ABSTRACT

Title of Document: QUEUEING NETWORK APPROXIMATIONS FOR MASS DISPENSING AND VACCINATION CLINICS.

Ali Pilehvar, Master of Science, May 2007

Directed By: Dr. Jeffrey Herrmann, Department of Mechanical Engineering and Institute for System Research

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 before such an event occurs requires determining clinic capacity and estimating queueing performance.

Due to the nature of these facilities, we model a clinic as an open queueing network and estimate the time that county residents will spend at each workstation in such facilities. County residents are the customers, and the servers are the clinic staffs, who are the critical resource. Residents arrive according to an external (not necessarily Poisson) arrival process. When a resident arrives, he goes to the first workstation. Based on his information the resident moves from one workstation to another in the clinic.

We decompose the queueing network by estimating the performance of each

workstation using a combination of exact and approximate models. There is a network of nodes and directed arcs. The nodes represent service facilities (workstations) and the arcs represent residents' flows through the clinic. We characterize each workstation by the first two moments of the interarrival time and service time distributions and consider it as a *G/G/m* queueing system. Congestion measures for the entire network are obtained by assuming as an approximation that the nodes are stochastically independent given the approximate flow parameters.

A key contribution of this thesis is to introduce approximations for workstations with batch arrivals and multiple parallel servers, for workstations with batch service processes and multiple parallel servers, and for self service workstations.

We validated the models for likely scenarios using data collected from emergency preparedness exercises and from simulation experiments. Although this research was motivated by this specific application, it should be applicable also to the design and analysis of manufacturing systems with batch service processes.

QUEUEING NETWORK APPROXIMATIONS FOR MASS DISPENSING AND VACCINATION CLINICS.

By

Ali Pilehvar.

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
[Master of science]
[2007]

Advisory Committee:
Associate Professor Jeffrey W. Herrmann, Chair
Associate Professor Subramanian Raghavan
Associate Professor David Lovell

Acknowledgements

I would like to express a great deal of appreciation to Dr. Jeffrey Herrmann for his careful guidance, support, patience and significant suggestions throughout all of my research specially to accomplish this thesis. I also wish to thank staff of Montgomery County (Maryland) Public Health Services for their support. Furthermore, I am very grateful to all of my colleagues in Computer Integrated Manufacturing (CIM) laboratory at University of Maryland. Lastly, I express my sincere gratitude to my family for their devotions especially to my father for his unfailing support.

Cooperative Agreement Number U50/CCU302718 from the CDC to NACCHO supported this publication. Its contents are solely the responsibility of the University of Maryland and the Advanced Practice Center for Public Health Emergency Preparedness and Response of Montgomery County, Maryland, and do not necessarily represent the official views of CDC or NACCHO.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

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.

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 it would have to set up mass dispensing and vaccination clinics. Counties across the United States are creating plans for this type of response.

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.

While the study of queueing networks has resulted in numerous results, the need to model queueing networks with batch service processes performed by multiple parallel servers and self service stations led us to develop the model presented here. Motivated by the setup of typical clinics, we assume that there is no re-entrant flow.

The fundamental problem is to evaluate the capacity and queueing of a given clinic design, 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.

The queueing network operates in the following manner. When a resident arrives, s/he goes to the first workstation. Based on that resident's personal information (including current state of health and medical allergies), the resident moves from one workstation to another in the clinic. Most residents will receive treatment (medication or vaccination) and then leave. However, some residents will leave without receiving treatment, and others will be transported to a hospital.

Most of the workstations in a clinic have multiple and 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 in a clinic have batch service 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 the 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. Such processes also cause batch (bulk) arrivals at subsequent stations.

There are also 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.

In this thesis we develop an analytical model for queueing networks that have batch arrivals, batch size variability, batch service processes and self service stations. This model yields approximations for queueing network performance. Using data collected from emergency preparedness exercises we preformed the results of a set of simulation experiments in order to assess the accuracy of our proposed analytical model and evaluate these approximations for typical scenarios by comparing their performance to the results of the discrete event simulation models.

1.1 Motivation

In engineering, performing experiments on a real system is often infeasible –for instance, it may be expensive to take a manufacturing system offline to investigate different setup options. On the other hand, traditional discrete-event simulation, which permits accurate analysis of the performance of a wide array of systems is also often time consuming.

An interesting alternative to represent most real world queueing systems is to use analytical models based on queuing theory, although some of them may be difficult to solve mathematically. If the model has been verified and validated then it can be accepted as a dependable substitute for the real system.

Among all different types of queueing networks, the presence of batch (bulk) arrivals, batch size variability, batch service systems performed by multiple parallel servers and the existence of self service workstations within open networks make approximating the queueing network an interesting problem.

There has been extensive research on queueing systems with batch arrivals, queueing systems with infinite number of server (to represent a self service station), and queueing systems with batch service mechanisms in different areas such as manufacturing, communication and computer systems. These studies have mainly introduced general intricate approaches and series of sophisticated mathematics for the queueing systems being studied. Most of the papers in this regard indicate their corresponding queueing model under assumptions of Poisson arrival and exponential service. Unfortunately, these results are not useful in real-world problem settings where relevant performance estimates are needed.

In other words, in most scholars considering the batch arrivals, batch service process and self service problems, there are no useful studies leading to some sets of closed formulas to specifically calculate the batch arrivals, batch service and self service measuring performance applicable practically in real engineering problems such as clinic planning which is our main concern in this thesis.

Since we were unable to find previously proposed models that apply to the situation addressed in our clinic model, we intuitively and experimentally introduced some new concepts and methods to find out queueing network approximations presented in this thesis build on existing and studied models and include novel contributions as well.

A significant contribution of this thesis is the synthesis of a variety of existing and new proposed models into a systematic approach for the type of queueing network explained in this thesis. For example, one of the studied models, never studied before, is a queueing system having both batch arrivals with batch size variability and a batch service process whose batch size is bigger than the arrival batches. Moreover, including self service workstations in models of a real mass dispensing and vaccination clinics as well as studying their behavior is another unique contribution of this thesis.

1.2 Thesis outline

The remaining part of this thesis is organized into four chapters. Chapter 2 provides background about mass dispensing and vaccination clinics and queueing theory in general and reviews briefly the existing approaches for queueing network modeling as well as queueing networks under steady state condition. Then, we introduce different types of batches and waiting time which might exist in the mass dispensing and vaccination clinics. Moreover, at the end of the Chapter 2, we describe two different types of simulations we are carrying out in this thesis.

Chapter 3 includes the existing model of the mass dispensing and vaccination clinics and states its limitations compared to the models that this thesis proposes. The results and findings in this chapter are from Mark Treadwell's thesis (2006).

In Chapter 4, we presents the results of computational experiments completed to evaluate and find different estimates for wait-in-batch-time, self service interdeparture time variability, and batch formation process including batch formation variability and estimation for average waiting to form the batches or wait-to-batch-time. Moreover, we describe our batch branching approach and its results at the end of this chapter.

In Chapter 5, we bring our findings and formulas from Chapter 4 and integrate them with other existing models for queueing system. Then, in order to construct our final model of the mass dispensing and vaccination clinic, we divide the clinic queueing systems into 6 different types of stations. Additionally, at the end of Chapter 5, we validate our clinic model by running some long-run simulation for specific clinic examples and comparing the simulation results with the estimates obtained from our mathematical equations.

Finally, Chapter 6 concludes the thesis and recommends areas for future investigation.

Chapter 2: Literature Survey

Recent intentional and natural disease outbreaks in the United States, caused by a terrorist act or a natural occurrence, such as the 2001 anthrax attacks and the 2003 influenza season, have focused increased attention on the ability of state and local public health authorities to provide affected individuals and communities with rapid, reliable access to medications or vaccination.

Fortunately, guidelines and standards provided by different Federal or non-federal health organizations do exist to aid planners of the clinics in their work. Moreover In order to design the best policy of managing the clinics and give the personnel training under real working conditions, local governments sometimes run full-scale disaster simulations. During these exercises, the performance measures recorded there were used to build a computer simulation model and construct the several pieces of software and spreadsheets. These software packages along with their related tools are basically constructed based on the employment of an operations research discipline called queueing theory which is mainly used to approximate the performance of the queueing networks like what we have in mass dispensing and vaccination clinics.

Since there is plenty of room for improvement in the currently available software tools, particularly with regard to their ability to adapt their models to a particular situation, the role of queueing network theory in updating the existing models as well as introducing the new queue approximations by utilizing more exact approaches is undeniable.

2.1 Mass Dispensing and Vaccination clinics

In light of the substantial health risks posed by anthrax, influenza, smallpox and other bacteria, the U.S. Federal government has called on all states especially the regions that are densely populated or strategically important, such as the nation's capital to devise comprehensive mass preventive plans and policies to ensure that civilian populations have timely access to necessary antibiotics and/or vaccines in the event of future outbreaks.

Although different prophylaxis plans 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 case of an emergency, county residents will visit clinics to receive treatment. The building housing the clinic may be a school, a recreation center, a concert hall, or some other facility 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 in the effected area. 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.

The last couple of years have seen a major expansion of Federal assets to assist local public health providers in the planning and execution of mass prophylaxis campaigns for bioterrorism and epidemic outbreak response. Although each county has its own plan for setting up and operating a clinic, many are planning to setup clinics similar to that shown in Figure 1 in case of smallpox. (This design is based upon federal guidelines.) Each box in Figure 1 represents stations 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. Moreover, holding room, symptoms room and consultation can have residents exiting without receiving vaccinations.

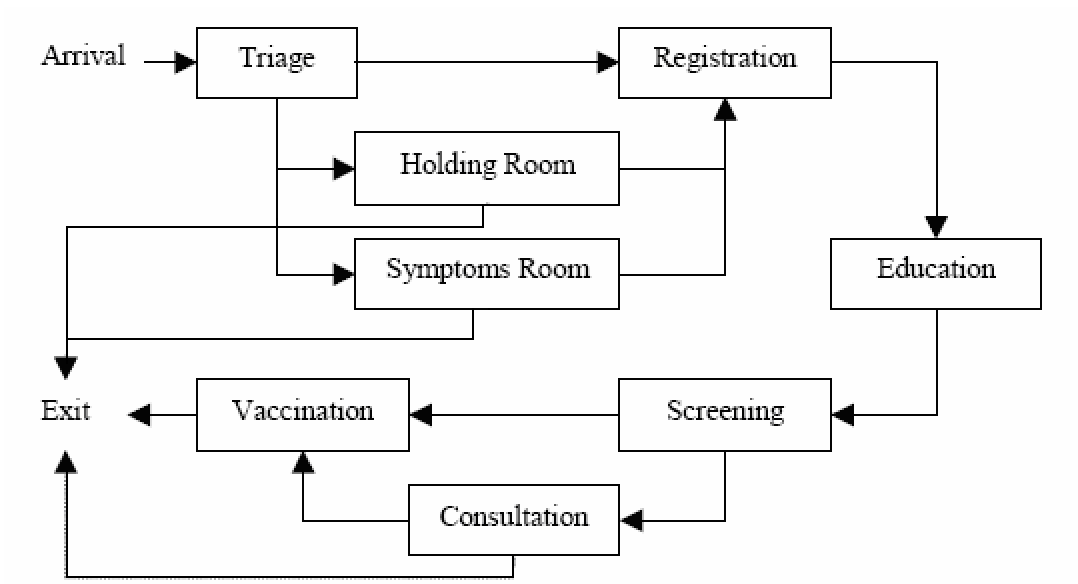

Figure 1. Flowchart of resident flow (Pilehvar et al. 2006)

State, county, and local health authorities have been charged with the development of their suitable mass prophylaxis plans, with financial and technical support of the Department of Health and Human Services Office of Public Health Emergency Preparedness (OPHEP) as well as the Centers for Disease Control and Prevention (CDC).

According to the “Community-Based Mass Prophylaxis” guide, there are five main components to outbreak response: surveillance, supply and stockpiling, distribution, dispensing, and follow-up (AHRQ, 2004). When surveillance teams have identified a disease outbreak, medication from the Strategic National Stockpile (SNS) will be distributed at the federal and state levels. Receiving and dispensing this medication is the responsibility of local public health authorities.

Dispensing of antibiotics and/or vaccines is a key activity of any mass prophylaxis campaign against outbreaks of preventable disease. Without the ability to safely dispense large volumes of medications or vaccines to community-based individuals, efforts to improve surveillance, stockpiling, or distribution capacity will not translate into improved public health response. Conversely, dispensing operations are critically dependent on these surveillance, stockpiling, and distribution functions for defining the prophylaxis mission to be accomplished and for supplying the medical materiel necessary for its successful completion.

There are two possible approaches to mass prophylaxis: “push” and “pull”. The “push” approach, exemplified by the recent Memorandum of Agreement between the Department of Health and Human Services (DHHS) and the U.S. Postal Service, consists of bringing medicine directly to individuals or homes in an affected community. The “pull” approach, in contrast, requires that individuals leave their homes or places of work in order to travel to specially designated centers where they can receive medications or vaccinations. Each approach has strengths and weaknesses. The “push” approach may enable faster and more widespread coverage of an affected community, but it has little flexibility to handle medical evaluation for contraindications or dosage adjustment and may be infeasible for vaccination campaigns. On the other hand, the “pull” approach may increase efficient use of scarce health care providers and resources, enable medical evaluation of potential victims, and provide opportunities for centralized data; however, these advantages must be weighed against the delays and logistical challenges of setting up sufficient dispensing clinics to handle high patient volumes.

In this thesis, we study the “pull” approach, which means the individuals will visit clinics to receive treatment. In the “pull” model of mass prophylaxis, the Dispensing/Vaccination Clinics is the principal operational unit of the dispensing function of community-wide disease outbreak response.

2.2 Queueing Theory in general

Queueing theory is generally considered a branch of operations research, and it is simply the science of waiting. Since jobs “stand in line” while waiting to be processed, waiting to move, waiting for parts, and so on, queueing theory is a powerful tool for studying and modeling any system having a queue inside such as a manufacturing, transportation, and telecommunication system.

The theory enables mathematical analysis of several related processes, including arriving at the queue (arrival or input process), waiting in the queue (waiting process) and being served at the workstations (service process). Each workstation consists of units which provide service to the arriving entities such as jobs or customers. These units are usually called servers and can be either people or machines.

A queueing system combines the components that have been considered so far: an arrival (input) process, a queue, and a service process. For the arrival process, in most cases, the arrival process is the product of external factors. Therefore, the best way, one can do is to describe the arrival process in terms of random variables which can represent either the number of arrivals during a time interval or the time interval between successive arrivals. In this way, the arrival process can stem from several streams whose arrival probability distributions are different and independent. In the meanwhile, if entities (jobs or customers) arrive in groups, their size can be a random variable as well.

For a queue, the possible queueing discipline can be first-come first-served (FCFS), last-come first-served (LCFS), shortest process time (SPT), earliest due date (EDD), or any of a host of priority schemes. In many situations customers in some classes get priority in service over others. In this thesis, for all workstation, we have the FCFS service discipline without having any kind of priority scheme for a specific class of customers.

Additionally, for the service process, the workstations can have different number of servers; the various batch processing sizes (number of customers getting served at once), the service time and mode of service. The serving time is a random variable which can be generated from any probability distribution. Although, in most workstations, the batch processing size is a fixed number, it can be a random variable and follows a probability distribution in some cases.

The basic notation widely used in queueing theory for a queueing system is made up symbols representing three elements: input/service/number of servers. For instance, using M for Poisson or exponential, D for deterministic (constant), and G for general distribution.

The whole objective of studying the queueing behavior of a queueing system is to estimate some useful performance measures such as average waiting time in the queue or the system, the expected number of customers waiting or receiving service and the probability of encountering the system in certain states, such as empty, full, having an available server.

Another important issue in queueing systems is capacity. How many customers can wait at a time in a queueing system is a significant factor for consideration. If the waiting room is large, one can assume that for all practical purposes, it is infinite. But a real world queueing system such as a telephone system tells us that the size of the buffer that is able to accommodate our calls while waiting to get a free line is finite and important to know.

A queueing network is simply composed of several queuing systems. Queues can be chained to form queueing networks where the departures from one queue enter the next queue. Queueing networks can be classified into two categories: open queueing networks and closed queueing networks. Open queueing networks have an external input and an external final destination. Closed queueing networks include some customers circulating continually in the network with no leaving. In this thesis, we are dealing only with the open queueing networks.

In the following sections, we will discuss briefly some of important approaches, areas and concepts in queueing theory we are intending to employ in this thesis.

2.2.1 Approximate approaches to model open queue network

Since it is difficult to obtain exact analytical solutions for complex problems with general service and arrival time distributions, bulk arrival and batch service process, an alternative is to have an approximate analytical solution to a more realistic model.

The approximation models for analyzing job shops using open queueing networks can broadly be classified into four categories: decomposition methods, diffusion approximations, mean value analysis, and operational analysis. The procedure that has been employed with considerable success to analyze the open network such as a manufacturing system is the decomposition approach. Only recently diffusion models have been utilized to study scheduling and operational control problems arising in manufacturing. Operational analysis (example: Denning and Buzen, 1978) has been applied primarily to computer system models, and mean value analysis (example: Reiser and Lavenberg, 1980) is concerned with closed queueing networks. Because of the importance of decomposition and diffusion, we will delve into them in separate sections.

2.2.1.1 Decomposition approach

The decomposition approach is an approximation method that leads to acceptable results in a wide variety of open networks.

The overall approach is to decompose a system into small components, model these components, and then integrate the general system by the appropriate combination of these components. In other words, in the decomposition approach, the network is broken down into several workstations (nodes). The decomposition approach makes two basic assumptions: (a) the nodes can be treated as being stochastically independent; and (b) the input to each queue is a renewal process characterized by the mean and variance (two parameter approximation) of the interarrival time distributions of customers. Often, we use the square coefficient of variation (SCV), which equals the variance divided by the square of the mean. The output and input to each node is linked to customer routings. The linking of outputs and inputs can be solved to obtain performance at each node. The three main steps in the approximation are as follows:

- Decomposition of the network into individual nodes.
- Analysis of each node and the interaction between the nodes.
- Re-composition of the individual results to compute the network performance.

One type of decomposition approach is the parametric decomposition approach (PDA) which has been very effective in estimating the first moment of the queue length in general networks. Reiser and Kobayashi (1974) and Kuehn (1976, 1979) were among the first proponents of the parametric decomposition approach, which was later used by Shanthikumar and Buzacott (1981) for single product networks and by Whitt (1983a, b). This approach, which is also utilized in this thesis, generalizes the notion of independence and product form solutions of Jackson type networks to more general models. In this method, the arrival process at each station is approximated by a renewal process. Additionally, the interarrival time SCVs at each station are computed approximately. The performance measures such as mean number of jobs or customers and queue lengths at each station are estimated based on these SCVs.

2.2.1.2 Diffusion approach

Diffusion approximations are based on the heavy traffic limit theories (Reiman(1984), Chen and Mandelbaum(1991)). These approximations are valid when the traffic intensities at the workstations are close to one (traffic intensity is defined as the ratio of the arrival rate to the total processing rate). They use reflected Brownian motion to approximate the queueing network, requiring a large number of partial differential equations to be solved. The concept of Brownian motion is taken from the field of physics, where it is used to model the random movement of small particles.

Since the characteristics of job shop and our clinic problem we are studying in this thesis are comparable; we will utilize one of the existing approaches, decomposition (parametric decomposition approach), for analyzing the job shop to model the mass dispensing and vaccination clinics.

In Chapter 5, we will concentrate more on decomposition, since decomposition is a suitable approach for our model.

2.2.2 Analysis of queueing systems under the steady state condition

Most analytical results in qeueing theory are for queueing systems with steady state condition. The steady state condition is reached as the time from system initialization becomes very large and the initial conditions no longer have any effect on the performance measures. The literature emphasizes this type of analysis because the equations involved are considerably simplified in the limit, and relatively straightforward techniques such as balance equations and Little’s laws can then be used.

In steady state condition, some time has elapsed after the system is started or initiated. This initial situation is often identified as a transient state, start-up or warm-up period. One of the good reasons that make the steady state condition a strong method of analyses in queueing network theory is the independence between the initial condition of the queueing systems and long-run performance measures.

Nevertheless, such steady state analyses are inappropriate in many real world situations since the time horizon of operation naturally terminates, or steady-state measures of system performance simply cannot be reached.

For example, a bank has a definite closing time each day, and the repairmen at a service facility will leave at some point. For such problems, an appropriate analysis would be transient, i.e., it would describe the system's operation for a fixed, finite amount of time (or for a fixed number of "customers") and take into account the initial conditions of the system.

However, transient results can be quite difficult to obtain, tend to be rather complicated, are available only for a fairly restricted class of models, and usually assume "empty and idle" initial conditions for the system. That is why; it is not usually employed in most previous researches studying queueing systems because of its complexity.

One simple condition for steady state in a queueing system is that the customer arrival rate to the system is less than the service rate. This means that if our system runs for an infinite amount of time, it will not blow up, that is, the number of customers in the system will remain finite. For example, The *M/M/m* queue experiences poisson arrivals at rate λ, has a single first-in, first-out (FIFO) queue feeding s parallel servers, each providing exponential service at rate μ; all interarrival and service times are assumed to be independent of each other. The steady-state behavior of this system is well known (see Gross and Harris, 1974). Assume that $u = \frac{\lambda}{m\mu}$, for having steady state u should be less than 1.

Because of all the afore-mentioned reasons, queueing models are generally constructed to represent the steady state of a queueing system and analyze the performance measures under steady state condition. That is, they evaluate the typical, long-run or average state of the system. As a consequence, these are stochastic models that represent the probability that a queueing system will be found in a particular configuration or state.

A general procedure for constructing and analyzing such queueing models is:

1. Identify the parameters of the system, such as the arrival rate, service time, queue capacity, and perhaps draw a diagram of the system.

2. Identify what are the system states. (A state will generally represent the integer number of entities such as customers or jobs in the system and may or may not be limited.)

3. Draw a state transition diagram that represents the possible system states and identify the rates to enter and leave each state. This diagram is a representation of a Markov chain.

4. Because the state transition diagram represents the steady state situation between states there is a balanced flow between states so the probabilities of being in adjacent states can be related mathematically in terms of the arrival and service rates and state probabilities.

5. Express all the state probabilities in terms of the empty state probability, using the inter-state transition relationships.

6. Determine the empty state probability by using the fact that all state probabilities always sum to 1.

Since our clinics have to run for couple of days to vaccinate all of the population of a region, in other words, they run for long enough period, in this thesis, we will study our clinic models under the steady state condition (stable queueing systems). Additionally, when we design some simulation experiments to validate our constructed models and new approximations, we take into account acceptable warm-up (transient) periods before reaching the steady sate condition to guarantee having exact simulation results within the given confidence intervals.

2.2.3 Waiting times

One the of most important performance measurers that queueing theory is used to describe is the time a customer or job spends waiting to find an idle server.

To cover all of the cases, we have to find an approximations that satisfy the cases with general arrival and process distribution in which we have multiple servers working in parallel to serve several customers at once. Sakasegawa (1977) proposed an approximation for this queueing time for *G/G/m*, with *m* representing the number of servers, given in Formula 1. Moreover c_a^2 and c_e^2 respectively represents the interarrival time and the service time variability (SCV). When $m = 1$, this equation reduces to the *G/G/1* approximation.

The *G/G/m* approximation for queueing time is:

$$CT_q = \left(\frac{c_a{}^2 + c_e^2}{2} \right) \frac{u^{\sqrt{2m+2}-1}}{m(1-u)} t \quad \textbf{(Formula 1)}$$

2.2.4 Batch (bulk) arrival process

In batch arrival queueing systems, customers or jobs arrive in batches in which an arrival can be a group (of random size) of items. Items (customers or jobs) might be batched for the purpose of having more economical and easier transportation among workstations.

One of the important causes of flow variability in a queueing network is a batch arrivals process. This is one of the strongest motivations for studying how the batch arrival affects the performance measures in most of the published papers in this area.

We can consider a queuing system in which arrivals occur according to a general distribution in batches of varying size and stations have service times distributed according to another general statistical distribution. All service channels can have similar or different identical statistical properties. In this thesis, we assume that all of the service channels (servers) are completely identical. The arrival batches can be served individually or in batch size bigger or smaller than batch arrival size. After completion of service at one service center, a job or customer and a group of jobs or customs may leave the queuing network or may move to another service center for further service.

We study stations that have both batch (bulk) and individual arrivals in mixed-arrival sections of the model formulation.

We also can have a finite or infinite number of servers in a station for a batch arrival queueing system. In this thesis, we assume that the number of servers (service channels) is limited. Thus, based on the notation introduced by Kendall, we use $G^{[X]}/G^{[X]}/m$ to show bulk arrival process with multiple parallel servers possessing batch service process discipline[1].

An example for this queueing system with batch service process size of one (individual service process) is the following behavior: when residents arrives in a group by buses to the mass dispensing and vaccination clinics, they go to the first workstation (triage or greeting) in a batch size of the bus capacity. Based on that resident's personal information, the resident is served individually and is guided to

[1] Batch process size can be bigger or equal to 1. For individual processing the batch process size is 1.

another workstation in the clinic. In this example, although the residents' arrival is in a batch, however, all residents receive service one by one at the first station.

Generally, it is difficult, if not impossible, to find tractable expressions for the waiting time probabilities of individual customers. It is, therefore, useful to have easily computable approximations for these probabilities. Although there are many papers studying methods for the computation of the waiting time distribution, however, these methods apply only for special conditions such as having specific service time distributions or batch interarrival time distribution, they and are, in general, not suited for routine calculations in practice.

To analyze batch arrivals, we study queueing systems in which customers arrive at a station in batches but are processed as individuals. There are two ways of handling them. The first method of unbatching is to treat them as individuals arriving in a process with an extremely high SCV; the arrival variability of individuals out of a batch is given below (Curry, 2002), where the processing time SCV of a batch is denoted by $c_{b,a}{}^2$ and k is the arriving batch size:

$$c^2{}_a = kc_{b,a}{}^2 + k - 1 \quad \textbf{(Formula 2)}$$

The second way of dealing with "unbatching" which is mainly used in this thesis, is to find the time that the batch spends in queue (CT_q) with other batches, then add the time that individuals spend waiting once the batch they arrived in is "opened," referred to as wait-in-batch time (WIBT) (Hopp and Spearman, 2001).

To explain WIBT more, since there are k items in the batch, the items have different delays while awaiting their turn at service. The first item served from a batch has no additional delay due to waiting for others from the same batch, while the second item serviced waits for the first item; the third item waits for the first two selected items, and so on.

$$CT_q = \left(\frac{c_{b,a}{}^2 + \frac{c_e{}^2}{k}}{2} \right) \frac{u}{1-u} kt \quad \textbf{(Formula 3)}$$

$$WIBT = \frac{(k-1)t}{2} \quad \textbf{(Formula 4)}$$

These formulas are for a queueing system with a fixed size arrival batch size and a single server with individual service process ($G^{[X]}/G/1$). $c_e{}^2$ and t are respectively service time SCV and average service time for each arriving customer or item to the workstation. Moreover, $c_{b,a}{}^2$ is the batch interarrival time SCV.

Curry and Deuermeyer (2002) compared these two unbatching strategies and found that the approach suggested by Hopp and Spearman (second way) gave results that were significantly better when compared to a simulation. However, neither Hopp and Spearman nor Curry and Deuermeyer (2002) considered the case of unbatching at a station with multiple servers.

In this thesis, we study the second unbatching strategy for stations with multiple servers. The only application of the first unbatching strategy is in Section 5.2.9, for a self service station with mixed arrivals. Additionally, we also bring two methods to calculate the WIBT for multiple server stations with several batch arrival streams and individual service.

2.2.5 Batch service process

Another type of batching is a frequently encountered batch service process. In the batch service process, the servers in workstation can serve a group of jobs or customers at once. There are many reasons to have batch service process for one workstation. For instance, sometimes, due to the slow processing rates of a workstation, large capacity machines have been developed that can process several units of an items simultaneously. At the completion of service, the batch is removed from the server and the units either as a group or individually is sent to their next workstation.

One of the necessary processes before each batch service process workstation is batch forming at the same size of the batch service process size of downstream stations. Coming items (jobs or customers) should wait in the incomplete batch until the proper quantity has accrued and then the full batch is formed and transported to the workstation waiting to serve these batches.

To model the batch forming procedure, several aspects of the problem will have to be considered. First, the batch forming time as it contributes to each individual item, or the average item delay, needs to be computed. Then the arrival stream characteristics for the batch receiving workstation need be developed. That is, the mean arrival rate for batches and the interarrival time SCV.

When customers arrive at a batch service process, they must first wait while the other customers in the batch arrive, then wait as a batch for the server to become available. Hopp and Spearman (2001) refer to this first delay as wait-to-batch time (WTBT), and define it for a single server station as:

$$WTBT = \frac{k-1}{2\lambda} \quad \textbf{(Formula 5)}$$

In this formula, k is the number of customers or jobs should wait to form a complete batch to be processed in the downstream station. Furthermore, λ is the arrival rate of individuals to the batch service process workstation.

After the batch is formed, queueing can be approximated using the formulas previously discussed, substituting parameters in regard to the batch for the individual parameters. The SCV as the batches are formed and arrive at the process is obtained by dividing the individual interarrival time SCV by k (Hopp and Spearman, 2001).

To analyze batch service process stations in our clinic models, we need to have wait-to-batch time (WTBT) and interarrival time SCV for both arriving individuals and batches from different arrival branches.

From Hopp and Spearman (2001), we have only results for individual arrivals, therefore, we will study the wait-to-batch time (WTBT) and formed batch variability for batch arrivals with batch size variability from multiple arrival streams in the Chapter 4. Additionally, we discuss the effect of the branching process and the formation of arrival batches of random size after the batch service process stations at downstream stations in Chapter 4.

2.2.6 Batch move (transfer)

The third type of batch that has been studied is the batch moves. A batch move is merely for purpose of having more convenient transportation. To have a better understanding of batch move modeling for this type of application, we bring an example.

Consider a queueing system where batches are formed after individual service process and are transported to the next work-station. At the second workstation, batches wait in the queue until service on individual items within the batch begins. Items leave as individuals as soon their service in the station has been completed. In this batch move model, items should arrive at the downstream workstation in batches of fixed size *k*, but are served individually.

A representation of this queueing system is given in Figure 2.

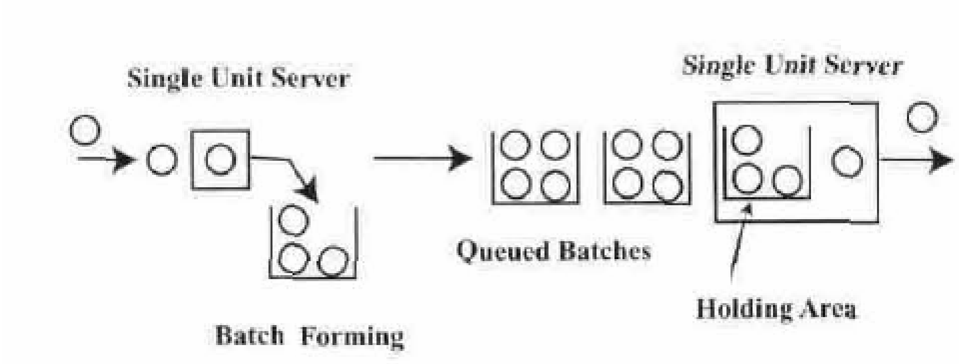

Figure 2. A simple batch move model (Curry and Deuermeyer, 2002)

The general approach for modeling departures from $G/G/1$ workstations is to approximate the interdeparture process by a renewal process (Albin and Kai. 1986).

$$c_d^2 = c_a^2(1-u^2) + u^2 c_e^2 \quad \textbf{(Formula 6)}$$

In this formula, c_d^2, c_a^2 and c_e^2 are respectively the interdeparture time SCV, the interarrival time SCV and the process time SCV for the *G/G/*1. Moreover, u is the notation for utilization.

Since the interdeparture time SCV for a station will be the interarrival time SCV for downstream stations, it is necessary to have approximations of that for different queueing systems. In Chapters 5, we present our approach to calculate the interdeparture time SCV for other types of queueing systems needed for our clinic model.

2.2.7 Self service stations

One of the most important contributions of this thesis is analyzing the behavior of stations in which customers or jobs complete some activities without having any kind of assistance from real servers. The only important concern for self service station is studying the interdeparture time SCV that can affect the behavior of next stations considerably.

In this type of stations, jobs or customers can arrive individually or in batch to the workstation. The customers perform the process themselves without any external resources. In this domain, an example from our clinic would 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. Thus, the workstation can be modeled as a $G/G/\infty$ queueing system.

The idea of utilizing $G/G/\infty$ systems in modeling the self service station is very simple. As we know in $G/G/\infty$, because of the unlimited number of servers, there is always an idle server for each arriving entity and of course there is no waiting time in queue. On the other hand, in a self service station, there is no waiting time in queue for arriving entities because they can immediately initiate serving or completing a process upon arrival. Therefore, the behavior of $G/G/\infty$ queueing system can be similar to the self service's performance from the perspective of interdeparture process which is the main goal of studying the self service stations.

To estimate the interdeparture time variability, we first take into account the following facts. For a $G/D/\infty$ system, the interdeparture time variability equals the interarrival time 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 departure variability equals 1(Burke, 1958; Mirasol, 1963). For a $G/G/\infty$ system, Whitt (1983) suggests that the interdeparture time 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 time variability to equal the interarrival time variability. These imply that, in the general case (a $G/G/\infty$ system with moderate load), the interdeparture time SCV will be somewhere between the interarrival time SCV and one.

In Chapter 4, we study the behavior of the self service station in detail by carrying out some simulation sets for different scenarios and consequently we formulate approximately the extracted and observed trend of self service stations to be utilized in modeling of our clinic in this thesis.

2.3 Model evaluation

As part of developing and testing the queueing models, we will use simulation discrete-event models of queueing systems in various ways such as validation and experimentation.

Discrete-event simulation models carried out in this thesis were all created by Rockwell Software's Arena 5.0 ®. The Process Analyzer software included with Arena was used to manage the running of multiple scenarios and the tabulation of their results. These results included the calculation of a 95% confidence interval on all measured responses.

To construct the simulation models in order to either validate the queueing systems or extract some experimental equations among parameters, transient operation is readily observable, but true steady-state behavior is very difficult to observe, in general. This difficulty stems from the inability to initialize the simulation according to a steady-state distribution, which is presumably unknown if we are conducting a steady-state simulation.

A standard tactic is to initialize the simulation in some "reasonable" way, allow it to run "long enough" for the effect of these initial conditions to have dissipated, and then collect observations during the ensuing "steady-state" portion of the run. The length of the "warm-up" period will certainly depend on the method of initialization, and we would like to initialize in a way which would promote rapid convergence to steady-state operation.

Although for steady state condition, the initial conditions for the simulation models have no effect on the performance measures in long-run, we assume that we have zero customers at the time of zero in our all simulation models in this thesis.

2.3.1 Validation

In order to validate our constructed queueing models and new findings, simulation can be an appropriate tool to assist us to check the exactness of our modeling approach and results. The simulation run lengths and numbers of replications were chosen in order to ensure that confidence intervals were less than 5% of the associated response.

2.3.2 Experimentation

We run simulation models with a range of parameter values, study the results, and then extract trends to get insight into relationships and motivate the models and estimate their parameters.

In Chapter 4, we use simulation models to determine experimentally the behavior of different parameters for WIBT, the batch interarrival time after being formed in batch formation process, and the self service interdeparture time variability.

Chapter 3: Previous work on mass dispensing and vaccination clinics

In this chapter, we briefly review what has been done on mass dispensing and vaccination clinics before I started to do research in this area. The results and findings in this chapter are from Mark Treadwell's thesis (2006).

I will use some of his findings for the model in Chapter 5. On the other hand, most of them are not applicable and suitable for the general cases we are studying in thesis, which is why, intuitively I will follow different approaches and methods in Chapter 4 to integrate them in Chapter 5 to build up the new models for some specific situations we might face in a real clinic.

Although the goal of his research was to compare analytical models of queueing processes to discrete-event simulations in order to determine which models are the most accurate for use with a general set of inputs, Treadwell (2006) mainly focused problems of planning and modeling mass dispensing and vaccination clinics.

In other words, most of his work included modeling the clinics, designing a spreadsheet, and implementing the software targeting public health officials in order to assist them to plan and manage setting up the mass dispensing and vaccination clinics upon the emergent events with more preparedness and effectiveness.

Since in this thesis, we focus on the mass dispensing and vaccination clinics from the perspective of queueing network analysis and mathematical modeling, I merely mention briefly the queueing network approximations that have been previously studied.

3.1 Batch arrivals, individual service process with multiple servers

As mentioned in Chapter 2, Hopp and Spearman (2001) suggested, and Curry and Deuermeyer (2002) demonstrated, that a batch arrival process can be accurately modeled by representing the batches of size k as customers of a process with service time kt, and scaling the process and interarrival time SCV by $1/k$. In order to extend this result to a station with m servers, the service time must be scaled to the new mean of kt/m.

The $1/k$ terms in the SCV actually cancel with the additional k in the service time, and it turns out that the average time a batch spends waiting in queue is the same amount of time that an individual customer would spend in the queue. We also replace the basic utilization term with the multiple-server form given by Sakasegawa (1977). The approximation for WIBT must be adjusted to accommodate a station with multiple servers, again by scaling the mean service time.

$$CT_q = \frac{(c_a^2 + c_e^{\ 2})u^{\sqrt{2m+2}-1}t}{2(1-u)m} \quad \textbf{(Formula 7)}$$

$$WIBT = \frac{(k-1)t}{2m} \quad \textbf{(Formula 8)}$$

To demonstrate the accuracy of this approximation, it is compared to an equivalent simulation model. The results of the simulation for confidence interval 95% are given in Table 1, along with the values obtained using the proposed approximations for both portions of the waiting time. The magnitude of error between the two is given as a percentage of the simulation value which Arena calculated for each of the performance measures.

Table 1. Results for exponential batch arrivals to exponential service process (Treadwell, 2006).

k	m	u	t(min)	Simulation WIBT (min)	Predicted WIBT (min)	Error	Simulation CT_q (min)	Predicted CT_q (min)	Error
5	1	99%	0.033	0.067	0.067	0.0%	10.2	9.901	2.9%
		95%	0.033	0.067	0.067	0.0%	1.92	1.900	1.0%
		90%	0.033	0.067	0.067	0.0%	0.901	0.900	0.1%
		80%	0.033	0.067	0.067	0.0%	0.399	0.400	0.3%
		50%	0.033	0.067	0.067	0.0%	0.1	0.100	0.0%
5	3	99%	0.100	0.066	0.067	1.0%	9.9	9.819	0.8%
		95%	0.100	0.065	0.067	1.0%	1.86	1.821	2.1%
		90%	0.100	0.063	0.067	11.1%	0.84	0.825	1.8%
		80%	0.100	0.058	0.067	15.5%	0.36	0.332	7.6%
		50%	0.100	0.045	0.067	38.9%	0.06	0.056	6.1%

The WIBT from Formula 8 provides an exact match to values obtained from the simulation for *m*=1; however, for *m*>1, the degree of error increases as utilization decreases. This is an interesting result; the discrepancy is caused by the increased likelihood that a batch will find more than one server idle when it arrives at the service process. Despite this discrepancy, the model still provides a useful upper bound on WIBT, and is reasonably accurate for u >90%.

The approximation for batch queueing time given by Formula 7 provides excellent results for Markovian arrival and service processes with a single server, even outside the stated limits on utilization mentioned by Hopp and Spearman (2001); at 99% utilization, the predicted value is within 3% of the simulation result. When multiple servers are present, the issue discussed above leads to a corresponding reduction in the mean service time for batches, and hence in the time batches spend in queue. Therefore, this reduces the model's accuracy somewhat.

The results of this test generally follow the form of the experiment with exponential service times; for a single server, WIBT is exact and the predicted queue time gives a good estimate of the simulated queue time. For multiple servers, the accuracy of the models (Formula 8) is reduced.

In Chapter 4, we will take simulation results for WIBT and *CTq* for various scenarios and compare them with the new formulas for WIBT and *CTq* for multiple server stations and more general cases.

For the WIBT with multiple servers, we can say that when multiple servers are processing residents who arrive in batches, there is some probability that more than one server will be idle when a batch arrives. When this happens, the WIBT for the members of that batch is reduced accordingly, and the queue time for subsequent batches is affected. Although, there is some non-exact open formulas for WIBT with multiple servers which are only applicable in the spread sheet models, there exists no efficient closed formula for WIBT with multiple servers that can be easily implemented. Therefore, we will spend much time in Chapter 4 studying the behavior of WIBT with multiple servers under the general distribution and different scenarios.

At this point, we can see that we need to construct some new formulas and approaches to calculate WIBT and *CTq* for stations with multiple servers under the general cases ($G^{[X]}/G^{[X]}/m$). Although Formulas 7 and 8 can be two possible estimations for stations with batch arrival, individual service process and multiple servers, but since Treadwell (2006) originally has taken them out from queueing systems with a single server, the results of these formulas cannot be acceptable in so many cases. In other words, we need more dependable equations to calculate the performance measures than Treadwell (2006).

3.2 Multiple batch arrival streams

The approximations discussed above are applicable to a station with a single input stream of batches; however, in a queueing network, it is possible that batches will arrive from multiple stations, each with a different batch size. Models for a mixed input of this sort do not appear to exist, so one of the possible ways to aggregate the different batch size from various streams is utilizing the routing probabilities. The proposed equation for aggregate batch size is:

$$\overline{k}_{ai} = \sum_{j=1}^{i-1} k_j \frac{\lambda_j p_{ji}}{\lambda}$$

In this formula, $\overline{k}_{ai}$ and k_j are respectively the aggregate batch size and the arrival batch size from station j to i. Moreover $\lambda = \sum_{j=1}^{i-1} \lambda_j p_{ji}$ and p_{ji} is the routing probability from station j to i. We should say that in Table 2, all rates are in terms of minute and p_{ji} =1for j=1, 2, 3, 4.

This equation calculates the aggregate batch size from the perspective of customers in the batch, based on the proportion of the total flow rate associated with each batch size (this is slightly different from weighting batch sizes by their proportion of the total number of batches that arrive, which gives a mean batch size from an external perspective). This aggregate batch size gives an excellent performance in estimating

WIBT; Table 2 below gives the performance of several simulations with multiple batch arrival streams, along with the results predicted using aggregate batch size as an input to Formula 8 for a single server station. These experiments were performed on an *M/M/1* system.

Table 2. Experimental results for multiple batches arriving to a single server (Treadwell, 2006)

k_1	λ_1	k_2	λ_2	k_3	λ_3	k_4	λ_4	λ	$\bar{k}_a$	t (min)	Simulation WIBT (min)	Predicted WIBT (min)	Error
1	5	5	10	0	0	0	0	15	3.667	0.050	0.0670	0.0667	0.50%
1	10	10	10	0	0	0	0	20	5.500	0.048	0.1064	0.1069	0.45%
1	5	3	9	4	12	5	3	29	3.246	0.033	0.0396	0.0374	5.49%
1	5	20	20	20	40	10	15	80	16.938	0.012	0.0946	0.0946	0.03%

These results make clear that the aggregate batch size approach provides excellent estimates of the performance of a station where batches of different sizes arrive from multiple sources to an individual service process station with a single server.

3.3 Complete queueing modeling framework for the clinics

With the unusual situations accounted for, a complete framework for constructing queueing models can now be described. Demand for service is calculated with user inputs for the total number of customers to be served (population) and how long they have to be serviced (treatment time). We use *i* throughout the proposed queueing model to denote individual stations, with 0 referring to the bus arrival process, 1 through "*I*" referring to the stations in the clinic, and "*I* +1" referring to the exit.

Before presenting the model introduced by Treadwell (2006), we should say that, this model is quite limited. It can only model and calculate partially some of the performance measures of queueing systems such as $G^{[x]}/G/m$ (batch arrival, individual service process from different arrival streams with fixed batch size) and $G/G^{[x]}/m$ (individual arrival, batch service process). In the meanwhile, the results are not very good for $m_i > 1$ and multiple arrival streams.

Before introducing our notation for this section, we should mention at this point that some of our notation through the inputs, outputs and equation sections of this chapter might be changed into other formats to be consistent with other new findings, formulas and approaches brought from Chapter 4 to study our complete model formulation in Chapter 5 of this thesis.

3.3.1 Inputs

P = Size of population to be treated (residents)

L = Time allotted for treatment (days)

h = Daily hours of operation (hours per day)

N = Number of clinics

m_i = Number of staff at station i

t_i = Mean process time at station i (minutes)

σ^2_i = Variance of mean service time at station i (minutes^2)

k_i = Processing batch size at station i

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 = Bus arrival size

$\bar{k}_{ai}$ =Aggregate batch arrival size to the station i

c_{a1}^2 = interarrival time SCV at station 1

3.3.2 Outputs

TH' = Required throughput (residents per minute)

m_i' = Minimum staff at station i

$WTBT_i$ = Wait to batch time at station i (minutes)

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

CT_i = Cycle time at station i (minutes)

TCT = Total average time in clinic (minutes)

WIP = Average number of residents in clinic

λ_i = Batch arrival rate at station i (batches per minute)

c_{ai}^2 = Interarrival time SCV at station i

c_{ei}^2 = Process time SCV at station i

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

R= Clinic capacity (residents per minute)

CTq_i = Average queue time at station i (minutes)

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

Q_i = Average queue length at station i

u_i = Utilization at station i

3.3.3 Equations

The throughput required to treat the population in the given time is $TH' = \frac{P}{60LhH}$. If residents arrive individually, the user specifies the arrival variability c_{a1}^2. Else, the individual resident arrival variability is given as $c_{a1}^2 = k_0 - 1$.

All arriving residents go to the first station. We calculate the arrival rates for the other stations based on the routing probabilities:

$$\lambda_i = \begin{cases} TH^{i} & (i = 1) \\ \sum_{j=1}^{i-1} \lambda_j p_{ji} & (i > 1) \end{cases}$$

At each station after the first, we calculate arrival batch size based on the process batch size of the previous stations:

$$\overline{k}_{ai} = \begin{cases} k_0 & (i = 1) \\ \sum_{j=1}^{i-1} k_j \frac{\lambda_j p_{ji}}{\lambda} & (i > 1) \end{cases}$$

We use station arrival rates to determine the minimum staff at each station:

$$m_i' = \frac{\lambda_i t_i}{\bar{k}_{ai}}$$

We then use user-selected staff levels m_i to calculate station utilization:

$$u_i = \frac{\lambda_i t_i}{m_i k_i}$$

We calculate the variability of arrivals, processes, and departures from each station:

$$c_{ai}^2 = \sum_{j=1}^{i-1} \left(\left(c_{di}^2 - 1 \right) p_{ji} + 1 \right) \cdot \frac{\lambda_j p_{ji}}{\lambda_i}$$

$$c_{ei}^2 = \frac{\sigma_i^2}{t_i^2}$$

$$c_{di}^2 = (k_i - 1 + k_i) \left[1 + \left(1 - u_i^{\,2}\right) \left(\frac{c_{ai}^2}{k_i} - 1 \right) + \frac{u_i^{\,2}}{\sqrt{m_i}} \left(c_{ei}^2 - 1 \right) \right]$$

The average time spent waiting at station i depends upon the arrival and process batch sizes; denotes time waiting for service, while $WIBT_i$ represents time waiting in arrival batches and $WTBT_i$ represents time waiting to form a process batch.

$$\bar{k}_{ai} = \sum_{j=1}^{i-1} k_j \frac{\lambda_j p_{ji}}{\lambda}$$

$$CT_{qi} = \left(\frac{c_{ai}^2 + c_{ei}^2}{2} \right) \cdot \left(\frac{u_i^{\sqrt{2m_i + 2} - 1}}{m_i \left(1 - u_i\right)} \right) \cdot \frac{t_i}{m_i} \qquad k_i = 1, \bar{k}_{ai} > 1$$

$$CT_{qi} = \left(\frac{c_{a,i}^2 / k_i + c_{ei}^2}{2} \right) \cdot \left(\frac{u_i^{\sqrt{2m_i + 2} - 1}}{1 - u_i} \right) \cdot \frac{t_i}{m_i} \qquad k_i > 1, \bar{k}_{ai} = 1$$

$$WTBT_i = \frac{k_i - 1}{2\lambda_i}$$

$$WIBT_i = \frac{(\overline{k}_{ai} - 1)t_i}{2m_i}$$

The average time spent traveling to the next station after station *i* depend upon the routing probabilities and the average walking speed: $W_i = \frac{1}{60v}\sum_{j=i+1}^{I+1} p_{ij}d_{ij}$.

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

We weight the station cycle times by their arrival rates to calculate the total average time in clinic: $CT = \frac{1}{\lambda_1}\sum_{i=1}^{I} \lambda_i CT_i$

Other statistics we calculate include clinic capacity, the average queue length at each station, and the average clinic WIP:

$$R = \min_{i=1,\ldots,I} \left\{ \frac{m_i \lambda_1}{t_i \lambda_i} \right\}$$

$$WIP = \lambda_1 \cdot CT$$

$$Q_i = CT_{qi} \lambda_i$$

3.4 Limitations of the existing model

To explain the limitations of the models in this chapter, we note that the proposed formulas for this model are simple and the model itself is not sufficiently complete to satisfy all of our requirements in the mass dispensing and vaccination clinics.

