EMERGENCY MODELING IN TRANSPORTATION VIA QUEUING AND GAME THEORY

BY ZHE DUAN

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Industrial and Systems Engineering

Written under the direction of

Dr. Melike Baykal-Gürsoy

and approved by

New Brunswick, New Jersey

January, 2013

ABSTRACT OF THE DISSERTATION

Emergency Modeling in Transportation via Queuing and Game Theory

By ZHE DUAN

Dissertation Director: Dr. Melike Baykal-Gürsoy

In modern transportation network, emergencies on roadway and man-made emergencies in the infrastructure can incur enormous costs to society directly and indirectly. The direct costs include transportation cost due to incident delay and traffic congestion, and various risks brought upon the infrastructure by man-made emergencies. The indirect cost can include economic and psychological impacts on society.

Emergencies on roadways include accidents, disabled vehicles, adverse weather conditions, spilled loads, hazardous materials, etc. Under these cases, non-recurrent congestion will slow down the traffic flow on certain road link. In previous research, deterministic queuing models are often used for traffic flow modeling. However, due to the random environment of traffic flow, it is necessary to introduce stochastic elements into current traffic flow modeling. In our research, we use stochastic queuing models, such as Markov-modulated queuing systems, for traffic flow modeling under incidents. And non-recurrent and recurrent congestion models will be combined together to improve travel time estimation.

Man-made emergencies in the infrastructure are terrorist attacks, suicide bombings, etc. Human casualties are the major goal of intelligent adversaries. We use game theory in order to allocate first responders' resources inside the transit infrastructure to minimize human casualty. In the static zero-sum game model, we show that both the adversary and the first responder choose the same group of locations to attack and defend. In the dynamic case when the first responder is mobile while the adversary is hidden in a cell, the equilibrium solution for the first responder becomes the best patrol policy within the infrastructure. This model utilizes partially observable Markov decision process (POMDP) in which the payoff functions depend on an exogenous people flow, thus, are time varying. People flow are modeled as an open queuing network in the infrastructure. And illustration example is shown to provide insight into the competitive nature of this game between first responder and adversary.

Acknowledgements

I would like to express the deepest appreciation to my advisor, Dr. Melike Baykal-Gürsoy, whose expertise, understanding, generous guidance and support made it possible for me to work on such a rich and interesting topic. Especially when I was facing difficulties, she presented great patience and confidence in me, and without those, this work will not be possible.

My sincere thanks are extended to Dr. Tayfur Altiok, Dr. Mohsen A. Jafari, Dr. Endre Boros, and Dr. Ali Maher for serving as my degree committee members and providing me with valuable suggestions to improve this research. And special thanks to Dr. Tuğrul Özel for being on my committee after Dr. Tayfur Altiok passed away. May Dr. Altiok rest in peace.

I am also grateful to Dr. Susan L. Albin for her assistance of course selection and various aspects of my life at Rutgers. I am indebted to Dr. David W. Coit, Dr. Hoang Pham and other professors for their instructions and inspiration in my courses.

And finally, I would like to express my appreciation to Ms. Cindy Ielmini, and Ms. Helen F. Smith-Pirrello for their great assistance over my years at Rutgers. Also, I thank Mr. Hang Zhang, Mr. Zhe Liang, Mr. Dayhim Muhammad, Ms. Yada Zhu, Mr. Wei Zeng, and many other fellow graduate students, for their friendship and assistance.

Dedication

To

My Parents

My Wife

Table of Contents

List of Figures

Chapter 1
Introduction

In modern transportation network, emergencies on roadway and transit infrastructure can incur enormous cost to society directly and indirectly. The direct costs include transportation cost due to incident delay and traffic congestion, and economic cost due to infrastructure damage caused by man-made emergencies, etc. The indirect costs include economic and psychological impacts of such events. In this research, I plan to investigate both under impacts of emergencies. In the first part, I will introduce model to represent traffic flow influenced by incidents. Then I will consider the man-made emergencies on transit infrastructure. The objective of this research is to help decision makers improve performance and security level of transportation network, and then reduce the emergency burden upon transportation network.

The following two sections will briefly discuss my research on traffic flow modeling with incidents and man-made emergency in transit infrastructure.

1.1 Traffic Flow Modeling under Incidents

Due to increasing oil prices, population and economic growth, changes in the lifestyles (employees living far from their workplace), etc., the demand for transportation has increased exponentially. On the other hand, the infrastructure has not followed this trend, as prohibitive investment costs and environmental concerns make it hard to expand current highway systems. Increased traffic flow on existing roadways results in an inevitable rise in congestion. Congestion leads to delays, decreasing flow, higher fuel consumption and has negative environmental effects.

In this research, queuing models, especially Markov-modulated service queues, are utilized to analyze the traffic flow on roadways in random environment. This research will open up the analysis and control of systems from transportation networks to telecommunication networks where service goes through deterioration. Current models are not sufficient to realistically incorporate the impact of incidents on these systems. This research makes contributions i) to the queuing theory literature; ii) to the decision making in traffic congestion; and iii) in laying the framework in extended applications of queuing theory.

1.2 Man-made Emergency in Transit Infrastructure

In recent years, more and more people are choosing public transit as their first option to commute. The skyrocketing energy prices are one of the main reasons. Due to this increasing trend towards public transit usage, safety and security of the public transit infrastructure become a critical issue. We have witnessed terrible disasters happened in recent years that took the life of hundreds of people. For example, the suicide bombings in London subway train and bus system on July 7 2005, cost 52 innocent lives. The 2004 Madrid train bombings caused almost 200 casualties. These human-made attacks significantly threaten ordinary people's life and transit infrastructure, so in this research, we will analyze the risks within certain transit infrastructure and effects of some security strategies.

For a meaningful and effective emergency analysis and security improvement, one should define precise risk measures before performing any analysis for the optimization of resources and personnel. Therefore, at first, the objective of this research is to provide risk measures for various transit infrastructures under man-made attack. We will focus on the airport, train and subway infrastructures with huge pedestrian flows since the objectives of man-made attacks are to create maximum chaos. Man-made emergencies will cause significant costs, that might be direct (facility destruction, loss of human lives, etc.), and indirect (transportation delay, economic and psychological impacts, etc.). Since under severe man-made attack, human life is the primary target, in this research, we will focus on human casualties. Thus the occupancy level and pedestrian flow in a facility during an emergency become the major concern.

In this research, transit infrastructure are described in grid structure with cells of various characteristics. Then, game theory and partially observable Markov decision processes(POMDP) will be applied to model the infrastructure security problem, and to improve security strategies for first responders. And also, pedestrian flow are modeled using open queuing network. Based on this pedestrian flow, risk measures can be developed, then used to evaluate first responders security strategies.

Chapter 2

Literature Review

This literature review focuses on two areas. We first review traffic flow modeling with incidents, including both queuing and other methods. Especially, queuing models with Markov-modulated service speeds, which represent incident effects on roadway, are investigated thoroughly. Secondly, for man-made emergency, search theory and game theory models applied into infrastructure security are reviewed, and problems with various types of first responders and adversaries are investigated too.

2.1 Traffic Flow Modeling

Increasing traffic flow on existing roadways results in an inevitable rise in congestion. Congestion leads to delays, decreasing flow rate, higher fuel consumption and thus has negative environmental effects. The cost of total delay in rural and urban areas is estimated by the USDOT to be around $1 trillion per year NCTIM (2002). Researchers from widely varying disciplines have been paying attention to modeling the vehicular travel in order to improve the efficiency of the current highway systems.

2.1.1 Classical Theory

Classical traffic models are mostly based on the treatment of interacting vehicles, their statistical distribution, or their average velocity and density as a function of time and space. Main modeling approaches can be classified as microscopic (particle-based), mesoscopic (gas-kinetic), and macroscopic (fluid-dynamic, deterministic queue) (see Helbing (2001)).

Microscopic approach was developed based on driver's acceleration and deceleration behaviors due to the interaction of vehicles nearby, called as the car-following model (Richards

(1956), Gazis et al. (1959), Gazis et al. (1961)), and Newell (1961) introduced an optimal velocity model by considering a distance-dependent velocity to reflect the safety restriction. Recently, Helbing et al. (1999) proposed an intelligent driver model, taking all the aspects into account for microscopic traffic modeling such as relative velocity and safe driving distance.

Macroscopic traffic theory explains the traffic behavior in terms of average parameters, such as average velocity and average traffic density, based on continuity flow equation, in contrast to microscopic traffic modeling. Lighthill and Whitham (1955) developed the so-called L-W model on the assumption that there is no interruption to the traffic system and they obtained the fundamental diagram. Based on the continuity equation, Whitham (1974) derived the nonlinear wave equation for the propagation kinetic wave and developed the basis of shock wave theory. Whitham (1974) presented the Burgers equation by introducing a diffusion term into wave equation based on the relationship between velocity and density. Kuhne (1987) introduced a viscosity term in the Burgers equation for the negative drivers' reaction to the gradient of traffic flow and a Navier-Stokes velocity equation was obtained. Payne (1971) transformed microscopic variables to macroscopic scale and obtained the Payne's velocity equation, which described the reaction of individual vehicles to the surroundings and adaptation of individual velocity to the equilibrium velocity.

In the mesoscopic approach, driving vehicles are treated as the interacting particles in gas environment. By the continuity equation in phase space, Prigogine and Andrews (1960) modeled the acceleration and overtaking behaviors and obtained the critical density of the phase transition from free flow to congestion (see also Prigogine and Herman (1971)). Paveri-Fontana (1975) improved this model by introducing various driver types. Helbing (1995) included a term for adaptation to the road condition and his approximation managed to explain the increase in velocity variance before a phase transition.

Some other authors considered congestion around planned road work and incidents. Gas-kinetic models are introduced to describe the behaviors at bottleneck areas by Shvetsov and Helbing (1999) and Kerner (2004). Redner et al. (1994) introduced the ballistic agglomeration to model one-lane traffic flow and clustering. Lia et al. (2008) developed a traffic control plan based on empirical data, the Dynamic Late Lane Merge System (DLLMS), to improve traffic flow volume and solve potential traffic congestion problems close to work zones.

Kuhne et al. (2002) and Mahnke et al. (2005) developed a stochastic model to describe the traffic behavior and the general master equation was constructed. Combining Markov process and optimal velocity model, they concluded that the formation of traffic congestions was due to stochastic perturbation and dissolution of cluster depended on the cluster size. In analogy to nucleation mechanism, they developed a multi-cluster model on one-lane circular road.

Helbing (2003) proposed a deterministic queueing model for traffic network by dividing the road into free road and congested sections. He estimated the average traveling time and congestion pattern, assuming a fundamental diagram with linear free and congestion branches. Lammer et al. (2008) introduced a model to anticipate the queueing process at the traffic lights and estimate the waiting time. Based on different evolutions of queue length at green, yellow and red lights, they derived the hybrid dynamical equations to obtain the required green time to clear the queue. Lammer and Helbing (2008) proposed a self-organized traffic-light control at intersections with live data input that minimizes the total waiting time.

2.1.2 Stochastic Queueing Models

The arrival process in roadway traffic is modeled as a singly arriving Poisson process (Darroch et al. (1964), Tanner (1953)), and as platoons to represent the behavior of the vehicles moving between traffic signals (Alfa and Neuts (1995), Daganzo (1994), Dunne (1967), Lehoczky (1972)). Daganzo (1994) presented a cell transmission model, representing traffic on a highway with a single entrance and exit, which can be used to predict changes in the traffic pattern over time and space. Initially, queuing analysis has been mainly utilized for performance evaluation using deterministic (fluid-dynamic) models (May and Keller (1967), Newell (1971)), and synchronization of traffic-lights (Newell (1965)). Stochastic queues were studied by Cheah and Smith (1994) that explored the generality and usefulness of finite server queuing models with state dependent service rate (traveling speed) for modeling pedestrian traffic flows. As an extension, Jain and Smith (1997) used such queues for modeling and analyzing vehicular traffic flow on a roadway segment that can accommodate a finite number of vehicles. In the Jain and Smith model, arrivals are assumed to follow Poisson process (M), travel times are assumed to be generally (G) distributed random variables, and if the link is full, new arrivals should leave and find alternate paths. Consider vehicles traveling on a link as shown in Figure 2.1.

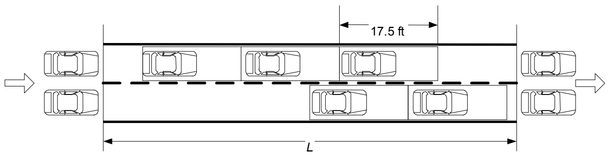

Figure 2.1: A Two-Lane Roadway Link

The space occupied by an individual vehicle on the road segment can be considered as one queuing "server", which starts service as soon as a vehicle joins the link and carries the "service" (the act of traveling) until the end of the link is reached. A "server" in this context is the moving albeit virtual vehicle-space including the safe distance to the vehicle in front. Thus, the maximum number of vehicles that can be accommodated on the link provides the number of servers in the model. Although there are several different types of vehicles utilizing the roadway, in Jain and Smith (1997) they are all assumed to be identical and considered as a passenger car equivalent. In practice, the service rate (traveling speed) is assumed to be a decreasing function of the number of vehicles on the link to represent congestion caused by traffic volume. Such a queue is called an $M/G/C/C$ model with the first M denoting Poisson arrival process, G representing general service times and finally C denoting the number of servers/roadway capacity.

Heidemann (1996) used $M/M/1$ where the second M represents the exponentially distributed service times, and $M/G/1$ queues to model uninterrupted traffic flows. Note that in all queuing models, both deterministic and stochastic, the link is considered as a point queue (or vertical queue, see Daganzo (1997)). Rakha and Zhang (2005) show the consistency of the total delay and total travel time estimates in the gas-kinetic and deterministic queuing models. While in the multi-server case, a link is separated into cells, contrary to the cell transmission model; there is no interdependence between the service times. However, we emphasize that these models have been shown to be effective in representing traffic flow. Woensel and Vandaele (2006) and Woensel et al. (2006) validate the use of queueing models via empirical data and simulation, respectively. They conclude that $M/G/1$ queueing models are the best models to describe normal traffic flow on a highway, while state-dependent $GI/G/m$ queues were more realistic for the congested traffic. Heidemann (2001) studied the transient behavior of $M/M/1$ queues to analyze non-stationary traffic flow. Vandaele et al. (2000) also used $M/M/1$ and $M/G/1$ queues to model traffic flow. Although some of these queuing models consider congestion, they all ignore the impact of random incidents on traffic flow.

2.1.3 Modeling Traffic Flow Interrupted by Incidents

Consider vehicles traveling on a roadway link, as shown in figure 2.1, which is subject to traffic incidents. During an incident, traffic deteriorates such that both the number of servers and the service rate of all servers decrease. Once an incident occurs, the incident management system sends a traffic restoration unit to clear the incident. The number of servers and the service rate of all servers are restored to their normal level when the incident is resolved. The negative impact of incident involves both congestion and reduction of road capacity. In this study, a lower service rate, affecting every server will be used to represent the impact of congestion caused by incidents. The type of service system with batch interruptions is also considered as a Markov-modulated service mechanism. Note that, the concepts in this paper also cover, the $M/M/1$ queuing model considered in Heidemann (1996), Heidemann (2001) and Vandaele et al. (2000). Consider a queue with C servers working at free speed service rate μ, subject to random interruptions of exponentially distributed durations. During interruptions, the free speed service rates of these C servers drop from μ to μ'. As soon as the interruption is cleared, the service rate of all servers are restored to μ. We assume that interruptions arrive according to a Poisson process with rate f, and the repair time is exponentially distributed with rate r. The customer arrivals are in accordance with a homogeneous Poisson process with intensity λ. The service times are assumed to be independent and identical exponentially distributed. The interruption and customer arrival processes, and the service and repair times are all assumed to be mutually independent. We would like to emphasize that the Poisson assumption for vehicle arrivals (Woensel and Vandaele (2006) and Woensel et al. (2006)) and exponential interarrival times for the incidents (Skabardonis et al. (1998)), are shown to be reasonable. Although the exponential service times may not seem realistic, in our setting, the total time to traverse a link (overall service time) is not going to be exponential. Thus, our model may be considered as having a generally distributed service time.

Queues with Service Interruptions

The study of queuing systems with service interruptions has received significant attention by researchers in the field. One type of service interruption has already been considered in the context of "vacation" queues where interruptions only happen as soon as the queue becomes empty, or a service is completed. In general, queues with server vacations are used to model non-preemptive priority systems where customers receive service according to their priority level. The server continuously serves low priority customers until higher priority customers arrive. When a high priority customer arrives, the server starts serving the new customer upon completion of the service of one, a number of, or all of the low priority customers. Thus, in these models only complete service breakdowns that happen at the instant of service completion are considered. These vacation models in steady-state are shown to exhibit the stochastic decomposition property. This fundamental result establishes the relationship between a performance measure (system size distribution, waiting time distribution, sojourn time distribution, etc.) for the queuing system with vacations, and the same performance measure for the same queuing system without vacations (Yadin and Naor (1963), Cooper (1970), Levy and Yechiali (1975), Fuhrmann and Cooper (1985), Shanthikumar (1986), Altiok (1987), Doshi (1990), Chao and Zhao (1998)).

Queues with Random Service Interruptions

In the traffic flow models that we consider in this research, incidents happen randomly, independent of service completions. The literature on queues with this type of interruptions is relatively scarce. White and Christie (1958) studied a single server queue with preemptive resume discipline, and related such queues to queues with random service interruptions. Gaver (1962) and Keilson (1962) also studied a single server queue with random interruptions. Gaver (1962) obtained the generating functions for the stationary waiting time and the number in the system in an $M/G/1$ queue. Avi-Itzhak and Naor (1963) derived the expected queue length for $M/G/1$ queue with server breakdown, also see, Halfin (1972), Fischer (1977), Federgruen and Green (1986) and Federgruen and Green (1988). Mitrani and Avi-Itzhak (1968) analyzed $M/M/C$ queue where each server may be down independently of the others for an exponential amount of time. They obtained an explicit form of the moment generating function of the queue size for one-server and two-server systems, and gave a computational procedure for cases with more than two servers. In the above models, servers fail independently of each other and failures are complete service breakdowns. $M/G/\infty$ queue with alternating renewal breakdowns was studied in Jayawardene and Kella (1996); who show that the decomposition property, a well known property of vacation type queues, holds for such queues: the stationary number of customers in the system can be interpreted as the sum of the state of the corresponding system with no interruptions and another nonnegative discrete random variable. Considering the case of partial failure, the $M/M/1$ system in a two-state Markovian environment where the arrival as well as the service process are affected, is analyzed via generating functions first by Eisen and Tainiter (1963), then by Yechiali and Naor (1971), and by Purdue (1973). Such queues, in general, in n-state Markovian environments are said to have Markovian arrival processes (MAP) and Markovian service process (MSP), and might be represented in Kendall notation as $MAP/MSP/1$. Yechiali (1973) considered the general $MAP/MSP/1$ queue. Neuts (1981) studied $M/M/1$ and briefly $M/M/C$ queues in a random environment using matrix-geometric computational methods. O'Cinneide and Purdue (1986) analyzed the n-state $MAP/MSP/\infty$ queue, where ∞ represents the number of servers as infinite. In infinite server queues, customers need not wait because a server is always available. For all these queuing models no closed form solution was given. Keilson and Servi (1993) studied a matrix $M/M/\infty$ system in which both the arrival and service processes are Markov-modulated. They obtained the generating function of the stationary number of customers in the system in terms of Kummer functions. A single server queue with Markov-modulated arrival and service processes is analyzed in (Adan and Kulkarni (2003)) and asymptotic results are presented. For the special case of $M/M/\infty$ queue with two-state Markov-modulated arrival process, they showed that the decomposition property holds, and provided the explicit solution. Baykal-Gürsoy and Xiao (2004) considered the $M/M/\infty$ system with the two-state Markov-modulated service process, e.g., $M/MSP/\infty$ queue. Using the method introduced in Keilson and Servi (1993), they proved that this model also exhibits a stochastic decomposition property, and gave the stationary distribution in closed form.

For the infinite server queue with a two-state service mechanism, Jayawardene and Kella (1996) in the case of complete breakdown, and Baykal-Gürsoy and Xiao (2004) also in the case of partial failure, are the first papers showing the validity of the decomposition property. The latter paper has created a renewed interest in infinite server queues in Markovian environment (D'Auria (2007), Yechiali (2007), and Pang and Whitt (2009)).

2.2 Man-Made Emergency in Transit Facilities

Due to increasing population, economic growth, changes in the lifestyles (employees living far from their workplace), etc., the demand for transportation has increased exponentially. The U.S. is one of the most mobile nations in the world, providing over 4 trillion miles of passenger travel annually Fed (1999). Further, every workday, about 14 million Americans use some form of public transit USG (2002). Public transit users made more than 9 billion unlinked trips using the more than 6,000 transit properties in 2001[1]. Due to this high volume, almost a third of terrorist attacks worldwide target public transit[2]. Recent attacks in New York, Madrid, London and Mumbai have forced governments to

[1] http://www.apta.com/research/stats/

[2] Congressional Research Service, Transportation Issues in the 107th Congress, Washington, D.C., 2002

devote significant time and resources to secure transportation infrastructure.

The suicide bombings in three London subway trains and one bus in the morning rush hours on July 7, 2005 affected the whole population of London. Before this terrible disaster, there were two other significant attacks on public transit facilities: the sarin gas attack in the Tokyo subway on March 20, 1995, and Madrid train bombings on March 11, 2004. Recently 7 bomb blasts hit Mumbai's commuter rail network during the rush hour on July 11, 2006, killing more than 200 passengers. The dramatic and devastating nature of these attacks point to the vulnerabilities of the public transit system.

On the other hand, public transit facilities are being used more frequently by ordinary citizens. Therefore, increasing transit usage puts public-transit facilities under much more significant security threat. The threat could be substantially reduced by analyzing the risk associated to each transit infrastructure, planning for emergency preparedness, and employing best prevention and response policies. Most transit officials accept "that the public transit systems are open, dynamic, and inherently vulnerable to terrorist attacks and cannot be closed and secured like other parts of the transportation system" (Loukaitou-Sideris et al. (2006)). There are other similar systems. Last attacks in Mumbai on a number of hotels/community centers simultaneously, have broadened the concern to other public infrastructures. Management of man-made emergencies is an important social task that has profound effects on the safety and well being of society.

Emergency management problems have the following characteristics: (1) They are stochastic. The resource (personnel, funds, etc.) availability, as well as the occurrences and characteristics of emergencies, the size of the affected population are all subject to randomness. (2) They are resource-allocation problems. Resources should be allocated wisely among individual response centers to minimize the risk.

2.2.1 Problem Description

The game we consider here is called the hide and seek game. The hider in the problem is an attacker who is trying to damage transportation facilities by putting bombs in certain critical locations, meanwhile, avoiding the security staff. The seeker, i.e. security staff, first responder, allocate available resources to the different places in facilities to search for bombs or attackers. In this research, we will consider two game forms, one step noncooperative game and another dynamic noncooperative game.

In the one-step noncooperative game, adversary is trying to cause as much damage as possible by an attack, on the other hand, the responder is trying to minimize the damages. Each place has different importance relative to its occupancy level, so damages will not be the same. Both zero-sum and nonzero-sum game can be built between the responder and adversary based upon the information shared between them.

We also consider dynamic games. In this game, the adversary is a suicide bomber, moving around the cells at every discrete time period, the responder reallocates resources during the same period. After each stage of the game, the information available for responder and adversary about the opponent will evolve.

The following sections will discuss prior work on related topics, from both search theory and security problem standpoints.

2.2.2 Search Theory

Our problem is different from the classic search game. In a typical search game(see book Alpern and Gal (2003)), the map or domain for the player to hide and search is much more important, usually it is a network, a tree or even continuous three dimensional area. So lots of research on search theory devote themselves into generating a good path or route in a network or map to find the hider. This is not what we want in this research, because our problem does not involve constant movement from one place to another within a short time period. The problem here only involve movement in dynamic case and it can only happen at certain time period. And we already generalized our problem domain from map into cells, possibly involving simplified structure, so we can focus on the game itself and pay little attention to the search map.

