ANALYSIS OF MILITARY ENTRY CONTROL POINT QUEUEING

THESIS

MARCH 2016

Justin J. Dwyer, Major, USA

AFIT-ENS-MS-16-M-103

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-16-M-103

ANALYSIS OF MILITARY ENTRY CONTROL POINT QUEUEING

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Justin J. Dwyer, BS

Major, USA

March 2016

DISTRIBUTION STATEMENT A.

AFIT-ENS-MS-16-M-103

ANALYSIS OF MILITARY ENTRY CONTROL POINT QUEUEING

THESIS

Justin J. Dwyer, BS

Major, USA

Committee Membership:

Dr. Jeffery D. Weir
Chair

Dr. Darryl K. Ahner, PhD
Member

AFIT-ENS-MS-16-M-103

Abstract

Military Entry Control Facilities (ECFs) are unique service queues that are constrained by space, receive high peak traffic flow, and have a customer base that must receive service. Due to complexity of the interactions within the system, simulations provide input that would be impractical for quantitative experimentation. Our research examines relationships within the ECF in order to develop insights that could lead to more efficient daily operations. We focus the research on interactions that generate a queue length that would interfere with traffic flow surrounding the base. Examining the interactions between multiple arrival rates and service times as well as the layout and model of the ECF we establish criterion for Officers in Charge (OICs) to make changes within the constraints of the ECF to their operations to better serve the customers and prevent ECF traffic from interfering with the community outside the military base.

Dedication

This work is dedicated to my lovely wife and our two incredible children. They have been extremely supportive and have given me the encouragement and strength to succeed.

Acknowledgments

I would like to thank my advisor, Dr. Jeffery Weir for his direction, support, and guidance throughout my research. I also wish to thank Dr. Darryl Ahner for his feedback and contributions.

Justin J. Dwyer

Table of Contents

Page

List of Figures

Page

List of Tables

Page

ANALYSIS OF MILTARY ENTRY CONTROL POINT QUEUEING

I. Introduction

Background

The purpose and mission behind military Entry Control Facilities (ECFs) is to provide security to the instillation from unauthorized access and intercept contraband while maximizing traffic flow. The design of an ECF should maximize traffic flow without compromising security, safety or causing undue delays that may affect off-instillation public highway users or instillations operations (Department of Defense, 25 May 2005). Due to fluctuations within the processing times and the arrival rates of customers, the customers waiting for service from the ECF may exceed its capacity, resulting in traffic overflow into the surrounding community traffic. Once instillation traffic begins to overflow the capacity of the ECF, the queue is no longer strictly instillation traffic. Civilian traffic not desiring to enter the instillation may become part of the ECF queue simply because they are travelling on roadways that surround the instillation. This additional traffic causes a faster growth in queue length, which leads to more traffic interference as the queue grows.

Military Entry Control Points (ECPs) have unique characteristics that are different from other customer service queues. One of those characteristics is that every customer must receive service in order to enter the military base. This is similar to traffic tolls or amusement park entry points, as every customer on the highway or in line with a ticket needs to utilize the service queue. However, there is a choice to not utilize the toll booth and take a different, possibly longer route to work, or an amusement park customer may choose to sell off their tickets to another customer that hasn't bought a ticket yet or come back a different day in order to avoid the congestion. Customers that need to access the base for work or other services must utilize a Military ECP, there is no other alternative.

Two other factors that Military ECPs face are receiving high peak traffic hours and being constrained by space. High traffic flow during peak hours is not exclusive to Military ECPs, but combining this with the requirement that all customers must be served increases the overall queue length more than other customer service queues. Other customer service queues constrained by space have the ability to rearrange their queues to better utilize the space, expand the amount of queues, or move the whole facility to a larger, better-designed location. In addition, the majority of these customer service queues have a direct tie to profit. When most people think of customer service lines, they think about fast-food restaurants and drive-thrus, bank tellers, tollbooths, amusement parks, and other cashiers. For those queues, construction costs and other costs to change and improve the facility may be recuperated with higher customer throughput. Military ECPs do not have a tie to profit, so costs to conduct changes to the facility will not be recuperated through higher ECP throughput.

Although Military ECPs do not have a tie to profit, it does not mean that their processes should not be examined or improved. We found three ways to profit, non-financially, that would be significant to military bases and beneficial to military ECPs: reduce manpower required to work the ECP, increase customer satisfaction, and reduce the back up of vehicles (overall queue length) from the ECP interfering with traffic outside the base. Reducing manpower to work ECPs saves the base money if the base is using contracted guards. If a base utilizes active duty military members to work the

ECPs then there would be no significant cost savings, only man-hour savings. While this could be of interest, each base is able to run their ECPs utilizing different manning requirements, so studying this savings may not be useful to all bases. Increasing customer satisfaction could be a result of improvement to ECP throughput, which reduces the customer wait time or from improving the quality of service received from the ECP worker. Although ECP workers are instructed to be professional and respectful, the service provided is for the security of the base, not the customer's satisfaction. Reducing the number of vehicles waiting in line for the ECP will reduce the interference with traffic outside the base. Both customer satisfaction and overall queue length are affected by increasing the throughput of a given ECP. This research examines some of the interactions that influence the throughput and overall queue length of vehicles waiting to be processed by the ECP.

Purpose of Research

In order to prevent military base ECP traffic from interfering with the civilian population outside the base, an analysis of the interactions between the controllable factors within the ECP provides valuable insight to the ECP OICs on how to increase throughput. Our research shows the effects of changing operations strategies within the ECP to increase throughput and prevent traffic overflow into the surrounding base traffic.

The factors that we focus on are the arrival rate of the vehicles, the processing time of the ECP guards, deploying tandem servers on one lane or multiple parallel service gates.

Overview

This chapter discusses the overview of Military ECPs and how they are unique customer service queues. Chapter II outlines a review of previous studies and research that study customer service, general queueing theory, and queueing theory specifically related to traffic. Chapter III concentrates on the methodology used to analytically solve and build an appropriate simulations model to study the interactions that occur at Military ECPs. Chapter IV presents the results of simulations that could provide input to the OICs running each ECP to improve their throughput. Chapter V discusses the results of the simulations models, limitations of the current model, and recommendations for future research.

II. Literature Review

Overview

This chapter begins with the overview of queueing theory and the queue model system design. Next, we discuss the characteristics of queue models focusing on the important processes that assist in describing the overall system. Looking into more detail of the system, we show the actions the customers may take during the wait process that disrupt or change the queueing process as well as show different designs of a queue. We then cover mitigation techniques used to better serve the customer and allow for more efficient processing in the queueing system. Then we focus on traffic flow literature provides historical overview on the study of traffic that restrict the flow of traffic. We also cover the Military ECP specific characteristics that may differ from normal queues. Finally, we compare analytical and simulation techniques used to solve and provide insight to service queue issues.

Queueing Theory

In a simple explanation, queueing theory is considered the mathematical study of waiting lines with respect to length and time. A queueing theory model is constructed in order to predict queue length, service times, and waiting times (Allen, 2014). A queueing theory model is normally defined using Kendall's notation, which is a triplet, $A/S/c$ that consist of a set of letters describing the overall model. The model notation has expanded since Kendall's research to six descriptors for the model, $A/S/c/K/N/D$ (Banks et al., 2000) (Allen, 2014). Each of the letters represents a different characteristic of the queueing model:

- *A*: The arrival process (or distribution) of the customers to the systems.
- S: The distribution of time of the service of the customer in the system.
- *c*: The number of separate servers or service channels within the system
- *K*: The capacity of customer the system can hold.
- *N*: The size of population where the customers come.
- *D*: The service discipline within the queue system.

In order to shorten notation for most queue model descriptions, when the three last parameters (*K*/*N*/*D*) are not utilized in the model description, it is assumed that $K = \infty$, $N = \infty$, and *D* = first in, first out (FIFO) (Pinto, 2011).

Queue Model System Design

The queue system design has many aspects to consider when looking at the overall system processing. The simplest design to consider is a single-channel, single-phase queueing system commonly seen at older fast food drive-thru facilities or a single-family dentist office. Adding a second (or multiple) service facility in tandem would change the system to a single-channel, multiphase system found in dual-window fast food drive-through facilities (Heizer & Render, 2010).

The previous systems focused on single servers in one or multiple phases. Adding a second (or multiple) server to a one-phase system is changed to a multichannel, single-phase system that is found at most banks, barbershops, or post office service facilities. Finally, adding a second (or multiple) server to any or all of the service phases is transformed to a multichannel, multiphase system such as college registration systems or a military recruits physical in which recruits need blood draw, eye exam, speak with a psychiatrist, and receive a physical exam from a medical doctor that has multiple technicians or medical providers at each phase (Heizer & Render, 2010). Figure 1 illustrates the different queue system designs previously discussed.

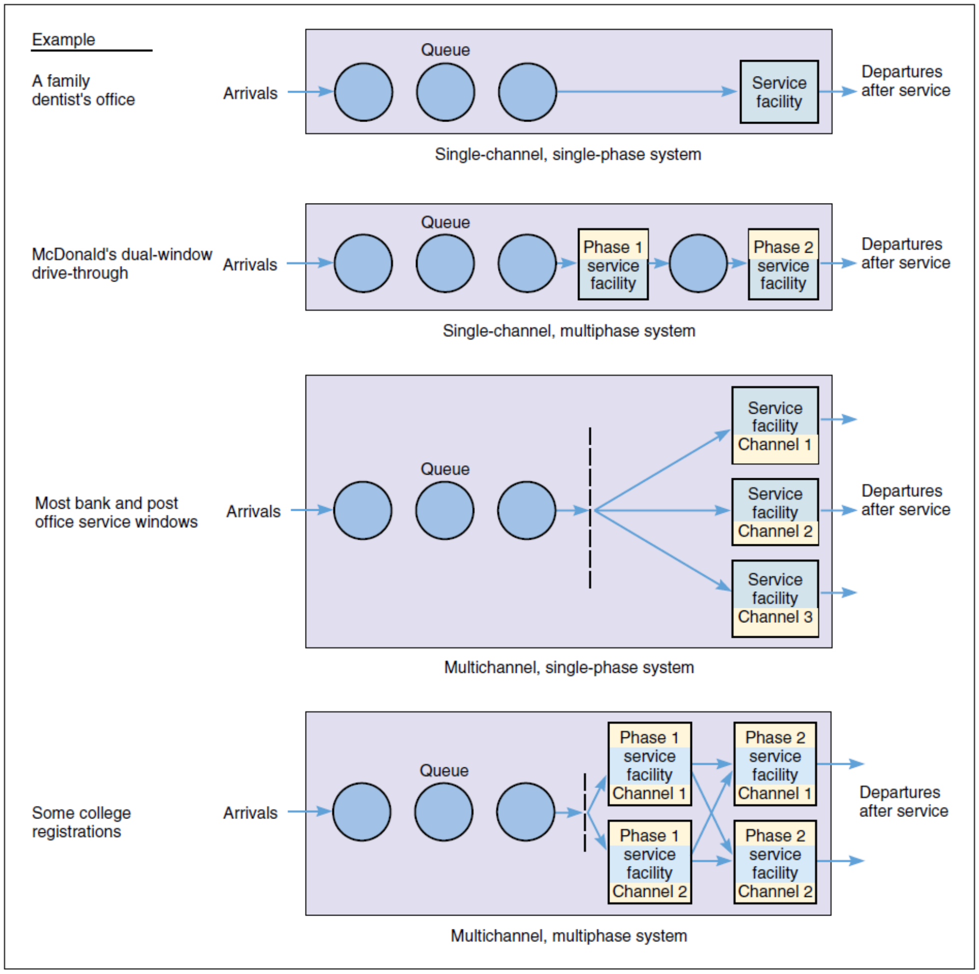


Figure 1: Queue System Design

Characteristics of a Queueing Model

Arrival Process

The arrival process for a queueing model is normally characterized in terms of the interarrival times between consecutive customers. Interarrival times occur on set schedule, occur at constant time intervals, or at random times intervals. When the times are random, the interarrival time is normally characterized with a probability distribution (Banks et al., 2000).

The Poisson arrival process, denoted as *M* in Kendall's Notation, is typically used to model a large population from which customers make independent decisions about when to arrive for a service. The process has been successfully employed to model the arrival of customers to banks, restaurants, and telephone calls to call centers (Banks et al., 2000). Two other common arrival processes used in Kendall's notation include *G*, which represents a general distribution with known mean and variance, and *D*, which represents a deterministic or constant arrival rate (Render, Stair, & Hanna, 2012).

Service Time Distribution

Services time to complete a service for a customer that enters the system may be either constant or random. If the service time for a customer is considered constant or deterministic, it is represented by *D* in Kendall's notation. Constant service time is typically found in machine-performed services such as automatic carwashes or rollercoasters. Regularly most queue systems service times are randomly distributed, and in many cases, the assumed random service time is described by the negative exponential probability distribution, represented by *M* in Kendall's notion. (Heizer & Render, 2010).

Although the exponential distribution is most commonly used, two other distributions may be more valid: normal and Erlang distributions. The Erlang distribution is common when a process has a series of stations that must be passed through before the next customer may enter the process (Banks et al., 2000) while normally distributed service times are found in automobile repair shops (Render, Stair, & Hanna, 2012).

Queue Discipline

Many queue models use a queue discipline known as the FIFO rule. These systems operate so that the first customer in line received the first service (Heizer & Render, 2010). Examples of the FIFO system include banks tellers, tollbooths, and super markets. Although customers may not view the overall check out system at a super market that has multiple checkouts each with their own individual queue as a FIFO process, each checkout operates under the discipline of FIFO (Heizer & Render, 2010). A hospital emergency room is an example of a system that primarily operates on a FIFO discipline but also has separate priorities that preempt the queue due to severity of injury or illness (Connelly & Bair, 2004). However, upon closer inspection, each sub-category of severity operates under the FIFO discipline.