About the formulas we can say that most of them are for the cases with a single server station. Since there has been no further research for the multiple server station cases so far, Treadwell (2006) added some factors to the formulas to create some estimates for multiple server stations. His results for multiple server stations with general cases are not very good.

On other hand, the models in this chapter cannot satisfy some of the cases that are needed to analyze a real clinic completely. For example, in Treadwell (2006), the clinic model doesn't include different types of queueing networks having batch (bulk) arrivals with random size from different arrival streams to a batch or individual processing station performed by multiple parallel servers. Moreover, the model does not include self service stations.

Since we were unable to find previously proposed models that apply to the situation addressed in our clinic model, in Chapters 4 and 5, we intuitively and experimentally introduce some new concepts and methods to create queueing network approximations that build on existing models and include novel contributions as well.

The clinic model in Chapter 5 is more complete than the model in this chapter in following ways:

- We consider arrival batch size variability and its effect on the aggregation process, batch interarrival time variability, and other performance measures for a station.

- We consider departure batch size variability and splitting (branching process) and its effect on the aggregation process, batch interarrival time variability and other performance measures for downstream stations.

- We study in detail the behavior of WIBT and waiting time in queue for multiple server stations with multiple batch arrival streams under the general cases arriving to an individual process station.

- We study in detail the behavior of the batch interarrival time SCV after the batch formation process for a batch process station.

- We study in detail the behavior of self service stations and their interdeparture time variability.

- By studying and analyzing many papers and sources from various branches of queueing theory, we extract some findings for the aggregation process, interdeparture time variability and splitting process which will be suitable for the different kinds of stations introduced in Chapter 5.

Chapter 4: Our new applied approach and results

This chapter presents the results of computational experiments completed to evaluate different estimates for wait-in-batch-time, self service interdeparture time variability, and batch formation process including batch formation variability and estimation for WTBT. In regard to WIBT, we study it under two different cases: first, we study cases when batch size is larger than number of servers; second, we analyze WIBT for more general cases.

Moreover, we will go through analyzing our batch branching approach and its results at the end of the chapter. While these results suggest that some approximations are better than others, we cannot guarantee their accuracy. Additional work would be useful to characterize their accuracy in other scenarios and to seek better approximations for those scenarios where they perform poorly.

The approaches and extracted equations studied in this chapter are needed to thoroughly model mass dispensing and vaccination clinics. All of these findings are employed in the model formulation of Chapter 5 for analyzing different type of queueing systems in the clinics.

In this section, we will define some of our notation needed for a specific purpose. In Chapter 5, which is our final model formulation section, we will bring again some of this notation along with other new notation to be able to have a complete model of a clinic.

4.1 Wait-in-batch-time (batch size larger than number of servers)

This section considers the case with a general arrival process. Residents arrive to the workstation in batches and individually. The arrival batches may come from different batch service process workstations, and the batch sizes from each workstation may vary due to the routing probabilities. There are also individual arrivals from individual service process 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.

A key quantity is the estimate of the wait-in-batch-time, the average time that a resident spends in the batch from the time that the batch "opens" until the resident begins service.

4.1.1 First type of formulas for WIBT

As an important point, we should say that the formula extracted in this section is applicable for the scenarios in which the arrival batch sizes is larger than the number of servers.

We will use the following notation:

m_i = Number of staff at station i

t_i = Mean process time at station i (minutes)

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

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

u_i = Utilization at station *i*

$p_n(i)$ = Steady-state probability of having *n* residents in station *i*.

U_i = Steady-state probability of all of the servers at station *i* being busy

X_i = Average number of residents that wait in the batch at station *i*.

$WIBT_i$ =Average wait in batch time at station *i* (minutes)

We can estimate the wait-in-batch-time for multiple servers (see Section 3.1) as follows[1]:

$$WIBT_i = \frac{(\bar{K}_{Ai} - 1)t_i}{2m_i} \quad \textbf{(Formula 9)}$$

As we will see, this is not a good approximation, so we will derive a new formula for the wait-in-batch-time. To do so, we start by calculating the following terms:

$$u_i = \frac{\lambda_{Ai}\bar{K}_{Ai}t_i}{m_i}$$

$$U_i = \sum_{n=m_i}^{\infty} p_n(i) = 1 - \sum_{n=0}^{m_i-1} p_n(i)$$

It will be useful to note the following:

$$\sum_{n=0}^{m_i-1} np_n(i) = m_i u_i - m_i \sum_{n=m_i}^{\infty} p_n(i) = m_i\left(u_i - U_i\right)$$

[1] Formula 9 is exactly Formula 8 but in terms of the new notation introduced in Chapter 4.

If, when the batch arrives, the number of residents, who are already in the system, is greater than or equal to the number of servers, all of the servers are busy, so the batch waits in the queue. Eventually, the batch is at the head of the queue and one of the servers completes a resident. Then the batch opens, one resident begins service without waiting in batch, and all of the others wait in the batch.

If, when the batch arrives, the number of residents, who are already in the system, is less than the number of servers, one or more servers are idle, so the batch opens and one or more residents begin service immediately.

From this we estimate X_i as follows:

$$\begin{aligned} X_i &= \sum_{n=0}^{m_i-1} p_n(i)\left(\bar{K}_{Ai} - m_i + n\right) + \sum_{n=m_i}^{\infty} p_n(i)\left(\bar{K}_{Ai} - 1\right) \\ &= \bar{K}_{Ai} - m_i\left(1 - U_i\right) + m_i\left(u_i - U_i\right) - U_i \\ &= \bar{K}_{Ai} - m_i + m_i u_i - U_i \end{aligned}$$

Thus, $\bar{K}_{Ai} - X_i$ residents go to servers immediately. For them the wait-in-batch-time is zero. Assuming that the servers, when busy, complete a resident every $\frac{t_i}{m_i}$ minute, the first resident of those remaining must wait-in-batches for $\frac{t_i}{m_i}$ minutes. The second waits $\frac{2t_i}{m_i}$ minutes, and so forth. The last resident in the batch waits for $\frac{X_i t_i}{m_i}$ minutes. Then we can estimate the average wait-in-batch-time as follows:

$$WIBT_i = \frac{1}{\bar{K}_{Ai}} \sum_{n=1}^{X_i} \frac{nt_i}{m_i} = \frac{X_i\left(X_i+1\right)}{2\bar{K}_{Ai}} \cdot \frac{t_i}{m_i}$$

$$= \frac{\left(\bar{K}_{Ai} - m_i + m_i u_i - U_i\right)\left(\bar{K}_{Ai} - m_i + m_i u_i - U_i + 1\right)}{2\bar{K}_{Ai}} \cdot \frac{t_i}{m_i} \qquad \textbf{(Formula 10)}$$

The only remaining task is to estimate U_i. Following Shore (1988) and dropping the station subscript for the moment, we let $E(N_c)$ be the mean number of customers in the system and $E(N_1)$ be the mean number in of customers in the corresponding GI/G/1 queue having the same traffic intensity.

$$E(N_c) = m_i u_i + \left[\frac{u_i^{\sqrt{2m_i+2}}}{1-u_i}\right]\left[\frac{c^2{}_{ai} + c^2{}_{ei}}{2}\right]$$

$$E(N_1) = m_i u_i + \left[\frac{u_i^{2}}{1-u_i}\right]\left[\frac{c^2{}_{ai} + c^2{}_{ei}}{2}\right]$$

Shore (1988) shows that

$$U_i = u_i(E(N_c) - m_i u_i)/(E(N_1) - u_i)$$

From this, we intuitively extracted that $U_i = u_i^{\sqrt{2m_i+2}-1}$. Since this is not affected by the arrival variability, we will use this result for our batch arrival case. Going back to the original notation, we have

$$U_i = u_i^{\sqrt{2m_i+2}-1}$$

4.1.2 Wait-in-batch-time experiments

To evaluate Formulas 9 and 10, we conducted a set of computational experiments using a discrete-event simulation model of the station. In the simulation model, batches hold in a queue until a server becomes available, at which point they are "opened" and individual entities enter the server's queue (Figure 3). This extra step in the simulation logic allows the components of waiting time to be examined separately.

To explain in more detail, after releasing the batches from holding area by finding at least an idle server, they split to individual entities and the average time that each of these entities should wait in queue (broken batches) until it gets served, is defined as wait-in-batch-time (WIBT). We can understand easily that for the at least the first entity in each broken batches there is the WIBT of zero, since it goes directly to an idle server after breaking the batches.

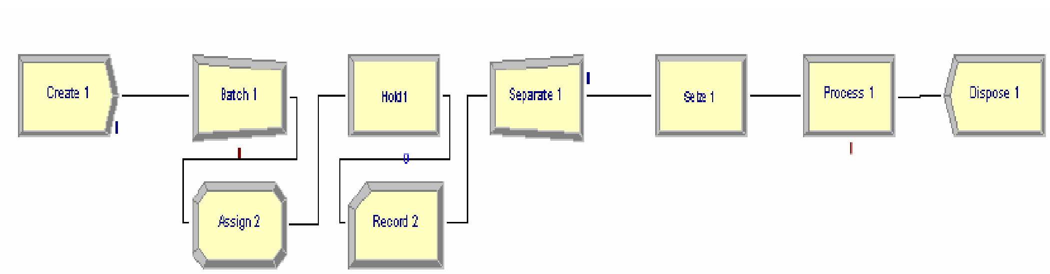

Figure 3. Simulation logic for dividing "waiting time" into queue time as a batch and WIBT.

Throughout the simulation experiments, each scenario had the arrival batches with fixed size (either 5 or 20), and the interarrival times were exponentially distributed. The mean interarrival time varied from 0.1684 minutes to 0.3333 minutes. The distribution of the processing times was an exponential distribution or a gamma distribution. For the exponential distributions, the mean was either 0.0333 minutes or 0.10 minutes. For the gamma distributions, α was always 0.5, while β was set to 0.0167, 0.050, 0.0667, and 0.20. The number of servers was either 1 or 3. Table 3, 4 and 5 describe the scenarios.

Table 3. Scenarios with exponentially distributed process times.

Scenario	Batch size	Mean Interarrival Time (mins)	Mean Processing Time (mins)	Number of servers
E-5-1-99	5	0.1684	0.0333	1
E-5-1-95	5	0.1754	0.0333	1
E-5-1-90	5	0.1852	0.0333	1
E-5-1-80	5	0.2083	0.0333	1
E-5-1-50	5	0.3333	0.0333	1
E-5-3-99	5	0.1684	0.1000	3
E-5-3-95	5	0.1754	0.1000	3
E-5-3-90	5	0.1852	0.1000	3
E-5-3-80	5	0.2083	0.1000	3
E-5-3-50	5	0.3333	0.1000	3

Table 4. Scenarios with gamma distribution process times and 1 server.

Scenario	Batch size	Mean Interarrival Time (mins)	Mean Processing Time (mins)	Number of servers
G-5-1-99	5	0.1684	0.0333	1
G-5-1-95	5	0.1754	0.0333	1
G-5-1-90	5	0.1852	0.0333	1
G-5-1-80	5	0.2083	0.0333	1
G-5-1-50	5	0.3333	0.0333	1
G-20-1-99	20	0.1684	0.0083	1
G-20-1-95	20	0.1754	0.0083	1
G-20-1-90	20	0.1852	0.0083	1
G-20-1-80	20	0.2083	0.0083	1
G-20-1-50	20	0.3333	0.0083	1

Table 5. Scenarios with gamma distribution process times and 3 servers.

Scenario	Batch size	Mean Interarrival Time (mins)	Mean Processing Time (mins)	Number of servers
G-5-3-99	5	0.1684	0.1000	3
G-5-3-95	5	0.1754	0.1000	3
G-5-3-90	5	0.1852	0.1000	3
G-5-3-80	5	0.2083	0.1000	3
G-5-3-50	5	0.3333	0.1000	3
G-20-3-99	20	0.1684	0.0250	3
G-20-3-95	20	0.1754	0.0250	3
G-20-3-90	20	0.1852	0.0250	3
G-20-3-80	20	0.2083	0.0250	3
G-20-3-50	20	0.3333	0.0250	3

For each scenario, we ran a simulation model with 100 replications, each 30,000 minutes long with a warm-up period of 25,000 minutes. From the simulation model we could calculate the average wait-in-batch-time of residents. We also used Formula 9 and Formula 10 to estimate the average wait-in-batch-time. Tables 6, 7, and 8 show the results.

For each scenario, the table lists the average wait-in-batch-time from the simulation model, the estimate from Formula 9, and the estimate from Formula 10. Also listed are the relative errors for the estimates. We see that Formula 10 provides a much better estimate than Formula 9.

Table 6. Results for scenarios with exponentially distributed process times.

Scenario	WIBT from simulation (mins)	WIBT from Formula 9 (mins)	Relative error, Formula 9	WIBT from Formula 10 (mins)	Relative error, Formula 10
E-5-1-99	0.0667	0.0667	0.050%	0.0667	0.050%
E-5-1-95	0.0667	0.0667	0.050%	0.0667	0.050%
E-5-1-90	0.0667	0.0667	0.050%	0.0667	0.050%
E-5-1-80	0.0667	0.0667	0.050%	0.0667	0.050%
E-5-1-50	0.0667	0.0667	0.050%	0.0667	0.050%
E-5-3-99	0.0660	0.0667	1.010%	0.0663	0.475%
E-5-3-95	0.0660	0.0667	1.010%	0.0649	1.720%
E-5-3-90	0.0600	0.0667	11.111%	0.0630	4.959%
E-5-3-80	0.0600	0.0667	11.111%	0.0590	1.748%
E-5-3-50	0.0480	0.0667	38.889%	0.0453	5.717%

Table 7. Results for scenarios with Gamma distributed process times with 1 server

Scenario	WIBT from simulation (mins)	WIBT from Formula 9 (mins)	Relative error, Formula 9	WIBT from Formula 10 (mins)	Relative error, Formula 10
G-5-1-99	0.0665	0.0667	0.251%	0.0667	0.251%
G-5-1-95	0.0665	0.0667	0.251%	0.0667	0.251%
G-5-1-90	0.0665	0.0667	0.251%	0.0667	0.251%
G-5-1-80	0.0665	0.0667	0.251%	0.0667	0.251%
G-5-1-50	0.0665	0.0667	0.251%	0.0667	0.251%
G-20-1-99	0.0790	0.0792	0.211%	0.0792	0.211%
G-20-1-95	0.0790	0.0792	0.211%	0.0792	0.211%
G-20-1-90	0.0790	0.0792	0.211%	0.0792	0.211%
G-20-1-80	0.0790	0.0792	0.211%	0.0792	0.211%
G-20-1-50	0.0790	0.0792	0.211%	0.0792	0.211%

Table 8. Results for scenarios with Gamma distributed process times with 3 servers

Scenario	WIBT from simulation (mins)	WIBT from Formula 9 (mins)	Relative error, Formula 9	WIBT from Formula 10 (mins)	Relative error, Formula 10
G-5-3-99	0.0660	0.0667	1.010%	0.0663	0.475%
G-5-3-95	0.0643	0.0667	3.681%	0.0649	0.878%
G-5-3-90	0.0621	0.0667	7.354%	0.0630	1.409%
G-5-3-80	0.0576	0.0667	15.741%	0.0590	2.346%
G-5-3-50	0.0429	0.0667	55.400%	0.0453	5.491%
G-20-3-99	0.0787	0.0792	0.593%	0.0729	7.373%
G-20-3-95	0.0779	0.0792	1.626%	0.0729	6.422%
G-20-3-90	0.0770	0.0792	2.814%	0.0729	5.328%
G-20-3-80	0.0750	0.0792	5.556%	0.0729	2.803%
G-20-3-50	0.0688	0.0792	15.068%	0.0729	5.956%

4.2 Wait-in-batch-time (general case)

We now consider cases in which the batch arrival size is less than number of servers. Formula 10 cannot be acceptable for this case. That is why we are trying to seek a new method and formula. Moreover, we seek a new formula that will be useful for any number of servers.

4.2.1 Second type of formulas for WIBT

In order to come up with an universal formula than can satisfy all of scenarios, our methodology is to run simulations with various specifications for all the cases to compare the results, and to extract a general formula for WIBT among the possible variables such as batch size, number of servers, process time and utilization.

It should be said that we will again use exactly the notation and relationships from Section 4.1.

4.2.2 Wait-in-batch-time experiments

In this section, we conducted 3 different types of computational experiments using a discrete-event simulation model of the station. Each type of experiment consists of a number of sets and each set includes several scenarios. Experiment type one consisted of 8 sets that we named set 1 to set 8. Experiment type two included 2 sets that we named sets 9 and 10. Finally, experiment type three used sets 11 and 12.

The purpose of carrying out these experiment types is finding the approximate behavior of WIBT versus factors such as the number of servers, utilization, arrival batch size and process time to come up with a general WIBT formula that corresponds to all of the cases.

For each scenario, we ran a simulation model with 10 replications and a confidence interval of 95%, each 1,000,000 minutes long with the warm-up periods of 500,000 minutes.

As we said in previous section, a key issue is the estimate of the wait-in-batch-time, the average time that a resident spends in the batch from the time that the batch "opens" until the resident begins service. In the simulation, batches hold in a queue until a server becomes available (waiting time in queue), at which point they are "opened" and individual entities enter the server's queue.

4.2.2.1 First type of experiment

In the first type of simulation, we have different sets with a constant number of servers, processing time, and utilization. The arrival batch size and batch arrival rate are the variables within each set of scenarios. The purpose of this experiment is to find a relationship between WIBT and the arrival batch size.

Among the 8 sets of simulation of scenarios, the utilization ranges from 25% up to 93%. In each set, the arrival batch size varied from 1 to 13 or 16, the number of servers had a one of the fixed size of 3, 4, 6, 8, 10, 12, and the interarrival times and processing time were exponentially distributed.

Tables 9 to 16 show the average wait-in-batch-time from the simulation model with its upper and lower bound of 95% of confidence interval for each scenario in Set 1 to 8. The tables also describe other specifications for the scenarios and the name of the scenarios.

Additionally, Figures 4 to 11 demonstrate the average wait-in-batch-time from the simulations for each set. Since the difference between upper and lower bound is small especially for the low utilization sets, we didn’t show them in these figures.

Table 9. Specifications and simulations results for Scenarios 1-1-1 to 1-1-13 (Set 1)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	WIBT From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-1-1	0.50	4	1	25.00%	2	0	0	0.007
1-1-2	0.25	4	2	25.00%	2	0.03	0.002	0.058
1-1-3	0.16	4	3	25.00%	2	0.104	0.081	0.127
1-1-4	0.12	4	4	25.00%	2	0.191	0.091	0.291
1-1-5	0.1	4	5	25.00%	2	0.337	0.247	0.427
1-1-6	0.084	4	6	25.00%	2	0.511	0.451	0.571
1-1-7	0.072	4	7	25.00%	2	0.713	0.624	0.802
1-1-8	0.0625	4	8	25.00%	2	0.919	0.917	0.921
1-1-9	0.056	4	9	25.00%	2	1.142	1.032	1.252
1-1-10	0.05	4	10	25.00%	2	1.367	1.237	1.497
1-1-11	0.046	4	11	25.00%	2	1.594	1.454	1.734
1-1-12	0.042	4	12	25.00%	2	1.826	1.726	1.926
1-1-13	0.039	4	13	25.00%	2	2.054	1.914	2.194

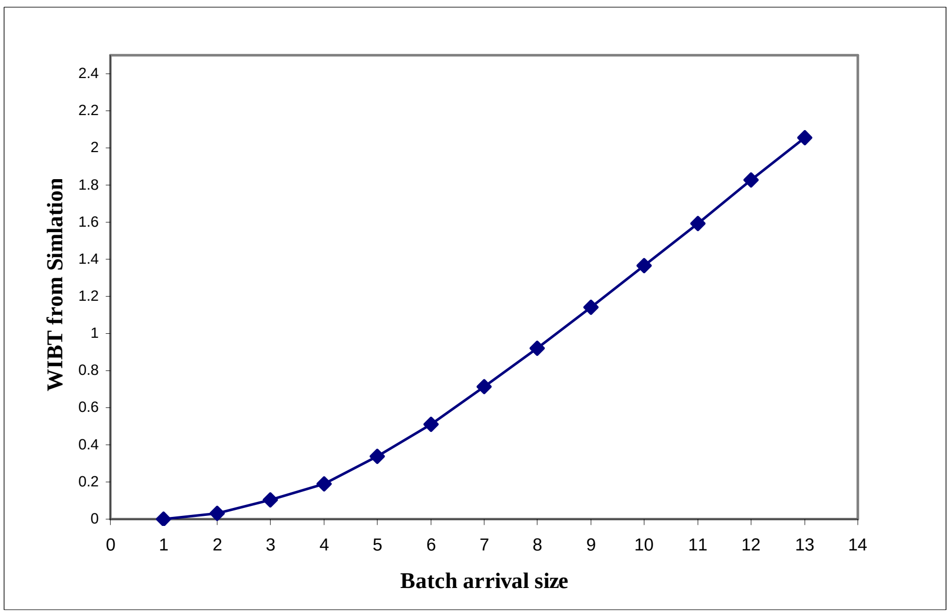

Figure 4. Simulation results for Scenarios 1-1-1 to 1-1-13 (Set 1)

Table 10. Specifications and simulations results for Scenarios 1-2-1 to 1-2-13 (Set 2)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	WIBT From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-2-1	0.50	6	1	50.00%	6	0	0	0.009
1-2-2	0.25	6	2	50.00%	6	0.126	0.056	0.196
1-2-3	0.16	6	3	50.00%	6	0.327	0.304	0.35
1-2-4	0.12	6	4	50.00%	6	0.586	0.486	0.686
1-2-5	0.1	6	5	50.00%	6	0.907	0.817	0.997
1-2-6	0.084	6	6	50.00%	6	1.185	1.125	1.245
1-2-7	0.072	6	7	50.00%	6	1.554	1.465	1.643
1-2-8	0.0625	6	8	50.00%	6	1.944	1.942	1.946
1-2-9	0.056	6	9	50.00%	6	2.343	2.233	2.453
1-2-10	0.05	6	10	50.00%	6	2.76	2.63	2.89
1-2-11	0.046	6	11	50.00%	6	3.203	3.063	3.343
1-2-12	0.042	6	12	50.00%	6	3.657	3.557	3.757
1-2-13	0.039	6	13	50.00%	6	4.103	3.963	4.243

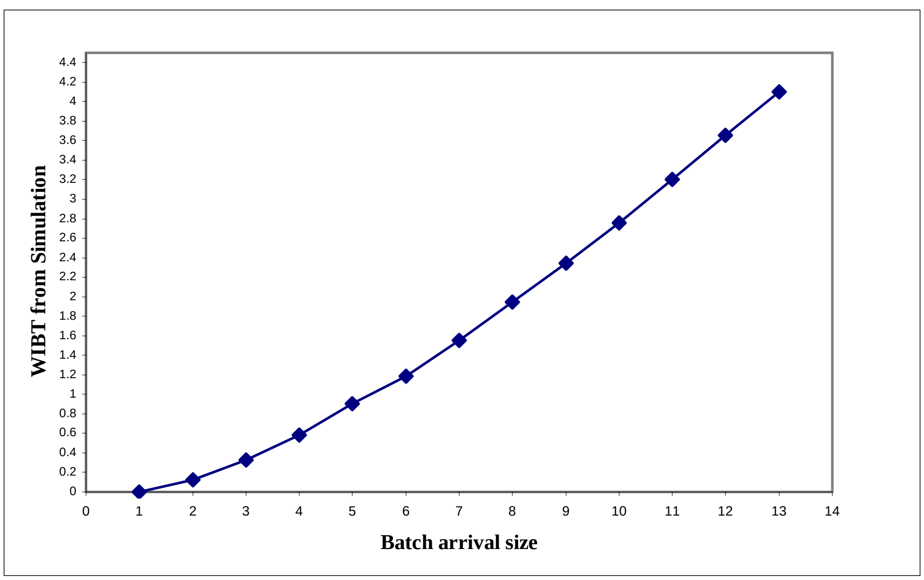

Figure 5. Simulation results for Scenarios 1-2-1 to 1-2-13 (Set 2)

Table 11. Specifications and simulations results for Scenarios 1-3-1 to 1-3-13 (Set 3)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	WIBT From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-3-1	0.50	3	1	58.33%	3.5	0	0	0.002
1-3-2	0.25	3	2	58.33%	3.5	0.337	0.267	0.337
1-3-3	0.17	3	3	58.33%	3.5	0.733	0.68	0.733
1-3-4	0.13	3	4	58.33%	3.5	1.192	1.092	1.192
1-3-5	0.10	3	5	58.33%	3.5	1.708	1.618	1.708
1-3-6	0.08	3	6	58.33%	3.5	2.234	2.174	2.234
1-3-7	0.07	3	7	58.33%	3.5	2.77	2.671	2.77
1-3-8	0.06	3	8	58.33%	3.5	3.337	3.247	3.337
1-3-9	0.06	3	9	58.33%	3.5	3.883	3.773	3.883
1-3-10	0.05	3	10	58.33%	3.5	4.464	4.414	4.464
1-3-11	0.05	3	11	58.33%	3.5	4.994	4.874	4.994
1-3-12	0.04	3	12	58.33%	3.5	5.583	5.433	5.583
1-3-13	0.04	3	13	58.33%	3.5	6.138	5.998	6.138

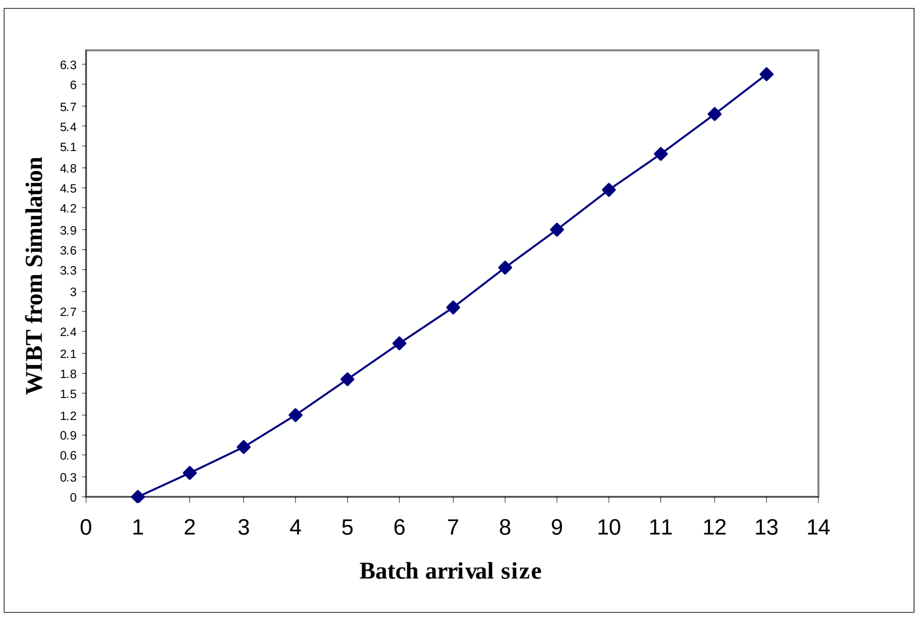

Figure 6. Simulation results for Scenarios 1-3-1 to 1-3-13 (Set 3)

Table 12. Specifications and simulations results for Scenarios 1-4-1 to 1-4-16 (Set 4)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	WIBT From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-4-1	0.75	12	1	62.50%	10	0	0	0.023
1-4-2	0.38	12	2	62.50%	10	0.09	0.01	0.17
1-4-3	0.25	12	3	62.50%	10	0.235	0.145	0.325
1-4-4	0.19	12	4	62.50%	10	0.423	0.363	0.483
1-4-5	0.15	12	5	62.50%	10	0.639	0.55	0.728
1-4-6	0.13	12	6	62.50%	10	0.878	0.876	0.88
1-4-7	0.11	12	7	62.50%	10	1.142	1.032	1.252
1-4-8	0.09	12	8	62.50%	10	1.426	1.296	1.556
1-4-9	0.08	12	9	62.50%	10	1.699	1.559	1.839
1-4-10	0.08	12	10	62.50%	10	2.013	1.913	2.113
1-4-11	0.07	12	11	62.50%	10	2.31	2.08	2.54
1-4-12	0.06	12	12	62.50%	10	2.632	2.432	2.832
1-4-13	0.06	12	13	62.50%	10	2.942	2.762	3.122
1-4-14	0.05	12	14	62.50%	10	3.281	3.121	3.441
1-4-15	0.05	12	15	62.50%	10	3.643	3.503	3.783
1-4-16	0.05	12	16	62.50%	10	4.013	3.903	4.123

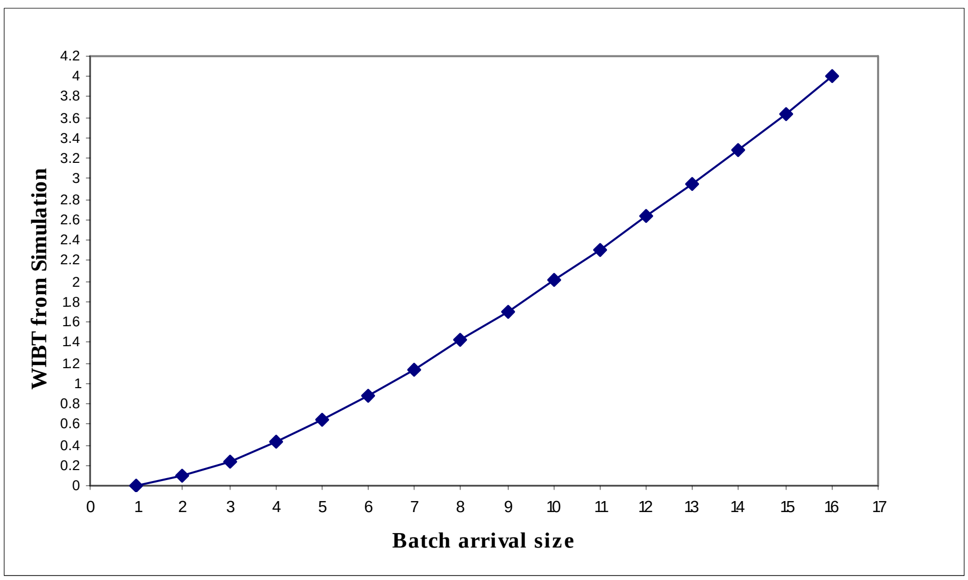

Figure 7. Simulation results for Scenarios 1-4-1 to 1-4-16 (Set 4)

Table 13. Specifications and simulation results for Scenarios 1-5-1 to 1-5-13 (Set 5)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	WIBT From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-5-1	0.50	4	1	68.75%	5.5	0	0	0.007
1-5-2	0.25	4	2	68.75%	5.5	0.416	0.346	0.486
1-5-3	0.17	4	3	68.75%	5.5	0.912	0.822	1.002
1-5-4	0.13	4	4	68.75%	5.5	1.453	1.353	1.553
1-5-5	0.10	4	5	68.75%	5.5	2.001	1.881	2.121
1-5-6	0.08	4	6	68.75%	5.5	2.6	2.46	2.74
1-5-7	0.07	4	7	68.75%	5.5	3.221	3.121	3.321
1-5-8	0.06	4	8	68.75%	5.5	3.899	3.81	3.988
1-5-9	0.06	4	9	68.75%	5.5	4.552	4.362	4.742
1-5-10	0.05	4	10	68.75%	5.5	5.166	4.956	5.376
1-5-11	0.05	4	11	68.75%	5.5	5.86	5.58	6.14
1-5-12	0.04	4	12	68.75%	5.5	6.523	6.353	6.693
1-5-13	0.04	4	13	68.75%	5.5	7.157	6.857	7.457

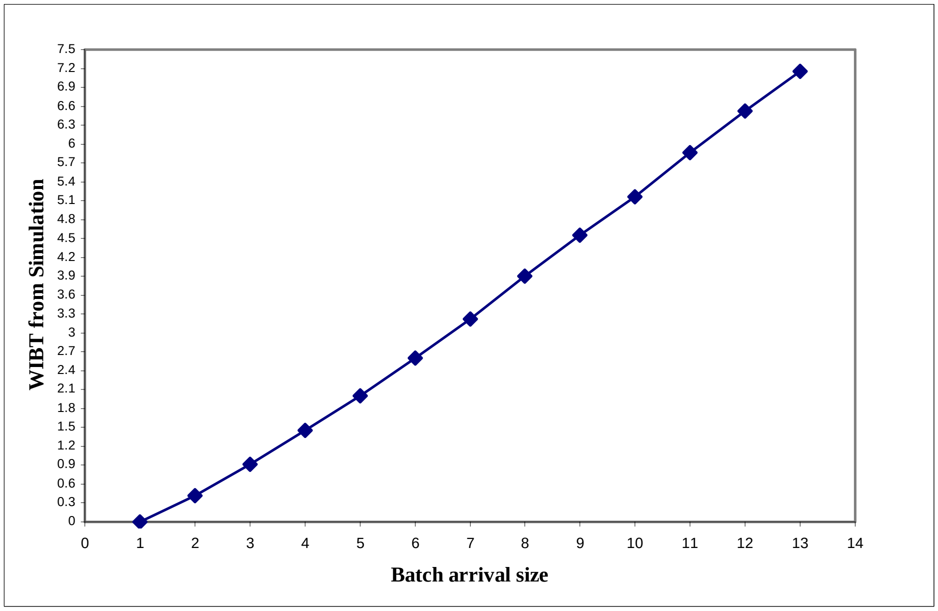

Figure 8. Simulation results for Scenarios 1-5-1 to 1-5-13 (Set 5)

Table 14. Specifications and simulations results for Scenarios 1-6-1 to 1-6-13 (Set 6)

Scenario name	**Batch arrival rate per min**	**Number of servers**	**Arrival batch size**	**Utilization**	**Mean Process time(min)**	**WIBT From simulation**	**Lower bound for 95% confidence interval**	**Upper bound for 95% confidence interval**
1-6-1	0.75	10	1	75.00%	10	0	0	0.087
1-6-2	0.38	10	2	75.00%	10	0.23	0.14	0.32
1-6-3	0.25	10	3	75.00%	10	0.524	0.424	0.624
1-6-4	0.19	10	4	75.00%	10	0.86	0.74	0.98
1-6-5	0.15	10	5	75.00%	10	1.223	1.033	1.413
1-6-6	0.13	10	6	75.00%	10	1.577	1.478	1.676
1-6-7	0.11	10	7	75.00%	10	1.986	1.886	2.086
1-6-8	0.09	10	8	75.00%	10	2.382	2.202	2.562
1-6-9	0.08	10	9	75.00%	10	2.813	2.583	3.043
1-6-10	0.08	10	10	75.00%	10	3.222	2.922	3.522
1-6-11	0.07	10	11	75.00%	10	3.67	3.38	3.96
1-6-12	0.06	10	12	75.00%	10	4.089	3.989	4.189
1-6-13	0.06	10	13	75.00%	10	4.538	4.368	4.708

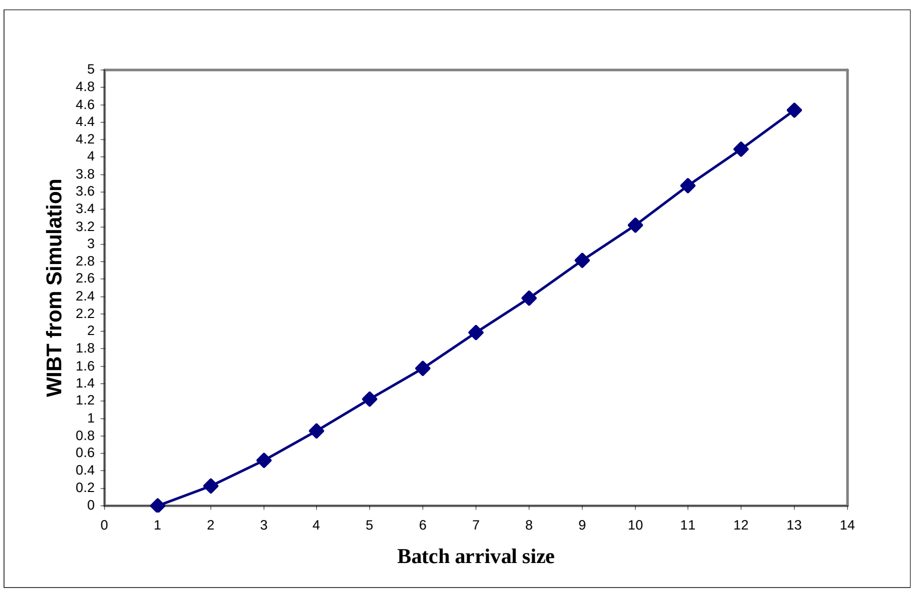

Figure 9. Simulation results for Scenarios 1-6-1 to 1-6-13 (Set 6**)**

Table 15. Specifications and simulations results for Scenarios 1-7-1 to 1-7-13 (Set 7)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	WIBT From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-7-1	0.50	4	1	81.25%	6.5	0	0	0.009
1-7-2	0.25	4	2	81.25%	6.5	0.618	0.598	0.638
1-7-3	0.17	4	3	81.25%	6.5	1.312	1.295	1.329
1-7-4	0.13	4	4	81.25%	6.5	2.006	1.916	2.096
1-7-5	0.10	4	5	81.25%	6.5	2.732	2.632	2.832
1-7-6	0.08	4	6	81.25%	6.5	3.501	3.381	3.621
1-7-7	0.07	4	7	81.25%	6.5	4.252	4.102	4.402
1-7-8	0.06	4	8	81.25%	6.5	5.008	4.888	5.128
1-7-9	0.06	4	9	81.25%	6.5	5.759	5.66	5.858
1-7-10	0.05	4	10	81.25%	6.5	6.598	6.464	6.732
1-7-11	0.05	4	11	81.25%	6.5	7.378	7.148	7.608
1-7-12	0.04	4	12	81.25%	6.5	8.233	7.893	8.573
1-7-13	0.04	4	13	81.25%	6.5	8.993	8.763	9.223

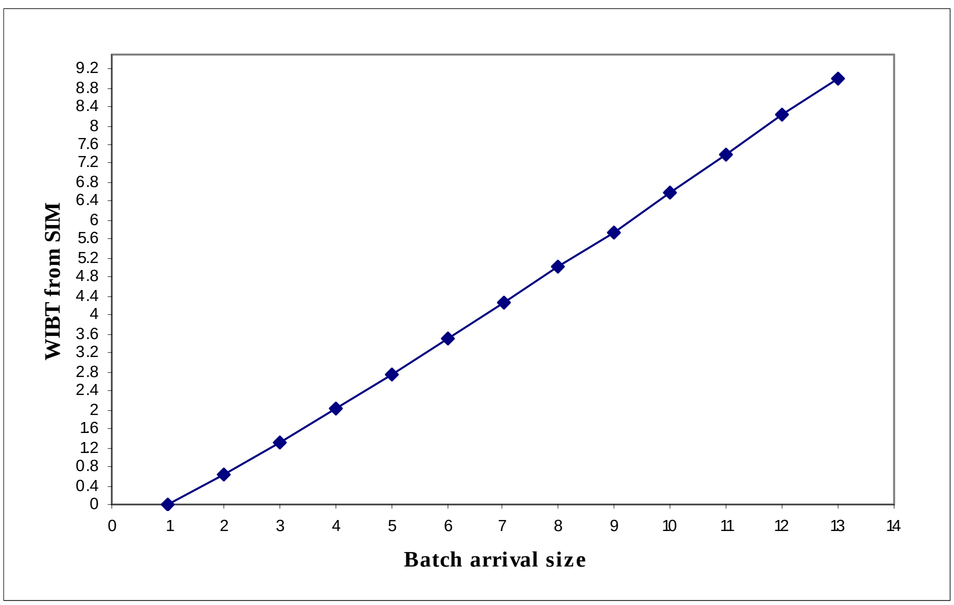

Figure 10. Simulation results for Scenarios 1-7-1 to 1-7-13 (Set 7)

Table 16. Specifications and simulations results for Scenarios 1-8-1 to 1-8-13 (Set 8)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	WIBT From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-8-1	0.75	8	1	93.75%	10	0	0	0.009
1-8-2	0.38	8	2	93.75%	10	0.55	0.48	0.62
1-8-3	0.25	8	3	93.75%	10	1.103	1.08	1.126
1-8-4	0.19	8	4	93.75%	10	1.707	1.607	1.807
1-8-5	0.15	8	5	93.75%	10	2.269	2.179	2.359
1-8-6	0.13	8	6	93.75%	10	2.863	2.803	2.923
1-8-7	0.11	8	7	93.75%	10	3.455	3.366	3.544
1-8-8	0.09	8	8	93.75%	10	4.064	4.062	4.066
1-8-9	0.08	8	9	93.75%	10	4.667	4.557	4.777
1-8-10	0.08	8	10	93.75%	10	5.269	5.139	5.399
1-8-11	0.07	8	11	93.75%	10	5.883	5.743	6.023
1-8-12	0.06	8	12	93.75%	10	6.536	6.436	6.636
1-8-13	0.06	8	13	93.75%	10	7.137	6.997	7.277

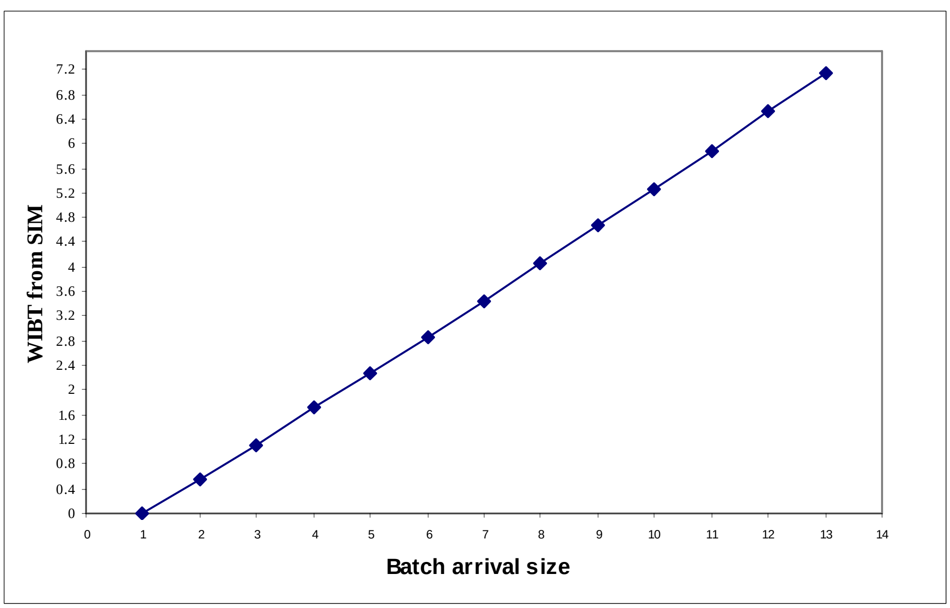

Figure 11. Simulation results for Scenarios 1-8-1 to 1-8-13 (Set 8)

4.2.2.2 Second type of experiment

In the second type of simulation, we have 2 different sets with constant arrival batch sizes, processing time and utilization. On the other hand, our variable here is the number of servers, which changes in each scenario. The purpose of this experiment is to find a relationship between the WIBT and the number of servers.

In this experiment, the utilization is either 60% or 80%. In each set, the number of servers varied from 1 to 13, the arrival batch was 4 or 6, and the interarrival times and processing time were exponentially distributed.

Table 17 and 18 show the average wait-in-batch-time from the simulation model with its upper and lower bound of 95% of confidence interval for each scenario. The tables also describe other simulations' specifications of the scenarios and the name of the scenarios.

Additionally, Figures 12 and 13 demonstrate the average wait-in-batch-time from simulations for each set (sets 9 and 10). Since the difference between upper and lower bound is small especially for the low utilization sets, we don't show them in these figures.

Table 17. Specifications and simulations results for Scenarios 2-1-1 to 2-1-13 (Set 9)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	WIBT From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
2-1-1	0.10	1	6	60.00%	1	2.495	1.885	3.105
2-1-2	0.20	2	6	60.00%	1	1.099	0.889	1.309
2-1-3	0.30	3	6	60.00%	1	0.646	0.583	0.709
2-1-4	0.40	4	6	60.00%	1	0.434	0.405	0.463
2-1-5	0.50	5	6	60.00%	1	0.315	0.225	0.405
2-1-6	0.60	6	6	60.00%	1	0.241	0.191	0.291
2-1-7	0.70	7	6	60.00%	1	0.192	0.142	0.242
2-1-8	0.80	8	6	60.00%	1	0.158	0.088	0.228
2-1-9	0.90	9	6	60.00%	1	0.132	0.112	0.152
2-1-10	1.00	10	6	60.00%	1	0.11	0.1	0.12
2-1-11	1.10	11	6	60.00%	1	0.094	0.085	0.103
2-1-12	1.20	12	6	60.00%	1	0.082	0.0729	0.0911
2-1-13	1.30	13	6	60.00%	1	0.071	0.069	0.073

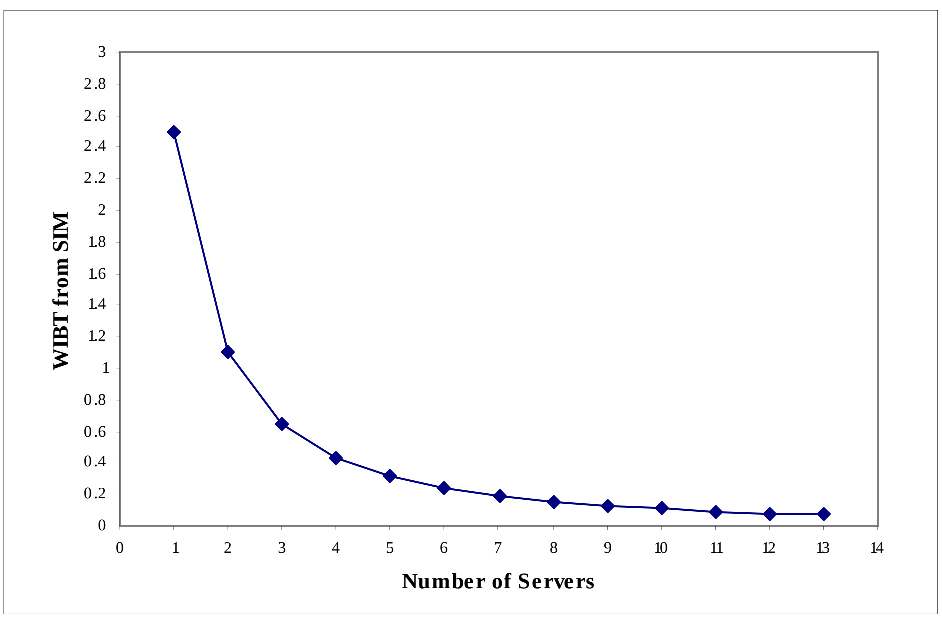

Figure 12. Simulation results for Scenarios 2-1-1 to 2-1-13 (Set 9)

Table 18. Specifications and simulations results for Scenarios 2-2-1 to 2-2-13 (Set 10)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	WIBT From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
2-2-1	0.20	1	4	80.00%	1	1.504	1.394	1.614
2-2-2	0.40	2	4	80.00%	1	0.689	0.599	0.779
2-2-3	0.60	3	4	80.00%	1	0.427	0.404	0.45
2-2-4	0.80	4	4	80.00%	1	0.303	0.213	0.393
2-2-5	1.00	5	4	80.00%	1	0.232	0.152	0.312
2-2-6	1.20	6	4	80.00%	1	0.188	0.178	0.198
2-2-7	1.40	7	4	80.00%	1	0.156	0.146	0.166
2-2-8	1.60	8	4	80.00%	1	0.129	0.079	0.179
2-2-9	1.80	9	4	80.00%	1	0.11	0.09	0.13
2-2-10	2.00	10	4	80.00%	1	0.098	0.088	0.108
2-2-11	2.20	11	4	80.00%	1	0.088	0.079	0.097
2-2-12	2.40	12	4	80.00%	1	0.075	0.0659	0.0841
2-2-13	2.60	13	4	80.00%	1	0.071	0.069	0.073

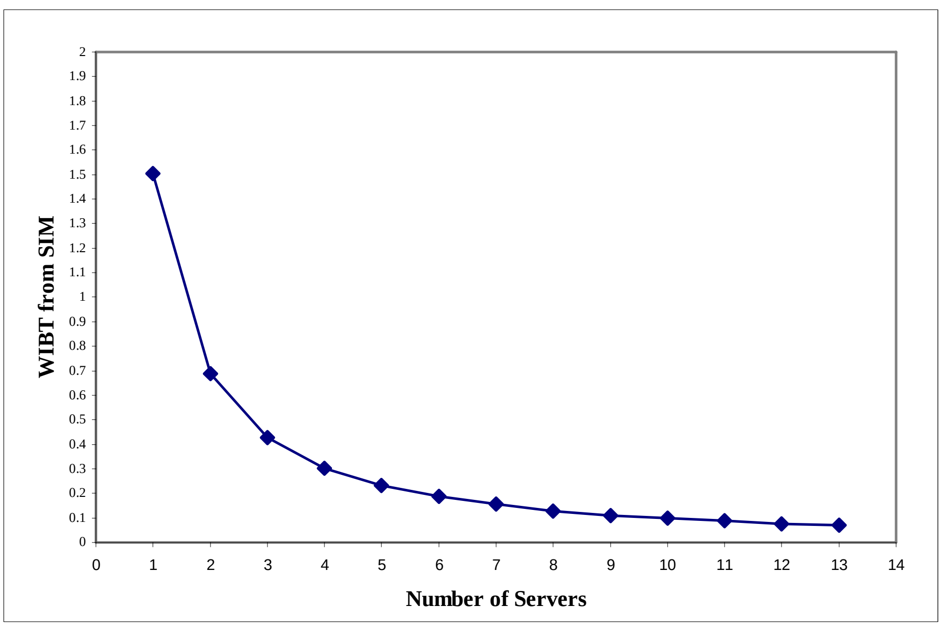

Figure 13. Simulation results for Scenarios 2-2-1 to 2-2-13 (Set 10)

4.2.2.3 Third type of experiment

In the third type of simulation, we have 2 different sets with constant arrival batch sizes, processing time .On the other hand, our variable here is the number of servers and utilization, which changes in each scenario. Since the batch interarrival time doesn't vary as we had in experiment type 2, the utilization is a variable in addition to the number of the servers.

