# Optimal Evacuation Plans for Network Flows over Time Considering Congestion

Edward Pye Chamberlayne

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Industrial and Systems Engineering

Douglas R. Bish, Chair
C. Patrick Koelling
Hesham A. Rakha
Hanif D. Sherali

June 9, 2011
Blacksburg, Virginia

Keywords: network flows, optimization, microscopic traffic simulation, INTEGRATION, congestion, evacuation planning, capacity drop

# Optimal Evacuation Plans for Network Flows over Time Considering Congestion

Edward Pye Chamberlayne

(ABSTRACT)

This dissertation seeks to advance the modeling of network flows over time for the purposes of improving evacuation planning. The devastation created by Hurricanes Katrina and Rita along the Gulf Coast of the United States in 2005 have recently emphasized the need to improve evacuation modeling and planning. The lessons learned from these events, and similar past emergencies, have highlighted the problem of congestion on the interstate and freeways during an evacuation. The intent of this research is to develop evacuation demand management strategies that can reduce congestion, delay, and ultimately save lives during regional evacuations. The primary focus of this research will concern short-notice evacuations, such as hurricane evacuations, conducted by automobiles. Additionally, this dissertation addresses some traffic flow and optimization deficiencies concerning the modeling of congested network flows.

This dissertation is a compilation of three manuscripts. Chapters 3 and 4 examine modeling network flows over time with congestion. Chapter 3 demonstrates the effects of congestion on flows using a microscopic traffic simulation software package, INTEGRATION. The flow reductions from the simulation are consistent with those found in several empirical studies. The simulation allows for the examination of the various contributing factors to the flow reductions caused by congestion, including level of demand, roadway geometry and capacity, vehicle dynamics, traffic stream composition, and lane changing behavior. Chapter 4 addresses some of the modeling and implementation issues encountered in evacuation planning and presents an improved modeling framework that reduces network flows due to congestion. The framework uses a cell-based linear traffic flow model within a mixed integer linear program (MILP) to model network flows over time in order to produce sets of decisions for use within an evacuation plan. The traffic flow model is an improvement based upon the Cell Transmission Model (CTM) introduced in Daganzo (1994, 1995) by reducing network flows due to congestion. The flow reductions are calibrated according to the traffic simulation studies conducted in Chapter 3. The MILP is based upon the linear program developed in Ziliaskopoulos (2000); however, it eliminates the "traffic holding" phenomenon where it cannot be implemented realistically within a transportation network. This phenomenon is commonly found in mathematical programs used for dynamic traffic assignment where the traffic is unrealistically held back in order to determine an optimum solution. Lastly, we propose additional constraints for the MILP that improve the computational performance by over 90%. These constraints exploit the relation of the binary variables based on the network topology. Chapter 5 applies the improved modeling framework developed in Chapter 4 to implement a demand management strategy called *group-level staging* – the practice

of evacuating different groups of evacuees at different times in order to reduce the evacuation duration. This chapter evaluates the benefits of group-level staging, as compared to the current practice of simultaneous evacuation, and explores the behavior of the modeling framework under various objective functions.

The computational work was performed with support from the Simulation and Optimization Laboratory in the Grado Department of Industrial and Systems Engineering at Virginia Tech. This work has been partially supported by the *National Science Foundation* under Grant Number 0825611.

# Dedication

*to my parents who inspired me to begin my studies in engineering with, "What are you going to do with a liberal arts degree?!"*

# Acknowledgments

I have truly enjoyed my time back on campus at Virginia Tech studying and conducting research. For this, I have to first thank Dr. Doug Bish. His patience and work ethic has been absolutely amazing. We have worked extremely well together, completely in a creative and collaborative environment; it has been fun bouncing ideas off of each other and teaching each other in the elusive ways of AMPL, OPL, and LaTeX.

I would like to thank Dr. Pat Koelling for his support and guidance during my time in the ISE Department. His positive attitude and sense of humor has made this process much more enjoyable. I have thoroughly enjoyed working Dr. Hesham Rakha; his demanding traffic flow theory course sparked my interest in combining my background in Civil Engineering with operations research problem solving techniques. We have worked closely together on the development of the manuscript in Chapter 3 while discussing traffic flow theory, simulation, and, at times, politics and religion! It has been a distinct honor to have been a student in Dr. Hanif Sherali's non-linear optimization class. I thank him for his guidance and comments during the conduct of this research. Lastly, I would like thank Dr. Barbara Fraticelli for her exceptional teaching of Integer Programming and Advanced Topics in Mathematical Programming; one of the best professors I have ever had and she inspired me to continue working on my research.

It is only through the support, love, and encouragement of my wife, Allison, that I was able to complete this research and dissertation; you have allowed me to do difficult things during my career only because I knew how strong and capable you are. To my children, Emma and Eddie – I am extremely proud of the both of you and I thank you for your patience over the past few years.

I would like to thank LTC/Dr. Brian Lunday for setting an impossible example to follow and for his thoughtful advice and guidance during my entire PhD process here at Va Tech. He is an amazing officer, leader, intellectual, and friend. To the ISE Grad Student posse: a huge thanks to Andrew Henry who really welcomed me to the ISE Department; my office-mate Victor Pereira who has always been generous with his time, advice, and cachaça; to Jason Judd for his help with OPL programming and coding in general; huge thanks Hussein Tarhini for his assistance and development of a network flow simulation within MATLAB; to Fadel Megahed for his *LONG* emails and Hokie spirit; and to Adria

Markowski, Anna Schuh, Esra Ağca, and everyone else that helped me through my classes and research here – thank you and good luck!

I would like to thank the US Army and the Systems Engineering Department at the United States Military Academy for financially supporting my studies at Virginia Tech; I specifically want to thank BG Tim Trainor and Mr. Pat Driscoll for their support of my research and career.

In regards to proper attribution, I would like to acknowledge Dr. Hesham A. Rakha for writing the INTEGRATION overview section and general editing of the manuscript in Chapter 3. He also contributed to the editing of the manuscript in Chapter 4. As my advisor, Dr. Douglas R. Bish was involved in every step of our collaborative research; he contributed remarks and edits to Chapter 3, he led the research and provided frequent guidance and feedback in the development of the manuscripts in Chapter 4 and 5.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

This dissertation is dedicated to advancing the optimization of network flows over time in the development of emergency evacuation plans. This chapter discusses the motivation behind this research, the primary research contributions, and concludes with a description of the organization of the remainder of the dissertation.

## 1.1 Motivation

The coastal portions of the United States from Texas through Virginia are at high risk for hurricane events, while the rest of the Atlantic Coast is at moderate risk (USGS, 2005). These areas have been subject to frequent and costly hurricane evacuation events where traffic congestion has contributed to both the danger and loss. The images of incredibly congested highways and freeways have become common during large-scale emergency evacuations. Congestion negatively impacts the evacuation process in three main ways: 1) congestion slows the evacuation and can potentially expose evacuees to the very threat causing the evacuation; 2) congestion places evacuees in stressful and potentially life threatening traffic for many hours; and 3) congestion decreases compliance with local and regional evacuation orders.

There is a large need to develop plans and procedures to reduce congestion and better manage the immense demand on our critical freeways during an emergency evacuation. There have been many lessons learned from the Hurricane Katrina and Rita events; among them are recommendations to clearly communicate the hazard, when to order an evacuation, and

how to prioritize different populations in order to address the problem of congestion and to improve evacuation safety (Litman, 2006). These lessons learned, the need for better evacuation plans, and the unfortunate certainty that emergency evacuations will be more numerous in the future all serve as the motivation for the research within this dissertation. This research seeks to advance the modeling of network flows over time for the purposes of improving evacuation planning.

In order to combat the problem of congestion during an evacuation, transportation network managers can use supply and/or demand management strategies to either expand capacity or to best use the existing capacity of the network. Currently, most evacuation plans incorporate some use of supply management; the most common technique is contra-flow, or the practice of reversing lanes to direct all traffic lanes away from an impending hazard area. Although building new interstate systems or additional freeway lanes or reversing flow through contra-flow will certainly assist in addressing the problem of congestion, it is infeasible (if not impossible) that enough supply could ever be constructed or be made available to meet the peak demand during an evacuation due to economic and geographic constraints.

Demand management must be considered in addition to supply management in order to fully address the problem of congestion. Its use could save lives and shorten the duration of an evacuation. An important demand management strategy is to evacuate different groups of evacuees at different times in order to reduce the evacuation duration; this strategy is known as *staging*. Staging can be planned at the household-level or the group-level. A phrase coined in Sbayti and Mahmassani (2006) is that it is truly **"faster to wait"** during an evacuation; a plan that uses group-level staging to make some groups wait while other groups evacuate can reduce travel times, increase speed and flow, reduce congestion, and reduce the duration of an evacuation.

Although the concept of group-level staging makes intuitive sense, this research area is not well studied and there are not many analytical tools to implement this strategy into evacuation plans. The intent of this research is to present a group-level staging strategy, implemented within a network flow optimization model, that can be used to generate evacuation plans and policies that are an improvement to the current state of practice. This research is primarily focused on improving regional evacuations conducted by automobiles for short-notice evacuations – those evacuations in advance of a hurricane or wildfire where the time and location of impact are known (approximately) with some advance notice. However,

the modeling framework, insights, and conclusions could very well be applied to no-notice evacuations (e.g., nuclear reactor meltdown).

The two main tools available to develop network flow models for an evacuation plan are simulation and optimization. Traffic simulation is a powerful tool to use due to the ability to model detailed traffic behavior and evaluate key components of an evacuation plan. There have been a few simulation studies that have evaluated staging strategies in order to reduce the duration of an evacuation (Mitchell and Radwan, 2006; Chen and Zhan, 2008). However, these simulations only concluded that *some* staging strategies can shorten the duration of an evacuation over a simultaneous evacuation (the practice of evacuating all residents of a region at once); they could not determine the best possible staging strategy but only the best out of the set of strategies evaluated.

In contrast with simulation, an optimization model has the potential to produce an optimal staging strategy. An optimization can use macroscopic traffic flow models to calculate network flows in order generate an optimal set of decisions for use within an evacuation plan. Various different evacuation metrics (e.g., evacuation duration or travel time) can be used as objectives to drive the optimization toward a desirable set of decisions. The use of optimization to generate evacuation plans with demand management has been studied in Liu et al. (2006); Bish and Sherali (2011); Bish et al. (2011b) but more study is required specifically in the area of group-level staging. A disadvantage of optimization models is that they can over-generalize certain traffic behaviors; a common technique is use simulation to evaluate an evacuation plan developed through optimization.

The decisions made by an optimization model must be implementable based on existing traffic controls and within the existing operating budget. Decisions that give an optimal solution but are impossible to implement are not helpful for evacuation planning. The level of control given to the optimization model must be constrained to those areas where decisions should be made during an evacuation. For instance, model output that requires large changes in traffic flow over time along sections of freeway without traffic signals or intersections would be very difficult to implement within an evacuation plan.

Additionally, the optimization model must accurately determine traffic flows that realistically depict the conditions during an evacuation. As previously mentioned, congestion during an evacuation can be especially severe and can put evacuees at great risk and exposure. The traffic flow model embedded within the optimization model must account for the effects

of congestion to reduce ***both the speed and the flow*** of network traffic. This will lead to more realistic network flow estimates produced by the network flow optimization and will produce better evacuation plans. For instance, we find that by ignoring the flow reduction created by congestion, a network planner could underestimate the amount of time required to clear a population from an evacuation zone by 30-40%. By including the effects of congestion into an optimization model, better decisions will be made and better evacuations plans will be developed.

The research contained in this dissertation uses both simulation and network flow optimization in order to develop a group-level staging strategy for evacuation planning.

## 1.2 Summary of Research Contributions

The following paragraphs review research contributions and key findings made during the conduct of this research.

**Microscopic Traffic Simulation:** We were able to validate that the microscopic traffic simulation software package, called INTEGRATION, accurately reproduces the "capacity drop" phenomenon and results consistent with several well known empirical studies. This phenomenon occurs when the peak flow is considerably less in the congested regime than in free-flow regime. This is normally measured in empirical studies by measuring average queue discharge flows after congestion has formed upstream and comparing the flows prior to congestion forming. The findings from this research were that 1) the "capacity drop" can be simulated while using a continuous traffic flow model, 2) heavy truck percentages and realistic vehicle acceleration rates are important contributors, along with lane changing behavior, to creating congestion that reduces queue discharge flows, and 3) the onset of congestion is probabilistic in nature and can be shown using the INTEGRATION simulation software. In regard to the evacuation research, INTEGRATION was used to approximate the reduction in flow due to congestion.

**Cell Transmission Model with Flow Reductions (CTM-FR):** A macroscopic traffic model, based on the Cell Transmission Model (CTM) (Daganzo, 1994, 1995), was developed to reduce flows due to the effects of congestion. The CTM-FR is a more general form of

the CTM that includes a linear function that *reduces flow and speed* with increased density within the congested regime of the model. The model is equivalent to the CTM, as it was originally proposed, when this linear function predicts maximum flow capacity at jam density. The linear function can be calibrated by simulation results produced by INTEGRATION.

**The Flow Reduction Model (FRM):** The CTM-FR has been implemented within a mixed integer linear program, called the Flow Reduction Model (FRM), that calculates network flows for the purposes of dynamic traffic assignment (the process of routing traffic between specific origins and destinations). In addition to modeling the effects of congestion, the FRM also eliminates "traffic holding" where appropriate. Traffic holding is a direct result of using mathematical programs for dynamic traffic assignment; it occurs when traffic is artificially held back in a roadway segment in order to achieve an optimal value for a specific objective function. The development of the FRM allows for demand management strategies to be modeled and evaluated – specifically, group-level staging strategies. Additionally, improvements were made to the computational performance of the FRM in order to enable practical use of the mathematical program for planning and management applications. The performance of the FRM has been illustrated with two common freeway bottleneck examples.

**Demand Management with the FRM:** Additional constraints were developed for the FRM in order to implement a group-level staging strategy. This strategy produces an evacuation plan that is superior to a simultaneous evacuation in terms of evacuation time, travel time, and/or network clearance time. Several different objective functions were developed and evaluated with the FRM; a key finding is that the minimization of evacuation duration and network clearance time are NOT equivalent when the effects of congestion are considered and fixed loading curves are used to load demand onto the evacuation network. Additionally, the optimal sizing of control groups for the purposes of group-level staging is demonstrated. Lastly, further improvements are made to the FRM to remove traffic holding completely from intersections and to introduce some controls over intersection traffic priority.

## 1.3 Organization of Dissertation

This dissertation is organized in a manuscript format; the primary chapters are separate manuscripts. Chapter 2 reviews several areas of research that the work in this dissertation is based upon; some of the research areas will be reviewed in more detail within the manuscript chapters. Chapter 3 demonstrates that flow is reduced due to congestion within a microscopic traffic simulation software package; the results are consistent with several empirical traffic studies. Chapter 4 develops a mathematical program that reduces flows due to congestion for use as part of a network flow optimization model. This chapter develops a macroscopic traffic flow model based upon the Cell Transmission Model (Daganzo, 1994, 1995) and incorporates it within a mixed integer linear program. Chapter 5 builds upon the mathematical program developed in Chapter 4 within the context of an emergency evacuation. A group-level staging strategy is implemented within the network flow optimization model and intersection controls are added to the model. Several evacuation test networks are used to demonstrate the features of the model. Chapter 6 summarizes the main conclusions of the research and recommends some future research directions. The dissertation ends with a bibliography that includes all references cited within all of the chapters of the dissertation.

# Chapter 2

# Literature Review

This chapter reviews several areas of research that have contributed and influenced the research contained in the next chapters of this dissertation. The review begins with traffic flow modeling and discusses the primary strengths and weaknesses of each approach. The review then briefly discusses dynamic traffic assignment (DTA) and a particular example later used in this dissertation. Next, common evacuation modeling techniques are reviewed that build upon traffic flow models, simulation tools and DTA models. Lastly, some network flow optimization and demand management approaches are discussed that form the basis for the research in this dissertation.

## 2.1 Traffic Flow Modeling

According to Drew (1968), there are three main approaches to the modeling of traffic flows: a microscopic, a macroscopic, and a human factors approach. Microscopic traffic modeling involves the interactions between individual cars where macroscopic traffic modeling considers the movement of a traffic stream across entire transportation networks. A detailed discussion of the human factors approach is beyond the scope of this dissertation, however, microscopic modeling does consider a driver sensitivity factor. Additionally in Drew (1968), microscopic and macroscopic approaches are concisely described; "...[the microscopic approach] solves a differential equation of stream motion and a differential equation of continuity, both expressed in terms of speed 'u' and density 'k' to obtain an equation of state (the equation of the fundamental q-u-k traffic surface). The latter combines the differential equation of

motion for an individual vehicle together with the appropriate boundary conditions to obtain an equation of state." The following paragraphs will describe commonly used terminology, applications of each approach, and the strengths and weaknesses of each approach.

### 2.1.1 Microscopic Traffic Modeling

As stated above, microscopic traffic modeling concerns the interactions between individual cars. The models developed to approximate these interactions are called *car-following models*. In order to review some of the most significant car-following models and the application of them, we must first review commonly terminology. The primary reference for much of the following material was Rakha (2009a, Lecture Notes) and May (1990, Chapter 6).

#### 2.1.1.1 Terminology

Spacing $\{s_n(t)\}$, or distance headway, is the distance from a selected point on the lead vehicle to the same point on the following vehicle. It incorporates vehicle length and a stationary gap ($s_j$) (normally between rear bumper of the lead vehicle and front bumper of following vehicle). Another important parameter is time headway $\{h_n(t)\}$ which represents the physical separation in time between vehicles. It is the ratio of spacing and vehicle speed. Most models incorporate a driver sensitivity factor ($\lambda$). The variables normally used in car following models are vehicle position, speed, and acceleration. Note that location is a function of time, speed is a function of location and time, and acceleration is a function of speed, location, and time. The following are macroscopic traffic modeling parameters, for a given road segment, but are included here since microscopic modeling references them. Free-flow speed ($u_f$) is a steady-state speed between the speed limit and design speed. Jam density ($k_j$) is the maximum density of vehicles where speed and flow are equal to zero. Capacity ($q_c$) is the maximum flow possible. Speed-at-capacity ($u_c$) is the speed at which at which maximum flow is reached. Sometimes an optimum or critical density ($k_c$) is used which is the density that corresponds with the speed-at-capacity ($u_c$) and the maximum flow ($q_c$).

#### 2.1.1.2 Car-Following Models

Examples of car-following models are the Pipes model, GM models, and the Van Aerde - Rakha model. The Pipes model is a linear car following model in which the relationship between vehicle speed and spacing is linear. The model requires the calibration of free-flow speed, spacing, and the driver-sensitivity factor to local conditions. A form of the model is:

$$u_n(t) = min(\lambda[x_{n-1}(t-T) - x_n(t-T) - s_j], u_f) \tag{2.1}$$

where $u_n$ is speed of vehicle $n$, $x_n$ is the position of vehicle $n$, $t$ is time, and $T$ is reaction time. Since the Pipes model is linear, it does not always accurately model traffic behavior in various different conditions (Rakha and Crowther, 2002).

The GM models are a collection of five separate models which are car-following models that are non-linear functions of spacing and vehicle speed. These models are very flexible and can be adapted to various conditions much more easily than linear car-following models. Additionally, the Pipes model and Greenshields model can be represented by a GM model by varying two of the exponential variables.

The Van Aerde - Rakha car-following model is a combination of the Pipes and Greenshields car-following models (Rakha and Crowther, 2002). The model attempts to capitalize on the strengths of both models while eliminating the weaknesses. The principal weakness of the Pipes model is that it assumes that speed is insensitive to both traffic density and flow in the uncongested regime (Rakha, 2009b). The principal weakness of the Greenshields model is that it assumes the speed-flow relationship to be parabolic. The challenge of using the Van Aerde - Rakha model is that the calibration of the model requires the estimation of four parameters that are functions of the macroscopic parameters $q_c$, $k_j$, $u_f$, and $u_c$. The functional form this model will be introduced in the macroscopic traffic modeling section.

The most common application of microscopic traffic models is in simulation. All traffic simulation software uses one of the numerous car-following models (or a combination). The Pipes model is used in the CORSIM (Corridor Simulation) and VISSIM (Verkehr In Städten - SIMulationsmodell) simulation packages. INTEGRATION uses the Van Aerde - Rakha car-following model; this software package will be reviewed further in Chapter 3. Simulation software is the best way to accurately predict how actual traffic flows will behave. An example of when to use simulation would be to test signal timings at intersections or to evaluate adding

additional lanes to a specific section of roadway. The primary strengths of microscopic traffic modeling and simulation are model accuracy and realism. The disadvantages are the calibration efforts required to estimate model parameters, collecting and entering the data required to accurately model the environment in which the microscopic traffic model will be used, and computational time and effort required to execute the simulations.

### 2.1.2 Macroscopic Traffic Modeling

Macroscopic traffic modeling involves traffic stream models that are more concerned with the accurate modeling of traffic streams across entire networks rather than the individual car-following behavior. These traffic stream models are expressed in terms of jam density ($k_j$, expressed in veh/mile/lane or vpmpl), capacity flow ($q_c$, expressed in veh/hour/lane or vphpl), free-flow speed ($u_f$, expressed in miles/hr or mph), and speed-at-capacity ($u_c$, expressed in mph). All models use the fundamental relationship that traffic flow equals the product of density and speed: $q = ku$. The distinction between microscopic modeling and macroscopic modeling is not always clear since most macroscopic traffic stream models can be derived from microscopic car-following models. This was first proved by Greenberg when he related his macroscopic traffic model to a GM-3 car-following model (May, 1990, Chapter 10). However, the following sections will review some significant single regime and multi-regime traffic stream models. The primary references used for the background on traffic stream models were May (1990, Chapter 10) and Rakha (2009b, Lecture Notes).

The first single-regime model was developed by Greenshields in 1934. Single-regime means that all traffic conditions from free-flow to congested flow is represented by a single relationship. Greenshields theorized that there existed a linear relationship between speed and density. The Greenshields Model requires estimates for $u_f$ and $k_j$. Both are difficult to estimate but $u_f$ is approximately between the speed limit and the design speed and $k_j$ is in the range between 185-250 vpmpl. A disadvantage of the model is that it represents the speed-flow relationship as parabolic (Rakha and Crowther, 2002). The equation for the Greenshields Model is:

$$u = u_f - \frac{u_f}{k_j}k \tag{2.2}$$

Note: Since the Greenshields Model is based on the linear relationship between speed and

density, the following relationships can be derived: $u_c = \frac{u_f}{2}$ and $q_c = \frac{k_j u_f}{4}$.

The Greenberg Model was developed after the Greenshields Model and proposed that the relationship between speed and density was nonlinear. The model requires estimates for $u_c$ and $k_j$; both estimates are difficult to estimate from a given traffic stream but $u_c$ is approximately one half of the design speed. A large disadvantage of the Greenberg Model is that free-flow speed ($u_f$) is infinity. The advantage is the nonlinear relationship between speed and density is more realistic than in the Greenshields Model. In order to correct for the error in free-flow speed estimation, a third single-regime model was proposed by Underwood. The model requires estimates for $u_f$ and $k_c$; the optimum density is extremely difficult to observe and estimate. A large disadvantage of this model is that speed never reaches zero and $k_j$ is infinity. There were other single regime models developed that would generate an entire family of traffic stream models. Models developed by Drew and Pipes-Munjal are examples of these.

In an attempt to improve upon the deficiencies of the single-regime models, multi-regime models were developed. May (1990) lists the primary weakness of single-regime models as, "their inability to track faithfully the measured field data near capacity conditions." The free-flow and congested regimes are modeled separately within a two-regime model. Similarly, the free-flow, transition, and congested regimes are modeled separately in a three-regime model. In 1961, Edie proposed a two-regime model that combined the benefits of the Underwood model in the free-flow regime and the Greenberg model in the congested regime. A team from Northwest University then proposed three multi-regime models. The Type I was a two-regime model that used the Greenshields model for both regimes separately. The Type II was a constant speed model for the free-flow regime and the Greenberg model for the congested regime. The Type III used a Greenshields model in three-regimes separately. The multi-regime models were in general seen as an improvement over the single-regime models but the advantages and disadvantages had to be weighed for the specific traffic conditions of a given traffic stream.

In 1995, the Van Aerde - Rakha traffic stream model was introduced. It is considered to be a single-regime nonlinear model. This model was already introduced with the car-following models. The terms of the model can be directly related to either car-following terminology or traffic stream model terminology.

The Cell Transmission Model (CTM) was introduced in Daganzo (1994, 1995) and is a

first-order macroscopic traffic stream model which uses cells to represent roadway segments and links between the cells to represent flow. It uses linear flow-density relationship to model traffic flow between 'sending cells' and 'receiving cells'. A more detailed review of the CTM will be provided in Chapter 4. The primary advantages of using the CTM, and other first-order macroscopic models, are its ability to properly model the spillback propagation of congestion and its tractability within mathematical programs.

However, there are several disadvantages to consider when using first-order models. In Papageorgiou (1998), it is argued that first-order models cannot be used effectively to develop strategies for controlling flows feeding into a freeway (known as "ramp metering") since first-order models will always predict maximum capacity flow regardless of the level of congestion. Gomes and Horowitz (2006) states that the use of first-order models is justified in a network where the bottlenecks are closely preceded by an off-ramp and when ramp metering cannot significantly affect congestion due to limited on-ramp storage space. However, Gomes and Horowitz (2006) also criticizes the use of first-order traffic models, especially when used within network optimization models, because they do not model the effects of congestion properly and are, "...incapable of exploiting the benefits of increasing bottleneck flow." One of the effects of congestion is called the "capacity drop" phenomenon; this occurs when the maximum flow achieved during congestion is lower than the maximum flow achieved during free-flow conditions. The literature concerning capacity drop will be reviewed in more detail in Chapter 3.

## 2.2 Dynamic Traffic Assignment

Regardless of what traffic flow model is used, in order to use a network flow optimization approach, a method of assigning traffic to routes and destinations, over time, needs to be used. The field of dynamic traffic assignment (DTA) is a broad area of research that addresses this method. Current research in the area of dynamic traffic assignment (DTA) is very active and has been approached in a variety of ways. In Peeta and Ziliaskopoulos (2001), DTA models were categorized into two groups: analytical approaches and simulation-based approaches. The analytical approaches were further divided into mathematical programming, optimal control, and variational inequality approaches. The focus of this literature review will be on the mathematical programming approaches.

The first DTA mathematical programs were presented in Merchant and Nemhauser (1978b,a). These programs were nonlinear and nonconvex, however, a piecewise linear program version was developed that could be solved more efficiently. These papers inspired a stream of mathematical programs. Carey (1987) reformulated Merchant and Nemhauser's mathematical program into a convex nonlinear program which could be solved much more easily. In Ziliaskopoulos (2000), a linear program for DTA was introduced that assigned traffic to routes using the CTM. The LP used a system-optimal (SO) objective function that minimized total travel time through the network. In Li et al. (2003), a Dantzig-Wolfe decomposition algorithm was used to reduce the Ziliaskopoulos SO LP formulation into minimum cost flow problems that are efficient and simpler to solve. In Tuydes and Ziliaskopoulos (2006), heuristics and tabu search algorithms were used in conjunction with the SO LP in order to find the best use of network capacity during an evacuation. In Ukkusuri and Waller (2008), the results from a SO and a user equilibrium (UE) linear programs are compared when applied to a network evacuation problem.

The CTM and the SO LP developed by Ziliakopoulos will be reviewed in Chapter 4 in more detail.

## 2.3 Evacuation Planning and Modeling

Evacuation modeling is the evaluation of evacuation plans and policies through the use of macroscopic and/or microscopic traffic simulation software. Hardy and Wunderlich (2007), a technical report prepared for the US Department of Transportation, presents a figure, shown in Figure 2.1, that summarizes some of the simulation packages used in evacuation modeling.

The current state-of-the-art in macroscopic evacuation modeling is found in OREMS (Oak Ridge Evacuation Modeling System) and ETIS (Evacuation Traffic Information System). A well written comprehensive review of these systems is found in Wolshon, Urbina, Wilmot, and Levitan (2005). OREMS was developed in 1995 and uses the CORFLO macroscopic traffic simulator. The model outputs link flows, total evacuation time, and evacuation routing and requires time-dependent travel demand as input (Wilmot and Mei, 2004). The development of OREMS is sponsored by the Federal Emergency Management Agency (FEMA) and is intended to replace DYNEV/I-DYNEV. DYNEV (Dynamic Evacuation Net-

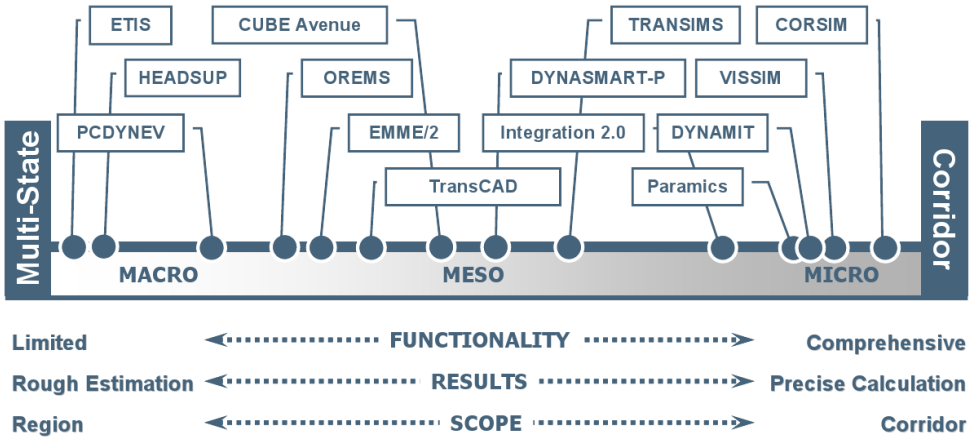



Figure 2.1: The Spectrum of Evacuation Modeling (used with permission of Matt Hardy, 2011)

work) is a macroscopic simulation model developed in the 1980's to evacuate populations from point disasters such as nuclear power plant meltdowns. Future development of OREMS is planned to incorporate microscopic traffic modeling (Abkowitz and Meyer, 2007). ETIS was developed after the damaging Hurricane Floyd in 2001 and is a web-based macroscopic model that estimates evacuation demand and traffic congestion. Both OREMS and ETIS involve the use of geographic information systems (GIS) to display results of the evacuation plan simulations and to perform some of the analysis.