Two alternate queue disciplines that are considered are last-in, first-out (LIFO) and priority scheduling. LIFO is common in inventory management when shelf life is a negligible factor and it is easiest to grab the last item in stock, which is typically the first item on the shelf or in line. Priority scheduling is common in computer programs and server bandwidth when one system is more significant than the others. An example is a company wants to prioritize the payroll computer once paychecks are due to employees (Heizer & Render, 2010).

Waiting-Line Characteristics

Queue Constraints

As customers wait in line or in a queue for a service they are normally staged in a facility or line that has limited area or capacity that would prevent a queue from going beyond a certain limit. Exceeding the capacity of the queue area causes congestion and disruption beyond the designated queue area (Buckley & Yagar, 1974) or it causes dissatisfied customers within the queueing system (Larson, Cahn, & Shell, 1993). In some situations, a system exceeds the designated queue capacity giving the idea of an infinite queue. Even if a system has a finite queue, most models follow the assumption of an infinite queue and then the probability of the infinite queue going beyond the finite queue length is calculated to find the number of customers that would not have been served (Huebner, 1998), (Ishizaki & Takine, 1999).

Customer Actions

When a customer arrives to a queue and is dissatisfied with the length of the queue they have the option of choosing not to enter the system or queue which is considered balking (Ancker Jr. & Gafarian, 1963), (Heizer & Render, 2010), (Rue & Rosenshine, 1981). There is also the option for a customer to enter the queue and wait for a period of time in the queue but exit the system before the service is conducted which is considered reneging (Ancker Jr. & Gafarian, 1963), (Heizer & Render, 2010). These actions are not considered in our study, as all customers must receive service in order to enter the base.

A concept that relies less on the customer’s satisfaction with the queue, but relies more on the actions that occur with the service is called either a re-service or a second (*n*th) service for the system (Madan, 2000), (Madan, Al-Nasser, & Al-Masri, 2004). The re-servicing of a customer may be needed if the service failed or was not satisfactory. For Military ECPs, there is not a re-service option; either the identity of the driver and passengers is verified before the vehicle enters the base or it’s not. In order to maintain security for the base, the service must be provided at the ECP correctly the first time. There are occasions when the customer does not provide proper identification at the gate. In this instance, the driver is still processed through the queue but is escorted out an exit of the base to retrieve proper identification and re-enter the queue. This would not be considered a re-service or an *n*th service because the customer does not directly re-enter the queue or move to a different queue to be service again. The customer re-enters the queue at a later time, performing as a new customer.

Queue Characteristics

Bulk-Service Queues

Not all service queues are limited to serving one customer at a time. Service queues that service more than one customer at a time are considered bulk-service queues. A frequently studied problem studied as a bulk service queue is the Fixed-Cycle Traffic-Light (FCTL) where multiple vehicles are able to be service at a time when the traffic light turns green. (Van Den Broek, Van Leeuwaarden, Adan, & Boxma, 2006). Another example of a bulk service queue is an amusement park ride where multiple guests are able to board the ride at a time.

Virtual Queues

Virtual queues are becoming more frequently as technology continues to provide us with new means of tracking our customers. Utilizing a virtual queue frees a customer from physically standing in a line to wait for a service (Dickson, Ford, & Laval, 2005). Doing this allows a customer to conduct other activities that decreases the perceived wait time for a service. Virtual queues are prevalent in amusement parks as guest have the ability to check-in to a ride and then return at given time window to go directly to the front of the line as if they were physically waiting in line for the ride (Dickson, Ford, & Laval, 2005), (Disney Parks & Travel), (Lemaster, 2015).

Queue Mitigation Techniques

There are multiple ways that the throughput of a system may be changed making simple adjustments to the overall system. The queueing system may be altered by dynamically adjusting the service rate during the process (George & Harrison, 2001). This method is utilized to keep customers from balking or reneging from the system if the queue begins to get to large. A technique used when the server is underutilized is to allow the queue to grow to ca certain point before a server is added to the system (Balachandran, 1973). This is executed based on three different policies: an *N*-policy, when the queue size reaches *N* customers; a *D*-policy, when the total work to be done reaches a value of *D*; or a *T*-policy, when a time of *T* units has passed after the end of the last busy period. (Balachandran, 1973), (Heyman, 1977).

Focusing specifically with the customer service industry there have been numerous studies that involve the perception of the wait time (Jones & Peppiatt, Managing perceptions of waiting times in service queues, 1996), (Jones & Dent, 1994), (Dickson, Ford, & Laval, 2005). In one study, it found that in service operations with a wait time of less than five minutes, the perceived wait time is up to 40 percent greater than actual wait time (Jones & Peppiatt, 1996). This difference in perceived versus actual wait time is reduced in combination with customer by occupying the customers wait time with menus, television, or readings to take their mind off the wait (Jones & Peppiatt,

1996) (Dickson, Ford, & Laval, 2005). Another technique to remove the hardship of waiting for a service is known as virtual waits. This technique allows customers to check-in to a service which has a queue, and receive a new time to come back to receive the service. This is most recognizable at amusement parks such as Walt Disney World Resort with the utilization of FastPass+ to wait for rides outside of the standard queue occupying themselves somewhere else within the park (Disney Parks & Travel) (Dickson, Ford, & Laval, 2005), but it is also found at restaurants that utilize long range buzzers or texting to recall customers to let them know their table is ready or cruise lines and all-inclusive resorts with certain long wait on-board activities (Dickson, Ford, & Laval, 2005).

Another technique utilized by Walt Disney World known as the “Magic Kingdom’s E Rides Night” used a technique known as demand shifting. Customers could purchase tickets to stay three hours after the park closed and ride the nine most popular attractions. Not only did this benefit the customers who purchased the tickets allowing them to better budget their time during the day elsewhere in the park, it also benefited the other customers in the park bring the queues on the most popular rides down during the peak hours (Dickson, Ford, & Laval, 2005).

Traffic Congestion

Vehicle traffic issues concerning either flow or congestion have been studied for many years. Traffic congestion is broken into simpler types of situations that are combined in combination to cause most traffic issues. William Vickery from Columbia University distinguished six types of congestion that included simple interaction, multiple interaction, bottleneck, triggerneck, network and control, and general density (Vickrey, 1969). While all six interactions are investigated in various studies, bottleneck and triggerneck are the main types of congestion when considering restricted traffic flow. The pure bottleneck occurs when a short segment of a route has a fixed capacity smaller relative to traffic demand than preceding or succeeding segments (Vickrey, 1969). A triggerneck situation occurs because of a bottleneck when the queue from the bottleneck interferes with traffic flow where not intended to interfere (Vickrey, 1969).

On-ramps and off-ramps are high probability locations for a triggerneck situation to cause many issues with traffic flow (Buckley & Yagar, 1974) (Wu, Jin, & Horowitz, 2008). This concern is not only for the immediate area, but for other portions of the freeway that adversely affected as the freeways overall productivity is reduced, the level of service and passenger satisfaction is reduced, and accidents, pollution, and fuel consumption are increased due to the congestion (Oviaci & May, 1974).

Some studies have looked into the effect of tollbooths on traffic flow which is a specific source of bottlenecking common in many countries around the world (Chau, Xu, & Liu, 2002), (Huang & Huang, 2002), (Wu, Jin, & Horowitz, 2008). Although traffic jams occur more frequently near tollbooths than any other part of the highway (Huang & Huang, 2002), they are needed for two main purposes: collect tolls and regulate traffic. This may seem strange at first, but in addition to increased revenue for governments, tollbooths (or road pricing) “is also considered to be one of the most efficient approaches to reducing congestion and has been investigated currently by both economists and transportation researchers” (Huang & Huang, 2002), (Yan & Lam, 1996), (Ferrari, 1995) (Small, 1992).

Military Entry Control Facility

Prior to 2001 military entry control facilities varied by instillation. Most instillations had ECFs, but they lacked the features and functionality that is mandated for the current force protection standards. Other instillations had limited or no entry control in place (Surface Deployment, 2008). The events on September 11, 2001 made it a necessity for immediate entry control. The focus of entry control was to address security that met anti-terrorism and force protection needs, but lacked the infrastructure to address traffic flow and safety for motorists as well as guards (Surface Deployment, 2008).

The traffic flow through military instillation ECFs directly depends on the number of lanes available for traffic, the number of guards working each lane, the method of identification inspection (visual or handheld device), the traffic allowed access through the specific gate, and the Force Protection Condition (FPCON) category. The number of lanes available at each ECF as well as the workforce at each ECF is determined by the instillation and is not consistent between instillations. Currently there is not a mandate on what method of identification must be used at instillation ECFs (visual or handheld). There is also no mandate of what equipment is being utilized at each base for handheld checks. Across the United States, there are 23 military installations and 16 U.S. ports that utilize the handheld device, Defense ID, produced by Intellicheck Mobilisa, Inc., which is only a fraction of military bases across the United States (Intellicheck Mobillisa, Inc., 2015).

UFC 4-022-01 classifies ECFs into four “use” classifications: primary, secondary, limited use, and pedestrian access (Department of Defense, 25 May 2005). The Surface Deployment and Distribution Command Transportation Engineering Agency

(SDDCTEA) have added three more categories to the four classifications for a total of seven classifications seen in Table 1.

Table 1: ECP Classifications (Surface Deployment, 2008)

Use Classification	Traffic Volumes	Typical Hours of Operation	Highest FPCON Operation	Typical Operation
Primary	High	24/7 - open continuously	Open thru Delta	Vehicle registration/visitor pass capacity. Regular operations, visitors with authorization. Could also be designated as truck and delivery ECF/ACP.
Secondary	High-Moderate	Regular hours, closed at times	Potentially closed at or above Charlie	Regular operations, visitors with authorization. Could also be designated as truck and delivery ECF/ACP.
Low-Volume	Low	Regular hours, closed at times	Potentially closed at or above Charlie	Regular operations, visitors with authorization. Could be located near installation housing areas. Per the Army Standard for Access Control Points the peak hour entering traffic volume is 290 vph or less.
Limited Use	Low	Open for special purpose	NA	Tactical vehicles, HAZMAT, special events, etc.
Commercial Vehicle-Only ECF	Moderate-Low	Regular hours, closed at times	Potentially closed at or above Charlie	Commercial/contractor access only. Visitors may also be processed.
Internal ECP	Low	Regular hours, closed at times	NA	Dependent upon installation mission. UFC does not apply, refer to SDDCTEA for guidance.
Pedestrian Access	NA	Varies	Potentially closed at or above Charlie	Personnel only, could be located near installation housing areas, near schools, or as part of a Primary or Secondary ECF

The classification of each ECF varies due to the inspection process that occurs for each form of traffic. Primary, Secondary, and Low-Volume ECF traffic can be viewed in the same manner due to the traffic through each ECF consisting of the similar vehicles.

The instillation FPCON level dictates the level of identification and inspection requirements at each ECF at the instillation. Currently the military operates under five main levels with one sub-level for six total FPCON levels: Normal, Alpha, Bravo, Bravo+, Charlie, and Delta. SDDCTEA Pamphlet 55-15 describes the typical processing characteristics for each ECF using Table 2.

Table 2: Force Protection Conditions (Surface Deployment, 2008)

FPCON	Application	Description	Typical Processing Characteristics
NORMAL	Applies when a general threat of possible terrorist activity exists but warrants only a routine security posture.	The baseline posture.	No direct checks; considered an open installation.
ALPHA	Applies when there is an increased general threat of possible terrorist activity against personnel or facilities, the nature and extent are unpredictable.	The measures must be capable of being maintained indefinitely.	Vehicle identification only.
BRAVO (including BRAVO+ used at some installations)	Applies when an increased or more predictable threat of terrorist activity exists.	The measures must be capable of being maintained for weeks without causing undue hardship or extreme traffic delays, affecting operational capability, or aggravating relations with local authorities.	Bravo: vehicle and driver identification, random vehicle inspections. Bravo+: All occupants identified, vehicle identification; random vehicle inspection.
CHARLIE	Applies when an incident occurs or intelligence is received indicating some form of terrorist action against personnel or facilities is likely.	Implementation of this measure for more than a short period may create hardship and affect the peacetime activities of the unit and its personnel.	Identification of vehicle and all vehicle occupants, more frequent random vehicle inspection.
DELTA	Applies in the immediate area where a terrorist attack has occurred or when intelligence has been received that terrorist action against a specific location or person is imminent. Normally, FPCON DELTA is declared as a localized warning.	Measures to be implemented in response to local warning and not intended to be sustained for lengthy periods of time.	ID checks of all vehicle occupants and complete inspections of all vehicles. Generally, only mission-essential personnel report for duty.

"In accordance with DOD O-2000.12-H [DoD Antiterrorism Handbook], the security measures employed during FPCON Bravo must be capable of being maintained for weeks without causing undue hardship, affecting operational capability, or aggravating relations with local authorities" (Department of Defense, 25 May 2005) . In order to adhere to this mandate ECFs need to be able to process vehicles into the instillation in timely manner when FPCON is BRAVO+ or below.

Analytical Solutions

The use of analytical solutions is a great tool when then system is straightforward with few processes interacting with each other. If the problem can be solved analytically, then there is no need to use simulation (Banks et al., 2000). In steady state queueing processes, Little's Law tells us that there is a strict relationship between the expected (long-term average) number of customers in the system (L), the expected time spent by a customer in the system (W), and the expect arrival rate (λ) within the system (Little, 1961) seen in Equation 1.

$$L = \lambda W$$

Equation 1: Little's Law

This method is useful in order to find a baseline of service for our problem; when one worker is servicing a queue with an open queue for traffic to enter.

Some frequently used queues such as M/M/1 and M/M/c have certain properties that we are able to utilize for some analytical solutions regarding steady state operations, stability, average number in queue, and average response time (Allen, 2014). We will go into more detail of the specific equations we utilize in Chapter III.