Gal (1979) considered search games in which the searcher moves along a continuous trajectory until he captures the hider, in either a network or a two (or more) dimensional region. The mobile and immobile hider cases were both analyzed. For some of the games, a complete solution was found, while for others upper and lower bounds were given. In Alpern and Gal (2002), the aim of the hider are not known to the searcher, and the hider can be either a cooperator or an evader. It produces a continuum of search problems, linking a zero-sum search game to a rendezvous problem. Therefore, these models provide a bridge between Search and Rendezvous Search. In Gal and Howard (2005), searching happens in two boxes. With probability p, the hider wishes to be found, and he tries to evade the search with probability $1 - p$. An associated zero-sum game is solved to develop the associated strategies for the searcher and each type of agent, and a continuous value function $v(p)$, giving the expected time until the agent is discovered, is obtained. In Alpern et al. (2008), network search games with immobile hider is considered, and starting point for searcher is not designated. The searcher's objective is to minimize the time to find the hider. They extended previous results to a wider class of networks, by showing that network is simply searchable. In this case, the optimal searching strategy is any random *Chinese Postman (CP)* path.

For dynamic situations, search theory usually considers that the movement can happen at many time points, even continuously. But in our problem, we only consider discrete time point movement, this is based on the features of our problem. For example, Thomas and Washburn (1991) give a description of a dynamic search game on the cells that need to be searched in a finite set, but the search time is continuous. Searchers are trying to find the hider as soon as possible. Probability structure is given in the problem, and expected rewards or costs for players are discounted by time. We will use similar ideas in our dynamic game to deal with the cost functions. For a thorough review of search theory, see the book by Alpern and Gal (2003) and a search theory survey by Dobbie (1968).

In one step hide and seek game, Baston and Garnaev (2000) develop a model suitable for a search game with a protector. In this game, they assume that an object has already been hidden in one of a finite number of cells. Then, a protector and a searcher will allocate their own resources into cells respectively. The protector wants to protect the object from being found, and the searcher wants to detect the object. Both players have exactly the same information about the probability distribution of where the object is hidden. Then, they show the existence of a Nash equilibrium under certain assumptions. A heuristic algorithm is developed and an example is given. Alpcan and Basar (2006) present two-player zero-sum stochastic game which models the interaction between malicious attackers to a system and the Intrusion Detection System(IDS). Sensor network is used to capture the attack information for IDS, and several information cases are discussed. The methods used are based on Markov decision process and Q-learning. In Hohzaki (2007), a search allocation game (SAG) is analyzed. The resources a searcher allocates can last a certain period of time and can also affect other areas around the drop points. Linear programming formulation is used to solve the problem.

For the dynamic case, Hespanha et al. (2000) give a sophisticated framework for dynamic cell-search game, different from Thomas and Washburn (1991). This paper considers time as discrete time points, and it is similar to our problem. A partial information Markov game is considered. A good framework for dynamic game is constructed, but the objectives for pursuer and evader are greedy, which means that they only consider a static game for the current situation. Under this framework, they consider one step game first, one step game is transferred into a matrix game that is easy to solve. For dynamic case, simulation is proposed to solve the problem. Actually, there are only a few articles dealing with the dynamic case. Alpern et al. (2009) considers a facility as a graph, and the attacker will choose several consecutive time periods, uninterrupted by patroller, to commit an attack. Patroller can follow any path on the graph during a given time period. It is assumed that, when the patroller and the attacker are at the same node at the same time, the attacker will be found. The patroller's objective is to minimize the time to detect the attacker. Optimal patrolling strategies are determined for various classes of graphs, and used to support decisions on facility reinforcement. In Jotshi and Batta (2008), a dynamic hide and seek game on a network is discussed, and an algorithm is also developed. In this game, the attacker is immobile, and the searcher tries to select paths in the network to detect the attacker as soon as possible.

2.2.3 Security Problems

There are also some papers focusing on the infrastructure security related issues. To detect and prevent the terrorist operations, a main challenge is the fusion of information, from different sources (U.S. or foreign), and of different types (signals, human intelligence, etc.). In Paté-Cornell (2002a), the author focus on merging contents of signals from various sources and types, and Bayesian approach is used to update the probability of an event, as new signals or information are received. This method allows effective fusion of signals and information, and account for both types of error probabilities (false positives and false negatives). In Paté-Cornell (2002b), the overarching model for setting priorities among homeland security countermeasures was presented. First, diverse kinds of information are gathered, then based on probabilistic risk analysis, decision analysis, and game theory, priorities among countermeasures were set. The dynamic competition between the U.S. goverment and terrorist are also considered.

In Pita et al. (2008), a game theoretic model is used to secure Los Angeles International Airport. A software decision support system called ARMOR is developed, that casts the police patrolling/monitoring problem as a Bayesian Stackelberg game. ARMOR has been deployed at the Los Angeles International Airport to randomly assign checkpoints on the roadways entering the airport and canine patrol routes within the airport terminals (see Paruchuri et al. (2005), Paruchuri et al. (2006), Paruchuri et al. (2007), and Paruchuri et al. (2008)). In this game, leaders are the police, and followers are the attackers. First, police set up checkpoints, and then, the attackers choose their actions based on the current set of checkpoints. However, in this research, we will consider a game between a responder and an adversary with simultaneous actions. Other recent papers include Nie et al. (2009a) and Nie et al. (2009b), where a passenger classification problem is analyzed, based on their scores at the checkpoints. The same authors also discuss the optimal placement of suicide bomber detectors within a grid structure Nie et al. (2007). Note that, Nie et al. (2007), Nie et al. (2009a) and Nie et al. (2009b) only involve single controller optimization models.

Chapter 3

Queuing Models and their Application to Roadway Traffic

This chapter will present several stochastic queuing systems with Markov-modulated arrival or Markov-modulated service processes, which later will possibly be applied into traffic flow modeling with incidents. First, a $M/M/C$ queuing system with finite number of servers and Markov-modulated service processes are analyzed. Secondly, stochastic decomposition results for stationary number in the system are presented for a single server queue with Markov-modulated arrival and service processes. Third, a queue with two service speeds represented by generally distributed service time is studied, service completion time for customers in such a queue is studied and Laplace transform of the completion time is obtained. Finally, a queuing system with Markov-modulated service process is applied into traffic flow modeling with incidents, and the results are validated through comparison with simulation results.

3.1 $M/M/C$ Queues with Markov-modulated Service Processes

Motivated by the need to study traffic flow affected by incidents we consider $M/M/C$ queuing system where servers operate in a Markovian environment. When a traffic incident happens, either all lanes or part of a lane is closed to the traffic. As such, we model these interruptions either as complete service disruptions where none of the servers work or partial failures where all servers work at some reduced service rate. We analyze the system with multiple failure states in steady state and present a scheme to obtain the stationary number of vehicles on a link. The special case of single breakdown case is further analyzed and performance measures in close form are obtained.

3.1.1 Mathematical Model

Consider a road link as shown in Figure 2.1 with C servers that are subject to random system interruptions of exponentially distributed durations. We assume that there is buffer space available in front of the link so that the vehicles that cannot get a server can wait for service. As the most general case we consider $M/M/C$ queues with n types of server states. The server states are denoted as $S_1, \ldots, S_n$ that have associated service rates $\mu_1, \ldots, \mu_n$ respectively. Service times are assumed to be independent and identically distributed (i.i.d.)

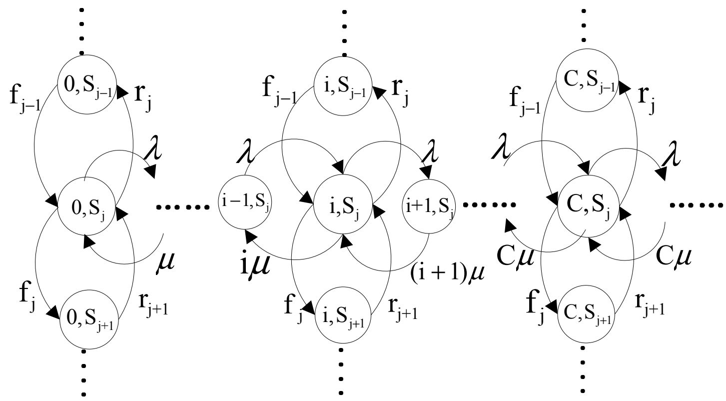

Figure 3.1: State transitions for $M/M/C$ queue with deteriorating service

exponentials. The vehicle arrivals are in accordance with a homogeneous Poisson process with intensity λ irrespective of the server state. Movements between server states include only the moves to the adjacent states as one state example shown in Figure 3.1. The state transitions at the boundary states could be presented respectively. This example represents the case where S_1 corresponds to the normal state and the server state deteriorates to the next state with each interruption and the previous server state is restored with each clearance action. At server state S_j, the interruptions arrive according to a Poisson process with rate f_j for $j = 1, \ldots, n-1$, and the clearance times are i.i.d. exponentials with rate r_j for $j = 2, \ldots, n$. Here $f_n = 0$ and $r_1 = 0$. The model considered above also includes the case that from the normal state with different types of failures the server state goes to either the moderate failure state or to the severe failure state depending on the severity of the incident. The clearance times of these incidents also depend on the incident type. Figure 3.2 presents this case where the server states are represented as N corresponding to the normal road conditions, M corresponding to the moderate incident and F corresponding to the severe incident conditions. The interruption and vehicle arrival processes, and the service and clearance times are all assumed to be mutually independent. In Figure 3.1 and 3, $0 < i < C$ and $k \geq C$. Note that, for $C = 1$ the system considered here is a special case of the $MAP/MSP/1$ queue studied in Yechiali (1973), since in the later one the server state can go into any of the other server states.

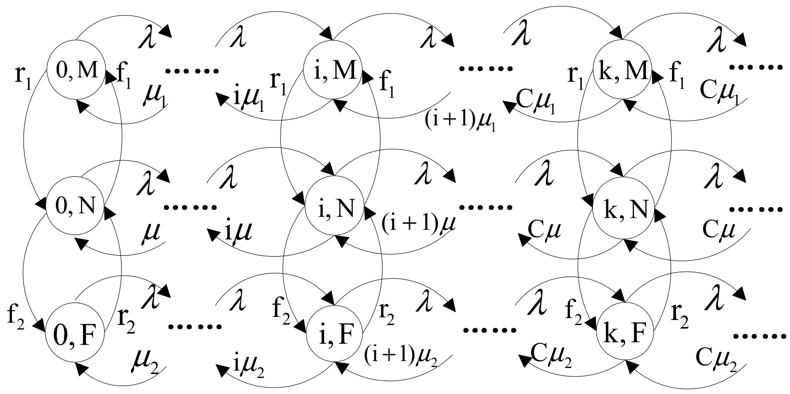

Figure 3.2: State transitions for M/MSP/C queue with three server states

The stochastic process $\{X(t), Y(t)\}$ describes the state of the link at time t, where $X(t)$ denotes the number of vehicles on the link at t, and $Y(t)$ denotes the server state.

Balance Equations

The steady-state balance equations are given below,

State S_1,

$$(\lambda + f_1)P_{0,S_1} = \mu_1 P_{1,S_1} + r_2 P_{0,S_2} \tag{3.1}$$

$$(\lambda + f_1 + i\mu_1)P_{i,S_1} = (i+1)\mu_1 P_{i+1,S_1} + r_2 P_{i,S_2} + \lambda P_{i-1,S_1} \quad (1 \leq i \leq C-1) \tag{3.2}$$

$$(\lambda + f_1 + C\mu_1)P_{i,S_1} = C\mu_1 P_{i+1,S_1} + r_2 P_{i,S_2} + \lambda P_{i-1,S_1} \quad (i \geq C) \tag{3.3}$$

State S_n,

$$(\lambda + r_n)P_{0,S_n} = \mu_n P_{1,S_n} + f_{n-1}P_{0,S_{n-1}} \tag{3.4}$$

$$\begin{aligned}(\lambda + r_n + i\mu_n)P_{i,S_n} &= (i+1)\mu_n P_{i+1,S_n} + f_{n-1}P_{i,S_{n-1}} \\ &\quad +\lambda P_{i-1,S_n} \quad (1 \leq i \leq C-1)\end{aligned} \tag{3.5}$$

$$\begin{aligned}(\lambda + r_n + C\mu_n)P_{i,S_n} &= C\mu_n P_{i+1,S_n} + f_{n-1}P_{i,S_{n-1}} \\ &\quad +\lambda P_{i-1,S_n} \qquad (i \geq C)\end{aligned} \tag{3.6}$$

State S_j $(j = 2, ...n-1)$,

$$\begin{aligned}(\lambda + f_j + r_j)P_{0,S_j} &= \mu_j P_{1,S_j} + r_{j+1}P_{0,S_{j+1}} \\ &\quad +f_{j-1}P_{0,S_{j-1}}\end{aligned} \tag{3.7}$$

$$\begin{aligned}(\lambda + f_j + r_j + i\mu_j)P_{i,S_j} &= (i+1)\mu_j P_{i+1,S_j} + r_{j+1}P_{i,S_{j+1}} \\ &\quad +f_{j-1}P_{i,S_{j-1}} + \lambda P_{i-1,S_j} \qquad (1 \leq i \leq C-1)\end{aligned} \tag{3.8}$$

$$\begin{aligned}(\lambda + f_j + r_j + C\mu_j)P_{i,S_j} &= C\mu_j P_{i+1,S_j} + r_{j+1}P_{i,S_{j+1}} \\ &\quad +f_{j-1}P_{i,S_{j-1}} + \lambda P_{i-1,S_j} \qquad (i \geq C)\end{aligned} \tag{3.9}$$

Generating Function

We will use the partial generating functions,

$$G_j(z) = \sum_{i=0}^{\infty} z^i P_{i,S_j},$$

to write the overall generating function as,

$$G(z) = \sum_{j=1}^{n} G_j(z). \tag{3.10}$$

By multiply the balance equations with z^i, and summing all equations for state S_j, we obtain,

$$\begin{aligned}
&[\lambda z(1-z) + f_1 z + C\mu_1(z-1)]G_1(z) - r_2 z G_2(z) \\
&\quad = \sum_{i=0}^{C-1} (z-1)(C-i)\mu_1 P_{i,S_1} z^i, && (3.11) \\
&[\lambda z(1-z) + r_n z + C\mu_n(z-1)]G_n(z) - f_{n-1} z G_{n-1}(z) \\
&\quad = \sum_{i=0}^{C-1} (z-1)(C-i)\mu_n P_{i,S_n} z^i, && (3.12) \\
&[\lambda z(1-z) + r_j z + f_j z + C\mu_j(z-1)]G_j(z) - r_{j+1} z G_{j+1}(z) \\
&\quad -f_{j-1} z G_{j-1}(z) \\
&\quad = \sum_{i=0}^{C-1} (z-1)(C-i)\mu_j P_{i,S_j} z^j, \qquad (j = 2, 3, ...n-1). && (3.13)
\end{aligned}$$

In these n equations, there are nC unknown probabilities, and we can use the balance equations to reduce them to only n unknowns, P_{0,S_j}, for $j = 1, \ldots, n$.

Proposition 1 *For the n-state M/MSP/C queue, the stability condition is,*

$$\lambda < \frac{\sum_{j=1}^{n} C\mu_j \cdot \left(\prod_{i=1}^{j-1} f_i \cdot \prod_{i=j+1}^{n} r_i\right)}{\sum_{k=1}^{n} \left(\prod_{i=1}^{k-1} f_i \cdot \prod_{i=k+1}^{n} r_i\right)}. \tag{3.14}$$

□

Proof We know that $G_j(1)$ corresponds to the probability that the system is in server state S_j in the long run. If we aggregate all states (i, S_j) in server state S_j as a mega state, then we can easily obtain the long-run probability that the system is in state S_j as,

$$G_j(1) = \frac{\prod_{i=1}^{j-1} f_i \cdot \prod_{i=j+1}^{n} r_i}{\sum_{k=1}^{n} \left(\prod_{i=1}^{k-1} f_i \cdot \prod_{i=k+1}^{n} r_i\right)}. \tag{3.15}$$

Thus, the stability condition for this system is,

$$\lambda < \sum_{j=1}^{n} C\mu_j \cdot G_j(1), \tag{3.16}$$

giving the required inequality 3.14. ■

In the next part, we will show that the denominator of $G(z)$ has $n-1$ distinct real roots that are unstable. These poles have to be eliminated by the zeros of $G(z)$, thus, giving $n-1$ equations in addition to $G(1) = 1$ to solve for the n unknowns. To this end, following the notation and the method introduced in Mitrani and Avi-Itzhak (1968), let,

$$
\begin{aligned}
g_1(z) &= \lambda z(1-z) + f_1 z + C\mu_1(z-1), \\
g_j(z) &= \lambda z(1-z) + r_j z + f_j z + C\mu_j(z-1), \\
&\qquad\qquad\qquad\qquad (j = 2, 3, ...n-1), \\
g_n(z) &= \lambda z(1-z) + r_n z + C\mu_n(z-1).
\end{aligned}
$$

Further let,

$$
A(z) = \begin{bmatrix} g_1(z) & -r_2 z & 0 & \cdots\cdots & 0 & 0 \\ -f_1 z & g_2(z) & -r_3 z & \cdots\cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots\cdots & -f_{n-1} z & g_n(z) \end{bmatrix}.
$$

$$
\vec{b}(z) = \begin{bmatrix} \sum_{i=0}^{C-1}(C-i)\mu_1 P_{i,S_1} z^i \\ \sum_{i=0}^{C-1}(C-i)\mu_2 P_{i,S_2} z^i \\ \vdots \\ \sum_{i=0}^{C-1}(C-i)\mu_n P_{i,S_n} z^i \end{bmatrix}, \quad \vec{G}(z) = \begin{bmatrix} G_1(z) \\ G_2(z) \\ \vdots \\ G_n(z) \end{bmatrix}.
$$

Equations 3.11-3.13 can be written in the following compact form,

$$
A(z)\vec{G}(z) = (z-1)\vec{b}(z).
$$

It is easy to show that $A(z)$ has a singularity at $z = 1$. Since $|A(z)|$ is a polynomial (degree of $2n$) in z, we may write,

$$
|A(z)| = (z-1)Q(z), \tag{3.17}
$$

where $Q(z)$ is a polynomial of degree $2n-1$. Using Cramer's rule, for all values of z at which $A(z)$ is nonsingular, we have,

$$
|A(z)|\, G_j(z) = |A_j(z)|(z-1), \qquad i = 1, 2, ...n. \tag{3.18}
$$

Here, matrix $A_j(z)$ is obtained by replacing the jth column of $A(z)$ with $\vec{b}(z)$. The equation 3.18 must hold for all $z \in [0,1]$ since all functions in 3.18 are continuous and bounded in $[0,1]$, in addition the polynomial $|A(z)|$ may have only a finite number of roots in this interval.

The following Lemma would be needed in the proof of Theorem 1.

Lemma 1 $Q(1) > 0$. □

Proof Using 3.17, equation 3.18 may be rewritten as,

$$Q(z)G_j(z) = |A_j(z)| \qquad j = 1, 2, ...n. \tag{3.19}$$

Taking the derivative of equation 3.17 with respect to z, then letting $z = 1$ gives,

$$Q(1) = \left.\frac{d|A(z)|}{dz}\right|_{z=1}. \tag{3.20}$$

Let $\vec{a}_j(z)$ be the jth row vector of matrix $A(z)$. We know that,

$$\left.\frac{d|A(z)|}{dz}\right|_{z=1} = \begin{vmatrix} \vec{a}'_1(1) \\ \vec{a}_2(1) \\ \vdots \\ \vec{a}_n(1) \end{vmatrix} + \begin{vmatrix} \vec{a}_1(1) \\ \vec{a}'_2(1) \\ \vdots \\ \vec{a}_n(1) \end{vmatrix} + \cdots\cdots + \begin{vmatrix} \vec{a}_1(1) \\ \vec{a}_2(1) \\ \vdots \\ \vec{a}'_n(1) \end{vmatrix}. \tag{3.21}$$

Using the definition of $A(z)$, we obtain,

$$\begin{vmatrix} \vec{a}_1(1) \\ \vdots \\ \vec{a}'_j(1) \\ \vdots \\ \vec{a}_n(1) \end{vmatrix} = (C\mu_j - \lambda) \cdot \prod_{i=1}^{j-1} f_i \prod_{i=j+1}^{n} r_i.$$

Then, from 3.20 and 3.21, we have,

$$Q(1) = \sum_{j=1}^{n} (C\mu_j - \lambda) \cdot \prod_{i=1}^{j-1} f_i \prod_{i=j+1}^{n} r_j. \tag{3.22}$$

The result follows from Proposition 1. ■

From equations 16, 17, and the definition of G(z), clearly, the generating function of the number of customers in the system is,

$$G(z) = \frac{\sum_{j=1}^{n} |A_j(z)|}{Q(z)}. \tag{3.23}$$

Letting $z = 1$ in equation 3.19 gives,

$$|A_j(1)| = Q(1)G_j(1) \qquad j = 1, 2, ..., n. \tag{3.24}$$

$Q(1)$ and $G_j(1)$ are given by equations 3.22 and 3.15. The $n-1$ of the n equations in 3.24 are all redundant since multiplying $\frac{f_j}{r_{j+1}}$ to the jth equation of 3.24 will give the $(j+1)st$ equation. On the other hand, since $|A_j(z)|$ must be zero whenever $Q(z) = 0, 0 \leq z < 1$, the next theorem proves that the generating function has $n-1$ unstable poles. Thus, the remaining equations will be obtained by equating the nominator of the generating function to zero at these unstable poles.

Theorem 1 *The polynomial $Q(z)$ exactly has $n-1$ distinct real roots in the interval (0,1).* □

By Lemma 1, we have $Q(1) > 0$. Then, the proof follows from Mitrani and Avi-Itzhak (1968) since A matrix has a similar structure as the model considered in Mitrani and Avi-Itzhak (1968).

3.1.2 Special Cases

In this section we consider the case with a single failure state. Thus this case reduces to the queue with two-state Markov-modulated service process considered in Baykal-Gürsoy et al. (2009). Since there is only one failure state we will use the failure and repair rates without any subscript as f and r. The service rate under normal conditions is denoted as μ and when the system failure occurs the service rate reduces to μ'. Also, let N denote the normal state, and F denote the failure state. Baykal-Gürsoy et al. (2009) obtained the generating function as,

$$G(z) = \frac{\begin{array}{c}[\lambda z(1-z) + C\mu'(z-1) + (r+f)z]\sum_{i+0}^{C-1}\mu z^i P_{i,N} \\ +[\lambda z(1-z) + C\mu(z-1) + (r+f)z]\sum_{i+0}^{C-1}\mu' z^i P_{i,F}\end{array}}{\begin{array}{c}\lambda^2 z^3 - (\lambda^2 + C\lambda\mu + \lambda f + C\lambda\mu' + \lambda r)z^2 \\ +(C\lambda\mu + C\lambda\mu' + C^2\mu\mu' + Cf\mu' + C\mu r)z - C^2\mu\mu'\end{array}}. \tag{3.25}$$

In this case, the stability condition is given as

$$\lambda < \frac{r}{r+f}C\mu + \frac{f}{r+f}C\mu'.$$

By finding the roots of the denominator one of which is inside $(0,1)$, we can obtain all of the unknown probabilities in the generating function. The expected number in the system is then obtained from $G'(1)$. Finding the single unstable root parametrically so that the generating function is obtained in closed form is elusive. Thus, in the case of partial failures $\mu' > 0$, this procedure is numerical. As an example, consider an $M/M/3$ queue subject to interruptions that reduce the service rate to a half of its normal value. The expected number of vehicles on the link versus the service rate μ is plotted in Figure 3.3. In this figure, $\lambda = 1.0, \mu = 2\mu'$, and f and r take some particular values. It can be seen

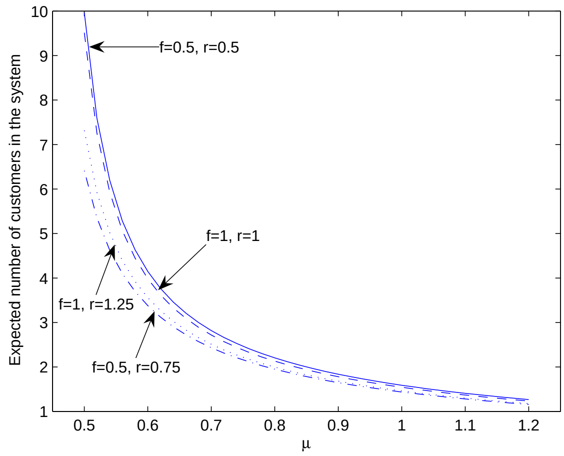

Figure 3.3: Case: $\lambda = 1.0, \mu = 2\mu'$

from Figure 3.3 that the number of vehicles on the link decreases as the service rate increases. Note that, for the two top most cases the stability condition requires that $\mu > 4/9\lambda$, since $r/(r+f) = 1/2$. If the service rate does not change, higher incident frequency or slower clearance rate would lead to more vehicles on the link. Figure 3.4 is used to show the effect of μ', where μ is fixed at 2 and μ' is increased independently. Similar to Figure 3.3, we let $\lambda = 1$, and f and r vary over a range. It can be seen that the expected number of customers also decrease as μ' increases, more significantly than in Figure 3.3, where μ and μ' increase simultaneously. Clearly, the stationary number of vehicles on the link when no incident occurs will constitute the lower bound.

