribution (min)	Batch process size k_i
1	15	1.237	0.73	GAMM(1.38,0.9)	1
2	5	1.752	0.52	GAMM (1.91,0.92)	1
3	7	1.154	0.40	GAMM (2.5,0.46)	1
4	n.a.	6	0.56	GAMM (1.8,3.33)	1
5	1	4	0.25	GAMM (4.1)	50
6	5	24	0.01	GAMM (144,0.17)	20
7	4	24	0.01	GAMM (144,0.17)	30

Table 12. Comparison of total cycle time for the clinic (Test 3-1)

Scenario	Arrival rate to the clinic (residents/min)	Total cycle time from simulation	Total cycle time from queueing network model	Percentage error %
1	10.42	113.12	54.41	51.90%
2	10.00	61.85	47.79	22.73%
3	9.62	49.74	45.22	9.08%
4	9.09	44.30	41.99	5.20%
5	8.70	42.14	39.64	5.93%
6	7.04	39.07	37.91	2.97%
7	5.26	39.47	38.55	2.32%
8	4.17	41.25	40.27	2.37%

Cycle time from simulation (min) with lower and upper bound (confidence interval for 95%)
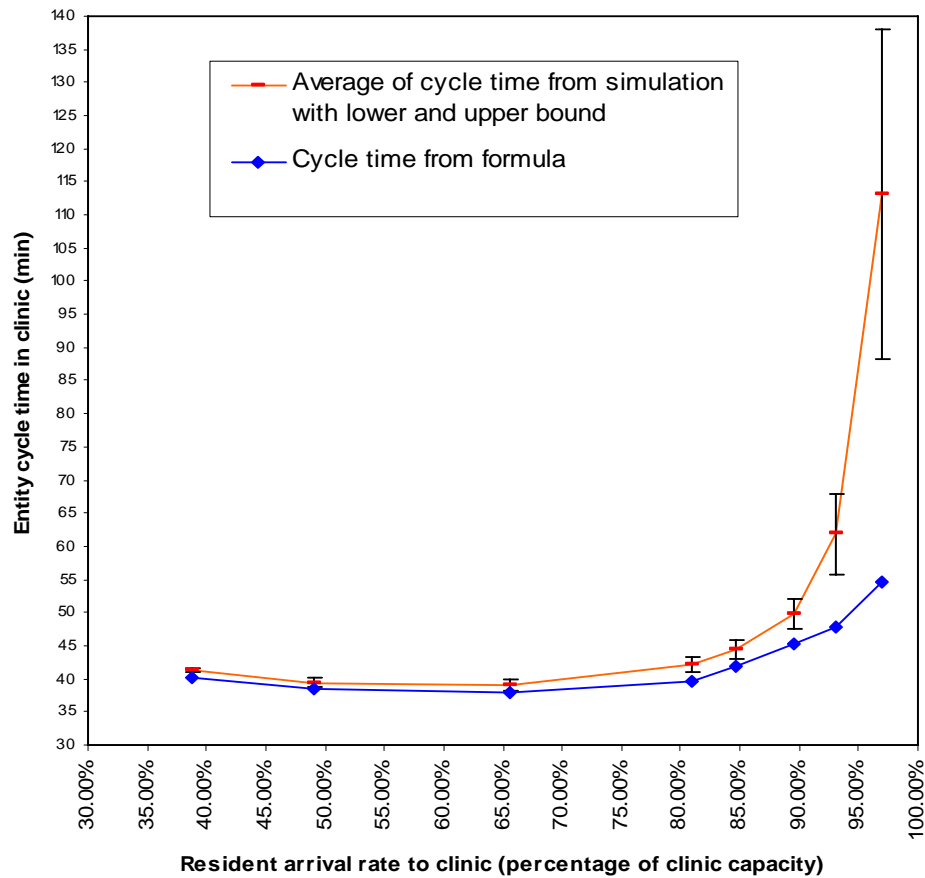
Figure 7. Comparison of total cycle time for the clinic (Test 3-1)

Results for Test 3-2

In Test 3-2, process time SCV of self service station is 1. Table 13 describes each of the seven stations in the clinic for Test 3-2. Table 14 shows the average total cycle time (in minutes) from the simulation results and from our queueing network model. Figure 8 shows the total cycle time estimates from the queueing network model and the discrete event simulation for a variety of arrival rates in Test 3-2. The plot for the simulation results includes error bars showing the 95% confidence interval on each estimate.