One of the earliest macroscopic evacuation modeling systems developed was NETVACl (Network Evacuation Large) introduced in Sheffi et al. (1982). NETVACl was developed for evacuation away from nuclear power plants similar to DYNEV. The system included a simulator that used a linear speed-density relationship as in the Greenshields traffic stream model. Interestingly, the author mentions, "The specific model used, however, is not going to affect substantially the results since all traffic flow models exhibit similar behavior for very heavy traffic and the linear model is the simplest to use." This may be true for macroscopic models but we now have the ability use more accurate models within microscopic simulations. NETVACl inspired the development of MASSVAC (Hobeika and Jamei, 1985); MASSVAC was another macroscopic simulation but was intended for use in network evacuations in response to hurricanes and other large scale emergencies.

The use of microscopic simulation within evacuation modeling has increased recently as the computers that run the simulations have improved. Recently CORSIM and VISSIM

(based on the Pipes car-following model) have been used to evaluate lane reversal strategies as part of evacuation plans (Williams et al., 2007). In Mitchell and Radwan (2006), INTEGRATION was used to evaluate evacuee staging strategies in order to reduce the overall evacuation clearance time. The reason macroscopic models were first used because it was possible to get a feasible result for the purposes of regional evacuation plans; microscopic simulation (especially in the 1980's) was simply too expensive in time and resources to use for anything beyond small local examples. This is still a problem today as mentioned in Williams et al. (2007), "... there are scalability issues with broad area microsimulation both in terms of computation time and file size. Therefore, although microsimulation proved useful for offline plan analysis, modeling support for real-time management will likely come from macroscopic simulation."

## 2.4 Network Flow Optimization

The use of macroscopic and microscopic traffic simulation alone is not adequate for evacuation planning and modeling. The power of mathematical programs to optimize the components of evacuation plans have recently come into their own. Since the development of a system optimal (SO) linear program of the CTM introduced in Ziliaskopoulos (2000), multiple different mathematical programs have been developed for the purposes of evacuation planning and modeling. Daganzo's CTM was developed with the intention to be used for evacuation modeling, among other uses (Daganzo, 1995).

Before reviewing a few mathematical programs used for network flow optimization, it is important to note that these programs are not a replacement for simulation but a decision making tool to assist in forming evacuation plans. The best method to evaluate evacuation plans is to use microscopic traffic simulation or macroscopic traffic simulation when microscopic simulation is too cost prohibitive. Mathematical programs typically optimize the total evacuation time, total delay, total evacuee travel distance, or network clearance time (or a combination).

Some examples of the decision sets that can be produced through network flow optimization are:

1. Optimal times to issue evacuation orders
2. Optimal lane based routing plans (Cova and Johnson, 2003)

3. Time dependent Merge Priorities and Diverge Percentages
4. Evacuation shelter locations (Sherali et al., 1991)
5. Evacuation time estimates
6. Contraflow strategies (Tuydes and Ziliaskopoulos, 2006)

For example, in Cova and Johnson (2003) a mixed integer program (MIP) was used to calculate the optimal lane based routing plans as part of a minimum cost flow problem. Then the MIP results were evaluated within a microscopic traffic simulator (Paramics). This shows the inter-relationship between mathematical programs and microscopic traffic simulation that can create state-of-the-art evacuation models and ultimately evacuation plans.

As stated previously, much of the research involving the use of mathematical programs in evacuation planning and modeling started with the work performed by Ziliaskopoulos. In Ziliaskopoulos (2000), the CTM was integrated into a single, system-optimal (SO) linear program (LP) that minimized total travel time through the network using a set of linear inequality constraints. This model will be reviewed in more detail in Chapter 4.

In Liu et al. (2008), an integrated emergency evacuation system is developed with five components similar in structure to the original NETVACl but updated with the recent advances. The system consists of five modules: an input module, database module, optimization module, online simulator, and an output module. The optimization module is based on an LP that incorporates the CTM and generates the optimal routing, demand loading, and signal timing in order to meet the objective evacuation clearance time. The online simulator module uses the VISSIM microsimulation software.

In Xie et al. (2009) simulation-based and optimization-based evacuation planning methods are discussed. "Past experiences show that simulation was only used for evaluating and assessing, but not for generating evacuation planning scenarios. ... An optimization-based evacuation planning model is typically of a network flow or design functional form and can be directly used to search for optimal evacuation solutions." This paper supports the earlier claim that optimization models can produce options for an emergency evacuation plan and simulation can be used to test the effectiveness of these plans. The primary contribution of the paper was an optimization model that minimized total evacuation time while developing optimal contraflow and crossing elimination strategies. The optimization model also included an user-optimal LP that is based on the CTM. Due to the two optimization problems, the overall model is called a bi-level optimization model. At the conclusion of the

optimization, the authors used VISTA, a mesoscopic traffic simulation package, to evaluate the solution.

Two recent papers demonstrate that evacuation staging strategies can effectively reduce total evacuation and network clearance time through the use of traffic simulation. Mitchell and Radwan (2006) demonstrates the utility in using staging to reduce Clearance Time Estimates (CTEs) over simultaneous evacuation. The paper used INTEGRATION to test several different staging strategies for a generic grid network. At low traffic demands, simultaneous evacuation yielded the optimal CTE. However, at higher demands, a few of the staging strategies produced CTEs lower than from simultaneous evacuation. The same result was shown in Chen and Zhan (2004, 2008); the paper tested an enumeration of the different staging sequences of four evacuation zones and compared the results to simultaneous evacuation. The evaluations were performed through simulation, using Paramics V4.0, for a generic grid, generic ring road, and real road (case study) network.

In Liu et al. (2006), an optimization model (LP) is presented for staged evacuation using an S-curve logit-function for demand loading at each origin. The model is based upon the SO LP introduced by Ziliaskopoulos (2000) which incorporates the CTM (Daganzo, 1994, 1995). The primary contribution of the model is the addition of a binary decision variable that indicates whether or not an evacuation order is issued to an evacuation zone. It also includes model parameters that specify the latest time to initiate an evacuation order and the latest time to clear each evacuation zone. In regard to the demand loading function used in this paper, a good review of demand loading functions is provided in Yazici and Ozbay (2008). This paper reviews three different loading curves to load demand onto a network and then uses the SO LP developed in Ziliaskopoulos (2000) to route traffic through the network.

In Chiu et al. (2007), an optimization model (LP), called the Joint Evacuation Destination-Route-Flow-Departure (JEDRFD), is presented for a no-notice evacuation. A distinction is made between no-notice and short-notice evacuations; for no-notice evacuations, all demand is loaded immediately into each origin within the area to be evacuated whereas for short-notice evacuations, dynamic loading patterns can be used such as S-shaped logistic curves. The model also uses the SO LP introduced in Ziliaskopoulos (2000). The primary contribution of the paper is a procedure to transform a given evacuation network to a single-destination network that is efficiently optimized by the JEDRFD model; additionally, an origin-destination table for demand is not required as an input but just origin demands based upon population estimates. Additional papers are reviewed concerning this topic in

Chapter 5.

## 2.5 Supply and Demand Management within the Context of an Evacuation

In Bish and Sherali (2011), a network flow optimization model is developed that incorporates aggregated demand management strategies to improve existing evacuation plans based in simultaneous evacuation. Additionally, the authors state there may be some advantage to modeling flows both in the free-flow and congested traffic regimes when developing evacuation strategies and plans.

This section will review various tools for managing an evacuation that can be implemented within an evacuation strategy. For each tool, we will discuss both the modeling challenges and the issues involved with implementation within an evacuation plan. The modeling challenges will involve the ease or difficulty in modeling each tool in regard to complexity and computational performance. In regard to implementation, the true value in developing a network flow optimization model is that it can produce a set of decisions that can be directly integrated into an evacuation plan. However, these decisions should be ones that are viable and can be successfully implemented within a plan.

By the use of evacuation management tools, an optimization model can produce an optimal evacuation strategy to assist evacuation planners in how to best manage supply and demand during an emergency. There are many strategies that can improve upon what is normally performed today during an evacuation. In regards to improving an evacuation plan, this would be in terms to agreed upon performance measures such as those listed in Table 2.1.

There are different evacuation management tools that can be used within an optimization model framework to develop evacuation plans and procedures. Table 2.2 summarizes some of the demand and supply management tools that can be considered.

In order to illustrate the benefits of using demand management, consider two examples. Example 1 is a typical hurricane evacuation strategy. It uses simultaneous evacuation where all regions are notified at once to evacuate and no form of staging is used. In addition, although evacuation routes are identified and labeled in advance, the exact routing is unde-

Table 2.1: Evacuation Performance Measures

- minimize total system travel time (TTT)
- minimize network clearance time (NCT)
- minimize evacuation zone (EZ) clearance time
- minimize total delay at each origin (or at specific origins or groups of origins)
- time the evacuation order is issued for each EZ
- minimize evacuation time span (EZ clearance time - evac order time)
- minimize average travel time within an EZ: total time divided by total demand going through zone
- minimize average waiting time at origins: total time spent waiting at origins divided by total demand generated from EZ
- minimize evacuee travel distance and/or travel time

Table 2.2: Evacuation Management Tools

| **Category** | **Type** | **Management Tool** | |
|---|---|---|---|
| Demand Management | Staging | Household-level Staging | |
| | | Group-level Staging | Geographic |
| | | | Vehicle-type |
| | | | Dwelling-type |
| | | Simultaneous Evacuation | |
| | Routing | Household-level Routing | |
| | | Group-level Routing | |
| | | Undefined Routing | |
| Supply Management | Contraflow | Cell-based Contraflow | |
| | | Segmented Contraflow | |
| | Road Construction | | |

fined. Contraflow is often included to add more capacity and meet the evacuation demand but requires a large amount of coordination and control measures in order to be successful. This implementation of this strategy would be overall fairly simple - it would only require issuing the evacuation order; contraflow would require careful simulation in advance and traffic control measures during the evacuation which would complicate implementation.

Example 2 is strategy that uses group-level staging and routing where different geographical regions are issued evacuation orders at different times and are issued specific routing instructions either ahead of time or while enroute during the evacuation. This strategy has the potential to reduce the amount of congestion typically encountered during evacuations and perform better than the strategy in Example 1 (with respect to the chosen performance measures). However, the modeling challenges associated with this strategy are greater than those from Example 1; in this strategy the model would determine the optimal evacuation order issue times and routing (unless *a priori* routing is used). Implementation of this strategy would also certainly be more challenging since it would involve communicating separate evacuation orders and routes to distinct groups of evacuees. Nevertheless, these modeling and implementation challenges may lead to superior evacuation performance in regards to congestion, evacuation times, and lives saved. In the following paragraphs, we will discuss the demand management tools in more detail.

### 2.5.1 Demand Management: Staging

The principal consideration when determining how to use staging in regard to an evacuation is the size of the control unit to receive instructions. One extreme is to issue a separate evacuation orders and departure times to each household; the other extreme is to issue only one evacuation order to an entire region. The following paragraphs will review the range of considerations.

**Simultaneous Evacuation (SE):** When staging is not used, it is referred to as simultaneous evacuation. It is easy to implement since it only involves issuing a single evacuation order. The performance of various evacuation strategies are normally compared to simultaneous evacuation to measure improvement.

**Group-level Staging (GS):** This form of staging groups evacuees by either geographic location, vehicle-type, and/or by dwelling-type. Evacuating neighborhoods or evacuation zones in stages is an example of geographic group-level staging. In this research, it is assumed that the delineation of evacuation zones is predetermined; however, some recent research has been dedicated to determining an optimal size of an evacuation zone (Church and Cova, 2000). Evacuating heavy transport trucks, recreational vehicles, and busses separately from passenger cars is an example of vehicle-type group-level staging, as recommended in Litman (2006). Evacuating all residents who live in mobile homes or trailers in advance to other residents would be an example of dwelling-type group-level staging. In group-level staging, the evacuation order is issued to each group separately in order to clear the hazard area before a forecasted impact. Demand loading curves are used to model the departure of evacuees from each origin within the evacuation network. The most common loading curves used are the behavioral response curves as described in Yazici and Ozbay (2008) in the form of a logistic curve (see Equation (2.3)) where $P(t)$ is the cumulative percentage of evacuees, $\alpha_r$ is the response parameter for origin $r$, and $H_r$ is the half-loading time for origin $r$.

$$P(t) = 1/\{1 + e^{-\alpha_r(t-H_r)}\} \tag{2.3}$$

Group-level staging is difficult to implement since separate evacuation orders must be given to each group, however, it can lead to improved performance over simultaneous evacuation. Geographic group-level staging was modeled in Liu et al. (2006) and Bish and Sherali (2011).

**Household-level Staging (HS):** Household-level staging is the most advanced use of staging. This determines exact departure times for each household within a region to be evacuated. Within a modeling context, this level of staging assumes all evacuees are ready and determines the optimal time for each household to departure their respective origins. This level of staging would be very difficult to implement (currently) but is easier to model than group-level staging. Once communication infrastructure and traffic control measures mature over the next few decades, this method of staging will drastically improve demand management and evacuations. Household-level staging and routing was modeled in Chiu et al. (2007), where it was used for a no-notice evacuation, and in Bish et al. (2011b).

### 2.5.2 Demand Management: Routing

In addition to staging, routing strategies consider various different levels of control for directing traffic through a transportation network to safety. There are two main routing tools within demand management: Household-level Routing and Group-level Routing. Another situation involves undefined routing where evacuation routes are specified but not directly assigned to either groups or households.

**Group-level Routing (GR):** This type routing develops routes for specific groups and/or vehicle-types from an origin to a specific destination. These routes can be developed *a priori* or can be constructed from the results of the optimization model - this will be referred to as dynamic routing in this research. A priori routing would only consider travel time and/or distance between specific origin and destinations. Dynamic routing would additionally consider traffic conditions during the evacuation and re-route when necessary. This concept is known as dynamic traffic assignment (DTA), as discussed in Merchant and Nemhauser (1978b,a); Carey (1987); Ziliaskopoulos (2000). A binary decision variable was used to control group-level routing in Bish and Sherali (2011).

**Household-level Routing (HR):** This type of routing develops specific evacuation routes for each individual household. As in group-level routing, *a priori* or dynamic routing could be used. Both types of household-level routing are currently very difficult to implement but will become easier to implement as more intelligent transportation systems (ITS) are integrated into our national transportation system. Most evacuation models that are based upon the SO LP introduced in Ziliaskopoulos (2000) use DTA for household-level routing. Another assumption that supports the use of household-level routing in modeling is that during an evacuation, evacuees are most concerned with getting out of the danger area so will travel to where they perceive less congestion.

### 2.5.3 Supply Management

The supply management tools are normally more costly than demand management tools and require more time to implement. They are contraflow and road construction.

**Contraflow:** The strategy involves reversing travel lanes in order to maximize the capacity flowing out of an evacuation zone (Wolshon, 2001).

The use of contra-flow can be considered on a cell basis by an optimization model - this is referred to as Cell-based Contraflow (CF-C). This would be nearly impossible to implement as a result directly from an optimization model; the results would have to be reviewed and modified by a someone familiar with the network structure before implementation. The use of contra-flow is vary contentious in many states and involves many layers of control mechanisms to allow for the safe passage of evacuees. Example: An optimization model for cell-based contraflow was introduced in Tuydes and Ziliaskopoulos (2004).

Another contraflow strategy is Segmented Contraflow (CF-S). For this strategy, network planners determine roadway segments where contraflow could be safely implemented. The optimization model then determines an optimal policy for where and when to implement contraflow. Although this contraflow strategy is easier to implement than CF-C, contraflow, in general, is very difficult to implement. The risks involved in the use of contraflow must be weighed against the expected benefits.

**Road Construction:** This strategy is constrained by budget, geography (space), time, and manpower. Transportation planners determine where to increase network capacity in order to meet growing demand. However, routine traffic flows normally drive these plans and not possible evacuation demands.

The research presented in Chapters 3, 4, and 5 are inspired by the works cited in the previous paragraphs and will build upon them. Additionally, each chapter will include a focused literature review for the specific research area of the manuscript.

# Chapter 3

# Modeling the Capacity Drop Phenomenon at Freeway Bottlenecks using the INTEGRATION Software

**Abstract:** Empirical studies have demonstrated that the discharge flow rate at a bottleneck is reduced following the onset of congestion. These flow reductions, also known as capacity drops, are typically measured by comparing the queue discharge flow rate to the maximum pre-queue discharge flow rate. This research demonstrates, through the use of the INTEGRATION microscopic traffic simulation software, that these empirically observed capacity drops can be simulated without enforcing a discontinuity in the steady-state car-following or fundamental diagram. Instead, these capacity drops may be captured by constraining vehicle acceleration levels, which in turn produces the desired macroscopic behavior. The study demonstrates that the INTEGRATION software produces capacity drops at the same level of magnitude as empirically observed. The study then uses the INTEGRATION software to demonstrate the empirically observed stochastic capacity and demonstrates how it is impacted by the level of acceleration that drivers are willing to exert, the lane changing behavior, and the percentage of heavy vehicles in the traffic stream.

## 3.1 Introduction

The fact that traffic congestion upstream of a freeway bottleneck reduces the discharge flow from the bottleneck has been well documented in several empirical studies. The problem in many of the empirical studies, however, is that it is very difficult to determine what exactly caused the congestion and, in turn, what caused the capacity drop and what factors affect

the magnitude of this drop.

The capacity drop phenomenon which refers to the original two capacity phenomenon was first introduced in Edie (1961). This phenomenon requires a discontinuity within the fundamental diagram. This paper will investigate the reductions in queue discharge irrespective of whether it is a result of a discontinuity in the fundamental diagram or discharging from a different location along the fundamental diagram. In order to quantify the capacity drop, the downstream flow during congestion has to be compared to a baseline flow condition prior to the formation of congestion. This is a difficult task for empirical studies and can be highly dependent on time-of-day, year, weather conditions, traffic configuration, and other factors.

This paper will validate and attempt to explore the sources of the capacity drop through the use of a microscopic traffic simulator, INTEGRATION. By using the simulation software, the composition of the traffic stream and the behavior of the traffic in terms of vehicle dynamics can be controlled. Through this level of control, the capacity drop can be quantified more accurately and attributed to various sources. The baseline flow conditions can equally be controlled and various situations can be modeled within a simulation environment. In addition, the software can be used to characterize the stochastic nature of segment capacity and develop curves that estimate the probability of breakdown as a function of the arrival flow.

The paper will briefly review the current applicable literature and empirical studies, then present two common freeway bottleneck examples, discuss the results, and present some conclusions related to the results of the simulations.

## 3.2 Background

The concept of two capacities or the "capacity drop" was first introduced in Edie (1961) where congested and uncongested flows in the Lincoln Tunnel in New York were analyzed. This empirical study provided the first support to the idea of flow reductions downstream of a congested bottleneck.

In Hall and Agyemang-Duah (1991), they concluded that once a queue forms upstream of a bottleneck, a capacity drop appears as a consequence of the way drivers acceler-

ate from the queue. Their conclusion was based on data collected from the Queen Elizabeth Way (QEW) outside of Toronto and the results showed a capacity drop of approximately 5 to 6% (Hall and Agyemang-Duah, 1991). In addition to the idea of capacity drop, the paper argued that the capacity drop should be measured within the bottleneck - not upstream or downstream of the bottleneck; otherwise only a reduction in flow would be measured. The authors demonstrated that measuring at this location yields a one branch occupancy-flow relationship. This branch consists of pre-queue flows and queue discharge flows. The vertical difference between these two curves presents the capacity drop. In this study, a truck percentage of 6% was reported in the traffic stream (Hall and Agyemang-Duah, 1991).

Cassidy and Bertini (1999) studied the capacity drop phenomenon along the same QEW section studied in Hall and Agyemang-Duah (1991). A capacity drop of 8 to 9% was found along the QEW, as compared to the 5 to 6% drop found in Hall and Agyemang-Duah (1991), by visually comparing transformed cumulative curves and cumulative occupancy against time created by data from downstream loop detectors (Cassidy and Bertini, 1999).

It was concluded in Banks (1991) that bottleneck capacities decrease when flow breaks down due to congestion, however this applies to the capacities of individual lanes. Four different bottlenecks at four sites in San Diego were investigated. It was shown that capacity decreased after breakdown by about 10% in the left lane. However, the capacity, when averaged across all lanes, decreased by only 3%.

In Chung et al. (2007), three different bottlenecks on different freeway segments in San Diego and San Francisco were investigated to examine the relationship between vehicular density and the drop in queue discharge flow rates. The three bottlenecks investigated were an on-ramp merge, a reduction in the number of travel lanes (or lane drop), and a horizontal curve on the freeway. The study concluded that a capacity drop, on average, of 10% resulted from the formation of a queue in the freeway shoulder lane upstream of the merge. A critical density of about 55 vehicles/km/lane was observed corresponding to the capacity drop. This merge was previously studied in a ramp metering study (Cassidy and Rudjanakanoknad, 2005) where loop detector occupancies in the range 22-27% were used as a threshold for detecting the capacity drop. A capacity drop of 5% was reported for the lane drop bottleneck, which also occurred when a critical density of approximately 46 vehicles/km was reached (Chung et al., 2007). In contrast to these results, Bertini and Leal (2005) reported a 10-12% capacity drop in two empirical studies of freeway lane drops in London and Minneapolis. Nevertheless, the conclusion that a specific value for critical

density causes the capacity drop requires further investigation.

In Kerner and Klenov (2006) and Kerner (2008), the flow breakdown phenomenon was found to be probabilistic, and not deterministic as concluded in Chung et al. (2007). Persaud et al. (1998) also found that the occurrence of a breakdown is probabilistic. To explore the probability of breakdown, traffic data were collected at three different sites in Toronto using surveillance cameras and automatic collection of speed and flow data from loop detectors. A capacity drop was found ranging from 11 - 17% at two of the sites, while at the third site the capacity drop was 26% (Persaud et al., 1998). Furthermore, the probability of breakdown at various traffic flow levels was investigated. In general for the three sites, the probability of breakdown increased by increasing the flow rates. It was found that by maintaining the pre-queue flows at the same level as those that occur after a queue forms, the probability of breakdown was almost negligible. By increasing these flows by 20% above the queue discharge flow, the probability of breakdown was only 10%. For higher flows, the probability of breakdown increased dramatically. Hence, it was suggested not to operate at pre-queue rates exceeding the queue discharge rate by more than 20% (Persaud et al., 1998).

In Treiber et al. (2006), conditions at a freeway merge bottleneck were simulated using three different car-following models modified by the variance-driven time (VDT) headway mechanism. The paper attempted to investigate the widely scattered behavior of flow-density plots within the congested regime. In the merge bottleneck simulation, both cars and trucks were used with two different free-flow speeds (126 km/h and 90 km/h respectively) with trucks making up 20% of the traffic stream composition. However, both vehicle types were assumed to have a vehicle length of 5m. The merge was constructed as a single lane on-ramp merging with a single lane freeway. Vehicle demand was supplied at a constant flow of 400 veh/hr at the on-ramp and at a variable rate for the freeway; freeway demand increased linearly from 300 veh/hr to 3000 veh/hr at t=2400s and then decreased linearly to 300 veh/hr at t=4800s. Although this paper did not report capacity drop percentages, the flow-density plots indicated a capacity drop did occur when the VDT approach was applied. However, only conditions upstream of the bottleneck were considered after the onset of congestion. Additionally, the criteria used to determine the onset of congestion was not made clear. The authors concluded that the velocity variance of the merging traffic and lane changing were significant contributors to decrease performance at freeway interchanges.

In Laval and Daganzo (2006), the lane drops that were empirically studied in Bertini and Leal (2005) were simulated using a kinematic wave traffic flow model that accounts

for lane changing and the results were consistent; a capacity drop of 9.3% as compared to 10-12% in the empirical studies. The primary contribution of this paper was to demonstrate the effect of lane changing and vehicle acceleration capabilities on the capacity drop at a bottleneck.

In Papageorgiou et al. (2008), capacity drops were concluded to be a result of vehicles accelerating out of (or away from) the bottleneck. The paper argued that since the capabilities of the various vehicles in a traffic stream are different, large gaps are created in the traffic stream that create flow reductions downstream of a bottleneck after the onset of congestion (Bertini and Leal, 2005).

As summarized in Table 3.1, the empirically observed capacity drops associated with bottlenecks are typically larger for merge bottlenecks compared to lane drop bottlenecks. These capacity drops range from 3 to 17% for on-ramp bottlenecks and range from 5 to 12% for lane drop bottlenecks.

Table 3.1: Capacity Drops Reported in Various Empirical Studies

| **Reference** | **Bottleneck Example** | **Location** | **Capacity Drop** |
|---|---|---|---|
| Hall and Agyemang-Duah (1991) | On-ramp Merge | QEW - Mississauga | 5-6% |
| Banks (1991) | On-ramp Merge | I-8 - San Diego (site 1) | 10.5% (left lane); 3% (across all lanes) |
| Cassidy and Bertini (1999) | On-ramp Merge | QEW - Mississauga | 8-9% |
| Cassidy and Rudjanakanoknad (2005); Chung et al. (2007) | On-ramp Merge | I-805 – San Diego | 10% |
| Chung et al. (2007) | Lane Drop | SR-24 – San Francisco | 5-8% |
| Persaud et al. (1998) | On-ramp Merge | Highway 401 & Gardiner Expwy - Toronto | 11-17% |
| Bertini and Leal (2005) | Lane Drop | M4 motorway – London | 10% |
| | Lane Drop | I-494 – Minneapolis | 12% |

Given that one explanation for the reduction in discharge flow after the onset of congestion relates to driver acceleration levels, a review of the literature in this area was also conducted. A study by Snare (2002), found that drivers normally accelerate at 60% of the maximum acceleration level of their vehicle; “Driver factors were observed that fit a

normal distribution with a mean of 0.6 and a standard deviation of 0.08" (Snare, 2002) . This confirmed a previous guideline published in 1954 by AASHTO (1954). This conclusion is important since capacity drops are influenced by how vehicles accelerate downstream of the bottleneck and thus it is important to capture typical driver behavior.

## 3.3 INTEGRATION Software Overview

This section provides a brief overview of the INTEGRATION software logic that relate to the research presented in this paper.

The INTEGRATION software is a microscopic traffic assignment and simulation software that was developed over the past three decades (Van Aerde and Rakha, 2007a,b; Van Aerde and Yagar, 1988a,b). It was conceived as an integrated simulation and traffic assignment model and performs traffic simulations by tracking the movement of individual vehicles every 1/10th of a second. This allows detailed analyses of lane-changing movements and shock wave propagations. It also permits considerable flexibility in representing spatial and temporal variations in traffic conditions. In addition to estimating stops and delays (Dion et al., 2004; Rakha et al., 2001, 2000b), the model can also estimate the fuel consumed by individual vehicles, as well as the emissions (Rakha et al., 2000a, 2004; Ahn et al., 2001, 2004). Finally, the model also estimates the expected number of vehicle crashes using a time-based crash prediction model (Avgoustis et al., 2004).

The INTEGRATION model updates vehicle speeds every deci-second using a user-specified steady-state speed-spacing relationship, the speed differential between the subject vehicle and the vehicle immediately ahead of it, and the vehicle acceleration constraints. In order to ensure realistic vehicle accelerations, the model uses the driver throttle input together with a vehicle dynamics model that estimates the maximum vehicle acceleration level. The INTEGRATION car-following model falls within the psycho-physical model formulations because it considers a driver within different regimes. Specifically, the vehicle will collide with a vehicle if its spacing from the lead vehicle is less than a safe distance that increases as the speed difference between the lead and following vehicle increases (negative speed differential). The driver typically attempts to converge to the steady-state behavior by either decelerating (collision avoidance) or accelerating. The driver is in free-flow mode once the spacing between the following and lead vehicle exceeds a threshold.