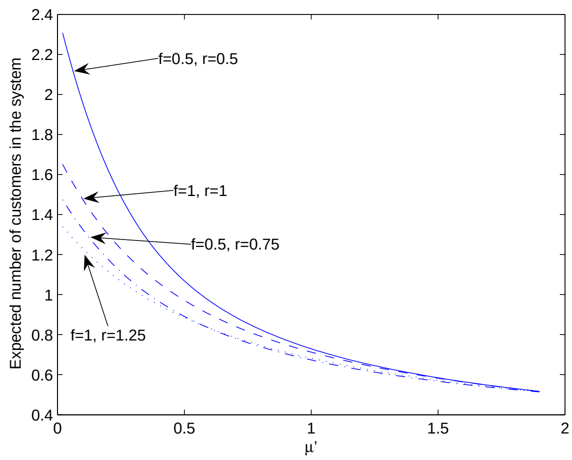

Figure 3.4: Case: $\lambda = 1.0, \mu = 2.0$

On the other hand, closed form solutions can be obtained for the complete breakdown case as will be shown in the next part.

$M/M/C$ Queues with System Breakdowns and Repairs ($\mu' = 0$)

For $M/M/C$ queues with complete server breakdowns and repairs($\mu' = 0$), the stability condition for this kind of system will be,

$$\lambda < \frac{r}{r+f} C\mu.$$

As we have said before, the $M/M/1$ queue under complete server breakdown has been studied by Mitrani and Avi-Itzhak (1968), and Gaver (1962). The generating function in this case can be written as,

$$G(z) = \frac{\frac{r}{r+f}(1-\rho\frac{r+f}{r})(1-\lambda/\delta z)}{(1-\rho z)(1-\lambda/\delta z) - \frac{f}{\delta}}, \tag{3.26}$$

where $\rho\frac{\lambda}{C\mu}$ with $C = 1$ and $\delta = \lambda + r + f$. Since the generating function of regular M/M/1 queue without interruptions is $G(z) = \frac{1-\rho}{1-\rho z}$, we see from 3.26 that contrary to Doshi (1990) this system does not exhibit the stochastic decomposition property.

For the $M/M/2$ queue, the generation function is given in (Baykal-Gürsoy et al. (2009)) as,

$$G(z) = \frac{\frac{r}{r+f}(1-\rho\frac{r+f}{r})(1-\lambda/\delta z)}{(1-\rho z)(1-\lambda/\delta z) - \frac{f}{\delta}} \cdot \frac{C+\eta z}{C+\eta}, \tag{3.27}$$

where $\eta = \frac{\lambda}{\mu}(\frac{\lambda+r+f}{\lambda+r})$.

Finally, we will consider the above $M/M/3$ queue with complete breakdowns. In this case, the generating function is,

$$G(z) = \frac{[\lambda(1-z)+r+f](3\mu P_{0,N} + 2\mu z P_{1,N} + \mu z^2 P_{2,N})}{\lambda^2 z^2 - (\lambda^2 + \lambda r + f\lambda + 3\lambda\mu)z + 3\mu(\lambda + r)}. \tag{3.28}$$

Since $G(1) = 1$, equation 3.28 provides,

$$3P_{0,N} + 2P_{1,N} + P_{2,N} = -\frac{\lambda}{\mu} + \frac{3r}{r+f}. \tag{3.29}$$

Using the balance equations, we evaluate,

$$\begin{aligned} P_{0,F} &= \frac{f}{\lambda + r} P_{0,N}; \\ P_{1,N} &= \frac{\lambda(\lambda + r + f)}{\mu(\lambda + r)} P_{0,N} = \eta P_{0,N}; \\ P_{1,F} &= \frac{f}{\lambda + r} P_{1,N} + \frac{\lambda f}{(\lambda + r)^2} P_{0,N} \\ &= \frac{\lambda f(\lambda + r + f + \mu)}{\mu(\lambda + r)^2} P_{0,N}; \\ P_{2,N} &= \frac{\lambda^2(\lambda + r + f)^2 + f\mu\lambda^2}{2\mu^2(\lambda + r)^2} P_{0,N} \\ &= \left(\frac{1}{2}\eta^2 + \frac{f\lambda^2}{2\mu(\lambda + r)^2}\right) P_{0,N}. \end{aligned}$$

By substituting the above probabilities in equation 3.29, we obtain,

$$P_{0,N} = \frac{3\frac{r}{r+f}(1 - \rho\frac{r+f}{r})}{3 + 2\eta + \frac{1}{2}\eta^2 + \frac{f\lambda^2}{2\mu(\lambda+r)^2}}.$$

We also have,

$$\begin{aligned} 3P_{0,N} + 2zP_{1,N} + z^2P_{2,N} = & \frac{3r}{r+f}\left(1 - \rho\frac{r+f}{r}\right) \\ & \cdot \frac{3 + 2\eta z + (\frac{1}{2}\eta^2 + \frac{f\lambda^2}{2\mu(\lambda+r)^2})z^2}{3 + 2\eta + \frac{1}{2}\eta^2 + \frac{f\lambda^2}{2\mu(\lambda+r)^2}}. \end{aligned}$$

Thus, the final form of generating function is,

$$G(z) = \frac{(1 - \frac{\lambda}{\delta}z)\frac{r}{r+f}(1 - \rho\frac{r+f}{r})}{(1 - \rho z)(1 - \frac{\lambda}{\delta}z) - \frac{f}{\delta}} \cdot \frac{3 + 2\eta z + (\frac{1}{2}\eta^2 + \frac{f\lambda^2}{2\mu(\lambda+r)^2})z^2}{3 + 2\eta + \frac{1}{2}\eta^2 + \frac{f\lambda^2}{2\mu(\lambda+r)^2}}. \tag{3.30}$$

As the number of servers increases, this system converges to an infinite server queue. Infinite server queues are more amenable to analysis even in the case of partial failures. It is shown in (Baykal-Gürsoy and Xiao (2004)), that the generating function has the following closed form,

$$G(z) = e^{(\lambda/\mu)(z-1)}\Psi(z), \tag{3.31}$$

where $\Psi(z)$ is the generating function of the mixture of two independent random variables. Depending on the value of μ' these two random variables are either in the form of generalized negative binomial (for the complete breakdown case) or Poisson with means distributed as truncated beta (for the partial failure case). Clearly, this system (3.31) exhibits the decomposition property.

3.1.3 Summary

The analysis of $M/MSP/C$ queue with n server states presented in this section clearly indicates that explicit solutions for the general case would be difficult to obtain. But, numerical methods as shown, could always be applied. For the special case of system breakdowns and repairs ($\mu' = 0$), the explicit solutions are obtained. Because breakdowns might happen during the service time of customers, the service completion time, i.e., dwell time on a link, will not remain exponential. So, the system we are solving could be considered as an $M/G/C$ queue with a special service structure. There is little known about $M/G/C$ queues that the closed form solutions obtained in Baykal-Gürsoy et al. (2009) and this section will help to fill this gap.

3.2 Stochastic Decomposition for MAP & MSP, and Retrial Queues

Stochastic decomposition property is common for many queueing systems. Decomposing a performance characteristic for the system into two or more simpler and independent random variables can make us understand the system more easily. This section will show you stochastic decomposition results for two useful queueing systems. These systems are utilized prominently in traffic problems and wireless systems.

Stochastic decomposition results have been found in many queueing models. Only several papers present very well on the complete distributions of the component random variables. The paper Baykal-Gürsoy and Xiao (2004) is one of them, it gives stochastic decomposition results for $M/M/\infty$ queues with Markov-modulated service rates, and explains the distributions of the random variables well. This paper will discuss more general cases with Markov-modulated arrival and service.

Next, I will give stochastic decomposition results on Markov-modulated arrival and service(2 states) $M/M/\infty$ queue, and $M/M/1/1$ retrial queue. Also, numerical results are given to show the comparison of probability mass functions between the component random variables.

3.2.1 Markov-Modulated Arrival and Service Queues

In this section, we consider an $M/M/\infty$ queueing system, the arrival process is Markov-modulated Poisson process, and the service process for all servers is also Markov-modulated(2 states). Then, we have the transition rate diagram as follows,

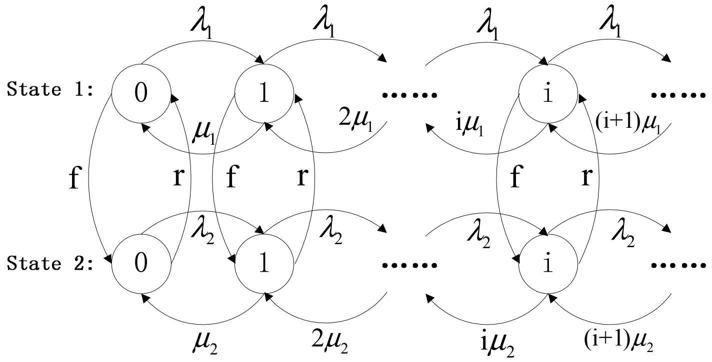

Figure 3.5: Transition Rate Structure for Markov-modulated $M/M/\infty$

Then we have the similar transition structure like in Keilson and Servi (1993). And we assume that $\mu_1, \mu_2 > 0$ and $\frac{\lambda_1}{\mu_1} < \frac{\lambda_2}{\mu_2}$ without loss of generality. So, we can see state 2 is the partial failure state, and we use f to denote the failure rate, and r to be repair rate. So,

$$\lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}; \qquad \mu = \begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix}; \qquad Q = \begin{bmatrix} -f & f \\ r & -r \end{bmatrix} \tag{3.32}$$

Then, by Keilson and Servi (1993), we can get the probability generating function of the number in the system,

$$\begin{aligned}\pi(u) = exp\left[\tfrac{\lambda_1}{\mu_1}(u-1)\right] \cdot \Big[\tfrac{f\mu_2 + r\mu_1}{\mu_2(f+r)} \cdot M\left(\tfrac{f}{\mu_1},\ \tfrac{f}{\mu_1} + \tfrac{r}{\mu_2},\ \left(\tfrac{\lambda_2}{\mu_2} - \tfrac{\lambda_1}{\mu_1}\right)(u-1)\right) \\ + \tfrac{r(\mu_2 - \mu_1)}{\mu_2(f+r)} \cdot M\left(\tfrac{f}{\mu_1},\ \tfrac{f}{\mu_1} + \tfrac{r}{\mu_2} + 1,\ \left(\tfrac{\lambda_2}{\mu_2} - \tfrac{\lambda_1}{\mu_1}\right)(u-1)\right)\Big]\end{aligned} \tag{3.33}$$

Using equation 13.4.3 in Abramowitz and Stegun (1964), we can derive another form for this probability generating function,

$$\begin{aligned}\pi(u) = exp\left[\tfrac{\lambda_1}{\mu_1}(u-1)\right] \cdot \Big[\tfrac{f\mu_2 + r\mu_1}{\mu_1(f+r)} \cdot M\left(\tfrac{f}{\mu_1},\ \tfrac{f}{\mu_1} + \tfrac{r}{\mu_2},\ \left(\tfrac{\lambda_2}{\mu_2} - \tfrac{\lambda_1}{\mu_1}\right)(u-1)\right) \\ + \tfrac{f(\mu_1 - \mu_2)}{\mu_1(f+r)} \cdot M\left(\tfrac{f}{\mu_1} + 1,\ \tfrac{f}{\mu_1} + \tfrac{r}{\mu_2} + 1,\ \left(\tfrac{\lambda_2}{\mu_2} - \tfrac{\lambda_1}{\mu_1}\right)(u-1)\right)\Big]\end{aligned} \tag{3.34}$$

Theorem 2 *The stationary number of customers in the system*($\mu_1, \mu_2 > 0$ *and* $\frac{\lambda_1}{\mu_1} < \frac{\lambda_2}{\mu_2}$), X, *has the form*

$$X = X_\varphi + Y \tag{3.35}$$

where X_φ *and* Y *are independent,* X_φ *is a Poisson random variable with mean* $\varphi = \lambda_1/\mu_1$, *and*

$$P\{Y = n\} = pP\{Y_1 = n\} + (1-p)P\{Y_2 = n\} \tag{3.36}$$

(i) When $\mu_1 < \mu_2$, *using equation 3.33,* $p = (f\mu_2 + r\mu_1)/(\mu_2(f+r))$, *and* Y_1 *and* Y_2 *are conditionally Poisson distributed with random means that have truncated beta distributions* $B(a, b, -2\rho*)$ *and* $B(a, b+1, -2\rho*)$, *respectively, where,*

$$a = \frac{f}{\mu_1}, \qquad b = \frac{f}{\mu_1} + \frac{r}{\mu_2}, \qquad \rho* = \frac{1}{2}\left(\frac{\lambda_1}{\mu_1} - \frac{\lambda_2}{\mu_2}\right) \tag{3.37}$$

□

For probability mass function of Y_1 *and* Y_2, *please see Baykal-Gürsoy and Xiao (2004) for detail.*
(ii) When $\mu_1 > \mu_2$, *using equation 3.34,* $p = (f\mu_2 + r\mu_1)/(\mu_1(f+r))$, *and* Y_1 *and* Y_2 *are conditionally Poisson distributed with random means that have truncated beta distributions* $B(a, b, -2\rho*)$ *and* $B(a+1, b+1, -2\rho*)$, *respectively.* a, b *and* $\rho*$ *are same as in the above case.*

PROOF Since *Kummer's* function $M(a,b,c(u-1))$ is the generating function of Poisson random variable Y randomized by truncated beta $B(a,b,c)$, the above results is easy to get. The proof is very similar to Baykal-Gürsoy and Xiao (2004). ■

From Theorem 2, we can get the expectation and variance of number of customers in the system.

Corollary 1 *For the number of customers in the system ($\mu_1, \mu_2 > 0$ and $\frac{\lambda_1}{\mu_1} < \frac{\lambda_2}{\mu_2}$), X, its expectation in steady state is given as*

$$E(X) = \frac{\lambda_1}{\mu_1} + \frac{(\lambda_2\mu_1 - \lambda_1\mu_2)f(f+r+\mu_1)}{\mu_1(f+r)(f\mu_2 + r\mu_1 + \mu_1\mu_2)}. \tag{3.38}$$

its variance is derived as

$$\begin{aligned} Var(x) \quad &= \frac{\lambda_1}{\mu_1} + \frac{f(\lambda_2\mu_1 - \lambda_1\mu_2)(f+r+\mu_1)}{\mu_1(f+r)(f\mu_2 + r\mu_1 + \mu_1\mu_2)} \\ &+ \frac{fr(\lambda_2\mu_1 - \lambda_1\mu_2)^2 \cdot [(f+r)^2 + 2(f\mu_2 + r\mu_1 + \mu_1\mu_2) + f\mu_1 + r\mu_2]}{(f+r)^2(f\mu_2 + r\mu_1 + \mu_1\mu_2)^2 \cdot (f\mu_2 + r\mu_1 + 2\mu_1\mu_2)} \end{aligned} \tag{3.39}$$

□

PROOF Take 1st derivative and 2nd derivative of the generating function, then let $u = 1$, then use formulas in Abramowitz and Stegun (1964), we can get the 1st and 2nd moments of x. Finally, we can have the expected number and its variance. ■

And we have the following two special cases:

CASE 1 - Markov-modulated Arrival only ($\mu_1 = \mu_2 = \mu$): Then, since $\lambda_1/\mu_1 < \lambda_2/\mu_2$, we have $\lambda_1 < \lambda_2$. The generating function for the number of customers in the system will be,

$$\pi(u) = exp\left[\frac{\lambda_1}{\mu}(u-1)\right] \cdot M\left(\frac{f}{\mu},\ \frac{f+r}{\mu},\ \frac{\lambda_2 - \lambda_1}{\mu}(u-1)\right). \tag{3.40}$$

From this, we can easily get the stochastic decomposition results,

$$X = X_\varphi + Y \tag{3.41}$$

Here, X_φ is still a Poisson random variable with mean $\varphi = \lambda_1/\mu$. But Y is conditionally Poisson distributed with random means that have truncated beta distributions $B(a, b, -2\rho*)$, where

$$a = \frac{f}{\mu}, \qquad b = \frac{f+r}{\mu}, \qquad \rho* = \frac{\lambda_1 - \lambda_2}{2\mu}. \tag{3.42}$$

CASE 2 - Markov-modulated Service only ($\lambda_1 = \lambda_2 = \lambda$): Then, since $\lambda_1/\mu_1 < \lambda_2/\mu_2$, we have $\mu_1 > \mu_2$. The generating function for the number of customers in the system will be, (since $\mu_1 > \mu_2$, we use equation 3.34)

$$\begin{aligned} \pi(u) = \quad & exp\left[\frac{\lambda}{\mu_1}(u-1)\right] \cdot \left[\frac{f\mu_2 + r\mu_1}{\mu_1(f+r)} \cdot M\left(\frac{f}{\mu_1}, \ \frac{f}{\mu_1} + \frac{r}{\mu_2}, \ \left(\frac{\lambda}{\mu_2} - \frac{\lambda}{\mu_1}\right)(u-1)\right)\right. \\ & \left. + \frac{r(\mu_1 - \mu_2)}{\mu_1(f+r)} \cdot M\left(\frac{f}{\mu_1} + 1, \ \frac{f}{\mu_1} + \frac{r}{\mu_2} + 1, \ \left(\frac{\lambda}{\mu_2} - \frac{\lambda}{\mu_1}\right)(u-1)\right)\right] \end{aligned} \tag{3.43}$$

This is actually the exact results from Baykal-Gürsoy and Xiao (2004). Stochastic decomposition result is given in Baykal-Gürsoy and Xiao (2004), here, we will not explain again.

And then, here we will give the probability mass function graphs for each case. Intuitively, you can get some idea from these graphs.

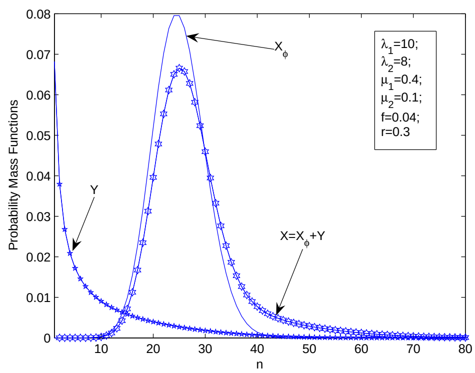

Figure 3.6: Probability Mass Function for the Number in the System (General Case)

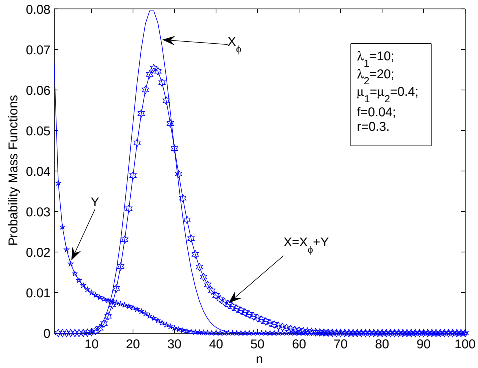

Figure 3.7: Probability Mass Function for the Number in the System (MAP)

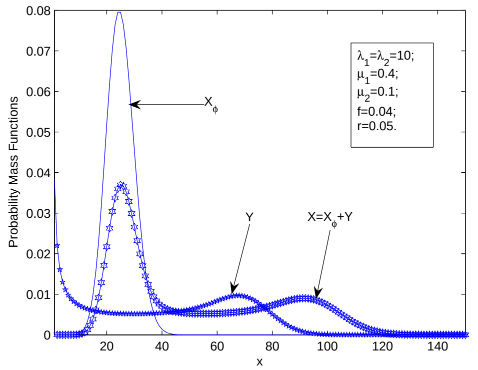

Figure 3.8: Probability Mass Function for the Number in the System (MSP, r=0.05)

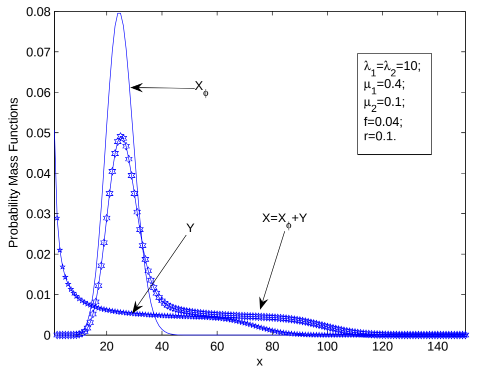

Figure 3.9: Probability Mass Function for the Number in the System (MSP, r=0.1)

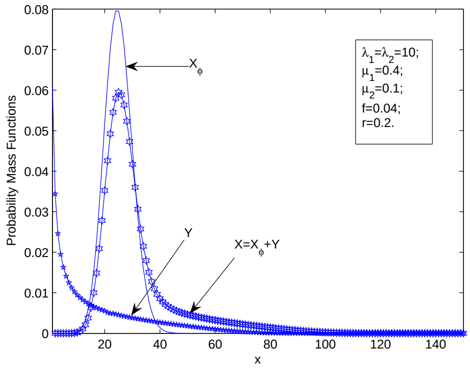

Figure 3.10: Probability Mass Function for the Number in the System (MSP, r=0.2)

From figure 3.8 to 3.10, we can see when r increases, PMF of X is influenced more by PMF of X_φ. That's because the probability that the system is in state 1, $\frac{r}{f+r}$ increases, and X_φ is a Poisson random variable with mean λ_1/μ_1(state 1).

We make the following assumption before, $\mu_1, \mu_2 > 0$. Here, we will discuss the case in which $\mu_1 > 0$ but $\mu_2 = 0$. This is the singular case that is also considered in Keilson and Servi (1993). Then, by Keilson and Servi (1993), we can easily have the probability generating function for the number in the system as,

$$\pi(u) = C_1 \cdot e^{\frac{\lambda_1}{\mu_1}u} \cdot [-\lambda_2\mu_1 u + (f + r + \lambda_2)\mu_1] \cdot [-\lambda_2\mu_1 u + \lambda_2\mu_1 + r\mu_1]^{-\frac{\mu_1+f}{\mu_1}} \tag{3.44}$$

Using $\pi(1) = 1$, the value of constant C_1 is,

$$C_1 = \frac{e^{-\frac{\lambda_1}{\mu_1}}(r\mu_1)^{\frac{\mu_1+f}{\mu_1}}}{(f+r)\mu_1}$$

Then,

$$\pi(u) = e^{\frac{\lambda_1}{\mu_1}(u-1)} \left\{ \frac{r}{r+f} \left[\frac{r/(\lambda_2+r)}{1-\lambda_2 u/(\lambda_2+r)} \right]^{\frac{f}{\mu_1}} + \frac{f}{r+f} \left[\frac{r/(\lambda_2+r)}{1-\lambda_2 u/(\lambda_2+r)} \right]^{\frac{f}{\mu_1}+1} \right\} \tag{3.45}$$

Here, $\left[\frac{r/(\lambda_2+r)}{1-\lambda_2 u/(\lambda_2+r)}\right]^{\frac{f}{\mu_1}}$ is the probability generating function of a negative binomial random variable with parameters $\frac{f}{\mu_1}$ and $\frac{r}{\lambda_2+r}$, and $\left[\frac{r/(\lambda_2+r)}{1-\lambda_2 u/(\lambda_2+r)}\right]^{\frac{f}{\mu_1}+1}$ is the probability generating function of a negative binomial random variable with parameters $\frac{f}{\mu_1}$ and $\frac{r}{\lambda_2+r}$.