Table 13. Parameters for clinic stations (Test 3-2)

Workstation	Staff	Mean service Time (min.)	Service time SCV	Processing time distribution (min)	Batch process size k_i
1	15	1.237	0.73	GAMM(1.38,0.9)	1
2	5	1.752	0.52	GAMM (1.91,0.92)	1
3	7	1.154	0.40	GAMM (2.5,0.46)	1
4	n.a.	6	1	GAMM (1,6)	1
5	1	4	0.25	GAMM (4.1)	50
6	5	24	0.01	GAMM (144,0.17)	20
7	4	24	0.01	GAMM (144,0.17)	30

Table 14. Comparison of total cycle time for the clinic (Test 3-2)

Scenario	Arrival rate to the clinic (residents/min)	Total cycle time from simulation	Total cycle time from queueing network model	Percentage error %
1	10.42	154.98	54.41	64.89%
2	10.00	58.45	47.77	18.27%
3	9.62	49.37	45.19	8.46%
4	9.09	44.23	41.96	5.14%
5	8.70	42.21	39.61	6.17%
6	7.04	38.93	37.86	2.72%
7	5.26	39.39	38.50	2.25%
8	4.17	41.24	40.23	2.45%

Cycle time from simulation (min) with lower and upper bound (confidence interval for 95%)
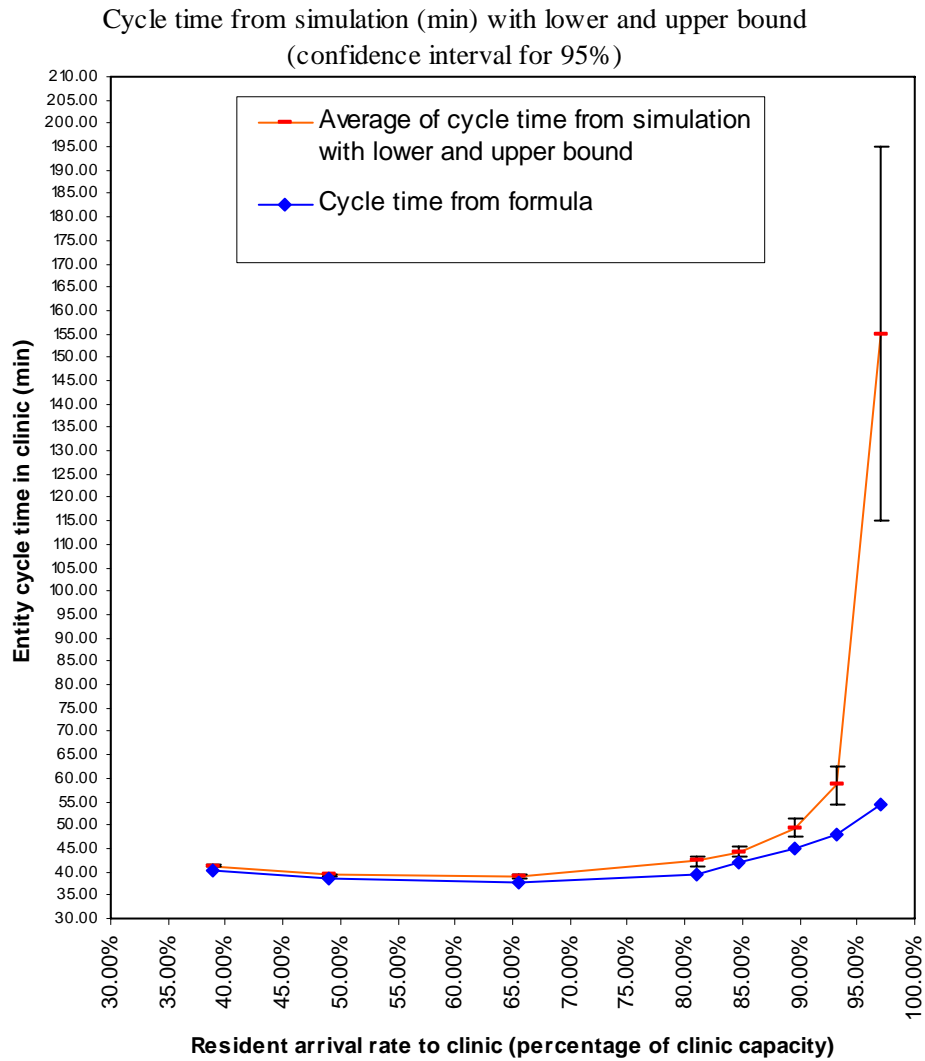