The purpose of this experiment is to find a relationship between the WIBT, the number of servers and the changes in utilization which are variables here.

In this experiment, the batch arrival rate of either 0.2 or 0.15 (batch/min). In each set, the arrival batches size varied from 1 to 13, the arrival batch size was 4 or 6 and the interarrival times and processing time were exponentially distributed.

Table 19 and 20 show the average wait-in-batch-time from the simulation model with its upper and lower bound of 95% of confidence interval for each scenario. The tables also describe other simulations specifications of the scenarios and the name of the scenarios. Moreover, Figure 14 and 15 demonstrate the average wait-in-batch-time from simulations for Set 11 and 12.

Since the difference between upper and lower bound is small especially for the low utilization sets, we don't show them in these figures.

Table 19. Specifications and simulations results for Scenarios 3-1-1 to 3-1-13 (Set 11)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	WIBT From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
3-1-1	0.20	1	4	96.00%	1.2	1.799	1.699	1.899
3-1-2	0.20	2	4	48.00%	1.2	0.692	0.593	0.791
3-1-3	0.20	3	4	32.00%	1.2	0.28	0.190	0.370
3-1-4	0.20	4	4	24.00%	1.2	0.11	0.080	0.140
3-1-5	0.20	5	4	19.20%	1.2	0.053	0.043	0.063
3-1-6	0.20	6	4	16.00%	1.2	0.023	0.013	0.033
3-1-7	0.20	7	4	13.71%	1.2	0.01	0.009	0.011
3-1-8	0.20	8	4	12.00%	1.2	0.004	0.002	0.006
3-1-9	0.20	9	4	10.67%	1.2	0.002	0	0.004
3-1-10	0.20	10	4	9.60%	1.2	0.001	0	0.002
3-1-11	0.20	11	4	8.73%	1.2	0	0	0
3-1-12	0.20	12	4	8.00%	1.2	0	0	0
3-1-13	0.20	13	4	7.38%	1.2	0	0	0

Figure 14. Simulation results for Scenarios 3-1-1 to 3-1-13 (Set 11)

Table 20. Specifications and simulations results for Scenarios 3-2-1 to 3-2-13 (Set 12)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	WIBT From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
3-2-1	0.15	1	6	90.00%	1	2.498	2.198	2.798
3-2-2	0.15	2	6	45.00%	1	1.036	0.837	1.235
3-2-3	0.15	3	6	30.00%	1	0.496	0.396	0.596
3-2-4	0.15	4	6	22.50%	1	0.244	0.144	0.344
3-2-5	0.15	5	6	18.00%	1	0.117	0.037	0.197
3-2-6	0.15	6	6	15.00%	1	0.055	0.035	0.075
3-2-7	0.15	7	6	12.86%	1	0.031	0.022	0.040
3-2-8	0.15	8	6	11.25%	1	0.017	0.008	0.026
3-2-9	0.15	9	6	10.00%	1	0.009	0.007	0.011
3-2-10	0.15	10	6	9.00%	1	0.004	0.003	0.005
3-2-11	0.15	11	6	8.18%	1	0.002	0.001	0.003
3-2-12	0.15	12	6	7.50%	1	0.001	0	0.002
3-2-13	0.15	13	6	6.92%	1	0.001	0	0.002

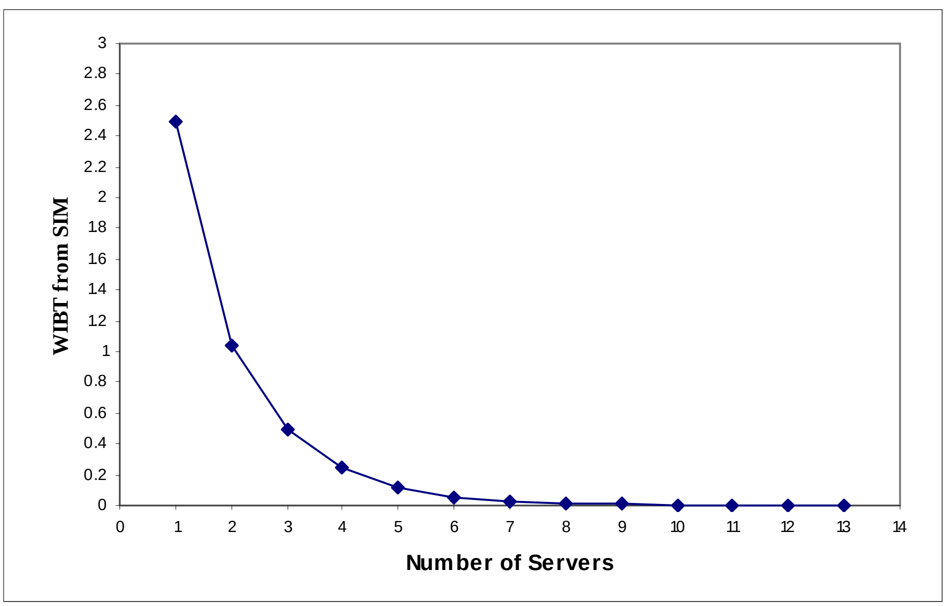

Figure 15. Simulation results for Scenarios 3-2-1 to 3-2-13 (Set 12)

4.2.3 Analysis of results from experiments

In this section, we try to extract some logical formulas and relationships among the observed WIBT and variables such as number of servers, batch arrival size, utilization and processing time. Our main goal is to find formulas satisfying all the scenarios for both $\bar{K}_{Ai} \geq m_i$ and $\bar{K}_{Ai} < m_i$.

One of the criteria for finding out the best formula is to have the least relative percentage or absolute error between the observed WIBT from the simulation results and the estimated WIBT from our formulas.

4.2.3.1 Analysis of results for the experiment type one

Here are the points have been taken out from the analysis of the results of 8 sets of simulation in experiment type one (Set 1 to 8) and our previous knowledge.

Point 1. We know that, when $\bar{K}_{Ai} = 1$, the WIBT should be zero, since there is no wait-in-batch-time anymore. Upon the availability of the first idle server, the arriving entity begins service directly without any waiting in batch time. So, we should have some factor of $\bar{K}_{Ai} - 1$ in our formula for all cases.

Point 2. We know from Hopp and spearman (Formula 4 in section 2.2.4) for the cases with a single server WIBT= $\frac{(\bar{K}_{Ai} - 1)\mathrm{t_i}}{2}$ for m_i =1.

Point 3. We see that the behavior of the WIBT in the experiment type one (which is merely a function of variable batch size) approaches to a linear behavior when the utilization is going up and getting close to 1. For the scenarios where utilization is close to 1 (Set 8), we see that the slope is roughly constant.

Figure 16 shows the linear behavior of WIBT in high utilization (Set 8). In this figure, one can notice the constant slope as the batch size arrival increases.

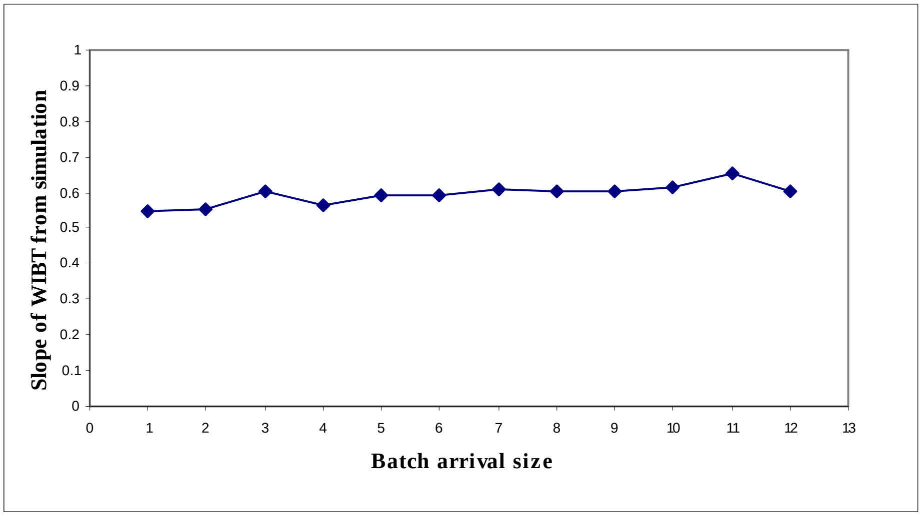

Figure 16. Simulation results for slopes for Scenarios 1-8-1 to 1-8-13 (Set 8)

In other words the WIBT should be a linear function of $\bar{K}_{Ai}$ in high utilization for all of the scenarios having both $\bar{K}_{Ai} \geq m_i$ and $\bar{K}_{Ai} < m_i$.

From points 1, 2, and 3, we can say that since for high utilization WIBT is the linear function of $\bar{K}_{Ai}$ and we should have the factor of $\bar{K}_{Ai} - 1$ in the formula, WIBT for high utilization (set 8) should approach $C_1(\bar{K}_{Ai} - 1)$. Here, C_1 should be a constant for all of the scenarios (1-8-1 to 1-8-13) whose number of servers, process time and utilization are the same.

Point 4. If we want to see the trend of the WIBT in regard to $\bar{K}_{Ai}$, again we have to construct the slope between the adjacent points to find out their relationships.

If we calculate the slopes for the Set 1 to 8, we can estimate the slope of the WIBT versus $\overline{K}_{Ai}$ by the following term.

$$Slope = \frac{\overline{K}_{Ai}^{\;\beta - u_i} u_i t_i}{m_i \alpha} \quad \textbf{(Formula 11)}$$

The α and β in Formula 11 is different for each set. In order to determine the best values of α and β, we constructed a summation of the squared absolute error between the slopes from the simulation results and Formula 11 for all scenarios within each set as an objective function. Then, we used the Microsoft Excel Solver to find the best values of α and β for each set by minimizing the constructed objective function and by changing α and β. By this way, we can obtain the best values for α and β for Set 1 to 8.

In Figures 17 to 24, one can see the slopes for Sets 1 to 8 and their estimated slopes from Formula 11, along with best calculated α and β.

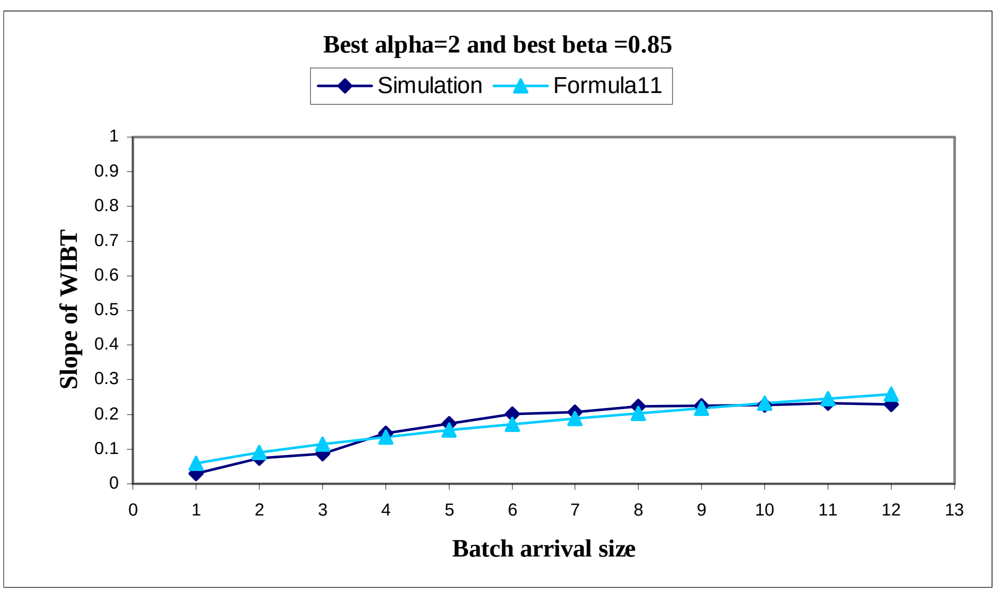

Figure 17. Simulation results for slopes for Scenarios 1-1-1 to 1-1-13 (Set 1)

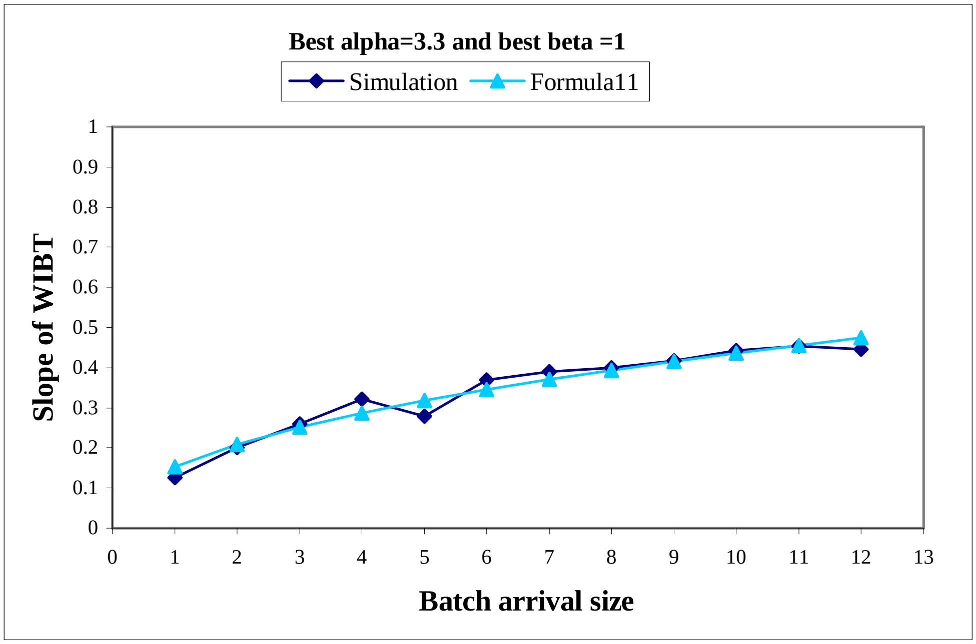

Figure 18. Simulation results for slopes for Scenarios 1-2-1 to 1-2-13 (Set 2)

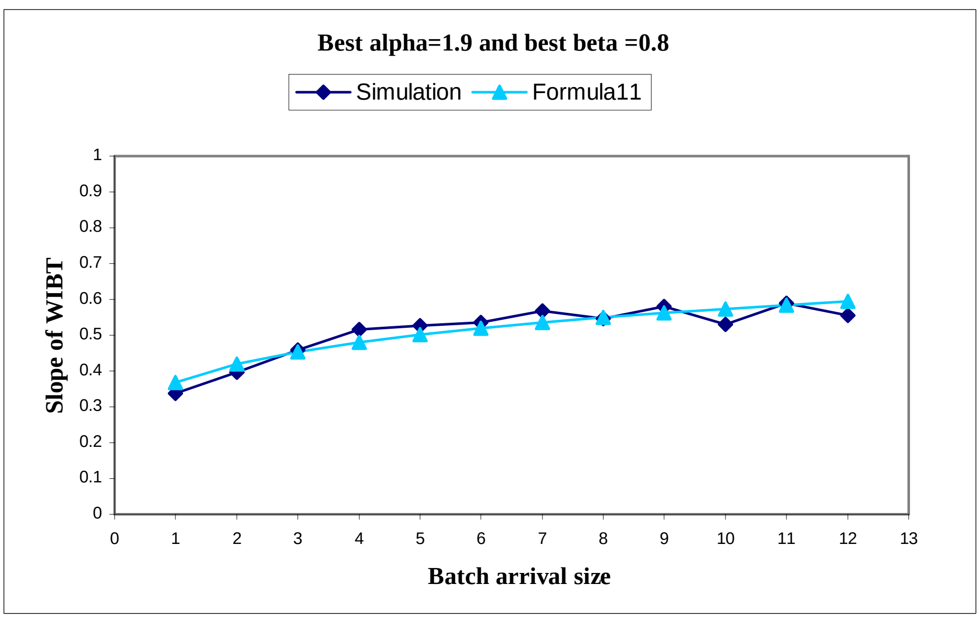

Figure 19. Simulation results for slopes for Scenarios 1-3-1 to 1-3-13 (Set 3)

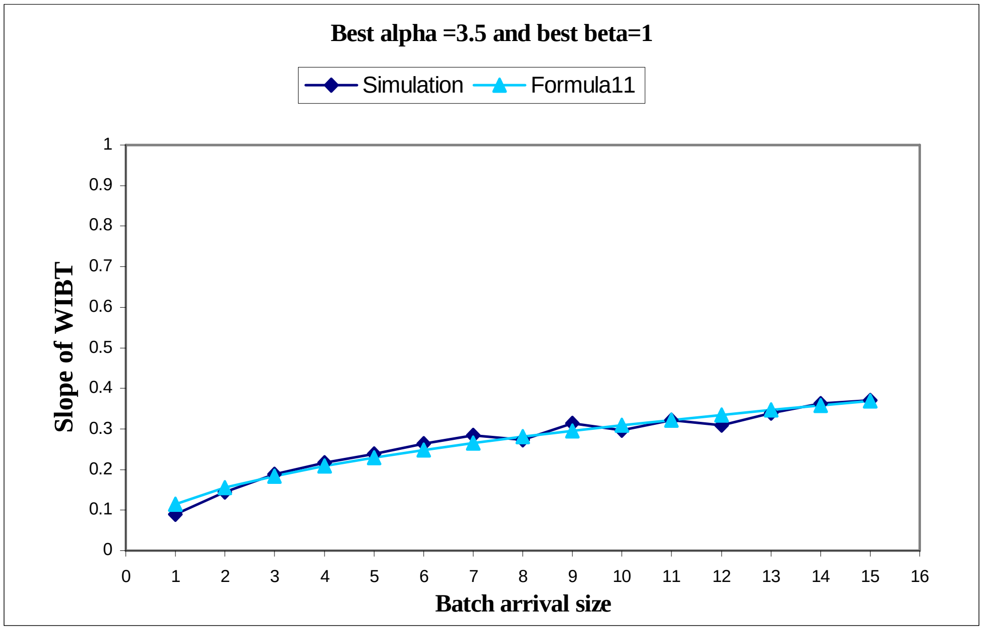

Figure 20. Simulation results for slopes for Scenarios 1-4-1 to 1-4-16 (Set 4)

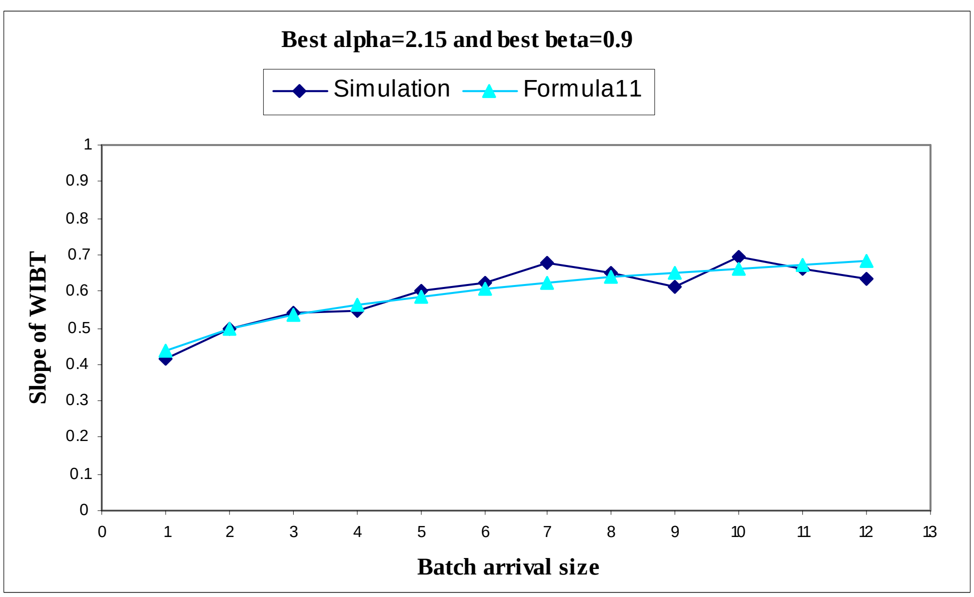

Figure 21. Simulation results for slopes for Scenarios 1-5-1 to 1-5-13 (Set 5)

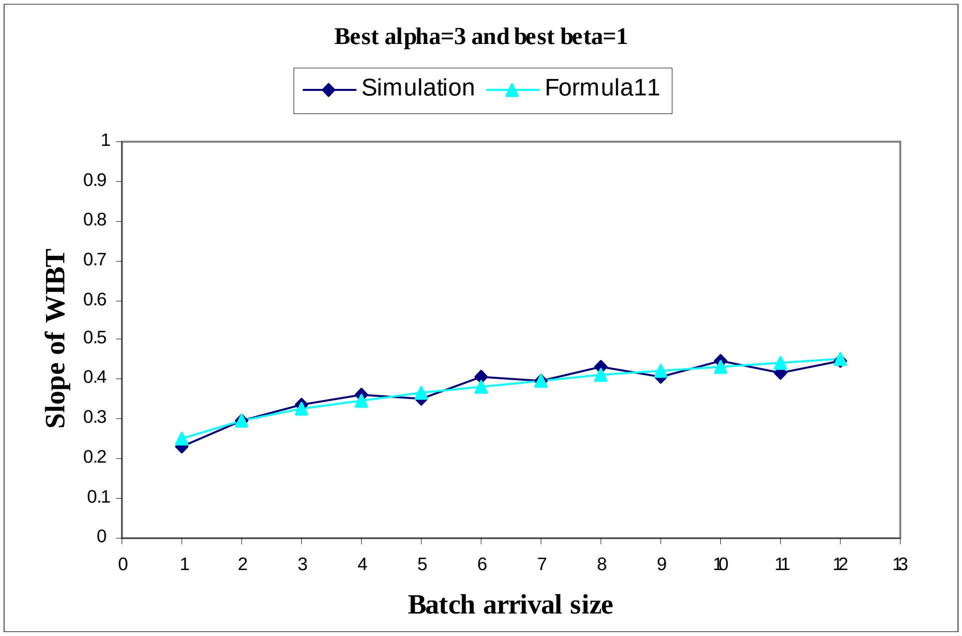

Figure 22. Simulation results for slopes for Scenarios 1-6-1 to 1-6-13 (Set 6)

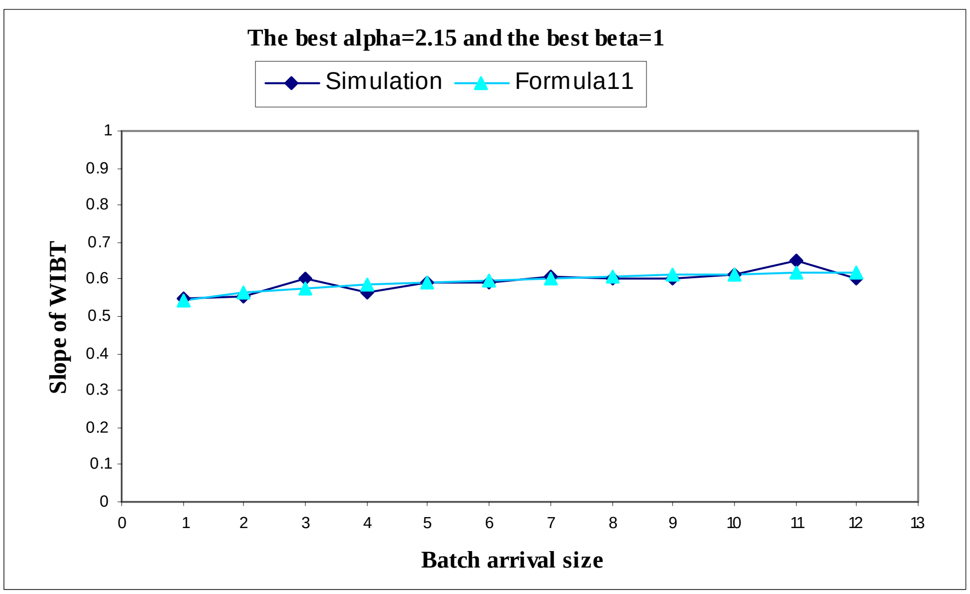

Figure 23. Simulation results for slopes for Scenarios 1-7-1 to 1-7-13 (Set 7)

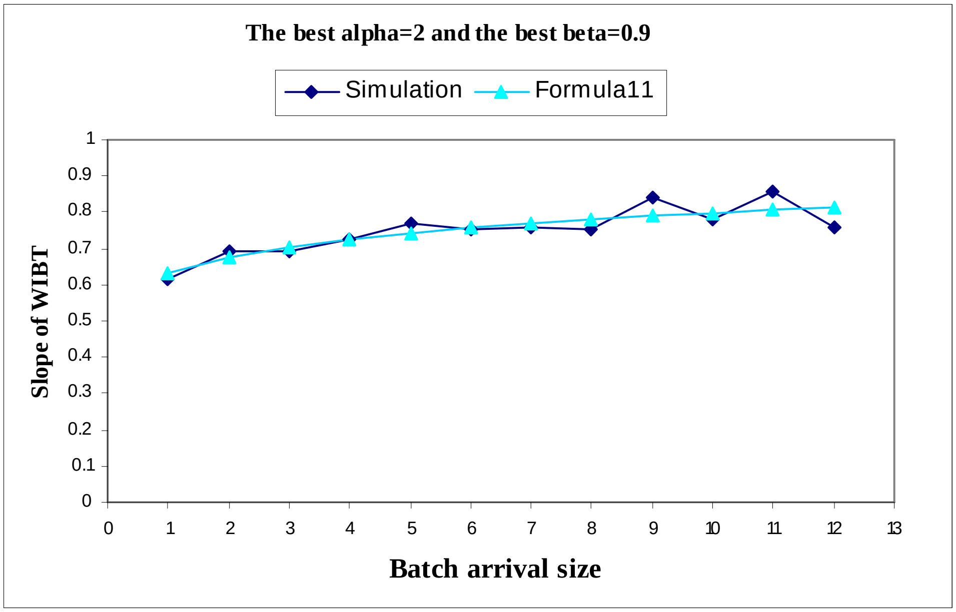

Figure 24. Simulation results for slopes for Scenarios 1-8-1 to 1-8-13 (Set 8)

According to the results of Figures 17 to 24, a good candidate to calculate the slope of each sets WIBT can be Formula 11 in which α and β is different for each set.

Presumably, if Formula 11 is a good approximation for the slope of the WIBT in which the only variable is arrival batch size, we can yield a formula for the WIBT by integrating Formula 11.

$$WIBT = \frac{\overline{K}_{Ai}^{\;1+\beta-u_i} u_i t_i}{m_i \alpha (1+\beta-u_i)} + C_2 \quad \textbf{(Formula 12)}$$

To calculate C_2, we know that, for $\overline{K}_{Ai}$ =1, the WIBT should be zero (Point 1 at the beginning of this section).

On the other hand, we know from before that for a single server

$WIBT = \frac{(\bar{K}_{Ai}\text{-}1)\,\mathrm{t_i}}{2}$ (Formula 4). Thus, one possible candidate for C_2 in Formula 12 is:

$$WIBT(\bar{K}_{Ai} = 1) = \frac{u_i t_i}{m_i \alpha (1 + \beta - u_i)} + C_2 = 0$$

$$C_2 = -\frac{u_i t_i}{m_i \alpha (1 + \beta - u_i)}$$

From obtained C_2, the Formula 12 will be changed to:

$$WIBT = \frac{(\bar{K}_{Ai}^{\;1+\beta-u_i} - 1) u_i t_i}{m_i \alpha (1 + \beta - u_i)} \quad \textbf{(Formula 13)}$$

For Sets 1 to 8, we note that α is different for each set and $0.8 \le \beta \le 1$.

Formula 13 with β =0.9 corresponds to the results from the high utilization set (Set 8) and behaves linearly as we expect. Since for β =0.9 in Set 8, $(\bar{K}_{Ai}^{\;0.9} - 1)$ is so close to $(\bar{K}_{Ai} - 1)$, we can justify the linear behavior of Formula 13 that approaches $WIBT = \frac{(\bar{K}_{Ai} - 1) t_i}{m_i \alpha}$ in which the best calculated α from Excel Solver is around 2.15 which can be rounded to 2 for simplicity.

In this way, Formula 13 for the sets with high utilization (such as Set 8) becomes:

$$WIBT = \frac{(\bar{K}_{Ai} - 1) t_i}{2 m_i} \quad \textbf{(Formula 14)}$$

This corresponds to the WIBT results from Set 8 very well.

The other justification for Formula 14 is that, when the utilization is close to 1, the m_i servers (whose process time is t_i) can be replaced by a single server whose process time is $\frac{t_i}{m_i}$.

Since we can estimate wait-in-batch-time for a single server from point 2 (Formula 4) in analysis of results for the experiment type one section, if we replace the process time t_i in Formula 4 by the new process time which is $\frac{t_i}{m_i}$, we can obtain a good approximation for the high utilization cases with the multiple servers which is:

$$WIBT = \frac{(\bar{K}_{Ai} - 1)t_i}{2m_i}$$

This formula is exactly Formula 14.

Figure 25 shows the trend of WIBT from Formula 14 and the simulation Set 8, whose utilization is the highest among the other sets and is close to 1.

This figure illustrates that the Formula 14 can be a good formula for scenarios with high utilization.

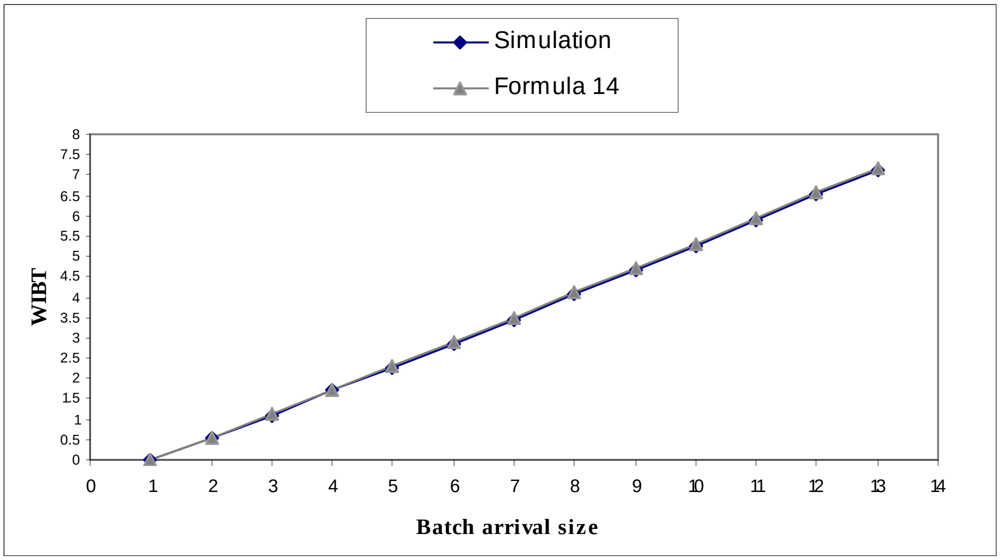

Figure 25. Simulation results and Formula 14 for Scenarios 1-8-1 to 1-8-13 (Set 8)

To determine whether Formula 13 is an acceptable formula for the other sets, Figure 26 to 32 demonstrates the simulation results and WIBT estimates from Formula 13 for Sets 1 to 7 respectively. It should be said again that, to use Formula 13, we made use of the best α and β have already obtained from Excel Solver (Figure 17 to 24).

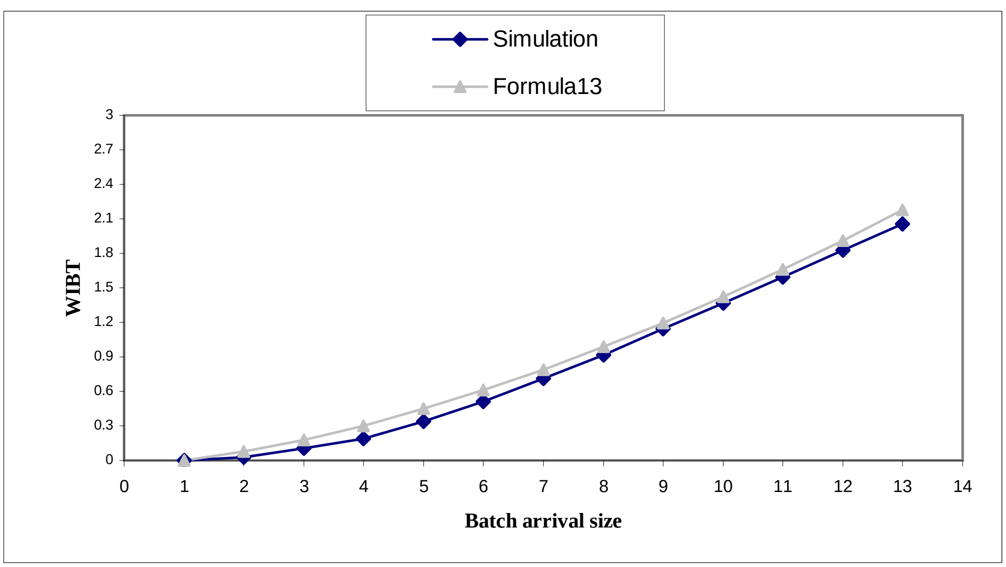

Figure 26. Simulation results and Formula 13 for Scenarios 1-1-1 to 1-1-13 (Set 1)

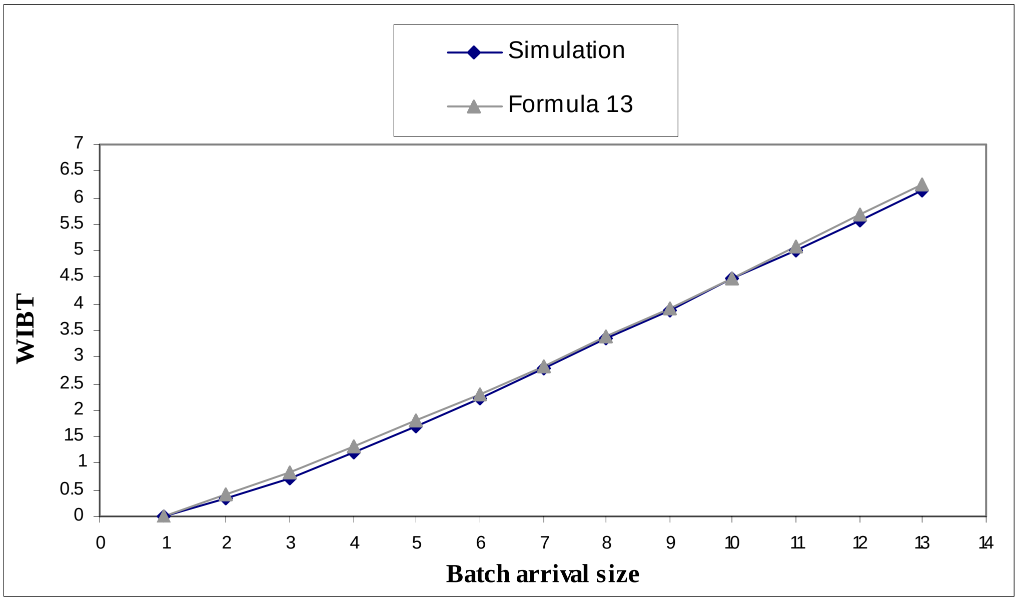

Figure 27. Simulation results and Formula 13 for Scenarios 1-2-1 to 1-2-13 (Set 2)

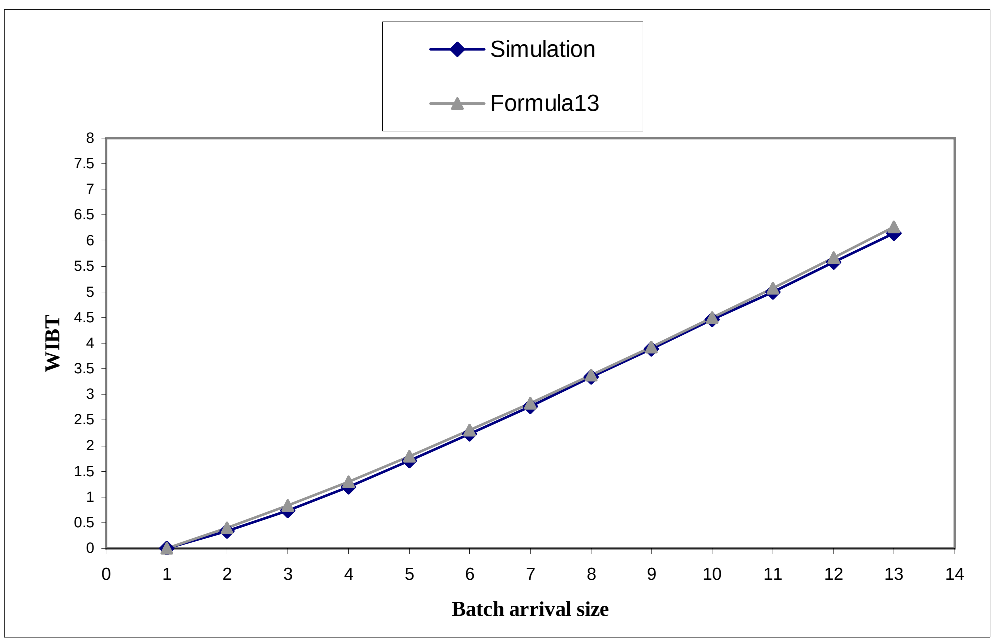

Figure 28. Simulation results and Formula 13 for Scenarios 1-3-1 to 1-3-13 (Set 3)

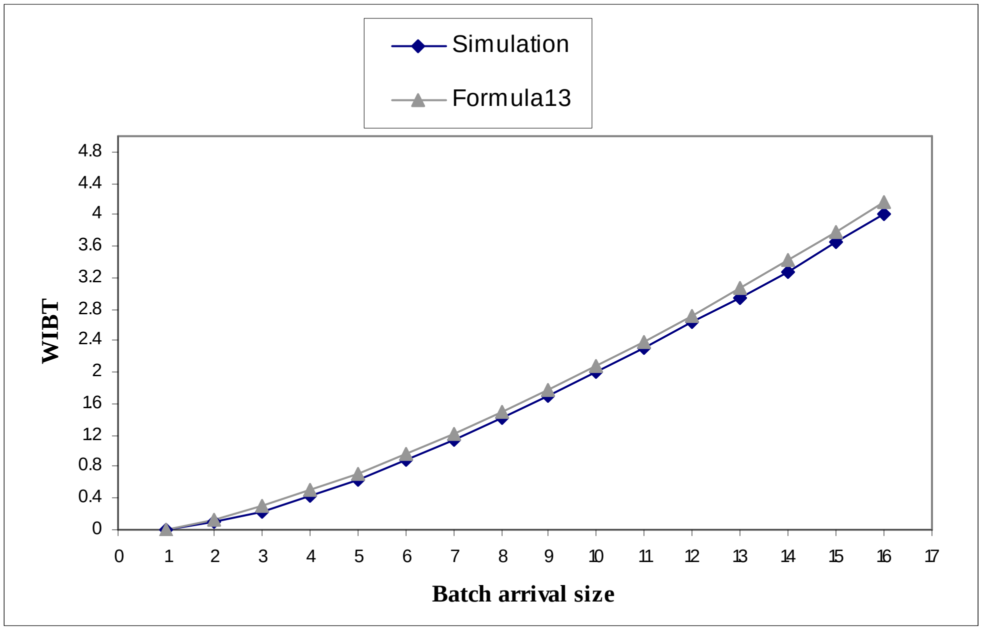

Figure 29. Simulation results and Formula 13 for Scenarios 1-4-1 to 1-4-16 (Set 4)

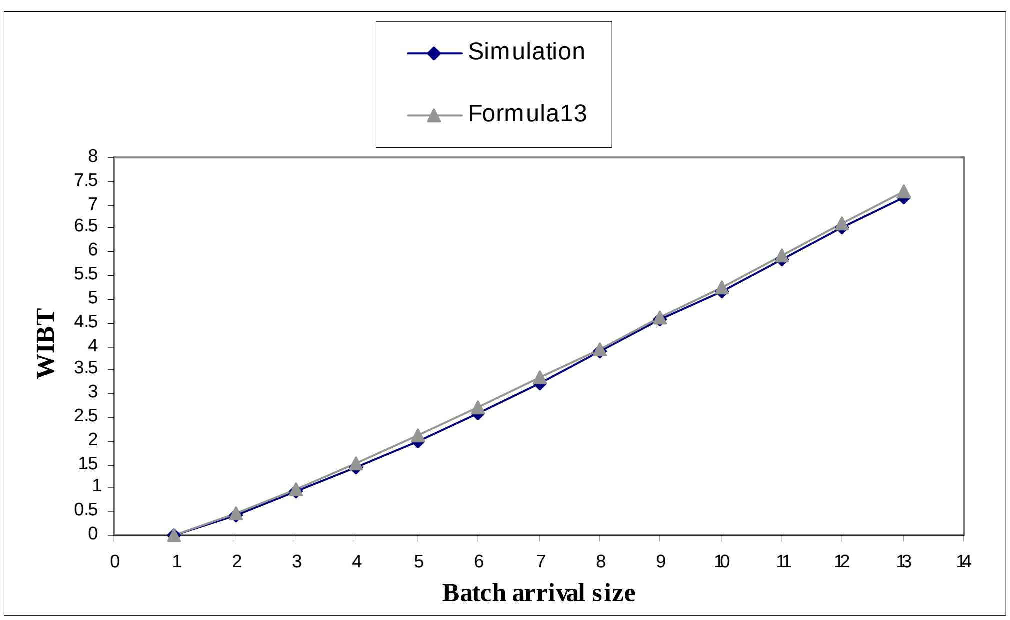

Figure 30. Simulation results and Formula 13 for Scenarios 1-5-1 to 1-5-13 (Set 5)

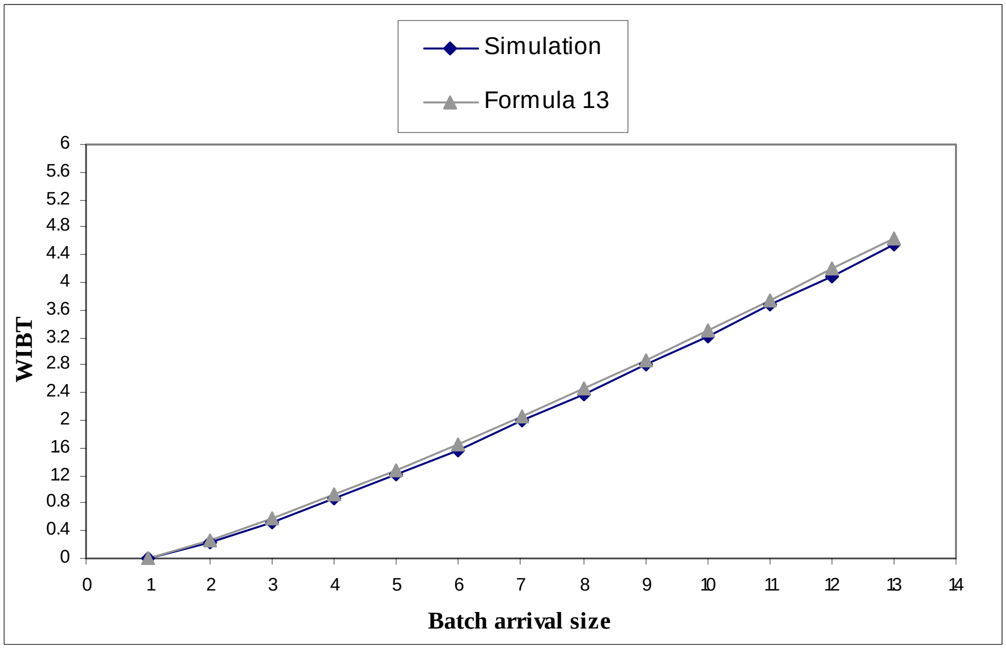

Figure 31. Simulation results and Formula 13 for Scenarios 1-6-1 to 1-6-13 (Set 6)

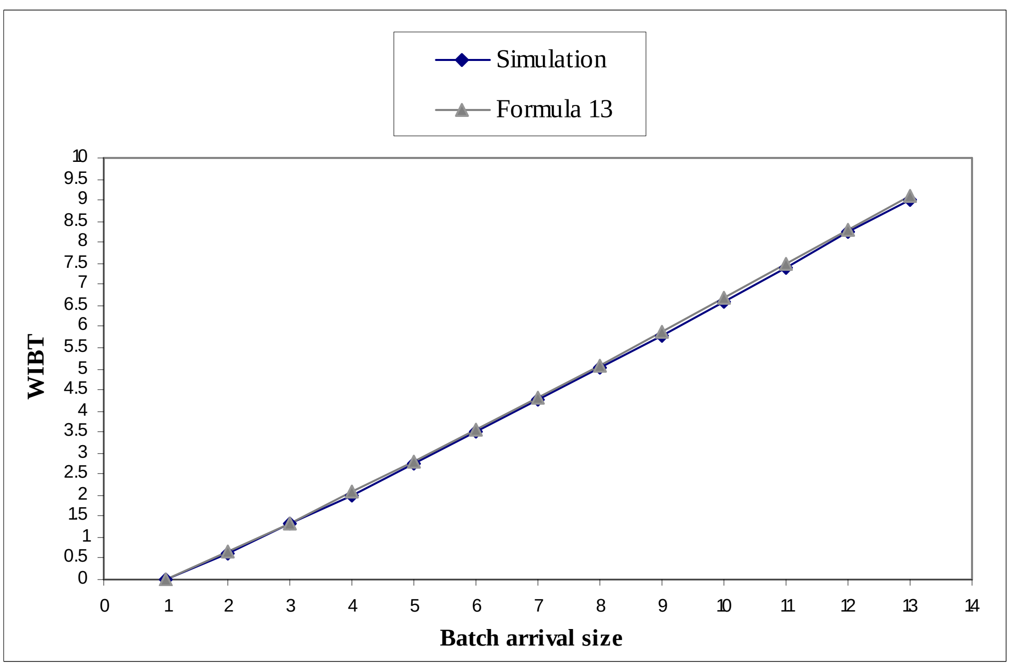

Figure 32. Simulation results and Formula 13 for Scenarios 1-7-1 to 1-7-13 (Set 7)

We can see from the Figures 25 to 32, Formula 13 can be acceptable equation for Sets 1 to 8 for the best obtained α and β from Excel Solver for each set.

4.2.3.2 Analysis of results for the experiment type two

If we see the simulation results of Set 9 and 10 in which the only variable is number of the servers and make use of some of the terms from Formula 13, we can notice with that the behavior of Set 9 and 10 in Figure 12 and 13 is roughly similar to the trend of $\frac{\bar{K}t_i}{m_i}$ by increasing the number of servers.

Figures 33 and 34 show the behavior of $\frac{\bar{K}t_i}{m_i}$ as a trend for the sets 9 and 10 versus the result of simulation.