Theorem 3 *The stationary number of customers in the Markov-modulated*

Arrival and Service system we described above ($\mu_1 > 0, \mu_2 = 0$ and $\frac{\lambda_1}{\mu_1} < \frac{\lambda_2}{\mu_2}$), X, has the form,

$$X = X_\varphi + Y \tag{3.46}$$

where X_φ and Y are independent, X_φ is a Poisson random variable with mean $\varphi = \lambda_1/\mu_1$, and

$$P\{Y = n\} = pP\{Y_1 = n\} + (1-p)P\{Y_2 = n\} \tag{3.47}$$

□

where $p = r/(r+f)$, Y_1 is $NB(f/\mu_1, r/(\lambda_2 + r))$ and Y_2 is $NB(f/\mu_1 + 1, r/(\lambda_2 + r))$.

PROOF Please see above description. ■

3.2.2 Retrial Queues

In this section, we will consider an $M/M/1/1$ retrial queue. In a wireless network, this is a one channel network, that means there is only one channel in each cell area. When a mobile device wants to connect to the network, but the channel is occupied by another device, it will retry again after a random interval or abandon trying with some rate. This is the basic system for wireless networks where multi-channel and other variations can be studied.

In this system, arrival process is Poisson with rate λ, and service times are $i.i.d$ exponential with rate μ. The retrial interval for each blocked customer is $i.i.d$ exponentially distributed with rate ζ, and each customer will renege from the system with a constant rate θ. This case is considered in Keilson and Servi (1993), please see more details in Keilson and Servi (1993). Then the probability generating function of the number in the orbit is

$$\pi(u) = \begin{cases} K\left[M(b-a,b,-d-e(u-1)) + \frac{\lambda\zeta}{\lambda\theta+\mu(\theta+\zeta)}M(b-a+1,b+1,-d-e(u-a))\right], & \text{if } \theta > 0 \\ \left(1+\frac{\lambda}{\mu}-\frac{\lambda}{\mu}u\right)\left\{\frac{1-\lambda/\mu}{1-(\lambda/\mu)u}\right\}, & \text{if } \theta = 0 \end{cases} \tag{3.48}$$

Here,

$$a = \frac{\mu}{\theta}, \qquad b = \frac{\lambda}{\theta+\zeta} + \frac{\mu}{\theta}, \qquad d = -\frac{\lambda\zeta}{\theta(\theta+\zeta)}, \qquad e = -\frac{\lambda}{\theta}.$$

Since $\pi(1) = 1$, we can get

$$K = \frac{\lambda\theta + \mu(\theta+\zeta)}{[\lambda\theta+\mu(\theta+\zeta)]M(b-a,b,-d) + \lambda\zeta M(b-a+1,b+1,-d)} \tag{3.49}$$

Then for the case $\theta > 0$, we can get this,

$$\pi(u) = p_1 \cdot \frac{M(b-a, b, -d-e(u-1))}{M(b-a, b, -d)} + p_2 \cdot \frac{M(b-a+1, b+1, -d-e(u-1))}{M(b-a+1, b+1, -d)} \tag{3.50}$$

Where,

$$p_1 = \frac{[\lambda\theta + \mu(\theta+\zeta)]M(b-a, b, -d)}{[\lambda\theta + \mu(\theta+\zeta)]M(b-a, b, -d) + \lambda\zeta M(b-a+1, b+1, -d)}, \tag{3.51}$$

$$p_2 = \frac{\lambda\zeta M(b-a+1, b+1, -d)}{[\lambda\theta + \mu(\theta+\zeta)]M(b-a, b, -d) + \lambda\zeta M(b-a+1, b+1, -d)}. \tag{3.52}$$

$b - a = \frac{\lambda}{\theta+\zeta} > 0$ and $-d = \frac{\lambda\zeta}{\theta(\theta+\zeta)} > 0$, so we can get $p_1 > 0,\ p_2 > 0$, and also we can easily have,

$$p_1 + p_2 = 1.$$

Let

$$g_1(u) = \frac{M(b-a, b, -d-e(u-1))}{M(b-a, b, -d)}, \qquad g_2(u) = \frac{M(b-a+1, b+1, -d-e(u-1))}{M(b-a+1, b+1, -d)}.$$

On the other hand, by Fitzgerald (2002), we know that $g_1(u)$ and $g_2(u)$ are both hyper-Poisson random variables with four parameters $(b-a, b, -d, -e)$ and $(b-a+1, b+1, -d, -e)$, respectively. The probability mass function for a hyper-Poisson random variable Y with four parameters (a, b, c, λ) is

$$P\{Y = n\} = \frac{M(a+n, b+n, c-\lambda)}{M(a, b, c)} \frac{(a)_n}{(b)_n} \frac{\lambda^n}{n!}. \tag{3.53}$$

If the first two parameters of hyper-Poisson random variables are equivalent, the hyper-Poisson will recover the Poisson distribution. For example, in our case, if $b - a = b$, then $\mu = 0$, that means there is no service able to be processed. Since $b - a = b$, $M(b-a, b, x) = e^x$, we can then have,

$$\begin{aligned}
p_1 &= \frac{\lambda\theta \cdot e^{-d}}{\lambda\theta \cdot e^{-d} + \lambda\zeta \cdot e^{-d}} = \frac{\theta}{\theta+\zeta} > 0; \\
p_2 &= \frac{\zeta}{\theta+\zeta}; \\
g_1(u) &= \frac{e^{-d-e(u-1)}}{e^{-d}} = e^{\frac{\lambda}{\theta}(u-1)}; \\
g_1(u) &= e^{\frac{\lambda}{\theta}(u-1)}.
\end{aligned}$$

So,

$$\pi(u) = e^{\frac{\lambda}{\theta}(u-1)}. \tag{3.54}$$

This is just the generating function of a Poisson random variable with parameter $\frac{\lambda}{\theta}$. Intuitively, this makes sense also, since the input for system is λ, and output is only the reneging rate θ, no service rate.

For another case, if $d = 0$, this means $\zeta = 0$. Then, we will have $p_1 = 1, \quad p_2 = 0$, and then,

$$\pi(u) = g_1(u) = M(b - a, b, -e(u - 1)).$$

So, this is the generating function of a Poisson random variable randomized by truncated beta $B(b - a, b, -e)$.

For the case $\theta = 0$, if $\lambda \geq \mu$, the number in the system will increase to ∞. So for the stable system, we need to let $\lambda < \mu$, that means $\frac{\lambda}{\mu} < 1$. Then we have

$$\pi(u) = p_1' + p_2' \cdot g(u) \tag{3.55}$$

where $p_1' = 1 - \frac{\lambda}{\mu}, p_2' = \frac{\lambda}{\mu}$, so we know $p_1' + p_2' = 1$ and $p_1', p_2' > 0$.
$g(u) = \frac{1-\lambda/\mu}{1-\lambda/\mu u}$ is the generating function of a Geometric random variable with parameter $1 - \frac{\lambda}{\mu}$

Then, we have the following theorem.

Theorem 4 *The number of customers in the retrial orbit, Y, has the probability mass function as,*

$$P\{Y = n\} = p_1 \cdot P\{Y_1 = n\} + p_2 \cdot P\{Y_2 = n\}, \text{ if } \theta > 0; \tag{3.56}$$

$$P\{Y = n\} = p_1' \cdot P\{Y_1' = n\} + p_2' \cdot P\{Y_2' = n\}, \text{ if } \theta = 0. \tag{3.57}$$

where Y_1 and Y_2 are hyper-Poisson random variables with four parameters $(b - a, b, -d, -e)$ and $(b - a + 1, b + 1, -d, -e)$, respectively. p_1 and p_2 are given in the above equations 3.51 and 3.52; Y_1' is a constant 0, Y_2' is Geometric random variables with parameter $1 - \frac{\lambda}{\mu}$, and $p_1' = 1 - \frac{\lambda}{\mu}, p_2' = \frac{\lambda}{\mu}$.□

Proof From the above descriptions, the results are easy to get. ■

The following figures are probability mass functions for number in the retrial orbit under several situations. From Figure 3.11 to 3.13, we can see that with increasing retrial rate ζ, form of the probability mass function of Y differs more from the form for probability mass function of Y_1, and Y_2 remain similar.

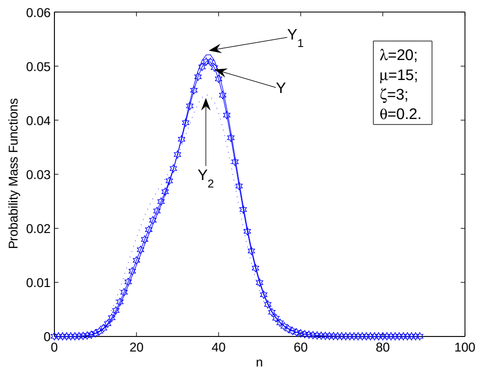

Figure 3.11: Probability Mass Function for the Number in the Orbit ($\zeta = 3$)

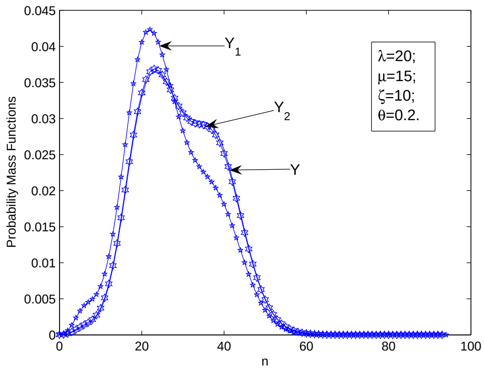

Figure 3.12: Probability Mass Function for the Number in the Orbit ($\zeta = 10$)

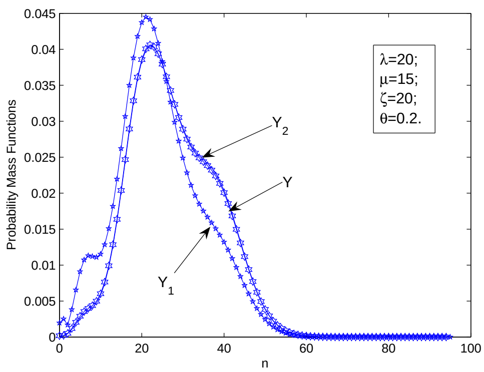

Figure 3.13: Probability Mass Function for the Number in the Orbit ($\zeta = 20$)

3.2.3 Summary

In this section, we give stochastic decomposition results for $M/M/\infty$ queue with Markov-modulated arrival and service(2 states), and present the expected number of customers in the system and its variance. Then, we present the stochastic decomposition results for a $M/M/1/1$ retrial queue with infinity capacity in the retrial orbit.

We have already discovered stochastic decomposition results for various queueing models. However, we believe that there are much more queueing models which have stochastic decomposition property. Next, we will focus on $M/G/\infty$ case and other general retrial queueing models.

We only considered the case in which there are only two modulation states. We will search for stochastic decomposition results for more general Markov-modulated queues. This will be another future research direction.

3.3 Completion Time Analysis of $M/G/\infty$ Queue under Two Service Speeds

As an expansion of previous research results on $M/M/\infty$ queues with different service speeds, in this section, an $M/G/\infty$ queuing system with two service speeds is discussed. The arrivals of customers is a Poisson process. The cumulative distribution function for service requirements of each customers is $F_S(t)$, probability density function is $f_S(t)$, and corresponding Laplace transform for density function is $L_S(s)$. Mean service requirement is $1/\mu$.

The system's normal service speed is 1, and when a failure happens to the system, the service speed will drop to $0 < \alpha < 1$. We call the periods that the system works with normal speed as up periods, and the periods that the system works with lower speed α as down periods. Since $\alpha > 0$, this kind of system is called partial breakdown system. The system will alternate between up and down periods. Up periods duration is exponentially distributed with mean $1/f$. However, the down periods duration is generally distributed with cdf $F_D(t)$, with mean duration $1/r$. The corresponding pdf is $f_D(t)$, and Laplace transform is $L_D(s)$.

And also, in this system, we assume that when a failure happens or a repair is finished, the service for current customers will restart with the same service requirements distribution.

3.3.1 Completion Time Analysis of $M/G/\infty$ with General Down Periods Partial Breakdown System

For completion time of an arrival, first, we divide the customers into two groups. The first group G^1 of customers are those who arrives during a system up period. The second group G^2 of customers are those who arrives during a system down period. Mean up periods and down periods for the system are $\frac{1}{f}$ and $\frac{1}{r}$, so, we can get the probability of any customer in G^1 or G^2. They are,

$$P\left\{G^1\right\} = \frac{r}{f+r}, \qquad P\left\{G^2\right\} = \frac{f}{f+r}. \tag{3.58}$$

In this system, we assumed that when a failure happens or a repair is finished, service for customers will restart with the same service requirements distribution. Then, for those two groups of customers we defined, we will get the Laplace transform of conditional completion time separately. Then combine them together to get Laplace transform for completion time C.

Customers in G^1

Here, we will assume our customers arrive during an up period. And up and down periods are U_i and D_i, and service requirements are S_i. And then we define the following events,

A_n there are n complete up periods, n complete down periods and another incomplete up period in the completion time ($n = 0, 1, 2, \ldots$). This actually means

$$\begin{aligned} &S_{2i-1} > U_i, \quad \frac{1}{\alpha} S_{2i} > D_i \text{ for } i = 1, 2, \ldots n \\ &\text{and } S_{2n+1} < U_{n+1}. \end{aligned} \tag{3.59}$$

E_n there are $n+1$ completed up periods, n completed down periods and another incompleted down period in the completion time ($n = 0, 1, 2, \ldots$). This actually means

$$\begin{aligned} &S_{2i-1} > U_i, \quad \frac{1}{\alpha} S_{2i} > D_i, \quad S_{2n+1} > U_{n+1} \text{ for } i = 1, 2, \ldots n \\ &\text{and } \frac{1}{\alpha} S_{2n+2} < D_{n+1}. \end{aligned} \tag{3.60}$$

In this system, we can get probabilities for each events,

$$\begin{aligned} P\left\{A_n | G^1\right\} &= P^n\{S > U\} \cdot P^n\{S > \alpha D\} \cdot P\{S < U\} \\ P\left\{E_n | G^1\right\} &= P^{n+1}\{S > U\} \cdot P^n\{1/\alpha S > D\} \cdot P\{1/\alpha S < D\} \end{aligned} \tag{3.61}$$

For each events stated above, we will try to get the Laplace transform for conditional completion time. The conditional completion time C will be,

$$C = \begin{cases} S_1 & \text{Event } A_0 \\ U_1 + \frac{1}{\alpha} S_2 & \text{Event } E_0 \\ U_1 + D_1 + S_3 & \text{Event } A_1 \\ \ldots & \ldots \\ \sum_{i=1}^{n} U_i + \sum_{i=1}^{n} D_i + S_{2n+1} & \text{Event } A_n \\ \sum_{i=1}^{n+1} U_i + \sum_{i=1}^{n} D_i + \frac{1}{\alpha} S_{2n+2} & \text{Event } E_n \\ \ldots & \ldots \end{cases} \tag{3.62}$$

Case A_n:

$$
\begin{aligned}
&E[e^{-sC}|A_n, G^1] \cdot P\{A_n|G^1\} \\
&= E^n[e^{-sU}|S > U] \cdot E^n[e^{-sD}|S > \alpha D] \cdot E[e^{-sS}|S < U] \cdot P^n\{S > U\} \cdot P^n\{S > \alpha D\} \cdot P\{S < U\} \\
&= \big(E[e^{-sU}|S > U]P\{S > U\}\big)^n \cdot \big(E[e^{-sD}|S > \alpha D]P\{S > \alpha D\}\big)^n \cdot E[e^{-sS}|S < U]P\{S < U\}
\end{aligned}
\tag{3.63}
$$

where, s is a complex number with positive real part, $U \sim Exp(f)$, S is the generally distributed service requirement random variable, and D is the generally distributed down period random variable.

First, find the conditional density function for corresponding random variables, then we will be able to get the following terms,

$$E[e^{-sU}|S > U]P\{S > U\} = \frac{f}{s+f}[1 - L_S(s+f)] \tag{3.64}$$

$$E[e^{-sD}|S > \alpha D]P\{S > \alpha D\} = L_D(s) = \int_0^\infty e^{-st} f_D(t) F_S(\alpha t) dt \tag{3.65}$$

$$E[e^{-sS}|S < U]P\{S < U\} = L_S(s+f) \tag{3.66}$$

Then, we can have the Laplace transform for conditional completion time, given that customer is in G^1 and event A_n happens.

$$
\begin{aligned}
E[e^{-sC}|A_n, G^1] \cdot P\{A_n|G^1\} = &\left\{\frac{f}{s+f}[1 - L_S(s+f)]\right\}^n \\
&\cdot \left\{L_D(s) - \int_0^\infty e^{-st} f_D(t) F_S(\alpha t) dt\right\}^n \cdot L_S(s+f)
\end{aligned}
\tag{3.67}
$$

Case E_n: Similar to A_n, we have,

$$
\begin{aligned}
E[e^{-sC}|E_n, G^1] \cdot P\{E_n|G^1\} = &\big(E[e^{-sU}|S > U]P\{S > U\}\big)^{n+1} \cdot \big(E[e^{-sD}|S > \alpha D]P\{S > \alpha D\}\big)^n \\
&\cdot E[e^{-s\frac{1}{\alpha}S}|S < \alpha D]P\{S < \alpha D\}
\end{aligned}
\tag{3.68}
$$

The first 2 terms in above equation are given in case A_n, and we used similar method to get the last term.

$$E[e^{-s\frac{1}{\alpha}S}|S < \alpha D]P\{S < \alpha D\} = L_S\left(\frac{s}{\alpha}\right) - \int_0^\infty e^{-s\frac{1}{\alpha}t} f_S(t) F_D\left(\frac{t}{\alpha}\right) dt \tag{3.69}$$

Then, we have,

$$E[e^{-sC}|E_n,G^1]\cdot P\{E_n|G^1\}=\left\{\frac{f}{s+f}[1-L_S(s+f)]\right\}^{n+1}\cdot\left\{L_D(s)-\int_0^\infty e^{-st}f_D(t)F_S(\alpha t)dt\right\}^n$$
$$\cdot\left[L_S(s/\alpha)-\int_0^\infty e^{-s\frac{1}{\alpha}t}f_S(t)F_D\left(\frac{t}{\alpha}\right)dt\right] \tag{3.70}$$

Summary for G^1 So, we can get the Laplace transform for the conditional completion time, given that the customer arrives during up periods.

$$\begin{aligned}E\left[e^{-sC}|G^1\right]&=\sum_{n=0}^\infty E\left[e^{-sC}|A_n,G^1\right]P\{A_n|G^1\}+\sum_{n=0}^\infty E\left[e^{-sC}|E_n,G^1\right]P\{E_n|G^1\}\\&=\frac{1}{1-V}\cdot\left\{L_S(s+f)+\frac{f}{s+f}\cdot[1-L_S(s+f)]\right.\\&\qquad\left.\cdot\left[L_S(s/\alpha)-\int_0^\infty e^{-s\frac{1}{\alpha}t}f_S(t)F_D\left(\frac{t}{\alpha}\right)\right]\right\}\end{aligned} \tag{3.71}$$

where

$$V=\frac{f}{s+f}[1-L_S(s+f)]\cdot\left[L_D(s)-\int_0^\infty e^{-st}f_D(t)F_S(\alpha t)dt\right]$$

Customers in G^2

Now, we will consider the customers in group G^2, similar method will be used. However, since down periods are generally distributed, more derivations are needed. Let Y be the remaining down time after arrival of the customer. Given that customer arrives during down period, conditional completion time will be,

$$C=\begin{cases}\frac{1}{\alpha}S_1 & \text{if } \frac{1}{\alpha}S_1<Y \qquad \text{Event } A_0\\ Y+S_2 & \text{if } \frac{1}{\alpha}S_1<Y \text{ and } S_2<U_1 \qquad \text{Event } E_0\\ Y+U_1+\frac{1}{\alpha}S_3 & \text{if } \frac{1}{\alpha}S_1<Y, S_2>U_1 \text{ and } \frac{1}{\alpha}S_3<D_2 \qquad \text{Event } A_1\\ \dots & \dots\\ Y+\sum_{i=1}^n D_i+\sum_{i=1}^n U_i+\frac{1}{\alpha}S_{2n+1} & \text{Event } A_n\\ Y+\sum_{i=1}^{n+1} D_i+\sum_{i=1}^n U_i+S_{2n+2} & \text{Event } E_n\\ \dots & \dots\end{cases} \tag{3.72}$$

First, for the remaining down time Y, probability density function is,

$$f_Y(t)=r\cdot[1-F_D(t)],\qquad t>0. \tag{3.73}$$

and Laplace transform for pdf of Y is,

$$L_Y(s) = \frac{r}{s}[1 - L_D(s)] \tag{3.74}$$

Similar method can be used to obtain Laplace transforms of completion time under each events,

$$E[e^{-sC}|A_0, G^2] \cdot P\{A_0|G^2\} = L_S\left(\frac{s}{\alpha}\right) - \int_0^\infty e^{-\frac{s}{\alpha}t} f_S(t) F_Y\left(\frac{t}{\alpha}\right) dt \tag{3.75}$$

$$E[e^{-sC}|E_0, G^2] \cdot P\{E_0|G^2\} = L_S(s+f) \cdot \left[L_Y(s) - \int_0^\infty e^{-st} f_Y(t) F_S(\alpha t) dt\right] \tag{3.76}$$

$$\begin{aligned}
&E[e^{-sC}|A_n, G^2] \cdot P\{A_n|G^2\} \\
&\quad = \left[L_Y(s) - \int_0^\infty e^{-st} f_Y(t) F_S(\alpha t) dt\right] \cdot \left[L_D(s) - \int_0^\infty e^{-st} f_D(t) F_S(\alpha t) dt\right]^{n-1} \\
&\qquad \cdot \left[\frac{f}{s+f}\left(1 - L_S(s+f)\right)\right]^n \cdot \left[L_S(s/\alpha) - \int_0^\infty e^{-s/\alpha t} f_s(t) F_D(t/\alpha) dt\right]
\end{aligned} \tag{3.77}$$

$$\begin{aligned}
&E[e^{-sC}|E_n, G^2] \cdot P\{E_n|G^2\} \\
&\quad = \left[L_Y(s) - \int_0^\infty e^{-st} f_Y(t) F_S(\alpha t) dt\right] \cdot \left[L_D(s) - \int_0^\infty e^{-st} f_D(t) F_S(\alpha t) dt\right]^n \\
&\qquad \cdot \left[\frac{f}{s+f}\left(1 - L_S(s+f)\right)\right]^n \cdot L_S(s+f)
\end{aligned} \tag{3.78}$$

Summary for G^2 By combining equations 3.75-3.78, we can get Laplace transform for completion time of customers in G^2,

$$\begin{aligned}
E[e^{-sC}|G^2] = &\frac{1}{1-V} \cdot \left[L_Y(s) - \int_0^\infty e^{-st} f_Y(t) F_S(\alpha t) dt\right] \\
&\cdot \left\{\frac{f}{s+f}[1 - L_S(s+f)] \cdot \left[L_S(s/\alpha) - \int_0^\infty e^{-s/\alpha t} f_S(t) F_D(t/\alpha) dt\right] + L_S(s+f)\right\} \\
&+ \left[L_S(s/\alpha) - \int_0^\infty e^{-s/\alpha t} f_S(t) F_Y(t/\alpha) dt\right]
\end{aligned} \tag{3.79}$$

Completion Time Laplace transform

By combining equations 3.71 and 3.79, for this system, the completion time Laplace transform is given by,

$$
\begin{aligned}
E[e^{-sC}] &= E[e^{-sC}|G^1]P\{G^1\} + E[e^{-sC}|G^2]P\{G^2\} \\
&= \frac{1}{1-V} \cdot \left\{ \frac{r}{r+f} L_S(s+f) + \frac{rf}{(r+f)(s+f)} [1 - L_S(s+f)] \cdot \left[L_S(s/\alpha) - \int_0^\infty e^{-s/\alpha t} f_S(t) F_D(t/\alpha) dt \right] \right. \\
&+ \left[L_Y(s) - \int_0^\infty e^{-st} f_Y(t) F_S(\alpha t) dt \right] \\
&\left. \cdot \left[\frac{f}{r+f} L_S(s+f) + \frac{f^2}{(r+f)(s+f)} [1 - L_S(s+f)] \cdot \left[L_S(s/\alpha) - \int_0^\infty e^{-s/\alpha t} f_S(t) F_D(t/\alpha) dt \right] \right] \right\} \\
&+ \frac{f}{r+f} \left[L_S(s/\alpha) - \int_0^\infty e^{-s/\alpha t} f_S(t) F_Y(t/\alpha) dt \right]
\end{aligned}
$$

Here,

$$
V = \frac{f}{s+f} [1 - L_S(s+f)] \cdot \left[L_D(s) - \int_0^\infty e^{-st} f_D(t) F_S(\alpha t) dt \right]
$$

3.3.2 Special Cases

In this section, several variants of the original problem are discussed, especially for the $M/M/\infty$ system with exponentially distributed up and down periods, and results for completion time here coincide with previous research results.