Analytic methods employ deductive reasoning of mathematics using limits, differential equations, and expected values to solve the model, which is useful and easier to follow because it follows a collection of mathematical equations to find a specific answer (Banks et al., 2000). This is limiting because each component of the system would require a new set of equations, consequently the number of equations used to examine a large complex system grows fast and becomes computationally difficult. Once computational results become too difficult, simulations are very useful.

Simulations

Overview

Processes that contain infinite queues are frequently amendable to be able to find exact analytical solutions while processes containing finite queues are not and may be more suitable for simulations (Huebner, 1998). Simulation are the appropriate tool to use for many reasons to include: verification of analytic solutions, animation to assist in visualizing the process or issues within the process, comparing multiple models from changes in inputs, and enables the study of and experimentation with, the internal interactions of a complex system or of subsystems (Banks et al., 2000).

Discrete and Continuous Simulation

Now that we have explored the use of simulations, we compare discrete and continuous simulations to find which method would be more useful. "A *discrete system* is one in which the state variable(s) change only at a discrete set of points in time" (Banks et al., 2000). While "a *continuous system* in one in which the state variables change continuously over time" (Banks et al., 2000).

Examples of a continuous system would be to examine the water level against a dam after a large rainfall which is continuously changing (Banks et al., 2000), finding the power produced by an engine when depressing the gas pedal, or the size of a polar ice cap as it melts into the ocean. Those examples show a response that is continuously changing throughout the simulation. An example of a discrete system to consider is a bank, which has state variables that change a discrete point: the number of customers in the bank changes only when a new customer walks into the bank or a service is complete and the customer no longer is in the system (Banks et al., 2000). Other examples that we previously discussed include: fast-food restaurant drive-thrus and traffic lights which follow the same concept of variables changing at certain times when services are complete or a new customer enters the queue. Our ECP problem aligns with the discrete system changing at set points in time.

Summary

This chapter covered the overview of the characteristics of service queues and service systems characteristics that are used to explain the overall queueing system. Then we also discussed mitigation techniques used to serve customers better allowing for processes that are more efficient. We covered the Military ECP specific characteristics that may differ from normal queues. Finally, we compared analytical and simulation techniques used to solve and provide insight to service queue issues.

III. Methodology

Chapter Overview

This chapter discusses the methodology we utilize to build the simulations to analyze military ECP operations and we compare them to analytical solutions of a similar queueing system to see if those equations are a good approximation in the future. First, we cover some of the notation and characteristics of our queueing system. Next, we discuss the assumptions that our models follow. Later, we discuss the creation of the model in SIMIO utilizing specific processes that lead to decisions made with the system. We then discuss the ranges for testing our experiment followed by the technique we use to verify the model comparing testing data to the current processing numbers utilized in the SDDC handbook. Finally, we discuss a few of the analytical solution methods from a similar queue system to compare to the simulations runs in order to verify if those equations are good approximations in the future to provide insights into the ECP operations.

Notation and Characteristics of our Queueing Systems

Throughout this chapter we use common notation when discussing our queueing system. The following notation is used:

- λ : mean arrival rate $\quad \lambda = \dfrac{1}{\mathrm{E}\left[\text{Inter-arrival time}\right]}$
- μ : mean service rate $\quad \mu = \dfrac{1}{\mathrm{E}\left[\text{Service time}\right]}$

- $\rho = \frac{\lambda}{\mu}$ utilization for a single server system which is equal to the probability that the server is busy
- c : number of servers
- $\rho_c = \frac{\rho}{c} = \frac{\lambda}{c\mu}$ utilization for a multi-server system which is equal to the probability that all servers are busy
- P_n : probability that there are n customers in the system
- L : mean number of customers in the system
- L_q : mean number of customers in the queue

Assumptions

Arrival Process

We use two primary arrival processes during our simulations. The first arrival process we utilize is a constant (deterministic) arrival rate used to verify the models are processing customers appropriately. The second arrival process used for the majority of simulations will follow the Poisson distribution. Due to the lack of empirical data and the success that the Poisson distribution has had when used in queue modeling (Allen, 2014), (Banks et al., 2000), we found this approach to be the most logical choice for our analysis.

Although utilizing the Poisson arrival rate is not ideal to model high peak traffic hours during which arrivals become time dependent, the Poisson arrival rate should be successful to model the majority of arrivals to the base.

Service Times

If service times within a queueing system are considered completely random, the exponential distribution is often used in the simulations process (Banks et al., 2000). Without empirical data, we thought to utilize the exponential distribution for service times, but there is an issue with this assumption.

Our processing time for customers includes the drive from the "on-deck" position behind the current customer being serviced to the position in front of the server (ECP worker). In order to find that distance for which the car would travel to move into the servicing position, we estimated that the car lengths and space between the cars would be similar to a parallel parking space. Each car would have to travel approximately 22 feet (Danbury City Council, 2016), (Planning Division, 2016), (Fort Worth City Council, 2016) from the "on-deck" position to the service position. In compliance with the safety standards and speed limits (Department of Defense, 25 May 2005) set forth in each ECP cars would travel this length at approximately 5 miles per hours (mph). Using this data, we approximate the travel of a vehicle over 22 feet at 5 mph to take 3 seconds. This would imply that the service time would take a minimum of 3 seconds.

If we were to use a strictly exponential distribution for the service times, from the characteristics of the exponential distribution (Ross, 2014) we know that the probability of an unknown value, X, being less than or equal to a given value, t, calculated using Equation 2 where λ is the mean of the distribution.

$$P(X \leq t) = 1 - e^{-\lambda t}$$

Equation 2: Probability for Exponential Distribution

Using Equation 2 for number of different means, λ, we see that in Table 3 that even for a large $\lambda = 20$, there is a significant amount of data, 13.9%, that would be modelled at a value less than 3 seconds which would be unlikely.

Table 3: P($X \leq 3$) for Exponential Distribution.

λ (mean)	P($X \leq 3$)
2	0.777
4	0.528
6	0.393
8	0.313
10	0.259
12	0.221
14	0.193
16	0.171
18	0.154
20	0.139

In order to correct our distribution to be more realistic but still use the exponential distribution due to the lack of empirical data we utilize an exponential distribution with mean, $\lambda - 3$, plus an additional 3 seconds which places a lower bound on the exponential distribution.

Customer Characteristics

We assume that all customers act logically when entering the queue unless otherwise stated. Acting rationally consists of the following: entering the shortest available queue, customers enter one queue and remain in that queue not disrupting other lanes of traffic, and customers do not leave a large gap between them and the vehicle in front of them.

Service Characteristics

Different FPCON levels dictate whether all of the vehicle occupant's identifications are checked or if only the driver's identification is checked (Surface Deployment, 2008). The majority of traffic serviced through ECPs during busy traffic hours are vehicles on their way to work. According to the United States Census Bureau, 88.3% of vehicles driving to work are single occupancy vehicles (McKenzie & Rapino, 2011). In order to represent the vehicles with more than one occupant, 11.7% of the vehicles, we use the right tail of the exponential service times to represent the infrequent cases when more than one occupant is in the vehicle.

ECP Operations

Contingent on the FPCON level, random vehicle inspections are conducted at different rates (Surface Deployment, 2008). To conduct vehicle inspections, a vehicle is removed from the queue to a separate (side) location to conduct the inspection process. We assume that random vehicle inspections will not interfere with traffic flow; therefore, vehicle inspections are not considered in our analysis.

Model Formulation

Service Times

As discussed previously, we consider the service time to be from when a vehicle arrives in front of the worker for inspection until when the vehicle following arrives in the same position as the previous vehicle for processing. We also established that we would be using the exponential distribution for processing time through the ECP. In order to find the mean specific service time, $\frac{1}{\mu}$, for the process we used data provided to us in SDDCTEA Pamphlet 55-15, Exhibit 2.5 which is shown in Table 4 (Department of Defense, 25 May 2005).

Table 4: ECF Processing Rates

Processing Technique	Assumed FPCON	Manual Checks		Checks Using Handheld Devices		Automated Lanes	
		Single Checks Per Lane	Tandem Checks Per Lane	Single Checks Per Lane	Tandem Checks Per Lane	Without Traffic Arms	With Traffic Arms (Up/Down For Each Vehicle)
		vphpl	vphpl	vphpl	vphpl	vphpl	vphpl
No identification	Normal	Capacity at Roadway	NA	NA	NA	NA	NA
Vehicle identification only	Alpha	800 to 1,200	NA	NA	NA	800 to 1,200	550 to 800
Vehicle and occupant identification	Bravo, Bravo+ and Charlie	300 to 450	400 to 600	275 to 375	350 to 475	400 to 450	325 to 350
Inspection of mission essential vehicles only	Delta	20 to 120	NA	20 to 120	NA	NA	NA

**Notes: vphpl = vehicles per hour per lane; NA - not applicable.*

Using this data, we were able to establish the low and high processing times for each processing technique for single checks per lane that we examine. We found the processing times using basic algebra seen in Equation 3.

$$\frac{450 \text{ vehicles per hour per lane (vphpl)}}{3600 \text{ seconds per hour}} = 0.125 \text{ vehicles per second per lane}$$

$$\frac{1}{0.125 \text{ vehicles per second per lane}} = 8 \text{ seconds per vehicle per lane}$$

Equation 3: Vehicle Processing Time

We were able to find all the service times for each method of service using the same process that is shown in Table 5.

Table 5 : Processing Time for ECP Service

	Processing Times											
	Manual Checks			Checks Using Handheld			Automated Lanes without			Automated Lanes with Arms		
	Low	Mid	High	Low	Mid	High	Low	Mid	High	Low	Mid	High
vplph	300	375	450	275	325	375	400	425	450	325	337.5	350
(1/µ)	12	9.6	8	13.091	11.077	9.6	9	8.471	8	11	10.667	10.286

As discussed in our assumptions, we are not using a pure exponential distribution for our modeling. We use an adjusted processing time, and a sample of this is shown in Table 6.

Table 6: Adjusted Processing Times

	Processing Times					
	Manual Checks			Checks Using Handheld Devices		
	Low	Mid	High	Low	Mid	High
(1/μ)	12	9.6	8	13.091	11.077	9.6
Exp(1/μ - 3) + 3	Exp(9)+3	Exp(6.6)+3	Exp(5)+3	Exp(10.091)+3	Exp(8.077)+3	Exp(6.6)+3

Tandem Servers

As discussed previously, we saw that the Military ECP was not a common queueing system; this is particularly evident when examining the use of tandem servers in a single lane. In Chapter II, Figure 1 showed the overall queueing system which had a single server per facility or process, but if there were multiple servers at a given process there was freedom to maneuver to the next server or out of the system. For ECPs, tandem servers are two servers in a single lane that provide the same service to the customer. Once a customer receives the service, has identification verified from one server, they are able to bypass the next server and continue onto the base. However, due to the security posture at the military ECPs, vehicles are unable to bypass vehicles in front of them keeping them in a single processing lane, as they get closer to each server.

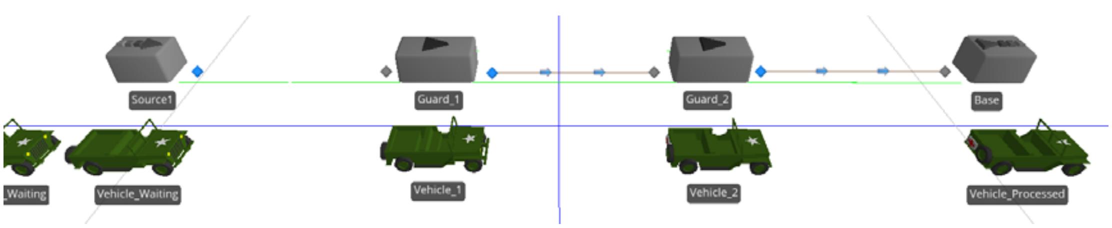


Figure 2: Tandem Server

Examining Figure 2 we see Guard 1 is servicing Vehicle 1 in this queue. Once Vehicle 1 occupant's identifications are verified they no longer need the service of Guard 2, so they are able to continue through the ECP onto the base without stopping to be serviced by Guard 2.

If we exam the process a little closer, we see that this set-up is an inefficient use of manpower. We look at two simple scenarios in order to recognize that the tandem server process is less efficient than the two parallel service gates.

For the start of each scenario, use Figure 2 as a visual reference. The first scenario uses the assumption that Vehicle 1 is processed faster than Vehicle 2 in both the tandem and the parallel servers. In the parallel server example, Vehicle 1 is free to move forward onto the base and allow the next vehicle to move forward and begin processing. Alternatively, in the tandem server, Vehicle 1 is blocked by Vehicle 2 and cannot move forward until Vehicle 2 has completed its service as shown in Figure 3.

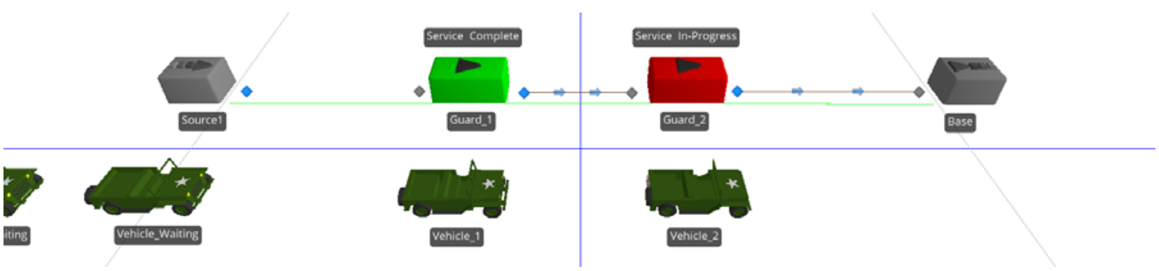


Figure 3 : Vehicle Blocked

The second scenario uses the assumption that Vehicle 2 is processed faster than Vehicle 1 in both the tandem and parallel servers. In both the parallel server and tandem server examples, Vehicle 2 is free to move forward onto the base. Once again, with the parallel servers when a vehicle is processed and moved forward the next vehicle in line is able to move forward and begin processing. However, in the tandem server, once Vehicle 2 moves forward onto the base, the next vehicle in queue is unable to move forward because it is blocked by Vehicle 1 until it has completed its service. Guard 2 sits unutilized (idle) until Vehicle 1 has concluded service seen in Figure 4.