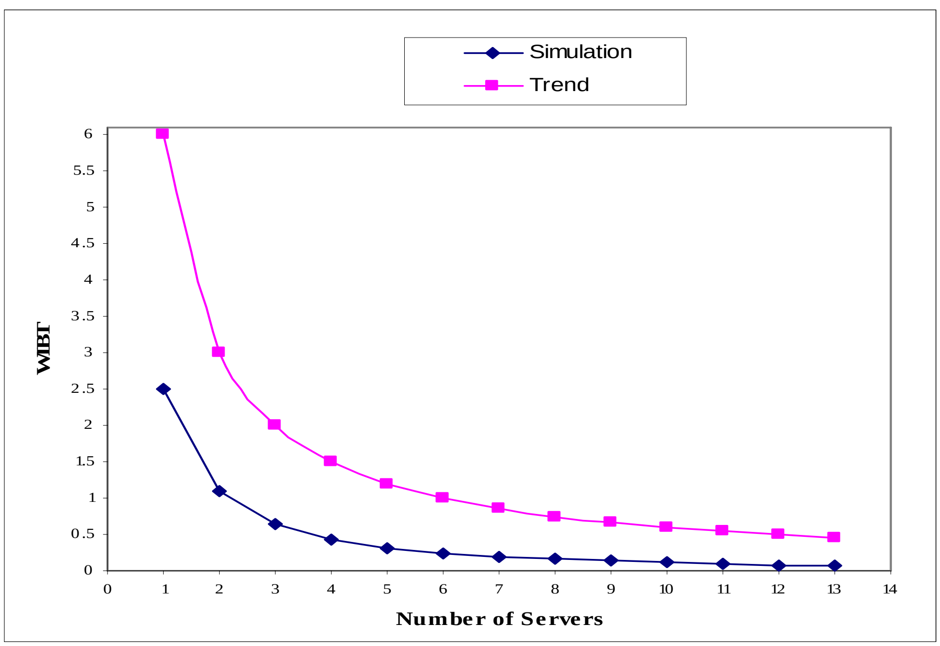

Figure 33. Simulation results and $\bar{K}t_i / m_i$ for Scenarios 2-1-1 to 2-1-13 (Set 9)

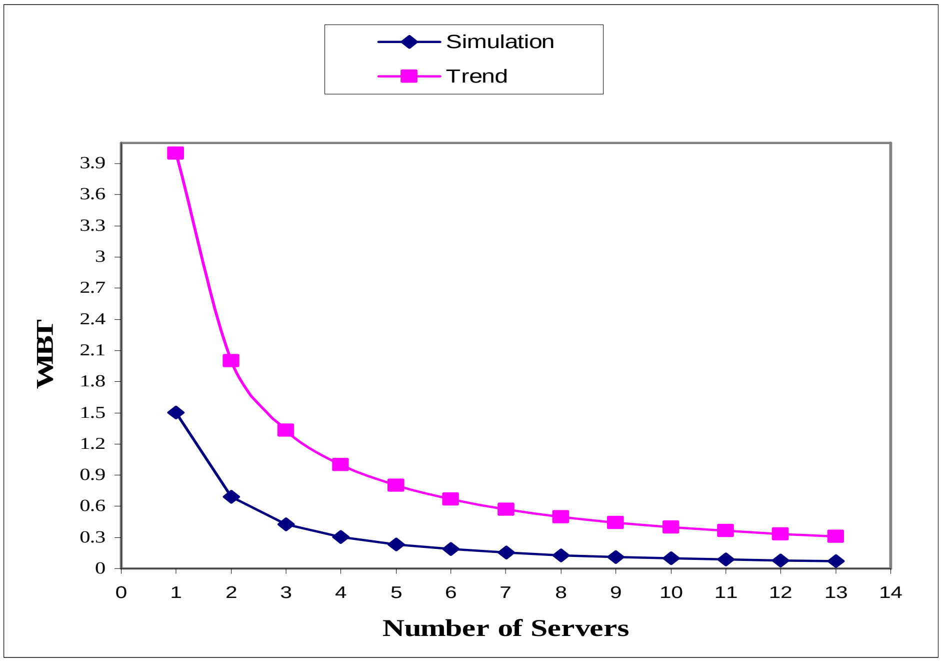

Figure 34. Simulation results and $\bar{K}t_i / m_i$ for Scenarios 2-2-1 to 2-2-13 (Set 10)

Therefore our new formula which can be workable for both set 9 and 10 should have the factor $\frac{\bar{K}t_i}{m_i}$. We need to add some coefficients or other factors to reduce the difference between the simulation results and our new formula.

Because the batch arrival size and utilization are constant in each set, according to Formula 13 and the fact that the only variable is the number of servers, we response a new formula for WIBT that is applicable for sets 9 and 10.This new formula is:

$$WIBT = \frac{(\bar{K}_{Ai}{}^{1+\beta-u_i} - 1)t_i}{m_i{}^{\lambda}\alpha(1+\beta-u_i)} \qquad \textbf{(Formula 15)}$$

For the single server case, we know that WIBT $= \frac{(\bar{K}_{Ai} - 1)t_i}{2}$ (Point 2 in section 4-2-3-1). Thus, we claim that for all of the Scenarios 2-1-1 to 2-1-13 and 2-2-1 to 2-2-13,

α should be 2, and β should be equal to the utilization to make $(1+\beta-u_i)$ equal to 1. So Formula 15 is changed to formula 16:

$$WIBT = \frac{(\bar{K}_{Ai} \; -1)t_i}{2m_i^{\lambda}} \quad \textbf{(Formula 16)}$$

Because we now know α and β, we need to calculate the best λ in Formula 16, which may be different for each simulation scenario in sets 9 and 10. In order to determine the best value for λ, we construct the square absolute error between the WIBT obtained from simulation results for each scenario and Formula 16 for that scenario.

Subsequently, the Microsoft Excel Solver can calculate the best λ for each scenario by minimizing the square absolute error and by changing λ by knowing this fact that α =2 and $\beta = u_i$ for each scenario. By this way, we can obtain the best value of λ for each scenario within each set.

We can see in Tables 21 and 22 the simulation results for WIBT and Formula 16 for that obtained λ from Excel Solver for each scenario within sets 9 and 10.

Table 21. Simulations results and Formula 16 for Scenarios 2-1-1 to 2-1-13 (Set 9)

Scenario Name	The best calculated λ from excel Solver	Number of servers	Arrival batch size	Utilization	Process time(min)	WIBT From simulation	WIBT from formula 16 with calculated λ	Error%
2-1-1	1.00	1	6	60.00%	1	2.495	2.500	0.20%
2-1-2	1.19	2	6	60.00%	1	1.099	1.099	0.00%
2-1-3	1.23	3	6	60.00%	1	0.646	0.646	0.00%
2-1-4	1.26	4	6	60.00%	1	0.434	0.434	0.00%
2-1-5	1.29	5	6	60.00%	1	0.315	0.315	0.00%
2-1-6	1.31	6	6	60.00%	1	0.241	0.241	0.00%
2-1-7	1.32	7	6	60.00%	1	0.192	0.192	0.00%
2-1-8	1.33	8	6	60.00%	1	0.158	0.158	0.00%
2-1-9	1.34	9	6	60.00%	1	0.132	0.132	0.00%
2-1-10	1.36	10	6	60.00%	1	0.11	0.110	0.00%
2-1-11	1.37	11	6	60.00%	1	0.094	0.094	0.00%
2-1-12	1.38	12	6	60.00%	1	0.082	0.082	0.00%
2-1-13	1.39	13	6	60.00%	1	0.071	0.071	0.00%

Table 22. Simulations results and Formula 16 for Scenarios 2-2-1 to 2-2-13 (Set 10)

Scenario Name	The best calculated λ from excel Solver	Number of servers	Arrival batch size	Utilization	Process time(min)	WIBT From simulation	WIBT from formula 16 with calculated λ	Error%
2-2-1	1.00	1	4	80.00%	1	1.504	1.500	0.27%
2-2-2	1.12	2	4	80.00%	1	0.689	0.689	0.00%
2-2-3	1.14	3	4	80.00%	1	0.427	0.427	0.00%
2-2-4	1.15	4	4	80.00%	1	0.303	0.303	0.00%
2-2-5	1.16	5	4	80.00%	1	0.232	0.232	0.00%
2-2-6	1.16	6	4	80.00%	1	0.188	0.188	0.00%
2-2-7	1.16	7	4	80.00%	1	0.156	0.156	0.00%
2-2-8	1.18	8	4	80.00%	1	0.129	0.129	0.00%
2-2-9	1.19	9	4	80.00%	1	0.11	0.110	0.00%
2-2-10	1.18	10	4	80.00%	1	0.098	0.098	0.00%
2-2-11	1.18	11	4	80.00%	1	0.088	0.088	0.00%
2-2-12	1.21	12	4	80.00%	1	0.075	0.075	0.00%
2-2-13	1.19	13	4	80.00%	1	0.071	0.071	0.00%

We note that, for each set, the best λ is close to $2-u_i$. In other words, since the utilization for set 9 and 10 are 60% and 80%, we can set λ to be equal to 1.4 and 1.2 respectively.

So our new formula for WIBT in which the only variable is number of servers is:

$$\textbf{WIBT}=\frac{(\bar{K}_{Ai}-1)t_i}{2m_i^{2-u_i}} \quad \textbf{(Formula 17)}$$

We see that the absolute error between the simulation results and Formula 17 is still small.

Table 23 and 24 demonstrate the simulation results for WIBT and Formula 17 for each scenario within sets 9 and 10.

Figures 35 and 36 demonstrate the results of WIBT from simulations and formula17 for sets 9 and 10 respectively.

Table 23. Simulations results and Formula 17 for Scenarios 2-1-1 to 2-1-13 (Set 9)

Scenario Name	The approximate λ which is $2-u_i$	Number of servers	Arrival batch size	Utilization	Process time(min)	WIBT From simulation	WIBT from Formula 17	Error%
2-1-1	1.4	1	4	60.00%	1	2.495	2.500	0.20%
2-1-2	1.4	2	4	60.00%	1	1.099	0.984	10.46%
2-1-3	1.4	3	4	60.00%	1	0.646	0.570	11.71%
2-1-4	1.4	4	4	60.00%	1	0.434	0.387	10.76%
2-1-5	1.4	5	4	60.00%	1	0.315	0.287	8.93%
2-1-6	1.4	6	4	60.00%	1	0.241	0.224	6.85%
2-1-7	1.4	7	4	60.00%	1	0.192	0.182	4.98%
2-1-8	1.4	8	4	60.00%	1	0.158	0.152	3.51%
2-1-9	1.4	9	4	60.00%	1	0.132	0.130	1.43%
2-1-10	1.4	10	4	60.00%	1	0.11	0.113	2.65%
2-1-11	1.4	11	4	60.00%	1	0.094	0.099	5.67%
2-1-12	1.4	12	4	60.00%	1	0.082	0.088	7.75%
2-1-13	1.4	13	4	60.00%	1	0.071	0.079	11.74%

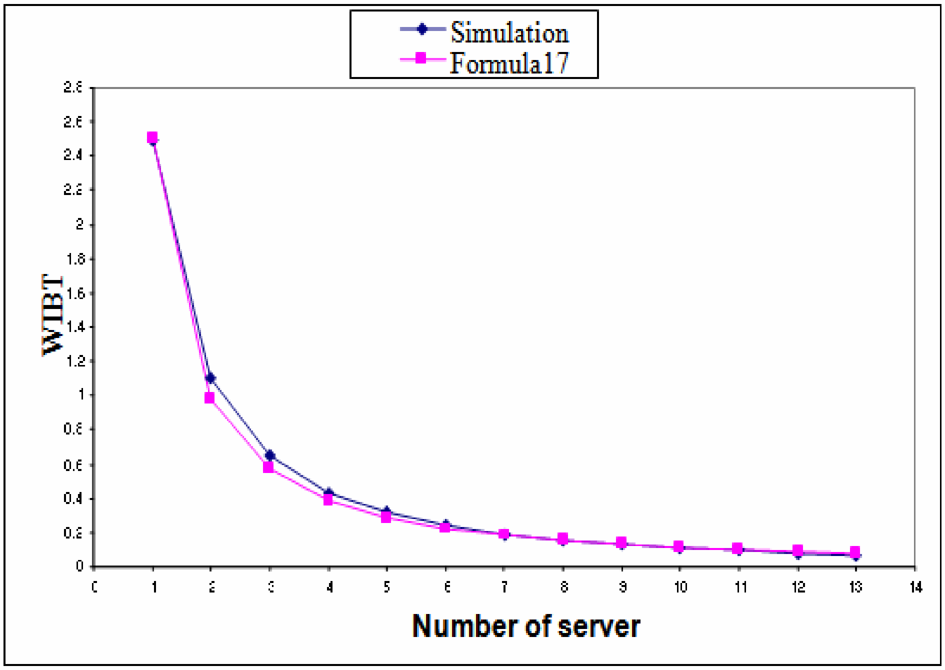

Figure 35. Simulations results and Formula 17 for Scenarios 2-1-1 to 2-1-13 (Set 9)

Table 24. Simulations results and Formula 17 for Scenarios 2-2-1 to 2-2-13 (Set 10)

Scenario name	The approximate λ which is $2-u_i$	Number of servers	Arrival batch size	Utilization	Process time	WIBT From simulation	WIBT from Formula 17	Error%
2-2-1	1.2	1	4	80.00%	1	1.504	1.500	0.27%
2-2-2	1.2	2	4	80.00%	1	0.689	0.662	3.96%
2-2-3	1.2	3	4	80.00%	1	0.427	0.410	3.99%
2-2-4	1.2	4	4	80.00%	1	0.303	0.292	3.67%
2-2-5	1.2	5	4	80.00%	1	0.232	0.224	3.33%
2-2-6	1.2	6	4	80.00%	1	0.188	0.181	3.81%
2-2-7	1.2	7	4	80.00%	1	0.156	0.151	3.36%
2-2-8	1.2	8	4	80.00%	1	0.129	0.129	0.18%
2-2-9	1.2	9	4	80.00%	1	0.11	0.112	1.86%
2-2-10	1.2	10	4	80.00%	1	0.098	0.099	0.96%
2-2-11	1.2	11	4	80.00%	1	0.088	0.088	0.46%
2-2-12	1.2	12	4	80.00%	1	0.075	0.080	6.37%
2-2-13	1.2	13	4	80.00%	1	0.071	0.073	2.23%

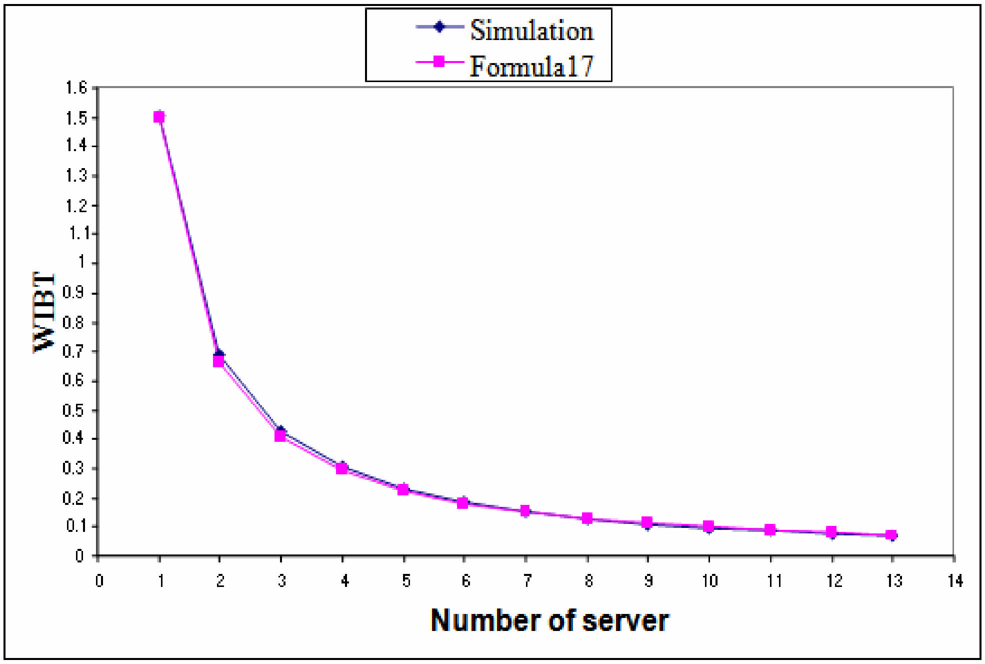

Figure 36. Simulations results and Formula 17 for Scenarios 2-2-1 to 2-2-13 (Set 10)

From these results, we see that the absolute error of Formula 17 is small for the scenarios only whose variable is the number of servers. Therefore, Formula17 is a good choice to estimate WIBT in sets 9 and 10.

4.2.3.3 Analysis of results for the experiment type three

Now consider the simulation results of sets 11 and 12 in which the only variable is number of the servers and utilization. We propose that Formula 17 might be a good estimate for sets 11 and 12.

$$\text{WIBT}=\frac{(\bar{K}_{Ai}-1)t_i}{2m_i^{2-u_i}} \quad \textbf{(Formula 17)}$$

The results of Formula 17 for simulation sets 11 and 12 are not that much exact like the results of simulation sets 9 and 10, but since the absolute error is so small; it can be a good approximation, although we have relatively the big percentage errors.

Tables 25 and 26 demonstrate the simulation results for WIBT and Formula 17 for each scenario within sets 11 and 12. Figures 37 and 38 demonstrate the results of WIBT from simulations and Formula 17 for sets 11 and 12 respectively.

Table 25. Simulations results and Formula 17 for Scenarios 3-1-1 to 3-1-13 (Set 11)

Scenario Name	Number of servers	Arrival batch size	Utilization	Process time	WIBT From simulation	WIBT from Formula 17	Percentage Error%	Absolute Error (min)
3-1-1	1	4	96.00%	1.2	1.799	1.80	0.06%	0.001
3-1-2	2	4	48.00%	1.2	0.692	0.63	9.30%	0.064
3-1-3	3	4	32.00%	1.2	0.28	0.28	1.52%	0.004
3-1-4	4	4	24.00%	1.2	0.11	0.16	42.64%	0.047
3-1-5	5	4	19.20%	1.2	0.053	0.10	85.04%	0.045
3-1-6	6	4	16.00%	1.2	0.023	0.07	189.57%	0.044
3-1-7	7	4	13.71%	1.2	0.01	0.05	379.71%	0.038
3-1-8	8	4	12.00%	1.2	0.004	0.04	802.41%	0.032
3-1-9	9	4	10.67%	1.2	0.002	0.03	1304.57%	0.026
3-1-10	10	4	9.60%	1.2	0.001	0.02	2145.29%	0.021
3-1-11	11	4	8.73%	1.2	0	0.02	#DIV/0!	0.018
3-1-12	12	4	8.00%	1.2	0	0.02	#DIV/0!	0.015
3-1-13	13	4	7.38%	1.2	0	0.01	#DIV/0!	0.013

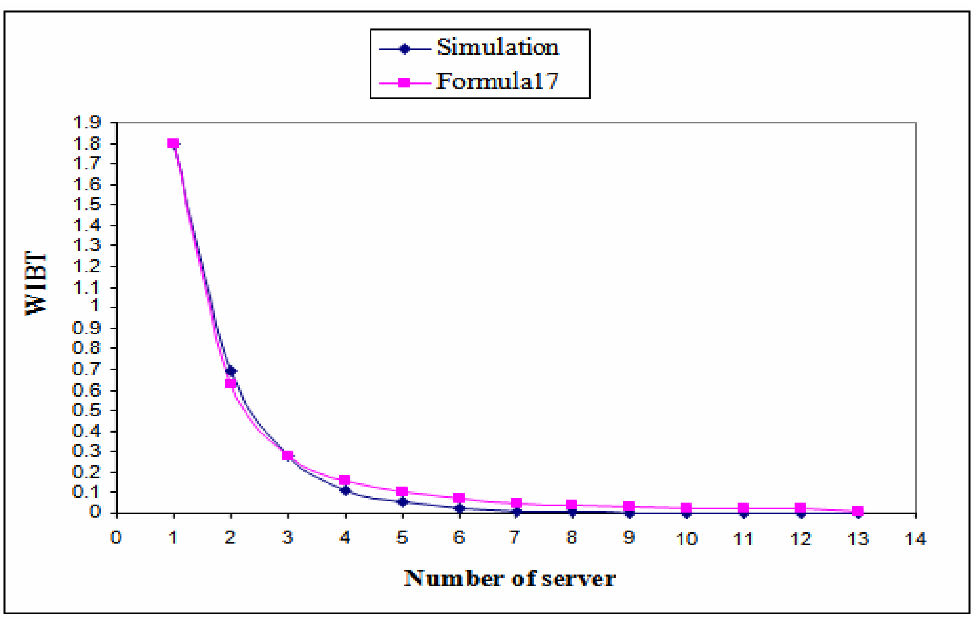

Figure 37. Simulations results and Formula 17 for Scenarios 3-1-1 to 3-1-13 (Set 11)

Table 26. Simulations results and Formula 17 for Scenarios 3-2-1 to 3-2-13 (Set 12)

Scenario Name	Number of servers	Arrival batch size	Utilization	Process time	WIBT From simulation	WIBT from Formula 17	Percentage Error%	Absolute Error (min)
3-2-1	1	6	90.00%	1	2.498	2.50	0.08%	0.002
3-2-2	2	6	45.00%	1	1.036	0.85	17.59%	0.182
3-2-3	3	6	30.00%	1	0.496	0.39	22.13%	0.110
3-2-4	4	6	22.50%	1	0.244	0.21	12.52%	0.031
3-2-5	5	6	18.00%	1	0.117	0.13	14.19%	0.017
3-2-6	6	6	15.00%	1	0.055	0.09	65.20%	0.036
3-2-7	7	6	12.86%	1	0.031	0.07	111.37%	0.035
3-2-8	8	6	11.25%	1	0.017	0.05	190.34%	0.032
3-2-9	9	6	10.00%	1	0.009	0.04	327.21%	0.029
3-2-10	10	6	9.00%	1	0.004	0.03	668.92%	0.027
3-2-11	11	6	8.18%	1	0.002	0.03	1156.98%	0.023
3-2-12	12	6	7.50%	1	0.001	0.02	1991.78%	0.020
3-2-13	13	6	6.92%	1	0.001	0.02	1666.74%	0.017

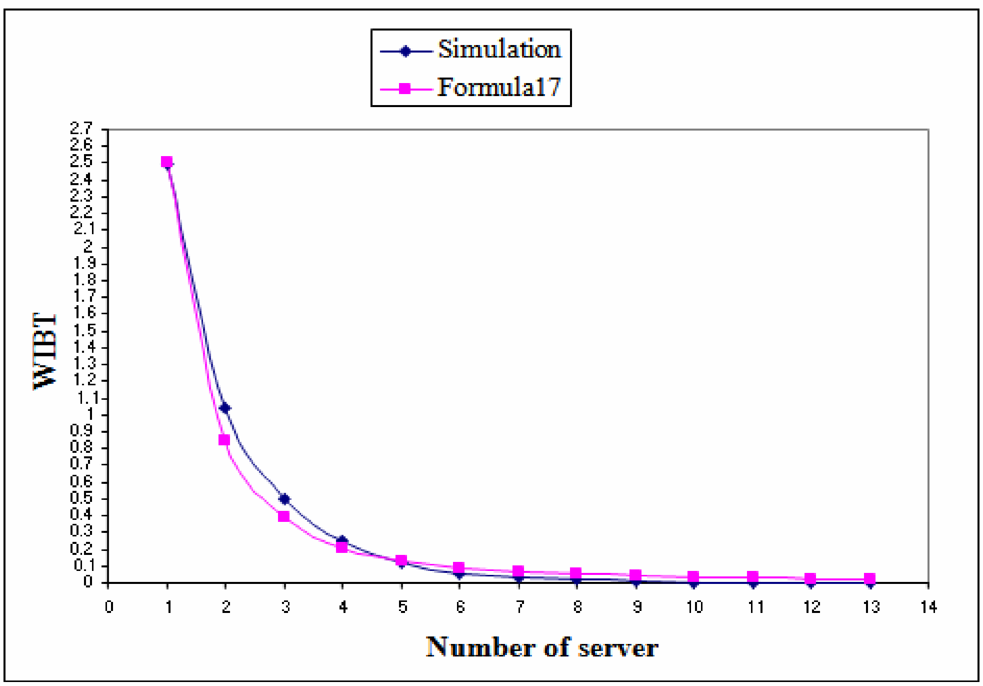

Figure 38. Simulations results and Formula 17 for Scenarios 3-2-1 to 3-2-13 (Set 12)

4.2.3.4 The final analysis of results from all experiments

We saw that, based on Formulas 13 and 17, WIBT is a function of the number of servers, arrival batch size, utilization and the process time.

According to Formula 13 and the simulation results from Sets 1 to 8, we conclude the Formula 13 can be a good approximation when the only variable is $\bar{K}_{Ai}$ and the other factors are constant within each set.

The main problem with Formula 13 is that we have different values for α and β in each Set 1 to 8, while our goal is to come up with a general formula satisfying all cases and conditions.

We can see from Sets 1 to 8 (first type of experiment) that the best β is between 0.8 and 1. So for simplicity, we can put β =1 in Formula13. Consequently, the factor $(\bar{K}_{Ai}{}^{1+\beta-u_i}-1)$ is changed to $(\bar{K}_{Ai}{}^{2-u_i}-1)$. Additionally, the $(1+\beta-u_i)$ in the denominator is changed to $(2-u_i)$.

So our new formula is:

$$WIBT = \frac{(\bar{K}_{Ai}{}^{2-u_i}-1)u_i t_i}{m_i \alpha (2-u_i)} \quad \textbf{(Formula 18)}$$

On the other hand, according to the results from Sets 9 to 12, Formula 17 is a good approximation for WIBT when the variables are m_i and u_i while the other factors are constant within each set.

The question here is what can be a general formula for WIBT that is acceptable when $\bar{K}_{Ai}$, m_i and u_i are all variables.

We construct a new formula to have some of factors from Formula 18 such as $(\bar{K}_{Ai}{}^{2-u_i} - 1)$ in the numerator and also to have the factor of the $m_i^{2-u_i}$ extracted from Formula 17 in the denominator. The process time (t_i), which is common between both Formulas 17 and 18, can also be one of the factors in our new formula. At the moment, we don't bring the $(2-u_i)$ and u_i from Formula 18 in our new formula since they don't exist in Formula 17 at all. In this way, our new formula is:

$$WIBT = \frac{(\bar{K}_{Ai}{}^{2-u_i} - 1)t_i}{\alpha m_i^{2-u_i}} \quad \textbf{(Formula 19)}$$

What is α ? We have the same α in Formula 18 and Formula 17, and we put α =2 in Formula 17.

If we construct Formula 19 for Sets 1 to 8 to calculate the best possible α for Sets 1 to 8 by utilizing the Microsoft Excel Solver to have the least error, as we did in previous sections, we set that α is a function of utilization.

For high utilization, α goes to 2, and, at the same time, $2-u_i$ approaches 1, which we expect for high utilization, where WIBT should be close to $\frac{(\bar{K}_{Ai} - 1)t_i}{2m_i}$ (Formula 14).

Table 27 lists the best α for Sets 1 to 8 for Formula 19.

Table 27. The best α obtained from simulation Sets 1 to 8 for Formula 19

Simulation Set #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Utilization (%)	25	50	58.33	62.5	68.75	75	81.25	93.75
Best α obtained from Excel Solver for Formula 19	8.78	4.8	4	3.83	3.3	3	2.7	2.2

Since, as shown in Table 27, α is a function of utilization, a good nominee for estimating α is $\frac{2.3}{u_i}$ which is relatively close to Table 27 results.

Table 28 shows the calculated α from $\alpha = \frac{2.3}{u_i}$ for all of the simulation Sets 1 to 8.

Table 28. The calculated α from $2.3/u_i$ for simulation Sets 1 to 8

Simulation set #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Best α obtained for Formula 19	8.78	4.8	4	3.83	3.3	3	2.7	2.2
α Obtained from $2.3/u_i$	9.2	4.6	3.9	3.7	3.3	3.1	2.8	2.5

By putting the $\alpha = \frac{2.3}{u_i}$ in Formula 19, our new formula is:

$$WIBT = \frac{(\bar{K}_{Ai}^{\,2-u_i} - 1)u_i t_i}{2.3 m_i^{\,2-u_i}} \quad \textbf{(Formula 20)}$$

As noted before, one of the important points is that for the high utilization WIBT=$\frac{(\bar{K}_{Ai} - 1)t_i}{2m_i}$ (Formula 14), so α should go toward 2. According to this point, another possible nominee for the α can be $\frac{3-u_i}{u_i}$. Since, when utilization approaches 1 α should go around 2 and $\frac{3-u_i}{u_i}$ satisfies this requirement. Table 29 shows the calculated α from $\alpha = \frac{3-u_i}{u_i}$ for all of the simulation sets1-8.

Table 29. The calculated α from $(3-u_i)/u_i$ for simulation Sets 1 to 8

Simulation set #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Best α obtained for Formula 19	8.78	4.8	4	3.83	3.3	3	2.7	2.2
α Obtained from $(3-u_i)/u_i$	11.00	5.00	4.14	3.80	3.36	3.00	2.69	2.20

Formula 21 modifies Formula 19 by substituting $\alpha = \frac{3-u_i}{u_i}$:

$$WIBT = \frac{(\overline{K}_{Ai}^{\,2-u_i} - 1)u_i t_i}{(3-u_i)m_i^{\,2-u_i}} \quad \textbf{(Formula 21)}$$

Tables 30 to 41 show the WIBT from the simulation results and the estimates from Formulas 20 and 21 for Sets 1 to 12. Since the absolute error between the simulation results and the formulas is small, we concluded that Formulas 20 and 21 are acceptable despite having a large percentage error in some scenarios.

Table 30. Simulations results and Formula 20 and 21 for Scenarios 1-1-1 to 1-1-13 (Set 1)

Scenario #	WIBT From simulation	WIBT from Formula 20	Percentage error% from Formula 20	Absolute error from Formula 20	WIBT from Formula 21	Percentage error% from Formula 21	Absolute error from Formula21
1-1-1	0	0.00	#DIV/0!	0.00	0.00	#DIV/0!	0.00
1-2-1	0.03	0.05	51.39%	0.02	0.04	26.61%	0.01
1-3-1	0.104	0.11	7.87%	0.01	0.09	9.78%	0.01
1-4-1	0.191	0.20	3.76%	0.01	0.17	13.22%	0.03
1-5-1	0.337	0.30	10.38%	0.03	0.25	25.04%	0.08
1-6-1	0.511	0.42	17.27%	0.09	0.35	30.81%	0.16
1-7-1	0.713	0.56	21.51%	0.15	0.47	34.35%	0.24
1-8-1	0.919	0.71	22.52%	0.21	0.60	35.20%	0.32
1-9-1	1.142	0.88	23.00%	0.26	0.74	35.60%	0.41
1-10-1	1.367	1.06	22.36%	0.31	0.89	35.07%	0.48
1-11-1	1.594	1.26	21.11%	0.34	1.05	34.02%	0.54
1-12-1	1.826	1.47	19.64%	0.36	1.23	32.79%	0.60
1-13-1	2.054	1.69	17.68%	0.36	1.41	31.15%	0.64

Table 31. Simulations results and Formula 20 and 21 for Scenarios 1-2-1 to 1-2-13 (Set 2)

Scenario #	WIBT From simulation	WIBT from Formula 20	Percentage error% from Formula 20	Absolute error from Formula 20	WIBT from Formula 21	Percentage error% from Formula 21	Absolute error from Formula21
1-2-1	0	0.00	#DIV/0!	0.00	0.00	#DIV/0!	0.00
1-2-2	0.126	0.16	28.79%	0.04	0.15	18.48%	0.02
1-2-3	0.327	0.37	13.89%	0.05	0.34	4.78%	0.02
1-2-4	0.586	0.62	6.01%	0.04	0.57	2.47%	0.01
1-2-5	0.907	0.90	0.39%	0.00	0.83	8.35%	0.08
1-2-6	1.185	1.22	2.58%	0.03	1.12	5.62%	0.07
1-2-7	1.554	1.55	0.06%	0.00	1.43	7.95%	0.12
1-2-8	1.944	1.92	1.26%	0.02	1.77	9.16%	0.18
1-2-9	2.343	2.31	1.52%	0.04	2.12	9.39%	0.22
1-2-10	2.76	2.72	1.53%	0.04	2.50	9.41%	0.26
1-2-11	3.203	3.15	1.68%	0.05	2.90	9.55%	0.31
1-2-12	3.657	3.60	1.54%	0.06	3.31	9.42%	0.34
1-2-13	4.103	4.07	0.78%	0.03	3.75	8.71%	0.36

Table 32. Simulations results and Formula 20 and 21 for Scenarios 1-3-1 to 1-3-13 (Set 3)

Scenario #	WIBT From simulation	WIBT from Formula 20	Percentage error% from Formula 20	Absolute error from Formula 20	WIBT from Formula 21	Percentage error% from Formula 21	Absolute error from Formula21
1-3-1	0	0.00	#DIV/0!	0.00	0.00	#DIV/0!	0.00
1-3-2	0.337	0.31	7.24%	0.02	0.30	11.72%	0.04
1-3-3	0.733	0.70	4.44%	0.03	0.67	9.05%	0.07
1-3-4	1.192	1.15	3.77%	0.04	1.09	8.41%	0.10
1-3-5	1.708	1.64	3.80%	0.06	1.56	8.44%	0.14
1-3-6	2.234	2.18	2.30%	0.05	2.08	7.02%	0.16
1-3-7	2.77	2.76	0.33%	0.01	2.63	5.14%	0.14
1-3-8	3.337	3.37	1.14%	0.04	3.21	3.75%	0.12
1-3-9	3.883	4.02	3.57%	0.14	3.83	1.43%	0.06
1-3-10	4.464	4.70	5.27%	0.24	4.47	0.19%	0.01
1-3-11	4.994	5.41	8.24%	0.41	5.14	3.02%	0.15
1-3-12	5.583	6.14	9.97%	0.56	5.84	4.66%	0.26
1-3-13	6.138	6.90	12.40%	0.76	6.57	6.97%	0.43

Table 33. Simulations results and Formula 20 and 21 for Scenarios 1-4-1 to 1-4-16 (Set 4)

Scenario #	WIBT From simulation	WIBT from Formula 20	Percentage error% from Formula 20	Absolute error from Formula 20	WIBT from Formula 21	Percentage error% from Formula 21	Absolute error from Formula21
1-4-1	0	0.00	#DIV/0!	0.00	0.00	#DIV/0!	0.00
1-4-2	0.09	0.14	57.92%	0.05	0.14	52.93%	0.05
1-4-3	0.235	0.31	33.94%	0.08	0.30	29.71%	0.07
1-4-4	0.423	0.51	20.75%	0.09	0.49	16.93%	0.07
1-4-5	0.639	0.73	13.65%	0.09	0.70	10.06%	0.06
1-4-6	0.878	0.96	9.17%	0.08	0.93	5.72%	0.05
1-4-7	1.142	1.21	5.59%	0.06	1.17	2.26%	0.03
1-4-8	1.426	1.47	2.87%	0.04	1.42	0.38%	0.01
1-4-9	1.699	1.74	2.44%	0.04	1.69	0.80%	0.01
1-4-10	2.013	2.03	0.63%	0.01	1.96	2.55%	0.05
1-4-11	2.31	2.32	0.51%	0.01	2.25	2.66%	0.06
1-4-12	2.632	2.63	0.14%	0.00	2.55	3.30%	0.09
1-4-13	2.942	2.94	0.08%	0.00	2.85	3.08%	0.09
1-4-14	3.281	3.27	0.34%	0.01	3.17	3.49%	0.11
1-4-15	3.643	3.60	1.07%	0.04	3.49	4.19%	0.15
1-4-16	4.013	3.95	1.65%	0.07	3.82	4.76%	0.19

Table 34. Simulations results and Formula 20 and 21 for Scenarios 1-5-1 to 1-5-13 (Set 5)

Scenario #	WIBT From simulation	WIBT from Formula 20	Percentage error% from Formula 20	Absolute error from Formula 20	WIBT from Formula 21	Percentage error% from Formula 21	Absolute error from Formula21
1-5-1	0	0.00	#DIV/0!	0.00	0.00	#DIV/0!	0.00
1-5-2	0.416	0.40	4.95%	0.02	0.39	5.46%	0.02
1-5-3	0.912	0.86	5.65%	0.05	0.86	6.16%	0.06
1-5-4	1.453	1.38	5.19%	0.08	1.37	5.71%	0.08
1-5-5	2.001	1.94	3.20%	0.06	1.93	3.72%	0.07
1-5-6	2.6	2.53	2.59%	0.07	2.52	3.12%	0.08
1-5-7	3.221	3.16	1.88%	0.06	3.14	2.41%	0.08
1-5-8	3.899	3.82	2.11%	0.08	3.80	2.64%	0.10
1-5-9	4.552	4.50	1.16%	0.05	4.48	1.69%	0.08
1-5-10	5.166	5.21	0.78%	0.04	5.18	0.23%	0.01
1-5-11	5.86	5.94	1.29%	0.08	5.90	0.74%	0.04
1-5-12	6.523	6.69	2.50%	0.16	6.65	1.94%	0.13
1-5-13	7.157	7.46	4.18%	0.30	7.42	3.61%	0.26

Table 35. Simulations results and Formula 20 and 21 for Scenarios 1-6-1 to 1-6-13 (Set 6)

Scenario #	WIBT From simulation	WIBT from Formula 20	Percentage error% from Formula 20	Absolute error from Formula 20	WIBT from Formula 21	Percentage error% from Formula 21	Absolute error from Formula21
1-6-1	0	0.00	#DIV/0!	0.00	0.00	#DIV/0!	0.00
1-6-2	0.23	0.25	9.90%	0.02	0.26	12.34%	0.03
1-6-3	0.524	0.54	3.17%	0.02	0.55	5.46%	0.03
1-6-4	0.86	0.85	0.70%	0.01	0.87	1.50%	0.01
1-6-5	1.223	1.19	2.89%	0.04	1.21	0.73%	0.01
1-6-6	1.577	1.54	2.44%	0.04	1.57	0.27%	0.00
1-6-7	1.986	1.90	4.10%	0.08	1.95	1.97%	0.04
1-6-8	2.382	2.28	4.12%	0.10	2.33	1.99%	0.05
1-6-9	2.813	2.68	4.90%	0.14	2.73	2.79%	0.08
1-6-10	3.222	3.08	4.48%	0.14	3.15	2.36%	0.08
1-6-11	3.67	3.49	4.90%	0.18	3.57	2.79%	0.10
1-6-12	4.089	3.91	4.32%	0.18	4.00	2.20%	0.09
1-6-13	4.538	4.34	4.29%	0.19	4.44	2.17%	0.10

Table 36. Simulations results and Formula 20 and 21 for Scenarios 1-7-1 to 1-7-13 (Set 7)

Scenario #	WIBT From simulation	WIBT from Formula 20	Percentage error% from Formula 20	Absolute error from Formula 20	WIBT from Formula 21	Percentage error% from Formula 21	Absolute error from Formula21
1-7-1	0	0.00	#DIV/0!	0.00	0.00	#DIV/0!	0.00
1-7-2	0.618	0.57	8.49%	0.05	0.59	3.79%	0.02
1-7-3	1.312	1.19	9.37%	0.12	1.25	4.71%	0.06
1-7-4	2.006	1.85	7.60%	0.15	1.95	2.85%	0.06
1-7-5	2.732	2.55	6.65%	0.18	2.68	1.85%	0.05
1-7-6	3.501	3.27	6.49%	0.23	3.44	1.68%	0.06
1-7-7	4.252	4.02	5.45%	0.23	4.23	0.59%	0.03
1-7-8	5.008	4.79	4.41%	0.22	5.03	0.51%	0.03
1-7-9	5.759	5.57	3.24%	0.19	5.86	1.73%	0.10
1-7-10	6.598	6.37	3.40%	0.22	6.70	1.57%	0.10
1-7-11	7.378	7.19	2.54%	0.19	7.56	2.47%	0.18
1-7-12	8.233	8.02	2.57%	0.21	8.43	2.44%	0.20
1-7-13	8.993	8.87	1.42%	0.13	9.32	3.65%	0.33

Table 37. Simulations results and Formula 20 and 21 for Scenarios 1-8-1 to 1-8-13 (Set 8)

Scenario #	WIBT From simulation	WIBT from Formula 20	Percentage error% from Formula 20	Absolute error from Formula 20	WIBT from Formula 21	Percentage error% from Formula 21	Absolute error from Formula21
1-8-1	0	0.00	#DIV/0!	0.00	0.00	#DIV/0!	0.00
1-8-2	0.55	0.49	11.45%	0.06	0.54	1.25%	0.01
1-8-3	1.103	0.99	10.22%	0.11	1.10	0.11%	0.00
1-8-4	1.707	1.50	11.88%	0.20	1.68	1.73%	0.03
1-8-5	2.269	2.03	10.69%	0.24	2.26	0.41%	0.01
1-8-6	2.863	2.56	10.75%	0.31	2.85	0.47%	0.01
1-8-7	3.455	3.09	10.58%	0.37	3.45	0.28%	0.01
1-8-8	4.064	3.63	10.71%	0.44	4.05	0.43%	0.02
1-8-9	4.667	4.17	10.61%	0.49	4.65	0.31%	0.01
1-8-10	5.269	4.72	10.43%	0.55	5.26	0.12%	0.01
1-8-11	5.883	5.27	10.42%	0.61	5.88	0.11%	0.01
1-8-12	6.536	5.82	10.90%	0.71	6.49	0.64%	0.04
1-8-13	7.137	6.38	10.60%	0.76	7.11	0.31%	0.02

Table 38. Simulations results and Formula 20 and 21 for Scenarios 2-1-1 to 2-1-13 (Set 9)

Scenario #	WIBT From simulation	WIBT from Formula 20	Percentage error% from Formula 20	Absolute error from Formula 20	WIBT from Formula 21	Percentage error% from Formula 21	Absolute error from Formula21
2-1-1	2.495	2.944	18.00%	0.45	2.822	13.09%	0.33
2-1-2	1.099	1.116	1.51%	0.02	1.069	2.72%	0.03
2-1-3	0.646	0.632	2.10%	0.01	0.606	6.18%	0.04
2-1-4	0.434	0.423	2.59%	0.01	0.405	6.65%	0.03
2-1-5	0.315	0.309	1.80%	0.01	0.296	5.89%	0.02
2-1-6	0.241	0.240	0.57%	0.00	0.230	4.71%	0.01
2-1-7	0.192	0.193	0.58%	0.00	0.185	3.61%	0.01
2-1-8	0.158	0.160	1.39%	0.00	0.154	2.84%	0.00
2-1-9	0.132	0.136	2.91%	0.00	0.130	1.38%	0.00
2-1-10	0.11	0.117	6.55%	0.01	0.112	2.11%	0.00
2-1-11	0.094	0.103	9.12%	0.01	0.098	4.57%	0.00
2-1-12	0.082	0.091	10.74%	0.01	0.087	6.12%	0.01
2-1-13	0.071	0.081	14.34%	0.01	0.078	9.57%	0.01

Table 39. Simulations results and Formula 20 and 21 for Scenarios 2-2-1 to 2-2-13 (Set 10)

Scenario #	WIBT From simulation	WIBT from Formula 20	Percentage error% from Formula 20	Absolute error from Formula 20	WIBT from Formula 21	Percentage error% from Formula 21	Absolute error from Formula21
2-2-1	1.504	1.488	1.06%	0.02	1.556	3.43%	0.05
2-2-2	0.689	0.648	6.00%	0.04	0.677	1.72%	0.01
2-2-3	0.427	0.398	6.75%	0.03	0.416	2.51%	0.01
2-2-4	0.303	0.282	6.96%	0.02	0.295	2.73%	0.01
2-2-5	0.232	0.216	7.03%	0.02	0.226	2.80%	0.01
2-2-6	0.188	0.173	7.81%	0.01	0.181	3.62%	0.01
2-2-7	0.156	0.144	7.67%	0.01	0.151	3.47%	0.01
2-2-8	0.129	0.123	4.87%	0.01	0.128	0.55%	0.00
2-2-9	0.11	0.107	3.14%	0.00	0.111	1.26%	0.00
2-2-10	0.098	0.094	4.20%	0.00	0.098	0.16%	0.00
2-2-11	0.088	0.084	4.84%	0.00	0.088	0.52%	0.00
2-2-12	0.075	0.075	0.58%	0.00	0.079	5.16%	0.00
2-2-13	0.071	0.069	3.48%	0.00	0.072	0.91%	0.00

Table 40. Simulations results and Formula 20 and 21 for Scenarios 3-1-1 to 3-1-13 (Set 11)

Scenario #	WIBT From simulation	WIBT from Formula 20	Percentage error% from Formula 20	Absolute error from Formula 20	WIBT from Formula 21	Percentage error% from Formula 21	Absolute error from Formula21
3-1-1	1.799	1.617	10.13%	0.18	1.823	1.33%	0.02
3-1-2	0.692	0.631	8.83%	0.06	0.576	16.79%	0.12
3-1-3	0.28	0.244	12.74%	0.04	0.210	25.11%	0.07
3-1-4	0.11	0.114	3.91%	0.00	0.095	13.41%	0.01
3-1-5	0.053	0.061	15.96%	0.01	0.050	5.02%	0.00
3-1-6	0.023	0.036	58.69%	0.01	0.030	28.52%	0.01
3-1-7	0.01	0.023	133.21%	0.01	0.019	87.36%	0.01
3-1-8	0.004	0.016	293.86%	0.01	0.013	214.54%	0.01
3-1-9	0.002	0.011	455.89%	0.01	0.009	341.89%	0.01
3-1-10	0.001	0.008	712.60%	0.01	0.006	543.59%	0.01
3-1-11	0	0.006	#DIV/0!	0.01	0.005	#DIV/0!	0.00
3-1-12	0	0.005	#DIV/0!	0.00	0.004	#DIV/0!	0.00
3-1-13	0	0.004	#DIV/0!	0.00	0.003	#DIV/0!	0.00

Table 41. Simulations results and Formula 20 and 21 for Scenarios 3-2-1 to 3-2-13 (Set 12)

Scenario #	WIBT From simulation	WIBT from Formula 20	Percentage error% from Formula 20	Absolute error from Formula 20	WIBT from Formula 21	Percentage error% from Formula 21	Absolute error from Formula21
3-2-1	2.498	2.417	3.23%	0.08	2.647	5.98%	0.15
3-2-2	1.036	1.007	2.78%	0.03	0.908	12.31%	0.13
3-2-3	0.496	0.404	18.62%	0.09	0.344	30.68%	0.15
3-2-4	0.244	0.193	21.08%	0.05	0.160	34.59%	0.08
3-2-5	0.117	0.105	10.36%	0.01	0.086	26.89%	0.03
3-2-6	0.055	0.063	14.27%	0.01	0.051	7.78%	0.00
3-2-7	0.031	0.040	30.41%	0.01	0.032	4.46%	0.00
3-2-8	0.017	0.027	61.49%	0.01	0.022	28.63%	0.00
3-2-9	0.009	0.019	116.16%	0.01	0.015	71.44%	0.01
3-2-10	0.004	0.014	256.71%	0.01	0.011	181.93%	0.01
3-2-11	0.002	0.011	438.21%	0.01	0.008	324.19%	0.01
3-2-12	0.001	0.008	731.43%	0.01	0.007	553.78%	0.01
3-2-13	0.001	0.007	555.18%	0.01	0.005	414.17%	0.00

As one see, for Sets 1 to 12, in general the absolute error for Formula 21 is smaller than Formula 20, so Formula 21 is more acceptable.

The worse absolute error occurs in scenarios that have a single server. We know that the WIBT for a single server is: $WIBT = \frac{(\overline{K}_{Ai} - 1)t_i}{2}$ (Formula 4)

Finally, we will use these formulas to estimate WIBT in this thesis:

$$WIBT = \frac{(\overline{K}_{Ai}^{\,2-u_i} - 1)t_i u_i}{(3-u_i)m_i^{\,2-u_i}}$$ **(Formula 21)** when $m_i > 1$.

$$WIBT = \frac{(\overline{K}_{Ai} - 1)t_i}{2}$$ **(Formula 4)** when $m_i = 1$.

4.3 Self service station (stations with infinite number of servers)

In this case, residents arrive individually to the workstation. The residents perform the process themselves without any external resources. In this domain, an example would be a workstation where each resident must complete a form. As we mentioned in Chapter 2, the self service workstation can be modeled as a $G/G/\infty$ queueing system. We will use the following notation:

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

ρ_i = Load

c_{ai}^2 = interarrival time SCV at station i

t_i = Mean process time at station i (minutes)

c_{ei}^2 = Process time SCV at station i

c_{di}^2 = Interdeparture time SCV at station i

We have mentioned some facts about the departure variability in regard to $G/G/\infty$ queueing systems in Chapter 2. Here, we summarize them as useful points as our initial knowledge about the station with infinite servers before any further experiments in this section.

- For a $G/D/\infty$ system, the time interdeparture variability equals the interarrival time variability.

- For a $M/G/\infty$ system, the departure process is a Poisson process; thus the interdeparture time variability equals 1.
- For a $G/G/\infty$ system, the interdeparture time variability approaches 1 as the load (the arrival rate divided by the service rate) goes to infinity.
- For $G/G/\infty$ system where the load is close to 0, the interdeparture time variability is equal to the interarrival time variability.