The model computes the vehicle speed as the minimum of three speeds, namely: the maximum vehicle speed based on vehicle dynamics considering the driver throttle level input,the desired speed based on the Van Aerde steady-state car-following model formulation, and the maximum vehicle speed that the vehicle can travel at ensuring that it does not collide with the vehicle ahead of it as shown in Equation (3.1).

$$u_n(t+\triangle t) = \min \left\{ \begin{array}{l} u_n(t) + 3.6\ \dfrac{F_n(t) - R_n(t)}{m} \triangle t, \\ \dfrac{-c_1 + c_3 u_f + \tilde{s}_n(t+\triangle t) - \sqrt{A}}{2c_3}, \\ \sqrt{u_{n-1}(t+\triangle t)^2 + 25920\ d_{max}\left(\tilde{s}_n(t+\triangle t) - \dfrac{1}{k_j}\right)} \end{array} \right\} \tag{3.1}$$

Notation:

| | | |
|---|---|---|
| $u_n(t)$ | : | speed of vehicle $n$ at time $t$ $(km/h)$ |
| $u_f$ | : | free-flow speed $(km/h)$ |
| $u_c$ | : | speed-at-capacity $(km/h)$ |
| $F_n(t)$ | : | vehicle tractive force $(N)$ which is computed using Equation (3.8) as will be described later |
| $R_n(t)$ | : | total resistance force $(N)$ |
| $m$ | : | vehicle mass $(kg)$ |
| $\triangle t$ | : | length of time interval |
| $c_1$ | : | fixed distance headway constant $(km)$ |
| $c_2$ | : | first variable headway constant $(km^2/h)$ |
| $c_3$ | : | second variable distance headway constant $(h)$ |
| $A$ | : | $A = [c_1 - c_3 u_f - \tilde{s}_n(t+\triangle t)]^2 - 4c_3[\tilde{s}_n(t+\triangle t)u_f - c_1 u_f - c_2]$ |
| $s_n(t)$ | : | vehicle spacing $(km)$ at time $t$ |
| $\tilde{s}_n(t+\triangle t)$ | : | predicted vehicle spacing at time $t+\triangle t$ considering that vehicle $n$ continues at its current speed where $\tilde{s}_n(t+\triangle t) = s_n(t) + [u_{n-1}(t) - u_n(t)]\triangle t + 0.5 a_{n-1}(t)\triangle t^2$ where $a_n(t)$ is the acceleration of vehicle $n$ at time $t$ |
| $d_{max}$ | : | is the maximum acceptable deceleration level the driver is willing to exert $(m/s^2)$ |

### 3.3.1 Steady-State Modeling

The Van Aerde nonlinear single-regime functional form that was proposed in Van Aerde (1995) and in Van Aerde and Rakha (1995), is formulated in Equation (3.2).

$$\tilde{s}_n(t+\triangle t) = c_1 + c_3 u_n(t+\triangle t) + \frac{c_2}{u_f - u_n(t+\triangle t)} \tag{3.2}$$

where $c_1$, $c_2$, and $c_3$ were defined earlier. Demarchi (2002) and Rakha et al. (2009) demonstrated that by considering three boundary conditions the model constants can be computed as shown in Equation (3.3).

$$c_1 = \frac{u_f}{k_j u_c^2}(2u_c - u_f); \quad c_2 = \frac{u_f}{k_j u_c^2}(u_f - u_c)^2; \quad c_3 = \left(\frac{1}{q_c} - \frac{u_f}{k_j u_c^2}\right) \tag{3.3}$$

In order to ensure that the speed estimates are realistic the square root term should be positive. This is achieved if the model parameters satisfy the condition of Equation (3.4) as was demonstrated in Rakha et al. (2009).

$$q_c \leq \frac{k_j u_f u_c}{2u_f - u_c} \tag{3.4}$$

As was demonstrated in Rakha and Crowther (2002) this continuous functional form amalgamates the Greenshields and Pipes car-following models. It should also be noted that given that the functional form is continuous the capacity is identical regardless of whether the vehicle approaches capacity from uncongested (upper arm) or congested (lower arm) conditions.

Typically the vehicle speed or the acceleration level is considered as the control variable. Consequently, solving Equation (3.2) for the vehicle speed results in Equation (3.5).

$$u_n(t+\triangle t) = \frac{-c_1 + c_3 u_f + \tilde{s}_n(t+\triangle t) - \sqrt{A}}{2c_3} \tag{3.5}$$

### 3.3.2 Collision Avoidance Modeling

In the case that the following vehicle is traveling at a higher speed than the lead vehicle (non-steady state conditions) the vehicle spacing should be sufficient to allow the following vehicle (vehicle $n$) to avoid a collision with the lead vehicle (vehicle $n-1$) if it were to decelerate at a feasible deceleration level. This deceleration level is assumed to be equal to $\mu f_b \eta_b g$, where $\mu$ is the coefficient of roadway friction, $f_b$ is the driver brake pedal input [0,1], $\eta_b$ is the brake efficiency [0,1], and $g$ is the gravitational acceleration (9.8067 $m/s^2$). The resulting minimum vehicle spacing can be computed using Equation (3.6).

$$\tilde{s}_n(t+\triangle t) \geq \frac{1}{k_j} + \frac{u_n(t+\triangle t)^2 - u_{n-1}(t+\triangle t)^2}{25920\ \mu f_b \eta_b g} \tag{3.6}$$

Again solving Equation (3.6) for the vehicle speed results in Equation (3.7).

$$u_n(t+\triangle t) = \sqrt{u_{n-1}(t+\triangle t)^2 + 25920\ \mu f_b \eta_b g\left(\tilde{s}_n(t+\triangle t) - \frac{1}{k_j}\right)} \tag{3.7}$$

### 3.3.3 Vehicle Acceleration Modeling

Vehicle acceleration is governed by vehicle dynamics. Vehicle dynamics models compute the maximum vehicle acceleration levels from the resultant forces acting on a vehicle (mainly vehicle tractive forces, which are a function of the driver throttle input and resistance forces).

The vehicle tractive effort is computed in Equation (3.8). Rakha and Lucic (2002) introduced the $\beta$ factor into Equation (3.8), in order to account for the gear shift impacts at low traveling speeds when trucks are accelerating. This factor is set to 1.0 for light duty vehicles. The $f_p$ factor models the driver throttle input level and ranges from 0.0 to 1.0. The vehicle resistance force is calculated as the sum of the aerodynamic, rolling, and grade resistance forces acting on the vehicle Rakha and Lucic (2002), as demonstrated in Equation (3.9).

$$F_n(t) = \min\left(3600 f_p\ \beta\ \eta_d\ \frac{P_n}{u_n(t)}, \tilde{m}_n\ g\ \mu\right) \tag{3.8}$$

$$R_n(t) = \frac{\rho}{25.92} C_d C_h A_f u_n(t)^2 + m_n g \frac{c_{r0}}{1000}\left(c_{r1} u_n(t) + c_{r2}\right) + mgG(t) \tag{3.9}$$

Notation:

| | | |
|---|---|---|
| $f_p$ | : | driver throttle input [0,1](unitless) (field studies have shown that it is typically 0.60) |
| $\beta$ | : | gear reduction factor (unitless) |
| $\eta_d$ | : | driveline efficiency (unitless) |
| $P$ | : | vehicle power ($kW$) |
| $\tilde{m}_n$ | : | mass of the vehicle on the tractive axle ($kg$) |
| $g$ | : | gravitational acceleration (9.8067 $m/s^2$) |
| $\mu$ | : | coefficient of road adhesion or the coefficient of friction (unitless) |
| $\rho$ | : | air density at sea level and a temperature of 15 °C (1.2256 $kg/m^3$) |
| $C_d$ | : | vehicle drag coefficient (unitless), typically 0.30 |
| $C_h$ | : | altitude correction factor (unitless) |
| $A_f$ | : | vehicle frontal area ($m^2$) |
| $c_{r0}$ | : | rolling resistance constant (unitless) |
| $c_{r1}$ | : | rolling resistance constant ($h/km$) |
| $c_{r2}$ | : | rolling resistance constant (unitless) |
| $m_n$ | : | total vehicle mass ($kg$) |
| $G(t)$ | : | roadway grade at instant $t$ (unitless) |

The vehicle acceleration is calculated as a ratio of the difference between the tractive forces and resistance forces and the vehicle mass (i.e., $a = (F - R)/m$). The vehicle speed at $t + \triangle t$ is then computed using Equation (3.10).

$$u_n(t + \triangle t) = u_n(t) + 3.6 \frac{F_n(t) - R_n(t)}{m} \triangle t \tag{3.10}$$

The computation of vehicle accelerations within the simulation environment is signif-

icant and will prove to be an important factor in determining capacity drops at the onset of congestion later in the paper.

The INTEGRATION model was chosen for this study due to its unique ability to model vehicle dynamics, specifically vehicle acceleration. In addition, the model has not only been validated against standard traffic flow theory (Dion et al., 2004; Rakha and Crowther, 2003, 2002; Rakha et al., 2001), but also has been utilized for the evaluation of real-life applications (Rakha et al., 2005, 1998; Rakha, 1991). The INTEGRATION lane-changing logic was validated against field data in an earlier publication (Rakha and Zhang, 2004). Furthermore, Rakha and Zhang (2006) demonstrated the validity of the INTEGRATION software for estimating the capacity of weaving sections by comparing to field observed weaving section capacities.

## 3.4 Experimental Design

In order to study and quantify the impact of various factors on the empirically observed capacity drop phenomenon, two simple examples are modeled. Before describing the specific examples, the following section describes the general assumptions made and the scenarios considered in each example.

A 30-second sampling interval was used at virtual loop detectors within our simulations. This was the same sampling interval used in Hall and Agyemang-Duah (1991) where the authors compiled average flow and occupancy measurements at loop detectors at stations upstream and downstream of the bottleneck during an empirical study.

In determining the onset of congestion, occupancy measurements at the upstream station were monitored. Congestion was considered to have formed once the occupancy exceeded a specific threshold for a 3-minute time period. A 3-minute time period was considered in order to be consistent with earlier empirical studies (Hall and Agyemang-Duah, 1991). Equation (3.11) gives the relationship between the critical density ($k_c$), average vehicle length ($\bar{L}_v$), loop detector length ($L_D$), and detector occupancy ($O_D$).

$$k_c = \frac{1000}{\bar{L}_v + L_D} O_D \tag{3.11}$$

$$k_c = \frac{q_c}{u_c} = \frac{q_c}{u_f} \tag{3.12}$$

$$O_D^* = \frac{q_c}{u_f} \frac{\bar{L}_v + L_D}{1000} \tag{3.13}$$

Equation (3.12) calculates $k_c$ by dividing the saturation flow rate (or capacity), $q_c$, by the speed-at-capacity. The speed-at-capacity, $u_c$, was set equal to the free-flow speed, $u_f$, to model the commonly known triangular flow-density relationship (fundamental diagram) or the Pipes car-following model. By substituting Equation (3.12) into Equation (3.11) and solving for the critical detector occupancy $O_D^*$, a threshold for the start of congestion can be solved for, Equation (3.13). Using the average vehicle length of 5 m, a detection zone length of 5 m, a capacity of 2400 veh/h/lane, and free-flow speed of 108 km/h, the threshold for detector occupancy, $O_D$, corresponds to 23%; this threshold is consistent with Cassidy and Rudjanakanoknad (2005). The onset of congestion corresponded to the instant at which the upstream detector occupancy exceeded 23% for at least a 3-minute period. The corresponding downstream flows were temporally offset by the travel time from the upstream to the downstream detector at free-flow speed. Reductions in flow were measured relative to the base downstream flow, as will be described later. See Figure 3.1 for an illustration of how the capacity drops were calculated.

Similar to the work performed in Treiber et al. (2006), two vehicle types were used in the simulation runs; a light duty composite vehicle to represent passenger cars and a medium truck to represent heavy-duty vehicles. The light duty vehicle has a mass of 1326 kg, is 4.8 m in length, and 109 kW of power. The medium truck has a mass of 31,751 kg, is 16 m in length, and 261 kW of power. However, in contrast with Treiber et al. (2006), the different vehicle lengths and acceleration capabilities were used in the simulation; additionally, the free-flow speeds were set as a function of the roadway segment and not the vehicle type.

A vehicle speed coefficient of variation of 5% (speed standard deviation divided by mean speed) was coded within the INTEGRATION software to reflect typical differences in driver speeds for the same vehicle spacing. The impact of the speed variability factor was explored in Farzaneh and Rakha (2006) and was found to range between 5-10% based on

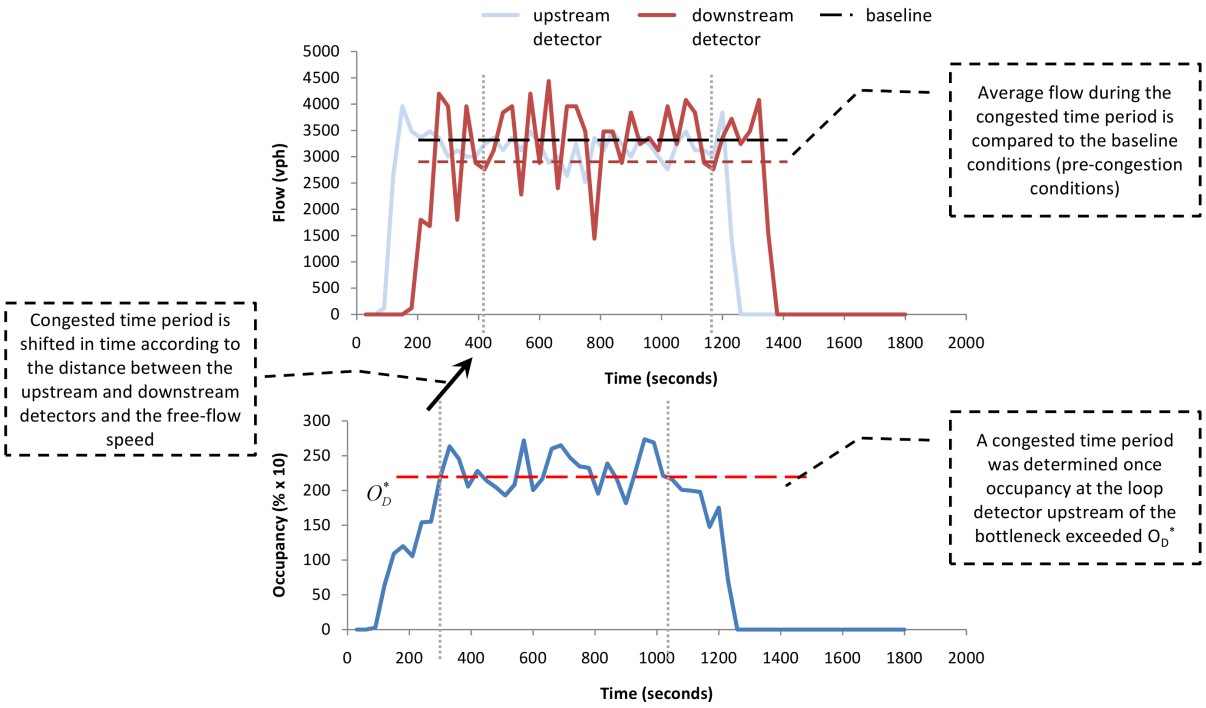



Figure 3.1: Computation of the Capacity Drop

empirical observations.

To capture the probabilistic nature of traffic behavior and the formation of congestion, 20 different random seeds were considered. This was done in order to produce statistically significant capacity drops for each demand and heavy truck percentage combination. In addition, the random instances also allow for the characterization of the probabilistic nature of traffic flow breakdown, as will be demonstrated later in the paper.

### 3.4.1 Scenarios

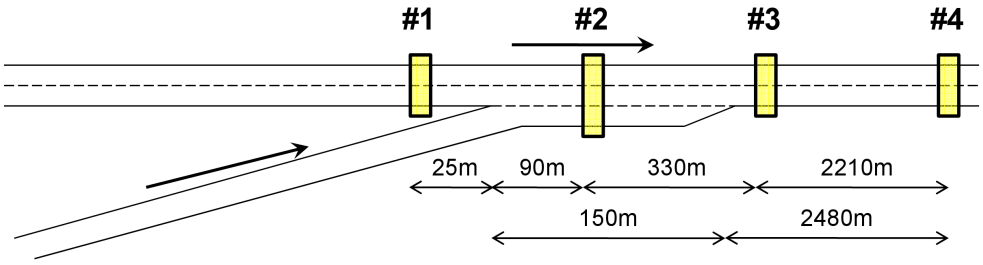
For the two freeway bottleneck examples, the following traffic stream composition scenarios were considered.

- Scenario 1 - The entire vehicle stream was composed of light duty vehicles accelerating at full throttle (100% of their maximum capability) [i.e., $f_p = 1.0$ in Equation (3.8)].
- Scenario 2 - The entire vehicle stream was composed of light duty vehicles that accel-

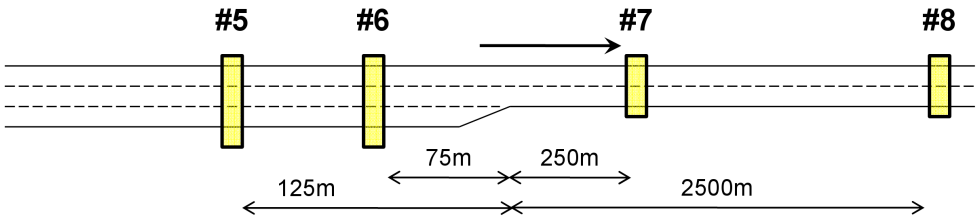
erate at typical acceleration levels (60% of their maximum capability) [i.e., $f_p = 0.6$ in Equation (3.8)].

- Scenarios 3 through 6 - The traffic stream was composed of 6, 10, 15, and 20% heavy duty vehicles and the remainder traffic stream was light duty vehicles. All drivers were assumed to accelerate at the typical throttle level of $f_p = 0.6$.

The change from $f_p = 1.0$ to $f_p = 0.6$ was inspired by Snare (2002) and previous research cited therein. The 6% heavy vehicle composition was taken from the empirical data used in Hall and Agyemang-Duah (1991); Cassidy and Bertini (1999); and the 20% heavy vehicle composition was used in Laval and Daganzo (2006); Treiber et al. (2006).




(a) Freeway On-ramp Bottleneck




(b) Freeway Lane Drop Bottleneck

Figure 3.2: Example Networks used in INTEGRATION

### 3.4.2 On-ramp Bottleneck Example

The first bottleneck example modeled is a freeway merge section, shown in Figure 3.2a. This example considers a 2-lane freeway segment with a 1-lane on-ramp merging onto the freeway. Loop detectors were used to measure traffic stream parameters upstream at Station 1, within the bottleneck at Station 2, and downstream of the bottleneck at Stations 3 and 4.

The following macroscopic parameters were used in this example. Typical values reported in the Highway Capacity Manual (HCM) for freeway sections are used, $q_c = 2400$ veh/hr/lane and $u_f = 108$ km/hr. For the on-ramp, $q_c = 1800$ veh/hr/lane and $u_f = 72$ km/hr. A jam density ($k_j$) of 150 veh/km/lane was used for both the freeway and on-ramp. The speed-at-capacity, $u_c$, was assumed to equal $u_f$ (Pipes model). It should be noted that the capacity that was coded is the base roadway capacity. The actual capacity is lower than the base capacity and depends on the level of lane-changing, percentage of trucks, and level of driver variability.

For each of the traffic stream composition scenarios, twelve to eighteen demand cases were considered. The freeway demand was fixed while the on-ramp demand was increased at increments of 100 veh/hr. Three levels of freeway demand were considered:

- HIGH - A fixed freeway demand of 3600 veh/hr was considered along with on-ramp demands ranging from 100 to 1200 veh/hr were considered. The total demand loaded onto the freeway ranged between 3700 and 4800 veh/hr (the base capacity of the freeway).
- MEDIUM - A fixed freeway demand of 3000 veh/hr was considered along with on-ramp demands ranging from 100 to 1800 vph were considered. The total demand loaded onto the freeway ranged between 3100 and 4800 veh/hr (the base capacity of the freeway).
- LOW - A fixed freeway demand of 2600 veh/hr was considered along with on-ramp demands ranging from 100 to 1800 veh/hr were considered. The total demand loaded onto the freeway ranged between 2700 and 4400 veh/hr.

The maximum on-ramp flow that did not result in the formation of congestion (occupancy of 23% for 3 minutes or longer) at the upstream detector station was used to establish the base pre-queue maximum flow. The average flow for the pre-queue scenario at the downstream detector was then used as the base reference flow for the computation of discharge flow reductions. This approach is consistent with ramp metering logic where the on-ramp flow is controlled in order to ensure that no breakdown occurs on the freeway mainline (i.e. no congestion forms upstream of the on-ramp). This approach is in contrast with Treiber et al. (2006) where the on-ramp demand was fixed and the freeway demand varied.

### 3.4.3 Lane Drop Bottleneck Example

The second example is a freeway lane drop where 3-lanes reduce to 2-lanes, as shown in Figure 3.2b. As in the on-ramp bottleneck example, loop detectors were located upstream, within, and downstream of the bottleneck; Station 5 is upstream, Station 6 is within the bottleneck, and Stations 7 and 8 are downstream of the bottleneck.

The following macroscopic parameters were used to simulate this example: base $q_c$ = 2400 veh/hour/lane, $u_f$ = 108 km/hr, and $k_j$ = 150 veh/km/lane.

As was the case with the on-ramp bottleneck example, several different demand loading profiles were tested in each of the four scenarios. The freeway flow was incrementally increased from 2800 to 4800 veh/hr (the base capacity of the 2-lane section) at increments of 100 veh/hr. Two separate baseline flow conditions were used in this example, as follows:

- Baseline 1: was determined by first computing the average downstream flow for each demand loading prior to the onset of congestion (based on the 23% occupancy threshold at the upstream station) and then taking the maximum of the set of average downstream flows collected. A separate Baseline 1 flow was calculated for each of the 20 random instances. This baseline was used in both freeway bottleneck examples.
- Baseline 2: was determined by using the average downstream flow for a 2-lane section of freeway (without a lane drop). This base flow eliminates any lane changing behavior effects on the baseline flow. This was intended to simulate traffic behavior along straight freeway sections when the traffic consolidates into 2-lanes well before the lane drop from 3-lanes to 2-lanes occurs.

## 3.5 Study Findings

This section summarizes the results of the simulations for each of the two bottleneck examples. The simulation results reveal reductions in discharge flow rates after the onset of congestion for each of the two bottleneck examples. The following sections analyze and discuss the results for each bottleneck/scenario combination.

### 3.5.1 On-ramp Bottleneck Example Results

In Scenario 1, a capacity drop was not observed given that vehicle accelerations were at maximum throttle. In Scenario 2, capacity drops were observed but were minimal given that no trucks were part of the traffic stream. This behavior demonstrates that typical acceleration levels have the potential of producing the empirically observed capacity drops. In Scenarios 3-6, significant capacity drops were observed that are consistent with various empirical study observations, which confirms that the combination of trucks and typical acceleration levels produces the observed capacity drops. These results are displayed in Tables 3.2 - 3.4 and Figures 3.3 - 3.5; three sets of results corresponding to the HIGH, MEDIUM, and LOW fixed freeway demand scenarios are presented. Note that the Figures display all of the simulation results where as the tables only show key portions (but not all) of the simulation results. The observed capacity drop reported is the average capacity drop from the 20 random instances for each on-ramp demand scenario and each traffic composition scenario. The figures only display the results from traffic composition Scenarios 3 - 6 since these scenarios use heavy vehicle percentages consistent with the empirical studies reviewed in Section 3.2.

Table 3.2: Capacity Drops for the On-ramp Bottleneck Example (HIGH freeway demand)

| **Total Demand** | **3800** | **4000** | **4200** | **4400** | **4600** | **4800** |
|---|---|---|---|---|---|---|
| On-ramp Demand | 200 | 400 | 600 | 800 | 1000 | 1200 |
| Scenario 3: 6% trucks | 5.90% | 11.56% | 14.99% | 16.89% | 17.14% | 16.93% |
| Scenario 4: 10% trucks | 4.58% | 10.73% | 15.05% | 15.57% | 16.20% | 16.84% |
| Scenario 5: 15% trucks | 5.20% | 12.69% | 15.23% | 16.47% | 16.78% | 16.04% |
| Scenario 6: 20% trucks | 5.39% | 12.68% | 17.47% | 17.50% | 18.55% | 18.36% |

In the case of the HIGH fixed freeway demand, there are significant capacity drops in the range of 15 - 18% for a total demand in excess of 4400 veh/hr across all of the traffic stream composition scenarios (Scenarios 3 - 6) (see Table 3.2 and Figure 3.3). The capacity drops increase slightly for each scenario. However, the different heavy vehicle percentages do not seem to affect the magnitude of the capacity drops significantly; i.e., the capacity drops from each scenario are all within 1-2% of each other. Consequently, while the introduction of heavy vehicles at the 6% level has a significant impact on the capacity drops, further increases in the heavy vehicle percentage does not significantly increase the capacity drops. This is due to the high-level of the freeway demand; the introduction of traffic from the on-ramp

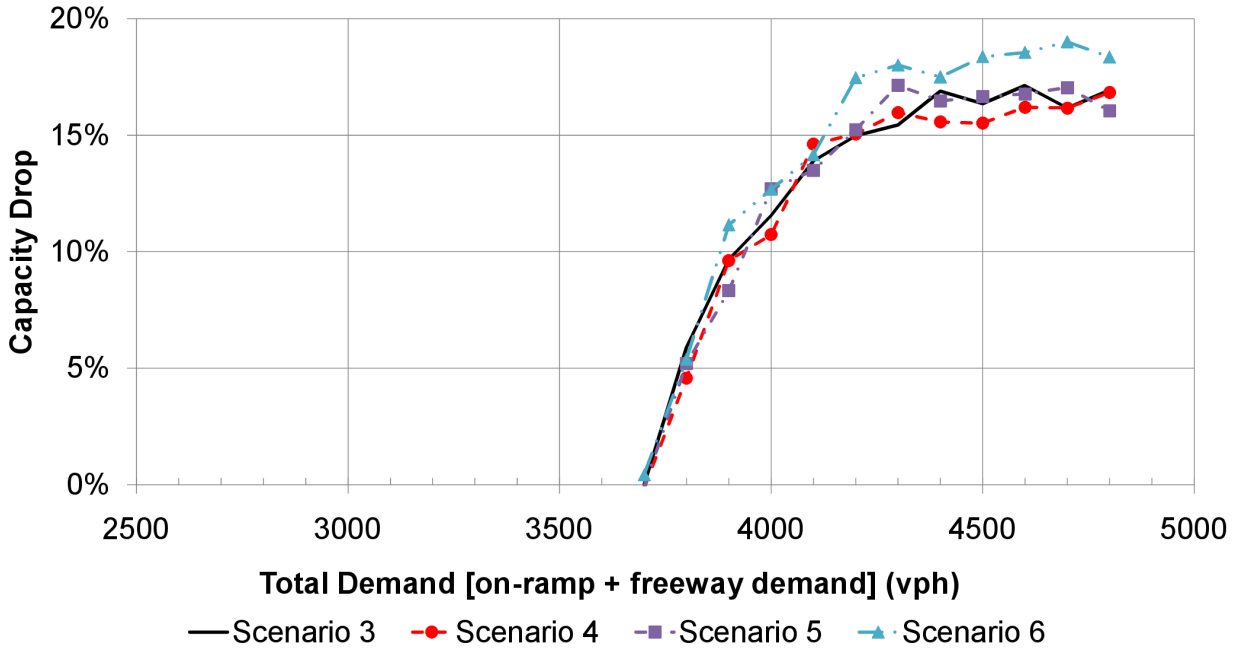



Figure 3.3: Capacity Drops vs. Total Demand for the On-ramp Bottleneck Example (HIGH freeway demand)

creates the capacity drops due to congestion created by lane-changing behavior. The impact of different heavy vehicle percentages does not seem to be as important as the lane-changing behavior in the creation of congestion and ultimately the flow reductions downstream under HIGH freeway demand conditions.

For the MEDIUM freeway demand levels, the capacity drop was found to be more distinct across the four traffic stream composition scenarios (see Table 3.3 and Figure 3.4). Specifically, the capacity drop is in the range of 6-17% for a total demand in excess of 4000 veh/hr; and varies over a larger range compared to the HIGH freeway demand scenario. Consequently, the percentage of heavy vehicles contributes more significantly to the magnitude of the capacity drop compared to the HIGH freeway demand conditions. Lane-changing behavior is not as much of a factor since there are more gaps in the freeway traffic stream for vehicles to merge. The capacity drops in Scenario 6 are in the same range for the MEDIUM and HIGH freeway demand scenarios, which indicates that traffic with large heavy vehicle percentages ($\geq 20\%$) will experience large capacity drops ($> 15\%$) regardless of the level of freeway demand.

Lastly, for the LOW freeway demand levels, the capacity drop is clearly much less

than that observed for the HIGH and MEDIUM demand levels (see Table 3.4 and Figure 3.5). For example in Scenario 3, the capacity drop is minimal and not significant. However, for Scenarios 4-6, the capacity drop is in the range of 4-10% for a total demand in excess of 3600 veh/hr which is a tighter range compared to the MEDIUM freeway demand case. These findings suggest that neither lane-changing nor heavy vehicle percentages of 6% or less are significant factors in the creation of capacity drops under LOW freeway demands. However, at higher heavy vehicle percentages, both lane-changing and heavy vehicles are significant contributors to the capacity drops.