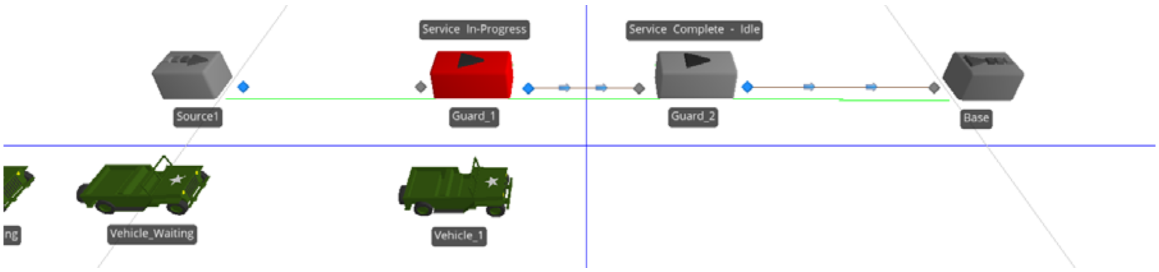


Figure 4: Queue Blocked

In both scenarios, we see that parallel servers allow for efficient throughput since each lane is essentially independent and is not affected by the processing time on the other lane. However, we see that with the tandem servers the maximum of the two service times for Vehicle 1 and Vehicle 2 dictates ECP throughput.

In the utilized software, SIMIO, there is not built in logic for tandem servers. We created our own process that allows a vehicle to bypass the first server when entering into tandem servers; drive to the front available server. We were able to accomplish this using a Decide and Transfer loop within the Processes section of SIMIO for our model. This loop, Figure 5, requires each entity (vehicle) to decide if both guards are open before transferring to the front guard when it is available. This processes leaves out the possibility of infeasibly transferring (leap frogging) to the front guard past an occupied guard still servicing a vehicle.

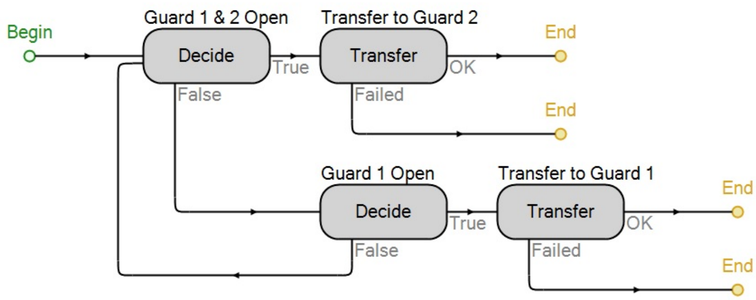


Figure 5: Tandem Loop

Model 1: Split to Individual Queue Model

Since military ECPs are laid out and manned for base security, there is not always freedom to maneuver within the queue of the ECP. For a more controlled approach, ECP sometimes use a technique that emplaces serpentine bollards to slow vehicle speed, Figure 6.

Figure 6: ECP Serpentine (Surface Deployment, 2008)

The serpentine is good for security, but not for traffic flow. This technique takes two lanes of traffic and makes it only one lane, which changes the vehicle capacity of the ECP. To study the effects the serpentine has on an ECP we produce a model that has an individual queue length that is shorter than the overall ECP queue length. The individual queue length begins where the serpentine ends (Figure 7).

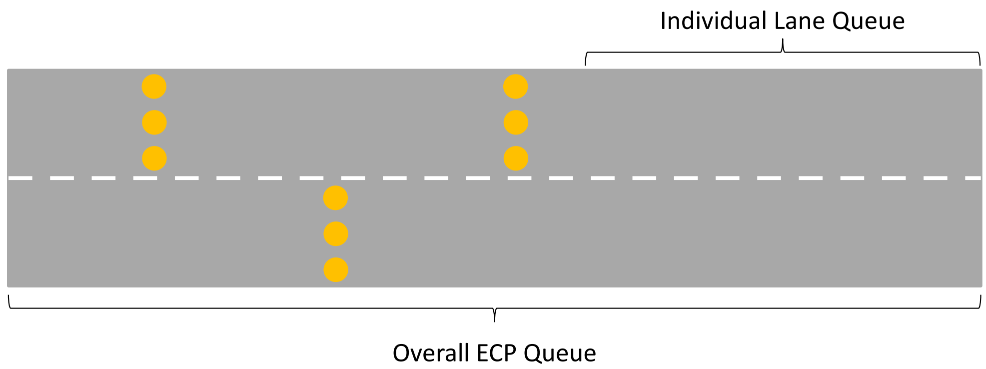


Figure 7: Serpentine for Model

Model 2: Illogical Customer

Most queue systems model queues that have logical customers that go to the next available server. Not all customers at an ECP act logically; this may be due to the customer being concerned about which direction they have to turn after they pass through the ECP or a number of different reasons. We create a model that examines the effect of the illogical customer on the queue length. For this model, we combine the split to individual lane queues with having a percent of customers that stick to the left lane for servicing regardless of the status of the right server, acting illogically. The remainder of customers act logically and enter the server with the fewest in the queue. This process is shown in Figure 8.

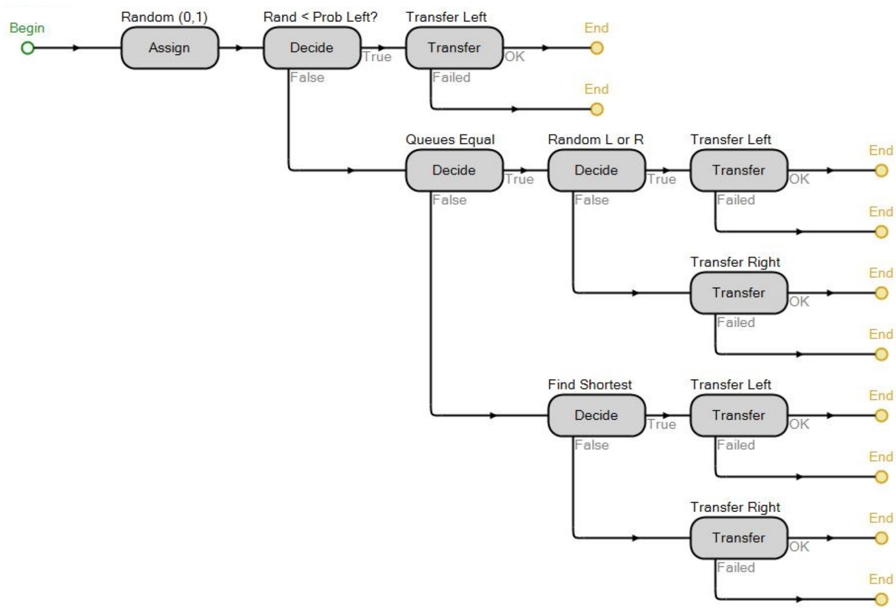


Figure 8: Illogical Customer Process

Model 3: Open Additional Server

A mitigation factor to improve throughput that was previously discussed was opening an additional server. Ideally, an ECP would have all lanes occupied with workers in order to maximize throughput, but this could be a waste in man-power. We would like to find out at what queue length to open an additional server in order to keep the overall queue length from increasing past the ECP lanes capacity.

In order to model this decision, we utilized a process that would check the overall queue length of the system and then open the additional server if the open criterion has been reached. In addition to opening the system, we record the time that it happened for later analysis shown in Figure 9.

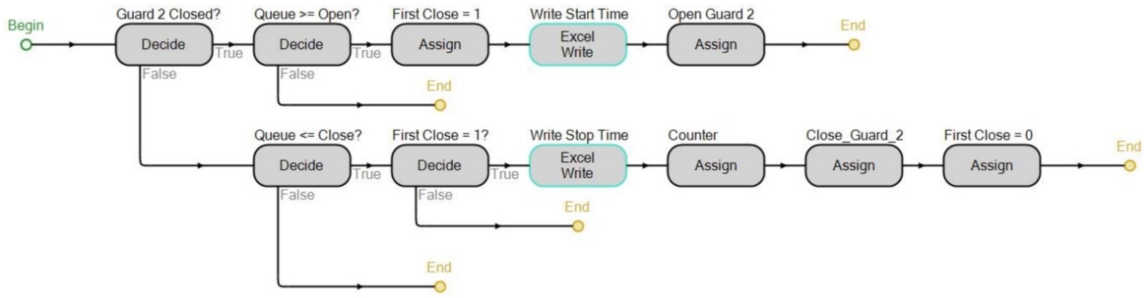


Figure 9: Open New Server Process

The process is the same for all models that open a second server for both tandem and parallel servers.

Ranges for Testing

After choosing a slightly modified exponential distribution for the service time and a Poisson distribution for the arrival time, we use some of the guidelines already in place for M/M/1 and M/M/c servers in order to minimize our testing values.

Arrival Rate (λ)

We wanted the testing to be realistic for ECP workers to calculate and decipher. With this in mind, we use arrival rates that refer to the number of cars that enter the ECP per minute. Using these values makes it easier for workers to adjust while working; a simple count of the cars that entered the previous minute estimate the current arrival rate.

Parallel Servers

In order for an M/M/c server to be considered stable, it must have a utilization rate, $\rho_c < 1$, otherwise the queue length will grow infinitely over time as the server cannot keep up with the arrival rate. For the majority of testing, we examine the ranges where the utilization rate for a singer server is great than 1, $\rho_1 > 1$, and the utilization rate for two servers is less than 1, $\rho_c < 1$. These values are shown in Table 7.

Table 7: Parallel Server Testing

Service Rate (μ) (seconds)	Arrival Rate (λ) vehicles per minute	ρ	ρ_2
8	7	0.93	0.47
8	8	1.07	0.53
8	9	1.2	0.6
8	10	1.33	0.67
8	11	1.47	0.73
8	12	1.6	0.8
8	13	1.73	0.87
8	14	1.87	0.93
8	15	2	1
9.6	6	0.96	0.48
9.6	7	1.12	0.56
9.6	8	1.28	0.64
9.6	9	1.44	0.72
9.6	10	1.6	0.8
9.6	11	1.76	0.88
9.6	12	1.92	0.96
9.6	13	2.08	1.04
9.6	4	0.64	0.32
12	5	1	0.5
12	6	1.2	0.6
12	7	1.4	0.7
12	8	1.6	0.8
12	9	1.8	0.9
12	10	2	1

Tandem Servers

We know that tandem servers should outperform an M/M/1 queue, but underperform compared to an M/M/2 server. This assists in narrowing the scope of testing for the tandem model. After a few initial experiments, our testing uses the range where the utilization rate for a singer server is great than 1 and less than 1.6, $1 < \rho_1 < 1.6$.

Verification of Baseline Model

To ensure that our adjusted distribution followed the current expected throughput for the lanes we ran our model for 50 replications with a constant, overwhelming arrival rate that would ensure a full queue for processing. We then compare our numbers to the estimations provided in Table 4 earlier in this chapter; we do this through visual inspection and not statistical testing as we are only verifying if our model is a good estimation of the current process.

Table 8: Baseline Vehicle Throughput

	Vehicle Throughput					
	Manual Checks			Checks Using Handheld		
	Low	Mid	High	Low	Mid	High
Assumed vplph	300	375	450	275	325	375
Model vplph	303.68	376.22	449.5	278.46	327.38	376.22
Difference	3.68	1.22	-0.5	3.46	2.38	1.22

We see from Table 8 that using the adjusted exponential service time does not affect the overall throughput of the server.

Simulation Specifics

Simulation Overview

The models examined utilize SIMIO for all simulations. Each of the individual variants of the models is replicated 30 times. For example, Model 1 will run 30 replications for a single arrival rate (eight vehicles per minute), single service time (eight seconds per vehicle), and single individual queue length (five vehicles). In order to remove the variance of the arrival rate from influencing the results, the starting seed for the arrival rate is set for consistency between experiments. An internal process to SIMIO will record all start and end times into a spreadsheet (Appendix A: SIMIO Processes) that is used to analyze the data after all experiments are completed.

Recovery Time

The focus of the simulations is to monitor when queue length interferes with traffic outside the base. In order to record this data we needed to set standards for a recovery once the queue reaches a length that interferes with civilian traffic surrounding the base. The simple standard would be once the queue is equal to its capacity it is recovered, but due to the rate of arrivals involved with the queueing system it could be very likely that the system would reach a length over capacity within seconds. We determined an adequate queue length for recovery was two vehicles less than the capacity of the ECP. This would allow for two arrivals before queue length would interfere with outside traffic again. The recovery time is defined in Equation 4.

$$\text{Recovery Time} = \left(\text{Time}_{\text{Queue} < (\text{Capacity - 2})}\right) - \left(\text{Time}_{\text{Queue} > \text{Capacity}}\right)$$

Equation 4: Recovery Time

Unrecoverable Queue

For some of the systems we know there will be a time when the queue length becomes unrecoverable with the processing and arrival rates tested. We establish a method to determine if the queue is considered unrecoverable. We defined a queue as unrecoverable if the end time of the simulation minus the last time the queue length was observed within the capacity of the ECP was greater than the average recovery time for that experiment with the same arrival and processing rates.