According to afore-mentioned points, in the general case (a $G/G/\infty$ system with moderate load), the interdeparture time variability will be somewhere between the interarrival time variability and one. Therefore, we conducted experiments to characterize this relationship and to examine various weights for interpolating between the interarrival time variability and one as a function of the load $\rho_i = r_i t_i$. The general form of the interpolation is

$$c_{di}^2 = (1-\omega)c_{ai}^2 + \omega \qquad 0 \le \omega \le 1$$

Note that, if the arrival variability equals 1, then (for any weight), the interdeparture time variability equals 1. The purpose of the experiments was to evaluate various functions that could be used to determine the weight for this interpolation. Three candidates were tried:

$$\omega_a = \frac{\rho_i^2 c_{ei}^2}{\left(1+\rho_i\sqrt{c_{ei}^2}\right)^2}$$

$$\omega_b = \frac{\rho_i^2 c_{ei}^2}{1+\rho_i^2 c_{ei}^2}$$

$$\omega_c = \frac{\rho_i c_{ei}^2}{1+\rho_i c_{ei}^2}$$

All of these have the following desirable properties:

1. As the process time variability goes to 0, the weight goes to 0, and the interdeparture time variability approaches the arrival variability.
2. As the load goes to 0, the weight goes to 0, and the interdeparture time variability approaches the interarrival time variability.
3. As the load goes to infinity, the weight goes to 1, and the interdeparture time variability approaches 1.

Based on the results (discussed in the next section), we decided to use ω_a, which yields the following approximation:

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

In this section, we will only study cases with individual arrivals to the self service station. However, in our modeling section in Chapter 5, we include the case of mixed arrival to the self service station.

In this type of the queueing system, we will assume that the arrival batch size doesn't have any variability. In other words, if we have mixed arrival with the average batch size of $\overline{K}_{Ai}$, its variability (SCV) is close to zero, that we can ignore it in our calculations. Therefore, according to the afore-mentioned c_{di}^2 for individual arrival to self service using ω_a and also Formula 2 (first method unbatching) from Section 2.2.4, we can approximate the interdepartue time SCV from self service stations with batch arrivals having batch interarrival time SCV of c_{bi}^2 and the average arrival batch size of $\overline{K}_{Ai}$ with nearly zero variability (SCV).

The formula is this estimate:

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

4.3.1 Self service experiments

To evaluate these weights, we conducted sets of computational experiments using a discrete-event simulation model of the station. The simulation model has only stations with a simple delay. In other words, whenever a customer comes to the self service station, s/he will be held in the station by the time that the delay time ends.

In all experiments, we ran five replications and measured the interdeparture times of the residents. We then calculated the interdeparture time SCV for each replication and calculated 95% confidence intervals. The run lengths and warm-up periods were proportional to the mean interarrival time as indicated below.

In the first set (which we denote as Set DE), the interarrival times were constant, and the processing times were exponentially distributed. The mean interarrival time went from 0.0006 minutes to 100 minutes. The mean processing time was 3 minutes in all scenarios. Thus, the load varied from 0.03 to 5000. The run length was set equal to 260,000 times the mean interarrival time, and the warm-up period was set equal to 200,000 times the mean interarrival time.

In the second set (Set GE), the interarrival times had a gamma distribution, and the processing times were exponentially distributed. The mean interarrival time went from 0.04 minutes to 40 minutes. α parameter was always equal to 0.2, so the interarrival time variability was always 5. The mean processing time was 3 minutes in all scenarios. Thus, the load varied from 0.075 to 750. The run length was set equal to 110,000 times the mean interarrival time, and the warm-up period was set equal to 50,000 times the mean interarrival time.

In the third set (Set EG), the interarrival times were exponentially distributed, and the processing times had a gamma distribution. The mean interarrival time was always 4 minutes. The mean processing time varied from 0.05 to 2000 minutes. α parameter was always equal to 0.5, so the processing time variability was always 2. Thus, the load varied from 0.0125 to 500. The run length was set equal to 865,000 times the mean interarrival time, and the warm-up period was set equal to 800,000 times the mean interarrival time.

In the fourth set (Set GG), the interarrival times had a gamma distribution, and the processing times had a gamma distribution. The mean interarrival time was always 4 minutes. α parameter was always 0.1, so the interarrival time variability was always 10. The mean processing time varied from 0.25 to 2000 minutes. α parameter was always equal to 0.5, so the processing time variability was always 2. Thus, the load varied from 0.0625 to 500. The run length was set equal to 1,315,000 times the mean interarrival time, and the warm-up period was set equal to 1,250,000 times the mean interarrival time.

In the fifth set (Set UG), the interarrival times had a uniform distribution, and the processing times had a gamma distribution. The interarrival time distribution was always between 3 and 5 minutes. Thus, the arrival variability was always 0.02. The mean processing time varied from 0.25 to 2000 minutes. α parameter was always equal to 0.5, so the processing time variability was always 2. Thus, the load varied from 0.0625 to 500. The run length was set equal to 140,000 times the mean interarrival time, and the warm-up period was set equal to 75,000 times the mean interarrival time.

Tables 42 to 46 present the results for sets DE, GE, EG, GG and UG. For each scenario, the table lists the load, the interdeparture time SCV from the simulation, the lower and upper bound on the confidence interval. In addition, it provides the three interdeparture time SCV estimates (one using each weight) and the relative error for each.

Figures 39 to 43 compare the estimates of c_{di}^2 using ω_a (Formula A) to the simulation results.

Table 42. Interdeparture time variability results for Set DE

Load	SCV (from sim)	Lower bound	Upper bound	SCV with ω_a	Relative error	SCV with ω_b	Relative error	SCV with ω_c	Relative error
5000	1.0055	0.027	1.0325	0.9996	0.590%	1.0000	0.550%	0.9998	0.570%
3000	1.0081	0.0258	1.0339	0.9993	0.873%	1.0000	0.807%	0.9997	0.840%
1000	1.0014	0.03	1.0314	0.9980	0.343%	1.0000	0.143%	0.9990	0.243%
500	0.9956	0.023	1.0185	0.9960	0.043%	1.0000	0.443%	0.9980	0.243%
400	1.0027	0.0268	1.0295	0.9950	0.771%	1.0000	0.275%	0.9975	0.523%
300	1.0077	0.0271	1.0348	0.9934	1.420%	1.0000	0.763%	0.9967	1.092%
200	0.9882	0.0245	1.0127	0.9901	0.195%	1.0000	1.197%	0.9950	0.696%
100	1.0012	0.0303	1.0316	0.9803	2.092%	0.9999	0.134%	0.9901	1.113%
60	0.9776	0.019	0.9966	0.9675	1.032%	0.9997	2.266%	0.9836	0.617%
30	0.9545	0.0158	0.9703	0.9365	1.879%	0.9989	4.656%	0.9677	1.392%
6	0.8499	0.0199	0.8697	0.7347	13.553%	0.9730	14.484%	0.8571	0.855%
3	0.7564	0.0215	0.7778	0.5625	25.630%	0.9000	18.991%	0.7500	0.841%
2	0.6796	0.0176	0.6972	0.4444	34.599%	0.8000	17.722%	0.6667	1.898%
1.5	0.5933	0.0143	0.6076	0.3600	39.325%	0.6923	16.683%	0.6000	1.126%
0.75	0.3957	0.0099	0.4056	0.1837	53.578%	0.3600	9.012%	0.4286	8.319%
0.5	0.2753	0.0045	0.2798	0.1111	59.642%	0.2000	27.355%	0.3333	21.07%
0.3	0.1446	0.0013	0.1458	0.0533	63.160%	0.0826	42.881%	0.2308	59.64%
0.2	0.0741	0.0008	0.0749	0.0278	62.513%	0.0385	48.095%	0.1667	124.9%
0.15	0.0434	0.0007	0.0441	0.0170	60.754%	0.0220	49.239%	0.1304	200.8%
0.12	0.0280	0.0006	0.0286	0.0115	59.058%	0.0142	49.372%	0.1071	282.1%
0.1	0.0195	0.0004	0.0199	0.0083	57.639%	0.0099	49.251%	0.0909	365.9%
0.03	0.0018	0.0003	0.0018	0.0008	51.524%	0.0009	48.618%	0.0291	1564.%

Table 43. Interdeparture time variability results for Set GE

Load	SCV (from sim)	Lower bound	Upper bound	SCV with ω_a	Relative error	SCV with ω_b	Relative error	SCV with ω_c	Relative error
750	0.9995	0.9755	1.0234	1.0106	1.115%	1.0000	0.051%	1.0053	0.583%
250	1.0074	0.9657	1.0491	1.0318	2.423%	1.0001	0.728%	1.0159	0.847%
125	1.0299	1.0144	1.0453	1.0632	3.237%	1.0003	2.878%	1.0317	0.179%
100	1.0075	0.9505	1.0645	1.0788	7.078%	1.0004	0.705%	1.0396	3.186%
75	1.0689	1.0372	1.1007	1.1046	3.337%	1.0007	6.379%	1.0526	1.522%
50	1.0607	1.042	1.0795	1.1553	8.921%	1.0016	5.572%	1.0784	1.672%
25	1.1834	1.1294	1.2374	1.3018	10.003%	1.0064	14.958%	1.1538	2.497%
15	1.3269	1.2303	1.4235	1.4844	11.868%	1.0177	23.303%	1.2500	5.795%
7.5	1.8037	1.5534	2.0539	1.8858	4.552%	1.0699	40.685%	1.4706	18.468%
1.5	3.7072	3.2701	4.1444	3.5600	3.971%	2.2308	39.826%	2.6000	29.866%
0.75	4.328	3.8798	4.7763	4.2653	1.449%	3.5600	17.745%	3.2857	24.082%
0.5	4.5708	4.1191	5.0226	4.5556	0.334%	4.2000	8.112%	3.6667	19.781%
0.375	4.6949	4.2422	5.1475	4.7025	0.161%	4.5068	4.005%	3.9091	16.738%
0.187	4.8676	4.3642	5.371	4.9003	0.671%	4.8642	0.071%	4.3684	10.255%
0.125	4.9324	4.4765	5.3882	4.9506	0.369%	4.9385	0.123%	4.5556	7.640%
0.075	4.9721	4.5159	5.4284	4.9805	0.170%	4.9776	0.111%	4.7209	5.052%

Table 44. Interdeparture time variability results for Set EG

Load	SCV (from sim)	Lower bound	Upper bound	SCV with ω_a	Relative error	SCV with ω_b	Relative error	SCV with ω_c	Relative error
500	0.9892	0.9729	1.0056	1.00	1.089%	1.00	1.089%	1.00	1.089%
250	0.9924	0.9737	1.0112	1.00	0.761%	1.00	0.761%	1.00	0.761%
125	0.9993	0.9714	1.0273	1.00	0.063%	1.00	0.063%	1.00	0.063%
87.5	1.0132	0.9892	1.0373	1.00	1.308%	1.00	1.308%	1.00	1.308%
62.5	0.9946	0.9686	1.0206	1.00	0.543%	1.00	0.543%	1.00	0.543%
50	0.9898	0.9769	1.0027	1.00	1.029%	1.00	1.029%	1.00	1.029%
37.5	0.9933	0.9717	1.015	1.00	0.670%	1.00	0.670%	1.00	0.670%
25	0.9972	0.962	1.0324	1.00	0.278%	1.00	0.278%	1.00	0.278%
12.5	0.9897	0.9704	1.0091	1.00	1.037%	1.00	1.037%	1.00	1.037%
6.25	0.9947	0.9679	1.0215	1.00	0.531%	1.00	0.531%	1.00	0.531%
2.5	0.9947	0.9831	1.0064	1.00	0.527%	1.00	0.527%	1.00	0.527%
1.25	0.9893	0.9764	1.0022	1.00	1.080%	1.00	1.080%	1.00	1.080%
0.625	0.9904	0.9694	1.0115	1.00	0.960%	1.00	0.960%	1.00	0.960%
0.125	0.9995	0.9802	1.0188	1.00	0.046%	1.00	0.046%	1.00	0.046%
0.062	1.0010	0.9822	1.02	1.00	0.107%	1.00	0.107%	1.00	0.107%
0.01	1.0021	0.984	1.0203	1.00	0.214%	1.00	0.214%	1.00	0.214%

Table 45. Interdeparture time variability results for Set GG

Load	SCV (from sim)	Lower bound	Upper bound	SCV with ω_a	Relative error	SCV with ω_b	Relative error	SCV with ω_c	Relative error
500	1.0172	0.9984	1.0359	1.0254	0.806%	1.0000	1.689%	1.0090	0.807%
250	1.0557	1.0254	1.086	1.0507	0.474%	1.0001	5.269%	1.0180	3.574%
125	1.0935	1.0588	1.1281	1.1010	0.683%	1.0003	8.524%	1.0359	5.271%
87.5	1.158	1.1211	1.1949	1.1437	1.233%	1.0006	13.593%	1.0511	9.228%
62.5	1.2288	1.1585	1.2991	1.2002	2.324%	1.0012	18.526%	1.0714	12.807%
50	1.2752	1.1668	1.3837	1.2493	2.034%	1.0018	21.440%	1.0891	14.593%
37.5	1.4536	1.2319	1.6752	1.3300	8.500%	1.0032	30.985%	1.1184	23.059%
25	1.7081	1.3121	2.1041	1.4883	12.868%	1.0072	41.034%	1.1765	31.124%
12.5	2.723	1.667	3.779	1.9379	28.831%	1.0287	62.222%	1.3462	50.564%
6.25	4.2706	2.6963	5.845	2.7365	35.922%	1.1137	73.921%	1.6667	60.973%
2.5	6.9288	4.9615	8.8961	4.5312	34.604%	1.6667	75.946%	2.5000	63.919%
1.25	8.5373	6.5603	10.514	6.3286	25.871%	3.1818	62.730%	3.5714	58.167%
0.625	9.7565	7.7646	11.748	8.0188	17.811%	6.0526	37.963%	5.0000	48.752%
0.125	10.730	8.7557	12.704	9.7969	8.697%	9.7273	9.346%	8.2000	23.579%
0.062	10.839	8.868	12.810	9.9415	8.283%	9.9313	8.377%	9.0071	16.903%

Table 46. Interdeparture time variability results for Set UG

Load	**SCV (from sim)**	**Lower bound**	**Upper bound**	**SCV with ω_a**	**Relative error**	**SCV with ω_b**	**Relative error**	**SCV with ω_c**	**Relative error**
500	0.9884	0.9591	1.0178	0.9972	0.894%	1.0000	1.173%	0.9990	1.075%
250	0.9938	0.9752	1.0125	0.9945	0.069%	1.0000	0.623%	0.9980	0.427%
125	0.9809	0.9551	1.0067	0.9890	0.827%	1.0000	1.944%	0.9961	1.549%
87.5	0.979	0.9636	0.9943	0.9844	0.548%	0.9999	2.139%	0.9944	1.577%
62.5	0.9727	0.9474	0.998	0.9782	0.567%	0.9999	2.794%	0.9922	2.008%
50	0.9703	0.9604	0.9802	0.9729	0.266%	0.9998	3.041%	0.9903	2.062%
37.5	0.9564	0.9403	0.9724	0.9641	0.804%	0.9997	4.522%	0.9871	3.212%
25	0.9402	0.9211	0.9594	0.9469	0.710%	0.9992	6.277%	0.9808	4.318%
12.5	0.8928	0.8797	0.9059	0.8980	0.578%	0.9969	11.657%	0.9623	7.789%
6.25	0.8288	0.8157	0.842	0.8111	2.139%	0.9876	19.163%	0.9275	11.905%
2.5	0.6933	0.6852	0.7013	0.6158	11.175%	0.9275	33.776%	0.8368	20.699%
1.25	0.5455	0.5358	0.5552	0.4203	22.957%	0.7626	39.803%	0.7202	32.032%
0.625	0.3736	0.3661	0.3811	0.2364	36.730%	0.4503	20.527%	0.5648	51.181%
0.125	0.0758	0.0738	0.0777	0.0429	43.369%	0.0505	33.375%	0.2167	185.83%
0.062	0.0366	0.0358	0.0374	0.0273	25.443%	0.0284	22.348%	0.1296	254.17%

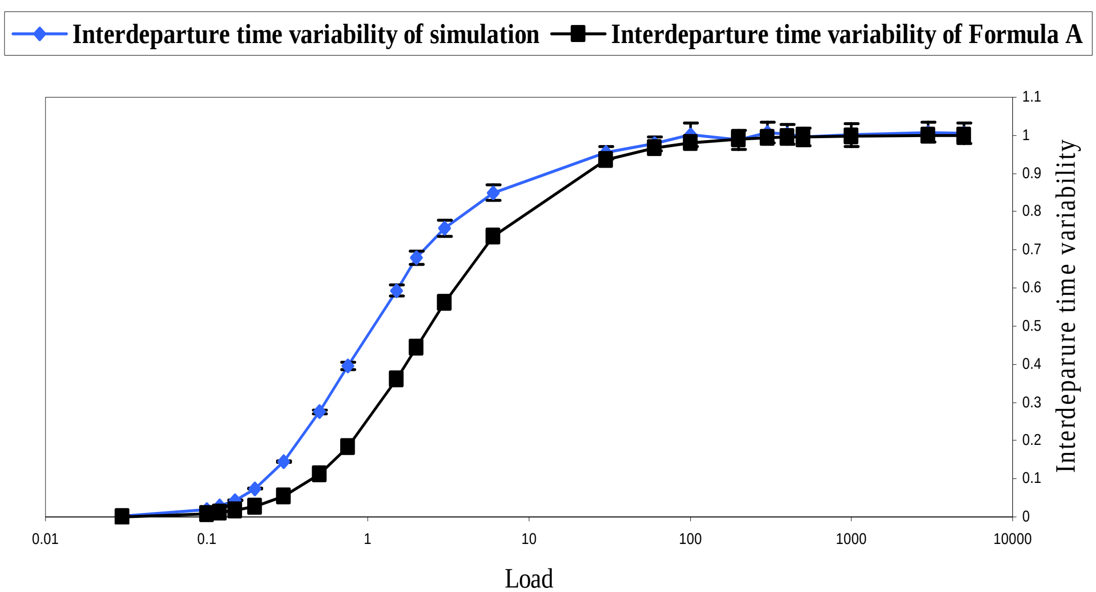

Figure 39. Interdeparture time variability results for Set DE

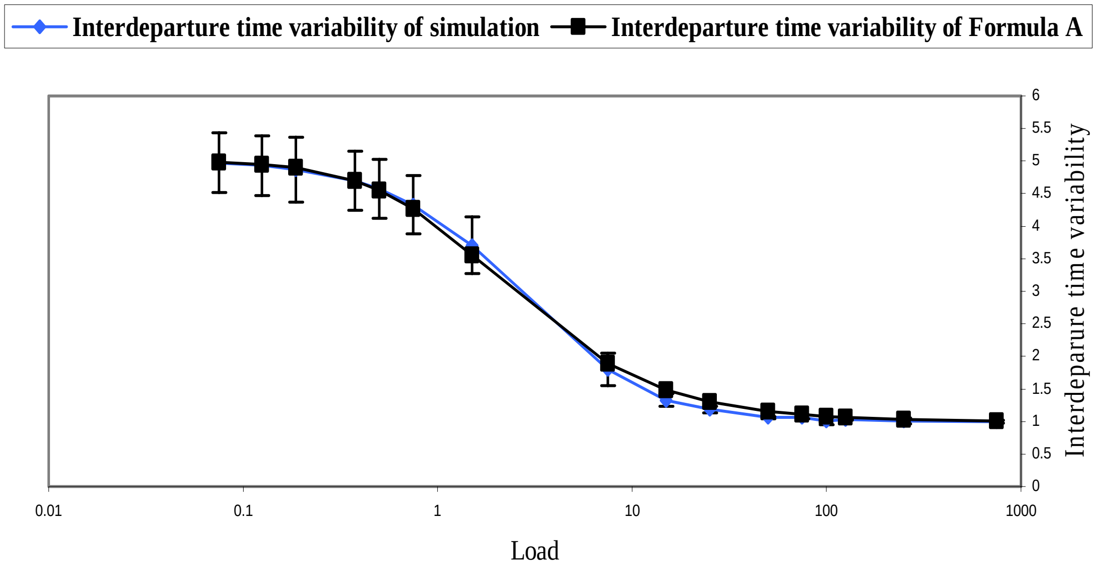

Figure 40. Interdeparture time variability results for Set GE

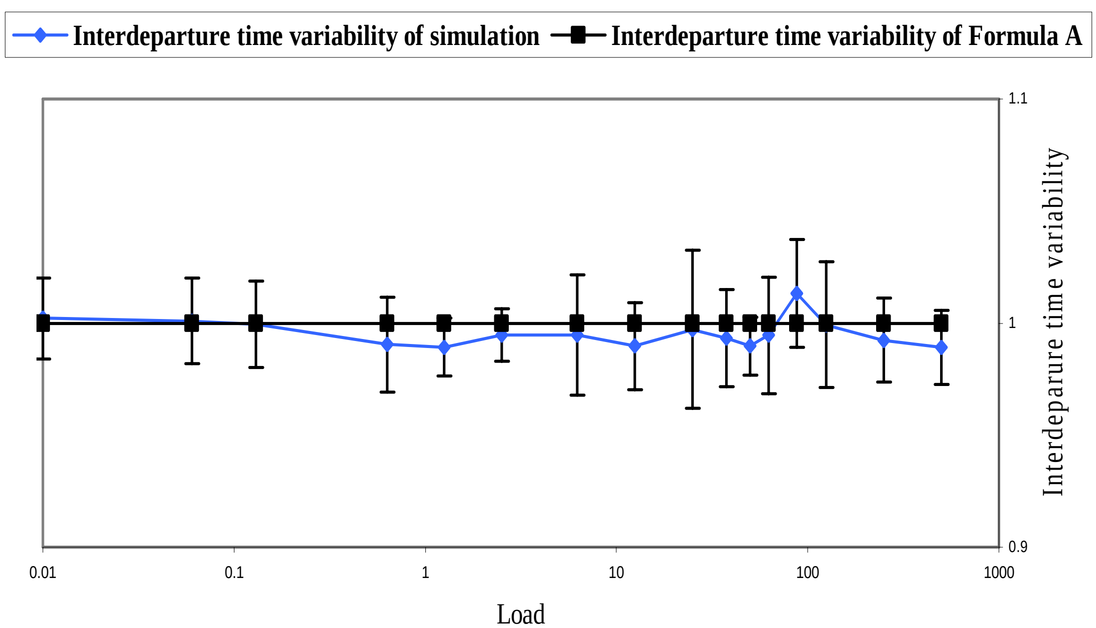

Figure 41. Interdeparture time variability results for Set EG

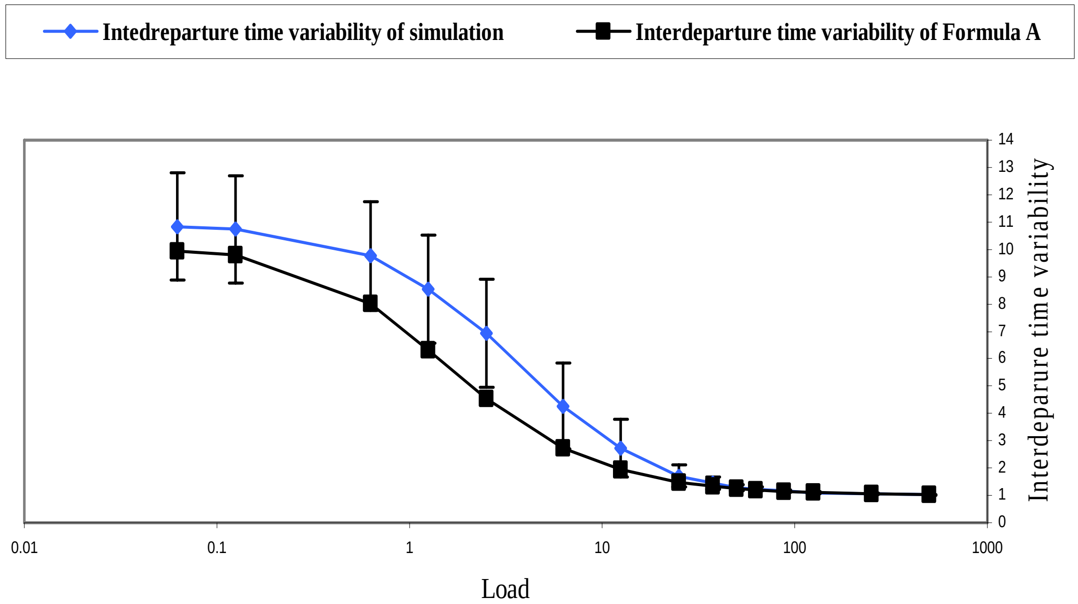

Figure 42. Interdeparture time variability results for Set GG

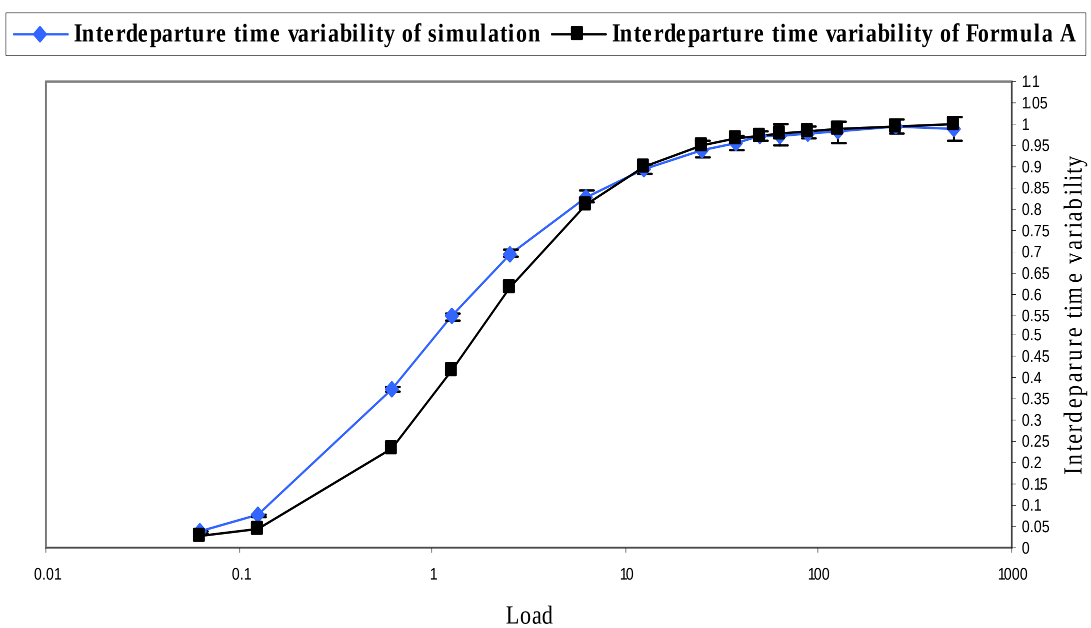

Figure 43. Interdeparture time variability results for Set UG

4.4 Batch formation process

As noted in Chapter 2, it is necessary at a batch service workstation to form process batches. Arriving items (jobs or customers) wait in an incomplete batch until the proper quantity has accumulated, and then the full batch is formed and joins the queue for service. This waiting time to form a batch was defined in Chapter 2 as wait-to-batch time (WTBT).

To analyze batch service workstations in our clinic models, we need to have both wait-to-batch time (WTBT) and batch interarrival time SCV for formed batches after the batch formation process. After the batch is formed, queueing can be approximated using the formulas previously discussed, substituting parameters pertaining to the batch for the individual parameters.

For individual arrivals, the interarrival time SCV of formed batches is obtained by dividing the individual arrival SCV by k (Hopp and Spearman, 2001). However, to model our clinic thoroughly, in this chapter, we calculate the interarrival time variability for batches formed from any kind of arrivals from more than one stream with batch size variability. In this thesis, we assume that the process batches are larger than the arrival batches

Additionally, again from Chapter 2, we had Formula 5 to calculate WTBT for cases with individual arrivals. In this section, we introduce our new approach to estimate WTBT for the cases with batch arrivals from multiple arrival streams.

4.4.1 Batch formation variability

In this case, residents arrive in batches (arrival batches) to the workstation. The arriving batches may come from multiple workstations and may have different batch sizes with batch size variability. Arriving residents are grouped into process batches of a given size to perform the process. There may be multiple servers that can process different batches in parallel. We assume that the process batches are larger than the arrival batches. In this domain, an example is a workstation where residents must view an educational video. The process batch is the group of residents watching the video at the same time.

Arriving residents enter a batch formation queue. A process batch is formed whenever there are k_i residents waiting in this queue. These residents then leave this queue, and the newly formed process batch enters a process queue, where it waits for a server to process it.

We will use the following notation:

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

c_{bi}^2 = Interarrival time SCV for process batches at station i (after being formed)

k_i = Processing batch size at station i

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

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

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

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

c_{Bji}^2 = Interarrival time SCV for batches that come to station i from station j

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

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

A key quantity for estimating the performance of such a workstation is the variability associated with the formation of process batches. The time between two consecutive process batches forming is a random variable with a SCV of c_{bi}^2, which we call the batch formation variability (SCV). There is no established estimate for this term. Thus, we developed and tested four different estimates X_i^h, for h = 1, 2, 3, and 4.

Next, we consider two special cases. First, if all of the residents arrive individually, then it is easy to see that the variability is pooled (Hopp and Spearman, 2001):

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

Second, if all of the arrival batches have exactly $\bar{K}_{Ai}$ residents, then each process batch has exactly $k_i / \bar{K}_{Ai}$ arrival batches:

$$c_{bi}^2 = X_i^1 = \frac{\bar{K}_{Ai} c_{ai}^2}{k_i}$$

In general, however, the size of the arrival batches varies and has SCV of the batch size of $\bar{C}_{Ai}^2$. Thus, in the general case, the above equation is only an approximation.

Intuitively it is clear that the arrival batch size variability $\bar{C}_{Ai}^2$ affects the batch formation variability. Therefore, we decided to create and test a second estimate:

$$X_i^2 = \frac{\bar{K}_{Ai}\left(c_{ai}^2 + \bar{C}_{Ai}^2\right)}{k_i}$$

4.4.1.1 SCV of the batch size of all batches that come to a station

Before running some tests to calculate the values of X_i^2, we need to have SCV of the batch size of all batches that come to a station ($\bar{C}_{Ai}^2$). We calculate the variability of the arriving batch size by adapting a formula from Fowler et al. (2002), who calculate the process time SCV 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 batch size of coming batches which is:

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

The next section will discuss the results of the tests calculating the value of X_i^2 for some scenarios.

4.4.1.2 Initial batch formation experiments

To evaluate the two estimates (X_i^1, X_i^2), we conducted sets of computational experiments using a discrete-event simulation model of the station. Each simulation replication was 150,000 to 600,000 minutes long, with a warm-up period of 100,000 to 400,000 minutes. Ten replications were conducted for each scenario. Simulation results are shown as 95% confidence intervals.

Initially, seven scenarios were tested. In all of the scenarios, three workstations (1, 2, and 3) sent batches to a fourth workstation, which is the workstation of interest. This forms three arrival streams, one from each workstation. In this set of scenarios, the batch size for each arrival stream is a constant (that is, all of the batches in each arrival stream has the same number of residents), and the interarrival times are exponentially distributed. The batch sizes and mean interarrival times for each stream were changed. The process batch size k_i varied as well. Table 47 describes the seven scenarios, and table 48 describes the results for the scenarios.

Table 47. Description of Scenarios 1 to 7

Scenario	k_i	Mean interrarrival times (mins)			Arrival batch size		
		1	2	3	1	2	3
1	10	6	7	10	1	2	2
2	10	6	7	10	1	3	5
3	10	6	7	10	2	4	6
4	10	6	7	10	8	1	5
5	15	6	4	10	8	7	6
6	30	6	4	10	6	2	12
7	30	6	4	10	3	11	7

Table 48. Results for Scenarios 1 to 7

Scenario	Batch formation variability estimates		Confidence interval on batch formation variability from simulation results			Relative error	
	X_i^1	X_i^2	c_{bi}^2 from Simulation	Lower bound	Upper bound	X_i^1	X_i^2
1	0.159	0.174	0.173	0.165	0.182	8.128%	0.609%
2	0.267	0.361	0.37	0.361	0.379	27.718%	2.468%
3	0.367	0.435	0.452	0.443	0.461	18.689%	3.641%
4	0.483	0.673	0.722	0.713	0.732	33.183%	6.746%
5	0.467	0.474	0.504	0.496	0.512	7.377%	6.002%
6	0.236	0.313	0.325	0.318	0.332	27.277%	3.753%
7	0.221	0.272	0.283	0.275	0.292	21.907%	4.088%

Next, we tested scenarios in which the interarrival time distributions of each arrival stream in Scenario 4 were changed in order to vary the variability in each arrival stream. The mean interarrival times and other parameters remained as specified for Scenario 4, and the other two arrival streams kept exponentially distributed interarrival times. Scenarios 4.1.1 to 4.1.8 changed the first arrival stream as shown in Table 49. (Note Scenario 4.1.4 is the same as the original Scenario 4.)

Table 49. Description of Scenarios 4.1.1 to 4.1.8

Scenario	Interrarrival time distribution	Interarrival time variability (SCV)
4.1.1	Constant	0
4.1.2	Gamma(2, 3.5)	0.5
4.1.3	Gamma(4/3, 21/4)	0.75
4.1.4	Exponential	1
4.1.5	Gamma(2/3, 21/2)	1.5
4.1.6	Gamma(1/2, 14)	2
4.1.7	Gamma(1/4, 28)	4
4.1.8	Gamma(1/5, 35)	5

Scenarios 4.2.1 through 4.2.8 modified the interarrival time distributions of the second arrival stream to increase the arrival variability in the same way. The mean interarrival time remained 7 minutes for these eight scenarios. Likewise, Scenarios 4.3.1 through 4.3.8 modified the interarrival time distributions of the third arrival stream to increase the arrival variability in the same way. The mean interarrival time remained 10 minutes for these eight scenarios.

Table 50. Results for Scenarios 4.1.1 to 4.1.8

Scenario	Batch formation variability estimates		Confidence interval on batch formation variability from simulation results			Relative error	
	X_i^1	X_i^2	c_{bi}^2 from Simulation	Lower bound	Upper bound	X_i^1	X_i^2
4.1.1	0.29	0.48	0.21	0.201	0.219	36.5%	127.6%
4.1.2	0.38	0.58	0.50	0.495	0.513	23.7%	14.2%
4.1.3	0.43	0.62	0.63	0.620	0.638	31.1%	0.7%
4.1.4	0.48	0.67	0.72	0.713	0.731	33.2%	6.7%
4.1.5	0.58	0.77	0.90	0.893	0.911	35.6%	14.4%
4.1.6	0.68	0.87	1.04	1.027	1.045	34.5%	16.0%
4.1.7	1.07	1.26	1.36	1.355	1.373	21.4%	7.4%
4.1.8	1.27	1.46	1.48	1.472	1.490	14.4%	1.5%

Table 51. Results for Scenarios 4.2.1 to 4.2.8

Scenario	Batch formation variability estimates		Confidence interval on batch formation variability from simulation results			Relative error	
	X_i^1	X_i^2	c_{bi}^2 from Simulation	Lower bound	Upper bound	X_i^1	X_i^2
4.2.1	0.31	0.51	0.717	0.708	0.726	56.2%	29.5%
4.2.2	0.40	0.59	0.707	0.698	0.716	43.7%	16.7%
4.2.3	0.44	0.63	0.727	0.718	0.736	39.4%	13.1%
4.2.4	0.48	0.67	0.722	0.713	0.731	33.2%	6.7%
4.2.5	0.57	0.76	0.739	0.730	0.748	23.3%	2.5%
4.2.6	0.65	0.84	0.737	0.728	0.745	11.6%	14.3%
4.2.7	0.99	1.18	0.754	0.745	0.763	31.0%	56.4%
4.2.8	1.16	1.35	0.743	0.714	0.771	55.6%	81.4%

Table 52. Results for Scenarios 4.3.1 to 4.3.8

Scenario	Batch formation variability estimates		Confidence interval on batch formation variability from simulation results			Relative error	
	X_i^1	X_i^2	c_{bi}^2 from Simulation	Lower bound	Upper bound	X_i^1	X_i^2
4.3.1	0.36	0.56	0.62	0.612	0.636	41.5%	10.9%
4.3.2	0.42	0.61	0.68	0.660	0.701	37.7%	9.6%
4.3.3	0.45	0.64	0.70	0.683	0.715	35.2%	7.9%
4.3.4	0.48	0.67	0.73	0.716	0.742	33.8%	7.6%
4.3.5	0.54	0.73	0.77	0.756	0.785	29.7%	5.0%
4.3.6	0.60	0.79	0.78	0.734	0.831	23.3%	1.1%
4.3.7	0.84	1.03	0.86	0.817	0.901	2.7%	19.5%
4.3.8	0.95	1.14	0.90	0.861	0.935	6.2%	27.5%

Scenarios 7.1.1 through 7.1.8 modified the interarrival time distributions of the first arrival stream in Scenario 7 to increase the arrival variability, but the mean interarrival time remained 6 minutes. Scenarios 7.2.1 through 7.2.8 modified the interarrival time distributions of the second arrival stream to increase the arrival variability, but the mean interarrival time remained 4 minutes.

Table 53. Results for Scenarios 7.1.1 to 7.1.8

Scenario	**Batch formation variability estimates**		**Confidence interval on batch formation variability from simulation results**			**Relative error**	
	X_i^1	X_i^2	c_{bi}^2 **from Simulation**	**Lower bound**	**Upper bound**	X_i^1	X_i^2
7.1.1	0.18	0.23	0.27	0.257	0.275	33.5%	14.5%
7.1.2	0.20	0.25	0.27	0.266	0.280	27.1%	8.6%
7.1.3	0.21	0.26	0.28	0.273	0.297	26.2%	8.5%
7.1.4	0.22	0.27	0.27	0.268	0.282	19.5%	1.2%
7.1.5	0.24	0.29	0.29	0.257	0.325	16.3%	1.0%
7.1.6	0.27	0.32	0.29	0.276	0.308	9.2%	8.0%
7.1.7	0.35	0.40	0.31	0.290	0.332	14.0%	30.2%
7.1.8	0.40	0.45	0.31	0.294	0.330	27.6%	43.7%

Table 54. Results for Scenarios 7.2.1 to 7.2.8

Scenario	**Batch formation variability estimates**		**Confidence interval on batch formation variability from simulation results**			**Relative error**	
	X_i^1	X_i^2	c_{bi}^2 **from Simulation**	**Lower bound**	**Upper bound**	X_i^1	X_i^2
7.2.1	0.19	0.24	0.126	0.118	0.134	49.3%	89.3%
7.2.2	0.20	0.26	0.213	0.204	0.222	4.0%	19.7%
7.2.3	0.21	0.26	0.251	0.242	0.260	15.1%	5.0%
7.2.4	0.22	0.27	0.283	0.271	0.295	21.8%	3.9%
7.2.5	0.24	0.29	0.337	0.328	0.347	29.5%	14.5%
7.2.6	0.25	0.30	0.377	0.354	0.400	32.5%	19.1%
7.2.7	0.32	0.37	0.474	0.466	0.481	32.3%	21.6%
7.2.8	0.35	0.40	0.504	0.502	0.506	29.8%	19.8%

The next step was to look at the impact of varying the arrival rates. To do this, we created Scenarios 4.4.1 through 4.4.5 and Scenarios 4.5.1 through 4.5.5 from the original Scenario 4. The interarrival time distributions of the first and second arrival streams remained as exponential distributions. For Scenarios 4.4.1 through 4.4.5, the interarrival times for the third arrival stream were constant.

For Scenarios 4.5.1 through 4.5.5, the interarrival times for the third arrival stream had a gamma distribution with α equal to 0.5. Thus, the interarrival time SCV equals 2. The mean interarrival times were varied as shown in table 55.

Table 55. Description of Scenarios 4.4.1 to 4.4.5 and Scenarios 4.5.1 to 4.5.5

Scenario	**Interrarrival time means (mins)**		
	Arrival stream 1	**Arrival stream 2**	**Arrival stream 3**
4.4.1 (4.5.1)	6	10	10
4.4.2 (4.5.2)	7	5	10
4.4.3 (4.5.3)	6	7	10
4.4.4 (4.5.4)	15	10	6
4.4.5 (4.5.5)	10	15	5

In addition, we created Scenarios 7.3.1 through 7.3.5 and Scenarios 7.4.1 through 7.4.5 from the original Scenario 7. In Scenarios 7.3.1 to 7.3.5, the interarrival time distributions of the second and third arrival streams remained as exponential distributions, but the interarrival times for the first arrival stream had a gamma distribution with α equal to 2.

In Scenarios 7.4.1 to 7.4.5, the interarrival time distributions of the first and third arrival streams remained as exponential distributions, but the interarrival times for the second arrival stream had a gamma distribution with α equal to 2/3. The mean interarrival times were varied as shown in table 56.

Table 56. Description of Scenarios 7.3.1 to 7.3.5 and Scenarios 7.4.1 to 7.4.5

Scenario	Interrarrival time means (mins)		
	Arrival stream 1	Arrival stream 2	Arrival stream 3
7.3.1 (7.4.1)	3	12	3
7.3.2 (7.4.2)	6	6	12
7.3.3 (7.4.3)	12	10	20/3
7.3.4 (7.4.4)	4	20	15
7.3.5 (7.4.5)	12	4	15

Table 57. Results for Scenarios 4.4.1 to 4.4.5

Scenario	Batch formation variability estimates		Confidence interval on batch formation variability from simulation results			Relative error	
	X_i^1	X_i^2	c_{bi}^2 from Simulation	Lower bound	Upper bound	X_i^1	X_i^2
4.4.1	0.38	0.54	0.63	0.623	0.641	39.141%	13.993%
4.4.2	0.33	0.55	0.57	0.558	0.576	42.506%	2.572%
4.4.3	0.60	0.79	0.62	0.611	0.636	3.732%	26.886%
4.4.4	0.29	0.43	0.31	0.295	0.319	5.564%	40.660%
4.4.5	0.24	0.34	0.35	0.331	0.375	32.410%	2.595%

Table 58. Results for Scenarios 4.5.1 to 4.5.5

Scenario	Batch formation variability estimates		Confidence interval on batch formation variability from simulation results			Relative error	
	X_i^1	X_i^2	c_{bi}^2 from Simulation	Lower bound	Upper bound	X_i^1	X_i^2
4.5.1	0.67	0.83	0.81	0.800	0.829	17.796%	1.707%
4.5.2	0.51	0.73	0.74	0.730	0.758	31.971%	1.526%
4.5.3	0.60	0.79	0.77	0.752	0.796	22.463%	2.201%
4.5.4	0.59	0.73	0.81	0.801	0.819	27.133%	9.625%
4.5.5	0.78	0.88	0.88	0.869	0.884	11.043%	0.958%

Table 59. Results for Scenarios 7.3.1 to 7.3.5

Scenario	Batch formation variability estimates		Confidence interval on batch formation variability from simulation results			Relative error	
	X_i^1	X_i^2	c_{bi}^2 from Simulation	Lower bound	Upper bound	X_i^1	X_i^2
7.3.1	0.20	0.25	0.27	0.249	0.293	26.611%	8.006%
7.3.2	0.22	0.28	0.30	0.283	0.318	25.971%	5.678%
7.3.3	0.23	0.27	0.29	0.278	0.302	19.323%	5.334%
7.3.4	0.14	0.20	0.19	0.184	0.202	25.015%	4.874%
7.3.5	0.28	0.32	0.35	0.338	0.353	19.547%	7.798%

Table 60. Results for Scenarios 7.4.1 to 7.4.5

Scenario	Batch formation variability estimates		Confidence interval on batch formation variability from simulation results			Relative error	
	X_i^1	X_i^2	c_{bi}^2 from Simulation	Lower bound	Upper bound	X_i^1	X_i^2
7.4.1	0.24	0.29	0.26	0.193	0.317	6.749%	13.046%
7.4.2	0.24	0.31	0.41	0.383	0.427	39.675%	24.625%
7.4.3	0.25	0.29	0.34	0.332	0.356	28.100%	16.311%
7.4.4	0.16	0.22	0.25	0.217	0.275	33.411%	9.983%
7.4.5	0.32	0.36	0.49	0.472	0.500	33.833%	25.479%

4.4.1.3 Discussion of initial batch formation experiments

Based on these results, we see that X_i^1, the first estimate for batch formation variability is generally much worse than X_i^2, the second estimate for batch formation variability. The latter estimate is, however, only acceptable when all of the arrival streams have interarrival time distributions with moderate variability, which occurs in Scenarios 1 to 7, Scenarios 4.1.3 to 4.1.5, Scenarios 4.2.3 to 4.2.5, Scenarios 4.3.3 to 4.3.5, Scenarios 7.1.3 to 7.1.5, and Scenarios 7.2.3 to 7.2.5.

Clearly, changes to arrival variability affect the batch formation variability. Moreover, changes in the arrival variability of smaller batches have less impact than changes in the arrival variability of larger batches. For example, the batch formation variability changes much more across Scenarios 4.1.1 to 4.1.8 (which modifies the arrival stream with the largest batch size) than it does across Scenarios 4.2.1 to 4.2.8, which modifies the arrival stream with the smallest batch size.

Similarly, the batch formation variability changes much more across Scenarios 7.2.1 to 7.2.8 (which modifies the arrival stream with the largest batch size) than it does across Scenarios 7.1.1 to 7.1.8, which modifies the arrival stream with the smallest batch size.

However, X_i^2, the second estimate for batch formation variability, does not include information about the batch sizes. Based on these observations, we developed two more estimates that replace the aggregate batch arrival variability term with terms that explicitly incorporate batch size information:

$$X_i^3 = \frac{\overline{K}_{Ai}}{k_i}\left(\frac{\sum_{j \in S_i} \overline{K}_{Bji}\lambda_{Bji}c_{Bji}^2}{\sum_{j \in S_i} \overline{K}_{Bji}\lambda_{Bji}} + \overline{C}_{Ai}^2\right)$$

$$X_i^4 = \frac{\overline{K}_{Ai}}{k_i}\left(\frac{\sum_{j \in S_i} \overline{K}_{Bji}^2\lambda_{Bji}c_{Bji}^2}{\sum_{j \in S_i} \overline{K}_{Bji}^2\lambda_{Bji}} + \overline{C}_{Ai}^2\right)$$

4.4.1.4 Evaluation of additional estimates

To evaluate these two new estimates (X_i^3 , X_i^4), we calculated them for the scenarios discussed in Section 4.4.1.2. Tables 61 to 69 show the results, along with the X_i^2 estimates previously calculated. These results show the fourth estimate X_i^4 is more accurate than the others. It is especially good with the interarrival time variability is moderate (between 0.5 and 1.5).

Table 61. Results for Scenarios 4.1.1 to 4.1.8

Scenario	**Batch formation variability estimates**			**Relative errors**		
	X_i^2	X_i^3	X_i^4	X_i^2	X_i^3	X_i^4
4.1.1	0.48	0.348	0.287	127.6%	65.8%	36.6%
4.1.2	0.58	0.511	0.480	14.2%	2.2%	3.9%
4.1.3	0.62	0.592	0.577	0.7%	6.0%	8.4%
4.1.4	0.67	0.674	0.674	6.7%	6.4%	6.4%
4.1.5	0.77	0.837	0.867	14.4%	7.1%	3.7%
4.1.6	0.87	0.999	1.060	16.0%	3.9%	2.0%
4.1.7	1.26	1.651	1.834	7.4%	21.4%	34.9%
4.1.8	1.46	1.976	2.221	1.5%	33.5%	50.1%

Table 62. Results for Scenarios 4.2.1 to 4.2.8

Scenario	**Batch formation variability estimates**			**Relative errors**		
	X_i^2	X_i^3	X_i^4	X_i^2	X_i^3	X_i^4
4.2.1	0.51	0.639	0.669	29.5%	10.9%	6.7%
4.2.2	0.59	0.656	0.671	16.7%	7.2%	5.1%
4.2.3	0.63	0.665	0.672	13.1%	8.5%	7.5%
4.2.4	0.67	0.674	0.674	6.7%	6.7%	6.7%
4.2.5	0.76	0.691	0.676	2.5%	6.5%	8.5%
4.2.6	0.84	0.709	0.679	14.3%	3.9%	7.9%
4.2.7	1.18	0.778	0.689	56.4%	3.2%	8.6%
4.2.8	1.35	0.813	0.694	81.4%	9.5%	6.5%

Table 63. Results for Scenarios 4.3.1 to 4.3.8

Scenario	Batch formation variability estimates			Relative errors		
	X_i^2	X_i^3	X_i^4	X_i^2	X_i^3	X_i^4
4.3.1	0.56	0.552	0.583	10.9%	11.0%	6.0%
4.3.2	0.61	0.613	0.628	9.6%	9.9%	7.6%
4.3.3	0.64	0.643	0.651	7.9%	8.1%	7.0%
4.3.4	0.67	0.674	0.674	7.6%	7.7%	7.7%
4.3.5	0.73	0.735	0.719	5.0%	4.6%	6.6%
4.3.6	0.79	0.796	0.764	1.1%	2.0%	2.0%
4.3.7	1.03	1.040	0.946	19.5%	20.9%	10.0%
4.3.8	1.14	1.162	1.036	27.5%	29.1%	15.1%

Table 64. Results for Scenarios 7.1.1 to 7.1.8

Scenario	Batch formation variability estimates			Relative errors		
	X_i^2	X_i^3	X_i^4	X_i^2	X_i^3	X_i^4
7.1.1	0.23	0.24	0.26	14.5%	12.9%	4.4%
7.1.2	0.25	0.25	0.26	8.6%	6.1%	1.9%
7.1.3	0.26	0.26	0.27	8.5%	6.2%	4.2%
7.1.4	0.27	0.27	0.27	1.2%	0.6%	0.6%
7.1.5	0.29	0.29	0.28	1.0%	0.1%	4.0%
7.1.6	0.32	0.31	0.29	8.0%	6.2%	1.7%
7.1.7	0.40	0.38	0.31	30.2%	22.8%	0.6%
7.1.8	0.45	0.42	0.33	43.7%	34.5%	4.9%

Table 65. Results for Scenarios 7.2.1 to 7.2.8

Scenario	Batch formation variability estimates			Relative errors		
	X_i^2	X_i^3	X_i^4	X_i^2	X_i^3	X_i^4
7.2.1	0.24	0.17	0.14	89.3%	36.2%	8.5%
7.2.2	0.26	0.22	0.20	19.7%	4.1%	4.2%
7.2.3	0.26	0.25	0.24	5.0%	1.7%	5.2%
7.2.4	0.27	0.27	0.27	3.9%	4.0%	4.0%
7.2.5	0.29	0.32	0.34	14.5%	4.6%	0.6%
7.2.6	0.30	0.37	0.41	19.1%	1.4%	7.8%
7.2.7	0.37	0.57	0.68	21.6%	20.6%	42.7%
7.2.8	0.40	0.67	0.81	19.8%	33.3%	61.0%

Table 66. Results for Scenarios 4.4.1 to 4.4.5.