Figure 8. Comparison of total cycle time for the clinic (Test 3-2)

When the arrival rate is low to moderate, there is a small difference between the estimates from the queueing network model and the discrete event simulation. As the arrival rate approaches the clinic capacity, the difference between the two estimates is large due almost entirely to different estimates for the cycle time at station six, which has the highest utilization (since it is the clinic bottleneck).

3.4 Experiment Four

In this experiment, the clinic has six workstations. The first workstation has an individual service process. The next four workstations are batch process workstations. The sixth station is a self-service station.

The model was tested at several levels of resident arrival rates, from 55% to 98% of clinic capacity under the different scenario. We ran 10 replications of 4000 hours, with 1000 hours of warm-up time allowed to achieve steady state for each scenario. Data was recorded for mean total time and mean queueing time at each node, as well as mean time in system and mean system WIP.

This experiment consisted of two tests: Test 4-1 and Test 4-2. In both tests, the average arrival batch size equals 40. The batch interarrival times were exponentially distributed, and all of the service process distributions had gamma distributions to allow different process time SCV. In Test 4-1, the SCV of the arrival batch size is 0.024; in Test 4-2, the SCV of the arrival batch size is 0.25.

Table 15 describes each of the seven stations in the clinic. Table 16 shows the routing probabilities from one station to another. Table 17 lists the capacity of each station and its bound on the clinic capacity.

In this experiment, station number 1 is the bottleneck station, and the clinic capacity is 4.85 residents per minute. Each test in this experiment has eight scenarios with different arrival rates.

Table 15. Parameters for clinic stations in Experiment Four.
(Station 6 is a self-service station.)

Workstation	Staff	Mean service Time (min.)	Service Time SCV	Processing time distribution (min.)	Batch processing size k_i
1	6	1.237	1.11	GAMM(1.38,0.9)	1
2	2	14	0.52	GAMM (1.96,7.14)	30
3	2	18	0.40	GAMM (9.53, 1.89)	40
4	2	21	0.56	GAMM (6.3, 3.33)	50
5	2	23	1.00	GAMM (8.82,2.61)	60
6	n.a.	6	0.44	GAMM (3,2)	1

Table 16. Routing table for the clinic in Experiment Four

From	To					
	2	3	4	5	6	Exit
1	0.20	0.30	0.50	0	0	0
2		0.25	0	0.35	0.4	0
3			0.45	0.55	0	0
4				0	1.00	0
5					0	1.00
6						1.00

Table 17. Table capacity for the clinic's stations in Experiment Four

Workstation	Station capacity (residents/min)	Relative Throughput	Bound on clinic capacity (residents/min)
1	4.850	1.000	4.850
2	4.286	0.200	21.429
3	4.444	0.350	12.698
4	4.762	0.658	7.242
5	5.217	0.263	19.876
6	n.a.	0.738	n.a.