Analytic Methodology

There are two primary formulas within M/M/1 and M/M/c queueing systems that we later utilize to compare analytical solutions to our model results. We use the mean (expected) number of customers in the queue during steady state operations. These formulas are shown in Equation 5and Equation 6.

$$L_q = \frac{\rho^2}{1-\rho}$$

Equation 5: M/M/1 Expected Number of Customers in the Queue

$$L_q = \frac{P_0\left(\frac{\lambda}{\mu}\right)^c \rho_c}{c!\left(1-\rho_c\right)^2}, \quad \text{where} \quad P_0 = \left[\sum_{m=0}^{c-1}\frac{\left(c\rho_c\right)^m}{m!} + \frac{\left(c\rho_c\right)^c}{c!\left(1-\rho_c\right)}\right]^{-1}$$

Equation 6: M/M/c Expected Number of Customers in the Queue

Scoring Measure – Fitness Functions

After we obtain the numerical results from the simulations, other than visual inspection, we have not set any criteria to assist in choosing the best option for application. We discuss how we used numerical results in choosing the best options below.

Since our attention is on the length of the queue as the driving decision for our assignment, we choose to focus on unrecoverable queues and the duration of overages that interfere with traffic. Using just those inputs, the results would focus purely on queue length and not on the stress of the worker for the additional ECP as the worker may occupy the server multiple times an hour for a short duration of time each opening. Additionally, any simulation that ended with 50% or more runs unrecoverable were eliminated from consideration regardless of score.

We explore a few different scoring options in order to place emphasis on different aspects of the results. We chose four different scoring equations and noted the top four in each category, which is the lowest score for that category. In order to keep the proportions similar for each category we normalized the data before adding them together. The option with the greatest number of top scores in the all categories is chosen as the best. The four equations we chose as scores are shown below.

$$(\text{\# unrecoverable})+(\text{\# over occurances})+(\text{average duration of overages})$$

Equation 7: Score 1

$$(\text{\# unrecoverable})+(\text{\# over occurances})+2(\text{average duration of overages})$$

Equation 8: Score 2

$$(\text{Score 1})+(\text{\# open occurances})$$

Equation 9: Score 3

$$(\text{Score 2})+(\text{\# open occurances})$$

Equation 10: Score 4

If there is a tie within a given queue system we will chose the option with the longest opening length as we will obtain similar results with a longer queue length before opening.

Summary

This chapter discussed many aspects of our models that we use to build our simulations to analyze ECP operations. We discussed which distributions to use to for our model as well as some of the processing decisions we use during our simulations. The ranges for our tests were determined based on our knowledge of a similar queueing system. We concluded with some of the analytic solution methods we use during our analysis.

IV. Analysis and Results

Chapter Overview

This chapter uses simulation results from the methods described previously to provide insight in our queueing system. We focus on four simulation models: the baseline model, split length to individual queues model, illogical customer model, and opening of second server model. For the baseline model, we compare our results to analytical solutions to verify if we can use them for later analysis. For the other three models, we explore the results from some of the more interesting simulations and discuss insights we take away from the simulations.

Baseline Model

We wanted to explore each queue system process, single server, tandem server, and parallel server, to compare them to some of the analytic solutions of similar queueing systems. Each server design ran for 30 replications with a one-hour warm up period in order to get the system into a steady state.

Single Server

We concentrated our simulations on arrival rates (λ) and processing rates (μ) that would result in utilizations that range from 0.60 to 1.00. This would give us an insight to the model in comparison with the M/M/1 queueing system.

The first test we conduct is the comparison of the expected queue length, L_q, utilizing a two sample t-test to verify if the means are equal. For the rejection criterion, we use a significance level of 0.05 to reject the hypothesis that the means are equal.

Table 9: Mean t-test for Single Service Model

Processing Time (1/μ)	Inter-arrival Time (1/λ)	ρ	Exponential Distribution L_{Eq}	Adjusted Distribution L_{Aq}	t-score	p-value	$L_{Eq} = L_{Aq}$
8	7	0.933	5.290	2.605	4.586	0.000	No
8	6	0.800	1.465	0.627	8.583	0.000	No
8	5	0.667	0.546	0.212	11.010	0.000	No
8	4	0.533	0.199	0.066	12.055	0.000	No
9.6	6	0.960	7.980	4.776	2.682	0.010	No
9.6	5	0.800	1.460	0.727	6.936	0.000	No
9.6	4	0.640	0.439	0.195	8.841	0.000	No
11.08	5	0.923	4.663	2.792	2.872	0.006	No
11.08	4	0.739	0.897	0.474	6.174	0.000	No
11.08	3	0.554	0.214	0.103	5.812	0.000	No
12	4	0.800	1.425	0.818	4.674	0.000	No
12	3	0.600	0.303	0.158	5.192	0.000	No
13.1	4	0.873	2.609	1.643	3.042	0.004	No
13.1	3	0.655	0.448	0.255	4.516	0.000	No

Looking at Table 9 we see that for all models we reject the hypothesis that the mean queue lengths are equal. Using these results, there is a low likelihood that our model is equivalent to the M/M/1 system. In order to estimate results of our queueing system, utilizing the M/M/1 system calculations would not be a good assumption.

Tandem Servers

The same testing is conducted for the tandem server, though there are no tandem server systems with established analytical solutions. The tandem simulations were conducted on single server utilizations that range from 0.70 to 1.20. For this comparison, we once again use a two-sample t-test to verify the equality of the means.

Table 10: Mean t-test for Tandem Service Model

Processing Time (1/μ)	Inter-arrival Time (1/λ)	ρ	Exponential Distribution L_{Eq}	Adjusted Distribution L_{Aq}	t-score	p-value	$L_{Eq} = L_{Aq}$
8	9	1.200	114.707	85.347	2.591	0.012	No
8	8	1.067	26.309	6.377	3.208	0.002	No
8	7	0.933	1.851	0.320	4.874	0.000	No
8	6	0.800	0.426	0.121	6.240	0.000	No
9.6	7	1.120	44.656	18.959	3.498	0.001	No
9.6	6	0.960	1.985	0.582	4.039	0.000	No
9.6	5	0.800	0.393	0.143	3.881	0.000	No
11.08	6	1.108	28.331	12.341	2.653	0.011	No
11.08	5	0.923	1.425	0.449	3.189	0.003	No
11.08	4	0.739	0.342	0.116	4.782	0.000	No
12	6	1.200	75.578	55.672	2.397	0.020	No
12	5	1.000	4.416	1.032	3.357	0.002	No
12	4	0.800	0.559	0.196	4.151	0.000	No
13.1	5	1.092	21.331	11.567	1.764	0.084	**Yes**
13.1	4	0.873	0.940	0.390	3.644	0.001	No
13.1	3	0.655	0.163	0.072	3.595	0.001	No

The results seen in Table 10 show that there is a low likelihood that the adjusted exponential distribution tandem model is equivalent to a strict exponential distribution as only 1 of 16 means were statistically equal.

Parallel Servers

Once again, the same testing is conducted for the parallel servers. Our simulations were conducted on arrival rates (λ) and processing rates (μ) that would result in utilizations that range from 0.60 to 1.00 for a multiple server queueing system. A two-sample t-test is used to verify the equality of the means.

Table 11: Mean t-test for Parallel Service Model

Processing Time (1/μ)	Inter-arrival Time (1/λ)	ρ	Exponential Distribution L_{Eq}	Adjusted Distribution L_{Aq}	t-score	p-value	$L_{Eq} = L_{Aq}$
8	15	1.000	37.955	28.699	1.230	0.224	**Yes**
8	14	0.933	9.928	4.942	2.820	0.008	No
8	13	0.867	3.355	1.643	3.897	0.000	No
8	12	0.800	1.735	0.837	3.815	0.000	No
8	11	0.733	0.912	0.424	5.399	0.000	No
8	10	0.667	0.477	0.201	6.846	0.000	No
9.6	12	0.960	13.755	8.464	1.618	0.112	**Yes**
9.6	11	0.880	3.682	2.059	2.891	0.006	No
9.6	10	0.800	1.470	0.779	4.404	0.000	No
9.6	9	0.720	0.684	0.343	5.339	0.000	No
9.6	8	0.640	0.340	0.162	5.186	0.000	No
11.08	11	1.016	37.104	32.086	0.804	0.425	**Yes**
11.08	10	0.923	5.418	3.380	2.292	0.027	No
11.08	9	0.831	1.828	1.049	3.418	0.001	No
11.08	8	0.739	0.769	0.430	3.293	0.002	No
11.08	7	0.646	0.341	0.174	4.257	0.000	No
12	10	1.000	24.588	20.016	0.900	0.372	**Yes**
12	9	0.900	3.857	2.423	2.700	0.010	No
12	8	0.800	1.287	0.783	2.618	0.012	No
12	7	0.700	0.540	0.300	3.654	0.001	No
12	6	0.600	0.239	0.125	4.350	0.000	No
13.1	9	0.983	16.856	12.954	1.148	0.256	**Yes**
13.1	8	0.873	2.646	1.695	2.432	0.019	No
13.1	7	0.764	0.925	0.558	2.942	0.005	No
13.1	6	0.655	0.388	0.222	3.704	0.001	No

Examining Table 11 we see that for values of ρ greater than 0.96 there is no evidence to reject that the means of the queue lengths are equal. This is most likely due to the high variance of the mean for these models. Our concentration for our simulations is on utilization values of less than one otherwise the models could become unstable and queue lengths would trend towards infinity and there are no closed for solutions for M/M/c systems unless $\rho < 1$. This information confirms that for all three models, there is a low likelihood that the adjusted exponential distribution models are equivalent to a strict exponential distribution. These results inform us not to use known M/M/c or M/M/1 system calculations as estimates for our models.

Model 1: Split to Individual Queue Model

Examining the split to individual queue model where customers act logically is comparable to an examination into the overall length of a single queue system with two servers. Using this information, we know that the length should increase continuously over time with utilization rates greater than or equal to one. We also discovered from the baseline model that once the utilization rate drops below 0.9 the average queue length is less than four customers that is well below the ECP capacity of twenty customers. We focus our analysis on the five models with $0.90 \leq \rho_2 < 1.00$

Results

Starting with the lowest two utilization models of $\rho_2 = 0.90$ and $\rho_2 = 0.92$, we found that if the individual queue length is greater than four, there were no occurrences where the overall queue length exceeded the capacity of the ECP. In order for the next model, $\rho_2 = 0.93$, to satisfy the same requirement, the individual queue length would have to be greater than eleven. Although that is the best case for this model, having an individual queue length of 1 resulting on only 0.4 overage occurrences per hour with an average recovery time of 3.0 ± 1.3 minutes.

For the next three models, we had to utilize numerical criteria to assist in our decision-making, as it was not as straightforward as the previous results. We focused on minimizing unrecoverable occurrences within each model as well minimizing all overage occurrences in the simulations. The choices for the remaining two models are shown in Table 12.

Table 12: Split Queue Results

Processing Time (1/μ)	Inter-arrival Time (1/λ)	Two Server $\rho_2 = (\lambda/2\mu)$	Length of Individual Queues	Unrecoverable (Percent)	Over Occurances (per hour)	Average Recovery Time (minutes)
13.1	6.67	0.98	> 14	3.33%	0.183333333	4.47 ± 3.85
9.6	5.00	0.96	> 10	3.33%	0.1	4.81 ± 2.85

Insights

When opening a second server it is important to open a second full lane. Although there are few models, when parallel servers are unable to maintain a steady control of the queue, fully opening the second lane adds a large buffer for any variations with the customers at the ECP.

Model 2: Illogical Customer Model

With the illogical customer model, we test probability values from 0.2 to 0.8 of turning left while the remaining customers act in a logical manner choosing the shortest queue to enter if it was accessible.

Results

We are able to conclude that a probability of turning left of 0.8 was detrimental to queues and ended every replication in an unrecoverable state. Most models with a probability of turning left equal to 0.7 were also unrecoverable except for two models, $\rho_2 = 0.70$ and $\rho_2 = 0.72$, which resulted in 33% and 40% of the models reaching an unrecoverable queue length. Once the probability dropped to 0.6 there were only five total models able to sustain a maintainable queue within the ECP capacity; all five sustainable models had a utilization value less than or equal to 0.80.

The models with a left turn probability of 0.5 showed the largest range throughout the models. Examining the average queue length in Figure 10, we see an almost linear trend to what individual queue length is required for the overall queue length to no longer be affected by the turning probability.

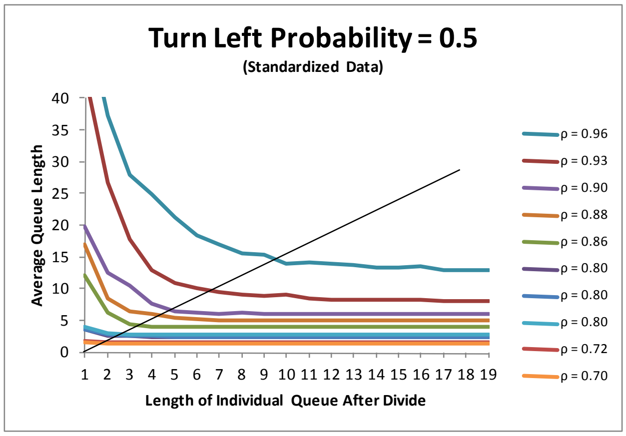


Figure 10: Illogical Customer Probability = 0.5

With the probability at or below 0.4 of turning left, all models could easily maintain an overall queue length within the ECP capacity. There were only three models, seen in Table 13, that required a separate individual queue while all other models with $\rho_2 < 0.90$ were able to sustain a single queue for the system and stay within the ECP capacity.

Table 13: Illogical Customer Probability = 0.40

Processing Time (1/μ)	Inter-arrival Time (1/λ)	Two Server $\rho_2 = (\lambda/2\mu)$	Length of Individual Queues
9.6	5.00	0.96	> 6
8	4.29	0.93	> 3
12	6.67	0.90	> 1

*All other models (ρ < 0.90) can sustain a joint single queue

Insights

When using multiple (parallel) servers it is vital to open both individual queue lanes as fully as possible in order to sustain any variability with the customers.