Scenario	Batch formation variability estimates			Relative errors		
	X_i^2	X_i^3	X_i^4	X_i^2	X_i^3	X_i^4
4.4.1	0.54	0.55	0.586	13.9%	12.9%	7.0%
4.4.2	0.55	0.53	0.553	2.6%	7.4%	3.0%
4.4.3	0.79	0.55	0.585	26.9%	10.6%	5.6%
4.4.4	0.43	0.33	0.366	40.7%	6.7%	18.1%
4.4.5	0.34	0.34	0.394	2.6%	1.9%	12.6%

Table 67. Results for Scenarios 4.5.1 to 4.5.5

Scenario	Batch formation variability estimates			Relative errors		
	X_i^2	X_i^3	X_i^4	X_i^2	X_i^3	X_i^4
4.5.1	0.83	0.82	0.784	1.7%	1.4%	3.2%
4.5.2	0.73	0.75	0.728	1.5%	1.8%	1.6%
4.5.3	0.79	0.80	0.767	2.2%	3.7%	0.4%
4.5.4	0.73	0.83	0.796	9.6%	2.6%	1.8%
4.5.5	0.88	0.89	0.838	0.9%	1.0%	4.8%

Table 68. Results for Scenarios 7.3.1 to 7.3.5

Scenario	Batch formation variability estimates			Relative errors		
	X_i^2	X_i^3	X_i^4	X_i^2	X_i^3	X_i^4
7.3.1	0.25	0.25	0.26	8.0%	8.8%	3.4%
7.3.2	0.28	0.27	0.29	5.7%	8.7%	4.3%
7.3.3	0.27	0.27	0.28	5.4%	7.5%	4.7%
7.3.4	0.20	0.18	0.20	4.9%	2.9%	6.8%
7.3.5	0.32	0.32	0.33	7.8%	8.8%	6.8%

Table 69. Results for Scenarios 7.4.1 to 7.4.5

Scenario	Batch formation variability estimates			Relative errors		
	X_i^2	X_i^3	X_i^4	X_i^2	X_i^3	X_i^4
7.4.1	0.29	0.30	0.31	13%	13.9%	19.3%
7.4.2	0.31	0.37	0.39	24.7%	10.4%	6.0%
7.4.3	0.29	0.34	0.35	16.3%	1.2%	3.7%
7.4.4	0.22	0.24	0.26	9.9%	2.6%	4.2%
7.4.5	0.36	0.44	0.46	25.5%	9.4%	6.7%

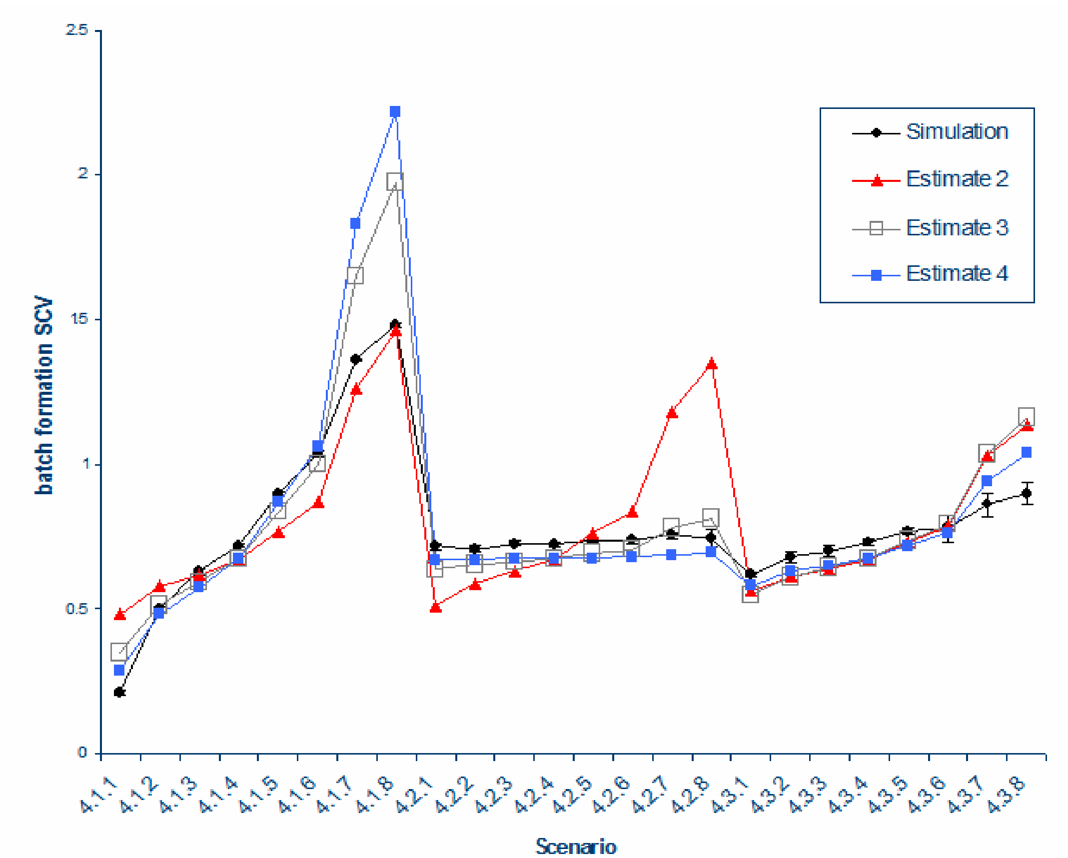

Figure 44. Results for Scenarios 4.1.1 to 4.3.8

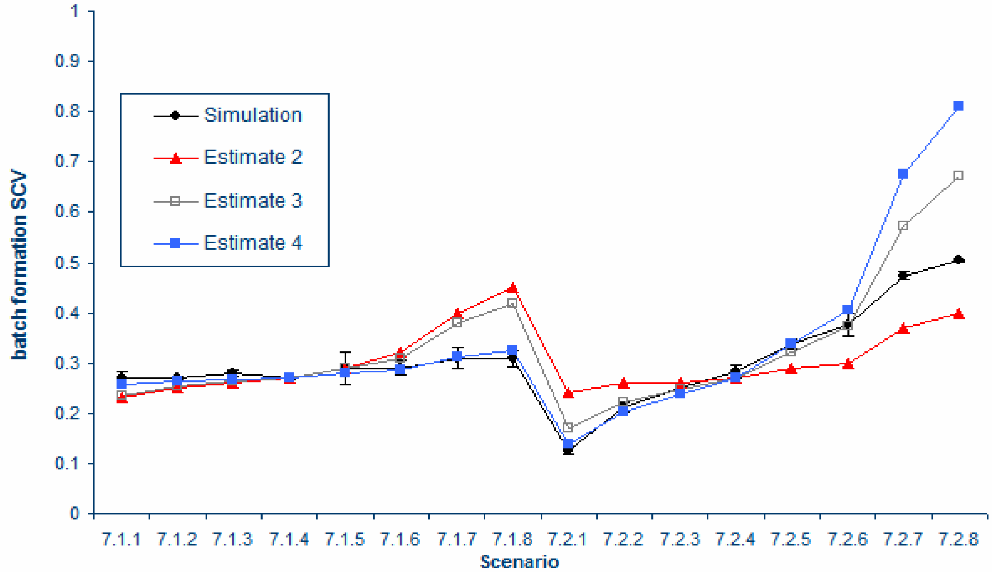

Figure 45. Results for Scenarios 7.1.1 to 7.2.8

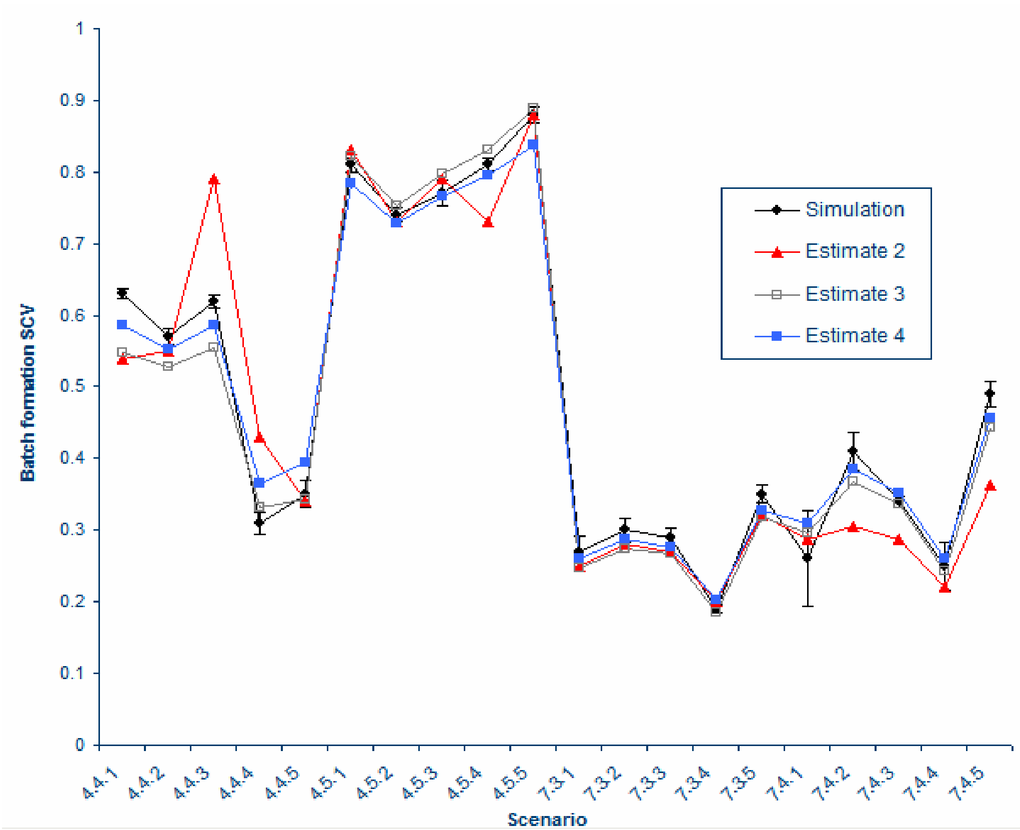

Figure 46. Results for Scenarios 4.4.1 to 4.5.5 and Scenarios 7.3.1 to 7.4.5

4.4.2 Wait time to form a batch (WTBT)

As noted in Chapter 2, when customers arrive at a batch service process, they must first wait while the other customers in the batch arrive, then wait as a batch for the server to become available. Hopp and Spearman (2001) refer to this first delay as wait-to-batch time (WTBT) and define it for a single server station with individual arrivals as: $WTBT = \frac{k-1}{2\lambda}$ (Formula 5 from Chapter 2).

In this formula, k is the number of customers or jobs needed to form a complete process batch and λ is the arrival rate of individuals to the batch process workstation.

In this section, we will study the WTBT for the general cases which can be employed in our model formulation section in the next chapter for the queueing with batch service process and any kind of the arrivals.

If we have multiple arrival streams with an average batch size of $\bar{K}_{Bji}$ and batch arrival rate of λ_{Bji}, the arrival rate of residents (customers) per min (r_i) is calculated by the following formula:

$$r_i = \sum_{j=1}^{i-1} \lambda_{Bji} \bar{K}_{Bji} \quad \textbf{(Formula 23)}$$

Therefore, similarly to $WTBT$ from Formula 5 which is for individual arrivals for a single server station, we replace the r_i instead of the individual arrival rate. In this way, the average time residents spend waiting to form a process batch is the following formula despite of having multiple servers in batch service station:

$$WTBT_i = \frac{k_i - 1}{2r_i} \quad \textbf{(Formula 24)}$$

As we will see from the simulation results in validation section of Chapter5, Formula 24 is an acceptable equation to estimate the WTBT even in general cases. Moreover, it includes Formula 5 for the individual arrival cases.

4.5 Branching process of individuals after batch processing

Modeling the flow of batches is an important question that we must answer before constructing our model completely. There are two different approaches to discuss. They define the batch flow rate and batch size distribution differently.

Before beginning to analyze these approaches, we define some notation needed for our discussion.

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

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

k_i = Processing batch size at station *i*

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)

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

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

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

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

In the first approach, the batch flow rate from station j to station i equals the rate at which batches depart station j. That is, $\lambda_{Bji} = \frac{r_j}{k_j}$.

The batch size distribution is a binomial distribution with k_j trials and a success probability of p_{ji}.

This approach allows batches to be of size 0, however. Thus, although the arrival rate doesn't change due to the branching process, we can have empty batches, which is not realistic.

In the second approach, the batch flow rate is reduced. The idea of the second approach, which is fundamental to the clinic models in Chapter 5, is from Curry et al. (2002).

Since the probability of all k_j individuals in a process batch leaving station j and going to a station other than station i is $(1-p_{ji})^{k_j}$, the probability of having a batch of at least individual moving to station i from station j is $1-(1-p_{ji})^{k_j}$. Therefore, the batch arrival rate to station j from station i is:

$$\lambda_{Bji} = \frac{r_j}{k_j} \times (1-(1-p_{ji})^{k_j}) \quad \textbf{(Formula 25)}$$

For stations with individual service processes, since the k_j is 1, $\lambda_{Bji} = r_j p_{ji}$.

The total batch arrival rate to station i is the aggregation of all batch arrivals from upstream stations which is:

$$\lambda_{Ai} = \sum_{j \in S_i} \lambda_{Bji} \quad \textbf{(Formula 26)}$$

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

In the case that the only upstream station for station i is station j, each batch arriving to station i, has a random batch size of 1 to k_j. Thus, the random batch size distribution should be binomial which is equal or bigger than 1.

If we define B_{ji} as the random batch size of batches that come to station i from station j, the probability distribution for B_{ji} is:

$$P\left(B_{ji} = n\right) = P_n = \binom{k_j}{n} \times \frac{p_{ji}{}^{n}(1-p_{ji})^{k_j - n}}{1-(1-p_{ji})^{k_j}} \quad n \in \{1, 2, \ldots k_j\}$$

It should be said that this probability distribution is a conditional binomial distribution given that the batch size is positive.

Moreover, if we define $\overline{K}_{Bji}$ (see the notation in Section 4.5) as average batch size of batches that come to station i from station j, the mean of this probability distribution is the expected batch size of batches that come from station j to station i:

$$\overline{K}_{Bji} = \frac{k_j P_{ji}}{1-\left(1-P_{ji}\right)^{k_j}} \quad \textbf{(Formula 27)}$$

Similarly, if we calculate the variance of this distribution and divide it by the square of average ($\bar{K}_{Bji}$), we have the SCV of the batch size of batches that come to station i from station j ($\bar{C}^2_{Bji}$):

$$\bar{C}^2_{Bji} = \frac{1 - P_{ji} - \left(k_j P_{ji} + 1 - P_{ji}\right)\left(1 - P_{ji}\right)^{k_j}}{k_j P_{ji}} \quad \textbf{(Formula 28)}$$

This equation for $\bar{C}^2_{Bji}$ is used to calculate $\bar{C}^2_{Ai}$ in Formula 22 in Section 4.4.1.1. Finally, to calculate $\bar{K}_{Ai}$, the average batch size of batches that come to station i from all upstream stations, we make use of the simplest way, which is the average based on the proportion of the total flow rate associated with each batch size. In this way,

$$\bar{K}_{Ai} = \frac{1}{\lambda_{Ai}} \sum_{j \in S_i} \lambda_{Bji} \bar{K}_{Bji} \quad \textbf{(Formula 29)}$$

$\lambda_{Ai} = \sum_{j \in S_i} \lambda_{Bji}$ is from Formula 26. Additionally, Formula 29 is similar to the formula introduced in Section 3.2 for multiple batch arrival streams.

4.6 Waiting time for mixed arrival and individual process service

As we have previously mentioned, for stations with mixed arrivals and individual processing, residents arrive to the workstation in batches and individually. The arrival batches may come from different batch service process workstations, and the batch sizes from each workstation may vary due to the routing probabilities. There are also individual arrivals from individual service process workstations. The workstation has multiple, parallel servers that serve residents individually.

To analyze this case, we model all of the arrivals as batches and divide the waiting times into two parts: *CTq* and WIBT.

CTq is the average time that batches hold in a queue until a server becomes available, and each batch must wait to get to the head of the queue, at which point they are "opened" and individual entities enter the server's queue

Wait-in-batch-time or WIBT is the average time that a resident spends in the batch from the time that the batch "opens" until the resident begins service.

In the Section 4.1 and 4.2, we estimated WIBT for some specific and general cases and concluded that Formulas 21 and 22 are suitable to calculate the WIBT for this type of queueing system.

In this section, we estimate *CTq*, the average time that batches hold in a queue until they get to the head of the queue.

We can estimate *CTq* for $G^{[X]}/G/1$ with Formula 3 from Section 2.2.4, while we don't have any estimate for $G^{[X]}/G/m$.

$$CT_{qi} = \frac{1}{2}\left(c_{ai}^2 + \frac{c_{ei}^2}{\bar{K}_{Ai}}\right)\left(\frac{u_i}{1-u_i}\right)\bar{K}_{Ai}t_i \qquad \textbf{(Formula 3)}$$

Our approach in this section to estimate the time that batches spend in the queue is that we model the $G^{[X]}/G/m$ 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.

In other words, since for a $G^{[X]}/G/1$ system, it takes $\bar{K}_{Ai}t_i$ min to serve a batch because of having a single server, in a $G^{[X]}/G/m$ system, it takes roughly T_i min between opening up a batch and leaving the last entity of the batch from the station.

By estimating T_i , we can make use of Formula 3 to estimate *CTq*, for $G^{[X]}/G/m$ as follows:

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

To estimate the T_i, our methodology is to run simulations with various specifications for some scenarios to compare the results for the *CTq*, and to extract a general formula for T_i.

The purpose of carrying out these experiment types is to find the approximate behavior of T_i versus factors such as the number of servers, utilization, arrival batch size and process time to come up with a general T_i formula that corresponds to all of the cases.

We will again use the notation and relationships from Section 4.1.

4.6.1 Experiments

In this section, we make use of the same 3 types of computational experiments carried out in Section 4.2 to estimate WIBT.

In these simulation experiments, we use a discrete-event simulation model of the station. Each type of experiment consists of a number of sets and each set includes several scenarios.

As a reminder, experiment type one consisted of 8 sets that we named set 1 to set 8. Experiment type two included 2 sets that we named sets 9 and 10. Finally, experiment type three used sets 11 and 12.

For each scenario, we ran a simulation model with 10 replications and a confidence interval of 95%, each 1,000,000 minutes long with the warm-up periods of 500,000 minutes. All the results in this section are in terms of minute.

4.6.1.1 First type of experiment

In the first type of simulation, we have different sets with a constant number of servers, processing time, and utilization. The arrival batch size and batch arrival rate are the variables within each set of scenarios.

Among the 8 sets of simulation of scenarios, the utilization ranges from 25% up to 93%. In each set, the arrival batch size varied from 1 to 13 or 16, the number of servers had a one of the fixed size of 3, 4, 6, 8, 10, 12, and the interarrival times and processing time were exponentially distributed.

Tables 70 to 77 show the *CTq* from the simulation model with its upper and lower bound of 95% of confidence interval for each scenario in Set 1 to 8. The tables also describe other specifications for the scenarios and the name of the scenarios.

Table 70. Observed *CTq* and specifications for Scenarios 1-1-1 to 1-1-13 (Set 1)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	*CTq* From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-1-1	0.50	4	1	25.00%	2	0.014	0.002	0.026
1-1-2	0.25	4	2	25.00%	2	0.045	0.01	0.08
1-1-3	0.16	4	3	25.00%	2	0.082	0.009	0.155
1-1-4	0.12	4	4	25.00%	2	0.138	0.048	0.228
1-1-5	0.1	4	5	25.00%	2	0.208	0.118	0.298
1-1-6	0.084	4	6	25.00%	2	0.273	0.203	0.343
1-1-7	0.072	4	7	25.00%	2	0.355	0.255	0.455
1-1-8	0.0625	4	8	25.00%	2	0.429	0.329	0.529
1-1-9	0.056	4	9	25.00%	2	0.505	0.415	0.595
1-1-10	0.05	4	10	25.00%	2	0.592	0.502	0.682
1-1-11	0.046	4	11	25.00%	2	0.679	0.619	0.739
1-1-12	0.042	4	12	25.00%	2	0.756	0.686	0.826
1-1-13	0.039	4	13	25.00%	2	0.014	0.002	0.026

Table 71. Observed *CTq* and specifications for Scenarios 1-2-1 to 1-2-13 (Set 2)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	*CTq* From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-2-1	0.50	6	1	50.00%	6	0.204	0.192	0.216
1-2-2	0.25	6	2	50.00%	6	0.466	0.376	0.556
1-2-3	0.16	6	3	50.00%	6	0.735	0.645	0.825
1-2-4	0.12	6	4	50.00%	6	1.062	0.952	1.172
1-2-5	0.1	6	5	50.00%	6	1.38	1.28	1.48
1-2-6	0.084	6	6	50.00%	6	1.844	1.754	1.934
1-2-7	0.072	6	7	50.00%	6	2.244	2.024	2.464
1-2-8	0.0625	6	8	50.00%	6	2.631	2.201	3.061
1-2-9	0.056	6	9	50.00%	6	3.014	2.894	3.134
1-2-10	0.05	6	10	50.00%	6	3.577	3.457	3.697
1-2-11	0.046	6	11	50.00%	6	3.967	3.567	4.367
1-2-12	0.042	6	12	50.00%	6	4.6	4.4	4.8
1-2-13	0.039	6	13	50.00%	6	4.875	4.775	4.975

Table 72. Observed *CTq* and specifications for Scenarios 1-3-1 to 1-3-13 (Set 3)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	*CTq* From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-3-1	0.50	3	1	58.33%	3.5	0.937	0.925	0.949
1-3-2	0.25	3	2	58.33%	3.5	1.574	1.354	1.794
1-3-3	0.17	3	3	58.33%	3.5	2.337	2.107	2.567
1-3-4	0.13	3	4	58.33%	3.5	2.941	2.621	3.261
1-3-5	0.10	3	5	58.33%	3.5	3.792	3.392	4.192
1-3-6	0.08	3	6	58.33%	3.5	4.393	4.173	4.613
1-3-7	0.07	3	7	58.33%	3.5	5.098	4.698	5.498
1-3-8	0.06	3	8	58.33%	3.5	6.183	5.983	6.383
1-3-9	0.06	3	9	58.33%	3.5	6.871	6.751	6.991
1-3-10	0.05	3	10	58.33%	3.5	7.677	7.577	7.777
1-3-11	0.05	3	11	58.33%	3.5	8.001	7.911	8.091
1-3-12	0.04	3	12	58.33%	3.5	9.231	9.131	9.331
1-3-13	0.04	3	13	58.33%	3.5	9.643	9.433	9.853

Table 73. Observed *CTq* and specifications for Scenarios 1-4-1 to 1-4-16 (Set 4)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	*CTq* From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-4-1	0.75	12	1	62.50%	10	0.212	0.2	0.224
1-4-2	0.38	12	2	62.50%	10	0.507	0.287	0.727
1-4-3	0.25	12	3	62.50%	10	0.85	0.62	1.08
1-4-4	0.19	12	4	62.50%	10	1.269	0.949	1.589
1-4-5	0.15	12	5	62.50%	10	1.67	1.27	2.07
1-4-6	0.13	12	6	62.50%	10	2.115	1.895	2.335
1-4-7	0.11	12	7	62.50%	10	2.66	2.26	3.06
1-4-8	0.09	12	8	62.50%	10	3.169	2.969	3.369
1-4-9	0.08	12	9	62.50%	10	3.695	3.575	3.815
1-4-10	0.08	12	10	62.50%	10	4.298	4.198	4.398
1-4-11	0.07	12	11	62.50%	10	4.748	4.658	4.838
1-4-12	0.06	12	12	62.50%	10	5.327	5.227	5.427
1-4-13	0.06	12	13	62.50%	10	5.938	5.838	6.038
1-4-14	0.05	12	14	62.50%	10	6.431	6.341	6.521
1-4-15	0.05	12	15	62.50%	10	7.152	7.072	7.232
1-4-16	0.05	12	16	62.50%	10	7.981	7.882	8.08

Table 74. Observed *CTq* and specifications for Scenarios 1-5-1 to 1-5-13 (Set 5)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	*CTq* From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-5-1	0.50	4	1	68.75%	5.5	1.785	1.773	1.797
1-5-2	0.25	4	2	68.75%	5.5	3.1	3.065	3.135
1-5-3	0.17	4	3	68.75%	5.5	4.248	4.175	4.321
1-5-4	0.13	4	4	68.75%	5.5	5.684	5.594	5.774
1-5-5	0.10	4	5	68.75%	5.5	6.872	6.782	6.962
1-5-6	0.08	4	6	68.75%	5.5	7.97	7.9	8.04
1-5-7	0.07	4	7	68.75%	5.5	9.265	9.165	9.365
1-5-8	0.06	4	8	68.75%	5.5	11.934	11.834	12.034
1-5-9	0.06	4	9	68.75%	5.5	13.29	13.2	13.38
1-5-10	0.05	4	10	68.75%	5.5	13.926	13.836	14.016
1-5-11	0.05	4	11	68.75%	5.5	15.81	15.75	15.87
1-5-12	0.04	4	12	68.75%	5.5	17.107	17.037	17.177
1-5-13	0.04	4	13	68.75%	5.5	17.922	17.822	18.022

Table 75. Observed *CTq* and specifications for Scenarios 1-6-1 to 1-6-13 (Set 6)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	*CTq* From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-6-1	0.75	10	1	75.00%	10	1.245	1.233	1.257
1-6-2	0.38	10	2	75.00%	10	2.216	2.181	2.251
1-6-3	0.25	10	3	75.00%	10	3.26	3.187	3.333
1-6-4	0.19	10	4	75.00%	10	4.509	4.419	4.599
1-6-5	0.15	10	5	75.00%	10	5.804	5.714	5.894
1-6-6	0.13	10	6	75.00%	10	6.632	6.562	6.702
1-6-7	0.11	10	7	75.00%	10	8.132	8.032	8.232
1-6-8	0.09	10	8	75.00%	10	9.105	9.005	9.205
1-6-9	0.08	10	9	75.00%	10	10.791	10.701	10.881
1-6-10	0.08	10	10	75.00%	10	11.95	11.86	12.04
1-6-11	0.07	10	11	75.00%	10	13.327	13.267	13.387
1-6-12	0.06	10	12	75.00%	10	14.853	14.783	14.923
1-6-13	0.06	10	13	75.00%	10	15.936	15.836	16.036

Table 76. Observed *CTq* and specifications for Scenarios 1-7-1 to 1-7-13 (Set 7)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	*CTq* From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-7-1	0.50	4	1	81.25%	6.5	5.451	5.439	5.463
1-7-2	0.25	4	2	81.25%	6.5	8.372	8.292	8.452
1-7-3	0.17	4	3	81.25%	6.5	11.83	11.73	11.93
1-7-4	0.13	4	4	81.25%	6.5	15.119	14.819	15.419
1-7-5	0.10	4	5	81.25%	6.5	17.377	17.077	17.677
1-7-6	0.08	4	6	81.25%	6.5	22.972	22.472	23.472
1-7-7	0.07	4	7	81.25%	6.5	25.271	25.071	25.471
1-7-8	0.06	4	8	81.25%	6.5	26.523	25.723	27.323
1-7-9	0.06	4	9	81.25%	6.5	28.651	27.751	29.551
1-7-10	0.05	4	10	81.25%	6.5	34.115	33.135	35.095
1-7-11	0.05	4	11	81.25%	6.5	36.037	35.137	36.937
1-7-12	0.04	4	12	81.25%	6.5	43.92	43.02	44.82
1-7-13	0.04	4	13	81.25%	6.5	47.043	45.843	48.243

Table 77. Observed *CTq* and specifications for Scenarios 1-8-1 to 1-8-13 (Set 8)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	*CTq* From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
1-8-1	0.75	8	1	93.75%	10	17.228	17.216	17.24
1-8-2	0.38	8	2	93.75%	10	23.076	22.996	23.156
1-8-3	0.25	8	3	93.75%	10	31.693	31.393	31.993
1-8-4	0.19	8	4	93.75%	10	42.508	42.008	43.008
1-8-5	0.15	8	5	93.75%	10	50.418	49.718	51.118
1-8-6	0.13	8	6	93.75%	10	61.89	61.09	62.69
1-8-7	0.11	8	7	93.75%	10	69.736	69.136	70.336
1-8-8	0.09	8	8	93.75%	10	81.86	81.16	82.56
1-8-9	0.08	8	9	93.75%	10	95.102	94.302	95.902
1-8-10	0.08	8	10	93.75%	10	92.161	91.181	93.141
1-8-11	0.07	8	11	93.75%	10	107.442	106.242	108.642
1-8-12	0.06	8	12	93.75%	10	121.138	120.238	122.038
1-8-13	0.06	8	13	93.75%	10	124.898	123.698	126.098

4.6.1.2 Second type of experiment

In the second type of simulation, we have 2 different sets with constant arrival batch sizes, processing time and utilization. On the other hand, our variable here is the number of servers, which changes in each scenario.

In this experiment, the utilization is either 60% or 80%. In each set, the number of servers varied from 1 to 13, the arrival batch was 4 or 6, and the interarrival times and processing time were exponentially distributed.

Table 78 and 79 show the *CTq* from the simulation model with its upper and lower bound of 95% of confidence interval for each scenario. The tables also describe other simulations' specifications of the scenarios and the name of the scenarios.

Table 78. Observed *CTq* and specifications for Scenarios 2-1-1 to 2-1-13 (Set 9)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	*CTq* From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
2-1-1	0.10	1	6	60.00%	1	5.196	5.106	5.286
2-1-2	0.20	2	6	60.00%	1	2.398	2.32	2.476
2-1-3	0.30	3	6	60.00%	1	1.383	1.293	1.473
2-1-4	0.40	4	6	60.00%	1	0.947	0.903	0.991
2-1-5	0.50	5	6	60.00%	1	0.701	0.658	0.744
2-1-6	0.60	6	6	60.00%	1	0.531	0.491	0.571
2-1-7	0.70	7	6	60.00%	1	0.418	0.398	0.438
2-1-8	0.80	8	6	60.00%	1	0.352	0.342	0.362
2-1-9	0.90	9	6	60.00%	1	0.293	0.263	0.323
2-1-10	1.00	10	6	60.00%	1	0.247	0.217	0.277
2-1-11	1.10	11	6	60.00%	1	0.207	0.167	0.247
2-1-12	1.20	12	6	60.00%	1	0.179	0.149	0.209
2-1-13	1.30	13	6	60.00%	1	0.154	0.114	0.194

Table 79. Observed *CTq* and specifications for Scenarios 2-2-1 to 2-2-13 (Set 10)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	*CTq* From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
2-2-1	0.20	1	4	80.00%	1	10.216	10.006	10.426
2-2-2	0.40	2	4	80.00%	1	4.899	4.769	5.029
2-2-3	0.60	3	4	80.00%	1	2.952	2.862	3.042
2-2-4	0.80	4	4	80.00%	1	2.085	1.985	2.185
2-2-5	1.00	5	4	80.00%	1	1.569	1.489	1.649
2-2-6	1.20	6	4	80.00%	1	1.316	1.226	1.406
2-2-7	1.40	7	4	80.00%	1	1.1	1.08	1.12
2-2-8	1.60	8	4	80.00%	1	0.857	0.847	0.867
2-2-9	1.80	9	4	80.00%	1	0.73	0.7	0.76
2-2-10	2.00	10	4	80.00%	1	0.684	0.654	0.714
2-2-11	2.20	11	4	80.00%	1	0.638	0.598	0.678
2-2-12	2.40	12	4	80.00%	1	0.5	0.47	0.53
2-2-13	2.60	13	4	80.00%	1	0.53	0.49	0.57

4.6.1.3 Third type of experiment

In the third type of simulation, we have 2 different sets with constant arrival batch sizes and processing time. On the other hand, our variable here is the number of servers and utilization, which changes in each scenario. Since the batch interarrival time doesn't vary as we had in experiment type 2, the utilization is a variable in addition to the number of the servers.

In this experiment, the batch arrival rate of either 0.2 or 0.15 (batch/min). In each set, the arrival batches size varied from 1 to 13, the arrival batch size was 4 or 6 and the interarrival times and processing time were exponentially distributed.

Table 80 and 81 show the *CTq* from the simulation model with its upper and lower bound of 95% of confidence interval for each scenario.

Table 80. Observed *CTq* and specifications for Scenarios 3-1-1 to 3-1-13 (Set 11)

Scenario name	**Batch arrival rate per min**	**Number of servers**	**Arrival batch size**	**Utilization**	**Mean Process time(min)**	***CTq* From simulation**	**Lower bound for 95% confidence interval**	**Upper bound for 95% confidence interval**
3-1-1	0.20	1	4	96.00%	1.2	86.281	85.941	86.621
3-1-2	0.20	2	4	48.00%	1.2	1.143	1.053	1.233
3-1-3	0.20	3	4	32.00%	1.2	0.267	0.257	0.277
3-1-4	0.20	4	4	24.00%	1.2	0.078	0.077	0.079
3-1-5	0.20	5	4	19.20%	1.2	0.026	0.025	0.027
3-1-6	0.20	6	4	16.00%	1.2	0.01	0.009	0.011
3-1-7	0.20	7	4	13.71%	1.2	0.004	0.0036	0.0044
3-1-8	0.20	8	4	12.00%	1.2	0.001	0.0008	0.0013
3-1-9	0.20	9	4	10.67%	1.2	0	0	0
3-1-10	0.20	10	4	9.60%	1.2	0	0	0
3-1-11	0.20	11	4	8.73%	1.2	0	0	0
3-1-12	0.20	12	4	8.00%	1.2	0	0	0
3-1-13	0.20	13	4	7.38%	1.2	0	0	0

Table 81. Observed *CTq* and specifications for Scenarios 3-2-1 to 3-2-13 (Set 12)

Scenario name	Batch arrival rate per min	Number of servers	Arrival batch size	Utilization	Mean Process time(min)	*CTq* From simulation	Lower bound for 95% confidence interval	Upper bound for 95% confidence interval
3-2-1	0.15	1	6	90.00%	1	31.637	31.2970	31.9770
3-2-2	0.15	2	6	45.00%	1	1.223	1.1330	1.3130
3-2-3	0.15	3	6	30.00%	1	0.32	0.3100	0.3300
3-2-4	0.15	4	6	22.50%	1	0.115	0.1140	0.1160
3-2-5	0.15	5	6	18.00%	1	0.047	0.0460	0.0480
3-2-6	0.15	6	6	15.00%	1	0.019	0.0180	0.0200
3-2-7	0.15	7	6	12.86%	1	0.008	0.0076	0.0084
3-2-8	0.15	8	6	11.25%	1	0.004	0.0037	0.0043
3-2-9	0.15	9	6	10.00%	1	0.002	0.0017	0.0023
3-2-10	0.15	10	6	9.00%	1	0.001	0.0007	0.0013
3-2-11	0.15	11	6	8.18%	1	0	0	0
3-2-12	0.15	12	6	7.50%	1	0	0	0
3-2-13	0.15	13	6	6.92%	1	0	0	0

4.6.2 Summary and results

In the section, we find some estimates for T_i and compare the numerical results from our best estimate with the simulation results brought in Tables 70 to 81.

The good initial guess for T_i is that for the high utilization when the servers are always busy, from the perspective of the servers, we can replace the batch arrivals with individual arrivals with the arrival rate of $\overline{K}_{Ai}$ times the batch arrival rate.

In this way, our $G^{[X]}/G/m$ system is changed to a $G/G/m$. Since we our assumption is to have a imaginary server with the serving time of T_i, this T_i for high utilization cases can be $\frac{\overline{K}_{Ai} t_i}{m_i}$.

Therefore, we propose that we must have a factor of utilization in T_i to be eliminated when it is close to 1.

In this way, T_i should be something like $\frac{\overline{K}_{Ai}t_i}{m_i}u_i^{\,C}$, which C can be itself a function of different factors such as utilization, number of servers or batch arrival size.

The point is that when the m_i =1, T_i should be $\overline{K}_{Ai}t_i$ (Formula 3). We found out so many possible guesses for C, but among them one of them is better than the others and it corresponds to all the cases in our clinics.

In this thesis, we assume the best possible C is $1-\frac{1}{m_i}$ and our chosen estimate for the T_i will be $\frac{\overline{K}_{Ai}t_i}{m_i}u_i^{(1-1/m_i)}$.

Tables 82 to 93 show the comparison between the simulation results for CTq from Tables 70 to 81 and our best estimate for CTq in which T_i will be $\frac{\overline{K}_{Ai}t_i}{m_i}u_i^{(1-1/m_i)}$.

Table 82. Simulations results and *CTq* estimate for Scenarios 1-1-1 to 1-1-13 (Set 1)

Scenario #	***CTq* from Simulation**	***CTq* for** $T_i = \frac{\bar{K}_{Ai} t_i}{m_i} u_i^{(1-1/m_i)}$	**Absolute error**	**Percentage error%**
1-1-1	0.014	0.059	0.045	320.90%
1-2-1	0.045	0.088	0.043	96.42%
1-3-1	0.082	0.118	0.036	43.72%
1-4-1	0.138	0.147	0.009	6.75%
1-5-1	0.208	0.177	0.031	15.01%
1-6-1	0.273	0.206	0.067	24.45%
1-7-1	0.355	0.236	0.119	33.60%
1-8-1	0.429	0.265	0.164	38.19%
1-9-1	0.505	0.295	0.210	41.66%
1-10-1	0.592	0.324	0.268	45.25%
1-11-1	0.679	0.354	0.325	47.93%
1-12-1	0.756	0.383	0.373	49.34%
1-13-1	0.837	0.412	0.425	50.72%

Table 83. Simulations results and *CTq* estimate for Scenarios 1-2-1 to 1-2-13 (Set 2)

Scenario #	***CTq* from Simulation**	***CTq* for** $T_i = \frac{\bar{K}_{Ai} t_i}{m_i} u_i^{(1-1/m_i)}$	**Absolute error**	**Percentage error%**
1-2-1	0.204	0.561	0.357	175.11%
1-2-2	0.466	0.842	0.376	80.65%
1-2-3	0.735	1.122	0.387	52.72%
1-2-4	1.062	1.403	0.341	32.12%
1-2-5	1.380	1.684	0.304	22.01%
1-2-6	1.844	1.964	0.120	6.52%
1-2-7	2.244	2.245	0.001	0.04%
1-2-8	2.631	2.526	0.105	4.01%
1-2-9	3.014	2.806	0.208	6.90%
1-2-10	3.577	3.087	0.490	13.71%
1-2-11	3.967	3.367	0.600	15.12%
1-2-12	4.600	3.648	0.952	20.70%
1-2-13	4.875	3.929	0.946	19.41%

Table 84. Simulations results and *CTq* estimate for Scenarios 1-3-1 to 1-3-13 (Set 3)

Scenario #	***CTq* from Simulation**	***CTq* for** $T_i = \frac{\bar{K}_{Ai} t_i}{m_i} u_i^{(1-1/m_i)}$	**Absolute error**	**Percentage error%**
1-3-1	0.937	1.140	0.203	21.70%
1-3-2	1.574	1.710	0.136	8.67%
1-3-3	2.337	2.281	0.056	2.41%
1-3-4	2.941	2.851	0.090	3.07%
1-3-5	3.792	3.421	0.371	9.79%
1-3-6	4.393	3.991	0.402	9.15%
1-3-7	5.098	4.561	0.537	10.53%
1-3-8	6.183	5.131	1.052	17.01%
1-3-9	6.871	5.702	1.169	17.02%
1-3-10	7.677	6.272	1.405	18.31%
1-3-11	8.001	6.842	1.159	14.49%
1-3-12	9.231	7.412	1.819	19.71%
1-3-13	9.643	7.982	1.661	17.22%

Table 85. Simulations results and *CTq* estimate for Scenarios 1-4-1 to 1-4-16 (Set 4)

Scenario #	***CTq* from Simulation**	***CTq* for** $T_i = \frac{\bar{K}_{Ai} t_i}{m_i} u_i^{(1-1/m_i)}$	**Absolute error**	**Percentage error%**
1-4-1	0.212	0.903	0.691	325.82%
1-4-2	0.507	1.354	0.847	167.08%
1-4-3	0.850	1.805	0.955	112.41%
1-4-4	1.269	2.257	0.988	77.84%
1-4-5	1.670	2.708	1.038	62.17%
1-4-6	2.115	3.160	1.045	49.39%
1-4-7	2.660	3.611	0.951	35.75%
1-4-8	3.169	4.062	0.893	28.19%
1-4-9	3.695	4.514	0.819	22.16%
1-4-10	4.298	4.965	0.667	15.52%
1-4-11	4.748	5.416	0.668	14.08%
1-4-12	5.327	5.868	0.541	10.15%
1-4-13	5.938	6.319	0.381	6.42%
1-4-14	6.431	6.770	0.339	5.28%
1-4-15	7.152	7.222	0.070	0.98%
1-4-16	7.981	7.673	0.308	3.86%

Table 86. Simulations results and CTq estimate for Scenarios 1-5-1 to 1-5-13 (Set 5)

Scenario #	***CTq* from Simulation**	***CTq* for** $T_i = \frac{\bar{K}_{Ai} t_i}{m_i} u_i^{(1-1/m_i)}$	**Absolute error**	**Percentage error%**
1-5-1	1.785	2.284	0.499	27.95%
1-5-2	3.100	3.426	0.326	10.51%
1-5-3	4.248	4.568	0.320	7.53%
1-5-4	5.684	5.710	0.026	0.45%
1-5-5	6.872	6.852	0.020	0.29%
1-5-6	7.970	7.994	0.024	0.30%
1-5-7	9.265	9.136	0.129	1.40%
1-5-8	11.934	10.278	1.656	13.88%
1-5-9	13.290	11.420	1.870	14.07%
1-5-10	13.926	12.562	1.364	9.80%
1-5-11	15.810	13.703	2.107	13.32%
1-5-12	17.107	14.845	2.262	13.22%
1-5-13	17.922	15.987	1.935	10.79%

Table 87. Simulations results and CTq estimate for Scenarios 1-6-1 to 1-6-13 (Set 6)

Scenario #	***CTq* from Simulation**	***CTq* for** $T_i = \frac{\bar{K}_{Ai} t_i}{m_i} u_i^{(1-1/m_i)}$	**Absolute error**	**Percentage error%**
1-6-1	1.245	2.316	1.071	86.00%
1-6-2	2.216	3.474	1.258	56.75%
1-6-3	3.260	4.631	1.371	42.07%
1-6-4	4.509	5.789	1.280	28.39%
1-6-5	5.804	6.947	1.143	19.69%
1-6-6	6.632	8.105	1.473	22.21%
1-6-7	8.132	9.263	1.131	13.90%
1-6-8	9.105	10.421	1.316	14.45%
1-6-9	10.791	11.578	0.787	7.30%
1-6-10	11.950	12.736	0.786	6.58%
1-6-11	13.327	13.894	0.567	4.25%
1-6-12	14.853	15.052	0.199	1.34%
1-6-13	15.936	16.210	0.274	1.72%

Table 88. Simulations results and *CTq* estimate for Scenarios 1-7-1 to 1-7-13 (Set 7)

Scenario #	***CTq*** **from Simulation**	***CTq*** **for** $T_i = \frac{\bar{K}_{Ai} t_i}{m_i} u_i^{(1-1/m_i)}$	**Absolute error**	**Percentage error%**
1-7-1	5.451	6.026	0.575	10.55%
1-7-2	8.372	9.039	0.667	7.97%
1-7-3	11.830	12.052	0.222	1.88%
1-7-4	15.119	15.065	0.054	0.35%
1-7-5	17.377	18.079	0.702	4.04%
1-7-6	22.972	21.092	1.880	8.19%
1-7-7	25.271	24.105	1.166	4.61%
1-7-8	26.523	27.118	0.595	2.24%
1-7-9	28.651	30.131	1.480	5.17%
1-7-10	34.115	33.144	0.971	2.85%
1-7-11	36.037	36.157	0.120	0.33%
1-7-12	43.920	39.170	4.750	10.81%
1-7-13	47.043	42.183	4.860	10.33%

Table 89. Simulations results and *CTq* estimate for Scenarios 1-8-1 to 1-8-13 (Set 8)

Scenario #	***CTq*** **from Simulation**	***CTq*** **for** $T_i = \frac{\bar{K}_{Ai} t_i}{m_i} u_i^{(1-1/m_i)}$	**Absolute error**	**Percentage error%**
1-8-1	17.228	17.721	0.493	2.86%
1-8-2	23.076	26.581	3.505	15.19%
1-8-3	31.693	35.441	3.748	11.83%
1-8-4	42.508	44.301	1.793	4.22%
1-8-5	50.418	53.162	2.744	5.44%
1-8-6	61.890	62.022	0.132	0.21%
1-8-7	69.736	70.882	1.146	1.64%
1-8-8	81.860	79.742	2.118	2.59%
1-8-9	95.102	88.603	6.499	6.83%
1-8-10	92.161	97.463	5.302	5.75%
1-8-11	107.442	106.323	1.119	1.04%
1-8-12	121.138	115.183	5.955	4.92%
1-8-13	124.898	124.044	0.854	0.68%

Table 90. Simulations results and *CTq* estimate for Scenarios 2-1-1 to 2-1-13 (Set 9)

Scenario #	***CTq* from Simulation**	***CTq* for** $T_i = \frac{\bar{K}_{Ai} t_i}{m_i} u_i^{(1-1/m_i)}$	**Absolute error**	**Percentage error%**
2-1-1	5.196	5.250	0.054	1.04%
2-1-2	2.398	2.033	0.365	15.21%
2-1-3	1.383	1.245	0.138	9.98%
2-1-4	0.947	0.895	0.052	5.52%
2-1-5	0.701	0.698	0.003	0.46%
2-1-6	0.531	0.572	0.041	7.66%
2-1-7	0.418	0.484	0.066	15.81%
2-1-8	0.352	0.420	0.068	19.24%
2-1-9	0.293	0.370	0.077	26.43%
2-1-10	0.247	0.332	0.085	34.21%
2-1-11	0.207	0.300	0.093	44.92%
2-1-12	0.179	0.274	0.095	53.03%
2-1-13	0.154	0.252	0.098	63.65%

Table 91. Simulations results and *CTq* estimate for Scenarios 2-2-1 to 2-2-13 (Set 10)

Scenario #	***CTq* from Simulation**	***CTq* for** $T_i = \frac{\bar{K}_{Ai} t_i}{m_i} u_i^{(1-1/m_i)}$	**Absolute error**	**Percentage error%**
2-2-1	10.216	10.000	0.216	2.11%
2-2-2	4.899	4.472	0.427	8.71%
2-2-3	2.952	2.873	0.079	2.69%
2-2-4	2.085	2.115	0.030	1.43%
2-2-5	1.569	1.673	0.104	6.63%
2-2-6	1.316	1.384	0.068	5.16%
2-2-7	1.100	1.180	0.080	7.26%
2-2-8	0.857	1.028	0.171	19.99%
2-2-9	0.730	0.911	0.181	24.82%
2-2-10	0.684	0.818	0.134	19.60%
2-2-11	0.638	0.742	0.104	16.33%
2-2-12	0.500	0.679	0.179	35.84%
2-2-13	0.530	0.626	0.096	18.12%

Table 92. Simulations results and *CTq* estimate for Scenarios 3-1-1 to 3-1-13 (Set 11)

Scenario #	***CTq* from Simulation**	***CTq* for** $T_i = \frac{\bar{K}_{Ai} t_i}{m_i} u_i^{(1-1/m_i)}$	**Absolute error**	**Percentage error%**
3-1-1	86.281	72.000	14.281	16.55%
3-1-2	1.143	0.959	0.184	16.07%
3-1-3	0.267	0.220	0.047	17.54%
3-1-4	0.078	0.081	0.003	4.12%
3-1-5	0.026	0.038	0.012	46.46%
3-1-6	0.010	0.021	0.011	106.81%
3-1-7	0.004	0.012	0.008	210.19%
3-1-8	0.001	0.008	0.007	699.86%
3-1-9	0.000	0.005	0.005	#DIV/0!
3-1-10	0.000	0.004	0.004	#DIV/0!
3-1-11	0.000	0.003	0.003	#DIV/0!
3-1-12	0.000	0.002	0.002	#DIV/0!
3-1-13	0.000	0.002	0.002	#DIV/0!