Completion Time for $M/G/\infty$ Partial Breakdown System With Restart Service

For this part, we will focus on $M/G/\infty$ system with the same partial breakdown rule as in previous section. However, the down periods duration will be exponentially distributed with mean r. After each repair or system partial breakdown, every customers in the system must restart their services with the same distribution $S(\cdot)$.

we can have Laplace transform for completion time for such an $M/G/\infty$ system,

$$\begin{aligned}E[e^{-sC}] &= E[e^{-sC}|G^1]P\{G^1\} + E[e^{-sC}|G^2]P\{G^2\} \\ &= \frac{1}{(f+r)(1-V)} \cdot \Bigg\{ rL_S(s+f) + \frac{fr}{s+f}\left[1 - L_S(s+f)\right] L_S\left(\frac{s+r}{\alpha}\right) \\ &+fL_S\left(\frac{s+r}{\alpha}\right) + \frac{fr}{s+r}\left[1 - L_S\left(\frac{s+r}{\alpha}\right)\right] L_S(s+f)\Bigg\} \qquad (3.80)\end{aligned}$$

where,

$$V = \frac{rf\left[1 - L_S(s+f)\right]\left[1 - L_S(\frac{s+r}{\alpha})\right]}{(s+f)(s+r)}$$

If the service requirement S is exponentially distributed with rate μ, then $L_S(s) = \frac{\mu}{s+\mu}$. Plug this into equation 3.80, and we will get the same results as in $M/M/\infty$ section.

Completion Time for $M/M/\infty$ Partial Breakdown System With General Down Periods

In this section, we made modification on original problem, the service time is changed to be exponentially distributed with mean $1/\mu$.

Then, we can get Laplace transform of completion time for this system,

$$\begin{aligned}E\left[e^{-sC}\right] =& \frac{f\mu\alpha\left[s+\mu\alpha - r(1 - L_D(s+\mu\alpha))\right]}{(f+r)(s+\mu\alpha)^2} \\ &+ \frac{1}{1-V} \cdot \left\{ \frac{r\mu\left[s+\mu\alpha + f(1+\alpha)(1 - L_D(s+\mu\alpha))\right]}{(f+r)(s+f+\mu)(s+\mu\alpha)} + \frac{f^2 r\mu\alpha\left[1 - L_D(s+\mu\alpha)\right]^2}{(f+r)(s+f+\mu)(s+\mu\alpha)^2} \right\} \quad (3.81)\end{aligned}$$

where,

$$V = \frac{f}{s+f+\mu} \cdot L_D(s+\mu\alpha) \qquad (3.82)$$

If down periods are exponentially distributed with rate r, we can get the same results as in section 3.3.2.

Completion Time for $M/M/\infty$ Partial Breakdown System

By making both the service time and down periods exponentially distributed, our system will became an $M/M/\infty$ system with Markov modulated service speeds. Then, we can have Laplace transform for completion time for such an $M/M/\infty$ system,

$$\begin{aligned} E[e^{-sC}] &= E[e^{-sC}|G^1]P\{G^1\} + E[e^{-sC}|G^2]P\{G^2\} \\ &= \frac{\mu}{f+r} \cdot \frac{r(s+f+r+\mu\alpha) + f\alpha(s+f+r+\mu)}{(s+f+\mu)(s+r+\mu\alpha) - fr} \end{aligned} \tag{3.83}$$

Stochastic Decomposition

For stochastic decomposition property, first we know for sure that $C = S+Y$, here S is the service requirement for the particular customer, and Y is another random variables we do not know. So, it is obvious that S and Y are not independent to each other, therefore, from Laplace transform of completion time, we can not clearly decompose it into multiplication of Laplace transforms of service requirement and another random variable. However, we can rewrite the Laplace transform, then get a sense about the completion time for this system. From equation 3.83, we can rewrite it into,

$$E[e^{-sC}] = \frac{r}{f+r} \cdot \underbrace{\frac{\mu(s+r+f\alpha+\mu\alpha)}{(s+f+\mu)(s+r+\mu\alpha)-fr}}_{A} + \frac{f}{f+r} \cdot \underbrace{\frac{\mu(s\alpha+f\alpha+r+\mu\alpha)}{(s+f+\mu)(s+r+\mu\alpha)-fr}}_{B}$$

Then, we can decompose completion time C into $C = \frac{r}{f+r}C_1 + \frac{f}{f+r}C_2$, and here, random variables C_1 and C_2 have Laplace transforms A and B respectively. After inverse, we can have density function for C_1 and C_2 as following,

$$f_{C_1}(t) = \frac{\mu(s_1+f\alpha+r+\mu\alpha)}{s_1-s_2}e^{s_1 t} - \frac{\mu(s_2+f\alpha+r+\mu\alpha)}{s_1-s_2}e^{s_2 t} \tag{3.84}$$

$$f_{C_2}(t) = \frac{\mu(s_1\alpha+f\alpha+r+\mu\alpha)}{s_1-s_2}e^{s_1 t} - \frac{\mu(s_2\alpha+f\alpha+r+\mu\alpha)}{s_1-s_2}e^{s_2 t} \tag{3.85}$$

Here, s_1 and s_2 are two roots for the following equation, also $s_1 > s_2$. On the other hand, it is easy to prove that these two roots are both real negative roots.

$$(s+f+\mu)(s+r+\mu\alpha) - fr = 0 \tag{3.86}$$

Comparison With Previous Results

In Gaver (1962), completion time for complete breakdown is analyzed, our completion time Laplace transform results (if $\alpha = 0$) coincide with its preemptive-repeate-different interruptions case. Since for Gaver's case all services start in an up period, we only need to compare Gaver's results with our results for group G^1. If $\alpha = 0$ for our problem, we will have exatly the same results as Gaver's.

And also, by little's theorem, we can get expected number in the system, this results also coincide with Baykal-Gürsoy and Xiao (2004). The expected completion time for this system is,

$$E[C] = \frac{1}{\mu} + \frac{f(1-\alpha)}{f+r} \cdot \frac{f+r+\mu}{\mu(f\alpha + r + \mu\alpha)} \tag{3.87}$$

3.3.3 Summary

Completion time for customers in an $M/G/\infty$ system with two service speeds is obtained. Two service speeds are modulated by generally distributed down periods and exponentially distributed up periods. Some special cases of this system are further analyzed, and the results coincide with previous research. In traffic flow modeling, the exponentially distributed up periods can be viewed as interarrival time of incidents, and the generally distributed down periods can be viewed as repair time of the road by first responders.

Further research can be extended to multiple service speeds with different partial breakdowns, and also, with extension to generally distributed up periods, the model can adapt into more complicated breakdown systems.

3.4 Application on Traffic Flow Modeling with Incidents

Highway congestion delay include recurrent delay and non-recurrent delay caused by accidents, vehicle breakdowns, and other random events. Recurrent delay is due to travel demand fluctuations, and the network structure of highway etc. Non-recurrent delay arises due to incidents.

In this section, a stochastic queuing traffic flow model which combines recurrent and non-recurrent congestion is presented. We compare the travel time estimate of this model with the incident only model via simulation experiments.

3.4.1 Combined Traffic Flow Modeling under Recurrent and Non-Recurrent Congestion

A modified $M/MSP/C/C$ queuing model is used to model the traffic flow subject to both incidents and recurrent congestion. The following diagram is the state transition diagram for such a queuing system. The two dimensional stochastic process $\{X(t), U(t)\}$ describes the state of the system at time t, where $X(t)$ is the number of customers in the system, and $U(t)$ is the status of the system at time t. If at time t, the system is experiencing an interruption, then $U(t)$ is equal to F (failure); otherwise, $U(t)$ is N (normal). The system is said to be in state (i, F), if there are i customers in the system which is experiencing an interruption, while the system is said to be in state (i, N), if there are i customers in the system which is functioning normally.

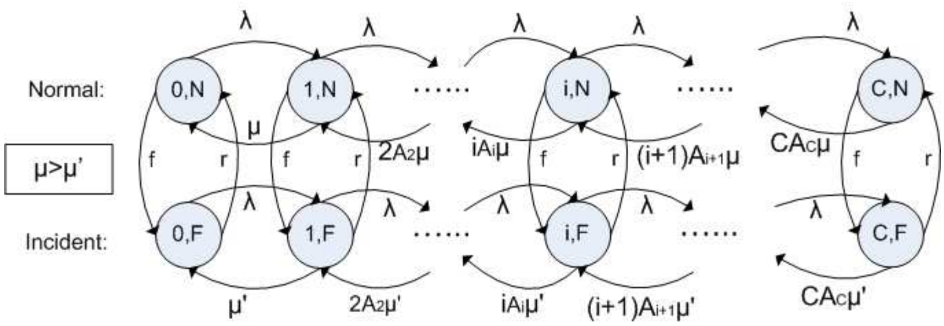

Figure 3.14: State Transition Diagram for Modified $M/MSP/C/C$ Model

In this system, we use the exponential congestion factor A_i to modify the $M/MSP/C/C$

queuing model with

$$A_i = \frac{V_i}{V_{free}}.$$

Here, V_i is the vehicle speed based on free speed V_{free} when there are totally i vehicles on the road link given by,

$$V_i = V_{free} \cdot \exp\left[-\left(\frac{i-1}{\beta}\right)^{\gamma}\right]$$

For detailed explanation about exponential congestion model, please refer to Jain and Smith (1997).

Table 3.1: Comparison of analytical models with simulation

C	λ	μ	f	r	Simulation	Incident only Model($M/MSP/C$)	Combined Model
400	0.3	0.015	0.0002	0.005	74.93	74.57(-0.48%)	85.81(14.52%)
200	0.3	0.03	0.0002	0.005	39.94	39.19(-1.88%)	47.40(18.68%)
100	0.3	0.06	0.0002	0.005	23.38	20.84(-10.86%)	28.85(23.4%)
50	0.3	0.12	0.0002	0.005	13.08	11.08(-15.29%)	13.57(3.75%)

Table 3.1 presents the expected travel times for the incident only model of Baykal-Gürsoy et al. (2009) and the combined model described in this section, together with the traffic simulation results. The relative errors between the simulation and analytic models are given inside the parenthesis in the last two columns. We use INTEGRATION to simulate a link with travel speed at full capacity $v = 57.5 km/hr (65 miles/hr)$. Note that INTEGRATION treats the arrival process as fluid, thus generating l vehicles per hour deterministically. We consider various arrival rates and link lengths on a two-lane roadway where minor incidents happen. Minor incidents take on the average of 7 minutes to clear (Skabardonis et al. (1998)). They report 0.5 incidents per hour for a one kilometer roadway. The values for f and r are chosen accordingly. For each setting, we run the simulation to obtain average travel times for 100 replications and each replication simulates a 12000-second period. Under congestion, each replication takes 5 minutes, thus each scenario takes more than four hours. This is very time-consuming compared to the analytical model. We would like to emphasize that in the simulation model as in real life, the service times are also neither independent nor exponential. The incident process is the only random process in the simulations. Also, here we take $m' = m/14$. From this set of results, we can conclude that, the incident only model will work well under light traffic situation such as most highway

traffic, and combined model will usually be effective under heavy traffic situation, such as city traffic.

3.4.2 Summary

We combined models for recurrent and non-recurrent congestion in traffic flow together, improved the travel time estimation results on a road link with heavy traffic situation. Recurrent congestion is represented by empirical exponential congestion model, and non-recurrent congestion is modeled by queuing system with Markov-modulated service processes.

Further research can be directed to use queuing network to represent the traffic network. In this system, various queuing systems can be used to represent different types of road links on the traffic network.

Chapter 4

Infrastructure Security Games

For man-made emergencies in transit infrastructure, such as terrorist attacks, suicide bombings etc. There are adversaries who are trying to do damages or hurt people in transit infrastructure, and then first responders who are trying to avoid these types of emergencies from happening. So games of security in infrastructure are constantly played between adversaries and first responders. In this chapter, we will introduce game theory into this topic, to help identify the nature of the problem, and also help first responders to design their resource allocation strategy or patrol strategy in order to avoid or minimize the damages caused by adversaries.

First, a one-step problem between adversary and first responder is considered, and it is a resource allocation problem for first responder who is trying to catch adversary with limited resources. Static game is designed to represent the problem, and resource allocation strategies under game equilibrium is obtained for first responder, also the mixed probability strategy for adversary.

Then, a dynamic game between adversary and first responder is considered, in which adversary is immobile, and first responder will patrol the infrastructure to detect the adversary. Game theory and partially observable Markov decision process(POMDP) models are used to analyze the competitive nature of the problem. Risk measure and patrol sequence strategies under equilibrium are obtained.

4.1 Background

First, let us look at the particular emergency situations to analyze them more deeply. On July 7 2005 in London, at 8:50 a.m., three bombs exploded within fifty seconds of each other on three London Underground trains, and a fourth bomb exploded on a bus nearly an hour later at 9:47 a.m. in Tavistock Square. The bombings killed 52 commuters and the four suicide bombers, also it caused disruption of country's transportation system and telecommunication system.

Bombers intended to detonate 4 bombs on 4 underground trains which left from King's Cross St. Pancras and headed to 4 different directions. But one of the bombers thought the Northern line he was intended to attack was suspended at that time according to reports. So, he turned away from the Underground, and took a bus, then finally detonated the bomb at Tavistock Square. And the other 3 bombs on the underground trains were detonated nearly at the same time, and they were detonated when two trains were crossing, thus resulting in more damages.

For most of the day, central London's public transport system was effectively crippled because of the complete closure of the underground system, the closure of the Zone 1 bus networks, and the evacuation of Russell Square. Most of the mainline train stations and bus networks reopened later that day, however, some train lines kept closed even until August. Not only the incident hit transportation system, but also caused intermittent unavailability of mobile and landline phone systems due to excessive usage during that time.

And also, there were immediate impacts on world economy in financial markets and exchange rate activity. The pound fell 0.89 cents to a 19-month low against U.S. dollar. The FTSE 100 Index fell by about 200 points in the two hours after the first attack. Stock market in Germany, France, Spain also closed about 1% down on that day. Then, they recovered on the second day since the damages were not as significant as originally thought.

Similar things happened during Madrid train bombing incident on March 11 2004, which killed 191 people, and injured around 2000 individuals. During these incidents, the attackers were not suicide bombers, instead, they used cell phones to detonate bombs. It also happened during the morning rush hour in order to cause mass casualties. More trains were involved, and more bombs were detonated in this incident.

In 1995 Tokyo Sarin gas attack on subway, plastic bags with Sarin gas were placed on the train, the attacker used an umbrella to poke a hole and then quickly left the train. The train continuing to operate caused all passengers to be affected by the gas. There were 12 people killed, and totally over a thousand people injured.

From the previous attacks on transit infrastructures, we can discover that situations and attack methods can be different from one another. It is hard to model this kind of emergency exactly according to what happened before, and it is also uncanny to do so since the attackers will definitely change their strategies to avoid detection next time. However, all these incidents have some characteristics in common, those are the attackers are trying to avoid any security detection, on the other hand, securities are trying their best to track down every footprints attackers left. Therefore, we shall generalize our problems into a noncooperative game between a responder and an adversary.

4.2 Resource Allocation Games

In this section, a one-step hide and seek game between adversary and first responder is analyzed. First responder allocates limited resources into different cells of infrastructure to maximize detection chances of adversary, while adversary is mixing up attacking position to avoid detection.

We are trying to work on the dynamic hide and seek game described in section 2.2.1. As a start point of this research, we will consider this one-step noncooperative game. We will assume that this static game is a zero-sum game, since the damage to the infrastructure are the payoff for adversary and cost for responder. The responder and the adversary will make their respective decisions simultaneously.

4.2.1 One Step Hide and Seek Game

We consider a grid structure composed of cells for the transit infrastructure. Cells have associated characteristics such as their importance to the adversary, vulnerability to attacks, and the measure of detection. These factors are all available to both the responder and adversary. Damage levels and detection are based on these factors and strategies of both players.

In the remaining part of this section, one step hide and seek game for responder and adversary will be modeled, characteristics of the game are analyzed, and an appropriate and efficient algorithm is developed to solve this problem.

Model Formulation

From now on, we will refer the adversary as hider, and refer responder as seeker. The hider is called player H, and seeker is called player S. The attack targets constitute a finite collection of cells $N_c := \{1, 2, \dots m\}$. For each cell, they have different detection possibility parameters for the seeker, λ_i. If λ_i is larger, bomb will be easier to detect in cell i. Each cell has a damage impact factor, C_i, if the bomb is detonated at cell i, damage cost for the seeker will be C_i. C_i is the positive payoff for the hider under this situation. So, this model is a zero-sum noncooperative two person game.

Now, we will look at the strategy space for hider and seeker. The strategy for hider will be the choice of cell to install and then detonate the bomb. In this model, the hider will choose a mixed strategy for placing the bomb. The mixed strategy is $\vec{p} := (p_1, \ldots, p_m)^T$, and p_i is the probability that the hider will hide and detonate the bomb in cell i. Strategy space for the hider will be $U_h := \{\vec{p} | \sum_{i \in N_c} p_i = 1, \vec{p} \in \Re^m_+\}$. For the seeker, we assume that there is a finite resource capacity X, then the strategy space for the seeker will be $U_s := \{\vec{x} = (x_1, \ldots, x_m) | \sum_{i \in N_c} x_i \leq X, \text{ and } \vec{x} \in \Re^m_+, \}$. Here, x_i is the amount of resource allocated to cell i by the seeker. Then we have the payoff function for hider,

$$J(\vec{p}, \vec{x}) = \sum_{i \in N_c} p_i \cdot e^{-\lambda_i x_i} \cdot C_i$$

Then, for this zero-sum noncooperative two person game, the existence and uniqueness of the saddle point is easy to prove. This is because the payoff function $J(\vec{p}, \vec{x})$ is continuous and strictly convex on $\vec{x} \in U_s$, and it is also continuous and strictly linear on $\vec{p} \in U_h$.

Solution of the Game

The following lemma will be used to find the solution of the one step game presented in the above section.

Lemma 2 *For saddle point of this game, $(\vec{p^*}, \vec{x^*})$, it satisfies the following two properties,*

1. *If $f_i(x_i^*) < \max_i f_i(x_i^*)$, then $x_i^* = 0$;*

2. $\sum_{i \in N_c} x_i^* = X$.

where $f_i(x_i) = e^{-\lambda_i x_i} \cdot C_i$ for $i \in N_c$. □

Proof Let $J := \{j | f_j(x_j^*) = \max_i f_i(x_i^*), j \in N_c\}$.

Assume for some $j \in J^c$, $x_j^* > 0$, then by continuity of function $f_j(x_j)$, we can find an $\epsilon > 0$, let $x_j' = x_j^* - \epsilon > 0$, such that,

$$f_j(x_j^*) < f_j(x_j') < \max_i f_i(x_i^*)$$

Then for each $i \in J$, we can find a $\delta_i > 0$, such that,

1. $\sum_{i\in J}\delta_i \le \epsilon$ and let $x'_i = x^*_i + \delta_i$;

2. $f_i(x'_i) < \max_i f_i(x^*_i) \quad$ for $i \in J$;

3. $f_j(x'_j) \le f_i(x'_i) \quad$ for $i \in J$.

Then, for $k \in J\ \{j\}$, let $x'_k = x^*_k$, so,

$$\max_i f_i(x'_i) < \max_i f_i(x^*_i)$$

And $\vec{p^*}$ is still the optimal response for seeker's strategy $\vec{x'}$. Then, $(\vec{p^*}, \vec{x'})$ is a better situation for seeker, and hider can not improve his/her own payoff under the new seeker's strategy $\vec{x'}$. So, finally we get that $(\vec{p^*}, \vec{x^*})$ is not a saddle point solution for this game. By contradiction, we know that, for any $j \in J^c$, $x^*_j = 0$, so the proof of part 1 is finished.

For the second part, $\sum_{i\in N_c} x^*_i = X$, we also use contradiction to prove it. Assume $\sum_{i\in N_c} x^*_i < X$, then let $K = X - \sum_{i\in N_c} x^*_i$, then for any $j \in J$, let $x'_j = x^*_j + \frac{K}{l}$, here $l = |J|$ is the number of elements in J. And for $i \in J^c$, let $x'_i = x^*_i$, then we have,

$$\sum_{i\in N_c} x'_i = X.$$

And so, it is obvious that

$$\max_i f_i(x'_i) < \max_i f_i(x^*_i).$$

This means that the seeker can reduce his/her cost by changing strategy unilaterally. Then, $(\vec{p^*}, \vec{x^*})$ will not be a saddle point solution. By contradiction, we know $\sum_{i\in N_c} x^*_i = X$, then the proof for part 2 is finished. ∎

For a more clear look at the saddle point solution for this one step game, we draw the graph for *security strategies* of both players in an example. *Security strategy* is the best strategy for one player, assuming that his/her opponent will apply the optimal response strategy corresponding to every one of his/her strategies. So, for the hider in this game, his/her problem will be a *Min-Max problem*, and for the seeker, it will be a *Max-Min problem*, based on $J(\vec{p}, \vec{x})$. The following two graphs are security strategies for hider and seeker for a 2 cells game.

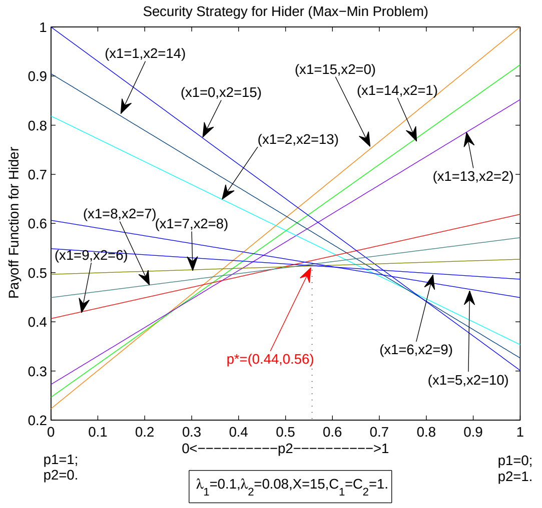

Figure 4.1: Hider's Security Strategy m=2 (Min-Max Problem)

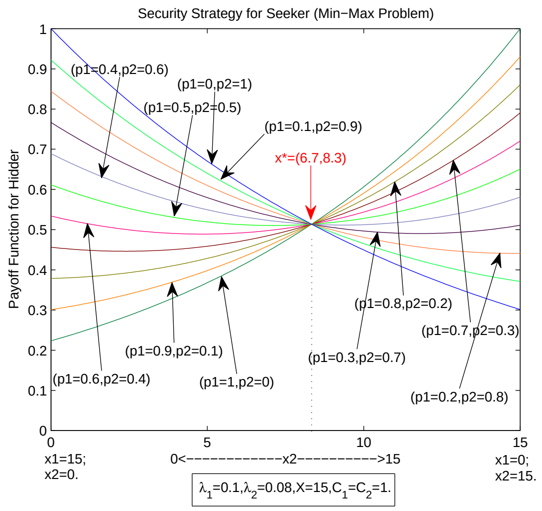

Figure 4.2: Seeker's Security Strategy m=2 (Max-Min Problem)

In Figure 4.1, it is a lower envelop for the hider, and the security strategy for hider will be the highest point of this lower envelop. And in figure 4.2, it is an upper envelop for the seeker, and then the security strategy for the seeker will be the lowest point of this upper envelop. And the security value for both players equal to each other, it is 0.5138, so the security strategies for both players obtained above are saddle point equilibrium solution.