Model 3: Open Additional Server Model

Tandem Server Results

After running multiple iterations of the tandem server, we notice that this system can be extremely sensitive. This is amplified by the fact we only used inter-arrival times related to whole car arrivals per minute. There is a very small gap of utilization values where operating an additional tandem server would be useful. From our results, the only time to utilize a tandem server as the only addition to your single server queueing system would be for the following single server utilization values: $0.93 < \rho < 1.12$.

Using the weighting scale from Chapter III, we were able to choose the best option of when to open the tandem server. The results of this testing is shown below, in Table 14.

Table 14: Results for Open Tandem Server

Processing Time (1/μ)	Inter-arrival Time (1/λ)	Single Server ρ = (λ/μ)	Queue Length To Open	Open Time (minutes)	Between Openings (minutes)	Over Occurances (per hour)	Queue Unrecoverable (Percent)
9.6	8.57	1.12	+	+	+	+	+
11.08^	10	1.11	5	7.22 ± 1.63	15.28	0.716666667	43.33%
13.1^^	12	1.09	6	11.69 ± 2.35	30.66	0.45	40.00%
8	7.5	1.07	6	7.38 ± 1.44	10.1	0.57	16.67%
9*	8.57	1.05	9	11.29 ± 2.19	24.01	0.333333333	20.00%
9.6	9.23*	1.04	7	7.64 ± 1.24	14.73	0.133333333	0.00%
12	12	1.00	8	9.82 ± 1.68	31.09	0.016666667	0.00%
9.6	10	0.96	12	8.99 ± 1.92	107.14	0.00	0.00%
8	8.57	0.93	-	-	-	-	-
11.08	12	0.92	18	12.14	3,587.86	0.00	0.00%
13.1	15	0.87	-	-	-	-	-
12	15	0.80	-	-	-	-	-

+ All open lengths resulted in > 50% unrecoverable

^ Only four options were consider as the rest resulted in > 50% unrecoverable

^^ Only nine options were consider as the rest resulted in > 50% unrecoverable

* Data added to fill gap in ρ values

- Single server sustained for 60 hours of simulation

Even though we were able to find solutions using our scoring criteria, we wanted to see if we would come to the same conclusions using visual inspection of the graphical results. We examine two of the results with the graphs; referring to the experiments by ρ value, we examine ρ = 1.07 and ρ = 1.00 from Table 14.

For the first experiment, ρ = 1.07, Figure 11 contains three separate graphs which contain the majority of the results we have been discussing. The top left graph shows the average ending queue length for each experiment with a queue length of 20 being our ECP capacity. The top right graph displays the statistics gathered from the second server opening, and the bottom graph shows the recovery statistics for the experiment. By examining the graphs in Figure 11 we would most likely choose a value between 4 and 6, and being conservative we would choose the open length of 4 which is the same result utilizing the scoring method.

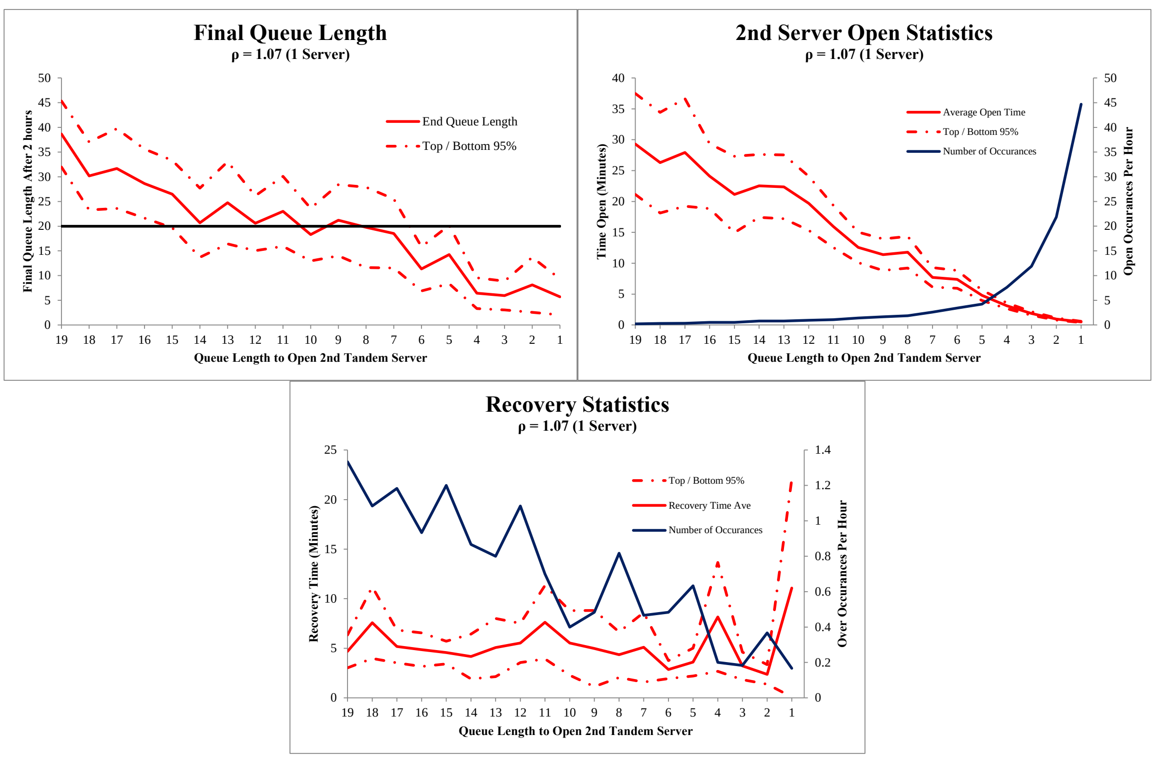


Figure 11: Tandem Experiment ρ = 1.07

The second ρ = 1.00 has graphs configured the same way as the previous experiment in Figure 12. Focusing specifically on the recover statistics, we would likely choose a queue length of 12-14 for the opening length. Looking back at our scoring method, we would have chosen 14 which strengths our confidence that the scoring methodology.

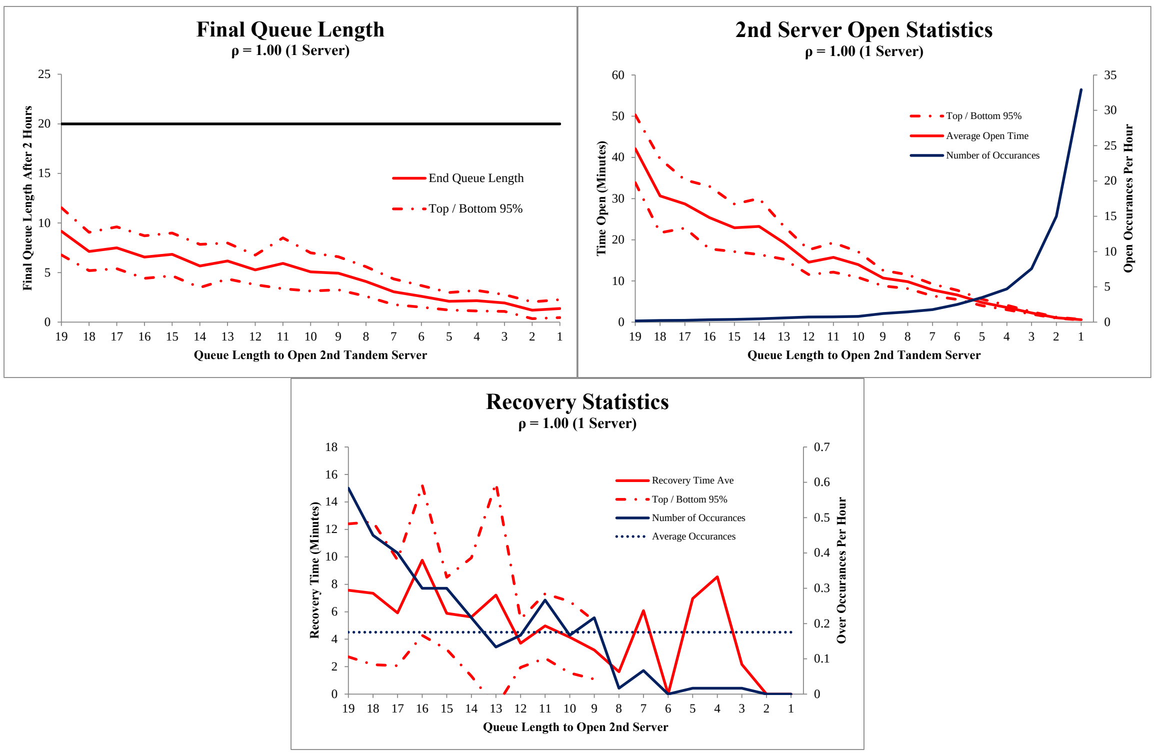


Figure 12: Tandem Experiment ρ = 1.00

Tandem Server Insight

While opening a second server in a tandem position may be beneficial to increasing throughput when faced with a high arrival rate, it is not beneficial in dropping the overall length of the queue during steady state operations.

When observing the operation of this queue during the simulation we noticed that the tandem servers acted like a batch server. This was due to the fact that both servers were occupied or blocked until all services were complete. That idea should hold for any number of tandem servers. The tandem servers do not act like a typical batch server, which has one service distribution for the batch. Instead, the service time for the "batch" is the max service time of the total of tandem servers.

Parallel Server Results

The results of the parallel server system seem to require a little less analysis due to the responsiveness of the queue as we cut the utilization in half when adding a second server because both servers are accessible to the queue.

We look at the same model over three different queue capacities. We first look at opening a second server but keeping a single queue so the capacity does not change from 20 when opening a second server. Then we look at opening a second server with a partially blocked second lane only opening half the available 2^{nd} queue increasing the capacity to 30. Lastly, we look at the case where we open a second server with the full lane accessible to the queue for a new capacity of 40 vehicles.

ECP Capacity Single Lane

When analyzing the single lane model for the best choice of opening queue length the scoring model was no longer the best method for models with $\rho < 1.6$. Using the scoring methodology these models would choose the small values of opening queue length when there was no significant benefit to opening the queue more often for shorter periods of time. These models also had zero unrecoverable replications. For models with $\rho < 1.6$ we chose the best option to be when the number of over occurrences per hour was less than 1. Table 15 shows our results for opening a parallel server.

Table 15: Results for Open Parallel Server (Capacity = 20)

Processing Time (1/μ)	Inter-arrival Time (1/λ)	Single Server ρ = (λ/μ)	Queue Length To Open	Open Time (minutes)	Between Openings (minutes)	Over Occurances (per hour)	Queue Unrecoverable (Percent)
8	4.29	1.87	5	4.39 ± 0.52	1.24	0.73	3.33%
	4.62	1.73	13	6.43 ± 0.38	2.8	0.52	0.00%
	5.00	1.60	16	5.31 ± 0.22	3.94	0.67	0.00%
	5.45	1.47	17	4.18 ± 0.14	5.15	0.42	0.00%
	6.00	1.33	18	3.51 ± 0.10	7.63	0.65	0.00%
	6.67	1.20	18	2.94 ± 0.10	12.58	0.17	0.00%
	7.50	1.07	19	2.61 ± 0.11	33.03	0.27	0.00%
9.6	5.00	1.92	7	10.19 ± 1.86	4.68	1.02	23.33%
	5.45	1.76	17	11.16 ± 0.78	4.91	2.08	3.33%
	6.00	1.60	17	6.50 ± 0.31	5.04	0.83	0.00%
	6.67	1.44	18	5.08 ± 0.18	7.25	1	0.00%
	7.50	1.28	18	3.93 ± 0.13	11.01	0.55	0.00%
	8.57	1.12	19	3.24 ± 0.13	24.88	0.47	0.00%
12	6.67	1.80	10	8.91 ± 0.85	3.72	0.483333333	3.33%
	7.50	1.60	17	8.15 ± 0.42	6.49	0.75	0.00%
	8.57	1.40	18	5.79 ± 0.24	10.07	0.6	0.00%
	10.00	1.20	19	4.70 ± 0.22	20.48	0.533333333	0.00%
11.08	6.00	1.85	8	8.28 ± 0.93	2.8	0.416666667	3.33%
	6.67	1.66	16	8.62 ± 0.49	5.44	0.616666667	0.00%
	7.50	1.48	17	5.95 ± 0.24	7.43	0.483333333	0.00%
	8.57	1.29	18	4.52 ± 0.18	12.46	0.333333333	0.00%
	10.00	1.11	19	3.66 ± 0.17	31.29	0.4	0.00%
13.1	6.67	1.97	11	23.16 ± 5.17	26.15	1.3	36.67%
	7.50	1.75	14	11.59 ± 0.92	5.72	0.766666667	0.00%
	8.57	1.53	17	7.31 ± 0.37	8.27	0.6	0.00%
	10.00	1.31	18	5.61 ± 0.25	13.75	0.316666667	0.00%
	12.00	1.09	19	4.42 ± 0.27	44.89	0.15	0.00%

ECP Capacity Single Plus Partial Lane

There was a significant change to the results if the ECP was able to achieve 1.5 times the original capacity opening a second lane. For all models with $\rho \leq 1.85$, the opening criteria was 19 in the queue (at capacity) as long as vehicles could begin to move freely into the second queue. For the remaining three models and their new recommended opening queue, lengths are shown in Table 16.