Table 3.3: Capacity Drops for the On-ramp Bottleneck Example (MEDIUM freeway demand)

| **Total Demand** | **3400** | **3600** | **3800** | **4000** | **4200** | **4400** |
|---|---|---|---|---|---|---|
| Onramp Demand | 400 | 600 | 800 | 1000 | 1200 | 1400 |
| Scenario 3: 6% trucks | 0.98% | 4.53% | 6.63% | 6.28% | 5.73% | 5.78% |
| Scenario 4: 10% trucks | 7.22% | 9.16% | 11.06% | 11.89% | 10.81% | 10.27% |
| Scenario 5: 15% trucks | 8.42% | 11.91% | 13.90% | 13.37% | 13.06% | 12.61% |
| Scenario 6: 20% trucks | 12.59% | 15.35% | 16.21% | 17.35% | 16.22% | 15.08% |

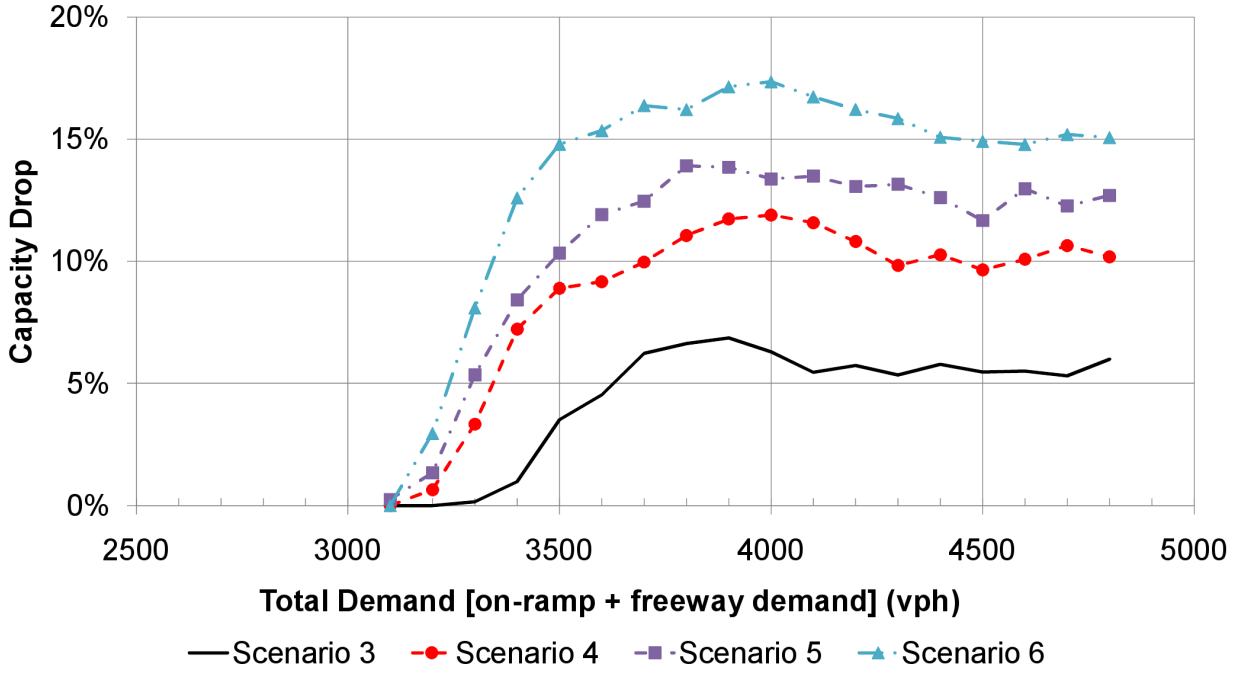



Figure 3.4: Capacity Drops vs. Total Demand for the On-ramp Bottleneck Example (MEDIUM freeway demand)

Table 3.4: Capacity Drops for the On-ramp Bottleneck Example (LOW freeway demand)

| **Total Demand** | **3200** | **3400** | **3600** | **3800** | **4000** | **4200** |
|---|---|---|---|---|---|---|
| On-ramp Demand | 600 | 800 | 1000 | 1200 | 1400 | 1600 |
| Scenario 3: 6% trucks | -0.18% | 1.42% | 1.72% | 0.64% | 0.25% | 0.40% |
| Scenario 4: 10% trucks | 3.52% | 4.58% | 4.54% | 3.78% | 2.57% | 3.56% |
| Scenario 5: 15% trucks | 7.48% | 8.96% | 7.63% | 5.81% | 6.07% | 5.79% |
| Scenario 6: 20% trucks | 6.92% | 10.20% | 7.64% | 9.06% | 9.75% | 5.32% |

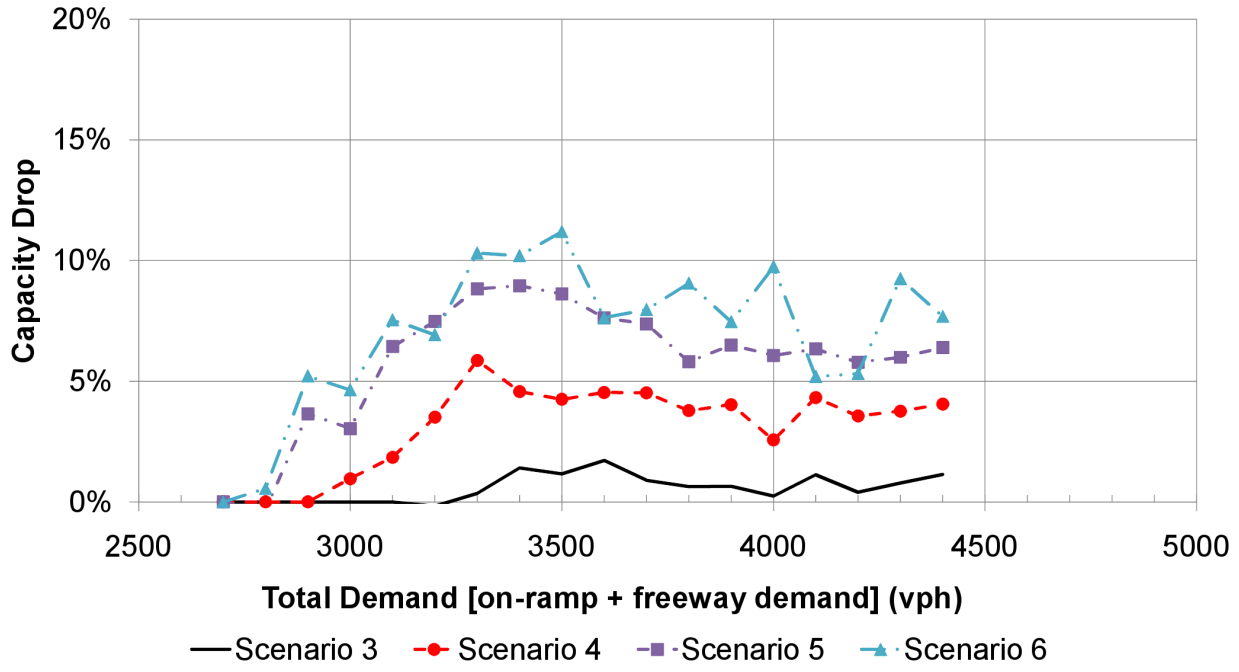
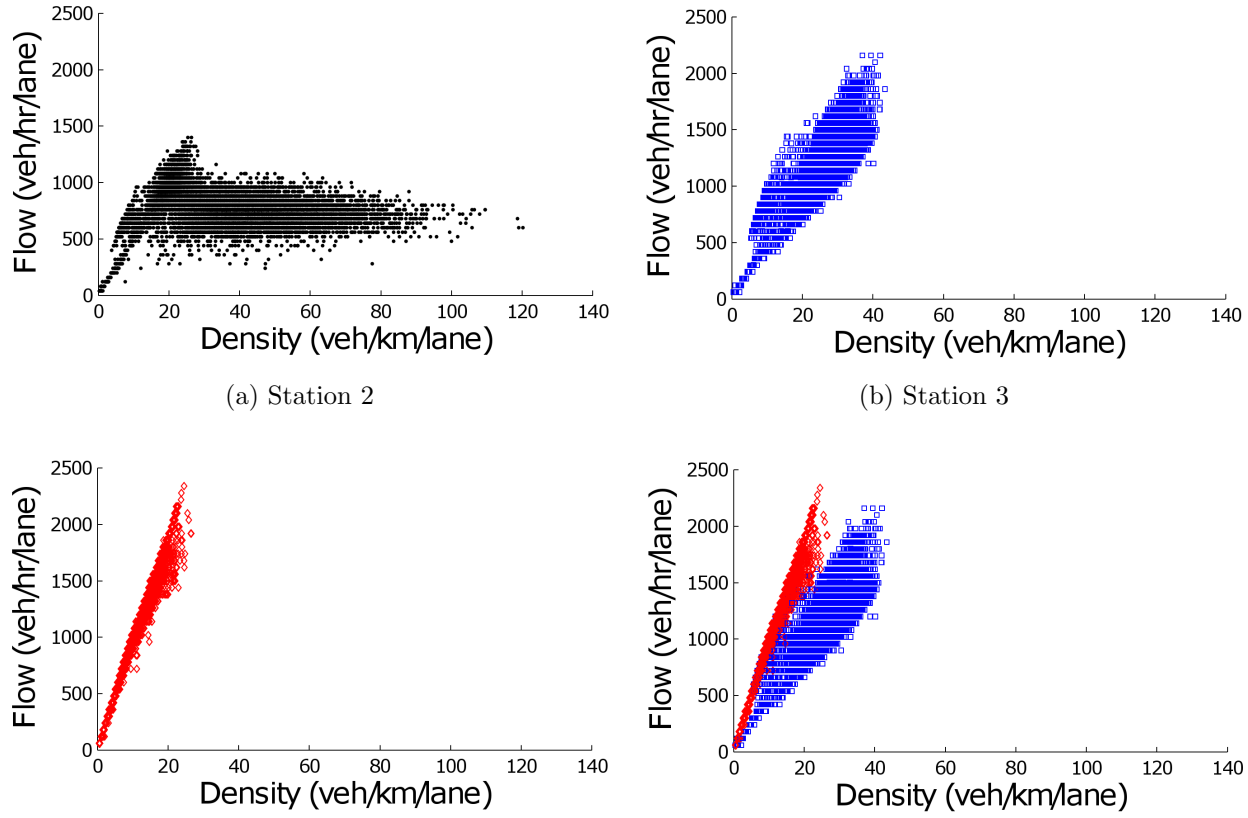



Figure 3.5: Capacity Drops vs. Total Demand for the On-ramp Bottleneck Example (LOW freeway demand)

In general, capacity drops are observed once typical driver acceleration levels were modeled and heavy vehicles are introduced into the traffic stream. These capacity drops are sensitive to the magnitude of total demand entering the network (mainline and on-ramp). This is not surprising since as total demand increases, the amount of congestion increases upstream of the merge. Increasing the percentage of heavy vehicles seems to increase the capacity drops observed. However, the degree to which lane-changing and heavy vehicles contribute to the capacity drop is dependent on the level of freeway demand. The simulated capacity drop is found to be consistent with empirical study findings, as summarized in




(a) Station 2

(b) Station 3

(c) Station 4

(d) Stations 3 and 4 (combined)

Figure 3.6: Flow-Density Plots for the On-ramp Bottleneck Example for Scenario 6 (20% heavy vehicles) and HIGH freeway demand

Table 3.1, since each study considered different loading conditions. These results appear to demonstrate the validity of the INTEGRATION model in replicating the field observed capacity drop after the onset of congestion.

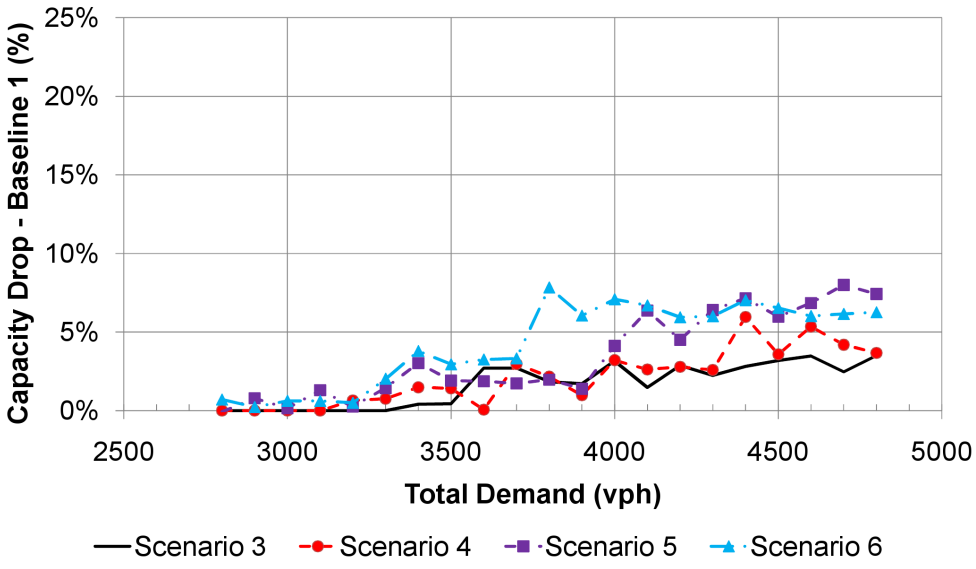
The effect of the level of vehicle acceleration and the percentage heavy vehicles are further illustrated in Figure 3.6. The figures illustrate the variation in the average flow rate at each location as a function of the average roadway density. It should be noted that the use of an average lane measure is required in order to normalize the data across the three stations given that the roadway section at Station 2 is three lanes wide while it is only two lanes wide at the other two stations. In Figure 3.6a, the capacity drop is clearly evident within the bottleneck at Station 2 (see Figure 3.2a for the locations of each station). The free-flow regime (traveling at an average speed of 55 km/hr) seems to end at a critical density of 25 veh/km/lane and a flow rate of 1400 veh/hr/lane (equivalent to a total flow of 4200

veh/hr). Alternatively, it is difficult to estimate an exact flow in the congested regime at the same critical density due to the dispersion of the data; however the average flow at a critical density of 25 veh/km/lane in the congested regime is clearly much less than 1400 veh/hr/lane. Additionally, Figure 3.6a exhibits the typical reverse-lambda shape observed in data from a section that experiences a capacity drop. Also, these results agree with the work in Hall and Agyemang-Duah (1991) where they argue that the capacity drop must be measured within the bottleneck. Just downstream of the bottleneck at Station 3, there is no congestion; however the speed, as was the case at Station 2, has been significantly reduced from the free-flow speed (108 km/hr) to around 52 km/hr, as illustrated in Figure 3.6b. Furthermore, the figure also demonstrates a high level of variability in the traffic stream parameters because the vehicles are accelerating. Further downstream at Station 4 in Figure 3.6c, the speed increases to approximately 95 km/hr. Furthermore, the figure also demonstrates a reduction in the level of noise in the data compared to Station 3. Figure 3.6d is a combined flow-density plot showing the conditions at Stations 3 and 4 which helps to show the differences in speed and the speed variability at the two stations.

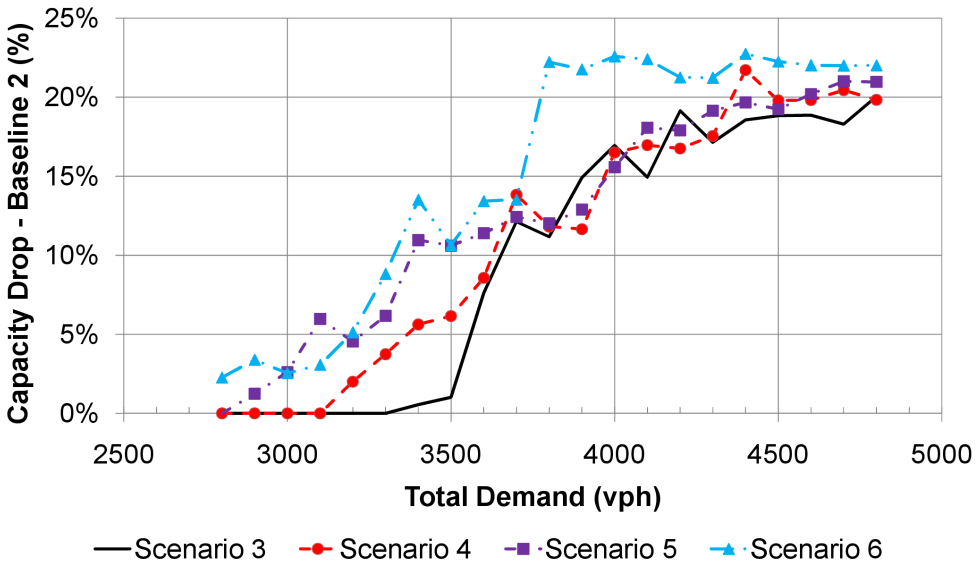
### 3.5.2 Lane Drop Bottleneck Example Results

The results for the lane drop bottleneck example are summarized in Table 3.5 and Figures 3.7a- 3.7b. They include two baseline flow conditions instead of just one condition as was the case with the on-ramp bottleneck example. This section summarizes the results from each of the four scenarios.

As in the On-ramp Example, the results from Scenarios 1 and 2 showed minimal capacity drops but the potential to model capacity drops by using typical vehicle acceleration levels. In Scenario 3, the capacity drops relative to Baseline 1 are initially insignificant. However, for demands in excess of 4400 veh/hr, the capacity drops are in the range of 4 to 8%. The capacity drops relative to Baseline 2 are more than twice those relative to Baseline 1; these capacity drops range from 18 to 22% for demands in excess of 4400 veh/hr. This difference between the flow reductions relative to each baseline flow demonstrates the impact of lane-changing behavior on the capacity drop; by using Baseline 2 to compute the capacity drops, we capture not only the effects of typical acceleration levels and heavy vehicles in the traffic stream but also the effect of lane-changing. This also explains why the capacity drops relative to Baseline 2 seem to converge to 21%, for demands in excess of 4400




(a) according to Baseline 1




(b) according to Baseline 2

Figure 3.7: Capacity Drops as a Function of Total Demand for the Lane Drop Bottleneck Example

Table 3.5: Capacity Drops for the Lane Drop Bottleneck Example

| **Total Demand** | **2800** | **3200** | **3600** | **4000** | **4400** | **4800** |
|---|---|---|---|---|---|---|
| Baseline 1 | | | | | | |
| Scenario 3: 6% trucks | 0.00% | 0.00% | 2.71% | 3.16% | 2.81% | 3.51% |
| Scenario 4: 10% trucks | 0.00% | 0.64% | 0.06% | 3.21% | 5.96% | 3.67% |
| Scenario 5: 15% trucks | -0.07% | 0.26% | 1.86% | 4.12% | 7.15% | 7.42% |
| Scenario 6: 20% trucks | 0.71% | 0.49% | 3.25% | 7.08% | 7.02% | 6.26% |
| Baseline 2 | | | | | | |
| Scenario 3: 6% trucks | 0.00% | 0.00% | 7.61% | 16.95% | 18.56% | 20.06% |
| Scenario 4: 10% trucks | 0.00% | 1.99% | 8.56% | 16.50% | 21.71% | 19.83% |
| Scenario 5: 15% trucks | -0.01% | 4.55% | 11.39% | 15.57% | 19.66% | 20.97% |
| Scenario 6: 20% trucks | 2.27% | 5.12% | 13.42% | 22.58% | 22.74% | 22.02% |

vph, regardless of the heavy vehicle percentage in the traffic stream, as seen in Figure 3.7b. At lower demands, the percentage of heavy vehicles clearly influences the magnitude of the capacity drops however once the demand increases to a level that the lane-changing behavior outweighs all other factors, the capacity drops converge across the different traffic stream scenarios. It should be noted that for each of the four scenarios the baseline flow decreases as the scenario increases. Consequently, the observed increases in discharge flow reductions are not attributed to the impact of trucks but rather are caused by the lower acceleration capabilities of these heavy vehicles.

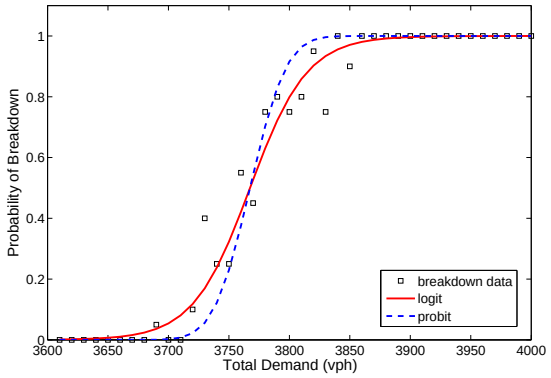
## 3.6 The Probability of Breakdown

In addition to the ability to replicate empirically observed capacity drops, the study also investigated the stochastic nature of the onset of congestion. The onset of congestion, or breakdown, occurred at different demand levels for each of the 20 simulation runs. It was observed that congestion formed for a total demand ranging between 3600 and 4000 veh/hr for the HIGH freeway demand on-ramp example. In order to analyze the probability of breakdown in more detail, additional simulation runs were made with on-ramp demands ranging between 10 - 400 veh/hr (at 10 veh/hr increments). When combined with the previous simulation runs (at 100 veh/hr increments), 40 demand profiles were used across the 4 traffic stream composition scenarios and the 20 random instances - a total of 2880

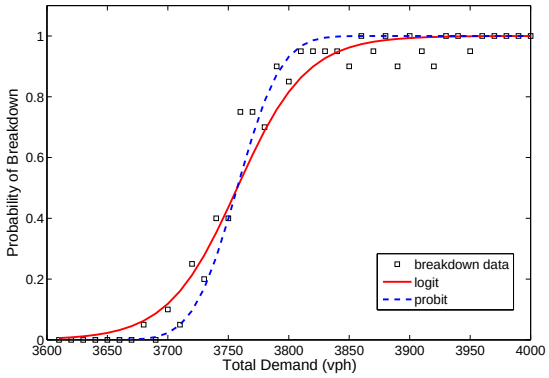
simulation runs. For each of the demand scenarios and for each of the random instances, breakdown was considered to have occurred when the detector occupancy exceeded the 23% threshold at the upstream detector station for more than 3 minutes. The total number of breakdowns were summed across the twenty random instances in order to compute the probability of breakdown for a particular demand scenario (e.g., if for a demand scenario there were 18 out of 20 breakdowns, then the probability of breakdown was 90%). Figure 3.8 displays plots of the probability of breakdown as a function of the total demand loaded onto the on-ramp bottleneck along with model fits using logit and probit models. These results are similar to those reported in Kerner and Klenov (2006); Kerner et al. (2002) where it was hypothesized that the probability for traffic stream breakdown can be approximated using a logistic curve. The figure demonstrates that a logit model appears to provide a better fit to the simulated data when compared to a probit fit.

A similar analysis was performed for the MEDIUM and LOW freeway demand conditions. Figures 3.9 - 3.11 display the fitted curves for each freeway demand level. The curves for the HIGH demand condition are intuitive up to Scenario 6; the probability curves occur at a lower total demand range as the percentage of heavy vehicles increase. However, with a traffic stream composed of 20% heavy vehicles, the breakdown occurs at a higher throughput compared to the 6% heavy vehicle scenario. The results seem to indicate that when the heavy vehicle volume exceeds a threshold, the system throughput increases. One potential explanation is that the trucks leave larger gaps ahead of them thus allowing the on-ramp demand to merge easier on to the freeway and thus increase the system throughput. However, this behavior was not observed in the case of the MEDIUM and LOW freeway demands. In the case of the LOW freeway demand, the goodness-of-fit of the curves was much worse than in the HIGH freeway demand condition. This is especially noticeable in the Scenario 6 probability of breakdown curve. It should be noted that the results also indicate that the variability in the pre-breakdown flow increases as the volume of heavy vehicles increases.

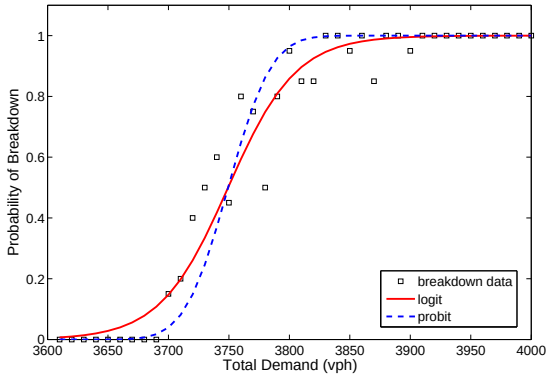
The observed breakdown behavior indicates that further research in this area is necessary to study the impact of heavy vehicles on the probability of breakdown and whether the probability of breakdown continues to follow a normal distribution at low fixed freeway demands.



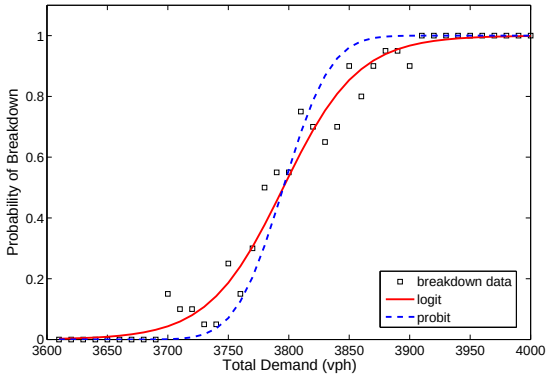

(a) Scenario 3 - 6% heavy vehicles




(b) Scenario 4 - 10% heavy vehicles




(c) Scenario 5 - 15% heavy vehicles




(d) Scenario 6 - 20% heavy vehicles

Figure 3.8: Probability of Breakdown Curves for the On-ramp Bottleneck Example (HIGH freeway demand)

## 3.7 Conclusions & Future Research Directions

The study conducted simulation runs on two common freeway bottleneck examples, namely an on-ramp and lane-reduction bottleneck. The study demonstrated that (a) a capacity drop can be created using an agent-based microscopic modeling approach without enforcing a discontinuity in the car-following steady-state relationship (fundamental diagram); (b) the INTEGRATION software is valid for replicating empirically observed capacity drops after the onset of congestion both in terms of the magnitude of these reductions and the probabilistic nature of the traffic stream failure; (c) these empirically observed capacity drops increase as vehicles accelerate less aggressively and as trucks are introduced into the traffic stream; (d) the increase of truck volumes beyond 6% does not necessarily produce

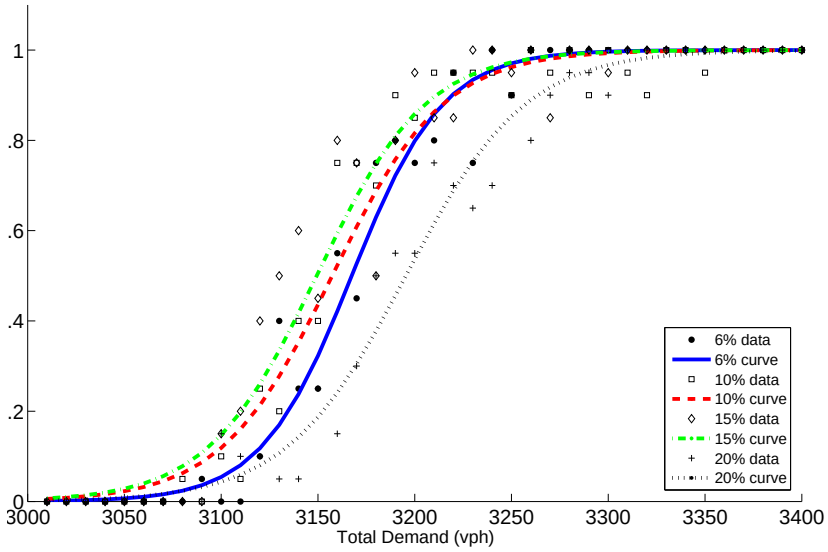



Figure 3.9: Probability of Breakdown Curves with HIGH Freeway Demand for the On-Ramp Bottleneck Example

significantly larger capacity drops especially for high freeway demand levels, however they play a more significant role for lower freeway demand levels; (e) the occurrence of traffic stream breakdown is probabilistic and generally follows a normal distribution; and (f) the probability of breakdown curves are influenced by the percentage of heavy vehicles differently depending on the level of freeway demand.

During the conduct of this research, further research would prove worthwhile in the following areas:

- Future research should include a sensitivity analysis of the effect of loop detector locations both upstream and downstream of bottlenecks on capacity drop measurements.
- Up to this point, we have used one variation of the Van Aerde - Rakha traffic model; further investigation may provide additional insight using different assumptions for the speed-at-capacity parameter (e.g., set $u_c = 0.8u_f$).
- In order to further validate the use of INTEGRATION, additional research is recommended to evaluate capacity drops at other types of lane drop, merging, diverging, and weaving intersection configurations.