Results for Test 4-1

In Test 4-1, the arrival batch size distribution is 1+Poisson (39), and the arrival batch size SCV equals 0.024.

Table 18 shows the average total cycle time (in minutes) from the simulation results (Test 2-1) and from our queueing network model. Figure 9 shows the total cycle time estimates from the queueing network model and the discrete event simulation for a variety of arrival rates in Test 4-1. The plot for the simulation results includes error bars showing the 95% confidence interval on each estimate.

Table 18. Comparison of total cycle time for the clinic (Test 4-1)

Scenario	Arrival rate to the clinic (residents/min)	Total cycle time from simulation	Total cycle time from queueing network model	Percentage error %
1	4.76	268.75	285.25	6.14%
2	4.55	118.00	119.14	0.97%
3	4.44	103.04	102.64	0.38%
4	4.35	92.78	93.28	0.53%
5	4.21	89.36	85.13	4.73%
6	4.00	81.13	78.32	3.46%
7	3.33	73.69	72.40	1.75%
8	2.67	76.13	75.15	1.28%
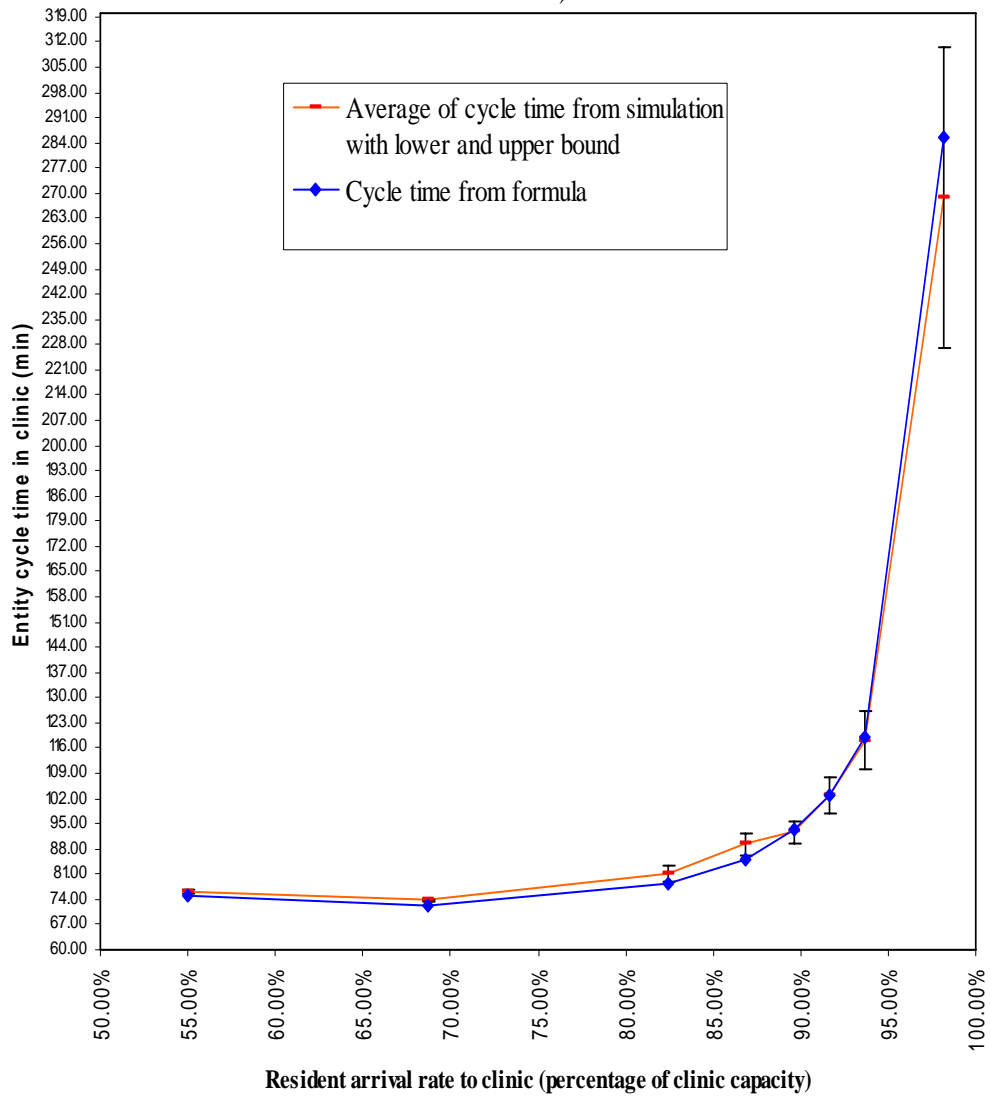
Figure 9. Comparison of total cycle time for the clinic (Test 4-1)