Table 16: Results for Open Parallel Server (Capacity = 30)

Processing Time (1/μ)	Inter-arrival Time (1/λ)	Single Server ρ = (λ/μ)	Queue Length To Open	Open Time (minutes)	Between Openings (minutes)	Over Occurances (per hour)	Queue Unrecoverable (Percent)
13.1	6.67	1.97	12	26.43 ± 6.18	27.3	0.65	16.67%
9.6	5.00	1.92	11	16.21 ± 2.74	9.14	0.52	10.00%
8	4.29	1.87	14	12.84 ± 1.51	4.98	0.27	3.33%

ECP Capacity Double Lane

If the ECP was able to open up two full lanes for a 40 vehicle capacity the only model that would not have opening criteria of a queue length of 19 would be the model with a single server $\rho = 1.97$ with an opening queue length equal to 16, shown in Table 17.

Table 17: Results for Open Parallel Server (Capacity = 40)

Processing Time (1/μ)	Inter-arrival Time (1/λ)	Single Server ρ = (λ/μ)	Queue Length To Open	Open Time (minutes)	Between Openings (minutes)	Over Occurances (per hour)	Queue Unrecoverable (Percent)
13.1	6.67	1.97	16	30.97 ± 7.21	44.03	0.316666667	10.00%

Parallel Server Insight

As we saw earlier, opening up the extra queue capacity within the ECP is crucial. While opening a second server is great advantage to processing, opening a second server and a second queue is the best way to keep the ECP queue from interfering with the outside traffic.

Scoring Sensitivity Analysis

Exploring the sensitivity of the scoring fitness functions, we examine the criteria in which we consider a queue unrecoverable. In our earlier fitness function, we defined an unrecoverable queue if the end time of the simulation minus the last time the queue length was observed within the capacity of the ECP was greater than the average recovery time for that experiment with the same arrival and processing rates. In order to cover a wider range of data, we explored an unrecoverable queue if the end time of the simulation minus the last time the queue length was observed within the capacity was greater than the average recovery time plus three standard deviations. Using Chebyshev's inequality, we know that if we include up to three standard deviations, we cover at least 88.89% of the data. Changing the fitness function with the new unrecoverable criteria, we only see a change in three of the 27 parallel model results seen in Table 18.

Table 18: Fitness Function Sensitivity Analysis Parallel

Processing Time (1/μ)	Inter-arrival Time (1/λ)	Single Server ρ = (λ/μ)	Queue Length To Open	Open Time (minutes)	Between Openings (minutes)	Recovery Time (minutes)	Queue Unrecoverable (Percent)	
8	4.29	1.87	8	7.41 ± 0.94	2.04	2.70 ± 0.73	6.67%	Mean
			3	2.37 ± 0.28	0.7	3.12 ± 1.22	0.00%	Mean + 3σ
9.6	5.00	1.92	5	6.93 ± 1.20	2.96	5.55 ± 1.80	23.33%	Mean
			3	3.66 ± 0.63	1.48	5.69 ± 1.69	10.00%	Mean + 3σ
13.1	6.67	1.97	4	7.70 ± 1.72	5.48	8.61 ± 2.92	33.33%	Mean
			6	11.08 ± 2.41	10.22	9.56 ± 4.11	0.00%	Mean + 3σ

With the 12 tandem models, the results did not change at all. The total change from using the new fitness function for the 39 experiments is less than eight percent. After seeing these results, we did not change the original fitness functions.

Summary

This chapter provided analysis for four queueing system models. We were also able to provide some insights into the overall model from the analysis conducted. Finally, in Chapter V, we discuss conclusions and recommendations for an ECP queueing system.

V. Conclusions and Recommendations

Chapter Overview

In this chapter, we summarize the insights found from our research and propose possible directions for future research. This analysis presented insights into the ECP queueing system that can prevent the system from interfering with surrounding traffic. With improved data and further exploration into multiple lane systems, policies and procedures could be implemented to improve ECP operations.

Conclusions of Research

Although our model utilized an adjusted exponential service distribution, not statistically the same as a pure exponential service distribution, we were able to use the utilizations, ρ, as references for our experimentation. This is beneficial for an ECP to utilize as a quick calculation to understand if their current processing layout is even feasible with the current arrival rate; the ECP cannot sustain a constant utilization greater than 1.

Illogical Customer

In order to prevent the illogical customer base from affecting the operations of the ECP it is imperative to open both queues back as far as possible in order to maximize the space for logical drivers to move to the shortest queue available. Not opening the entire queue can result in customers being unable to access an available server, which then cause the queue length to grow unnecessarily due to obstruction.

Tandem Server vs. Parallel Server

Based on the results from the two experiments there does not seem to be a time when utilizing tandem servers over parallel servers would be beneficial. The tandem server set-up was only able to recover the queue for utilization levels of $\rho \leq 1.11$, where the parallel system was able to recover for utilization levels of $\rho \leq 1.97$.

While adding a second tandem server does not significantly influence the overall length of the queue compared to the parallel server, in certain circumstances it can be beneficial. If there is not another lane to add an additional parallel server or the ECP is trying to shorten the wait time for customers in a system with utilization levels of $\rho \leq 1.11$, then utilizing the tandem server is beneficial.

Recommendations for Future Research

We suggest the following areas to improve and/or extend this research.

Verify Service Distribution

Without empirical data, we chose to utilize the most common service distribution, the exponential distribution. Gathering data from ECPs throughout the military would be beneficial to verify our assumption.

Multiple Lanes

This research focused on two lanes for an ECP system, which allowed us to focus on the effect of adding a second server to the ECP, not set policies. Expanding the research to larger systems could provide input to a variety of ECP arrangements allowing each military instillation to receive input for their specific ECF layout.

Closed Form Solution for Adding Additional Server

We utilize simulations for our research because we chose to use an unconventional service distribution that did not have analytical or closed form solutions. For queueing systems that operate with known distributions, utilizing the *N*-policy research from Balachandran, 1973, it may be possible to explore a closed form solution for adding and removing queueing systems that become overwhelmed.

Tandem Servers as Batch Service

As previously discussed, we observed tandem servers to operate similar to a batch service system with the service time equal to the maximum service time of the tandem servers in that lane. Exploring this further may lead to way to approximate the two service times as one batch service time. This would speed up the processing time of experiments utilizing fewer decisions in the simulations. This could also lead to better comparisons of the system if approximate batch processing time distributions are found.

Appendix A: SIMIO Processes

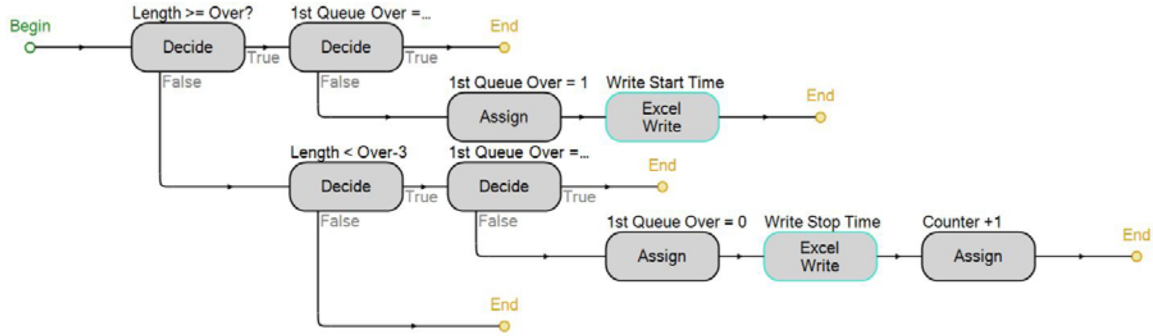


Figure 13: Record Length of Time for Recovery of Queue (All Models)

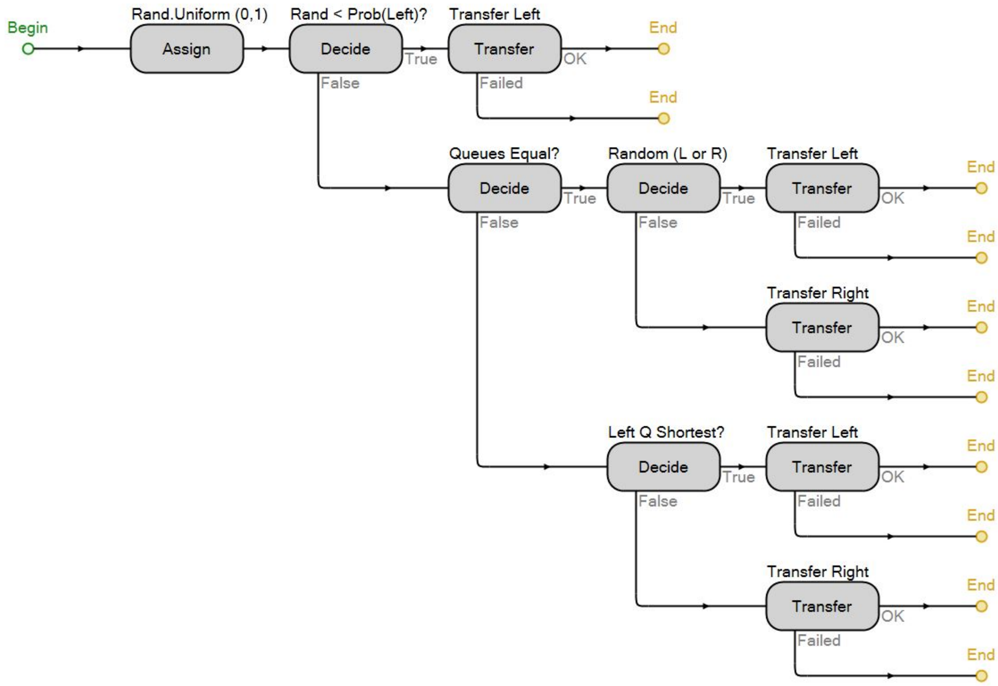


Figure 14: Illogical Customer Decision to Choose Left or Right Lane (Model 2)

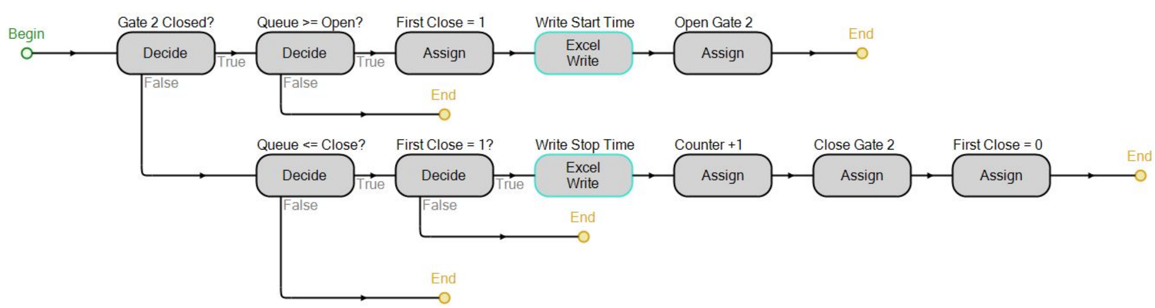


Figure 15: Record Length of Time 2nd Server Open (Model 3)

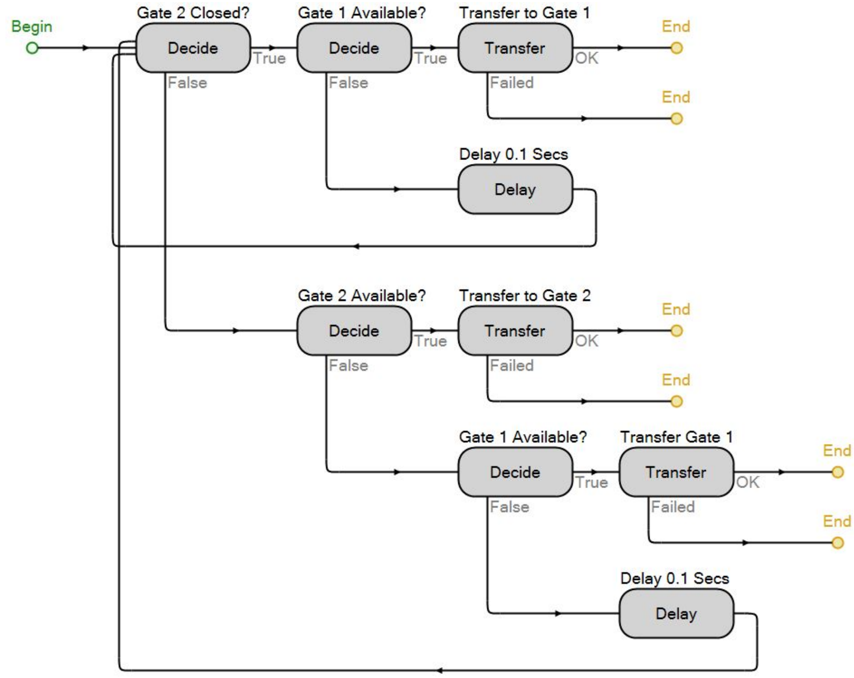


Figure 16: Transfer Vehicle to Open Gate (Model 3)