We also present a 3 cell example in Figure 4.3 and 4.4. The problem also admits the same security level, and the saddle point equilibrium is obtained.

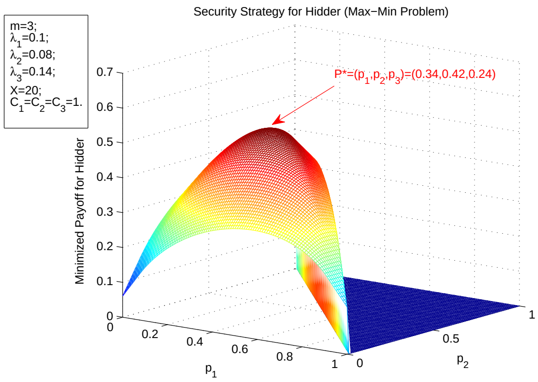

Figure 4.3: Hider's Security Strategy m=3 (Min-Max Problem)

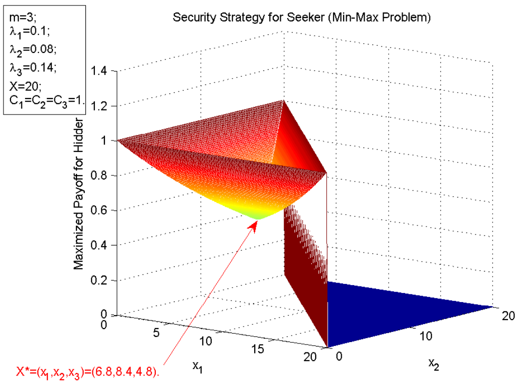

Figure 4.4: Seeker's Security Strategy m=3 (Max-Min Problem)

4.2.2 Summary

One-step hide and seek game is considered in this section. Next, dynamic game between first responder and adversary will be investigated for development of dynamic security strategies and adversary's attacking strategy.

4.3 Security Games

Dynamic infrastructure security game is developed in this section. A finite time period is set, adversary choose one cell to attack in the infrastructure, first responder will try to develop patrol sequence strategy to minimize possible risks. And in this problem, human life losses will be the risk measures considered.

4.3.1 Introduction

Due to ever increasing petroleum prices, population and economic growth, changes in lifestyles (employees living far from their workplace due to cheaper housing costs), etc., demand for mass transportation has increased exponentially. The U.S. is one of the most mobile nations in the world, providing over 4 trillion miles of passenger travel annually Fed (1999). Further, every workday, about 14 million Americans use some form of public transit USG (2002). Public transit users made more than 9 billion unlinked trips using the more than 6,000 transit properties in 2001[1]. As a consequence of this high volume, almost a third of terrorist attacks worldwide target public transit[2]. Successful and attempted terrorist attacks throughout the world such as New York, Madrid, London, Mumbai and Russia have forced governments to devote significant time, money and resources all in a concerted effort in order to secure transportation infrastructure.

Despite these vulnerabilities, public transit facilities are being used more frequently by ordinary citizens. As energy prices are skyrocketing, it is expected that, in the near future, public transit will carry ever more commuters to work, school and shopping. Therefore, increasing transit usage puts public-transit facilities under much more significant security threat. Public transit systems by design are open structural environments equipped to move large numbers of mass transit patrons in an effective and efficient manner. Therefore mass transit systems are considered *soft targets* and inherently are vulnerable and susceptible to terrorist attacks and because of the continuous hours of service cannot be closed and secured like other sectors of the area transportation system Loukaitou-Sideris et al. (2006). There

[1]http://www.apta.com/research/stats/

[2]Congressional Research Service, Transportation Issues in the 107th Congress, Washington, D.C., 2002

are other similar systems. The 2008 Mumbai attacks involved more than ten coordinated shootings and bombings across Mumbai, India's financial capital and its largest city. Attacks happened at various places, including hospital, hotels, railway station, cinema, and college. We can see from these events that attackers' prime target remains to be mass human casualties in addition to panic and chaos. The threat to any given infrastructural component or "infrastructure" could be substantially reduced by analyzing the risk associated to each transit infrastructure, mitigation planning, and employing best prevention and response policies.

To summarize, emergencies are natural or human-made and can be catastrophic events that have the potential to affect the entire transportation infrastructure such as major highways, bridges as well as mass transit stations, utilities, terminals, airports, etc. Emergency management problems have the following characteristics: a) *They are stochastic;* the resource (personnel, funds, etc.) availability, as well as occurrences and characteristics of emergencies, size of the affected population are all subject to randomness. b) *They are resource-allocation problems;* resources should be allocated wisely and rapidly among various locations, facilities in an infrastructure. c) *They are network problems;* adversary's attacks occur randomly over the infrastructure, and search patrol routes should be chosen carefully based on topology of the infrastructure.

In this section, we approach the infrastructure security problem via game theory by modeling it as a hide and seek game, Garnaev (2000). For a set of sites that could be the hiding place of an adversary or bomb, the probability of finding and capturing the adversary, i.e., probability of detection, depends on the conditions at the site, such as low visibility, heavy foliage, number of people, etc. Thus, the emergency-management problem is to allocate resources among the sites in an optimum way, to seek and dispose the threat in a timely fashion, and ultimately, minimize the risk. We discuss our model in general terms below.

In the first model, a first responder (emergency-management center) allocates resources (emergency personnel, or personnel-hours) to the sites of interest in an attempt to find an object (person or bomb, "adversary") that has been hidden, while the adversary selects a set of best sites to attack. Once the object is hidden, it cannot move during the search process. Similarly, the first responder can only act once, so a static game between first responder and adversary is formed. Throughout we use first responder and defender, and adversary and attacker synonymously.

In the next level of the problem, although the adversary is immobile, the responder can move among the sites. The probability of finding the adversary, and most importantly the risk associated with each site, are assumed to be known. Risk here might be defined as the expected cost of a successful attack. Thus, it is the multiplication of the threat (probability of an attack), the vulnerability of the infrastructure (probability that an attack at that site will be successful) and the consequence of an attack. Focusing on the severe attacks, we consider the loss of human life as the consequence. This measure typically depends on the occupancy level of the facility. In this paper, we assume that the occupancy level within the infrastructure can be estimated over time.

First, we briefly review the relevant literature. In section 4.3.3, we introduce the static game and present examples. In section 4.3.4, dynamic game between an immobile adversary and a mobile first responder is discussed. In section 4.3.5, numerical results are presented for an illustrative example. Finally, further applications and future research directions are discussed.

4.3.2 Related Work

There has been a recent interest in the infrastructure security related issues. Paté-Cornell (2002a) considers the fusion of intelligence information from various sources and types in a timely fashion. Paté-Cornell (2002b) presents a model for setting priorities among defense countermeasures; that combines different types of adversary and threat scenarios, the vulnerabilities of potential targets, and the consequences of different attack scenarios.

In Pita et al. (2008), a game theoretical model that could be utilized to secure an airport is developed. A software decision support system called ARMOR is introduced, that casts the police patrolling/monitoring problem as a Bayesian Stackelberg game. In such models, also called the *interdiction* models, the *leader* and the *follower* decide on their actions in a sequential manner. It is assumed that both the leader and the follower have perfect knowledge about system states and each other's actions. ARMOR has already been deployed at the Los Angeles International Airport to randomly assign checkpoints on roadways entering the airport, and assign canine patrol routes within the terminals Paruchuri et al. (2005), Paruchuri et al. (2006, 2007, 2008). In this game, leaders are the police, and followers are the attackers. Police set up checkpoints first, and the attackers will then choose their actions based on the current set of checkpoints. Brown et al. (2006) consider various Stackelberg games with the attacker or the defender as leader and give their mathematical programming formulations. Wood (1993), Morton et al. (2007), Lim and Smith (2007) look into the network interdiction models in which the reliability/survivability of a network becomes one of the performance measures. Zhuang and Bier (2007, 2009), Dighe et al. (2009), and Zhuang et al. (2010) study secrecy and deception in an attacker-defender resource allocation problem. Weaver et al. (2001) consider the terrorist's decision making process. Other recent papers include Nie et al. (2009a,b), where a passenger classification problem is analyzed. The same authors also discuss the optimal placement of suicide bomber detectors on a grid structure Nie et al. (2007). Hochbaum and Fishbain (2009) investigate the allocation of mobile sensors in an urban environment in order to detect "dirty bombs". Note that, the models in Nie et al. (2007, 2009a,b) only involve a single controller not multiple decision makers as in a game.

There are also other recent research on games related to our model, mostly on the static or dynamic hide and seek game. For a thorough introduction of search games, please see Garnaev (2000); Alpern and Gal (2003); Dobbie (1968). Here, we will review some of the prior work. A typical search game is a zero-sum game between a searcher and a hider where the objective of the searcher is to minimize the time to find the hider. It is assumed that the search terminates when the opponents are in close proximity. Gal (1979) considers search games in which the searcher moves along a continuous trajectory until s/he captures the hider, in either a network or a two (or more) dimensional space. The mobile and immobile hider cases are both analyzed, the upper and lower bounds of the value of the game are given. In Alpern et al. (2008), network search games with an immobile hider are presented assuming that the searcher can chose the starting point. The searcher's objective is to minimize the time to find the hider. In the case the network is simply searchable, they show that the optimal search strategy is a random *Chinese Postman (CP)* path Edmonds and Johnson (1973).

The dynamic search games Garnaev (2000); Alpern and Gal (2003), are sometimes also called pursuit-evasion games Isaacs (1965); Suzuki and Yamashita (1992). These games are, in general, defined in continuous-time and continuous state spaces, since they were initially developed for military applications. Thomas and Washburn (1991) describe a dynamic search game with discounting over a continuous time horizon. There are a finite number of cells to be searched, and the searchers try to find the hider as soon as possible. Recent approaches include integration of pursuit-evasion games with map building in an unknown environment Vidal et al. (2002). Alpern et al. (2009) considers a facility as a graph, where an attacker may choose several consecutive time periods, uninterrupted by patroller, to attack. Patroller can follow any path on the graph and can find the attacker when they are at the same node. The objective of the patroller is to minimize the time to detect the attacker. Baston and Garnaev (2000) analyze a static search game between a searcher and a protector, and present an algorithm and prove its effectiveness. In Hespanha et al. (2000), a dynamic hide and seek game on a cell based structure is presented. A greedy algorithm, that solves a one-step Nash Equilibrium is proposed. Hohzaki (2007) introduces a search allocation game (SAG). In Jotshi and Batta (2008), a dynamic hide and seek game on a network is discussed. In this game, the attacker is immobile, and the searcher tries to select paths in the network to detect the attacker as soon as possible.This problem is similar to ours. However, we focus on the consequence of an attack, not only the detection time. Alpcan and Basar (2006) consider a two-player zero-sum stochastic game which models the interaction between malicious attackers to a cyber-system, and the Intrusion Detection System(IDS).

4.3.3 Static Game Model

In this section, we consider the one-step security problem. The adversary and the first responder simultaneously choose their strategy over the potential sites. Payoff matrices for both responder and adversary are based on the occupancy level of each cell in the infrastructure. We assume that the adversary has the same information about occupancy and detection probability in each cell as the responder. We assume that impact of an attack will be based upon the occupancy level of the specific cell where the attack happens and can only endanger the human lives in that cell. People in neighbor cells will not be hurt directly due to this attack. As such, we utilize the expected number of casualties as the risk measure.

Assuming that the infrastructure is an m by n grid, let $N = mn$, be the total number of cells the responder and the adversary can occupy (possible actions for both players). Bold case letters will represent vectors, for example, the occupancy vector is denoted as $\mathbf{O}$ where $O(i)$ is the occupancy level of cell $i = 1, 2, \ldots N$. The expected number of casualties in cell i is evaluated as $C_i = \alpha(i) * O(i)$, where $\alpha(i)$ denotes the probability that a person in cell i will be hurt if an attack happens. If the adversary is detected, all these people are considered to be saved. Let (i, j) denote the location of the first responder and the adversary, respectively. In the case that both of them are in the same cell, (i, i), the adversary can be found by the responder with detection probability given by $P_D(i, i) \triangleq d_i$, otherwise $P_D(i, j) \triangleq 0$ for $i \neq j$. Therefore, the corresponding elements in the cost matrix for the responder when the responder is in cell i and the adversary is in cell j are,

$$(1 - P_D(i, j)) * C_j - P_D(i, j) * C_j = (1 - 2P_D(i, j)) * C_j, \quad \forall i, j = 1, 2, \ldots, N. \tag{4.1}$$

Then the cost matrix C for the first responder is an $N \times N$ matrix. The elements of this matrix are derived from equation (4.1) for the action pair (i, j) of the first responder and the adversary, respectively. For example, on a 2 by 2 grid infrastructure with 4 cells, we have the following cost matrix for the responder,

$$C = \begin{bmatrix} (1 - 2d_1)C_1 & C_2 & C_3 & C_4 \\ C_1 & (1 - 2d_2)C_2 & C_3 & C_4 \\ C_1 & C_2 & (1 - 2d_3)C_3 & C_4 \\ C_1 & C_2 & C_3 & (1 - 2d_4)C_4 \end{bmatrix} \tag{4.2}$$

Since the responder and the adversary have the same occupancy information, they will have the same risk measure C_i, and rewards to the responder will be considered as costs to the adversary. So, this game is a **zero-sum** matrix game, where the cost matrix for the adversary, C^a, is given by $C^a = -C$.

Next, we present a numerical example.

Example 1 Consider a 5×5 grid infrastructure with 25 cells. Assume that the occupancy vector $\mathbf{O}$, and the detection probability vector $\mathbf{d}$ on the grid are given in the 5×5 matrix form,

$$\mathbf{O} = \begin{bmatrix} 25 & 20 & 18 & 17 & 21 \\ 15 & 16 & 14 & 23 & 21 \\ 28 & 23 & 24 & 18 & 30 \\ 21 & 26 & 30 & 28 & 31 \\ 20 & 21 & 19 & 13 & 20 \end{bmatrix}, \qquad \mathbf{d} = \begin{bmatrix} 0.6 & 0.7 & 0.5 & 0.8 & 0.7 \\ 0.7 & 0.6 & 0.4 & 0.3 & 0.6 \\ 0.8 & 0.7 & 0.6 & 0.2 & 0.5 \\ 0.9 & 0.6 & 0.5 & 0.7 & 0.6 \\ 0.7 & 0.8 & 0.7 & 0.6 & 0.6 \end{bmatrix}.$$

And the casualty rate for all cells are $\alpha = 1$. Let $\mathbf{X}^*$ and $\mathbf{Y}^*$ denote the randomized Nash equilibrium decision vectors for the responder and the adversary, respectively. They are computed as:

$$\mathbf{X}^* = \begin{bmatrix} 0.0503 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.1007 & 0 & 0.0177 & 0 & 0.217 \\ 0 & 0.0804 & 0.217 & 0.115 & 0.2019 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{Y}^* = \begin{bmatrix} 0.138 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.0924 & 0 & 0.1438 & 0 & 0.138 \\ 0 & 0.1327 & 0.138 & 0.1056 & 0.1113 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The solution for the first responder prescribes the search of cells $\{1, 11, 13, 15, 17, 18, 19, 20\}$, while the same cells may be attacked with the given probabilities according to the adversary's decision rule $\mathbf{Y}^*$. □

For the two person zero-sum game, the mixed saddle-point strategies possess the ordered interchangeability property (refer to Basar and Olsder (1999)). This means that even if the Nash equilibrium is not unique, any combination of players' strategies provides the same game value.

Also, we notice that the search locations for first responder and adversary coincide with each other in most cases. The following proposition will provide some insights into this property.

Proposition 1 *For the two person zero-sum game under performance measure C given in equation 4.2, X^* and Y^* are the optimal mixed strategies for player 1 and 2, and Z^* is the value of the game. Then, if $X_j^* > 0$, then $Y_j^* > 0$.*

Proof For the two person zero-sum game in this problem, we can solve the following primal and dual linear programming models to obtain the optimal mixed strategy for players 1 and 2.

$$
\begin{array}{llr}
(P) & Min & Z \\
 & s.t. & C^T\mathbf{X} \leq Z \\
 & & \sum_{i=1}^{N} X_i = 1 \\
 & & \mathbf{X} \geq 0
\end{array}
\qquad\qquad
\begin{array}{lll}
(D) & Max & W \\
 & s.t. & C\mathbf{Y} \geq W \\
 & & \sum_{j=1}^{N} Y_j = 1 \\
 & & \mathbf{Y} \geq 0
\end{array}
$$

These are the primal and dual problems, $\mathbf{X}$ and Z ($\mathbf{Y}, W$) are the dual variables corresponding to the inequalities and the equality in the dual problem, respectively. At the optimal solution, $(\mathbf{X}^*, \mathbf{Y}^*)$, $Z^* = W^*$ gives the value of the game.

If $X_i^* > 0$, then by complementary slackness, $\mathbf{C}(i,\cdot)Y^* = W^*$, where $\mathbf{C}(i,\cdot)$ is the i^{th} row of C matrix. This implies,

$$\sum_{k=1}^{N} C_k Y_k^* - 2d_i C_i Y_i^* = W^*,$$

while for any other $j \neq i$, we have,

$$\sum_{k=1}^{N} C_k Y_k^* - 2d_j C_j Y_j^* \geq W^*.$$

There exists some l, $1 \leq l \leq N$, such that $Y_l^* > 0$, since $\sum_{j=1}^N Y_j = 1$. Then,

$$\sum_{k=1}^N C_k Y_k^* > \sum_{k=1}^N C_k Y_k^* - 2d_l C_l Y_l^* \geq W^* = \sum_{k=1}^N C_k Y_k^* - 2d_i C_i Y_i^*. \qquad \blacksquare$$

Clearly, the above inequality together with $Y_i^* \geq 0$, implies $Y_i^* > 0$.

Remark 1 Clearly, it also follows that, if $Y_j^* = 0$ then $X_j^* = 0$. The intuitive explanation for this property is that if the adversary does not attack certain cell j, the first responder should not allocate time or resources into that cell too. However, we can not validate the statement that if $Y_j^* > 0$ then $X_j^* > 0$. Instead, we have the following observations. $\square$

Remark 2 If the adversary chooses cell j, i.e., $Y_j^* > 0$, then the expected number of casualties in cell j should be at least equal to Z^*, i.e., $C_j \geq Z^*$. Furthermore, if $C_j > Z^*$, then $X_j^* > 0$. This could be seen again from complementary slackness arguments. If $Y_j^* > 0$, then $\mathbf{C}^T(j,\cdot)X_j^* = Z^*$, where, $\mathbf{C}^T(j,\cdot) = \mathbf{C}(\cdot,j)^T$. This implies,

$$C_j - 2d_j C_j X_j^* = Z^*,$$

giving $X_j^* = \frac{C_j - Z^*}{2d_j C_j}$. Since $X_j^* \geq 0$, this implies that $C_j \geq Z^*$. In fact, since for any constraint in the primal problem, we have,

$$C_j - 2d_j C_j X_i^* \leq Z^*,$$

if $C_j > Z^*$, $2d_j C_j X_i^* \geq C_j - Z^* > 0$, giving $X_j^* > 0$. $\square$

4.3.4 Dynamic Security Game

In this section, we consider a mobile first responder dynamically choosing cells to search for an immobile adversary. The first responder's objective is to develop a "best" patrol strategy to find the adversary with maximum reward or minimum cost. We assume that if the adversary is not caught within a finite time, say T, the adversary will launch the attack and destroy occupants in the cell s/he is in at time T. The first responder does not know the exact location of the adversary. Furthermore, the responder may not also have the current occupancy information. However, some sensory data are assumed to be available in order to develop a people flow model and make accurate estimates. Below, we discuss the exogenous flow model that will be used in the occupancy estimates. Then, we describe the methodology to obtain best strategies for both players.

People Flow Model

We first develop an exogenous people flow model that influences the decision making of the game players. Researchers have been using simulation models to describe the characteristics of pedestrian flow Yue et al. (2007), Hanisch et al. (2003). Yue et al. (2007) introduce a simulation model based on cellular automata on the square lattice with two-way and four-way pedestrian flow. In this simulation framework, pedestrian movement is more flexible and adaptive to dynamic conditions than vehicular flow. Hanisch et al. (2003) develop an online simulation tool for pedestrian flow in large public buildings, such as train stations, airports, shopping centers, etc. There is also research that concentrates on the occupancy estimation Meyn et al. (2009b); Niedbalski and Mehta (2007); Niedbalski et al. (2008). Meyn et al. (2009b) introduce a sensor-utility-network(SUN) method for occupancy estimation in buildings. Other studies have focused on the pedestrian flow in public buildings following special events, such as football matches, emergency fires, and terrorist attacks, etc., with crowd control and optimal evacuation as the main objectives Helbing et al. (2001); Klpfel et al. (2005); Klpfel and Meyer-Knig (2005); Deng et al. (2008); Meyn et al. (2009a). Alternating periods of congestion and slow movement dominate these cases, and most research on this topic utilize simulation models as in Bauer et al. (2007), and Deng et al. (2008).

For our purposes, we model people flow in a public building as a linear, stochastic dynamic system, and assume that some sensory information is available to be utilized in correcting the occupancy estimates. In this paper, we will not take account of the effects of special events, and also not consider crowd control problems.

At the microscopic level, people move like in an open queuing network where each cell is considered as a queueing station. Time is discretized, the time horizon is finite and is equal to T. Cells may have external arrivals and departures, from and to the outside, respectively, if there is direct connection to the outside, such as entrance doors, or train platforms from which they get on or off the train. Other cells in the building could be ticket offices, waiting rooms, food courts, shops, hall ways, etc.

For those cells with entrances, an arrival rate is estimated, vector $\boldsymbol{\lambda} \in \mathcal{R}^N$ represents arrival rate for each cell per unit time. Arrival rates may be time varying, as $\boldsymbol{\lambda}(m)$, representing peak and off peak hours during the day. At each time period, people move from one cell to the other according to the probabilities given by the routing matrix, $\mathbf{F}$. We assume that pedestrians are all similar, thus, they all have the same routing probability. These assumptions result in the following stochastic linear dynamic system of equations representing the pedestrian flow.

$$\mathbf{O_{m+1}} = \mathbf{F}^T \cdot \mathbf{O}_m + \lambda_{m+1} + \mathbf{W_{m+1}}, \qquad \mathbf{W_{m+1}} \sim \mathcal{N}(0, Q)$$

$$\mathbf{Z_m} = H \cdot \mathbf{O_m} + \mathbf{\Gamma_m}, \qquad \mathbf{\Gamma_m} \sim \mathcal{N}(0, R)$$

The first equation is the state equation, where $\mathbf{O_m}$ denotes the occupancy vector, and $\mathbf{W_m}$ denotes the process noise at time m, that is assumed to be normally distributed with covariance matrix Q. The second equation is the observation equation. $\mathbf{Z_m} \in \mathcal{R}^M$ is the measurement vector of actual occupancies at time m, H is the measurement matrix, and $\mathbf{\Gamma}$ denotes the measurement noise that is normally distributed with covariance matrix R. These measurements are obtained from video cameras, sensors, and other inspection methods. Here M may be less than N, meaning that not all cell occupancies may be available. However, we assume that the system is observable.