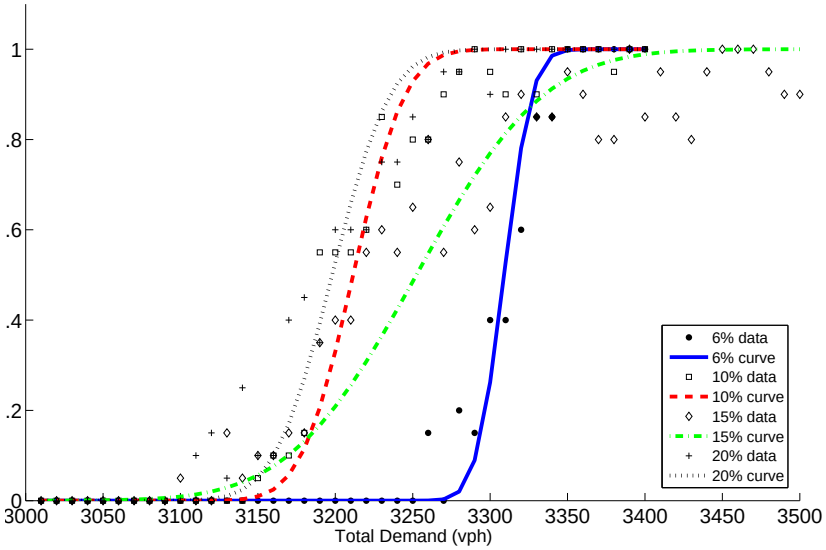



Figure 3.10: Probability of Breakdown Curves with MEDIUM Freeway Demand for the On-Ramp Bottleneck Example

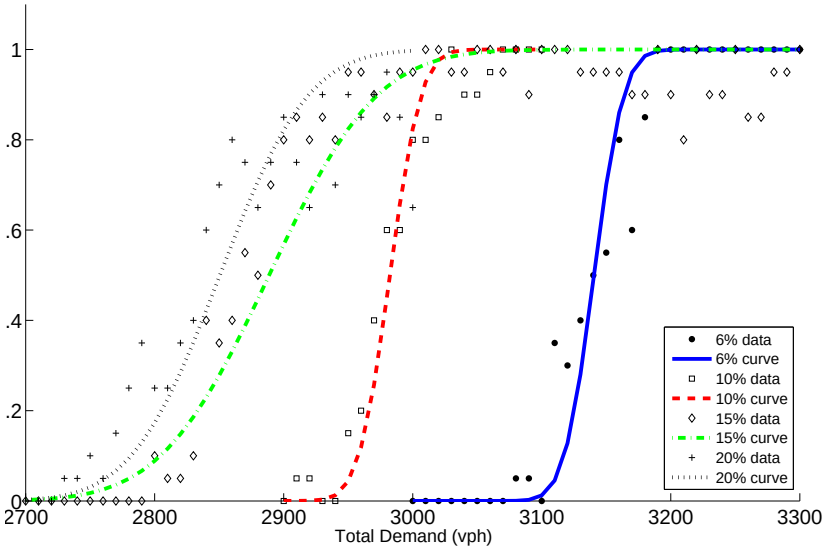



Figure 3.11: Probability of Breakdown Curves with LOW Freeway Demand for the On-Ramp Bottleneck Example

# Chapter 4

# Optimizing Network Flows with Congestion-based Flow Reductions

**Abstract:** When optimizing traffic systems using time-expanded network flow models, traffic congestion is an important consideration because it can decrease both the flow rate and speed of traffic. This paper proposes a mixed integer linear program (MILP) that captures the reduction in traffic flows due to the formation of congestion upstream of a bottleneck using a modified form of the Cell Transmission Model (CTM) as developed in Daganzo (1994, 1995). Although the MILP is more complex compared to the commonly used linear program (LP) developed in Ziliaskopoulos (2000), the MILP formulation does not produce the same "traffic holding" problem, which we consider an unrealistic level of control. This allows the MILP to produce an optimal solution for network flow that includes speed and flow reductions due to congestion; the improved mathematical program is called the MILP with flow reductions (MILP-FR). This improvement reflects more accurate estimates of total travel time and network clearance time and allows for the future application of the MILP-FR to optimize network flows with respect to congestion. Comparisons are made between the results from the MILP-FR, the LP as developed in Ziliaskopoulos (2000), and a microscopic traffic simulation. It is shown that the LP solution underestimates network clearance and total travel times by over 40% when compared to the simulation solutions for two freeway bottleneck examples. Additionally, it is demonstrated that the MILP-FR can be calibrated to match the solutions given by the simulation. Furthermore, this MILP-FR model can be solved efficiently by considering the relationships between the binary decision variables.

## 4.1 Introduction

Models that optimize network traffic flows over time are useful in many traffic planning applications; examples include dynamic traffic assignment (DTA) and regional evacuation planning (our main area of interest). The Cell Transmission Model (CTM) introduced in Daganzo (1994, 1995) and incorporated into a Linear Program (LP) in Ziliaskopoulos (2000) within the context of a DTA problem is such a model. The CTM is a first-order macroscopic traffic stream model that conforms to the hydrodynamic traffic flow and density relationships proposed by Lighthill and Whitham (1955) and Richards (1956) (commonly referred to as the LWR model). The CTM provides a promising modeling framework for traffic flow optimization because it models the spillback propagation and speed reduction caused by congestion, and can be incorporated, in an approximate manner, within an LP (see Ziliaskopoulos, 2000) which are generally computationally efficient. This modeling framework has been used or studied in numerous papers: DTA research (Li et al., 2003; Ukkusuri and Waller, 2008), ramp metering and traffic signal control (Lo, 2001; Lin and Wang, 2004; Gomes and Horowitz, 2006), and evacuation modeling and planning (Tuydes and Ziliaskopoulos, 2006; Sbayti and Mahmassani, 2006; Liu et al., 2006; Chiu et al., 2007; Yazici and Ozbay, 2008; Xie et al., 2009; Yao et al., 2009).

Optimization-based traffic flow models must include many simplifications in order to be tractable - traffic flow is a complex phenomena (see Kerner and Rehborn, 1996; Nagel et al., 2003; Schönhof and Helbing, 2007, for instance). Because of this the modeling framework must be based on appropriate simplifications so that important higher level effects are properly modeled, furthermore, results from an optimization-based framework should be tested using higher fidelity simulation models. Examining the CTM from this perspective, the most significant problem we find is that *it fails to capture the reduction in the flow discharge rate after the onset of congestion.* Traffic stream models typically have flows that initially increases with density until a critical density is reached and then flow decreases as density increases eventually reaching a flow of zero at jam density (May, 1990). The fact that congestion upstream of a bottleneck reduces downstream flows by somewhere between 10 to 20% has been extensively shown in various empirical studies (Banks, 1990; Hall and Agyemang-Duah, 1991; Cassidy and Bertini, 1999; Chung, Rudjanakanoknad, and Cassidy, 2007) and in traffic simulations (Chamberlayne, Rakha, El-Metwally, and Bish, 2011b; Chamberlayne, Rakha, and Bish, 2011a). In Papageorgiou (1998), the author states that LWR-based models like the CTM will seek to create congestion in order to maximize the

network outflow. This issue will be discussed further in Section 4.2 where we propose a modification of the CTM to better model congestion.

When incorporating the CTM into an LP (see Ziliaskopoulos, 2000) the resulting solution often exhibits *traffic holding* when the network contains a bottleneck. Traffic holding is an artifact of using a mathematical program to determine network flows with congestion without the proper constraints; the lack of proper constraints allows the model to have an unrealistic level of control and restricts the flow (i.e., "traffic holding") at locations where it is not realistic or reasonable. This problem will be discussed in more detail in Section 4.3.2, while the implications of traffic holding on our proposed modification of the CTM is discussed in Section 4.3.3, and a solution to eliminate unwanted traffic holding is presented in Section 4.3.4.

To use optimization-based models to develop traffic management strategies (see Gomes and Horowitz, 2006; Chiu et al., 2007, for instance), it is important that they model the effects of congestion sufficiently and do not allow traffic holding to produce an unrealistic solution. For instance, when considering network flows during an emergency evacuation, congestion can become especially severe, lasting for hours, and causing large reductions in the network flow capacity. For obvious reasons, this type of congestion should be avoided since it is not only an inconvenience but can threaten lives during an emergency evacuation. Furthermore, if a solution depends on traffic holding, yet there is no way to enforce this behavior, it will not be effective. This paper advances the understanding of this difficult problem of network flow over time with congestion by making the following specific contributions:

1. The development of a macroscopic, first order traffic model based on the CTM that reduces flows due to the effects of congestion at bottlenecks, called the "CTM with flow reductions" (CTM-FR). It should be noted that higher order models can reproduce these flow reductions however the solution of such problems is much more difficult to obtain.

2. The implementation of the CTM-FR within a mathematical program that eliminates unwarranted traffic holding, called the mixed integer linear program for flow reductions (MILP-FR).

3. Improvements to the computational performance of the MILP-FR are developed that enable practical use of the mathematical program. The performance of the CTM-FR and the MILP-FR is illustrated using two common freeway bottleneck examples and compared to a more detailed traffic simulation.

This paper seeks to capture the net effect of congestion within a mathematical program for the purposes of network flow optimization; it is not an attempt to advance traffic flow theory but is intended for use in transportation network planning.

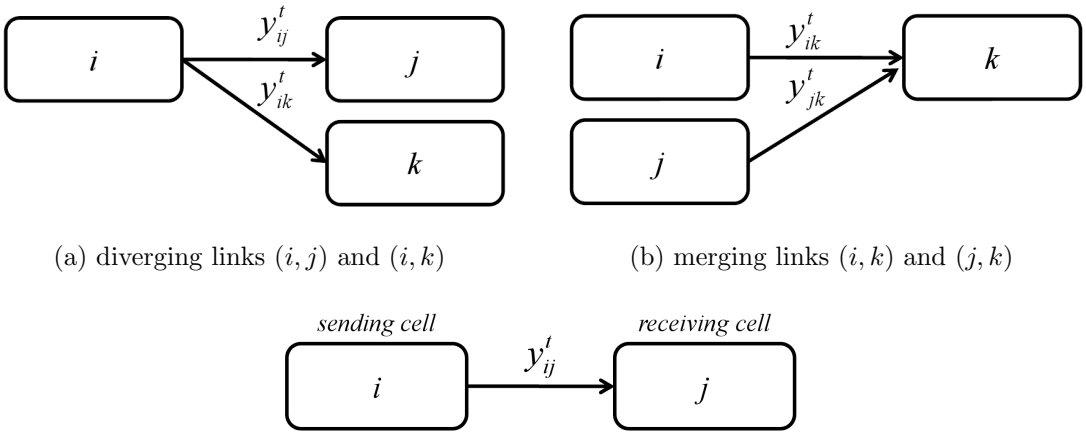
This paper is structured as follows. Section 4.2 reviews the CTM and proposes an enhancement to the CTM, called CTM-Flow Reduction (CTM-FR). This section ends by comparing the results of the CTM, CTM-FR, and a microscopic traffic simulation (called INTEGRATION) for a freeway lane drop example. Section 4.3 reviews the linear program (LP) presented in Ziliaskopoulos (2000) that incorporated the CTM, discusses the problem of "traffic holding," proposes a MILP incorporating the CTM-FR, and presents some constraints that reduce the computation effort required to solve the MILP. Section 4.4 demonstrates the impact and tractability of the MILP in a freeway merge example, while Section 4.5 presents the research conclusions.

## 4.2 Enhancing the Cell Transmission Model

In this section, we provide an overview of the CTM, discuss the effects of congestion, introduce a modification to the CTM that reduces the discharge flow (in addition to speed) at the onset of congestion, and demonstrate the results of this modification with a freeway bottleneck example and a microscopic traffic simulation.

The CTM utilizes a discrete time-expanded network of cells and links $(C, L)$ to represent a roadway system. The set of cells, $C$, is composed of the following subsets: source cells $(S_o)$, sink cells $(S_i)$, and roadway cells $(R)$. A roadway cell represents a section of roadway of length $\ell$ such that vehicles traveling at free-flow speed $(u_f)$ traverse the section in one time interval of length $\tau$; that is $\ell = u_f \times \tau$ for each cell. Additionally, $n_i$ represents the number of lanes in cell $i$. Besides free-flow speed, the model requires two other macroscopic parameters as inputs in order to compute the cell parameters: saturation flow rate $(q_c)$ and jam density $(k_{jam})$. Links represent allowable movements between cells. Links are classified as follows: ordinary links $(L_O)$, merging links $(L_M)$ (where multiple links enter a cell), and diverging links $(L_D)$ (where multiple links exit a cell). See Figure 4.1 for examples of these links.

Unlike Daganzo (1995), we categorize *links*, instead of cells, as ordinary, merging, and diverging because the flow constraints are at the link level. The notation used in this paper




(a) diverging links $(i, j)$ and $(i, k)$

(b) merging links $(i, k)$ and $(j, k)$

(c) ordinary link $(i, j)$

Figure 4.1: Link Flow Examples in the CTM

is as follows:

Sets:

- $C$ : set of cells within the network; subsets are sources $(S_o)$, roadway cells $(R)$, and sinks $(S_i)$
- $L$ : set of links between cells within the network; subsets are ordinary links $(L_O)$, merging links $(L_M)$, and diverging links $(L_D)$

Parameters:

- $T$ : number of time intervals of length $\tau$ in the planning horizon
- $Q_i$ : the maximum flow into or out of cell $i$, which is based on the saturation flow rate $(q_c)$, the length of time interval $(\tau)$, and the number of lanes in cell $i$ $(n_i)$ $[Q_i = q_c \times \tau \times n_i]$, $\forall i \in R$
- $N_i$ : the maximum number of vehicles that cell $i$ can hold, which is based on the cell length $(\ell_i)$, number of lanes in cell $i$ $(n_i)$, and jam density $(k_i^{jam})$ $[N_i = k_i^{jam} \times \ell_i \times n_i]$, $\forall i \in R$
- $\delta_i$ : is the ratio of wave speed to free-flow speed for cell $i$: $w/u_f$, $\forall i \in R$
- $d_i^t$ : number of vehicles that flow out of the source $i$ in time interval $t$, $\forall i \in S_o$, $t = 1, \cdots, T$

Decision Variables:

$x_i^t$ : number of vehicles in cell $i$ at the beginning of time interval $t$, $\forall i \in C$, $t = 1, \cdots, T$

$y_{ij}^t$ : number of vehicles that flow from cell $i$ to cell $j$ over link $(i, j)$ in time interval $t$, $\forall (i, j) \in L$, $t = 1, \cdots, T$

The $x_i^t$-variable is equivalent to cell density and the $y_{ij}^t$-variable is equivalent to flow in units of $vehicles/\tau$. Note that while $Q_i, N_i$ and $\delta_i$-parameters do not vary with respect to time in this paper, it is easy to include time-varying parameters. The $\delta$-parameter (the ratio of shockwave speed to free-flow speed) is set equal to 1 as long as the macroscopic input parameters are within an appropriate range; $Q_i/(N_i - Q_i)$ should be in the range of $1/6 - 1/4$ (see Dixit et al., 2008; Kalafatas and Peeta, 2006) for freeway modeling.

Equations (4.1) and (4.2) are adapted from Daganzo (1994). Equation (4.1) calculates the flow from cell $i$ to cell $j$ over an ordinary link $(i, j)$ (see Figure 4.1c), while Equation (4.2) calculates the traffic density as it changes over time.

$$y_{ij}^t = \min\{x_i^t, Q_i, Q_j, \delta_j[N_j - x_j^t]\} \tag{4.1}$$

$$x_j^t = x_j^{t-1} - \sum_{k:(j,k)\in L} y_{jk}^{t-1} + \sum_{i:(i,j)\in L} y_{ij}^{t-1} \tag{4.2}$$

Figure 4.2 (adapted from Daganzo, 1994) illustrates, from a cell-level perspective, how the properties of cell $j$, having ordinary links entering and exiting the cell, constrain the flow into and out of cell $j$. We emphasis constrain, because the actual flows into (out of) cell $j$ depend also on the properties of cell $i$ ($k$). Based on the properties of cell $j$, the flow from cell $i$ is limited by $y_{ij}^t(in) = \min\{Q_j, \delta_j[N_j - x_j^t]\}$ and the flow out of cell $j$ (to cell $k$) is $y_{jk}^t(out) = \min\{x_j^t, Q_j\}$. Thus $y_{jk}^t(in)$, see the left vertical axis, is constrained by the third and fourth terms of Equation (4.1) for link $(i, j)$; these terms are depicted by the horizontal solid line representing the $Q_j$-parameter and the dark solid line with a slope of $-\delta$ which represents the space available in the cell $j$. The flow $y_{jk}^t(out)$ is displayed along the right vertical axis in the figure and is constrained by the first two terms in Equation (4.1) for link $(j, k)$; these terms are depicted by the light solid line with a slope of 1 (which represents free-flow speed in terms of the CTM parameters) and the horizontal solid line representing

the $Q_j$-parameter.

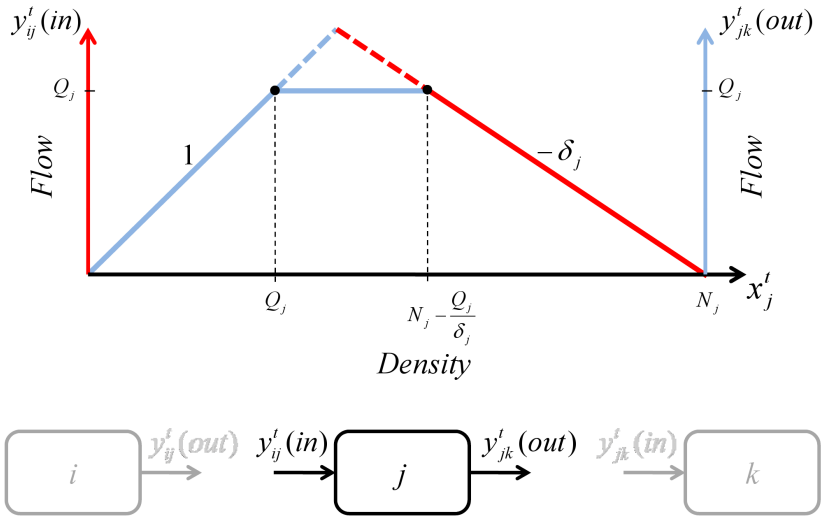



Figure 4.2: Flow-density relationship for cell $j$ for the generalized CTM (adapted from Daganzo, 1994).

Figures 4.3a and 4.3b, with their link-level perspective, illustrate (4.1) for an ordinary link $(i, j)$. In this case an ordinary link at a lane drop, i.e., the receiving cell has fewer lanes than the sending cell. These two figures show how the sending cell $i$ and receiving cell $j$ determine the flow on link $(i, j)$. Again, the potential flow out of sending cell $i$ is denoted as $y_{ij}^t(out) = \min\{x_i^t, Q_i\}$ (see Figure 4.3a). Because $Q_i$ is a parameter, and not a function of $x_i^t$, high densities in $i$ do not impact the potential flow out of $i$. The congested regime for cell $i$ does, however, cause a non-linear speed reductions as densities increase (see Figure 4.3c). The flow allowed into receiving cell $j$ is denoted as $y_{ij}^t(in) = \min\{Q_j, \delta_j[N_j - x_j^t]\}$ (see Figure 4.3b). Thus Equation (4.1) can be expressed as $y_{ij}^t = \min\{y_{ij}^t(out), y_{ij}^t(in)\}$.

Consider a network having a single bottleneck represented by cells $i$ and $j$, where $Q_i \geq Q_j$ (see Figures 4.3a and 4.3b) and the density in each cell in the network is initially in the free-flow regime ($x_k^1 \leq Q_k, \ \forall k \in R$). Given sufficient density in cell $i$ ($x_i^t \geq Q_j$) the flow out of $i$ will be $Q_j$, the maximum possible value. This is because Equation (4.1) has no mechanism for reducing the potential flow from cell $i$ based on the traffic density of cell $i$ and the only way to limit the flow from cell $i$ to a value less than $Q_j$ (once again, given sufficient demand) is for cell $j$ to become congested to such an extent ($x_j^t \geq N_j - Q_j/\delta_j$) that the term $\delta_j[N_j - x_j^t]$ minimizes Equation (4.1). As this is the only bottleneck, by definition every cell downstream of $j$ has a $Q$-value greater than or equal to $Q_j$ and thus cell $j$ will

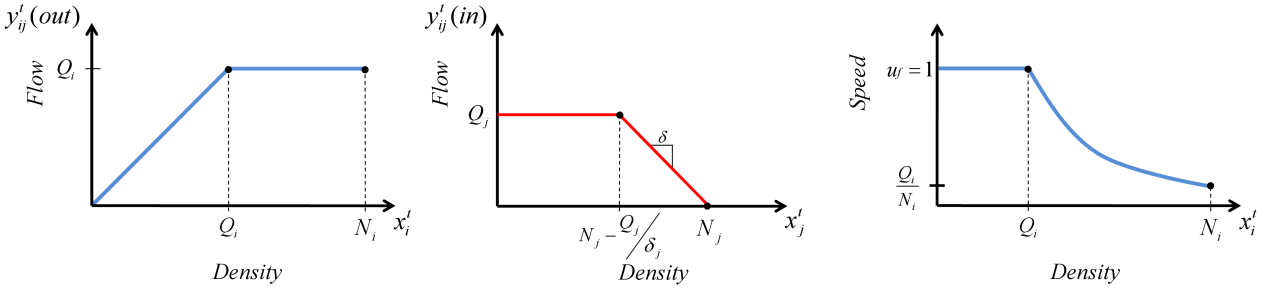



Figure 4.3: Depiction of (a) flow-density relationship for sending cell $i$, (b) flow-density relationship for receiving cell $j$, and (c) speed-density relationship for sending cell $i$.

remain in the free-flow regime and congestion will not build in $j$. The traffic density in cell $i$ will increase until a congestion equilibrium is reached (or demand lessens). Within the CTM, the congestion equilibrium density is equal $N_i - Q_j/\delta_i$ where $Q_j$ is the capacity of the system bottleneck and $N_i$ is the storage space of cell $i$ upstream of the bottleneck. However, the congestion equilibrium flow between cell $i$ and cell $j$ will be $Q_j$ thus congestion does not negatively impact the flow through the system bottleneck. While, the traffic speed at jam density ($x_j^t = N$) is $Q_i/N_i$, the speed at the equilibrium will be higher - $Q_i/\{N_i - Q_j/\delta_i\}$. In systems with multiple bottlenecks (i.e., restrictions) one set of bottlenecks will dominate and determine the flow rate out of the network, and using similar logic, this network flow rate is not reduced because of congestion. This has interesting implications, for instance, in an evacuation, this would imply *that congestion will not increase the overall duration of the evacuation* (i.e., the network clearance time). This contradicts evidence from actual evacuations and empirical data from non-evacuation studies; the following traffic studies show a flow reduction due to congestion (Chung et al., 2007; Cassidy and Bertini, 1999; Hall and Agyemang-Duah, 1991; Banks, 1990).

To model reductions in flow caused by congestion, we modify the CTM. It should be noted that others have attempted to capture this flow reduction through the use of higher order models. However, the use of a higher order model adds complexity to the problem including additional non-linearities. Here we are proposing a simple enhancement to an existing first order model in order to ensure that the problem can be solved efficiently. Specifically, we modify the second term in Equation (4.1) to be sensitive to the cell density of the *sending cell*. We will refer to this modified CTM as the CTM-FR for "flow reduction". To do this we introduce a new parameter that will govern how congestion impacts flow.

Additional Parameter:

$\Omega_i$ : the maximum flow rate out of cell $i$, when cell $i$ is at the maximum traffic density ($x_i^t = N_i$), $\Omega_i \in (0, Q_i]$

The $\Omega_i$-parameter will be expressed as a flow rate or as a percentage or proportion of $Q_i$ (all forms are equivalent). We replace the second term of Equation (4.1), which minimizes $y_{ij}^t(out)$ when $Q_i < x_i^t \leq N_i$ with a linearly decreasing function of density where $y_{ij}^t(out) = Q_i$ when $x_i^t = Q_i$ and $y_{ij}^t(out) = \Omega_i$ when $x_i^t = N_i$. This yields Equation (4.3), which replaces Equation (4.1) in the CTM-FR.

$$y_{ij}^t = \min\{x_i^t, x_i^t \frac{\Omega_i - Q_i}{N_i - Q_i} + N_i \frac{Q_i - \Omega_i}{N_i - Q_i} + \Omega_i, Q_j, \delta_j[N_j - x_j^t]\} \tag{4.3}$$

The $\Omega_i$-parameter has a maximum possible value of $Q_i$, which yields the CTM, thus the CTM is a special case of the CTM-FR. It also must have a value greater than zero. If $\Omega_i = 0$ then vehicles would be trapped in cell $i$ if the jam density were ever attained (this can only occur because of the initial state of the system or due to changes in the flow parameters over time). While theoretically flow at jam density is zero, conceptually, the traffic density for vehicles about to leave cell $i$ is determined by both the densities in $i$ and in the downstream receiving cell(s). Flow from cell $i$ should only equal zero because of the state of the receiving cell(s). Figure 4.4, which corresponds to Figure 4.3 for the CTM, illustrates the new constraint within the sending cell.

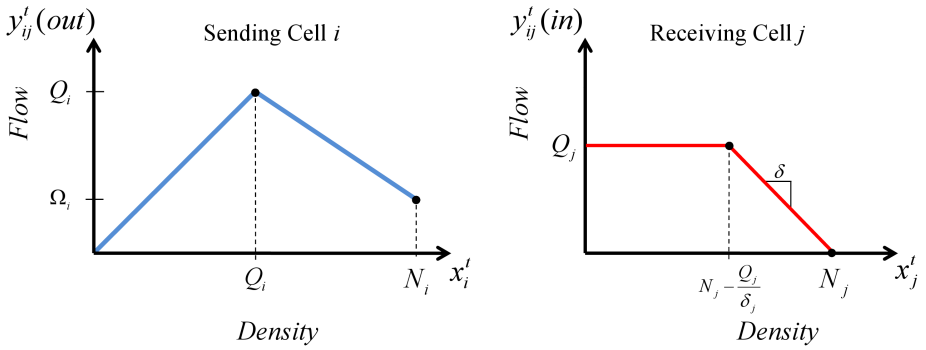



Figure 4.4: The flow-density relationship for the CTM-FR for flows over link $(i, j)$.

Congestion in the CTM-FR (as well as in the CTM) adheres to the following property.

**Property 1** *As congestion builds, the density shockwave propagates in a backward forming wave from the bottleneck (upstream), and as the congestion dissipates, the density shockwave propagates in a forward recovery wave towards the bottleneck (downstream).*

In order to describe this well known property, we return to the cell-level perspective. First we examine how the traffic density in cell $j$, an arbitrary roadway cell, determines the potential flows into and out of $j$. Figure 4.5 graphically displays the potential flows into and out of $j$ over the range of possible traffic densities (note: the vertical scale of the figure is exaggerated in order to depict the flows and densities of interest). There are two vertical axes: the left vertical axis depicts the potential flows into $j$ ($y_{ij}^t(in)$) and the corresponding flow constraints are displayed in the dark solid lines; the right vertical axis depicts the potential flows out of $j$ ($y_{jk}^t(out)$) and the corresponding flow constraints are displayed in the light solid lines.

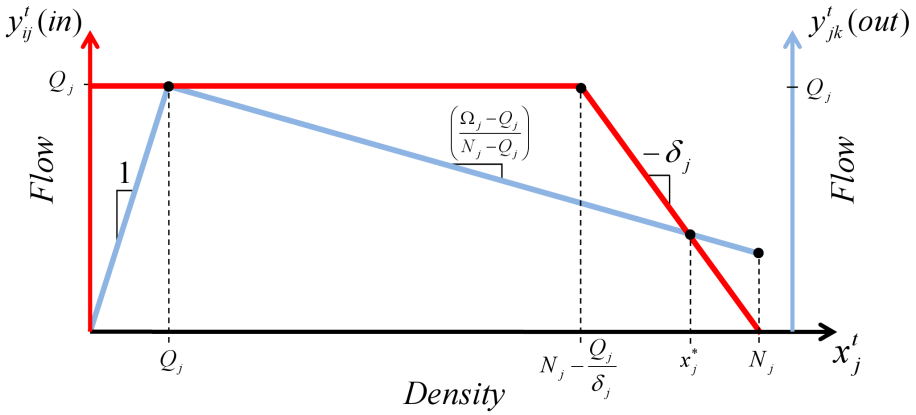



Figure 4.5: The flow into and out of cell $j$ for the CTM-FR and the equilibrium density, $x_j^*$

There exists a congestion equilibrium $x_j^*$ where $y_{ij}(in) = y_{jk}(out)$; this is calculated as the intersection of the flow constraints for $y_{ij}(in)$ and $y_{jk}(out)$, $x_j^* = \dfrac{\delta_j N_j^2 - N_j Q_j - \delta_j N_j Q_j + Q_j \Omega_j}{\delta_j (N_j - Q_j) - Q_j + \Omega_j}$. For densities less than $x_j^*$ we see from Figure 4.5 that $y_{ij}^t(in) > y_{jk}^t(out)$. This relationship holds in general due to restrictions on the terms in Equation (4.3). Furthermore, the $\delta_j$-parameter (usually a value of $\delta_j = 1$ is used (see Kalafatas and Peeta, 2006), but nevertheless $d_j \leq 1$) is used to correctly model the shockwave effect. This parameter ensures that the largest potential difference $y_{ij}^t(in) - y_{jk}^t(out)$ in the congested regime ($x_j^t > Q_j$) is less than or equal to $x_j^* - x_j^t$ thus ensuring the density does not overshoot the congested equilibrium density ($x_j^*$). The fact that $y_{ij}^t(in) \geq y_{jk}^t(out)$ drives the backward propagation of congestion as discussed in Property 1.

We now use the following analysis on flows through a bottleneck to demonstrate the veracity of Property 1. Consider roadway cells $i$, $j$, and $k$, which are connected in series ($i$-$j$-$k$), having maximum flows such that $Q_i = Q_j > Q_k$, thus $k$ is a bottleneck. Also assume that initial cell density place the cells in the free-flow regime and that there is sufficient demand to activate the bottleneck (i.e., the flow into cell $i$ is greater than $Q_k$ for a number of time intervals). If there are no bottlenecks downstream of cell $k$, then cell $k$ will always remain in the free-flow regime. The flow on link $(j, k)$ can be less than the potential flow out of cell $j$ because flow on this link must be no greater than $Q_k$. If $Q_k \geq x_j^* \frac{\Omega_j - Q_j}{N_j - Q_j} + N_j \frac{Q_j - \Omega_j}{N_j - Q_j} + \Omega_j$, then congestion can reduce flow through the bottleneck, else a new equilibrium will be found at a higher density than $x_j^*$. Cell $i$ can send a flow of up to $Q_i = Q_j$ into cell $j$ until the density in cell $j$ reaches $N_j - Q_j/\delta_j$, after which the fourth term of Equation (4.3) limits the flow. Thus the congestion density begins propagating backwards. If the flow into the system (i.e., into cell $i$) becomes less than the equilibrium flow, then the density reduction will propagate forward. This occurs since $y_{ij}^t(in) < y_{jk}^t(out)$.