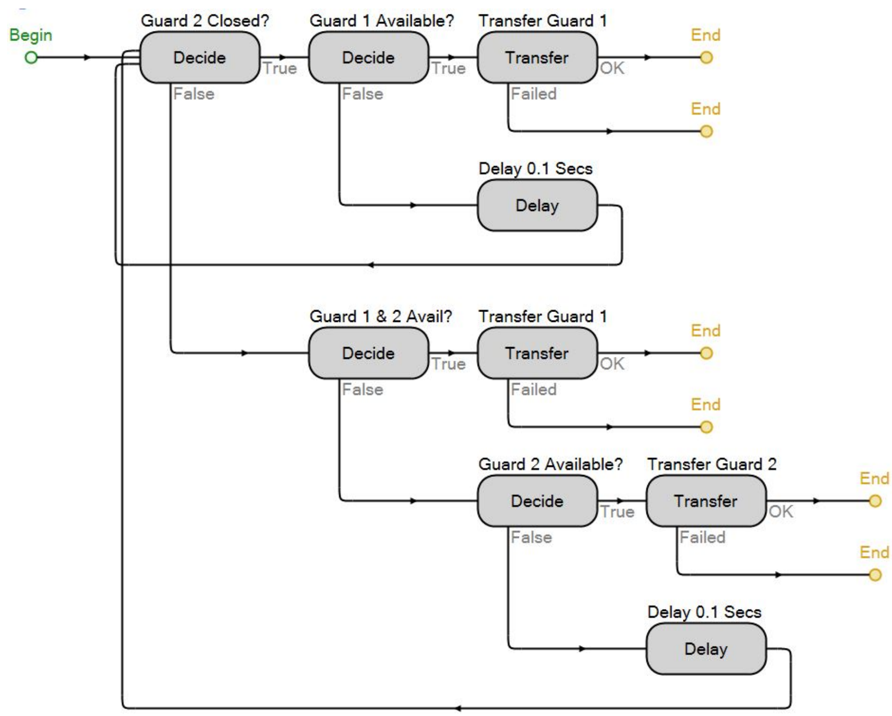
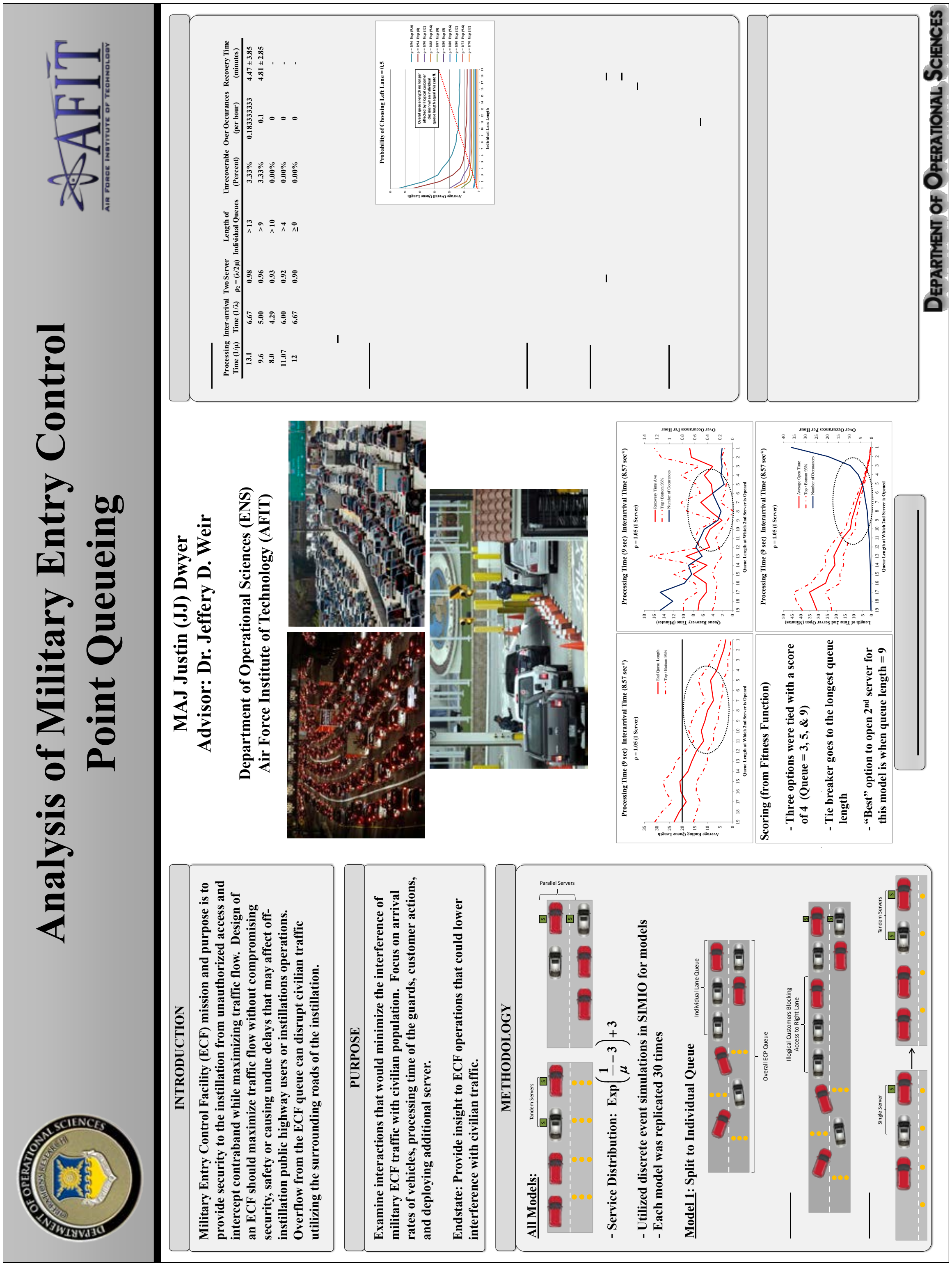


Figure 17: Transfer Vehicle to Open Server (Model 3 - Tandem)

Appendix B: Story Board



Bibliography

Allen, A. O. (2014). *Probability, statistics, and queueing theory.* Cambridge: Academic Press.

Ancker Jr., C., & Gafarian, A. (1963). Some Queuing Problems with Balking and Reneging. I. *Operations Research, 11*(1), 88-100.

Balachandran, K. (1973). Control policies for a single server system. *Management Science, 19*(9), 1013-1018.

Banks, J., Carson, J., Nelson, B., & Nicol, D. (2000). *Discrete-Event System Simulation* (Fifth Edition ed.). Englewood Cliffs, New Jersey, USA: Prentice Hall.

Buckley, D., & Yagar, S. (1974). Capacity funnels near on-ramps. *Transportation and Traffic Theory, Proceedings,* (pp. 105-123).

Chau, H., Xu, H., & Liu, L.-G. (2002). Exactly solvable single lane highway traffic model with tollbooths. *Physica A: Statistical Mechanics and its Applications, 303*(3), 534-542.

Connelly, L., & Bair, A. (2004). Discrete event simulation of emergency department activity: A platform for system-level operations research. *Academic Emergency Medicine, 11*(11), 1177-1185.

Danbury City Council. (2016, 1 12). Retrieved from City of Danbury, CT: http://www.danbury-ct.gov/filestorage/21015/21087/21123/23014/Page8-1112.pdf

Department of Defense. (25 May 2005). *UFC 4-022-01. Security Engineering: Entry Control Facilities / Access Control Points.*

Dickson, D., Ford, R., & Laval, B. (2005). Managing real and virtual waits in hospitality and service organizations. *Cornell Hotel and Restaurant Administration Quarterly, 46*(1), 52-68.

Disney Parks & Travel. (n.d.). *Walt Disney World Help Center - FastPass+*. Retrieved 11 19, 2015, from Walt Disney World Website: https://disneyworld.disney.go.com/help/fast-pass-plus/

Ferrari, P. (1995). Road pricing and network equilibrium. *Transportation Research Part B: Methodological, 29*(5), 357-372.

Fort Worth City Council. (2016, 1 12). *City of Fort Worth Development Standards and Guidelines.* Retrieved from City of Fort Worth: http://fortworthtexas.gov/uploadedFiles/Planning/Urban_Design/Development_Standards_and_Guidelines/ParkingLotDesign.pdf

George, J., & Harrison, J. (2001). Dynamic control of a queue with adjustable service rate. *Operations Research*, 720-731.

Heizer, J., & Render, B. (2010). *Operations Management* (10th ed.). Pearson Prentice Hall.

Heyman, D. (1977). The T-policy for the M/G/1 queue. *Management Science*, *23*(7), 775-778.

Huang, D.-W., & Huang, W.-N. (2002). The influence of tollbooths on highway traffic. *Physica A: Statistical Mechanics and its Applications, 312*(3), 597-608.

Huebner, F. (1998). On the accuracy of approximating loss probabilities in finite queues by erobabilities to exceed queue levels in infinite queues. *Global Telecommunications Conference, 1998. GLOBECOM 1998. The Bridge to Global Integration. IEEE. 1*, pp. 484 - 489. IEEE.

Intellicheck Mobillisa, Inc. (2015, October 13). US Army's Fort Sill Expands Adoption of Intellicheck's Defense ID. Jericho, New York, United States of America. Retrieved 10 31, 2015, from http://www.businesswire.com/news/

Ishizaki, F., & Takine, T. (1999). Loss probability in a finite discrete-time queue in terms of the steady state distribution of an infinite queue. *Queueing Systems, 31*(3-4), 317-326.

Jones, P., & Dent, M. (1994). Improving service: managing response time in hospitality operations. *International Journal of Operations & Production Management, 14*(5), 52-58.

Jones, P., & Peppiatt, E. (1996). Managing perceptions of waiting times in service queues. *International Journal of Service Industry Management, 7*(5), 47-61.

Larson, R., Cahn, M., & Shell, M. (1993). Improving the New York City arrest-to-arraignment system. *Interfaces*, *23*(1), 76-96.

Lemaster, A. (2015). *Analysis and Implications of Guest Attitudes Towards Queuing in Theme Parks.* Orlando: University of Central Florida.

Little, J. (1961). A Proof for the Queuing Formula: L = λW. *Operations Research, 9*(3), 383 - 387.

Madan, K. (2000). An M/G/1 queue with second optional service. *Queueing systems, 34*(1-4), 37-46.

Madan, K., Al-Nasser, A., & Al-Masri, A.-Q. (2004). On M [x] G1G21 queue with optional re-service. *Applied Mathematics and Computation, 152*(1), 71-88.

McKenzie, B., & Rapino, M. (2011, 1 13). *Commuting in the United Sstates: 2009.* US Department of Commerce, Economics and Statistics Administration, US Census Bureau . Retrieved from United States Census Bureau Website: https://www.census.gov/prod/2011pubs/acs-15.pdf

Oviaci, K., & May, A. (1974). Freeway priority entry control to favor multi-passenger vehicles. *Transportation and traffic theory, 6*, 125-160.

Pinto, R. (2011). *Operations research and management science handbook.* Florence: Taylor & Francis.

Planning Division. (2016, 1 12). *City of San Jose Document Center.* Retrieved from City of San Jose: https://www.sanjoseca.gov/DocumentCenter/Home/View/433

Render, B., Stair, R., & Hanna, M. (2012). *Quantitative Analysis for Management: Global Edition* (11th ed.). Pearson Higher Education.

Ross, S. M. (2014). *Introduction to Probability Models* (11th ed.). San Diego: Academic Press.

Rue, R., & Rosenshine, M. (1981). Optimal control for entry of many classes of customers to an M/M/1 queue. *Naval Research Logistics Quarterly, 28*(3), 489-495.

Small, K. (1992). Guest Editorial. *Transportation, 19*(4), 287-291.

Surface Deployment. (2008). *Distribution Command Transportation Engineering Agency, SDDCTEA. Pamphlet 55-15.* Military Surface Deployment and Distribution Command Transportation Engineering Agency.

Van Den Broek, M., Van Leeuwaarden, J., Adan, I., & Boxma, O. (2006). Bounds and approximations for the fixed-cycle traffic-light queue. *Transportation Science, 40*(4), 484-496.

Vickrey, W. (1969). Congestion theory and transport investment. *The American Economic Review*, 251-260.

Wu, J., Jin, X., & Horowitz, A. (2008). Methodologies for estimating vehicle queue length at metered on-ramps. *Transportation Research Record: Journal of the Transportation Research Board*(2047), 75-82.

Yan, H., & Lam, W. (1996). Optimal road tolls under conditions of queueing and congestion. *Transportation Research Part A: Policy and Practice*, *30*(5), 319-332.

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 074-0188*
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**			
1. REPORT DATE *(DD-MM-YYYY)* 24-03-2016	**2. REPORT TYPE** Master's Thesis		**3. DATES COVERED** *(From – To)* Oct 14 – Mar 16
4. TITLE AND SUBTITLE **Analysis of Military Entry Control Point Queueing**			**5a. CONTRACT NUMBER**
			5b. GRANT NUMBER
			5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Dwyer, Justin J., Major, USA			**5d. PROJECT NUMBER**
			5e. TASK NUMBER
			5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way WPAFB OH 45433-7765			**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT-ENS-MS-16-M-103
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Transportation Engineering Agency 1 Soldier Way Scott AFB, IL 62225 Phone: 618-220-5218 Email: darren.j.guttmann.civ@mail.mil ATTN: Darren J. Guttmann, PE			**10. SPONSOR/MONITOR'S ACRONYM(S)** TRANSCOM **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
12. DISTRIBUTION/AVAILABILITY STATEMENT DISTRUBTION STATEMENT A. Approved For Public Release; Distribution Unlimited			
13. SUPPLEMENTARY NOTES This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.			

14. ABSTRACT

Military Entry Control Facilities (ECFs) are unique service queues that are constrained by space, receive high peak traffic flow, and have a customer base that must receive service. Due to complexity of the interactions within the system, simulations provide input that would be impractical for quantitative experimentation. Our research examines relationships within the ECF in order to develop insights that could lead to more efficient daily operations. We focus the research on interactions that generate a queue length that would interfere with traffic flow surrounding the base. Examining the interactions between multiple arrival rates and service times as well as the layout and model of the ECF we establish criterion for Officers in Charge (OICs) to make changes within the constraints of the ECF to their operations to better serve the customers and prevent ECF traffic from interfering with the community outside the military base.

15. SUBJECT TERMS
Queueing, Entry Control Point

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Dr. Jeffery Weir, AFIT/ENS
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	**U**	**U**	**UU**	75	(937) 785-3636, x4523 Jeffery.Weir@afit.edu

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18