Table 93. Simulations results and *CTq* estimate for Scenarios 3-2-1 to 3-2-13 (Set 12)

Scenario #	***CTq* from Simulation**	***CTq* for** $T_i = \frac{\bar{K}_{Ai} t_i}{m_i} u_i^{(1-1/m_i)}$	**Absolute error**	**Percentage error%**
3-2-1	31.637	31.500	0.137	0.43%
3-2-2	1.223	0.960	0.263	21.46%
3-2-3	0.320	0.224	0.096	29.98%
3-2-4	0.115	0.083	0.032	27.83%
3-2-5	0.047	0.039	0.008	17.08%
3-2-6	0.019	0.021	0.002	11.49%
3-2-7	0.008	0.013	0.005	58.93%
3-2-8	0.004	0.008	0.004	104.96%
3-2-9	0.002	0.006	0.004	179.04%
3-2-10	0.001	0.004	0.003	296.36%
3-2-11	0.000	0.003	0.003	#DIV/0!
3-2-12	0.000	0.002	0.002	#DIV/0!
3-2-13	0.000	0.002	0.002	#DIV/0!

We note that, for Sets 1 to 12, in general the results are good. We have a large percentage error only when the absolute error is small. Therefore, according to these results, we make use of the obtained estimate for T_i which is $\frac{\bar{K}_{Ai} t_i}{m_i} u_i^{(1-1/m_i)}$ in this thesis to calculate the *CTq* for the stations with mixed arrival and individual process service.

4.7 Summary of the chapter

Because this chapter contains many tables and figures, it will be useful to summarize the findings and results in this chapter as follows:

1. Batch arrivals to the individual process stations with multiple numbers of servers.
2. Self service stations
3. Batch processing stations

4.7.1 Batch arrivals with individual process stations with multiple servers.

Since we divided the total waiting time in this type of station into two sections (*CTq* and WIBT), we summarize the results for each separately.

Recall that *CTq* is the average time that batches wait in queue before opening.

4.7.1.1 Summary for WIBT

- Tables 6 to 8 in Section 4.1.1 showed that the Formula 10 (the first formula for WIBT) would be a good estimate.
- Based on Figures 16 to 32 in Section 4.2.3.1, we conclude that Formula 13 is a good estimate for WIBT for the first type of experiment. In this experiment, we have different sets with a constant number of servers, processing time, and utilization. The arrival batch size and batch arrival rate are the variables within each set of scenarios and the purpose of this experiment is to find a relationship between WIBT and the arrival batch size.

- From Tables 21 to 22 in Section 4.2.3.2, we conclude that Formula 17 is a good estimate for WIBT for the second type of experiment. In this experiment, we have 2 different sets with constant arrival batch sizes, processing time and utilization. On the other hand, our variable here is the number of servers, which changes in each scenario. The purpose of this experiment is to find a relationship between the WIBT and the number of servers.

- From Tables 25 to 26 in Section 4.2.3.3, we conclude that Formula 17 is also a good estimate for WIBT for the third type of the experiment in which we have 2 different sets with constant arrival batch sizes, processing time. Our variable here is the number of servers and utilization, which changes in each scenario. Since the batch interarrival time doesn't vary as we had in experiment type 2, the utilization is a variable in addition to the number of the servers.

- Tables 30 to 41 in Section 4.2.3.4 showed that Formulas 21 and 4 were respectively good estimates for WIBT for all cases that had multiple and single number of servers in a batch arrival-individual process station.

4.7.1.2 Summary for *CTq*

From Tables 82 to 93 in Section 4.6, we found an estimate for T_i the average processing time of one fast server, by combing the multiple numbers of servers in a batch arrival-individual processing station. These tables also showed that utilizing this estimate in Formula 3 (instead of $\overline{K}_{Ai}t_i$) was a good approximation for *CTq*.

4.7.2 Self service station

From Tables 42 to 46 in Section 4.3, we extracted some estimates for the interdeparture time SCV for the self service station. These tables also demonstrated that the estimate having ω_a as the weight factor was a better approximation than the others.

4.7.3 Batch processing stations

From previous sections, for this type of the station, the arrival entities (jobs or items) should wait until they form a batch with the same size as the batch processing size.

After the batch formation, we have only the batches that have to wait in queue to get to the head of the line to be served in batches. To calculate the *CTq* for each formed batch, we must first estimate the batch formation variability.

Finally, after the batch processing, each served batch has to be broken into the smaller batches based on the routing probabilities, resident departure rates, and the size of the batches. Therefore, we need to follow an approach to study this process.

According to above, we summarize the findings and results for batch processing station into three sections.

4.7.3.1 Summary for WTBT (wait-to-batch-time)

In Section 4.4.2, we introduced Formula 24 which was the best estimate for WTBT for the cases with multiple batch arrivals from different streams.

4.7.3.2 Summary for batch formation variability

- Tables 50 to 60 in Section 4.1.1.1 compared the first two found estimates (X_i^1 and X_i^2) for the batch formation variability (SCV) at batch processing stations.
- Tables 61 to 69 in Section 4.4.1.4 showed that the fourth estimate (X_i^4) for batch formation variability was the best one.

4.7.3.3 Summary for our branching process

In Section of 4.5 of this chapter, we introduced our applied branching or splitting process of individuals after the batch processing station. It should be said again that the utilized probability distribution to model this process was a conditional binomial distribution with positive batch size, and, based on that, we were able to find out Formulas 25, 26, 27, 28 and 29 to model the our branching process completely.

Chapter 5: Model formulation

In this chapter, which is the main objective of this thesis, we bring our findings and formulas from Chapter 4 and integrate them with other existing models for queueing system. In order to construct our final model of the mass dispensing and vaccination clinic, we divide the clinic queueing systems into 6 different types of stations whose related equations and formulas can be either completely different or similar to each other.

These new types of queueing systems are defined based on all combinations of arrival process and service process. The arrival process may be individual or groups from multiple arrival streams with batch size variability. The service process may be individual service, batch service, or self service. As noted before, for queueing systems with the mixed arrival and batch service process, we study only the case where the average arrival batch size is less than the batch processing size.

In this chapter, first, we introduce the approach that we are going to utilize to construct our complete clinic models. Then, we analyze mathematically the behavior of all 6 types of queueing systems.

Finally, we validate our clinic model by running some long-run simulation for specific clinic examples and comparing the simulation results with the estimates obtained from our mathematical equations.

In order to have a consistent notation among all 6 types of queueing systems, we changed some of the notation from the previous chapter and added some new notation to be able to study the behavior of all types of queueing systems perfectly.

5.1 Our approach and assumptions

This section introduces our approach, which is generally a type of decomposition method, introduced in Chapter 2, as well as some assumptions required to model mass dispensing and vaccination clinics.

We will make use of parametric decomposition approach (PDA), which is a type of decomposition. Networks of queues have proven to be useful models to analyze the performance of complex systems such as computers, communications networks, and production job shops.

There is a network of nodes and directed arcs. The nodes represent service facilities, and the arcs represent flows of customers, jobs, or packets. There is also one external node, which is not a service facility, representing the outside world. Customers enter the network on directed arcs from the external node to the internal nodes, move from node to node along the internal directed arcs, and eventually leave the system on one of the directed arcs from an internal node to the external node. The flows of customers on the arcs are assumed to be random so that they can be represented as stochastic processes.

If all servers are busy at a node when a customer arrives, then the customer joins a queue and waits until a server is free. When there is a free server, that customer begins service, which is carried out without interruption. Successive service times at each node are assumed to be random variables, which may depend on the type of customer but which otherwise are independent of the history of the network and are mutually independent and identically distributed. After the customer completes service, s/he goes along some directed arc from that node to another node. The customer receives service in this way from several internal nodes and then eventually leaves the network. A picture of a typical network (without the external node) is given in Figure 47.

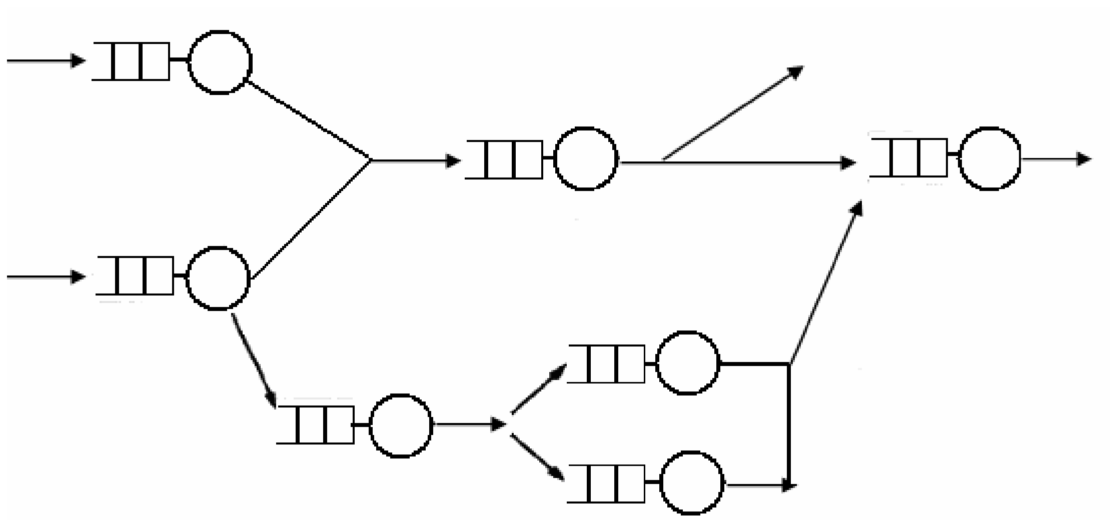

Figure 47. An open network of queues (modified from Whitt, 1983a)

An important feature of this model is that we have movement from node *j* to node *i* in forward flow not backward flow. In other words, as we have in our mass dispensing and vaccination clinics, customers or jobs cannot return to a node where they previously received service.

Our clinic model makes the following assumptions:

Assumption 1. The network is open rather than closed. Customers come from outside, receive service at one or more nodes, and eventually leave the system.

Assumption 2. There are no buffer capacity constraints. There is no limit on the number of customers that can be in the entire network, and each service facility has unlimited waiting space.

Assumption 3. There can be any number of servers at each node. They are identical independent servers, each serving either one customer or a batch of customers at a time. In other words, stations can have batch service processes with different sizes.

Assumption 4. Customers are selected for service at each facility according to the first-come, first-served discipline.

*Assumption 5***.** Customers can be created or combined at the node with the different coming batch sizes and also an arrival can cause more than one departure. In other words, we can have the multiple arrival streams for a workstation.

Assumption 6. The arrival batch size from each arrival stream can have variability. In addition, as we see in Section 5.5.1, the variability in batch size can be generated in superposition (aggregation or merging) and splitting (branching) process.

Assumption 7. The customers can arrive to the first workstation either one by one or in batches.

Assumption 8. Workstation service times and interarrival times follow a general distribution that is characterized by its first two moments.

Assumption 9. From Curry and Deuermeyer (2002), we know that it's better to have the batch move (transfer) after the stations with batch service process instead of splitting those served batches to individuals.

Assumption 10. As we have previously mentioned we only have forward flow not backward flow in our model.

Assumption 11. All of the analysis and calculations is under the steady state condition.

Our basic approach, which is a simplified version of one introduced by Whitt (1983a) as QNA[1], is to represent all the arrival processes and service-time distributions by a few parameters. However, Whitt doesn't have any kinds of mixed arrival, batch service process, and stations with infinite number of servers, which we do allow.

The congestion at each facility is then described by approximate formulas that depend only on these parameters. The parameters for the internal flows are determined by applying an elementary calculus that transforms the parameters for each of the three basic network operations: superposition (merging), splitting, and flow through a queue (departure).

These basic operations are depicted in Figure 48. In this figure (a) is superposition or merging, (b) is splitting (branching or decomposition) and (c) is departure or flow through a queue.

---

[1] Queueing network analyzer

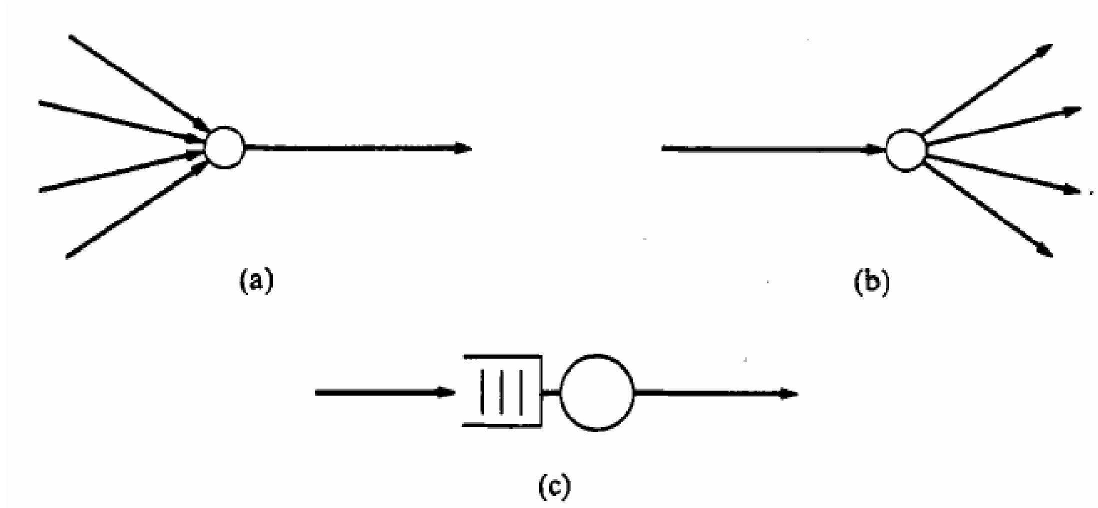

Figure 48. Basic network operations (Whitt, 1983a)

When the network has queues in series, the basic transformations can be applied successively one at a time, but in general it is necessary to solve a system of equations or use an iterative method. To summarize, there are four key elements in this general approach:

1. Parameters characterizing the flows and nodes that will be readily available in applications and that have considerable descriptive power in approximations of the congestion at each node.

2. Approximations for multiple servers queues based on the partial information provided by the parameters characterizing the arrival process and the service-time distribution at each node.

3. A calculus for transforming the parameters to represent the basic network operations: merging, splitting, and departure.

4. A synthesis algorithm to solve the system of equations resulting from the basic calculus applied to the network.

In this approach, we use two parameters to characterize the arrival processes and the service times, one to describe the rate and the other to describe the variability. For the service times, the two parameters are the first two moments. However, we actually work with the mean service time and the squared coefficient of variation, which is the variance of the service time divided by the square of its mean.

For the arrival processes, the parameters are associated with renewal-process approximations. The first two parameters are equivalent to the first two moments of the renewal interval (interval between successive arrival points[1]) in the approximating renewal process. The equivalent parameters we use are the arrival rate**,** which is the reciprocal of the renewal-interval mean, and the squared coefficient of variation**,** which is the variance of the renewal interval divided by the square of its mean.

To sum up, we can say there are three basic steps in the decomposition methods:

1. Characterization of the arrival process: At each station the arrival process resulting from the superposition of different streams arriving to that station is (approximately) determined.

2. Analysis of the queue: Based on the characteristics of the arrival process determined in step 1, the queueing effects at the station are (approximately) computed.

3. Determination of the departure process: The characteristics of the departure process of each product from the station are (approximately) determined. The departure streams in turn become arrivals at some other stations.

[1] Interarrival time

Several variants of the decomposition method can be developed by varying the implementation of the three steps. One of the most often used procedures is the parametric decomposition approach (PDA).

5.1.1 Parametric decomposition approach (PDA)

The approximation method in QNA is perhaps best described as a parametric-decomposition method. Under the parametric decomposition approach (PDA), in addition to assuming that each node can be treated as being stochastically independent (the decomposition assumption), the arrival process to, the departure process from, and the flow between each node are approximated by renewal processes. Further, it is assumed that two parameters: mean and variance of the interarrival and service time distributions are adequate to estimate the performance measures at each node. Hence to compute the performance measures we need to (i) approximate all the flows in the network, and (ii) compute the performance measures based on the first two moments of the interarrival and service times.

Accordingly, the description of the PDA will be in two parts: *flow analysis* and *estimation of performance measures.*

5.1.1.1 Flow analysis

As noted above, we have 3 main operations in decomposition approach: superposition (merging), departure and splitting. We will delve into each of them in this section.

Let p_{ji} be the probability of a job going to station *i*, upon completion of service at station *j*. For the renewal process approximating the flow from station *j* to station *i*, let L_{ji} and $c^2{}_{ji}$ be the mean and the squared coefficient of variation (SCV) of the renewal interval length (interarrival time).

Denote the flow rate from node *j* to *i* by r_{ji} which $r_{ji} = 1/L_{ji}$. In PDA, the superposition of the flows arriving at a node is further approximated by a renewal process. We let r_i and r_{0i} denote the total flow rate and the flow rate from external environment into *i* respectively. The flow rates r_{ji} are determined by the following traffic equations:

$$r_i = r_{0i} + \sum_{j=1}^{i-1} r_j p_{ji}$$
$$r_{ji} = r_j p_{ji}$$

While determining the flow rates is straightforward, approximations are needed for the SCVs. In particular we need procedures for approximating by a renewal process each of the following: (i) superposition of renewal processes, (ii) departure processes from queues, and (iii) flow along each arc out of a node (splitting the departure stream).

(i) Approximations for Superposition of Renewal Processes

In PDA only the mean and the variance of the approximating renewal interval need to be determined. The mean is straightforward to compute: the arrival rate of the approximating process must equal the arrival rate of the superposition process. Whitt

(1982) considers two basic procedures for determining the variance of the approximating process. He calls them micro and macro approaches.

Let S_n be the time of the n^{th} arrival after time 0, and $V(S_n)$ the variance of the random variable S_n. Under the macro approach the variance of the approximating renewal interval is set at $Lim_{n\to\infty} V(S_n)/n$.

The macro approach is also called the asymptotic method. Henceforth we refer to the limiting variance and SCV as the asymptotic variance and asymptotic SCV.

Under the micro approach the variance of the approximating renewal interval is set at $V(S_1)$.The time interval starting from 0 until the first arrival after 0 is referred to as the stationary interval of the superposition process. Henceforth we refer to $V(S_1)$ as the stationary interval variance and the corresponding SCV as the stationary interval SCV.

The asymptotic SCV can be computed readily from the SCVs of the interarrival times of each of the process being merged. We let $c^2{}_{ai}$ and $c^2{}_{ji}$ denote the total interarrival time SCV and interarrival time SCV from station *j* into *i* respectively. The asymptotic SCV of the arrivals to station i can be given by:

$$c^2{}_{ai} = \frac{\sum_{j=1}^{i-1} c^2{}_{ji} r_{ji}}{r_i}$$

When the two approaches, micro and macro, were used to estimate performance measures of queueing systems, Whitt (1982) and Albin (1981, 1984) found that neither method dominated. Based on their experiments they discovered that a convex combination of the SCVs provided by the micro and macro approaches yielded the best results. This approach has been called the hybrid approach (Albin, 1984).

If we let c^2_{as}, c^2_s and c^2_h, denote the asymptotic, stationary, and hybrid SCV respectively. Then, $c^2_h = wc^2_{as} + (1-w)c^2_s$, where $0 \le w \le 1$, and w is a function of the utilization of the server and the number of arrival streams being merged.

As the number of arrival processes being merged goes to infinity, the stationary interval is asymptotically correct. On the other hand as the utilization goes to 1, the asymptotic limit is asymptotically correct. The weighting factor w is so chosen that as the number of process being merged goes to infinity, w goes to zero, and as the utilization goes to 1, w goes to 1.

In the queueing network analyzer proposed by Whitt (1983a) the following approximation is used:

$$
\begin{aligned}
c_{ai}{}^2 &= 1 - w_i + \frac{w_i}{r_i} \sum_{j \in S_i} r_{ji} c_{ji}^2 \\
w_i &= \frac{1}{1 + 4(1-u_i)^2 (v_i - 1)} \\
v_i &= \left(\sum_{j \in S_i} \left(\frac{r_{ji}}{r_i} \right)^2 \right)^{-1} \\
u_i &= \text{Utilization of station i}
\end{aligned}
\qquad \textbf{(Formula 30)}
$$

We should say that, Formula 30 is our best choice to calculate the aggregate interarrival time SCV in this thesis.

(ii) Approximations for the departure process

The departure process from a queue is in general not a renewal process. However, in PDA it is approximated by a renewal process. The mean of the approximating renewal interval is easy to determine. Two alternatives have been considered for the variance: the stationary departure interval variance and the asymptotic limit.

Whitt (1984) shows that for *G/G/m* queues with utilization less than 1, the asymptotic variance of the departure process is the same as the variance of the interarrival times. Hence, once again the asymptotic limit is easy to determine. However the computational tests indicated that the stationary interval provides better approximation, and that was adopted by Whitt (1983a).

Unfortunately, determining the stationary interval distribution of the departure stream is not easy, and instead of computing the exact stationary interdeparture interval SCV, approximations are employed.

Combining the formula for the stationary interval due to Marshall (1968) with the Kraemer-Langenbach-Belz (1976) approximation for the expected waiting time, Whitt (1983a) obtains the following approximation formula for the interdeparture time SCV. If we let $c_{di}{}^2$ and $c_{ei}{}^2$ to be interdeparture time SCV from station *i* and service time SCV at station *i*, respectively, we have:

$$c_{di}{}^2 = u_i{}^2 c_{ei}{}^2 + (1 - u_i{}^2) c_{ai}{}^2$$

It can be concluded that if workstation i is always busy, so that if u_i =1, then $c_{di}{}^2 = c_{ei}{}^2$. Similarly, if the machine is (almost) always idle, so that if u_i =0, then $c_{di}{}^2 = c_{ai}{}^2$. For intermediate utilization levels, 0 < u_i < 1, $c_{di}{}^2$ is a combination of $c_{ai}{}^2$ and $c_{ei}{}^2$.

When there is more than one server (m>1), the following formula from Hopp and Spearman (2001) is a reasonable way to estimate $c_{di}{}^2$ for station i:

$$c_{di}{}^2 = 1 + (1 - u_i{}^2)(c_{ai}{}^2 - 1) + \frac{u_i{}^2 (c_{ei}{}^2 - 1)}{\sqrt{m_i}} \qquad \textbf{(Formula 31)}$$

Whitt (1983a) suggested modifying Formula 31 to Formula 32 to have a better estimate for interdeparture time SCV. Formula 32 shows the new formula for $c_{di}{}^2$, which we utilize in this thesis for stations with batch service process with any types of arrival and also stations with individual arrival and individual service process.

$$c_{di}{}^2 = 1 + (1 - u_i{}^2)(c_{ai}{}^2 - 1) + \frac{u_i{}^2 (\max(c_{ei}{}^2, 0.2) - 1)}{\sqrt{m_i}} \qquad \textbf{(Formula 32)}$$

For stations with batch arrival process and individual service process, the only existing formula for individual interdeparure time SCV is from Curry et al. (2002), for stations with single server. This formula for the station i is:

$$c_{di}{}^2 = \overline{K}_{Ai}(1 - u_i{}^2)c^2{}_{bi} + u_i{}^2 c_{ei}{}^2 + (\overline{K}_{Ai} - 1)(1 - u_i)^2 \qquad \textbf{(Formula 33)}$$

In Formula 33, $\overline{K}_{Ai}$ is the average batch arrival size to station i and $c^2{}_{bi}$ is the batch interarrival time SCV to station i.

From simulation results, we see that Formula 33 can be used approximately for stations with multiple servers as well. For simplicity, we can have a single server with a faster process time instead of having *m* servers. Additionally, a simple algebra with service time variance and mean of service time shows that the service time SCV for stations with a single server is equal to the service time SCV for stations with multiple servers. That is why Formula 33 can be used for all stations with any number of servers with batch arrival and individual service process approximately.

In the modeling section, we will make use of Formula 33 to calculate c_{di}^{2} for any stations with (mixed) batch arrival and individual service process discipline.

(iii) Approximations for flow along each arc (splitting)

If the routing is Markovian, and the departures from the station are approximated by a renewal process, the flow along each arc will be a renewal process. Under these assumptions, the interdeparture time along each arc out of the station will be the random sum of interdeparture times from the station. The number of interdeparture times (from the station) that have to be convoluted is of course geometrically distributed. Hence the SCVs for the flows along each arc can readily be expressed in terms of the interdeparture time SCV from the source station and the routing probabilities.

From Sevcik et al. (1977), the simplest interarrival time SCV from station j to station i (c^{2}_{ji}) equation is defined below as a function of c_{dj}^{2}:

$$c^{2}_{ji} = p_{ji}c_{dj}^{2} + 1 - p_{ji} \qquad \textbf{(Formula 34)}$$

If the departure process is Poisson (i.e., $c_{dj}^{2} = 1$), then Formula 34 is exact and gives $c^2{}_{ji}$ =1. Note that as $p_{ji} \to 1$, then Formula 34 results in $c^2{}_{ji} \to c^2{}_{dj}$. That is, as the expected departure rates from station j to station i tend to the merged expected departure rate from station j, the interarrival time SCV from station j to station i also tends to the merged interdeparture time SCV from station j.

Furthermore, as $p_{ji} \to 0$, then Formula 34 results in $c^2{}_{ji} \to 1$, indicating that as the proportion of flow between stations j and i tends to zero, the departure process between these two stations tends to a Poisson process.

Formula 34 can be a good estimation for analyzing the splitting process after stations with individual service process and one arrival stream. So in this thesis, it is a good formula for analyzing the splitting process after stations with mixed arrival and individual service process.

When we have a mixed arrival process, we change all these streams to one imaginary stream with the size of the average batch size of coming batches from all arrival streams. However, Formula 34 is not a good estimation for stations with the multiple arrival streams.

In order to improve Formula 34 and find a better estimate when we want to calculate the $c^2{}_{ji}$ after stations with multiple arrival streams, intuitively, we assume each of arrival streams to a station similarly can be considered as a different class of a product. Under this assumption, it's possible to make use of an alternative expression for the splitting process of multiple classes with deterministic routings introduced by Segal and Whitt (1989) for individual service process.

If we let ${ce_j}^2$ be the average of the external interarrival time SCV of the classes at station *j,* weighted by the expected number of visits of each class at station j, $c^2{}_{ji}$ can be:

$$c^2{}_{ji} = p_{ji}{c_{dj}}^2 + (1-p_{ji})p_{ji}{c_{aj}}^2 + (1-p_{ji})^2 ce^2{}_j \qquad \textbf{(Formula 35)}$$

$$ce^2{}_j = \frac{\sum_{n=1}^{j-1} r_{nj}c_{nj}^2}{r_j}$$

In this thesis, we assume that a station can have multiple arrival streams. So, we can make use of Formula 36 for all stations with individual arrivals and individual service process.

$$c^2{}_{ji} = p_{ji}{c_{dj}}^2 + (1-p_{ji})p_{ji}{c_{aj}}^2 + (1-p_{ji})^2 \frac{\sum_{n\in S_j} r_{nj}c_{nj}^2}{r_j} \qquad \textbf{(Formula 36)}$$

From Formula 30, we have only one stream arriving to station j, then $ce^2{}_j = {c_{aj}}^2$ and then interestingly Formula 35 will be simplified to Formula 37, which is similar to Formula 34. Formula 37 can be a good estimate for $c^2{}_{ji}$ in self service stations with any type of the arrival process.

$$c^2{}_{ji} = p_{ji}{c_{dj}}^2 + (1-p_{ji}){c_{aj}}^2 \qquad \textbf{(Formula 37)}$$

However, for analyzing the splitting process after stations whose batch processing size is bigger than 1 ($k_j > 1$), none of afore-mentioned formulas for $c^2{}_{ji}$ is a good estimation. So, we need to have a new estimation for this case, since Formulas 34, 35 and 36 are for stations with individual service process ($k_j = 1$).

According to the branching approach described in Section 4.5, Formula 37 can be changed to its general form for a station with batch service process and a batch process size of k_j.

$$c^2_{ji} = (1-(1-p_{ji})^{k_j})c_{dj}^2 + (1-p_{ji})^{k_j} c_{aj}^2 \qquad \textbf{(Formula 38)}$$

Formula 38 can be employed for the splitting process after stations with more than one arrival streams such as mixed arrival process, when according to Formula 35, $ce^2_j = c_{aj}^2$. A good example of this situation is a station with a batch service process and mixed arrivals. We need to have the batch formation process before it. This makes ce^2_j equal to c_{aj}^2 and changes both of them to c_{bj}^2 (the batch interarrival time SCV of station j after batch formation). In this way, we have to change Formula 38 to Formula 39, which has c_{bj}^2 instead of c_{aj}^2. Formula 39 can be a good estimation when we have mixed arrivals (more than one arrival stream) to a batch service process in this thesis.

We should emphasize again that we study only batch process where the average arrival batch size is less than the process batch process size.

$$c^2_{ji} = (1-(1-p_{ji})^{k_j})c_{dj}^2 + (1-p_{ji})^{k_j} c_{bj}^2 \qquad \textbf{(Formula 39)}$$

Until now, we studied the splitting process for all the kinds of stations in this section, the only remaining type which has to be considered is stations with one individual arrival stream and batch service process.

From Section 4.5 for the splitting process after the batch process stations and Formula 34, which is suitable for only one arrival stream, Formula 40 can be used for the splitting process after a station with an individual arrival process and a batch service process.

Formula 34 depends upon p_{ji}, the probability of an individual going from station *j* to station *i*. The probability of having a batch of at least individual moving to station *i* from station *j* is $1-(1-p_{ji})^{k_j}$.

Similarly, in Formula 34, the probability that the individual goes to some other station is $(1-p_{ji})$, while the probability of having no batches moving to station *i* from station *j* is $(1-p_{ji})^{k_j}$ .This leads to Formula 40:

$$c^2_{ji} = (1-(1-p_{ji})^{k_j})c_{dj}{}^2 + (1-p_{ji})^{k_j} \qquad \textbf{(Formula 40)}$$

<u>5.1.1.2 Estimation of performance measures</u>

The performance measures at each station are estimated using approximations that are based on the first two moments of the interarrival and service times. A wide variety of approximations have been proposed for the analysis of *G/G/m* queues. From Hopp and Spearman (2001) we have (as a reminder t_i is the average processing times of station *i*):

$$CT_{qi}(\text{G/G/m}) = \left(\frac{c_{ai}^2 + c_{ei}^2}{2}\right) \cdot \left(\frac{u_i^{\sqrt{2m_i+2}-1}}{m_i(1-u_i)}\right) \cdot t_i \qquad \textbf{(Formula 41)}$$

The *CTq* for *G/G/*1 is $\left(\frac{c_{ai}^2 + c_{ei}^2}{2}\right) \cdot \frac{u_i t_i}{(1-u_i)}$. Whitt (1984) and Bitran et al. (1989), propose a better approximation for *CTq* in *G/G/* 1:

$$\begin{aligned} &CT_{qi}(\text{G/G/1}) = \left(\frac{c_{ai}^2 + c_{ei}^2}{2}\right) \cdot \frac{u_i t_i}{(1-u_i)} . g_i \\ &g_i = \exp\left[\frac{-2(1-u_i)(1-c_{ai}{}^2)^2}{3u_i(c_{ai}{}^2 + c_{ei}{}^2)}\right] \qquad if\ c_{ai}{}^2 < 1 \qquad \textbf{(Formula 42)} \\ &g_i = 1 \qquad if\ c_{ai}{}^2 \geq 1 \end{aligned}$$

For *G/G/m* we can divide the t_i by m_i, so instead of Formula 41, we can create Formula 43:

$$\begin{aligned} &CT_{qi}(\text{G/G/m}) = \left(\frac{c_{ai}^2 + c_{ei}^2}{2}\right) \cdot \frac{u_i t_i}{m_i(1-u_i)} g_i \\ &g_i = \exp\left[\frac{-2(1-u_i)(1-c_{ai}{}^2)^2}{3u_i(c_{ai}{}^2 + c_{ei}{}^2)}\right] \qquad if\ c_{ai}{}^2 < 1 \qquad \textbf{(Formula 43)} \\ &g_i = 1 \qquad if\ c_{ai}{}^2 \geq 1 \end{aligned}$$

To improve Formula 43, we can the u_i in the numerator by $u_i^{\sqrt{2m_i+2}-1}$. In this way, we estimate formula *CTq* using Formula 44:

$$CT_{qi}(\text{G/G/m}) = \left(\frac{c_{ai}^2 + c_{ei}^2}{2}\right) \cdot \frac{u_i^{\sqrt{2m_i+2}-1} t_i}{m_i\left(1-u_i\right)} . g_i$$

$$g_i = \exp\left[\frac{-2(1-u_i)(1-c_{ai}{}^2)^2}{3u_i(c_{ai}{}^2 + c_{ei}{}^2)}\right] \quad if\ c_{ai}{}^2 < 1 \qquad \textbf{(Formula 44)}$$

$$g_i = 1 \quad if\ c_{ai}{}^2 \geq 1$$

Furthermore, when the queueing system is heavily loaded, or u_i approaches 1, the heavy traffic approximation (Kollerstrom, 1974) of the queueing system states that the distribution of steady state waiting time in queue in a *G/G/m* system is approximately exponential with mean value of:

$$CT_{qi}(\text{G/G/m}) = \frac{c_{ai}^2 / r_i + c_{ei}{}^2 t_i{}^2 r_i / m_i}{2(1-u_i)} \qquad \textbf{(Formula 45)}$$

In Formula 45, r_i is the arrival rate at station *i* (residents per minute). We use Formula 45 for stations with utilization higher than 90%.

Therefore, our model uses Formula 44 and 45 for high utilization stations. The formulas change slightly depending upon the type of arrival process and type of service process.

5.2 Model description

After introducing all of the approaches, findings, and formulas necessary to analyze the different types of stations, we present in this section the complete models for the 6 different types of stations and estimate their performance measures.

First, we bring in new more general notation to be able to manifest the behaviors of all 6 types of stations consistently. Most of the notation is similar to the notation used in previous sections. Finally, we present the 6 queueing systems that represent the types of stations found in mass dispensing and vaccination clinics.

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.

5.2.1 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

5.2.2 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

$\overline{C}_{Bji}^2$ = 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

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

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

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

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

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

c_{Bji}^2 = 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)

5.2.3 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 $TH' = \frac{P}{60H}$.

A key concept in the queueing network model is the flow of batches from on 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 interarrival times SCV), 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. All arriving residents go to the first station. Therefore, $p_{01} = 1$, while $p_{0i} = 0$ and $\lambda_{B0i} = 0$ for all i > 1. If the first station has a batch process, the SCV of batch size many be positive.

$$
\begin{aligned}
r_1 &= TH' \\
\lambda_{A1} &= \lambda_{B01} = TH' / k_0 \\
\bar{K}_{A1} &= k_0 \\
\bar{C}_{A1}^2 &= \bar{C}_{B01}^2
\end{aligned}
$$

We calculated the arrival rates for the other stations based on the routing probabilities.

$$r_i = \sum_{j=1}^{i-1} \lambda_{Bji} \bar{K}_{Bji} = \sum_{j=1}^{i-1} r_j p_{ji}$$ **(See Section 4.4.2, Formula 23)**

$$\lambda_{Ai} = \sum_{j \in S_i} \lambda_{Bji}$$ **(See Section 4.5, Formula 26)**

$$\bar{K}_{Ai} = \frac{1}{\lambda_{Ai}} \sum_{j \in S_i} \lambda_{Bji} \bar{K}_{Bji}$$ **(See Section 4.5, Formula 29)**

Following Whitt (1983a), we estimate the aggregate batch interarrival times SCV at each station as follows:

$$c_{ai}^2 = 1 - w_i + \frac{w_i}{\lambda_{Ai}} \sum_{j \in S_i} \lambda_{Bji} c_{Bji}^2$$
$$w_i = \frac{1}{1 + 4(1-u_i)^2 (v_i - 1)}$$
$$v_i = \left(\sum_{j \in S_i} \left(\frac{\lambda_{Bji}}{\lambda_{Ai}} \right)^2 \right)^{-1}$$
(See Section 5.1.1.1, Formula 30)

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

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

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

The average time spent traveling to the next station after station *i* depend 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).

1. Individual arrivals, individual service process

2. Individual arrivals, batch service process

3. Individual arrivals, self-service

4. Mixed arrivals, individual service process

5. Mixed arrivals, batch service process

6. Mixed arrivals, self-service

5.2.4 Individual arrivals, individual service process

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 interarrival time variability c_{ai}^2 can be determined as discussed in outputs (Section 5.2.3). In this case, $\overline{K}_{Bji} = 1$, $C_{Bji}^2 = 0$ and 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:

$$CT_{qi} = \left(\frac{c_{ai}^2 + c_{ei}^2}{2}\right)\left(\frac{u_i^{\sqrt{2m_i+2}-1}\, t_i}{m_i\left(1-u_i\right)}\right).g_i \qquad \textbf{(See Section 5.1.1.2, Formula 42)}$$

Where the parameter g_i, suggested by Whitt (1984) and Bitran et al. (1989), equals 1 if the interarrival time variability $c_{ai}^2 \geq 1$. However, if $c_{ai}^2 < 1$, then it can be determined as follows:

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

Additionally, from Section 5.1.1.2 for utilization higher than 90% from Section 5.1.1.2, waiting time for service can be estimated by following term as well.

$$CT_{qi} = \frac{c_{ai}^2 / \lambda_{Ai} + c_{ei}{}^2 t_i{}^2 \lambda_{Ai} / m_i}{2(1-u_i)}$$ **(See Section 5.1.1.2, Formula 45)**

The cycle time at station *i* is:

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

For the interdeparture time SCV, as we had in Section 5.1.1.1, we use the interdeparture time variability estimate from Hopp and Spearman (2001) and adapt results from Whitt (1983a). The batch flow from workstation *i* to a downstream workstation *n* is characterized as follows:

$$\lambda_{Bin} = r_i p_{in}$$ **(See Section 4.5, Formula 25)**

$$c_{di}^2 = 1 + \left(1 - u_i{}^2\right)\left(c_{ai}^2 - 1\right) + \frac{u_i{}^2}{\sqrt{m_i}}\left(Max(0.2, c_{ei}^2) - 1\right)$$ **(See Section 5.1.1.1, Formula 32)**

$$c_{Bin}^2 = p_{in} c_{di}^2 + \left(1 - p_{in}\right) p_{in} c_{ai}^2 + \frac{\left(1 - p_{in}\right)^2}{\lambda_{Ai}} \sum_{j=1}^{i-1} \lambda_{Bji} c_{Bji}^2$$ **(See Section 5.1.1.1, Formula 36)**

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

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

5.2.5 Individual arrivals, batch service process

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 processes 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 (see Section 4.4.2 for more details).

As before, the arrival rate r_i and interarrival time variability c_{ai}^2 can be determined as discussed in outputs (Section 5.2.3). In this case, $\bar{K}_{Bji} = 1$ and $\bar{C}_{Bji}^2 = 0$ for all upstream workstations j that send residents to workstation i, so consequently $\bar{K}_{Ai}$ equals to1.

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

$$WTBT_i = \frac{k_i - 1}{2r_i}$$ **(See Section 4.4.2, Formula 24)**

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

From Section 4.4.1, the best found formula experimentally to calculate c_{bi}^2, interarrival time SCV for process batches at station *i* after being formed, is:

$$c_{bi}^2 = \frac{c_{ai}^2}{k_i} \qquad \textbf{(See Section 4.4.1, Formula } X_i^1 \textbf{ when } \bar{K}_{Ai} = 1\textbf{)}$$

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

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

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

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

Similar to individual arrival/individual service process, for the cases with utilization higher than 90% from Section 5.1.1.2, 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 $\frac{\lambda_{Ai}}{k_i}$):

$$CT_{qi} = \frac{k_i c_{bi}^2 / \lambda_{Ai} + c_{ei}{}^2 t_i{}^2 \lambda_{Ai} / (k_i m_i)}{2(1-u_i)} \qquad \textbf{(See Section 5.1.1.2, Formula 45)}$$

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 (See our batch branching approach in Section 4.5).

Not all of the residents in a particular process batch go to the same station. The batch flow from workstation i to a downstream workstation n is characterized as follows:

$$\lambda_{Bin} = \frac{r_i}{k_i}\left(1-\left(1-p_{in}\right)^{k_i}\right)$$ **(See Section 4.5, Formula 25)**

$$c_{di}^2 = 1+\left(1-u_i^{\,2}\right)\left(c_{bi}^2-1\right)+\frac{u_i^{\,2}}{\sqrt{m_i}}\left(Max(0.2, c_{ei}^2)-1\right)$$ **(See Section 5.1.1.1, Formula 32)**

$$c_{Bin}^2 = (1-(1-p_{in})^{k_i})c_{di}^2 + (1-p_{in})^{k_i}$$ **(See Section 5.1.1.1, Formula 40)**

$$\bar{K}_{Bin} = \frac{k_i p_{in}}{1-\left(1-p_{in}\right)^{k_i}}$$ **(See Section 4.5, Formula 27)**

$$\bar{C}_{Bin}^2 = \frac{1-p_{in}-\left(k_i p_{in}+1-p_{in}\right)\left(1-p_{in}\right)^{k_i}}{k_i p_{in}}$$ **(See Section 4.5, Formula 28)**

5.2.6 Individual arrivals, self service

In this case, residents arrive individually to the workstation. The residents perform the process themselves without any external resources. In this domain, an example would be a workstation where each resident must complete a form. Thus, as we studied in Section 4.3, the workstation can be modeled as a *G/G/∞* queueing system.

The arrival rate r_i and interarrival time variability c_{ai}^2 can be determined as discussed in outputs (Section 5.2.3). The only point is that to calculate c_{ai}^2 for self service station, w_i in Formula 30 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 time variability, we first take into account the following facts. As a reminder, for a *G/D/∞* system, the interdeparture time SCV equals the interarrival time SCV because the departure process is simply the arrival process shifted by a constant equal to the processing time. For a *M/G/∞* system, the departure process is a Poisson process; thus the interdeparture time SCV equals 1. For a *G/G/∞* system, Whitt (1983a) suggests that the interdeparture time SCV 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 time SCV to equal the interarrival time SCV. These imply that, in the general case (a *G/G/∞* system with moderate load), the interdeparture time SCV will be somewhere between the interarrival time SCV and one, in which it will depend upon the load. Therefore, we conducted experiments to characterize this relationship and to examine various weights for interpolating between the arrival variability and one. Based on the results (Section 4.3.1), we decided to use the following approximation:

$$\rho_i = r_i t_i$$ **(See Section 4.3)**

$$c_{di}^2 = c_{ai}^2 \left(1 - \frac{\rho_i^2 c_{ei}^2}{\left(1 + \rho_i \sqrt{c_{ei}^2}\right)^2} \right) + \frac{\rho_i^2 c_{ei}^2}{\left(1 + \rho_i \sqrt{c_{ei}^2}\right)^2}$$ **(See Section 4.3)**

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.

$$\lambda_{Bin} = r_i p_{in}$$ **(See Section 4.5, Formula 25)**

$$c_{Bin}^2 = p_{in} c_{di}^2 + (1 - p_{in}) c_{ai}^2$$ **(See Section 5.1.1.1, Formula 34)**

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

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

5.2.7 Mixed arrivals, individual service process

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 process workstations, and the batch sizes from each workstation can be random varying due to the routing probabilities. There are also individual arrivals from individual process 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 calculated the variability (SCV) of the arriving batch size in Section 4.4.1.1 by adapting a formula from Fowler et al. (2002), who calculated the process time SCV for different products that arrive at different rates.

$$\bar{C}_{Ai}^2 = -1 + \frac{1}{\lambda_{Ai}\bar{K}_{Ai}^2}\sum_{j\in S_i}\lambda_{Bji}\left(1+\bar{C}_{Bji}^2\right)\bar{K}_{Bji}^2$$ **(See Section 4.4.1.1, Formula 22)**

The arrival rate r_i and arrival variability c_{ai}^2 can be determined as discussed in outputs (Section 5.2.3).

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 $\bar{C}_{Ai} + c_{ei}^2 / \bar{K}_{Ai}$ from Buzacott and Shanthikumar (1993).

$$T_i = \frac{\bar{K}_{Ai} t_i}{m_i} u_i^{(1-\frac{1}{m_i})} \quad \textbf{(See Section 4.6)}$$

$$CT_{qi} = \frac{1}{2}\left(c_{ai}^2 + \bar{C}_{Ai} + \frac{c_{ei}^2}{\bar{K}_{Ai}}\right)\left(\frac{u_i}{1-u_i}\right) T_i g_i \quad \textbf{(See Section 5.1.1.2, Formula 42 when } m_i \textbf{ =1)}$$

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 $c_{ai}^2 \geq 1$. Otherwise,

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

Similar to other cases, for the individual process stations with mixed arrival and the utilization higher than 90% from Section 5.1.1.2, 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:

$$CT_{qi} = \frac{c_{ai}^2 / \lambda_{Ai} + (\bar{C}_{Ai} + \frac{c_{ei}^2}{\bar{K}_{Ai}}) T_i^2 \lambda_{Ai}}{2(1-u_i)} \quad \textbf{(See Section 5.1.1.2, Formula 45)}$$

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.

According to our experiments and analysis for WIBT from Section 4.2 for general cases, the best WIBT yields from Formulas 4 and 21:

$$WIBT = \frac{(\overline{K}_{Ai} - 1)t_i}{2} \quad \text{if } m_i = 1$$ **(See Sections 4.2.3.4 and 2.2.4, Formula 4)**.

$$WIBT = \frac{(\overline{K}_{Ai}{}^{2-u_i} - 1)t_i u_i}{(3-u_i)m_i^{2-u_i}} \quad \text{if } m_i > 1$$ **(See Section 4.2.3.4, Formula 21)**

The cycle time at station *i* is $CT_i = CT_{qi} + WIBT_i + t_i + W_i$. Using the interdeparture time SCV approximation from Curry and Deuermeyer (2002) in Section 5.1.1.1 for multiple servers, the batch flow from workstation *i* to a downstream workstation *n* is characterized as follows:

$$\lambda_{Bin} = r_i p_{in}$$ **(See Section 4.5, Formula 25)**

$$c_{di}^2 = \overline{K}_{Ai} c_{ai}^2 \left(1 - u_i{}^2\right) + (1-u_i)^2 (\overline{K}_{Ai} - 1) + u_i{}^2 c_{ei}^2$$ **(See Section 5.1.1.1, Formula 33)**

$$c_{Bin}^2 = p_{in} c^2{}_{di} + 1 - p_{in}$$ **(See Section 5.1.1.1, Formula 34)**

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

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

5.2.8 Mixed arrivals, batch service process

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 can be random varying 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.

Just as a reminder, in this thesis, we only study the mixed arrival with batch service workstation in which the average arriving batch size to the station is equal or smaller than the batch processing size of the station.

As before, we calculated the SCV of the arriving batch size by adapting a formula from Fowler et al. (2002) in Section 4.4.1.1. 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 batch size for arriving batches:

$$C_{Ai}^2 = \frac{1}{\lambda_{Ai}\bar{K}_{Ai}^2}\sum_{j\in S_i}\lambda_{Bji}\left(1+C_{Bji}^2\right)\bar{k}_{ji}^2 - 1$$ **(See Section 4.4.1.1, Formula 22)**

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

$$WTBT_i = \frac{k_i - 1}{2r_i}$$ **(See Section 4.4.2, Formula 24)**

For the interarrival time SCV of process batches c_{bi}^2 (after they are formed), we will use the results from Section 4.4.1.1 to choose the best formula among them, which is:

$$c_{bi}^2 = \frac{\overline{K}_{Ai}}{k_i}\left(\frac{\sum_{j \in S_i} \overline{K}_{Bji}^2 \lambda_{Bji} c_{Bji}^2}{\sum_{j \in S_i} \overline{K}_{Bji}^2 \lambda_{Bji}} + \overline{C}_{Ai}^2\right)$$ **(See Section 4.4.1.1, Formula X_i^4)**

The arrival rate r_i and $\overline{K}_{Ai}$ can be determined as discussed in outputs (Section 5.2.3).