### 4.2.1 Example 1: Freeway Lane Drop Bottleneck

This example studies a five mile section of freeway that contains a lane drop from 3-lanes to 2-lanes, thus forming a bottleneck. We compare the results obtained from the CTM, CTM-FR, and a microscopic traffic simulation (see below for discussion). Figure 4.6 displays this example, which includes only ordinary links, along with the relevant input parameters. This example is similar to one introduced in Papageorgiou (1998) that was used to illustrate modeling deficiency of LWR model, such as the CTM. For comparison purposes we will use two common performance measures: total travel time (TTT) and network clearance time (NCT).

For this example we would expect congestion to cause a reduction of flow at the bottleneck. Factors involved in the formation of the congestion include: the obvious reduction in capacity from 3-lanes to 2-lanes (at Cell 9), lane changing behavior, traffic stream compositions (percentage of heavy vehicles), and vehicle dynamics (acceleration/deceleration losses). The congestion will reduce the speed of the traffic and the capacity of the bottleneck. The impact of various factors on traffic flows is explored further in Chamberlayne et al. (2011a).

To compare our results from the CTM and CTM-FR for this example, we use the

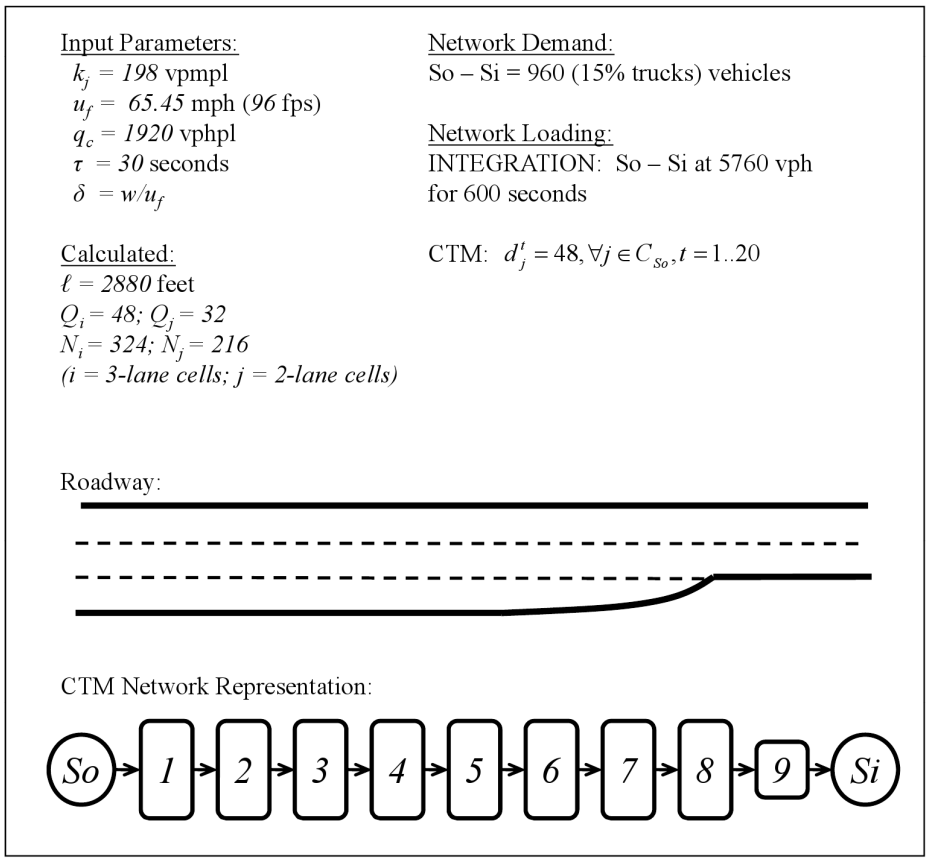
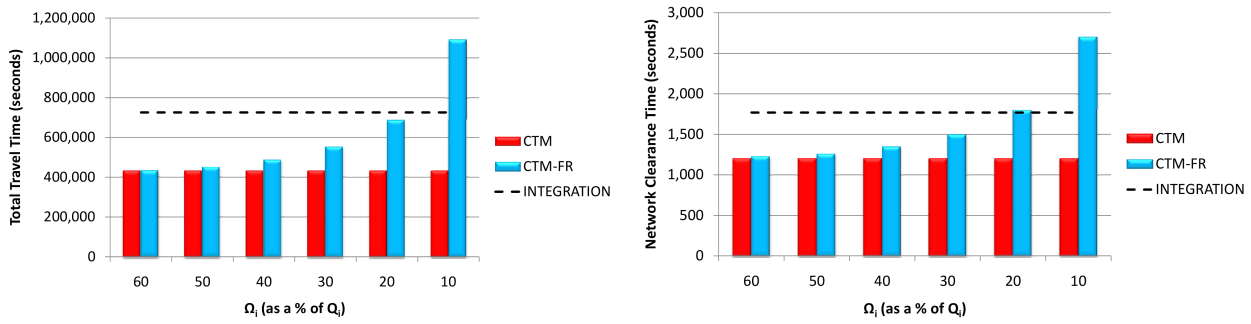



Figure 4.6: Example 1: Freeway Lane Drop Bottleneck.

microscopic traffic simulation package INTEGRATION. A conceptual discussion of INTEGRATION is provided in Van Aerde and Yagar (1988a,b), while a discussion of the actual software is provided in Van Aerde and Rakha (2007a,b). INTEGRATION is an integrated simulation and traffic assignment model and performs traffic simulations by tracking the movement of individual vehicles every 0.1 seconds. This allows for the detailed analysis of lane-changing movements and shock wave propagations and permits considerable flexibility in representing spatial and temporal variations in traffic conditions. INTEGRATION uses a realistic vehicle dynamics model and a psycho-physical model for car-following based on the Van Aerde steady-state car-following model formulation introduced in Van Aerde (1995); Van Aerde and Rakha (1995).

The INTEGRATION model has not only been validated against standard traffic flow theory (Dion et al., 2004; Rakha and Crowther, 2002, 2003; Rakha et al., 2001), but also has been utilized for the evaluation of real-life applications (Rakha, 1991; Rakha et al., 2005,

1998). The INTEGRATION lane-changing logic was validated against field data in an earlier publication (Rakha and Zhang, 2004). Furthermore, Rakha and Zhang (2006) demonstrated the validity of the INTEGRATION software for estimating the capacity of weaving sections by comparing to field observed weaving section capacities.




(a) (b)

Figure 4.7: Comparison of (a) TTT and (b) NCT estimates from the CTM, CTM-FR, and INTEGRATION for Example 1.

The results from the CTM-FR were evaluated at values of $\Omega_i$ that ranged from $0.10Q_i$ to $0.60Q_i$; values of $\Omega_i$ greater than $0.60Q_i$ were not considered since they are greater than the capacity of the 2-lane bottleneck. Figure 4.7 shows the comparisons of the TTT and NCT estimates from the CTM, CTM-FR, and INTEGRATION. Clearly the CTM is a macroscopic traffic model while INTEGRATION is a microscopic traffic simulator, therefore we expect INTEGRATION to produce results with behavior not possible in a macroscopic model such as the CTM or CTM-FR. For instance, INTEGRATION allows for the modeling of heavy vehicles within the traffic stream. The traffic composition used was 85% passenger cars and 15% trucks; typically the percentage of trucks along a freeway is in the range of 5-12% (USDOT, 2008), however, increasing this percentage to 15% accounts for buses and additional cargo that the evacuees may be carrying (Lim, 2003). From Figure 4.7 we see that $\Omega_i$ should be about $0.20Q_i$ in order for the CTM-FR to match the simulation results for our high-level performance measures.

The bottleneck flows for this example are shown in Figure 4.8 from the CTM, CTM-FR, and INTEGRATION. In all three models, congestion builds upstream of the bottleneck at Cell 8. However, in the CTM, the flow into the bottleneck is unaffected by the congestion in the cells upstream of the bottleneck. The CTM does not reduce the flow at the bottleneck based on the density in the sending cell. The calculation of flow, in accordance with Equa-

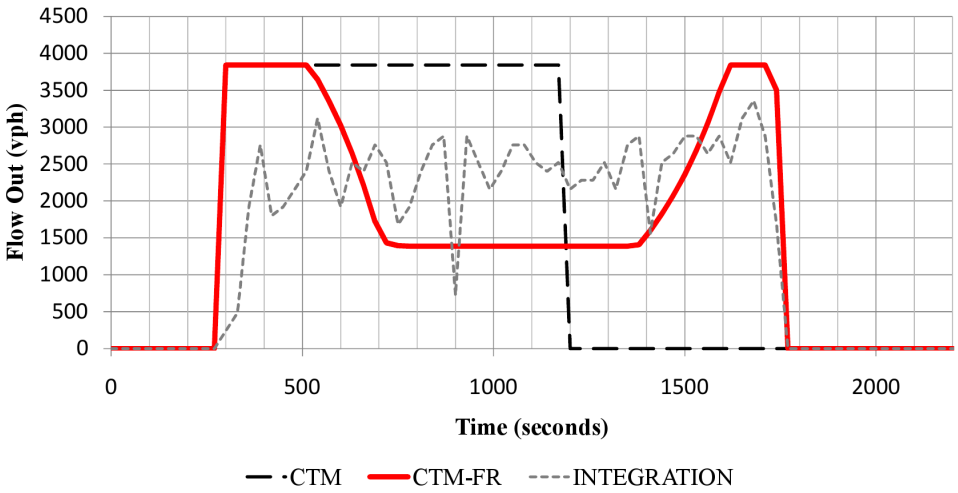



Figure 4.8: CTM, CTM-FR, and INTEGRATION flows out of the network over time.

tion (4.1), along link $(i, j)$ will be dominated by $Q_9$ since $Q_9 < Q_i$ and $Q_9 < x_i$, $\forall i = 1, \cdots, 8$ (assuming an active bottleneck exists). According to the CTM, flow at bottleneck between Cell 8 and Cell 9 will be at the maximum 2-lane value of $2 \times 1920 = 3840$ veh/hr (or $Q = 32$) irrespective of the level of congestion in the upstream cells. Papageorgiou (1998) discusses this same problem, stating that in macroscopic formulations based on the LWR model, "...the presence of a mainstream congestion maximizes the mainstream outflow." However, the CTM-FR reduces the bottleneck flows due to congestion for the majority of the time period. The large peaks in the CTM-FR flows around $t = 400$ and $t = 1700$ represent those time periods where the effects of congestion are not great enough to reduce the bottleneck flows. INTEGRATION produces more complex, microscopic traffic behavior such as lane changing, vehicle dynamics and car-following are not modeled which accounts for the differences between INTEGRATION and CTM-FR at these locations.

Table 4.1: Comparison of NCT and TTT from the CTM, CTM-FR and INTEGRATION

| Model | TTT (seconds) | Difference | NCT (seconds) | Difference |
|---|---|---|---|---|
| CTM | 432,000 | -40.5% | 1,200 | -32.2% |
| CTM-FR | 672,423 | -7.3% | 1,770 | 0% |
| INTEGRATION | 725,633 | n/a | 1,770 | n/a |

By calibrating the CTM-FR to INTEGRATION, we used an $\Omega_i = 0.2075 Q_i$ to produce a solution having TTT and NCT values similar to those from INTEGRATION. Table 4.1 summarizes the NCT and TTT estimates from the CTM, calibrated CTM-FR and INTE-

GRATION. In this example, the CTM underestimates TTT by 40% and NCT by 32%. The value of $\Omega_i$ used should be determined based upon the input parameters, network configuration, and calibrated to appropriate empirical data or simulation model data.

## 4.3 Incorporating the CTM-FR into a Mathematical Program

In this section, we provide an overview of a linear programming implementation of the CTM (see Ziliaskopoulos, 2000), discuss the problem of "traffic holding", and introduce a mathematical program that incorporates the CTM-FR without traffic holding.

### 4.3.1 A Linear Program that Incorporates the CTM

The LP introduced in Ziliaskopoulos (2000) approximates the CTM using a system of linear inequalities. The notation used when modeling the CTM within a mathematical program is the same as introduced in Section 4.2. The following formulation minimizes the total travel time (TTT) given a demand profile for each origin and assuming a single destination.

$$\text{Minimize} \quad \sum_{t=1}^{T} \sum_{j \in R} \tau x_j^t \tag{4.4}$$

subject to:

$$x_j^t = x_j^{t-1} + \sum_{i:(i,j)\in L} y_{ij}^{t-1} - \sum_{k:(j,k)\in L} y_{jk}^{t-1}, \quad \forall\ j \in C,\ t = 2, \cdots, T \tag{4.5}$$

$$\sum_{j:(i,j)\in L} y_{ij}^t = d_i^t, \quad \forall\ i \in S_o,\ t = 1, \cdots, T \tag{4.6}$$

$$\sum_{k:(j,k)\in L} y_{jk}^t \leq x_j^t, \quad \forall j \in C \setminus S_i,\ t = 1, \cdots, T \tag{4.7}$$

$$\sum_{k:(j,k)\in L} y_{jk}^t \leq Q_j, \quad \forall j \in R,\ t = 1, \cdots, T \tag{4.8}$$

$$\sum_{i:(i,j)\in L} y_{ij}^t \leq Q_j, \quad \forall j \in R,\ t = 1, \cdots, T \tag{4.9}$$

$$\sum_{i:(i,j)\in L} y_{ij}^t \leq \delta_j [N_j - x_j^t], \quad \forall j \in R,\ t = 1, \cdots, T \tag{4.10}$$

$$x_j^t \geq 0, \quad \forall j \in C,\ t = 1, \cdots, T \tag{4.11}$$

$$y_{ij}^t \geq 0, \quad \forall (i,j) \in L,\ t = 1, \cdots, T. \tag{4.12}$$

Objective Function (4.4) minimizes the total travel time (TTT) across the network. Equivalently, we could also minimize network clearance time, but the direct formulation of this objective is more complex. Constraint (4.5) calculates cell density based on the cell density and flows from the previous time interval. The initial state of the network must also be specified; we set $x_i^1 = 0,\ \forall i \in (R \cup S_i)$, and $x_i^1 = \sum_{t=1}^{T} d_i^t, \forall i \in S_o$, which places all evacuees into their respective origins. Constraint (4.6) sets the flow out of each origin to be equal to $d_i^t$, which represents a loading curve.[1] Constraints (4.7) - (4.10), which approximate Equation (4.1), determine the flows over links, with appropriate summations for merging and diverging link sets. Constraint (4.7) states the traffic flowing out a cell must be less or equal to the number of vehicles in the cell. Constraints (4.8) and (4.9) enforce the flow capacity

[1] In Liu et al. (2006), an inequality constraint is used (instead of an equality constraint). In Bish et al. (2011b) and Chiu et al. (2007), this constraint is eliminated allowing the model full control of network loading.

limit. Constraint (4.10) models the backward propagation of congestion. Constraints (4.11) and (4.12) enforce the logical non-negativity restrictions.

### 4.3.2 Traffic Holding

The unrealistic traffic flows that linear programs often produce (Peeta and Ziliaskopoulos, 2001) results from *traffic holding*. To illustrate traffic holding, consider the flow $y_{ij}^t$ over ordinary link $(i, j)$. Equation (4.1) determines the value of $y_{ij}^t$, and as it is non-linear, it is approximated by the linear inequalities (4.7) - (4.10), one for each term in Equation (4.1). If the flow $y_{ij}^t$ does not satisfy one of the constraints (4.7) - (4.10) as an equality, then the LP is essentially "holding" traffic in a manner that contradicts the CTM and thus is not considered realistic. Ziliaskopoulos (2000) discusses traffic holding as a possible set of traffic controls that could be implemented to optimize flows. But Bish, Sherali, and Hobeika (2011b) shows that for an LP implementation of the CTM, there is always an optimal solution without traffic holding; traffic holding just provides an alternate optimal solution. We will see in the CTM-FR, however, that traffic holding when left unchecked, yields the optimal solution. We somewhat agree with Ziliaskopoulos (2000); deliberate traffic holding at merge and diverge intersections is possible with appropriate traffic controls measures (e.g., ramp metering on a freeway on-ramp). But for ordinary links, such as a freeway section having no on-ramps or off-ramps, these types of flow controls are not very realistic. Regardless, one must be careful using traffic holding to represent flow controls; this is a topic beyond the scope of this paper, but here we assume that traffic holding is appropriate for merge and diverge link sets.

Traffic holding in the CTM literature is briefly reviewed. Liu, Lai, and Chang (2006) mentions traffic holding as an unrealistic behavior for an evacuation. A common approach to model a nonlinear minimum function is to use a mixed integer program as performed in Lo (2001). However, such an approach is normally computationally very expensive and can prove to be inefficient for large network problems. Shen et al. (2007) suggests a post-processing algorithm to remove traffic holding from a solution generated by an SO LP. Lin and Wang (2004) proposes the use of an alternative linear objective function to eliminate traffic holding by forcing at least one of the inequality constraints that approximate Equation (4.1) to be tight. We will use a mixed integer programming approach to eliminate traffic holding from ordinary links and will use some additional constraints to reduce the computational effort required to solve the MIP. The suggestions made in Lin and Wang (2004) will not work for

the CTM-FR since they were intended for the CTM; the post processing algorithm suggested in Shen et al. (2007) is not advisable for the CTM-FR since we want to capture the impact of congestion within the output (sets of decisions) of the optimization model that incorporates the CTM-FR.

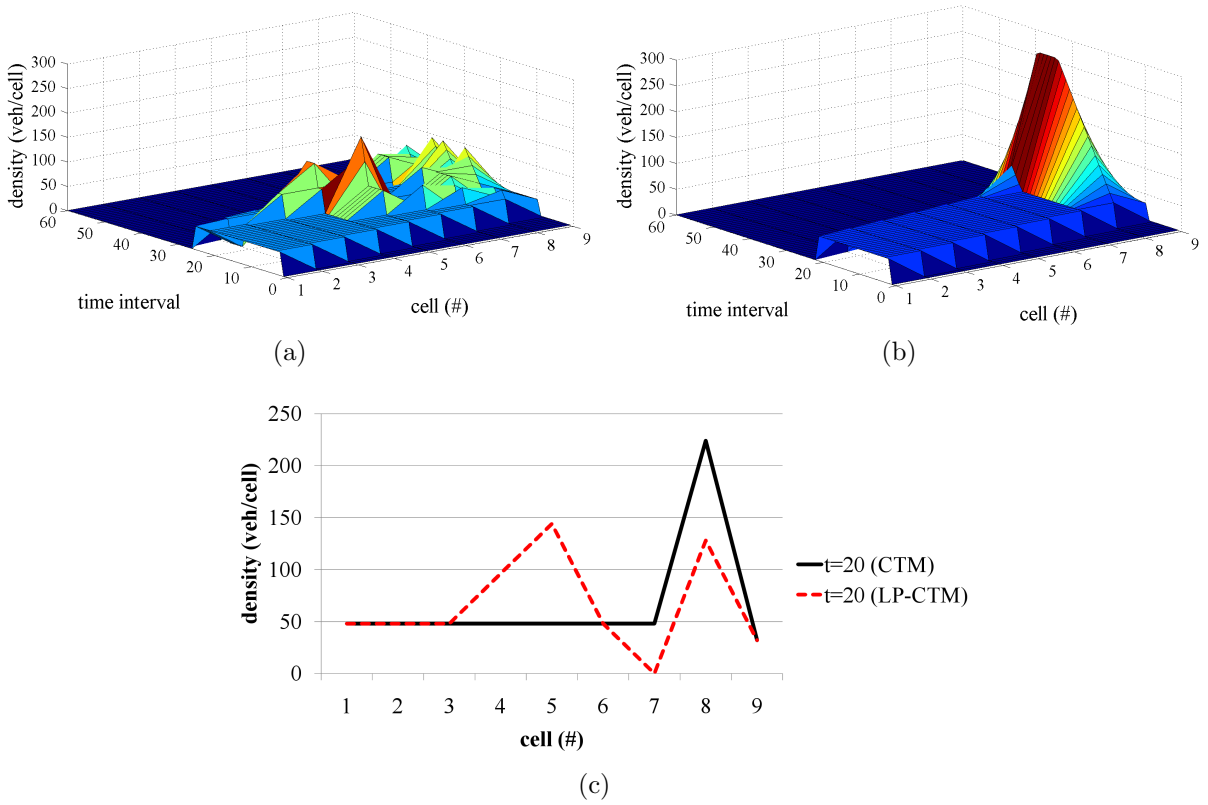



Figure 4.9: Plots of density as a function of location and time for the (a) LP and (b) CTM; (c) plots density-location for a specific time interval (t=20) for the LP and the CTM.

In order to illustrate the effects of traffic holding, we will use Example 1 from Section 4.2.1. Figure 4.9 graphically displays the traffic densities for the LP and the CTM solutions. As an orientation to the figures, the cells are numbered 1-9, where cell 9 is the bottleneck with 2-lanes (the rest of the cells have 3-lanes), the time intervals range from 0-60, and the cell densities range from 0-300 vehicles/cell. In Figure 4.9a, the LP solution has congestion forming at cell 8 and growing upstream, however, the density surface is irregular despite the regular flow into the system; this indicates the presence of traffic holding. In Figure 4.9b, traffic holding does not exist and the traffic flows are governed by Equation (4.1); here, the density within the cells far upstream of the bottleneck (cells 1-5) stays uniform at $x_i^t = Q_i = 48$; densities greater than this would indicate congestion. In contrast in Fig-

ure 4.9a, the density in cells 1-5 remains uniform for the first 20 intervals but afterwards fluctuates between $x_i^t = 48$ and $x_i^t = 100$ which shows congestion forming independent of the presence of a congestion shockwave and thus illustrates the traffic holding phenomenon. To enhance the discussion of the different behavior displayed in Figures 4.9a and 4.9b, Figure 4.9c compares the behavior of both the LP implementation of the CTM and the CTM for a specific time interval (t=20). While the sharp increase in cell density at cell 8 in the CTM is due to traffic congestion, the increases in density at cells 4 and 5 (and decrease at cell 7) in the LP are due to traffic holding. It is evident from these three figures that the LP is making decisions where there are no valid decisions to make; in this example, the decision variable $y_{ij}^t$ should solely depend on the relationships in Equation (4.1) for all links in the network.

However, the "no holding" results from the CTM in Figure 4.9b are more realistic than the results from the LP; these results represent how traffic would naturally flow and how congestion propagates. Furthermore, in this example, traffic holding does not improve the objective function - *the values for TTT and NCT from the LP and the CTM are identical.* The reason traffic holding does not impact TTT and NCT is that congestion in cell 8 does not reduce the flow from cell 8 to cell 9.

### 4.3.3 A Linear Program that Incorporates the CTM-FR

The logical first step in formulating the CTM-FR into a mathematical program is to simply replace Constraint (4.8) with Constraint (4.13) in the LP summarized in Section 4.3.1.

$$\sum_{j:(i,j)\in L} y_{ij}^t \leq x_i^t \frac{\Omega_i - Q_i}{N_i - Q_i} + N_i \frac{Q_i - \Omega_i}{N_i - Q_i} + \Omega_i, \quad \forall i \in R,\ t = 1, \cdots, T \tag{4.13}$$

However, this is often ineffective because the LP uses traffic holding to reduce the amount of congestion upstream of the bottleneck such that Constraint (4.13) does not limit flow. We demonstrate this using Example 1.

In Figure 4.10, the cell densities from the modified LP (with Constraint (4.13)) are compared to the cell densities calculated directly from the CTM-FR for an $\Omega_i = 0.2075 Q_i$. The NCT values for the solutions illustrated in Figures 4.9a, 4.9b, 4.10a are all equal to 40 time intervals. However, the CTM-FR solution in Figure 4.10b yields an increased NCT

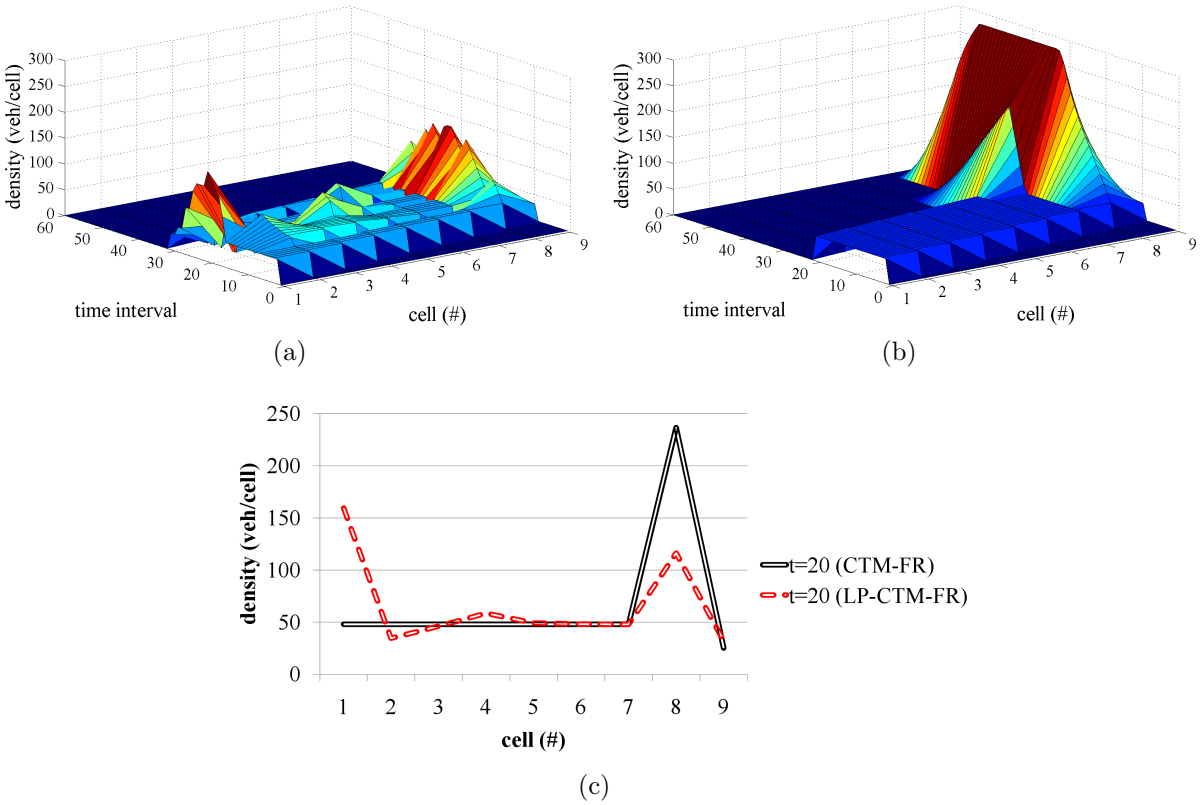



Figure 4.10: Plots of density as a function of location and time for (a) LP using the CTM-FR and (b) CTM-FR with $\Omega_i = 0.2075Q_i$; (c) plots density-location for a specific time interval (t=20) for the LP using the CTM-FR and the CTM-FR.

value of 60 time intervals. Thus, for the modified LP, traffic holding has more of an impact than just unrealistic modeling behavior between ordinary cells; it yields a superior objective function when compared to the non-holding solution by allowing the model to unrealistically avoid the effects of congestion and prevents the application of our flow reduction improvement to the CTM. As in Figure 4.9, Figure 4.10c is included to detail the difference between the cell densities in the CTM-FR and the LP implementation of the CTM-FR for a specific time interval (t=20). Traffic holding is evident in the LP implementation of the CTM-FR initially at cell 1 and at cell 4.

### 4.3.4 A Mixed Integer Linear Program that Incorporates the CTM-FR

In order to eliminate improper traffic holding, a mixed integer linear program, which we denote as MILP-FR, was developed. The MILP-FR eliminates traffic holding between ordinary cells and ensures that flow is governed by the CTM-FR (i.e., Equation (4.3)).

However, at merges and diverges, linear inequality constraints are used to model allowable traffic holding due to traffic metering or other traffic control measures. The constraints for merges and diverges are (4.7), (4.9), (4.10), and (4.13). These constraints can allow a limited amount of traffic holding but this holding will potentially cause congestion that limits flow.