Kalman filter is utilized to predict and correct the occupancy level estimates. The *Kalman filter* is an efficient recursive filter that estimates the state of a linear dynamic system from a series of noisy measurements Kalman (1960, 1962); Kalman and Bucy (1961); Gürsoy and Baykal-Gürsoy (2010). At each time period, the responder will observe, $\mathbf{Z}_m$, the occupancy vector, and then utilizes these measurements to correct occupancy level forecasts. The equations for the Kalman filter are as follows,

$$\hat{\mathbf{O}}_{m+1}^{-} = \mathbf{F}^T \cdot \hat{\mathbf{O}}_m + \lambda_{\mathbf{m+1}}; \tag{4.3}$$

$$\hat{\mathbf{O}}_{m+1} = \hat{\mathbf{O}}_{m+1}^{-} + K_{m+1} \cdot (\mathbf{Z_{m+1}} - H\hat{\mathbf{O}}_{m+1}^{-}); \tag{4.4}$$

$$P_{m+1}^{-} = \mathbf{F}^T \cdot P_m \cdot \mathbf{F} + Q; \tag{4.5}$$

$$P_{m+1} = (I - K_{m+1}H) \cdot P_{m+1}^{-} \tag{4.6}$$

where P_m is the posterior and P_m^{-} is the *a priori* estimation error covariance matrix, respectively. K_{m+1} is the Kalman gain given by:

$$K_{m+1} = P_{m+1}^{-} H^T \cdot (HP_{m+1}^{-} H^T + R)^{-1}.$$

In these equations, $\hat{\mathbf{O}}_{m+1}^{-}$ and P_{m+1}^{-} are the forecasted values that are used to determine the initial position and initial patrol sequence. $\hat{\mathbf{O}}_{m+1}$ and P_{m+1} are the corrected values after each measurement. They are used at the beginning of each time period to reevaluate and update the original patrol sequence.

Example 2 Here is an example representing the cell occupancies in a 2×2 grid. This

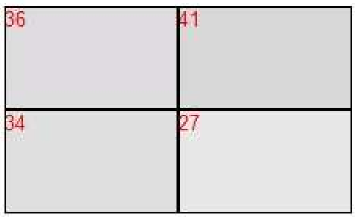

Figure 4.5: Cell Occupancy

figure is taken from the simulation package we have developed. The number in each cell represents the current occupancy level. Cells are numbered from 1 to 4, top left cell being 1, and bottom right cell being 4. Assuming that there are entrances in cells 1 and 4, we have the following arrival rate vector,

$$\boldsymbol{\lambda}^T = (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (\lambda_1, 0, 0, \lambda_4),$$

where bold letters denote vectors, and the superscript T denotes the vector transpose. Below is a 4×4 matrix representing the routing matrix.

$$\mathbf{F} = \begin{bmatrix} 0.2 & 0.3 & 0.3 & 0 \\ 0.2 & 0.3 & 0 & 0.5 \\ 0.4 & 0 & 0.1 & 0.5 \\ 0 & 0.3 & 0.2 & 0.2 \end{bmatrix}$$

From this matrix, notice that people leave the infrastructure from cell 1 with probability $1-0.2-0.3-0.3 = 0.2$, will go to cell 2 and 3 with the same probability 0.3, and will stay in cell 1 with 0.2 probability. Similarly, people leave cell 4 with probability $1-0.3-0.2-0.2 = 0.3$. Those cells that people can get in and out are the entrance doors, as well as the train platforms where they can get on and off the trains. □

Strategy Development

Occupancy estimates together with the detection probabilities establish the performance measure on the infrastructure grid. Note that only the first responder is mobile. Thus, after picking the initial locations the first responder can patrol the premises to find the adversary while the adversary remains at its initial location. Next, we describe the first responder's patrol strategy starting from an initial position. Then, we will discuss the first responder's and adversary's strategies for picking the best initial location.

First Responder's Strategy

Consider discrete time periods $\{m = 0, 1, 2, \ldots, T\}$ and the grid structure representing the infrastructure. The state of the system is given by $\{l^f_m, l^a_m, \mathbf{O_m}, P_m\}$, where l^f_m and l^a_m denote the position of the first responder and the adversary, respectively, at time m. Since the adversary is immobile, $l^a_m = l^a$. However, the adversary's location cannot be observed and the first responder only has information about his/her own location, i.e., l^f_m, and P_m, thus, arises the need to use the POMDP (partially observable Markov decision process) model to solve this problem.

Let the first responder's belief state be $\mathbf{b_m} = \{l^f_m, \mathbf{p^a_m}, \hat{\mathbf{O}}_\mathbf{m}, P_m\}$. Here, $\mathbf{p^a_m}$ is the vector of belief probabilities of the adversary's location. $\hat{\mathbf{O}}_\mathbf{m}$ is the estimated occupancy vector at time m. The value function for the responder at time $m = 0, \ldots, T-1$ is given by,

$$V_m^f(\mathbf{b_m}) = \max_{k \in A^f\{l_m^f\}} \Big\{ r(\mathbf{b_m}, k) + \gamma \cdot (1 - Pr\{\mathrm{D}|\mathbf{b_m}, k\}) \cdot V_{m+1}^f(l_{m+1}^f, \mathbf{p_{m+1}^a}, \hat{\mathbf{O}}_{\mathbf{m+1}}^-, P_{m+1}^-) \Big\} \tag{4.7}$$

where $r(\mathbf{b_m}, k)$ denotes the one-step expected reward function for the responder in belief state $\mathbf{b_m}$ when action k is applied, and γ denotes the discount factor with ($\gamma \leq 1$), k_m denotes the responder's action, i.e., the next cell in the patrol route, and $A^f\{l_m^f\}$ denotes the set of responder's possible actions at the next time period when the responder's current location is l_m^f. $Pr\{\mathrm{D}|\mathbf{b_m}, k\}$ is the probability of detecting the adversary given the belief state $\mathbf{b_m}$ and and applying action k. We assume that when the responder successfully finds the adversary, the game will end, and no more rewards will be earned. $\hat{\mathbf{O}}_{\mathbf{m+1}}^-$ and P_{m+1}^- are the corresponding forecasted values of the occupancy levels and the error covariance matrix. The one-step expected reward function is written as:

$$r(\mathbf{b_m}, k) = Pr\{\mathrm{D}|\mathbf{b_m}, k\}(\alpha(k)\hat{O}_m(k) + C) \tag{4.8}$$

where C is the terminal reward for apprehending the adversary, and $\alpha(k)$ is the casualty rate in cell k if attacked. Here $Pr\{\mathrm{D}|\mathbf{b_m}, k\}$ is equal to:

$$Pr\{\mathrm{D}|\mathbf{b_m}, k\} = d_k p_m^a(k). \tag{4.9}$$

Since the attack will materialize at the end of period T, the terminal value function is given as the negative of the expected casualty due to this attack,

$$V_T^f(b_T) = -\sum_{k=1}^{N} p_T^a(k) \cdot \alpha(k) \cdot \hat{O}_T(k). \tag{4.10}$$

Note that transitions from $\hat{\mathbf{O}}_{\mathbf{m}}$ and P_m to $\hat{\mathbf{O}}_{\mathbf{m+1}}$ and P_{m+1}, respectively, are associated with the people flow model and are given through equations (4.3 to 4.6) deterministically. Also because action k identifies the location of the first responder in the next time period, the transition probability for POMDP from state $\mathbf{b_m}$ to $\mathbf{b_{m+1}} = \{k, p_{m+1}^a, \hat{\mathbf{O}}_{\mathbf{m+1}}, P_{m+1}\}$ is as follows,

$$Pr\left\{\mathbf{b_{m+1}}|\mathbf{b_m},k\right\} = \begin{cases} 1 - Pr\{\mathrm{D}|\mathbf{b_m},k\}, & \text{for } \mathbf{p^a_{m+1}} \text{ as in equations (4.12,4.13)} \\ Pr\{\mathrm{D}|\mathbf{b_m},k\}, & \text{for } \mathbf{p^a_{m+1}} = \mathbf{e_k}. \end{cases} \tag{4.11}$$

In both cases above, the first responder moves to cell k to search for the adversary, so the location of first responder at time $m+1$ is cell k. In the first case, the first responder is not able to detect the adversary in cell k, so the belief probabilities will be updated through equations (4.12,4.13) given below. In the second case, the responder finds the adversary in cell k, so the belief probability vector is $\mathbf{e_k}$, the kth coordinate vector. Dynamic security game terminates after finding the adversary, all future value functions V_m^f will be 0, so the corresponding term in equation (4.7) is eliminated.

Here are the Bayesian update equations for belief probabilities. If the responder fails to detect the adversary at time m in cell k, then the belief probability of cell k is reduced while the belief probabilities of other cells are increased as given by:

$$p^a_{m+1}(k) = Pr\left\{l^a_{m+1} = k|\mathbf{p^a_m}, \text{Detection Failed}\right\} = \frac{(1-d_k)\cdot p^a_m(k)}{1-d_k\cdot p^a_m(k)}, \tag{4.12}$$

$$p^a_{m+1}(j) = \frac{p^a_m(j)}{1-d_k\cdot p^a_m(k)} \qquad \text{for } j \neq k, \tag{4.13}$$

where, d_k is the detection probability for cell k.

Bellman equation (4.7), representing POMDP, together with the terminal value function in equation (4.10), can be solved using dynamic programming algorithms. The solution provides a sequence of actions (patrol strategy) for the first responder, as well as the total expected risk or reward for the responder, i.e., the value function $V_m^f(\mathbf{b_m})$. In most cases, given the initial position of the first responder, the patrol strategy is deterministic. However, due to different detection results and changing occupancy levels, this patrol strategy will be reevaluated after each search based on the current flow information.

Next, we will discuss the initial position game between the responder and the adversary.

Responder and Adversary's Initial Position Game

The responder and the adversary must decide on their initial positions first. This is a static game between two players. To obtain the elements of the reward matrix, say (i, j)th element, first, the POMDP model generates an optimal patrol sequence starting from the initial position of the first responder (cell i). Then, given the initial position of the adversary(cell j), the expected rewards are calculated based on this patrol sequence. The expected rewards will be different from the value function obtained in the POMDP. The value function in the POMDP is the expected reward for every possible location of the adversary, however, the rewards we obtain for this static game is the expected rewards for fixed location combination (i, j) of the first responder and the adversary. In this initial position game, since future sensory information is not available at the time, forecasted occupancy levels are used. The elements of the reward matrix are obtained as follows,

$$R_f(i,j) = \sum_{m=0}^{T-1} Pr\{\text{D at time } m|\mathbf{S}, l^a = j\}\gamma^m \hat{O}_m(j) - Pr\{\text{No D within } T|\mathbf{S}, l^a = j\}\gamma^T \hat{O}_T(j), \tag{4.14}$$

where $\mathbf{S}$ is the patrol sequence obtained through the POMDP with $S(n)$ denoting the cell searched at time n, and $S(0) = i$. $Pr\{\text{No D within } T|\mathbf{S}, l^a = j\}$ denotes the probability that the adversary will not be detected, thus, the attack at time T will materialize, and is given as,

$$Pr\{\text{No D within } T|\mathbf{S}, l^a = j\} = \prod_{n=0}^{m-1} \left[1 - I\{S(n) = j\}d_j\right],$$

where $I\{\cdot\}$ denotes the indicator function for the event represented inside the parenthesis, i.e., if the event happens then the function takes value 1, otherwise its value is zero. $Pr\{\text{D at time } m|\mathbf{S}, l^a = j\}$ denotes the probability that detection of the adversary happens exactly at time m, given the first responder's patrol sequence and adversary's position, which can be written as,

$$Pr\{\text{D at time } m|\mathbf{S}, l^a = j\} = Pr\{\text{No D within } m-1|\mathbf{S}, l^a = j\} I\{S(m) = j\} d_j$$
$$= \prod_{n=0}^{m-1} [1 - I\{S(n) = j\} d_j] \cdot I\{S(m) = j\} d_j \qquad (0 < m < T)$$
$$Pr\{\text{D at time } 0|\mathbf{S}, l^a = j\} = I\{S(0) = j\} d_j$$

where $I\{S(m) = j\}$ is the indicator function of the first responder being in cell j at time m.

This static game is then solved to obtain the mixed strategy of initial position for first responder and the adversary. Note that with this randomized initial position for the first responder, the optimal patrol sequence will also be randomized.

4.3.5 Illustrative Example

In this section, we use an example to explain details of the POMDP model and initial position game. First, we describe the POMDP procedure, and then we present the results of the initial position game between the responder and the adversary. Together, an optimal patrol strategy for the first responder is developed.

A 3×3 grid infrastructure is considered, and cells are numbered from 1 to 9 from top left to bottom right. Total time periods are $T = 7$. The flow transition probabilities are given below as a 9×9 matrix,

$$\mathbf{F} = \begin{bmatrix} 0.1 & 0.3 & 0 & 0.2 & 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0.2 & 0.4 & 0 & 0.3 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0.1 & 0 & 0 & 0.4 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0.1 & 0.4 & 0 & 0.3 & 0 & 0 \\ 0 & 0.1 & 0 & 0.2 & 0.4 & 0.1 & 0 & 0.2 & 0 \\ 0 & 0 & 0.2 & 0 & 0.3 & 0.1 & 0 & 0 & 0.4 \\ 0 & 0 & 0 & 0.2 & 0 & 0 & 0.1 & 0.3 & 0 \\ 0 & 0 & 0 & 0 & 0.3 & 0 & 0.1 & 0.2 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 0.1 & 0 & 0.2 & 0.1 \end{bmatrix}$$

Arrival rates to each cell from outside is as follows,

$$\lambda = [25, \quad 0, \quad 20, \quad 0, \quad 0, \quad 0, \quad 10, \quad 0, \quad 15]^T$$

The number of arrivals per period is a Poisson distributed random variable with above arrival rates. From this, we can see that there are gates in cells 1, 3, 7 and 9 in this facility. Initial occupancy levels are given as,

$$\mathbf{O_0} = [50, \quad 32, \quad 41, \quad 42, \quad 80, \quad 35, \quad 51, \quad 45, \quad 39]^T$$

Casualty rate α is a one vector, which means the attack will kill every occupant in the cell.

Q is a diagonal matrix with diagonal elements given as

$$[25, \quad 36, \quad 50, \quad 40, \quad 37, \quad 28, \quad 48, \quad 35, \quad 40]$$

R is a diagonal matrix with diagonal elements as

$$[10, \quad 6, \quad 8, \quad 7, \quad 8, \quad 9, \quad 4, \quad 15, \quad 10]$$

P_0 is a diagonal matrix with diagonal elements as

$$[4, \quad 5, \quad 5, \quad 3, \quad 9, \quad 2, \quad 5, \quad 4, \quad 3]$$

And H is an identity matrix in this example, which means that the measurements are actually observations on the actual occupancy levels.

The detection probabilities for each cell are,

$$\mathbf{P}_D = [0.8, \quad 0.9, \quad 0.7, \quad 0.9, \quad 0.6, \quad 0.7, \quad 0.8, \quad 0.75, \quad 0.8]^T$$

The prior adversary-location belief probabilities are equally distributed among 9 cells.

Detection costs are all assumed to be negligible, and detection terminal reward is also zero. The discount factor is $\gamma = 1$, since this is a short period game, it is appropriate to assume that penalty and reward are not discounted in such a short period.

POMDP Results

First, assuming that the initial position of the responder is known, we use POMDP model to optimize the responder's patrol strategy for each time period. At each time period, the responder will observe, $\mathbf{Z}_m$, the occupancy vector through video camera, sensors or other methods, and then utilize these measurements to correct occupancy level forecasts. The actual occupancy levels and measurements for each time period are given as follows,

$$O = \begin{bmatrix} 50 & 97 & 52 & 42 & 47 & 45 & 46 & 39 \\ 32 & 39 & 67 & 63 & 62 & 65 & 63 & 51 \\ 41 & 92 & 59 & 79 & 54 & 64 & 50 & 58 \\ 42 & 54 & 44 & 44 & 48 & 43 & 42 & 53 \\ 80 & 79 & 91 & 111 & 113 & 115 & 113 & 112 \\ 35 & 35 & 56 & 45 & 57 & 44 & 50 & 43 \\ 51 & 60 & 34 & 32 & 16 & 45 & 27 & 30 \\ 45 & 42 & 59 & 50 & 49 & 53 & 53 & 53 \\ 39 & 71 & 50 & 74 & 64 & 64 & 59 & 59 \end{bmatrix}, \quad Z = \begin{bmatrix} 103 & 53 & 39 & 49 & 43 & 44 & 42 \\ 37 & 70 & 63 & 65 & 65 & 65 & 51 \\ 90 & 57 & 79 & 54 & 63 & 49 & 60 \\ 57 & 47 & 44 & 48 & 40 & 42 & 55 \\ 77 & 92 & 107 & 109 & 113 & 111 & 110 \\ 37 & 59 & 47 & 60 & 46 & 52 & 40 \\ 61 & 37 & 32 & 15 & 44 & 28 & 32 \\ 42 & 56 & 42 & 45 & 51 & 57 & 57 \\ 73 & 54 & 75 & 67 & 73 & 61 & 61 \end{bmatrix} \tag{4.15}$$

The actual occupancy levels O is a 9×8 matrix, formed by column vector O_0, O_1 to O_T. Measurements $\mathbf{Z_m}$ for each time period are given in matrix Z (9×7 matrix), and column m of Z, $\mathbf{Z_m}$, represents the measurements on actual occupancy levels at time m.

Let the first responder start from cell 4. If measurements in Z are not available, then the optimal patrol sequence for the responder is $\{4,5,8,9,6,3,2\}$. However, if they are available, then the first responder will correct the forecasts on occupancy levels, and then use the POMDP to develop a new patrol sequence which will only be used for the next move. The optimal patrol sequence in this case is $\{4,5,2,3,6,9,8\}$. This is different from the previous patrol sequence due to the Kalman filter correction procedure. Both patrol sequences will check same cells, but in different orders, thus will generate different expected rewards for the first responder.

The following two matrices contain patrol sequences for the first responder, starting from every initial cell, either without or with actual measurements. The first matrix gives the patrol sequences when Z is not available, and the second matrix are patrol sequences when Z is available. Each row represents one patrol sequence.

$$\begin{bmatrix} 1 & 4 & 5 & 2 & 3 & 6 & 9 \\ 2 & 1 & 4 & 5 & 8 & 9 & 6 \\ 3 & 2 & 1 & 4 & 5 & 8 & 9 \\ 4 & 5 & 8 & 9 & 6 & 3 & 2 \\ 5 & 8 & 9 & 6 & 3 & 2 & 1 \\ 6 & 9 & 8 & 5 & 4 & 1 & 2 \\ 7 & 4 & 1 & 2 & 5 & 8 & 9 \\ 8 & 9 & 6 & 3 & 2 & 5 & 4 \\ 9 & 8 & 5 & 4 & 1 & 2 & 3 \end{bmatrix} \qquad \begin{bmatrix} 1 & 4 & 5 & 2 & 3 & 6 & 9 \\ 2 & 1 & 4 & 5 & 6 & 9 & 8 \\ 3 & 2 & 1 & 4 & 5 & 8 & 9 \\ 4 & 5 & 2 & 3 & 6 & 9 & 8 \\ 5 & 8 & 9 & 6 & 3 & 2 & 1 \\ 6 & 9 & 8 & 5 & 2 & 1 & 4 \\ 7 & 4 & 1 & 2 & 5 & 8 & 9 \\ 8 & 9 & 6 & 3 & 2 & 5 & 4 \\ 9 & 8 & 5 & 4 & 1 & 2 & 3 \end{bmatrix} \tag{4.16}$$

Most of the patrol sequences are similar under both cases, except starting from initial cells 2, 4, and 6. The patrol sequences will be updated due to the information obtained through measurements. Updated patrol sequences usually generate better rewards for the first responder.

Initial Position Game

When the responder and the adversary decide on their initial positions, the responder does not have the information about future measurements provided in Z. So Z is not needed in this procedure. For each initial position of the first responder, the optimal patrol sequence is already given in the first matrix of equation 4.16, based only on the forecasted occupancy levels. The people flow model forecasts the future occupancy levels to provide expected risk measures throughout the finite time period T.

Given the position of both the first responder and the adversary, and the patrol sequence, the expected rewards can be easily calculated by using equation 4.14. So, R_f, the reward matrix for the first responder in the initial position game can be built.

Given the simplest case, that the initial position game is a zero-sum game, the following Nash Equilibrium strategies for the first responder and the adversary are obtained,

$$\begin{aligned} X^* =&(0.1464, \quad 0.0000, \quad 0.0000, \quad 0.4776, \quad 0.0000, \quad 0.0000, \quad 0.3556, \quad 0.0000, \quad 0.0204)^T \\ Y^* =&(0.0014, \quad 0.0000, \quad 0.4691, \quad 0.0000, \quad 0.0000, \quad 0.0038, \quad 0.5256, \quad 0.0000, \quad 0.0000)^T \end{aligned}$$

One can see that the first responder can choose from cell $1, 4, 7, 9$ as initial position, and the adversary can choose from cell $1, 3, 6, 7$ as attacking position. The game value for this initial position game is -5.1602. This means that the expected reward for the first responder is negative, so the first responder needs to improve the probability of detection or deploy more personnel to do the search.

Please note that this game value is calculated based on the forecasted occupancy levels without any correction. To calculate the actual expected reward of the first responder under strategy (X^*, Y^*), the actual occupancy levels O in equation 4.15 are used. Both the updated and non-updated patrol sequences are given in eq. 16. The actual expected reward obtained under updated patrol sequence is -3.2757, while it is -5.6281 when updates are not available. Clearly, updated patrol sequence is better. However, this is not always the case, there exists some cases in which updated sequences generate less expected reward than non-updated sequences. Generally speaking, when people flow experiences unusual shocks, such as sudden influx or out flux of people, updated patrol sequence will work much better.

On the other hand, if the first responder improves the detection probability to 1 for all cells, the actual expected rewards for non-updated and updated patrol sequences become 3.8816 and 3.9608. Updated patrol sequences generate slightly better reward for the first responder in this case too. They both are much better than the original case with lower detection probabilities. Of course, in this case, non-updated, updated patrol sequences and the mixed strategy for the initial position (X^*, Y^*) will all be different from the original case.

4.3.6 Summary

In this section, we consider static and dynamic game models for the infrastructure security problem. In these models, rewards and costs are based upon the occupancy level at each location in the infrastructure. Our models can be used in obtaining real time strategies for the infrastructure security personnel.

For the static game, we proved certain properties of the equilibrium. While for the dynamic game where the first responder is mobile we present a solution methodology that is based on the POMDP model. Throughout, examples are provided.

Next, we plan to study the two-controller resource allocation problem in which a number of sites (targets) are attacked by the adversary and are defended by the first responders. Depending on the players' objectives, such a problem can be modeled as a zero-sum stochastic game Baykal-Gürsoy (1989, 1991); Avsar and Baykal-Gürsoy (2006, 1999), or a Nash game Avsar and Baykal-Gürsoy (2002). Our approach will consider discrete and known environment, and incorporate risk measures into the objective function.

Chapter 5

Future Research

5.1 Stochastic Queuing Systems

5.1.1 Further Completion Time Analysis for Various Queuing Systems

In our research, completion time analysis has been done on one queuing systems. Since completion time in queuing systems correspond to travel time in traffic flow problems, research of completion time for other queuing systems is a clear direction for future research. And with more research on completion time for other queuing systems, different types of road links can be represented by these queuing systems, so travel time can be estimated with more accurate models. And also besides mean travel time, other characteristics of travel time can be derived from model results, such as standard deviation of the travel time, confidence interval etc.

5.1.2 Queuing Network Model for Transportation Network

In our research, single link travel time is estimated by using a queuing model with congestion and incidents. In future research, instead of a single link, travel time estimation in a transportation network will be our main task. Transportation network will be modeled as an open queuing network. In such a model, each road link on will be modeled as a Markov-modulated queue. Our previous queuing models can be adapted into this open queuing network, such as the finite server and capacity queue with incidents, infinite server queue with incidents, and state-dependent service rate queue for incidents and congestion combined. Different road conditions will be modeled as different queuing systems. For example, in city transportation network, finite server and capacity is a better fit for the situation, however in highway transportation network, an infinite server queue can be a good approximation for the system, since less congestion will happen, unless there is a major incident.