Results for Test 4-2

In Test 4-12 the arrival batch size distribution is 440-Poisson (400), and the arrival batch size SCV equals 0.25.

Table 19 shows the average total cycle time (in minutes) from the simulation results (Test 2-1) and from our queueing network model. Figure 10 shows the total cycle time estimates from the queueing network model and the discrete event simulation for a variety of arrival rates in Test 4-2.

Table 19. Comparison of total cycle time for the clinic (Test 4-2)

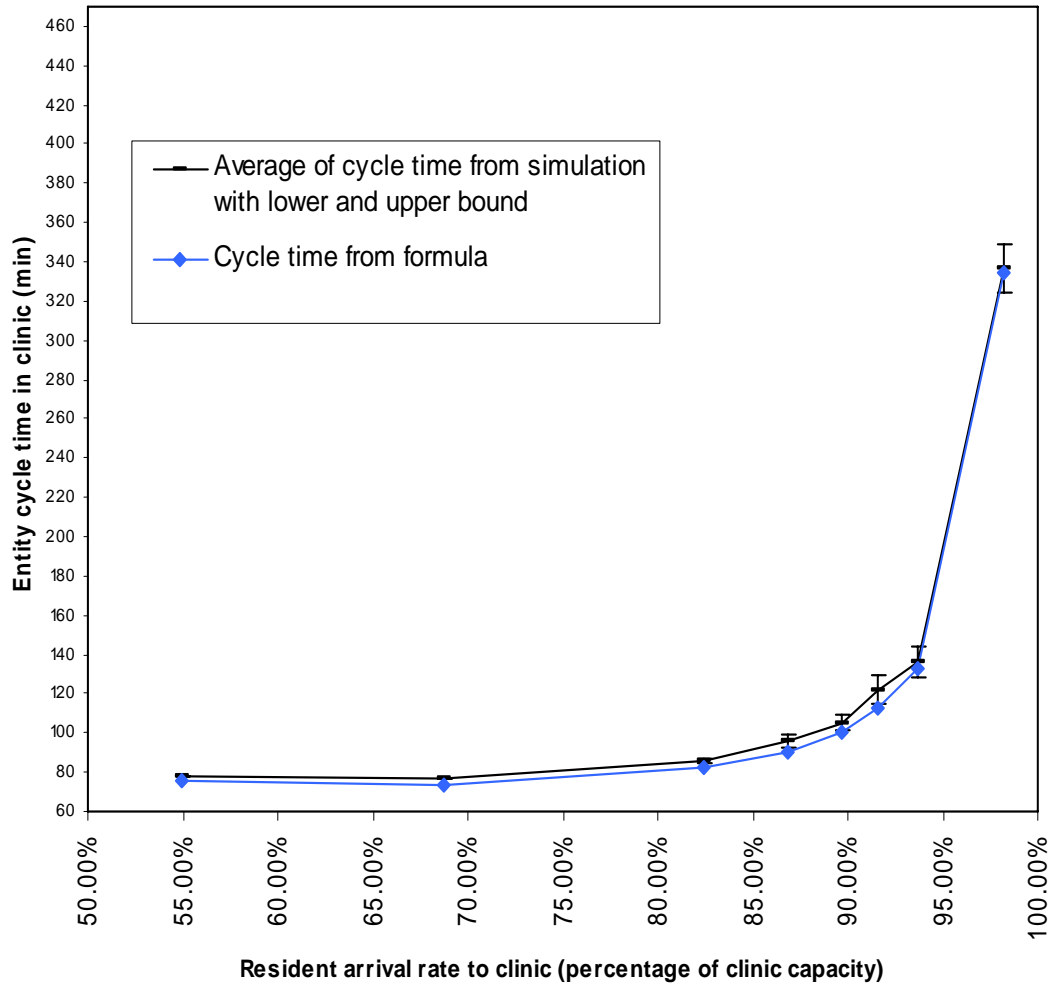| Scenario	Arrival rate to the clinic (residents/min)	Total cycle time from simulation	Total cycle time from queueing network model	Percentage error %
1	4.76	336.60	334.52	0.62%
2	4.55	135.95	132.27	2.70%
3	4.44	122.05	112.11	8.15%
4	4.35	105.20	100.62	4.35%
5	4.21	95.39	90.57	5.05%
6	4.00	85.59	82.04	4.14%
7	3.33	76.84	73.90	3.82%
8	2.67	78.14	75.84	2.94%