Then, we estimate the queueing in the resulting *G/G/m* system:

$$CT_{qi} = \left(\frac{c_{bi}^2 + c^2{}_{ei}}{2}\right)\left(\frac{u_i^{\sqrt{2m_i+2}-1}}{m_i(1-u_i)}\right)t_i g_i$$ **(See Section 5.1.1.2, Formula 42)**

The parameter g_i, suggested by Whitt (1984) and Bitran et al. (1989), equals 1 if the batch 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 Section 5.1.1.2, 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 $\frac{\overline{K}_{Ai}\lambda_{Ai}}{k_i}$):

$$CT_{qi} = \frac{k_i c_{bi}^2 / (\overline{K}_{Ai}\lambda_{Ai}) + c_{ei}{}^2 t_i{}^2 \overline{K}_{Ai}\lambda_{Ai} / (k_i m_i)}{2(1-u_i)}$$ **(See Section 5.1.1.2, Formula 45)**

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:

$$\lambda_{Bin} = \frac{r_i}{k_i}\left(1-(1-p_{in})^{k_i}\right)$$ **(See Section 4.5, Formula 25)**

$$c_{di}^2 = 1 + \left(1-u_i{}^2\right)\left(c_{bi}^2 - 1\right) + \frac{u_i{}^2}{\sqrt{m_i}}\left(Max(0.2, c_{ei}^2) - 1\right)$$ **(See Section 5.1.1.1, Formula 32)**

$$c_{Bin}^2 = (1-(1-p_{in})^{k_i})c_{di}^2 + (1-p_{in})^{k_i} c^2{}_{bi}$$ **(See Section 5.1.1.1, Formula 39)**

$$\overline{K}_{Bin} = \frac{k_i p_{in}}{1-(1-p_{in})^{k_i}}$$ **(See Section 4.5, Formula 27)**

$$\overline{C}_{Bin}^2 = \frac{1-p_{in} - (k_i p_{in} + 1 - p_{in})(1-p_{in})^{k_i}}{k_i p_{in}}$$ **(See Section 4.5, Formula 28)**

5.2.9 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 Formula 30 should be 1.

To estimate the interdeparture time variability, we adapt the estimate used in the self-service case (Section 4.3). The key change is that the interarrival time variability of individuals depends upon the batch size and the batch interarrival time variability.

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 don't have any effect of C_{Ai}^2 in the formula.

$$\rho_i = r_i t_i \quad \textbf{(See section 4.3)}$$

$$c_{di}^2 = \left(\bar{K}_{Ai} c_{ai}^2 + \bar{K}_{Ai} - 1\right)\left(1 - \frac{\rho_i^2 c_{ei}^2}{\left(1+\rho_i\sqrt{c_{ei}^2}\right)^2}\right) + \frac{\rho_i^2 c_{ei}^2}{\left(1+\rho_i\sqrt{c_{ei}^2}\right)^2} \quad \textbf{(See Section 4.3)}$$

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:

$$\lambda_{Bin} = r_i p_{in} \qquad \textbf{(See Section 4.5, Formula 25)}$$

$c_{Bin}^2 = p_{in} c_{di}^2 + (1 - p_{in}) c_{ai}^2$ **(See Section 5.1.1.1, Formula 34)**

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

$C_{Bin}^2 = 0$

We should say that we don't have any numerical results and experiments for the mixed arrival with self service station.

5.2.10 Clinic Performance Measures

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

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

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

$$TCT = \frac{1}{r_1} \sum_{i=1}^{I} r_i CT_i$$

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

$$WIP = r_1 TCT$$

5.3 Model validation

In order to evaluate this queueing network approximation, we compared the model's results to the results from a discrete-event simulation package using Rockwell Software's Arena® 5.00. To validate our formulas for our clinic modeling, we will carry out four different kinds of experiments including a few tests and scenarios for each one. In each of these experiments, we only have bus arrival process to the first station at the clinic. Moreover, we assume that the traveling time among stations is negligible in our calculations.

For the first experiment, we designed experiments for different scenarios for a mass smallpox vaccination clinic that includes batch processes. In this experiment, we relied on a time study of a mass smallpox vaccination clinic exercise to collect our needed data. In this exercise, we didn't have any self service stations, and we assumed that the arrival bus size was fixed.

For the remaining 3 experiments, we will have 2 tests for each one. In each of these tests, we design the test to have different combinations of stations to be able to validate thoroughly the 6 queueing cases proposed previously in Section 5.2. The major differences between the last 3 experiments and the first one are that, in the last 3 experiments, we will have self service workstations. Additionally, we will have the arrival bus size variability.

For each of these 4 experiments, we use the simulation results for the confidence interval of 95%.

5.3.1 First experiment

To obtain data for this experiment, we relied on 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.

The model was tested at several levels of resident arrival rates, from 20% to 97.5% of clinic capacity under the different scenario. We ran 10 replications of 2000 hours, with 500 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.

In our model of a mass smallpox vaccination clinic, residents arrived by bus. Each bus brought exactly 50 residents. Bus interarrival times were exponentially distributed. Table 94 describes each of the eight stations in the clinic. Table 95 shows the routing probabilities from one station to another (only station numbers are shown to save space). Table 96 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.

We should say that in this experiment, we have 21 scenarios in which arrival rate starts from nearly its maximum rate (5.123 residents per min) and continues to the lower rates.

Table 94. Parameters for Mass Smallpox Vaccination Clinic (first experiment)

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 95. Routing table for Mass Smallpox Vaccination Clinic

	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 96. Capacity for Mass Smallpox Vaccination Clinic's stations

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 97 shows the average total cycle time in terms of minute for each entity in the clinic from simulation results and our clinic mathematical models as well as the percentage of error between them.

Moreover, Figure 49 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.

Table 97. Comparison of total cycle time for mass smallpox vaccination clinic

Scenario	Arrival rate to the clinic (residents/min)	Total cycle time from simulation	Total cycle time from clinic mathematical model and formulas	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%

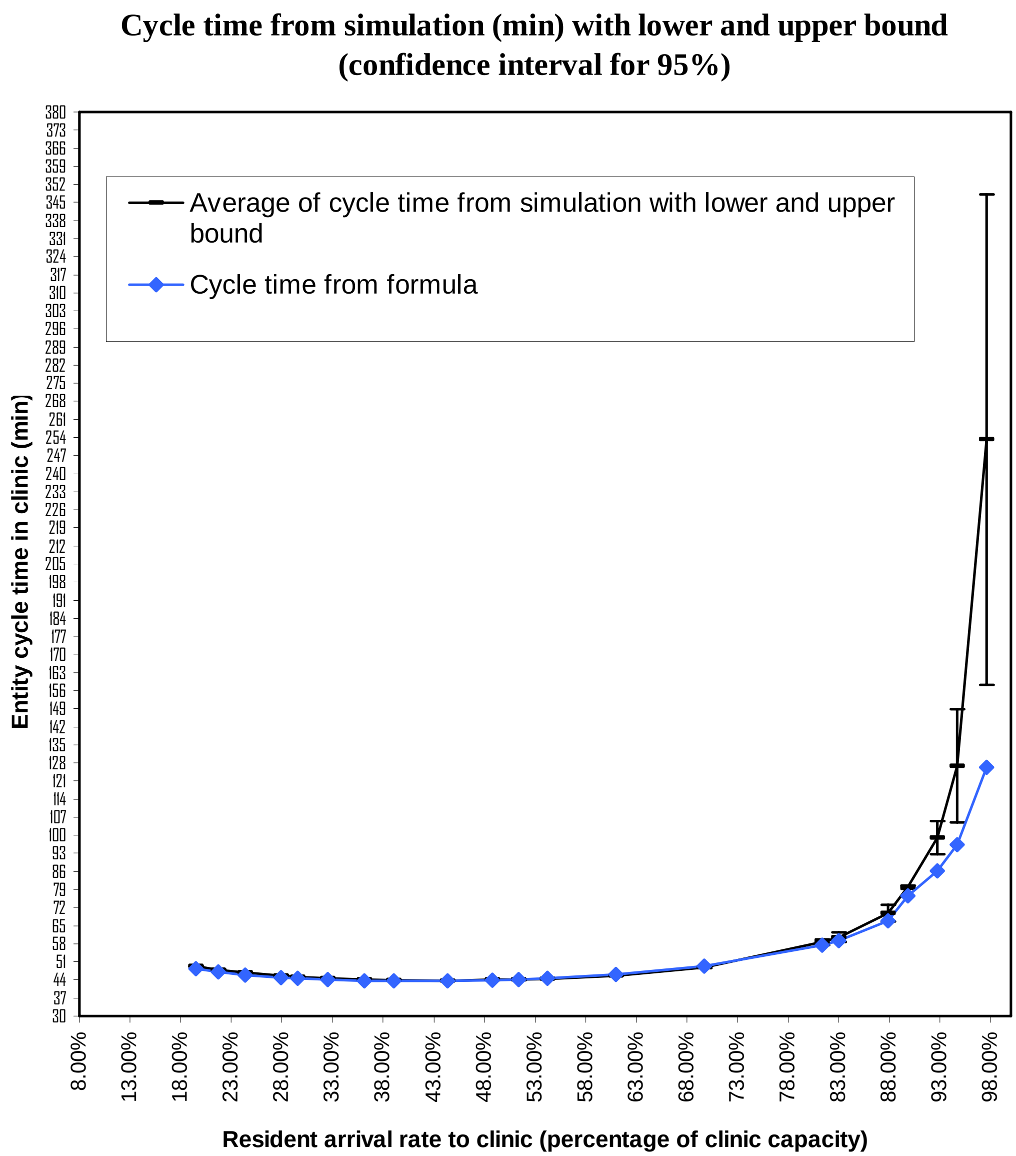

Figure 49. Comparison of total cycle time for mass smallpox vaccination clinic

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).

5.3.2 Second experiment

In this experiment, we have 7 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.

In this experiment consisting of 2 tests: Test 2-1 and Test 2-2, residents arrived by bus to the clinic. Each bus brought 20 residents with variability. Bus interarrival times were exponentially distributed and all of the service process distributions had gamma distributions to allow different process time SCV. The only difference between these 2 tests is that in Test 2-1, the SCV of bus size is 0.05 (almost zero), while in Test 2-2, the SCV of bus size is 0.2.

Table 98 describes each of the seven stations in the clinic. Table 99 shows the routing probabilities from one station to another. Table 100 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 8 scenarios in which arrival rate starts from nearly its maximum rate (10.704 residents per min) and continues to the lower rates.

Table 98. Parameters for constructed clinic (Second experiment)

Workstation Number	**Number of 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 (Self service)	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 99. Routing table for the clinic in the second experiment

	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 100. Table capacity for the clinic’s stations in the second experiment

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 (Self service)	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

We should mention this point that since station number 4 is a self service workstation and we don’t have any servers, the capacity is not applicable (n/a) for self service station.

5.3.2.1 Results for Test 2-1

In Test 2-1, the SCV of arrival bus size is 0.05. Therefore, we model the bus arrival batch size distribution with 1+Poisson (19). Table 101 shows the average total cycle time in terms of minute for each entity in the clinic from simulation results (Test 2-1) and from our clinic mathematical models as well as the percentage of error between them.

Moreover, Figure 50 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 101. 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 clinic mathematical model and formulas	Percentage error %
1	10.00	16.65	16.78	0.79%
2	9.09	11.96	12.83	7.30%
3	8.00	10.26	9.80	4.48%
4	6.67	9.34	9.12	2.32%
5	5.00	8.76	8.67	1.07%
6	3.33	8.50	8.42	0.95%
7	2.50	8.41	8.34	0.84%
8	2.00	8.37	8.30	0.91%

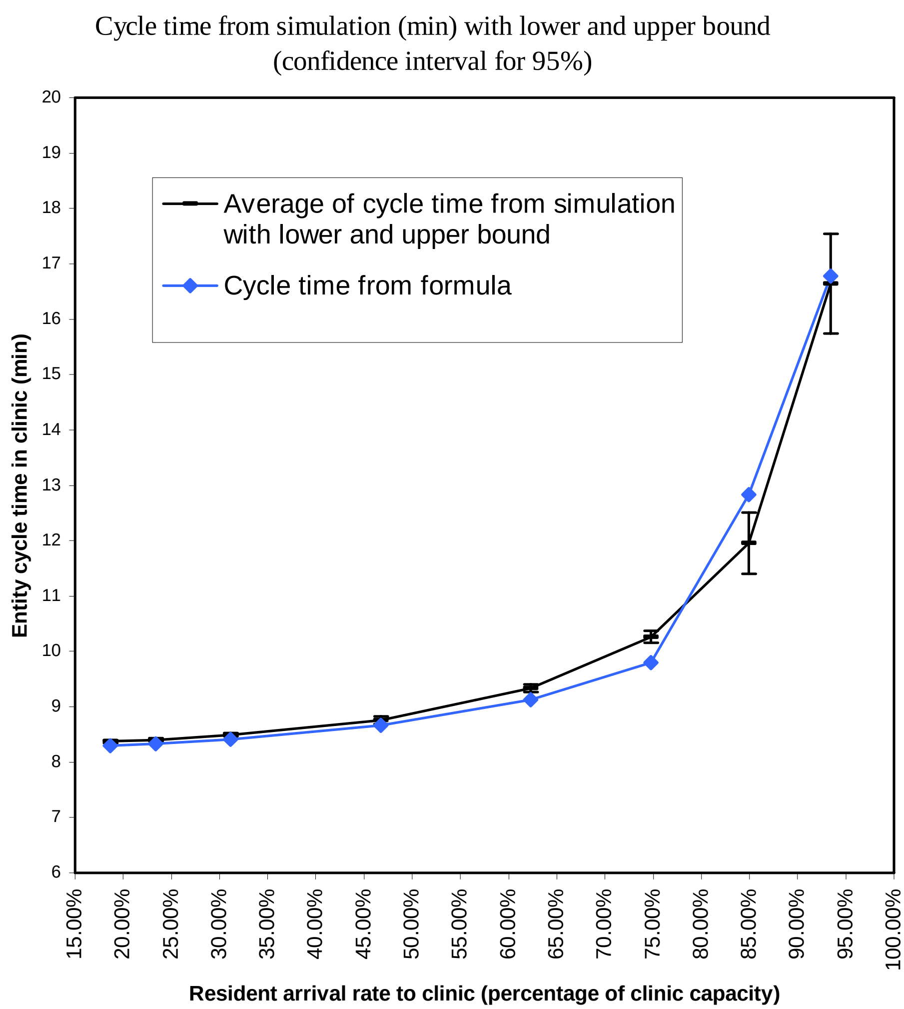

Figure 50. Comparison of total cycle time for the clinic (Test 2-1)

5.3.2.2 Results for Test 2-2

In Test 2-2 the SCV of arrival bus size is 0.2. Therefore, we model the bus arrival batch size distribution with 100-Poisson (80).

Table 102 shows the average total cycle time in terms of minute for each entity in the clinic from simulation results (Test 2-2) and from our clinic mathematical models as well as the percentage of error between them.

Moreover, Figure 51 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 102. 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 clinic mathematical model and formulas**	**Percentage error %**
1	10.00	18.80	18.11	3.65%
2	9.09	12.72	12.93	1.62%
3	8.00	10.71	9.86	7.89%
4	6.67	9.60	9.16	4.56%
5	5.00	8.95	8.68	3.00%
6	3.33	8.62	8.42	2.27%
7	2.50	8.51	8.34	2.04%
8	2.00	8.47	8.30	2.02%

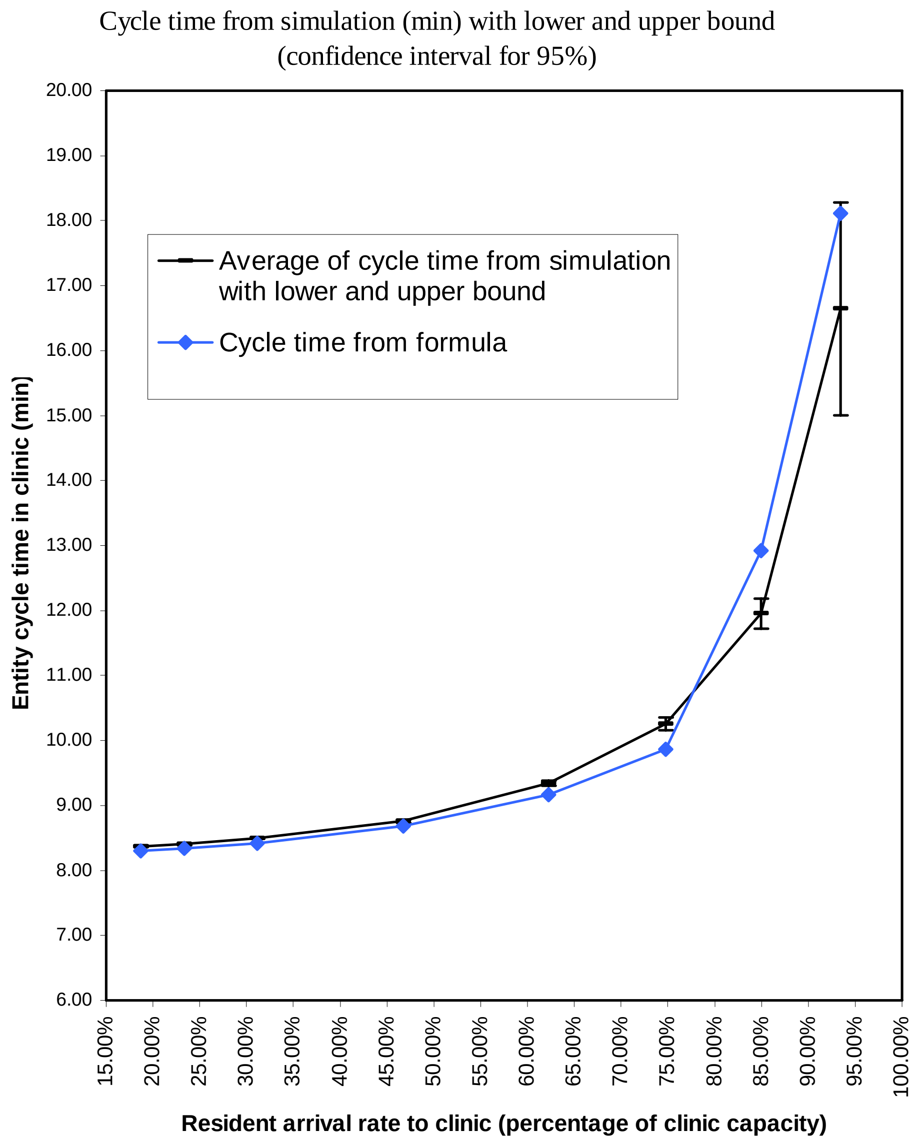

Figure 51. Comparison of total cycle time for the clinic (Test 2-2)

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.

5.3.3 Third experiment

In this experiment, we have 7 workstations with different specifications. The first 4 workstations have individual service process and the last three one are batch process workstations. Moreover, like experiment number 2, station 4 is a self service station with individual arrival.

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.

In this experiment consisting of 2 tests: Test 3-1 and Test 3-2, residents arrived by bus to the clinic. Each bus brought 50 residents with the variability of 0.02. Bus interarrival times were exponentially distributed and all of the service process distributions had gamma distributions to be able us to make different process time SCV. The only difference between these 2 tests is having 2 different process time variances for self service station. Therefore, in Test 3-1, the process time SCV of self service station is 0.56, while in Test 3-2, the process time SCV of self service station is 1.

The routing probabilities from one station to another in this experiment are the same as experiment number 2 (See Table 99). Table 103 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 number 6 is the bottleneck station, and the clinic capacity is 10.736 residents per minute.

In this experiment, for each test we have 8 scenarios in which arrival rate starts from nearly its maximum rate (10.736 residents per min) and continues to the lower rates.

Table 103. Table capacity for the clinic's stations in the third experiment

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 (Self service)	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

Since station number 4 is a self service workstation and we don't have any servers, the capacity is not applicable (n/a) for self service station.

Additionally, because the SCV of arrival bus size is 0.02 in this experiment, we model the bus arrival batch size distribution with 1+Poisson (49).

5.3.3.1 Results for Test 3-1

In Test 3-1, process time SCV of self service station is 0.56. Table 104 describes each of the seven stations in the clinic for Test 3-1. Table 105 shows the average total cycle time (in minutes) for each entity in the clinic from simulation results (Test 3-1) and from our clinic mathematical models as well as the percentage of error between them.

Moreover, Figure 52 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 104. Parameters for constructed clinic (Test 3-1)

Workstation #	Number of 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 (Self service)	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 105. 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 clinic mathematical model and formulas	Percentage error %
1	10.42	113.12	54.45	51.87%
2	10.00	61.85	47.80	22.71%
3	9.62	49.74	45.22	9.08%
4	9.09	44.30	41.98	5.22%
5	8.70	42.14	39.62	5.96%
6	7.04	39.07	37.88	3.05%
7	5.26	39.47	38.52	2.40%
8	4.17	41.25	40.24	2.44%

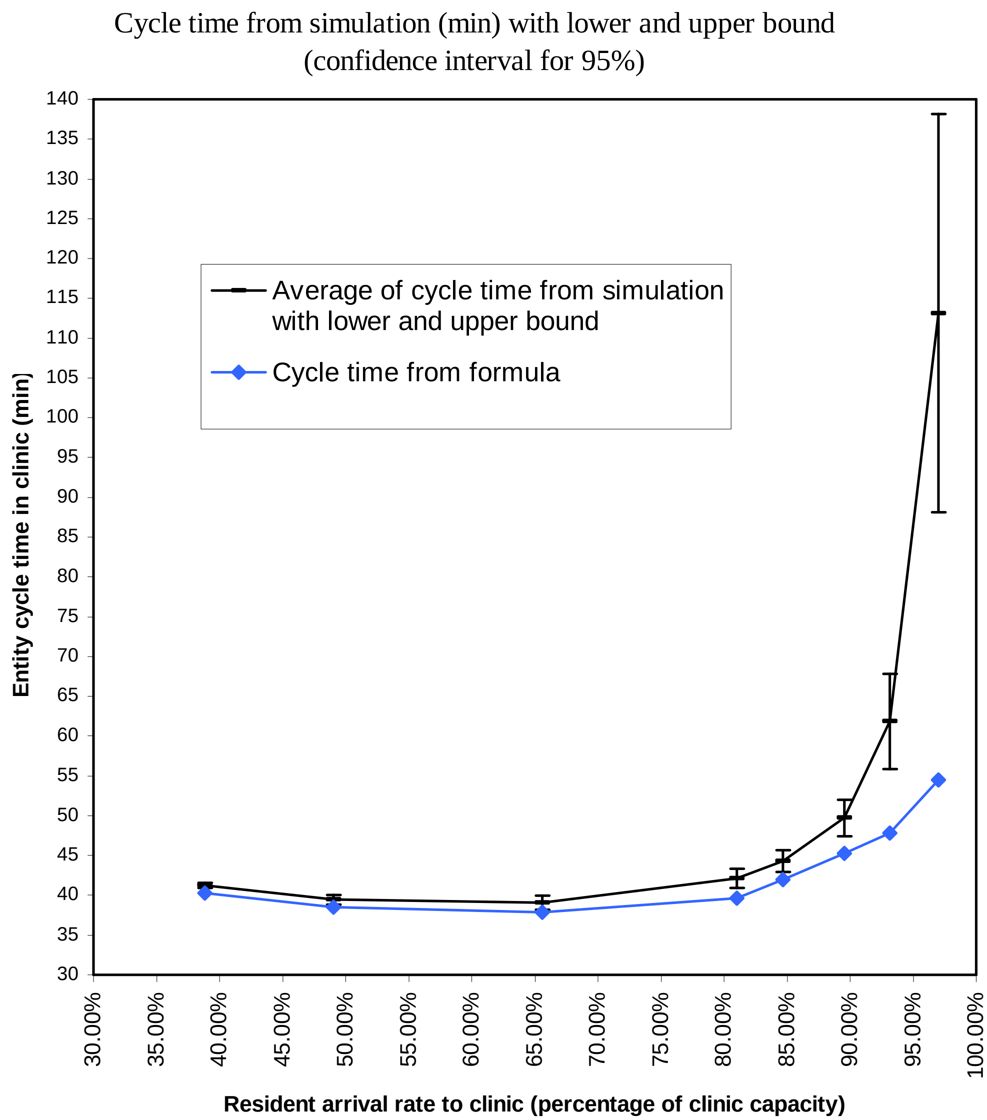

Figure 52. Comparison of total cycle time for the clinic (Test 3-1)

5.3.3.2 Results for Test 3-2

In Test 3-2, process time SCV of self service station is 1. Table 106 describes each of the seven stations in the clinic for Test 3-2. Table 107 shows the average total cycle time in terms of minute for each entity in the clinic from simulation results (Test 3-2) and from our clinic mathematical models as well as the percentage of error between them.

Moreover, Figure 53 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 106. Table Parameters for constructed clinic (Test 3-2)

Workstation Number	Number of 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 (Self service)	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 107. 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 clinic mathematical model and formulas	Percentage error %
1	10.42	154.98	54.36	64.92%
2	10.00	58.45	47.74	18.32%
3	9.62	49.37	45.17	8.50%
4	9.09	44.23	41.94	5.17%
5	8.70	42.21	39.60	6.19%
6	7.04	38.93	37.86	2.73%
7	5.26	39.39	38.50	2.25%
8	4.17	41.24	40.23	2.45%

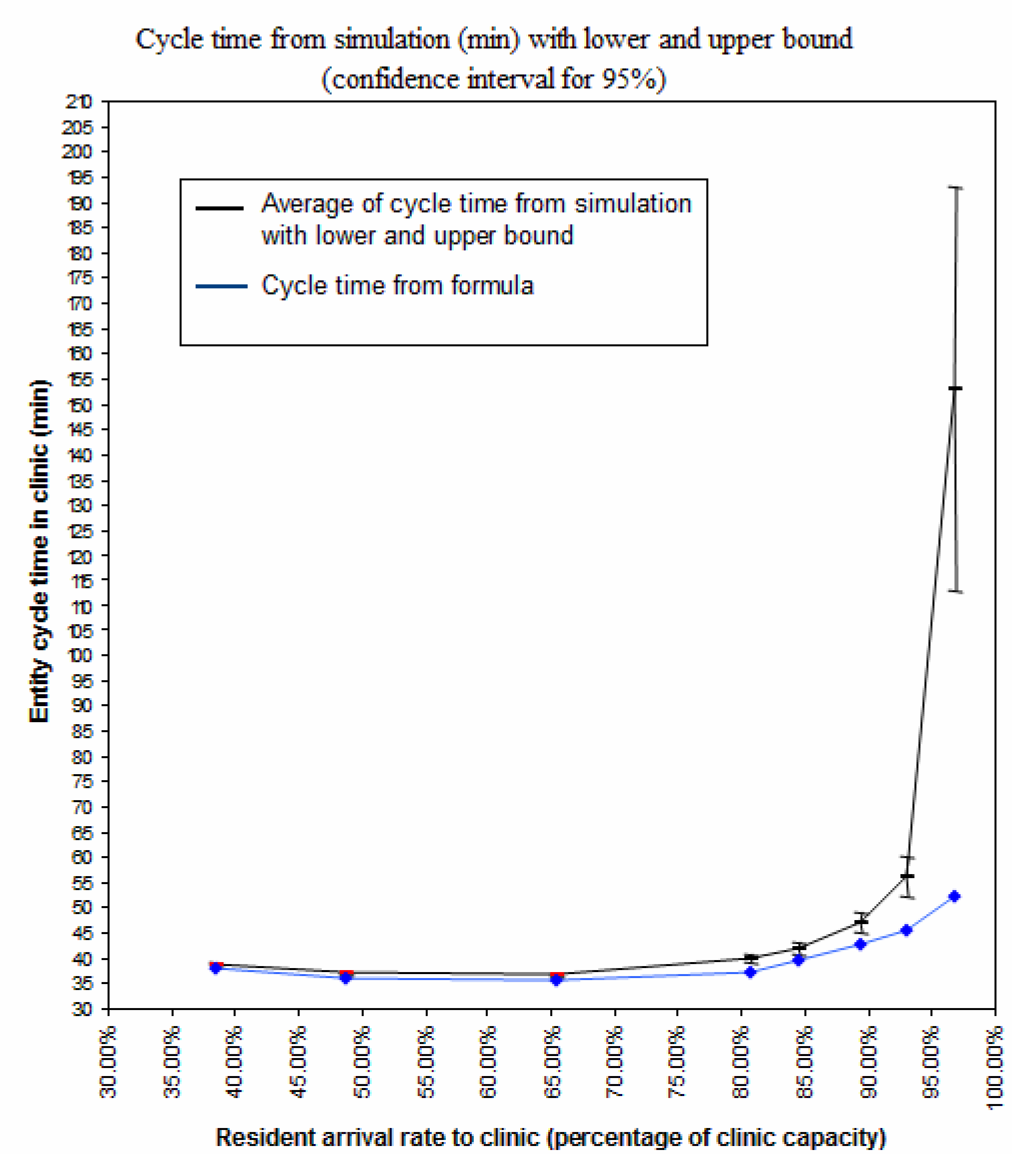

Figure 53. Comparison of total cycle time for the clinic (Test 3-2)

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 the station 6, which has the highest utilization (since it is the clinic bottleneck).

5.3.4 Fourth experiment

In this experiment, we have 6 workstations with different specifications in which all stations other than station 1 and 6 are batch process stations. Moreover, the station 6 is a self service station with individual arrival.

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.

In this experiment consisting of 2 tests: Test 4-1 and Test 4-2, residents arrived by bus to the clinic. Each bus brought 40 residents with the variability. Bus interarrival times were exponentially distributed and all of the service process distributions had gamma distributions to be able us to make different process time SCV. The only difference between these 2 tests is that in Test 4-1, the SCV of bus size is 0.024 (almost zero), while in Test 4-2, the SCV of bus size is 0.25.

Table 108 describes each of the seven stations in the clinic. Table 109 shows the routing probabilities from one station to another. Table 110 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. In this experiment, for each test we have 8 scenarios in which arrival rate starts from nearly its maximum rate (4.85 residents per min) and continues to the lower rates.

Table 108. Parameters for constructed clinic (Fourth experiment)

Workstation Number	Number of 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 (Self service)	n/a	6	0.44	GAMM (3,2)	1

Table 109. Routing table for the clinic in the fourth experiment

	To					
From	**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 110. Table capacity for the clinic's stations in the fourth experiment

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 (Self service)	n/a	0.738	n/a

Since station number 6 is a self service workstation and we don't have any servers, the capacity is not applicable (n/a) for self service station.

5.3.4.1 Results for Test 4-1

In Test 4-1, the SCV of arrival bus size is 0.024. Therefore, we model the bus arrival batch size distribution with 1+Poisson (39). Table 111 shows the average total cycle time in terms of minute for each entity in the clinic from simulation results (Test 4-1) and from our clinic mathematical models as well as the percentage of error between them.

Moreover, Figure 54 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 111. 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 clinic mathematical model and formulas**	**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%

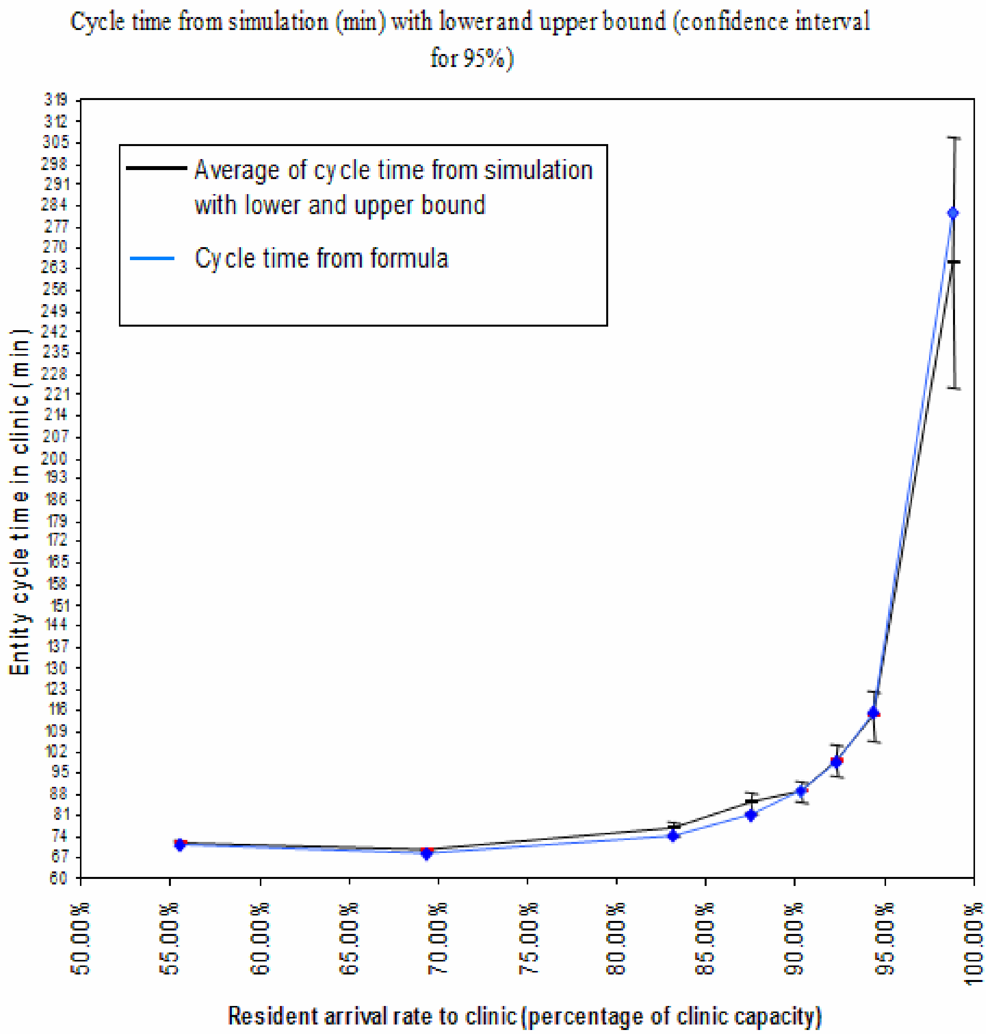

Figure 54. Comparison of total cycle time for the clinic (Test 4-1)

5.3.4.2 Results for Test 4-2

In Test 4-2 the SCV of arrival bus size is 0.25. Therefore, we model the bus arrival batch size distribution with 440-Poisson (400).

Table 112 shows the average total cycle time in terms of minute for each entity in the clinic from simulation results (Test 4-2) and from our clinic mathematical models as well as the percentage of error between them.

Moreover, Figure 55 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.

The plot for the simulation results includes error bars showing the 95% confidence interval on each estimate.

Table 112. 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 clinic mathematical model and formulas**	**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%

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

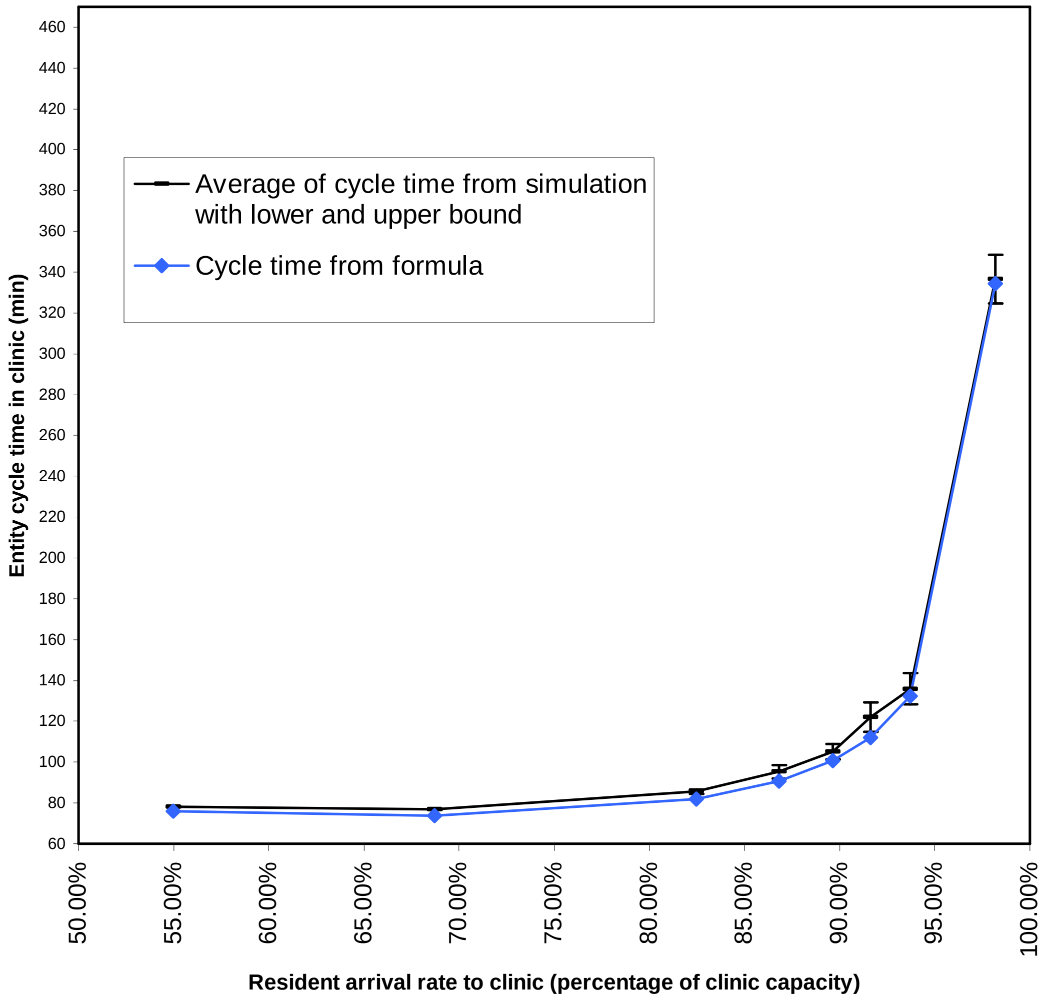

Figure 55. Comparison of total cycle time for the clinic (Test 4-2)

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.

5.4 Summary of the chapter

Since the whole objective of this thesis is to be able to construct a thorough model of a mass dispensing and vaccination clinic, to summarize this chapter, we only will review and analyze the results of Experiments 1 to 4 and our formulas from the analytical model for these experiments.

Before briefly discussing each experiment, note that batch processing and batch moves (transfer) make estimating the batch size more difficult. Therefore, the relative error between the simulation results and the formulas for performance measures such as waiting time or cycle time increases in cases with high arrival rate or when the utilization of the bottleneck station is very high.

In Experiment 1, which had no self service and batch size variability for arriving buses, we had 21 scenarios in which arrival rates varied from 20% to 97.5% of clinic capacity. From Table 97, we found out that relative error between the simulation results and the approximation for total cycle time was good except for scenarios with high arrival rate, because of the batch processing which was mentioned as an important point at the beginning of this section.

The other reason for a large relative error for scenarios with a high arrival rate is that the bottleneck station is the last station. Errors in the performance measures of the first five stations all affect simultaneously the cycle time estimate of the sixth station, which is the bottleneck station.

In Experiment 2, station 4 was a self service station. In addition, the clinic had batch size variability at the first station. In this experiment, we had 8 scenarios for each test in which arrival rates varied from 20% to 94.5% of clinic capacity. In Experiment 2, the clinic had only individual processing, and there was no batch processing with batch moves. From Tables 101 and 102, we see that the relative error between the simulation results and the approximations for total cycle time was good for all of the scenarios within each test, although the resident arrival rate to the clinic was very high and close to the clinic maximum capacity for some scenarios.

Additionally, from Figures 50 and 51, we see that the confidence intervals include the estimates from the cycle time approximation.

Similarly, for Experiment 3, station 4 was a self service station, and the first station had arrival batch size variability. In this experiment, we had 8 scenarios for each test in which arrival rates varied from 20% to 94.5% of clinic capacity. In Experiment 3, the clinic had stations with individual and batch processing. From Table 105 and 107, we see that the relative error between the simulation results and the approximation for total cycle time was good except for the cases in which the arrival rate was very high and close to the clinic maximum capacity.

Additionally, we see in Figures 52 and 53 that the confidence interval for high arrival rates didn’t include the cycle time estimate from formulas. In Experiment 3, the bottleneck station has arrivals from the first four stations. As we saw in Experiment 1, errors when estimating the performance measures of the first four stations all affect simultaneously the cycle time estimates of the sixth station. Thus, they increase the relative error between the simulation results and the approximations.

Experiment 4 was similar to Experiment 3. Because the bottleneck station in Experiment 4 is the first station, without any upstream stations, the batch processing and batch moves don't cause any errors in estimating the cycle time of the first station. From Tables 111 and 112, we see that, although we have batch processing and batch move throughout the clinic in Experiment 4, the relative error between the simulation results and the cycle time estimates is small even for the cases with high arrival rate, when the bottleneck utilization is close to 1.

Figure 54 and 55 shows that the fact that since the first station is bottleneck station, confidence intervals include the estimates from the formulas for cycle time even for the high arrival rate cases.

Chapter 6: Conclusion

The overall goal of this research has been to provide public health emergency preparedness and response planners with mathematical models that can help them to estimate the important performance measures such as total waiting or cycle time in the mass dispensing and vaccination clinic. With this information, planners become better informed when they have to make decisions regarding staff placement, POD layout, and other relevant concerns.

The proposed models in Chapter 5 correspond to clinics that consist of different kinds of stations with any kind of arrival process (individual or batch) or service process (individual, batch or self service). The recommended model in this thesis can also satisfy cases that we have batch size variability.

6.1 Conclusion

Although this research was motivated by a specific application in emergency planning area, it should be applicable also to the design and analysis of manufacturing systems with similar specifications to our clinic models.

Briefly, what we have done successfully in this thesis has been consistently modeling mass dispensing and vaccination clinics that can consist of diverse workstations with any kind of arrival processes or any type of service processes (such as individual processing, batch processing or self service) as an integrated and complete queueing network.

In order to synthesize a variety of existing and new proposed models into a systematic approach for the type of queueing network explained in this thesis, we have made significant and innovative contributions in this thesis. For example, one of the studied models in this thesis has been queueing systems with both batch arrivals with a positive SCV of a batch size and batch service process whose batch size is bigger than the arrival batch. Moreover, including self service workstations in a mass dispensing and vaccination clinic model as well as studying their behavior has been another unique and interesting section of this thesis.

However, in spite of having approximations including novel contributions and having been one of the first recommended mathematical models integrating all possible types of workstations into a single model, our estimates have some limitations that have to be mentioned at this point.

First, as we saw in our simulation results in Chapter 5, for the scenarios whose bottleneck station utilizations are close to 1 (more than 90%), the percentage error between the simulation results and the numerical results from our proposed formulas was relatively large.

Additionally, our proposed formulas to estimate CTq for all types of stations had more errors compared to the simulation results in which the average interarrival time SCV has been large (more than 4).

Finally, since in this thesis, we studied the behavior of workstations with batch arrival and batch processing in which the average arriving batch size is always less than the batch processing size, we cannot model cases in which the average arrival batch size to a workstation is greater than the batch processing size.

6.2 Future work for research

Several parts of this work reveal opportunities for further research to be performed in the future.

One highly critical concern that needs further research is creating some new estimates for waiting time in queue for all of the 6 different types of workstations, proposed in Chapter 5, in which the arrival process has high aggregated interarrival time SCV (more than 4) and when the utilization is relatively high (bigger than 90%).

Additionally, cases where the average arrival batch size to a workstation is more than the batch processing size need to be studied. It may be possible to model these cases similar to what we had in stations with batch arrival and individual process service. In this way, it is necessary to study again the trend of the WIBT and *CTq* as done in Chapter 4 by running simulation models with a range of parameter values, studying the results, and then extracting trends to get insight into relationships, motivate the models, and estimate their parameters.

The other interesting issue to investigate is to have a new model that can support the possibility of having different types of customer classes such as families in a mass dispensing and vaccination clinic. Since in a real clinic, each family may prefer to travel and spend times in all stations through the clinic together, families with different sizes can be considered as classes that can have their own specifications such as different process time and process time SCV in each station depending on the size of the families and other criteria.

In the meanwhile, it will be useful to take the analytical model from Treadwell (2006) and compare it with the results from the new model in this thesis. This comparison will show how the new model is more exact and complete and can include the cases that Treadwell (2005) cannot model.

Finally, as we mentioned in this thesis several times, we considered mass dispensing and vaccination clinics as open queueing networks, so we were able to adopt the parametric decomposition approach as a tool to study and analyze the behavior of the clinic. However, another possible approach by which one can investigate mass dispensing and vaccination clinic behavior is the diffusion approach, although it is more appropriate for closed queueing networks. Thus, diffusion approach can be another area for researchers to construct a new model of clinics with new estimates based on the assumptions and facts existing in this approach.

Bibliography

Albin, S. L. (1981), “Approximating queues with superposition arrival processes,” Phd dissertation Department of IE and OR, Columbia University, New York.

Albin, S. L. (1984), “Approximating a point process by a renewal process. II. Superposition arrival processes to queues,” Operations Research, 32, 1133-1162.

Albin. S. L., and Kai. S. R. (1986), “Approximation for a departure process of a queue in a network,” Naval Research Logistics Quarterly, 33, 129-143.

Agency for Healthcare Research and Quality (2004), “Community-based mass prophylaxis: a planning guide for public health preparedness,” AHRQ, Rockville, Maryland.

Bitran, G., and Tirupati, D. (1989), “Approximations for product departure from a single-server station with batch processing in multi product queues,” Management science, Vol. 35, No. 7.

Burke, P. J. (1958), “The output process of a stationary M/M/s queueing system,” Ann. Math Statist, 39, 1144-1152.

Buzacott, J. A., and Shanthikumar, J. G., (1993), Stochastic Models of Manufacturing Systems. Prentice-Hall, Inc., New Jersey

Chen, H., and Mandelbaum, A. (1991), “Stochastic Discrete Flow Networks: Diffusion Approximations and Bottlenecks, ” Annals of Probability, Vol. 19, No. 4, 1463-1519.

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

Denning, PJ., and Buzen, J. P. (1978), “The operational analysis of queueing network models,” Computing Surveys, 10, 225-261.

Fowler, J. W., Phojanamongkolkij, N., and Cochran, J. K. (2002), “Optimal batching in a wafer fabrication facility using a multi-product G/G/c model with batch service process,” International Journal of Production Research, Vol. 40, No. 2, 275-292.

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

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

Kaplan, E.H., Craft, D.L., and 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.

Kollerstrom, L. (1974), “Heavy Traffic Theory for Queues with Several Servers,” Journal of Applied Probability, Vol. 11, 544-552

Kraemer, W., and M. Langenbach-Belz, (1976), "Approximate formula for the delay in the queueing system GI/G/1," Congressbook, Eighth Int. Teletraffic congress, Melbourne, 235- 1/8

Kuehn, P. J. (1976), "Analysis of Complex Queueing Networks by Decomposition, " 8th ITC, Melbourne, 236.1-236.8.

Kuehn, P. J. (1979), "Approximate Analysis of General Networks by Decomposition," IEEE Transaction. Comm., COM-27, I1, 13-126.

Marshall, K.T. (1968), "Some inequalities in queueing, " Operations Research, 16, 651-665.

Mirasol , N. M. (1963), "The Output of an M/G/∞ Queueing system in Poisson, " Operations Research, 11, 282-284

Pilehvar, A., and Herrmann, J. W. (2006), "Constructing Queueing Network Approximations for Mass Dispensing and Vaccination Clinics, " Technical Report 2006-17, Institute for Systems Research, University of Maryland, College Park.

Pilehvar, A., Mark A. Treadwell, and Jeffrey W. Herrmann (2006), "Queueing network approximations for mass dispensing and vaccination clinics," submitted to IIE Transactions.

Reiman, M., (1984), "Open queueing networks in heavy traffic," Mathematics of Operations Research, Vol. 9, 441-458.

Reiser, M., and Lavenberg, S. S. (1980), “Mean value analysis of closed multichain queueing networks,” JACM, 27, 313-322.

Reiserm, A., and Kobayash, H. (1974), “Accuracy of Diffusion Approximation for Some Queueing Systems,” IBM J. Res. Develop., I8, 110-124.

Sakasegawa, H. (1977), “An approximate formula Lq=$\alpha\beta\rho/(1-\rho)$,” Annals of the Institute of Statistical Mathematics, 29, 67-75.

Segal, M., and Whitt, W. (1989), “A queueing network analyzer for manufacturing. Teletraffic Science for New Cost-Effective Systems, Networks and Services,” ITC-12, M. Bonatti (ed.), Elsevier, North-Holland, Amsterdam, 1146-1152.

Sevcik, K. C., and Levy, A. L., Tripathi, S. K., and Zahorjan, J. L. (1977), “Improving approximations of aggregated queueing network systems,” Computer Performance (eds. K Chandy and M. Reiser), North-Holland, 1-22.

Shanthikumadr, JG., and Buzacott, J. A. (1981), “Open queueing network models of dynamic jobshops,” International Journal of Production Research., 19(5), 255-266.

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. (1982), “Approximating a point process by a renewal process: Two basic methods,” Operations Research, 30, 125 -147.

Whitt, W. (1983a), “The queueing network analyzer,” Bell Systems Technical Journal, 62, 2779-2843.

Whitt, W. (1983b), “Performance of the queueing network analyzer,” Bell Systems Technical Journal, 63, 1911-1979.

Whitt, W. (1984), “Approximations for departure processes and queues in series,” Naval Research Logistics Quarterly, 31, 499 -521