For illustrative purposes, consider a simple network between an $O-D$ pair composed of two parallel routes, 1st route has two links, and the other one has only one link, as shown in Figure 5.1. Assume that 1st route is a local road, and 2nd route is a section of the freeway. Under this situation, we will model the 2nd route as a $M/M/\infty$ queue with Markov-modulated service rates and general repair times due to incident. Both mean and variance for travel time can be obtained from the completion time analysis. For 1st route, two finite server queues $M/M/C$ and $G/M/C/C$ with incidents will represent link 1 and 2 for this route respectively. For first link, since we assume no arrival of vehicle will be lost, we need to use an infinite buffer queue $M/M/C$ to model this link. For second link, there is no room for vehicles to wait, if there are more vehicles, they will stay in the first link, so $G/M/C/C$ is an appropriate model for this link. And since arrival for second link is dependent on output of first link, we use a general arrival model to analyze. Under this situation, we can consider 1st route as a tandem queue model with incidents involved.

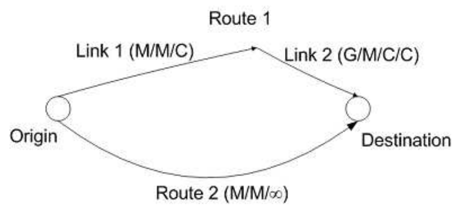

Figure 5.1: Analysis of Alternative Routes for 1 O-D pair with Queuing Model

In the previous example, 2nd route is freeway, but it is possible that there are multiple intersections along the way, such as in the following figure. Under this situation, the waiting time for each intersection along the road link will be estimated. Durations of red lights are constant, however for each individual car, when the car arrived at the intersection, remaining duration of the red light will be the car's waiting time. Once we have the arrival process for vehicles at the intersection, the distribution for waiting time in this intersection can be analyzed. After we have waiting times for each intersection, by combining with completion time analysis results for the whole road link without intersections, we can get the distribution for travel time on the road link with intersections.

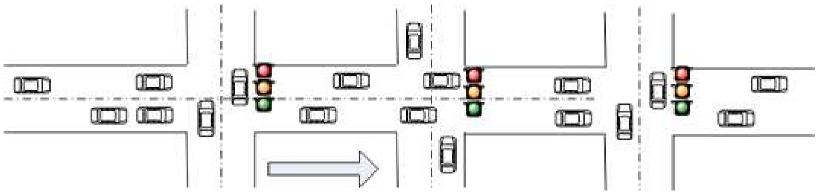

Figure 5.2: Freeway with Multiple Intersections

5.2 Infrastructure Security Games

5.2.1 Mobile Adversary Problems

Mobile adversary case should be considered to accommodate more general case of security problems. With mobile adversary, the problem is significantly complicated, and competitive POMDP model needs to be studied to represent this problem. Computation will be time consuming, and theoretical results will be even harder to achieve. However, this is the correct way to handle fast changing nature of adversary attacks. Simulation and heuristic algorithm can be a good way to reach satisfied results from the model.

5.2.2 Simulation Package

Simulation package should be developed to accommodate such a complicated problem with pedestrian flow simulation underneath. And real-time updates should be an option for improving the model results. It will certainly facilitate our research and application to real problems.

Bibliography

Conditions and Performance Report, chapter 1 Personal Mobility. Federal Highway Administration, Washington D.C., 1999.

Mass transit: Federal action could help transit agencies address security challenges. Technical report, US GAO, 2002.

M. Abramowitz and I. Stegun, editors. *Handbook of Mathematical Functions*. National Bureau of Standards, Washington D.C., 1964.

I. Adan and V. Kulkarni. Single-server queue with Markov-dependent inter-arrival and service times. *Queueing Systems*, 45:113–134, 2003.

A. Alfa and M. Neuts. Modelling vehicular traffic using the discrete time Markovian arrival process. *Transportation Science*, 29:109–117, 1995.

T. Alpcan and T. Basar. An intrusion detection game with limited observations. In *12th Int. Symp. on Dynamic Games and Applications, Sophia Antipolis, France*, 2006.

S. Alpern and S. Gal. Searching for an agent who may or may not want to be found. *Operations Research*, 50:311–323, 2002.

S. Alpern and S. Gal, editors. *The Theory of Search Games And Rendezvous*. Kluwer Academic Publishers, 2003.

S. Alpern, V. Baston, and S. Gal. Network search games with immobile hider, without a designated searcher starting point. *Int J Game Theory*, 37:281–302, 2008.

S. Alpern, A. Morton, and K. Papadaki. Optimizing randomized patrols. working paper, 2009, http://www.lse.ac.uk/collections/operationalResearch/research/workingPapers.htm, 2009.

T. Altiok. Queues with group arrivals and exhaustive service discipline. *Queueing Systems*, 2: 307–320, 1987.

B. Avi-Itzhak and P. Naor. Some queueing problems with the service station subject to breakdown. *Operations Research*, 11:303–320, 1963.

M. Z. Avsar and M. Baykal-Gürsoy. A decomposition approach for undiscounted two-person zero-sum games. *Mathematical Methods in Operational Research*, 49:483–500, 1999.

M. Z. Avsar and M. Baykal-Gürsoy. Inventory control under substitutable demand: A stochastic game application. *Naval Research Logistics*, 49:359–375, 2002.

M. Z. Avsar and M. Baykal-Gürsoy. A note on two-person zero-sum communicating games. *Operations Research Letters*, 34:4:412–420, 2006.

T. Basar and G. J. Olsder, editors. *Dynamic Noncooperative Game Theory.* SIAM Series in Classics in Applied Mathematics, Philadelphia,, 1999.

V. Baston and A. Garnaev. A search game with a protector. *Naval Research Logistics*, 47:85–96, 2000.

D. Bauer, S. Seer, and N. Brandle. Macroscopic pedestrian flow simulation for designing crowd control measures in public transport after special events. In *SCSC Proceedings of the 2007 summer computer simulation conference*, pages 1035–1042, 2007.

M. Baykal-Gürsoy. A sample-path approach to stochastic games. In *Proceedings IEEE Conference on Decision and Control, Tampa*, 1989.

M. Baykal-Gürsoy. Two-person zero-sum stochastic games. *Annals of Operations Research*, 28: 135–152, 1991.

M. Baykal-Gürsoy and W. Xiao. Stochastic decomposition in M/M/∞ queues with Markov-modulated service rates. *Queueing Systems*, 48:75–88, 2004.

M. Baykal-Gürsoy, W. Xiao, and K. M. A. Ozbay. Modeling traffic flow interrupted by incidents. *European Journal of Operational Research*, 195:127–138, 2009.

G. Brown, M. Carlyle, J. Salmeron, and R. K. Wood. Defending critical infrastructure. *Interfaces*, 36:530–544, 2006.

X. Chao and Y. Zhao. Analysis of multi-server queues with station and server vacations. *European Journal of Operational Research*, 110:392–406, 1998.

J. Cheah and J. Smith. Generalized M/G/C/C state dependent queuing models and pedestrian traffic flows. *Queueing Systems*, 15:365–385, 1994.

R. Cooper. Queues served in cyclic order: Waiting times. *Bell Syst. Tech. J.*, 49:399–413, 1970.

C. F. Daganzo. The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research-B*, 4:269–287, 1994.

C. F. Daganzo, editor. *Fundamentals of Transportation and Traffic Operations.* Pergamon-Elsevier, Oxford, U.K., 1997.

J. Darroch, G. Newell, and R. Morris. Queues for vehicle-actuated traffic light. *Operations Research*, 12:882–895, 1964.

B. D'Auria. Stochastic decomposition of the M/G/∞ queue in a random environment. *Operations Research Letters*, 35:805–812, 2007.

K. Deng, W. Chen, P. G. Mehta, and S. Meyn. Resource pooling for optimal evacuation of a large building. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, 2008.

N. S. Dighe, J. Zhuang, and V. M. Bier. Secrecy in defensive allocations as a strategy for achieving more cost-effective attacker deterrence. *International Journal of Performability Engineering*, 5: 31–43, 2009.

J. M. Dobbie. A survey of search theory. *Operations Research*, 16:525–537, 1968.

B. Doshi. Single server queues with vacations. In H. Takagi, editor, *Stochastic Analysis of Computer and Communication Systems*, pages 217–265. 1990.

M. Dunne. Traffic delays at a signalized intersection with binomial arrivals. *Transportation Science*, 1:24–31, 1967.

J. Edmonds and E. Johnson. Matching, Euler tours and Chinese postman. *Mathematical Programming*, 5:88–124, 1973.

M. Eisen and M. Tainiter. Stochastic variations in queuing processes. *Operations Research*, 11(6): 922–927, November-December 1963.

A. Federgruen and L. Green. Queueing systems with service interruptions. *Operations Research*, 34: 752–768, 1986.

A. Federgruen and L. Green. Queueing systems with service interruptions II. *Naval Research Logistics*, 35:345–358, 1988.

M. Fischer. An approximation to queueing systems with interruptions. *Management Science*, 24: 338–344, 1977.

D. Fitzgerald. Tricomi and kummer functions in occurrence, waiting time and exceedance statistics. *Stochastic Environmental Research and Risk Assessment*, 16:207–224, 2002.

S. Fuhrmann and R. Cooper. Stochastic decomposition in the M/G/1 queue with generalized vacations. *Operations Research*, 33:1117–1129, 1985.

S. Gal. Search games with mobile and immobile hider. *SIAM J. Control and Optimization*, 17: 99–122, 1979.

S. Gal and J. V. Howard. Rendezvous-evasion search in two boxes. *Operations Research*, 53:689–697, 2005.

A. Garnaev, editor. *Search Games and Other Applications of Game Theory.* Springer, 2000.

D. Gaver. A waiting line with interrupted service, including priorities. *Journal of the Royal Statistical Society. Series B (Methodological)*, 24:73–90, 1962.

D. Gazis, R. Herman, and R. Potts. Car-following theory of steady-state traffic flow. *Operations Research*, 7:499–505, 1959.

D. Gazis, R. Herman, and R. Rothery. Nonlinear follow the leader models of traffic flow. *Operations Research*, 9:545–567, 1961.

K. Gürsoy and M. Baykal-Gürsoy. Forecasting: State-space models and Kalman filter estimation. In *Wiley Encyclopedia of Operations Research and Management Sciences*. 2010.

S. Halfin. Steady-state distribution for the buffer content of an M/G/1 queue with varying service rate. *SIAM J. Appl. Math.*, 23:356–363, 1972.

A. Hanisch, J. Tolujew, K. Richter, and T. Schulze. Online simulation of pedestrian flow in public buildings. In *Proceedings of the 2003 Winter Simulation Conference*, 2003.

D. Heidemann. A queueing theory approach to speed-flow-density relationships. In *Proc. Of the 13th International Symposium on Transportation and Traffic Theory*, France, July 1996.

D. Heidemann. A queueing theory model of nonstationary traffic flow. *Transportation Science*, 35: 405–412, 2001.

D. Helbing. Theoretical foundation of macroscopic traffic models. *Physica A*, 219:375–390, 1995.

D. Helbing. Traffic and related self-driven many-particle systems. *Reviews of Modern Physics 73*, pages 1067–1124, 2001.

D. Helbing. A section-based queuing-theoretical traffic model for congestion and travel time analysis in networks. *J. Phys. A.*, 36:L593, 2003.

D. Helbing, A. Hennecke, and M. Treiber. Phase diagram of traffic states in the presence of inhomogeneities. *Phys. Rev. Letter*, 82, 1999.

D. Helbing, P. Moln'ar, I. J. Farkas, and K. Bolay. Self-organizing pedestrian movement. *Environment and Planning B-planning & Design*, 28:361–383, 2001.

J. Hespanha, M. Prandini, and S. Sastry. Probabilistic pursuit-evasion games: a one-step Nash approach. In *Decision and Control, 2000. Proceedings of the 39th IEEE Conference on*, 2000.

D. S. Hochbaum and B. Fishbain. Nuclear threat detection with mobile distributed sensor networks. *Annals of Operations Reserch*, 2009. published online.

R. Hohzaki. A search game taking account of attributes of searching resources. *Naval Research Logistics*, 55:76–90, 2007.

R. Isaacs, editor. *Differential Games*. John Wiley & Sons, 1965.

R. Jain and J. Smith. Modeling vehicular traffic flow using M/G/C/C state dependent queueing models. *Transportation Science*, 31:324–336, 1997.

A. K. Jayawardene and O. Kella. $M/G/\infty$ with alternating renewal breakdowns. *Queueing Systems*, 22:79–95, 1996.

A. Jotshi and R. Batta. Search for an immobile entity on a network. *European Journal of Operational Research*, 191:347–359, 2008.

R. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:35–45, 1960.

R. Kalman. Canonical structure of linear dynamical systems. *Proceedings of the National Academy of Sciences*, 3:596–600, 1962.

R. Kalman and R. Bucy. New results in linear filtering and prediction theory. *Trans. ASME J. Basic Eng.*, pages 83–95, 1961.

J. Keilson. Queues subject to service interruptions. *Ann. Math. Statistics*, 33:1314–1322, 1962.

J. Keilson and L. Servi. The matrix M/M/∞ system: Retrial models and Markov modulated sources. *Advances in Applied Probability*, 25:453–471, 1993.

B. Kerner. Three phase traffic theory and highway capacity. *Physica A*, 333:379, 2004.

H. Klpfel and T. Meyer-Knig. *Traffic and Granular Flow 03*, chapter Simulation of the Evacuation of a Football Stadium Using the CA Model PedGo, pages 423–428. Springer Berlin Heidelberg, 2005.

H. Klpfel, M. Schreckenberg, and T. Meyer-Knig. *Traffic and Granular Flow 03*, chapter Models for Crowd Movement and Egress Simulation, pages 357–372. Springer Berlin Heidelberg, 2005.

R. Kuhne, R. Mahnke, I. Lubashevsky, and J. Kaupuzs. Probabilistic description of traffic breakdowns. *Phys. Rev. E*, 65:6125, 2002.

R. D. Kuhne. In *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, 1987.

S. Lammer and D. Helbing. Self-control of traffic lights and vehicle flows in urban road networks. *J. Stat. Mech*, page 1742, 2008.

S. Lammer, R. Donner, and D. Helbing. Anticipative control of switched queueing systems. *Euro. Physics.J.B*, 63:341, 2008.

J. Lehoczky. Traffic intersection control and zero-switch queues. *J. of Appl. Prob.*, 9:382–395, 1972.

Y. Levy and U. Yechiali. Utilization of idle time in an M/G/1 queueing system. *Management Sc.*, 22:202–211, 1975.

G. Lia, D. Tapan, and H. Catherine. Dynamic late lane merge system at freeway construction work zones. *Transportation Research Record*, 2055:3, 2008.

M. Lighthill and G. Whitham. On kinematic waves: II. a theory of traffic on long crowded roads. In *Proc. Roy. Soc. London Ser. A 229*, pages 317–345, 1955.

C. Lim and J. C. Smith. Algorithms for discrete and continous multicommodity flow network interdiction problems. *IIE Trans.*, 39:1:15–26, 2007.

A. Loukaitou-Sideris, B. D. Taylor, and C. N. Y. Fink. Rail transit security in an international context. *Urban Affairs Review*, 41:727–748, 2006.

R. Mahnke, J. Kaupuzs, and I. Lubashevsky. Probabilistic description of traffic flow. *Phys.Rep.*, 408:1, 2005.

A. May and H. Keller. Non-integer car-following models. *Highway Res. Rec.*, 199:19–32, 1967.

S. P. Meyn, A. Surana, Y. Lin, and S. Narayanan. Anomaly detection using projective markov models in a distributed sensor network. In *CDC'09*, pages 4662–4669, 2009a.

S. P. Meyn, A. Surana, Y. Lin, S. M. Oggianu, S. Narayanan, and T. A. Frewen. A sensor-utility-network method for estimation of occupancy in buildings. In *CDC*, pages 1494–1500. IEEE, 2009b.

I. Mitrani and B. Avi-Itzhak. A many-server queue with service interruptions. *Operations Research*, 16:628–638, 1968.

D. P. Morton, F. Pan, and K. J. Saeger. Models for nuclear smuggling interdiction. *IIE Trans.*, 39:1: 3–14, 2007.

NCTIM. In *National Conference on Traffic Incident Management: A Road Map to the Future*, pages 2–4, March 2002.

M. Neuts. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. The John Hopkins University Press, 1981.

G. Newell. Approximation methods for queues with application to the fixed-cycle traffic light. *SIAM Rev.*, 7:2:223–240, 1965.

G. Newell. *Applications of Queueing Theory*. Chapman and Hall, London, 1971.

G. F. Newell. Nonlinear effects in the dynamics of car following. *Operations Research*, 9:209–229, 1961.

X. Nie, R. Batta, C. Drury, and L. Lin. Optimal placement of suicide bomber detectors. *Military Operations Research*, 12:65–78, 2007.

X. Nie, R. Batta, C. G. Drury, and L. Lin. The impact of joint responses of devices in an airport security system. *Risk Anal*, 29:2:298–311, 2009a.

X. Nie, R. Batta, C. G. Drury, and L. Lin. Passenger grouping with risk levels in an airport security system. *European Journal of Operational Research*, 194:574–584, 2009b.

J. S. Niedbalski and P. G. Mehta. Simulation and Estimation of Traffic Dynamics on a Graph. In *American Control Conference*, pages 5064–5069, 2007. doi: 10.1109/ACC.2007.4282795.

J. S. Niedbalski, K. Deng, P. G. Mehta, and S. Meyn. Model reduction for reduced order estimation in traffic models. In *American Control Conference*, pages 914–919, 2008. doi: 10.1109/ACC.2008.4586609.

C. O'Cinneide and P. Purdue. The M/M/∞ queue in a random environment. *J. Appl. Probab.*, 23: 175–184, 1986.

G. Pang and W. Whitt. Heavy-traffic limits for many-server queues with service interruptions. *Queueing Syst. Theory Appl.*, 61(2-3):167–202, Mar. 2009. ISSN 0257-0130. doi: 10.1007/s11134-009-9104-2. URL `http://dx.doi.org/10.1007/s11134-009-9104-2`.

P. Paruchuri, M. Tabme, F. Ordez, and S. Kraus. Safety in multiagent systems by policy randomization. In *SASEMAS*, 2005.

P. Paruchuri, M. Tambe, F. Ordez, and S. Kraus. Security in multiagent systems by policy randomization. In *International Conference on Autonomous Agents, Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, 2006.

P. Paruchuri, J. P. Pearce, J. Marecki, M. Tambe, F. Ordez, and S. Kraus. An efficient heuristic approach for security against multiple adversaries. In *AAMAS*, 2007.

P. Paruchuri, J. P. Pearce, J. Marecki, M. Tambe, F. Ordez, and S. Kraus. Playing games for security: An efficient exact algorithm for solving bayesian stackelberg games. In *AAMAS*, 2008.

E. Paté-Cornell. Fusion of intelligence information: A Bayesian approach. *Risk Analysis*, 22:445–454, 2002a.

E. Paté-Cornell. Probabilistic modeling of terrorist threats: a systems analysis approach to setting priorities among countermeasures. *Millitary Operations Research*, 7:5–20, 2002b.

S. Paveri-Fontana. On Boltzmann-like treatments for traffic flow: a critical review of the basic model and an alternative proposal for dilute traffic analysis. *Transportation Research*, 9:225, 1975.

H. J. Payne. Models of freeway traffic and control in g.a. bekey. *Mathematical Models of Public Systems*, 1, 1971.

J. Pita, M. Jain, J. Marecki, F. Ordez, C. Portway, M. Tambe, C. Western, P. Paruchuri, and S. Kraus. Deployed armor protection: the application of a game theoretic model for security at the los angeles international airport. In *International Conference on Autonomous Agents, Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems: industrial track*, 2008.

I. Prigogine and F. Andrews. A Boltzmann-like approach for traffic flow. *Operations Research*, 8(6): 789–797, 1960. ISSN 0030-364X.

I. Prigogine and R. Herman. *Kinetic Theory of Vehicular Traffic*. Elsevier, NY, 1971.

P. Purdue. The M/M/1 queue in a Markovian environment. *Operations Research*, 22:562–569, 1973.

H. Rakha and W. Zhang. Consistency of shock-wave and queuing theory procedures for analysis of roadway bottlenecks. In *TRB Annual Meeting CDROM, Paper # 05-1763*, 2005.

S. Redner, E. Ben-Naim, and P. Krapivsky. Kinetics of clustering in traffic flows. *Phys.Rev. E*, 50: 822, 1994.

P. Richards. Shock waves on the highway. *Operations Research*, 4:42–51, 1956.

J. Shanthikumar. On stochastic decomposition in M/G/1 type queues with generalized server vacations. *Operations Research*, 36:566–569, 1986.

V. Shvetsov and D. Helbing. Macroscopic dynamics of multilane traffic. *Phys.Rev. E*, 59:6328, 1999.

A. Skabardonis, K. Petty, P. Varaiya, and R. Bertini. Evaluation of the Freeway Service Patrol (FSP) in Los Angeles, ucb-its-prr-98-31. Technical report, California PATH Research Report, Institute of Transportation Studies, University of California, Berkeley, 1998.

I. Suzuki and M. Yamashita. Searching for a mobile intruder in a polygonal region. *SIAM Journal on Computing*, 21:5:863–888, 1992.

J. Tanner. A problem of interface between two queues, biometrica. *Biometrica*, 40:58–69, 1953.

L. C. Thomas and A. R. Washburn. Dynamic search games. *Operations Research*, 39:415–422, 1991.

N. Vandaele, T. VanWoensel, and N. Verbruggen. A queueing based traffic flow model. *Transportation Research-D: Transportation and Environment*, 5:121–135, 2000.

R. Vidal, O. Shakernia, H. L. Kim, D. H. Shim, and S. Sastry. Probabilistic pursuit-evasion games: Theory, implementation and experimental evaluation. *IEEE Trans. on Robotics and Automation*, 2002.

R. Weaver, B. G. Silverman, H. Shin, and R. Dubois. Modeling and simulating terrorist decision-making: A "performance moderator function" approach to generating virtual opponents. In *10 th CGF Proceedings, New York: SISO & IEEE*, 2001.

H. White and L. Christie. Queuing with preemptive priorities or with breakdown. *Operations Research*, 6:79–95, 1958.

G. B. Whitham, editor. *Linear and Nonlinear Waves.* Wiley, New York, 1974.

T. V. Woensel and N. Vandaele. Emprical validation of a queueing approach to uninterrupted traffic flows. *4OR-A Quart. J. of Operations Research*, 4(1):59–72, 2006.

T. V. Woensel, B. Wuyts, and N. Vandaele. Validating state-dependent queuing models for uninterrupted traffic flows using simulation. *4OR-A Quart. J. of Operations Research*, 4:159–174, 2006.

R. K. Wood. Deterministic network interdiction. *Math Comput Model*, 17:2:1–18, 1993.

M. Yadin and P. Naor. Queueing systems with a removable service station. *Operations Research Quart.*, 14:393–405, 1963.

U. Yechiali. A queueing-type birth-and-death process defined on a continuous-time Markov chain. *Operations Research*, 21:604–609, 1973.

U. Yechiali. Queues with system disasters and impatient customers when system is down. *Queueing Systems: Theory and Applications*, 56:195–202, 2007.

U. Yechiali and P. Naor. Queueing problems with heterogeneous arrivals and service. *Operations Research*, 19:722–734, 1971.

H. Yue, H. Hao, X. Chen, and C. Shao. Simulation of pedestrian flow on square lattice based on cellular automata model. *Physica A*, 384:567–588, 2007.

J. Zhuang and V. M. Bier. Balancing terrorism and natural disastersdefensive strategy with endogenous attacker effort. *Operations Research*, 55:976–991, 2007.

J. Zhuang and V. M. Bier. Secrecy and deception at equilibrium, with applications to anti-terrorism resource allocation. *Defence and Peace Economics*, 2009.

J. Zhuang, V. M. Bier, and O. Alagoz. Modeling secrecy and deception in a multiple-period attacker-defender signaling game. *European Journal of Operational Research*, 203:409–418, 2010.