Figure 10. Comparison of total cycle time for the clinic (Test 4-2)

As we see from the results for Test 4, when the arrival rate is low to moderate, there is a small difference between the estimates from the queueing network model and the discrete event simulation.

4   CONCLUSION

The model presented in this paper can evaluate a variety of mass dispensing and vaccination clinics with a diverse set of workstations, arrival processes, and service processes (including individual processing, batch processing, and self service) as an integrated and complete queueing network.

The model synthesizes a variety of existing and new proposed models into a systematic approach and includes significant and innovative contributions. One of the components is a model of a queueing system with variable arrival batches and a batch service process whose batch size is larger than the arrival batch. Another innovative component is a model that describes the departure process of self-service workstations.

This model is not limited to just one mass dispensing and vaccination clinic. The model takes into account arrival batches, process batches, parallel servers, stochastic routings, and self-service stations. The model provides good estimates for total cycle time in a typical mass dispensing and vaccination clinic scenario when resident arrival rates are low or moderate (less than 85% of capacity). Compared to results from a discrete event simulation model, the total cycle time estimates have an error of less than 8%. As arrival rates approach clinic capacity, the error in the total cycle time estimates is small in some cases but increases in others. It is not clear why the model performance varies in this way.

Despite the discrepancies that can occur in some scenarios, we believe that this model is an important contribution. We are aware of no other open queueing network approximations that can model such a wide range of processes. Thus, the approximations presented above, combined systematically in this way, represent a significant step towards a comprehensive model for open queueing networks. We continue testing to evaluate the approximations' accuracy for other types of clinics and are implementing the approximations in software for generating clinic planning models.

Although planning mass dispensing and vaccination clinics motivated the development of this queueing network approximation, we expect that the model will be useful in other applications where the queueing networks have similar characteristics. Testing using data relevant to these applications should be done to determine the accuracy of the approximations.

REFERENCES

Aaby, K., Herrmann, J.W., Jordan, C., Treadwell, M., and Wood, K. (2005) Improving Mass Vaccination Clinic Operations, Proceedings of the International Conference on Health Sciences Simulation, New Orleans, Louisiana, January 23-27, 2005.

Aaby, K., Herrmann, J.W., Jordan, C., Treadwell, M., and Wood, K. (2006a) Using Operations Research to Improve Mass Dispensing and Vaccination Clinic Planning, Interfaces, Volume 36, Number 6, pages 569-579.