The additional notation required for the MILP-FR are as follows:

Parameters:

$M$ : a large positive real number based on the cell and network input parameters; the term "$M$" is used only for notational simplicity

Variables:

$e1_i^{t\pm}$ : the difference in potential sending flow from cell $i$ between the first term in Equation (4.3) and $\min(Q_i, Q_j)$ for ordinary link $(i,j)$, $\forall (i,j) \in L_O,\ t = 1, \cdots, T$

$e2_i^{t\pm}$ : the difference in potential sending flow from cell $i$ between the second term in Equation (4.3) and $\min(Q_i, Q_j)$ for ordinary link $(i,j)$, $\forall (i,j) \in L_O,\ t = 1, \cdots, T$

$e3_j^{t\pm}$ : the difference in potential receiving flow between the first three terms and the last term in Equation (4.3) for ordinary link $(i,j)$, $\forall (i,j) \in L_O,\ t = 1, \cdots, T$

$z1_i^t$ : a binary variable that equals 1 over a range of densities in cell $i$ where $\min(Q_i, Q_j)$ dominates the first term in Equation (4.3) for ordinary link $(i,j)$, $\forall (i,j) \in L_O, t = 1, \cdots, T$

$z2_i^t$ : a binary variable that equals 1 over a range of densities in cell $i$ where the second term of Equation (4.3) dominates $\min(Q_i, Q_j)$ for ordinary link $(i,j)$, $\forall (i,j) \in L_O, t = 1, \cdots, T$

$z3_j^t$ : a binary variable that equals 1 over a range of densities in cell $j$ where the fourth term dominates all other terms in Equation (4.3) for ordinary link $(i,j)$, $\forall (i,j) \in L_O, t = 1, \cdots, T$

Figure 4.11 graphically illustrates the relationships between the new MILP-FR vari-

ables and Equation (4.3). Note, in order to simplify the figure, the time interval index $t$ has been removed from the time dependent variables and parameters. The flow out of cell $i$ is calculated by $y_i = x_i - e1_i^+ - e2_i^+$. The upper limit for $y_i$ is the minimum between $Q_i$ and $Q_j$; as pictured in the figure, $Q_j$ is the upper limit for $y_i$ in this case. In the figure, cell densities at either $x_i'$ or $x_i''$ will result in the $y_i$ as pictured. If the density in cell $j$ is $x_j'$, then $y_{ij} = y_i$. However, if the density in cell $j$ is $x_j''$, then $y_{ij} = y_i - e3_j^+$.

Next we introduce the constraints sets for ordinary links for the MILP-FR. In Subsection 4.3.4.1 we introduce some constraints that dramatically improve the MILP-FR run-time performance.

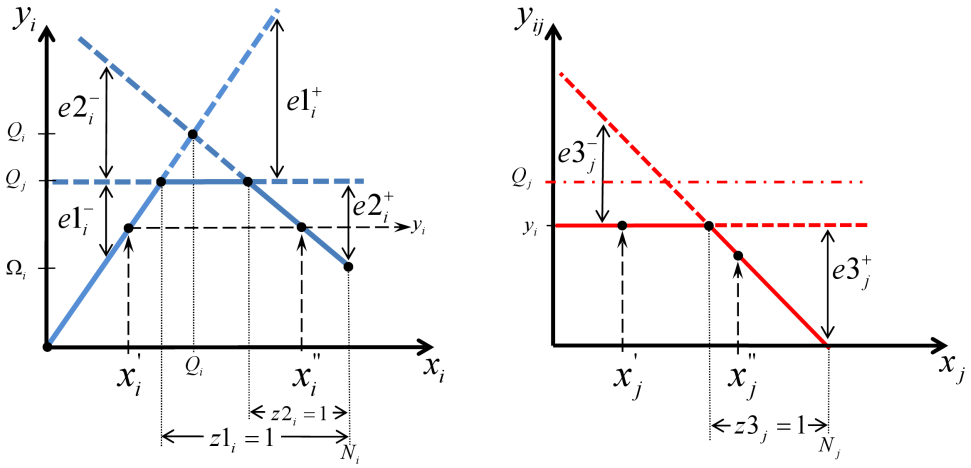



Figure 4.11: Diagram relating the MILP-FR and Equation (4.3) for flow along ordinary link $(i, j)$ with a capacity reduction where $y_i = \min\{y_{ij}^t(out), Q_j\}$

**Constraints for Ordinary Links** First we replace the inequality Constraints (4.7) - (4.10), and (4.13) with MILP constraints intended for flow on ordinary links. Refer to Figure 4.11, for guidance.

$$e1_i^{t+} \leq M(z1_i^t), \quad \forall (i,j) \in L_O, \ t = 1, \cdots, T \tag{4.14}$$

$$e1_i^{t-} \leq M(1 - z1_i^t), \quad \forall (i,j) \in L_O, \ t = 1, \cdots, T \tag{4.15}$$

$$e2_i^{t+} \leq M(z2_i^t), \quad \forall (i,j) \in L_O, \ t = 1, \cdots, T \tag{4.16}$$

$$e2_i^{t-} \leq M(1 - z2_i^t), \quad \forall (i,j) \in L_O, \ t = 1, \cdots, T \tag{4.17}$$

$$e3_j^{t+} \leq M(z3_j^t), \quad \forall (i,j) \in L_O, \ t = 1, \cdots, T \tag{4.18}$$

$$e3_j^{t-} \leq M(1 - z3_j^t), \quad \forall (i,j) \in L_O, \ t = 1, \cdots, T \tag{4.19}$$

$$x_i^t - \min(Q_i, Q_j) = e1_i^{t+} - e1_i^{t-}, \quad \forall (i,j) \in L_O, \ t = 1, \cdots, T \tag{4.20}$$

$$\min(Q_i, Q_j) - \left\{ x_i^t \frac{\Omega_i - Q_i}{N_i - Q_i} + N_i \frac{Q_i - \Omega_i}{N_i - Q_i} + \Omega_i \right\} = e2_i^{t+} - e2_i^{t-}, \quad \forall (i,j) \in L_O, \ t = 1, \cdots, T \tag{4.21}$$

$$(x_i^t - e1_i^{t+} - e2_i^{t+}) - \delta_j (N_j - x_j^t) = e3_j^{t+} - e3_j^{t-}, \quad \forall (i,j) \in L_O, \ t = 1, \cdots, T \tag{4.22}$$

$$y_{ij}^t = x_i^t - e1_i^{t+} - e2_i^{t+} - e3_j^{t+}, \quad \forall (i,j) \in L_O, \ t = 1, \cdots, T \tag{4.23}$$

$$e1_i^{t\pm} \geq 0, \quad z1_i^t \in \{0,1\}, \quad \forall (i,j) \in L_O, \ t = 1, \cdots, T \tag{4.24}$$

$$e2_i^{t\pm} \geq 0, \quad z2_i^t \in \{0,1\}, \quad \forall (i,j) \in L_O, \ t = 1, \cdots, T \tag{4.25}$$

$$e3_j^{t\pm} \geq 0, \quad z3_j^t \in \{0,1\}, \quad \forall (i,j) \in L_O, \ t = 1, \cdots, T. \tag{4.26}$$

Constraints (4.14) - (4.19) allow only one variable in each $e1, e2, e3$ pair to have a positive value. For instance, if $z1_i^t = 1$, then $e1_i^{t+} \geq 0$ and $e1_i^{t-} = 0$, otherwise $e1_i^{t+} = 0$ and $e1_i^{t-} \geq 0$. The M-parameter is used for notational simplicity and can be replaced by a parameter calculated for each constraint to improve run-time efficiency. For example, the $M$ used for $e1_i^{t+}$ is $N_i - \min(Q_i, Q_j)$ and for $e2_i^{t-}$ is $N_i \dfrac{Q_i - \Omega_i}{N_i - Q_i} + \Omega_i - \min(Q_i, Q_j)$ since these are the maximum values for each difference variable. Constraints (4.20) - (4.22) ensure that the flow between ordinary cells is governed by Equation (4.3), thus eliminating traffic holding, and determine the values for each difference variable $e1_i^{t\pm}, e2_i^{t\pm}, e3_j^{t\pm}$. The flows between cells are then calculated by Constraint (4.23). Note that Constraint (4.21) represents the modified sending cell constraint with the $\Omega_i$-parameter.

#### 4.3.4.1 MILP-FR Performance Improvements

The MILP-FR formulation adds additional constraints and decision variables to the LP formulation introduced in Section 4.3.1. Although our MILP-FR models network flows more accurately, the computational effort required is much greater than the LP. We have developed some constraints for the MILP-FR that take advantage of the properties of congestion to significantly decrease the computational effort required to solve the MILP.

**Proposition 1.** *For a link* $(i,j) \in L_O$*, if* $Q_i = Q_j$ *then* $z1_i^t = z2_i^t$.

**Proof:** By Constraint (4.20) and $Q_i = Q_j$ we have $e1_i^{t+} \geq 0$ when $x_i^t \geq Q_i$, thus requiring $z1_i^t = 1$ by Constraint (4.14). By Constraint (4.21) and $Q_i = Q_j$ we have $e2_i^{t+} \geq 0$ when $x_i^t \geq Q_i$, thus requiring $z2_i^t = 1$. Hence when $x_i^t \geq Q_i$, $z1_i^t = z2_i^t = 1$. □

This is essentially acknowledging that for ordinary links between cells having equal flow limits, fewer calculations are required to solve the MILP-FR by equating $z1_i^t$ and $z2_i^t$.

**Proposition 2.** *For a link* $(i,j) \in L_O$*, if* $N_i > Q_i + Q_i/\delta_i$ *then* $z1_i^t \geq z3_i^t$.

**Proof:** For cell $i$, $z3_i^t = 1$ when $\delta_i(N_i - x_i^t)$ determines the flow into $i$, i.e., when this term defines the min in Equation (4.3). The *largest possible range* of $x_i^t$-values for which $z3_i^t = 1$ is achieved when the flow into cell $i$ is $Q_i$ because of the negative slope of the $\delta_i(N_i - x_i^t)$ flow constraint. Thus, the largest range of $x_i^t$-values for which $z3_i^t = 1$ is $N_i - Q_i/\delta_i < x_i^t \leq N_i$. The *smallest possible range* of $x_i^t$-values for which $z1_i^t = 1$ is achieved when $Q_i \leq Q_j$, and this range extends from $Q_i < x_i^t \leq N_i$. Hence, the smallest possible range for which $z1_i^t = 1$ contains the largest possible range for which $z3_i^t = 1$ when $Q_i < N_i - Q_i/\delta_i$ or alternatively when $N_i > Q_i + Q_i/\delta_i$. Thus, when $N_i > Q_i + Q_i/\delta_i$ we have $z1_i^t \geq z3_i^t$. □

As mentioned earlier, a reasonable relationship between the values of $Q_i$ and $N_i$ is given by $1/6 \leq Q_i/(N_i - Q_i) \leq 1/4$ (Dixit et al., 2008; Kalafatas and Peeta, 2006), or $5Q_i \leq N_i \leq 7Q_i$. Since Proposition 2 proves that $z1_i^t \geq z3_i^t$ for $N_i > Q_i + Q_i/\delta_i$, clearly this will also hold true for realistic values of $N_i \geq 5Q_i$.

**Proposition 3** *For a link* $(i,j) \in L_O$, *if* $Q_i = Q_j$ *then* $z1_i^t \leq z3_j^t$.

**Proof:** Based on Property 1, the traffic density in the receiving cell $j$ will limit flow into cell $j$ before the traffic density in cell $i$ will limit the flow out of cell $i$. Since $z1_i^t = 1$ represents the range of traffic densities where the second term of Equation (4.3) would limit flow out of cell $i$ and $z3_j^t = 1$ represents the range of traffic densities where term four limits the flow into cell $j$, then the inequality $z1_i^t \leq z3_j^t$ must hold for a given time $t$. □

The inequality constraint must be used (instead of an equality constraint) in order to account for those instances during the formation and clearing of congestion when congestion exists in the receiving cell but has not formed in the sending cell.

These three propositions reduce the number of linear relaxations required to calculate the binary variables while solving the MILP-FR formulation. The key idea is that the binary variables within the network are *not* independent. Each binary variable is linked to upstream and downstream binary variables within the network. For identical ordinary cells, congestion builds upstream in a backward forming wave and clears downstream in a forward recovery wave.

When Propositions 1-3 are applied as constraints in the MILP-FR, they reduce the solution time considerably. This will be demonstrated in the next section [Section 4.4].

## 4.4 Example 2: Freeway Merge Bottleneck

In order to demonstrate the MILP-FR, we use an example based on a 2-lane section of freeway on-ramp intersection where a one lane on-ramp joins. Using this example, we compare the results from the MILP-FR with those from the LP introduced in Section 4.3.1 and the INTEGRATION simulation software.

In order to evaluate the results from MILP-FR, we tested values for $\Omega_i$ ranging from $0.90Q_i$ to $0.20Q_i$ in increments of $0.10Q_i$ (where $0.20Q_i$ best matches the results from INTEGRATION). At each value of $\Omega_i$, the results from the MILP-FR are compared to the results from LP in terms of two important evacuation metrics, TTT and NCT. The MILP-FR was coded in IBM ILOG Optimization Programming Language (OPL) Integrated Development Environment (IDE) and used CPLEX 12.1.0. All runs were made on a computer having

2 Intel 2.13 GHz Xeon [quad-core] processors and 4 GB of RAM using a 64-bit Windows 7 operating system. Each instance of the model considers 39,978 constraints and 38,988 variables (10,395 binary variables).

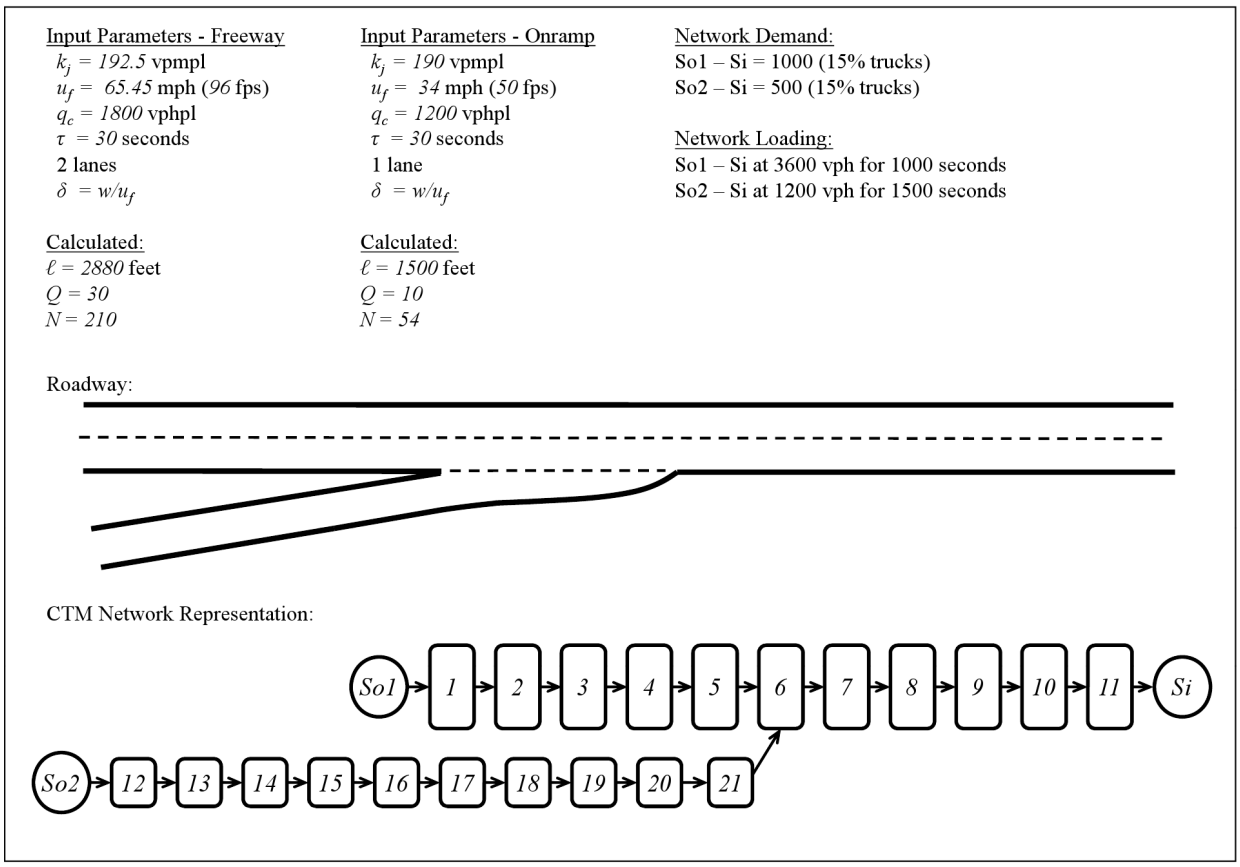
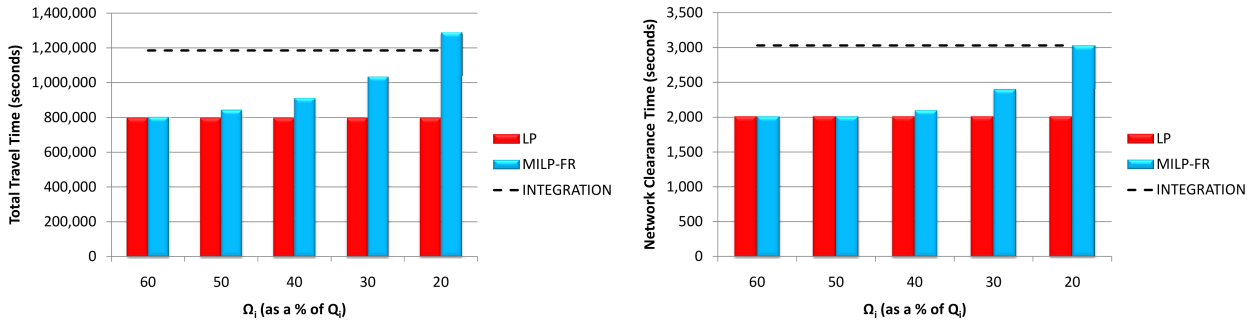



Figure 4.12: Example 2: Freeway Merge Bottleneck.

Figure 4.12 displays the example network with the input parameters used. In this example, the MILP-FR has the ability to make some decisions at the merge intersection in order to find the optimal solution; the model can adjust the merge priorities for both the freeway and on-ramp traffic flows in order to maximize flow through the network. Additionally, traffic holding is limited to just the cells upstream of the merge (cells 5 and 21 in this example); traffic holding has been eliminated from the ordinary links both upstream and downstream of this intersection.

The MILP-FR results are displayed in Figure 4.13 and are compared to the LP and INTEGRATION estimates for TTT and NCT. Note that in Section 4.2.1, we compared




(a) (b)

Figure 4.13: Comparison of (a) TTT and (b) NCT estimates from the LP, MILP-FR, and INTEGRATION for Example 2.

the estimates from the CTM, CTM-FR, and INTEGRATION; Figure 4.13 compares the estimates given by the two mathematical programs that *incorporate* the CTM and CTM-FR with INTEGRATION. In this example, it is clear that the $\Omega_i$-parameter should be in the range of $0.20Q_i$-$0.30Q_i$.

These results indicate the effectiveness of the MILP-FR and the CTM-FR improvement over the CTM. By successively decreasing the $\Omega_i$-parameter, we can increase our predicted TTT and NCT values to approximate the results from INTEGRATION. As can be seen in Figure 4.13, the LP underestimates TTT and NCT drastically when compared to INTEGRATION.

The computational performance of the MILP-FR is shown in Table 4.2. Here we compare the different solution times for freeway on-ramp example using different values of $\Omega_i$ for the MILP-FR *with and without* the constraints from Propositions 1 - 3. We set 3 hours (or 10,800 seconds) as an upper limit on the solution time. In addition, the objective function value (representing TTT) is given for the MILP-FR and the LP for each value of $\Omega_i$.

Table 4.2: Comparison of Solution Times with and without constraints from Propositions 1-3

| | Solve Time (seconds) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\Omega_i$ (as a proportion of $Q_i$) | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 |
| MILP-FR | 465.69 | > 10,800 | 733.99 | > 10,800 | > 10,800 | >10,800 | >10,800 | >10,800 |
| MILP-FR with Propositions 1-3 | 4.46 | 4.03 | 3.78 | 26.05 | 30.79 | 89.42 | 321.34 | 1,038.49 |
| MILP-FR TTT value | 795,600 | 795,600 | 795,600 | 802,445 | 843,119 | 911,837 | 1,035,415 | 1,290,067 |
| LP TTT value | 795,600 | 795,600 | 795,600 | 795,600 | 795,600 | 795,600 | 795,600 | 795,600 |

As can be seen in Table 4.2, the solution times for the MILP-FR exceed 3 hours for $\Omega_i$ values less than $0.70Q_i$. The solution times for the MILP-FR with constraints from Propositions 1-3 are dramatically lower and the solution times increase as $\Omega_i$ decreases. Additionally, the MILP-FR objective function value increases, as expected, with decreasing values of $\Omega_i$; at $0.20Q_i$, the MILP-FR yields a TTT value that overestimates the INTEGRATION TTT value of 1,184,822 by 9% as compared to the LP that underestimates the TTT by 33%.

## 4.5 Conclusions

The CTM-FR, and the MILP-FR that incorporates it, predicts more accurate and realistic evacuation times and will improve evacuation modeling and planning. A significant research contribution is the development the CTM-FR that reduces the sending flow capacity of the cell with respect to increasing traffic density. This reduces the overall flow between cells based on the effects of congestion upstream of a bottleneck(s). Since the CTM is commonly used in evacuation planning research, the CTM-FR improvement should enhance all of the recent work and research in this field. An additional research contribution is the formulation of a mixed integer linear program (MILP) to eliminate traffic holding among ordinary cells and successfully implement the CTM-FR. This will have a large impact when optimizing components of an evacuation plan such as network loading and routing because the MILP-FR will attempt to avoid congestion. This is not the case with the LP since the presence of congestion will maximize flow through the network. Lastly, we proposed three constraints for the MILP-FR that dramatically reduce the computation effort required to solve the mixed integer linear program.

Future research will use the MILP-FR formulation to examine various different existing and novel evacuation demand management strategies. The intent of this future research will be to use a network optimization model to determine optimal evacuation plans and policies that will perform better than existing evacuation plans in practice today. The development and use of the MILP-FR, when included in a larger evacuation management strategy formulation, will allow network planners to better avoid crippling congestion that can both frustrate and endanger evacuees during the conduct of an evacuation. It is expected that when using the MILP-FR with other constraint sets and additional decision variables for a particular evacuation strategy, that the computational performance will worsen; future research will build upon the computational benefits of Propositions 1 - 3 and define more

relationships between the existing and new binary variables to further decrease the computational effort required. Another research direction would be to exploit any special structures of the mathematical program (that includes the MILP-FR) and apply a decomposition strategy to solve the program more efficiently. Additionally, we will model the evacuation policy decisions generated by the network optimization model within the INTEGRATION simulation software in order to better evaluate the effectiveness of the decisions produced.

# Chapter 5

# Reducing Congestion during an Evacuation: Optimal Demand Management Strategies for Evacuation Planning

**Abstract:** Congestion negatively impacts the evacuation process in three main ways: 1) congestion slows the evacuation and can potentially expose evacuees to the very threat causing the evacuation; 2) congestion places evacuees in stressful and potentially life threatening traffic for many hours; and 3) congestion decreases compliance with local and regional evacuation orders. This paper presents a network flow optimization model, called the Flow Reduction Model (FRM), that uses an evacuation demand management strategy to improve the performance of an evacuation plan. The model captures the effects of congestion by reducing network flows at the onset of congestion; the model penalizes network flows when congestion is present and therefore drives the model to manage demand in order to *avoid* congestion. The demand management strategy used is termed *group-level staging* – a strategy that evacuates different group and different time intervals in order to reduce congestion and evacuation duration. The effects of an unrealistic modeling phenomenon called "traffic holding" are explored within the FRM; this occurs when a mathematical model used for dynamic traffic assignment artificially holds back traffic in order to achieve an optimal solution. Additionally, different objective functions are evaluated within the FRM and conclusions are offered concerning the use of group-level staging to form an evacuation plan.

## 5.1 Introduction

The images of incredibly congested highways and freeways have become common during large-scale emergency evacuations. Over the past few decades, the Atlantic and Gulf coasts of the United States have been subject to frequent and costly hurricane evacuation events where traffic congestion has contributed to both the danger and loss. Congestion negatively impacts the evacuation process in three main ways: 1) congestion slows the evacuation and can potentially expose evacuees to the very threat causing the evacuation; 2) congestion places evacuees in stressful and potentially life threatening traffic for many hours; and 3) congestion decreases compliance with local and regional evacuation orders.

There is a large need to develop plans and procedures to manage the immense demand on our critical freeways that lead away from the impending hazard. Although creating additional supply by building new interstate systems or additional freeway lanes or by reversing flow through contra-flow will certainly assist in addressing the problem of congestion during an evacuation, it is infeasible (if not impossible) that enough supply could ever be constructed to meet the peak demand during an evacuation due to economic and geographic constraints. There have been many lessons learned from the Hurricane Katrina and Rita events; among them are recommendations to clearly communicate the hazard, when to order an evacuation, and how to prioritize different populations in order to address the problem of congestion and to improve evacuation safety (Litman, 2006).

The main contributions of this paper are to present a network flow optimization framework that models congestion realistically and produces results that are implementable. The results from this model uses a demand management strategy that will better manage evacuation demand within the physical constraints of a current transportation network. We use a strategy where different geographic regions, or groups, are given separate and distinct priorities for evacuation in order to reduce congestion and evacuate the entire transportation network sooner – called *group-level staging*. This is in contrast to a simultaneous evacuation where a single evacuation order is issued across a large region or metropolitan area. It is also in contrast to household-level staging where each household is issued a separate evacuation order; this method would drastically reduce congestion (assuming a high compliance rate) but would be almost impossible to implement with *existing* communication and transportation infrastructure. There are many factors that effect how people respond to emergency events from a behavioral analysis perspective (Alsnih and Stopher, 2004) but that is not the focus

of this research or paper; will focus network modeling based on an assumed set of evacuee behaviors.

The context for the optimization model and demand management strategies presented in this paper is a short-notice emergency evacuation (such as an evacuation in advance of a hurricane) of a region by automobile. However, the optimization model could also be adapted for a no-notice evacuation, such as one in and around a nuclear reactor.

This paper is structured as follows. Section 5.2 reviews some of the current literature in the area of evacuation planning, modeling, and simulation. Section 5.3 introduces a mixed-integer linear program (MILP) that models network flows while reducing the network capacity in the presence of congestion, as presented in Bish, Chamberlayne, and Rakha (2011a), called the Flow Reduction Model (FRM). Section 5.4 reviews some commonly used objective functions that have been used to optimize network flows during an evacuation and will discuss how these different objective functions behave within the FRM framework. Section 5.5 uses the FRM staging decision variable to refine how demand is staged and loaded onto the network to improve the performance of an evacuation plan. Section 5.6 modifies the FRM formulation to remove traffic holding from key intersections and to model flow priorities explicitly. Section 5.7 adds a constraint to the FRM to both improve solution effort and to dictate evacuation plan performance. Section 5.8 concludes the paper with a demonstration of the benefits of an evacuation plan using group-level staging over a simultaneous evacuation. Lastly, Section 5.9 presents some of the key conclusions from the research and suggests some future research directions.

## 5.2 Background

Efficient emergency evacuation planning has recently been a very active research area. The research is concentrated in two main areas: traffic simulation and network flow optimization. The simulation research has the advantages of microscopic traffic flow modeling and improved realism over macroscopic modeling, however, simulation alone will not produce an optimal evacuation plan. In contrast, network flow optimization approaches can directly yield an optimal plan for an evacuation but may lack some of the realism of microscopic traffic simulation.

Within the simulation approaches to evacuation planning, Mitchell and Radwan

(2006) uses a microscopic traffic simulator, INTEGRATION, to evaluate various different staging strategies for a generic grid network. Under high demand conditions, a few staging strategies are shown to be superior to a simultaneous evacuation strategy. Although the research suggests that staging network demand can provide advantages over simultaneous evacuation, the work represents experimentation essentially and does not provide any insight on how to determine the best staging strategy. Similarly, in Chen and Zhan (2008), a microscopic, agent-based, traffic simulator is used (Paramics) to evaluate an enumeration of different staging strategies for three different network types: grid, ring-road, and real road network. The authors find the same conclusion as in Mitchell and Radwan (2006); under high network work demand, staged strategies perform better than simultaneous evacuation.

In Sbayti and Mahmassani (2006), a simulation-optimization approach is used to combine the benefits of the active areas of research. The paper presents an improved strategy to stagger the departure of evacuees from each origin (called "staging") in order to minimize TET. The authors compared the staged results to those from simultaneous evacuation (or where all evacuees depart origins at the same time). Under a high demand scenario, the staged results reduced the network clearance time (NCT) by 61%, TTT by 62%, and average travel time (ATT) by 65%. Based on this result, the authors conclude that there are significant benefits by staging evacuation demand over a simultaneous evacuation procedure. Although the results are sound, the approach made by the authors assumed a high level of control; the approach does not use a fixed demand loading curve but uses decision variables that control how demand is loaded on the network at each time interval. The approach also assumes that 100% compliance with the evacuation directives. Both of these factors guarantee that the traffic conditions remain in the free-flow regime and avoids congestion. Lastly, the authors use an objective function that minimizes TET but state that the objective function also minimizes NCT and ATT as well.

Much of the research in the area of network flow optimization is based upon a macroscopic traffic flow model, the Cell Transmission Model (CTM) introduced in Daganzo (1994, 1995), and a linear program that incorporates the CTM used for dynamic traffic assignment introduced in Ziliaskopoulos (2000). These papers were reviewed in more detail in Bish et al. (2011a).

In the area of demand management, the network flow optimization models normally concentrate on disaggregate-level or aggregate-level staging strategies. For the purposes of this paper, we will use the terms household-level and group-level staging. We will first review

some papers that use household-level staging.