Aaby, K., Abbey, R., Herrmann, J.W., Treadwell, M., Jordan, C., and Wood, K. (2006b) Embracing Computer Modeling to Address Pandemic Influenza in the 21st Century, Journal of Public Health Management and Practice, 12(4) 365-372.

Buzacott, J.A., Shanthikumar, J.G. (1993) Stochastic Models of Manufacturing Systems. Prentice Hall, Englewood Cliffs, N.J.

Curry, G.L., Deuermeyer, L.B. (2002) Renewal approximations for the departure processes of batch systems, IIE Transactions, 34, 95-104.

Fowler, J.W., Phojanamongkolkij, N., Cochran, J.K., Montgomery, D.C. (2002) Optimal batching in a wafer fabrication facility using a multi-product G/G/c model with batch processing, International Journal of Production Research, 40(2) 275-292.

Gross, D., and Harris, C. (1974) Fundamentals of Queueing Theory. John Wiley and Sons, New York.

Hall, R.W. (1991) Queueing Methods for Services and Manufacturing, Prentice Hall, Englewood Cliffs, New Jersey.

Herrmann, J.W. (2008) Disseminating Emergency Preparedness Planning Models as Automatically Generated Custom Spreadsheets, to appear in Interfaces.

Hopp, W.J., and Spearman, M.L. (2001) Factory Physics: Foundations of Manufacturing Management, Second Edition, Irwin/McGraw-Hill, New York.

Hu, M.-D., and Chang, S.-C. (1998) Translating output specifications to production flow requirements for re-entrant lines, Proceedings of the 37th IEEE Conference on Decision and Control, Tampa, Florida, December 16-18, 1998, Volume 2, 2160-2165.

Kamath, M., Sivaramakrishnan, S., and Shirhatti, G. (1995) RAQS: A software package to support instruction and research in queueing systems, 4th Industrial Engineering Research Conference Proceedings, IIE, Norcross, Georgia, pp. 944-953.

Kaplan, E.H., Craft, D.L., Wein, L.M. (2002) Emergency response to a smallpox attack: The case for mass vaccination, Proceedings of the National Academy of Sciences, 99(16), 10935-10940.

Pilehvar, A. (2007) Queueing Network Approximations for Mass Dispensing and Vaccination Clinics, M.S. Thesis, University of Maryland, College Park.

Pilehvar, A., and Herrmann, J.W. (2008) Approximating the Departure Variability of Self-Service Stations, working paper, Institute for Systems Research, University of Maryland, College Park, Maryland.

Segal, M. and Whitt, W. (1989) A queueing network analyzer for manufacturing. Proceedings of the 12th International

    Teletraffic Congress, Torino, Italy, June 1988, 1146-1152.

Shore, H. (1988) Simple approximations for the GI/G/c queue - I: steady state probabilities, Journal of the Operational

    Research Society, 39(3), 279-284.

Treadwell, M. (2006) Queueing Models and Assessment Tools for Improving Mass Dispensing and Vaccination Clinic

    Planning, M.S. Thesis, University of Maryland, College Park.

Whitt, W. (1983) The queueing network analyzer. Bell Systems Technical Journal, 62, 2779-2815.

Whitt, W. (1984) Approximation for the departure and queues in series. AT&T Laboratories, Holmdel, New Jersey.

ACKNOWLEDGMENTS

AUTHOR BIOGRAPHIES

ALI PILEHVAR is a Ph.D. student in the Robert H. Smith School of Business at the University of Maryland, College

Park. His e-mail address is `<apilehva@rhsmith.umd.edu>`.

JEFFREY W. HERRMANN is an associate professor at the University of Maryland, College Park, where he holds a

joint appointment with the Department of Mechanical Engineering and the Institute for Systems Research. He is the

director of the Computer Integrated Manufacturing Laboratory. He is a senior member of IIE and also a member of

INFORMS, ASME, SME, and ASEE. He received his Ph.D. in industrial and systems engineering from the University

of Florida. His e-mail address is `<jwh2@umd.edu>`.