In Chiu et al. (2007), a dynamic network flow model, based on the CTM and the linear program developed in Ziliaskopoulos (2000), is used to minimize total evacuation time (TET) through a "demand specification method". This model is applied to a no-notice evacuation as opposed to a short-notice evacuation. The method loads all demand at each origin and then determines the optimal flow of traffic out of each origin with respect to the objective function. This approach essentially creates a household-level staging and routing plan within the study area, which would have some significant implementation issues. The formulation allows a problem known as "traffic-holding" where the demand is held back within the network artificially in order to achieve an optimal objective value (see Peeta and Ziliaskopoulos (2001); Bish et al. (2011a) for a more detailed discussion of traffic holding). In Yao et al. (2009), the LP developed in Ziliaskopoulos (2000) is adapted and a robust optimization formulation is presented for use in evacuation planning where network demand is stochastic. The model is demonstrated by using the network introduced in Chiu et al. (2007).

Household-level staging and routing is also explored in Bish, Sherali, and Hobeika (2011b). This paper introduces an evacuation model, based on the CTM, that develops an evacuation plan while considering shelters with capacity restrictions. An alternate formulation is presented with an efficient structure that is easier to solve for large scale networks. The impacts on various objective functions are explored based on the use of the CTM with household-level staging and routing strategies.

Very few papers have explored the use of group-level staging and routing. In Liu et al. (2006), a MILP, also based on the CTM, is proposed to minimize a weighted total evacuation time in order to generate an optimal group-level staging strategy. This approach, in contrast to Chiu et al. (2007), uses fixed demand loading curves at each origin and determines the time interval when each fixed loading curve flows traffic out of each origin. This results in origin specific staging plans, instead of household-level staging plans, which is easier to implement and execute during an evacuation. The formulation uses a rather complicated system of artificial waiting cells to determine optimal staging time intervals in order to reduce congestion within the network. The formulation also allows the traffic holding phenomenon, as in Chiu et al. (2007).

In Bish and Sherali (2011), a MILP formulation based on the CTM is proposed that

generates optimal group-level staging and routing decisions to minimize a novel lexicographic objective function; the function minimizes NCT, the sum of evacuation route lengths, and average evacuation time. The formulation uses fixed loading curves, as in Liu et al. (2006), at each origin and limits the traffic flow to the free-flow regime. Some of the modeling extensions introduced are the ability to assign destination preferences for each origin and to give certain origins priority in the staging strategy. Two heuristics are developed to produce near-optimal solutions quickly to the MILP formulation. The paper demonstrates that minimizing evacuation time and network clearance time are not equivalent when using fixed loading curves at the origins.

In regard to network demand loading curves for use in network flow optimization models, a lot of study has been dedicated to the factors that affect evacuation response and behavior. An excellent table summarizing a wide variety of factors (to include age, risk perception, type of home, and pet ownership) discussed in the current body of literature is presented in Yazici and Ozbay (2008). The results of any network optimization model are highly dependent on the network demand loading curve used. A sequential logit model (SLM) is proposed in Fu and Wilmot (2004); Fu et al. (2007) that accounts for a variety of evacuee factors in determining a probability of if and when an evacuee will depart a danger zone. The SLM appears to yield promising results when compared to actual evacuation response curves (Hurricane Floyd - 1999, Hurricane Andrew - 1992). However, many evacuation modeling studies (Yazici and Ozbay, 2008; Liu et al., 2006; Mitchell and Radwan, 2006) use simpler sigmoid curves (based on the logistic function) to approximate the evacuation response as a demand loading curve.

The study of the effects of traffic congestion is important to understand and model correctly. A key problem of the CTM, is that it fails to capture the reduction in the queue discharge after the onset of congestion. Many of the network flow optimization models that depend on the CTM as the internal traffic model fail to develop strategies that avoid or reduce congestion for this reason. Congestion upstream of a bottleneck reduces the queue discharge flows between 10 to 20% and has been extensively shown in various empirical studies (Banks, 1990; Hall and Agyemang-Duah, 1991; Cassidy and Bertini, 1999; Chung, Rudjanakanoknad, and Cassidy, 2007) and in traffic simulations (Chamberlayne, Rakha, El-Metwally, and Bish, 2011b; Chamberlayne, Rakha, and Bish, 2011a).

## 5.3 Overview of the Flow Reduction Model

The Flow Reduction Model (FRM) was introduced in Bish et al. (2011a). The principal motivations in the development of the model were to include the effects of congestion on network flows (i.e., congestion reduces flow) and to remove traffic holding from those locations where traffic control measures cannot be implemented. This first motivation led to the development of a cell-based model similar to the CTM (called the Cell Transmission Model with Flow Reductions [CTM-FR]), but where the traffic density in a cell can reduce the cell's flow potential, i.e., congestion.

The CTM is a macroscopic traffic flow model that uses a discrete time-expanded network of cells and links $(C, L)$ to represent a roadway system. Within the CTM, flows travel over *merging links* from multiple cells and into a single receiving cell whereas flow travels from a single cell over *diverging links* into multiple receiving cells. Flow travels from a single sending cell to a single receiving cell of an *ordinary link*. The CTM-FR modifies the *sending cell* of the CTM to send reduced flows to the *receiving cell* once the sending cell has entered the congested regime. See Figure 5.1 and Equation 5.1.

$$y_{ij}^t = \min\{x_i^t, (N_i - x_i^t)\frac{Q_i - \Omega_i}{N_i - Q_i} + \Omega_i, Q_j, \delta_j[N_j - x_j^t]\} \tag{5.1}$$

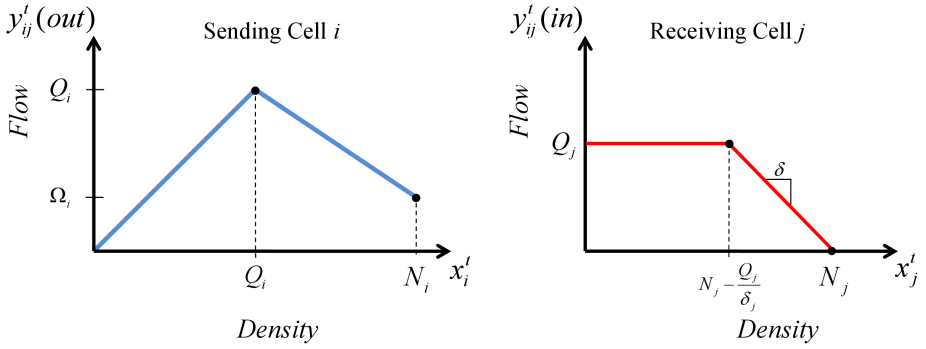



Figure 5.1: The flow-density relationship for the CTM-FR for flows over link $(i, j)$.

The $\Omega_i$-parameter is a measure of the reduction of flow at jam density. Currently the CTM assumes a $\Omega_i = Q_i$, the maximum flow possible at jam density. This makes the CTM-FR a more general form of the CTM that can model a range of different models depending on the level of importance given to congestion to reduce flow.

### 5.3.1 Notation and Definitions

The notation used in this paper is as follows:

Sets:

| | | |
|---|---|---|
| $C$ | : | set of cells within the network; subsets are sources ($S_o$), roadway cells ($R$), and sinks ($S_i$) |
| $noH$ | : | a subset of the $R$ where traffic holding is not allowed or realistic |
| $L$ | : | set of links between cells within the network; subsets are ordinary links ($L_O$), merging links ($L_M$), and diverging links ($L_D$) |

Parameters:

| | | |
|---|---|---|
| $T$ | : | number of time intervals of length $\tau$ in the planning horizon |
| $F_i$ | : | reduced planning horizon used for staging; [$F_i = T - W_i -$ min(time to travel between origin $i$ and each destination $j$)] |
| $Q_i$ | : | the maximum flow into or out of cell $i$, which is based on the saturation flow rate ($q_c$), the length of time interval ($\tau$), and the number of lanes in cell $i$ ($n_i$) [$Q_i = q_c \times \tau \times n_i$], $\forall i \in R$ |
| $N_i$ | : | the maximum number of vehicles that cell $i$ can hold, which is based on the cell length ($\ell_i$), number of lanes in cell $i$ ($n_i$), and jam density ($k_i^{jam}$) [$N_i = k_i^{jam} \times \ell_i \times n_i$], $\forall i \in R$ |
| $\delta_i$ | : | is the ratio of wave speed to free-flow speed for cell $i$: $w/u_f$, $\forall i \in R$ |
| $d_i^t$ | : | number of vehicles that flow out of the source $i$ in time interval $t$ according to a given loading curve, $\forall i \in S_o$, $t = 1, \cdots, T$ |
| $D_i$ | : | total demand or population at origin $i$, $\forall i \in S_o$ |
| $D$ | : | total network demand; $D = \sum_{i \in S_o} D_i$ |
| $\Omega_i$ | : | the flow rate out of cell $i$ at the maximum traffic density, $N_i$, $\Omega_i \in (0, Q_i]$ |
| $W_i$ | : | time duration of the demand loading curve for a given origin $i$ |
| $\alpha_i, \beta_i$ | : | loading curve response parameters for a given origin $i$ |

Variables:

| | | |
|---|---|---|
| $x_i^t$ | : | number of vehicles in cell $i$ at the beginning of time interval $t$, $\forall i \in C$, $t = 1, \cdots, T$ |
| $y_{ij}^t$ | : | number of vehicles that flow from cell $i$ to cell $j$ over link $(i, j)$ in time interval $t$, $\forall (i, j) \in L$, $t = 1, \cdots, T$ |
| $E^t$ | : | a binary decision variable that equals 1 when traffic is traveling on the network at time interval $t$ and 0 otherwise, $t = 1, \cdots, T$ |
| $\theta_i^t$ | : | a binary decision variable that equals 1 when the specific loading curve for origin $i$ is loaded onto the network at time interval $t$, $\forall i \in S_o, t = 1, \cdots, F$ |

$e1_i^{t\pm}$ : the difference between the first and second term in Equation (5.1), $\forall (i,j) \in L_O,\ t = 1, \cdots, T$

$e2_i^{t\pm}$ : the difference in potential sending flow, $y_i^{out}$, from cell $i$ and the potential receiving flow, $y_j^{in}$, into cell $j$, $\forall (i,j) \in L_O,\ t = 1, \cdots, T$

$e3_j^{t\pm}$ : the difference between the third and fourth term in Equation (5.1), $\forall (i,j) \in L_O,\ t = 1, \cdots, T$

$z1_i^t$ : a binary variable that equals 1 over a range of densities in cell $i$ where the second term dominates the first term in Equation (5.1) (or where $x_i^t \geq Q_i$), $\forall (i,j) \in L_O, t = 1, \cdots, T$

$z2_i^t$ : a binary variable that equals 1 when the potential sending flow exceeds the potential receiving flow (or when $y_i^{out} > y_j^{in}$), $\forall (i,j) \in L_O, t = 1, \cdots, T$

$z3_j^t$ : a binary variable that equals 1 over a range of densities in cell $j$ where the fourth term dominates the third term in Equation (5.1) (or where $x_j^t \geq N_j - Q_j/\delta_j$), $\forall (i,j) \in L_O, t = 1, \cdots, T$

### 5.3.2 The FRM Formulation

After the development and testing of the CTM-FR (Bish et al., 2011a), the traffic flow model was incorporated into an MILP, instead of an LP, in order to restrict the "traffic holding" problem to where it could be replicated by realistic traffic control measures. As discussed in Section 5.2, many papers acknowledge the existence of this problem but make some assumptions (to ignore the issue and) to keep their respective models as linear programs in order to take advantage of efficiency and calculation speed.

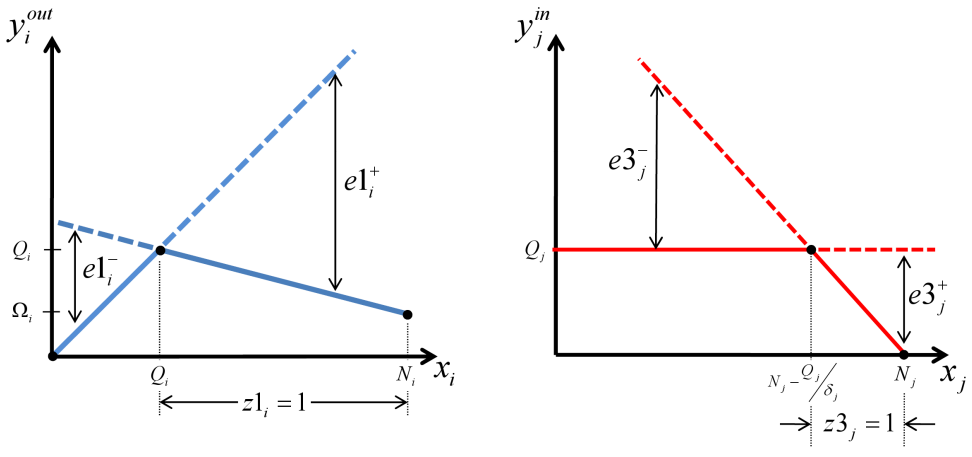



Figure 5.2: Diagram relating the FRM and Equation (5.1) for flow along ordinary link $(i,j)$

We developed a set of binary decision variables ($z1_i^t, z2_i^t, z3_i^t \in noH$) and a set of

difference variables ($e1_i^{t\pm}, e2_i^{t\pm}, e3_i^{t\pm}$) in order to restrict traffic holding to merge and diverge intersections. We also later used these variables to induce certain traffic behaviors into our demand management strategies. See Figure 5.2 for the relationship of these variables and the notation of the FRM and Bish et al. (2011a) for further explanation. In the following constraint set, we introduce the FRM and some demand management constraints for an emergency evacuation context.

$$x_j^t = x_j^{t-1} + \sum_{i:(i,j)\in L} y_{ij}^{t-1} - \sum_{k:(j,k)\in L} y_{jk}^{t-1}, \quad \forall\ j \in C,\ t = 2, \cdots, T \tag{5.2}$$

$$\sum_{j:(i,j)\in L} y_{ij}^t \leq x_i^t, \quad \forall i \in R,\ t = 1, \cdots, T \tag{5.3}$$

$$\sum_{j:(i,j)\in L} y_{ij}^t \leq (N_i - x_i^t)\frac{Q_i - \Omega_i}{N_i - Q_i} + \Omega_i, \quad \forall i \in R,\ t = 1, \cdots, T \tag{5.4}$$

$$\sum_{i:(i,j)\in L} y_{ij}^t \leq Q_j, \quad \forall j \in R,\ t = 1, \cdots, T \tag{5.5}$$

$$\sum_{i:(i,j)\in L} y_{ij}^t \leq \delta_j[N_j - x_j^t], \quad \forall j \in R,\ t = 1, \cdots, T \tag{5.6}$$

$$e1_i^{t+} \leq N_i(z1_i^t), \quad \forall i \in R,\ t = 1, \cdots, T \tag{5.7}$$

$$e1_i^{t-} \leq (N_i\frac{Q_i - \Omega_i}{N_i - Q_i} + \Omega_i)(1 - z1_i^t), \quad \forall i \in R,\ t = 1, \cdots, T \tag{5.8}$$

$$e2_i^{t+} \leq N_i(z2_i^t), \quad \forall i \in R,\ t = 1, \cdots, T \tag{5.9}$$

$$e2_i^{t-} \leq N_i(1 - z2_i^t), \quad \forall i \in R,\ t = 1, \cdots, T \tag{5.10}$$

$$e3_j^{t+} \leq N_i(z3_j^t), \quad \forall i \in R,\ t = 1, \cdots, T \tag{5.11}$$

$$e3_j^{t-} \leq N_i(1 - z3_j^t), \quad \forall i \in R,\ t = 1, \cdots, T \tag{5.12}$$

$$x_i^t - \left\{(N_i - x_i^t)\frac{Q_i - \Omega_i}{N_i - Q_i} + \Omega_i\right\} = e1_i^{t+} - e1_i^{t-}, \quad \forall i \in R,\ t = 1, \cdots, T \tag{5.13}$$

$$Q_i - \delta_i(N_i - x_i^t) = e3_i^{t+} - e3_i^{t-}, \quad \forall i \in R,\ t = 1, \cdots, T \tag{5.14}$$

$$(x_i^t - e1_i^{t+}) - (Q_j - e3_j^{t+}) = e2_i^{t+} - e2_i^{t-}, \quad \forall (i,j) \in L_O,\ t = 1, \cdots, T \tag{5.15}$$

$$y_{ij}^t = x_i^t - e1_i^{t+} - e2_i^{t+}, \quad \forall (i,j) \in L_O,\ t = 1, \cdots, T \tag{5.16}$$

$$\sum_{j:(i,j)\in L} y_{ij}^t = \sum_{s=1:s\leq t}^{F} \theta_i^s d_i^{t-s+1}, \quad \forall\ i \in S_o,\ t = 1, \cdots, T \tag{5.17}$$

$$\sum_{s=1}^{F} \theta_i^s = 1, \quad \forall\, i \in S_o \tag{5.18}$$

$$x_i^1 = 0, \quad \forall\, i \in R \cup S_i \tag{5.19}$$

$$x_i^1 = D_i, \quad \forall\, i \in S_o \tag{5.20}$$

$$x_j^t \geq 0, \quad \forall j \in C,\ t = 1, \cdots, T \tag{5.21}$$

$$y_{ij}^t \geq 0, \quad \forall (i,j) \in L,\ t = 1, \cdots, T \tag{5.22}$$

$$e1_i^{t\pm} \geq 0, \quad z1_i^t \in \{0,1\}, \quad \forall (i,j) \in L_O,\ t = 1, \cdots, T \tag{5.23}$$

$$e2_i^{t\pm} \geq 0, \quad z2_i^t \in \{0,1\}, \quad \forall (i,j) \in L_O,\ t = 1, \cdots, T \tag{5.24}$$

$$e3_j^{t\pm} \geq 0, \quad z3_j^t \in \{0,1\}, \quad \forall (i,j) \in L_O,\ t = 1, \cdots, T. \tag{5.25}$$

Constraint (5.2) calculates cell density based on the cell density and flows from the previous time interval. Constraints (5.3) - (5.6) are linear inequalities that approximate Equation (5.1) and determine the flows over roadways cells.

Constraints (5.7) - (5.16) prohibit traffic holding on roadways connected with ordinary links. Constraints (5.7) - (5.12) allow only one variable in each $e1, e2, e3$ pair to have a positive value. For instance, if $z1_i^t = 1$, then $e1_i^{t+} \geq 0$ and $e1_i^{t-} = 0$, otherwise $e1_i^{t+} = 0$ and $e1_i^{t-} \geq 0$. Constraints (5.13) - (5.15) ensure that the flow between ordinary cells is governed by Equation (5.1), thus eliminating traffic holding, and determine the values for each difference variable $e1_i^{t\pm}, e2_i^{t\pm}, e3_j^{t\pm}$. The flows between cells are then calculated by Constraint (5.16).

Constraint (5.17) enforces each origin's given evacuee loading curve and the demand management staging plan. Some papers use an inequality, as opposed to an equality, in this constraint (as discussed in Liu et al., 2006); an inequality would enable the model to modify the demand loading curve in order to achieve an optimal objective value, which gives network operators an unrealistic level of control. For this reason, we have used an equality constraint instead an inequality. The $\theta$-variables control when a specific origin begins to evacuate; the demand loading curve remains fixed regardless of the value of $\theta$. Note, simultaneous evacuation is defined as $\theta_i^1 = 1, \forall i \in S_o$. Constraint (5.18) ensures that each origin evacuates within the search horizon, $F_i$.

Full formulations of the different models referred to in this paper are listed in the Appendix. In addition to the FRM formulation, we also formulate the system optimal LP from Ziliaskopoulos (2000) with our added demand management constraints in the Appendix.

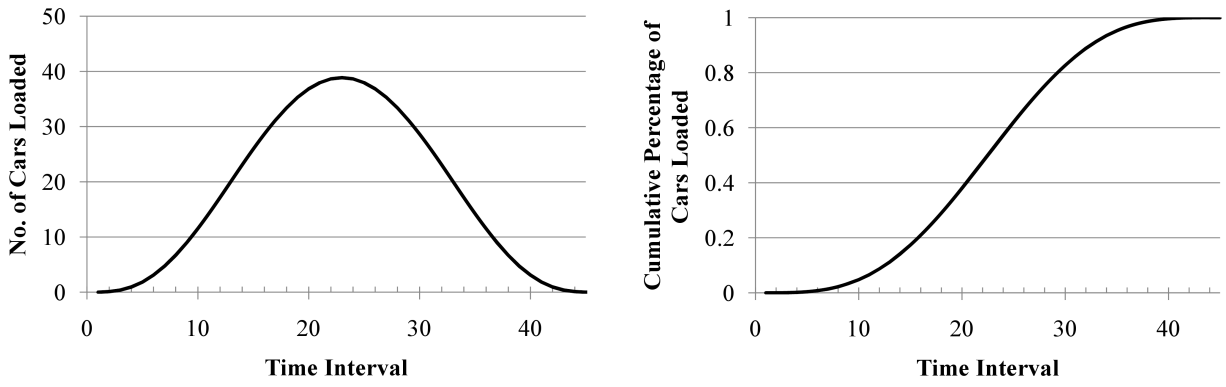
This formulation will be referred to as the CTM-MIP.

As consistent with the literature, we use a standard sigmoid curve to loading demand at each origin. We implement this using a beta distribution curve. This assumes that a specific evacuation time window is part of the evacuation instructions/order. The logistic curve, as introduced in Equation (5.26), has a tendency to load 50% of the demand by the half-loading time, $H$, but not necessarily 100% of the demand by $2H$ due to the behavior of the normal distribution.

$$P(t) = 1/\{1 + e^{-\alpha_i(t-H_i)}\}, \quad \forall\, i \in S_o \tag{5.26}$$

However, the beta distribution is, by definition, limited between 0 and 1. This is ideal to use, instead of the logistic function, in order to model different evacuation time windows and different response parameters. Equation (5.27) describes the beta distribution used to load demand onto the network; see Figure 5.3 for typical loading curves produced by Equation (5.27).

$$f(i,t) = \frac{D_i \left(\frac{t}{W_i}\right)^{\alpha_i - 1} \left(1 - \frac{t}{W_i}\right)^{\beta_i - 1}}{B(\alpha_i, \beta_i)}, \quad \forall\, i \in S_o,\ t = 1, \cdots, W_i \tag{5.27}$$




(a) (b)

Figure 5.3: Theoretical network loading curves (a) number of cars loaded at each time interval, (b) cumulative percentage of cars loaded during a given evacuation window.

Constraints (5.19) and (5.20) establish the initial conditions of the network; in this model, all demand, $D_i$, is loaded onto origin $i$ at the beginning of the first time interval

and then flows out of the origin based on the prescribed demand loading curve and staging plan. Constraints (5.21) - (5.25) are the non-negativity and binary restrictions on the model variables.

## 5.4 Objective Functions

There are various different objective functions that can be used to generate evacuation plans and policies. Some of the most common objectives are to minimize total travel time (or average travel time), network clearance time, and exposure to risk or casualties (Han et al., 2007). The type of emergency, whether short-notice or no-notice, will influence which type of objective is most appropriate. Additionally, the different objectives will drive different demand management solutions. This will be demonstrated later in this paper but first we will describe four objective functions.

During an evacuation, a key concern to many people is the length of time it will take to evacuate to safety. If a demand management staging strategy is used, then the evacuation time will consist of time waiting to evacuation and then the time traveling to the final destination.

$$\text{Minimize} \quad \sum_{t=1}^{T} \sum_{i \in S_o \cup R} \tau x_i^t \tag{5.28}$$

Objection Function (5.28) minimizes the total evacuation time (TET) by considering the time evacuees spend in the origins, $S_o$, waiting to evacuate and the time spent traveling on the roadway cells of the network, $R$. This objective function is one of the most commonly used in the literature. Ziliaskopoulos (2000) refers to this objective as minimizing total *travel* time instead of evacuation time. Since this formulation does not use staging to hold evacuees within the origins, travel time and evacuation time would be equivalent. Similarly, Chiu et al. (2007) uses an objective function to minimize total *system travel* time. In the context of evacuation modeling with staging, the term total evacuation time is more appropriate because it includes both the time waiting at the origin to evacuate and the travel time across the network. Yao et al. (2009) uses an objective to minimize *evacuee total threat exposure* but it is essentially the same as Objective (5.28) but weights each cell and time interval with a threat parameter. Although evacuation time includes additional time periods

prior to the issue of the evacuation order, it is a more appropriate term to use rather than travel time. The $\tau$-term can also be removed from objective function since it is a constant.

Instead of minimizing TET, the goal may be to just minimize total travel time (TTT). This would be effective if risk of exposure or hazard is low for a particular evacuation and the more time can be spent waiting in order to reduce congestion to an absolute minimum level.

$$\text{Minimize} \quad \sum_{t=1}^{T} \sum_{i \in R} \tau x_i^t \tag{5.29}$$

Objective Function (5.29) minimizes TTT since it only considers the time spent traveling across the network and excludes the time waiting at the origins. This objective, when used alone, is not an effective objective since it will increase the time spent waiting at the origins and extend the network clearance time in order to achieve an optimal TTT value.

Another important metric for measuring the effectiveness of an evacuation plan is network clearance time (NCT) (Han et al., 2007; Sbayti and Mahmassani, 2006).

$$\text{Minimize} \quad \sum_{t=1}^{T} E^t \tag{5.30}$$

$$E^t \geq 1 - \frac{1}{D} \sum_{i \in S_i} x_i^t, \quad t = 1, \cdots, T \tag{5.31}$$

Objective Function (5.30) minimizes the NCT by using a binary decision variable, $E^t$, that is set to 1 by Constraint (5.31) initially and then to 0 when the last vehicles clear the network, i.e., enters the sink. This constraint requires 100% of the population to evacuate before $E^t$ can equal 0, but this constraint can be easily modified to use either 90 or 95% of the total population depending on the assumptions made about the evacuee behavior.

### 5.4.1 The Equivalence of Minimizing TET and NCT

Under certain conditions, minimizing the TET produces a solution that also minimizes the NCT (Jarvis and Ratliff, 1982; Bish et al., 2011b). However, when loading curves are used and congestion limits flow at bottlenecks (Bish and Sherali, 2011) or when shelters with capacity restrictions are used (Bish et al., 2011b) this does not hold to be true.

In our formulation of the FRM, we assume that congestion will affect the queue discharge flow. This has been demonstrated through microscopic traffic simulation within INTEGRATION (Chamberlayne et al., 2011b,a). Additionally, we use loading curves at each origin in conjunction with a staging strategy. The following illustrative example demonstrates that the *min TET* and *min NCT* objective functions are not equivalent under the assumptions used within the FRM to build an optimal evacuation plan.

#### 5.4.1.1 Illustrative Example

Consider Network 1 consisting of a freeway and two on-ramps all leading to a single destination (an uncapacitated shelter). Origins O11 and O12 lead directly to a freeway while O2 and O3 lead to arterial roadways and then merge onto the freeway. Figure 5.4 is a cell and link representation of this network.

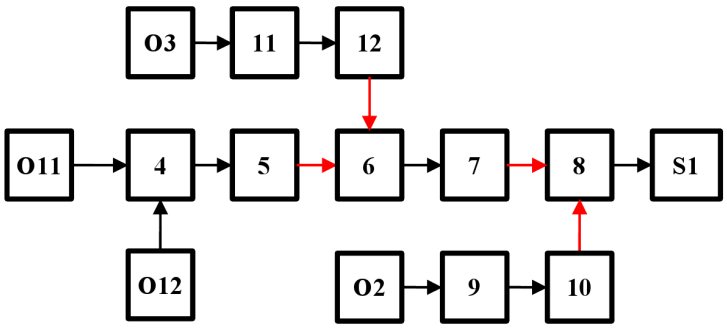



Figure 5.4: Network 1: Illustrative example test network

Realistic macroscopic parameters were used to represent traffic conditions across the network. For the entire network, a jam density, $k^{jam}$, of 264 vehicles/mile/lane (vpmpl) (or 20 ft/veh) was assumed. A free-flow speed, $u_f$, of 75 feet/second (fps) was assumed for the freeway; 37.5 fps for the arterials. A capacity flow rate, $q_c$, of 2400 vehicles/hour/lane (vphpl) was assumed for the freeway; 1800 vphpl for the arterials. Since the FRM is based upon the CTM-FR, the $\Omega$-parameter was fixed at $0.2Q_i$ based upon the previous freeway bottleneck simulations performed within INTEGRATION, a microscopic simulation software package (Van Aerde and Rakha, 2007a,b). A total demand of 2200 vehicles was loaded onto the network from the three origins; O11=600, O12=600, O2=500, O3=500. The beta function loading curve, as introduced in Section 5.3.2, was used to load the demand onto the network using a time window, $W_i$, of 75 intervals and response parameters, $\alpha_i, \beta_i$, equal to 4, $\forall i \in S_o$.

The loading curves produced a $\max(d_i^t)$ of 17.5 vehicles/interval at O11 and O12 and a $\max(d_i^t)$ of 14.5 vehicles/interval at O2 and O3. Lastly, a planning horizon, $T$, of 125 time intervals was used with a 30-sec time step ($\tau$). The network macroscopic parameters led to the FRM input parameters shown in Table 5.1.

Table 5.1: Network 1 input parameters

| | Q | N | $\ell$ (ft) |
|---|---|---|---|
| Freeway [Cells 4-8] | 40 | 225 | 2250 |
| Arterials [Cells 9 - 12] | 15 | 56.25 | 1125 |

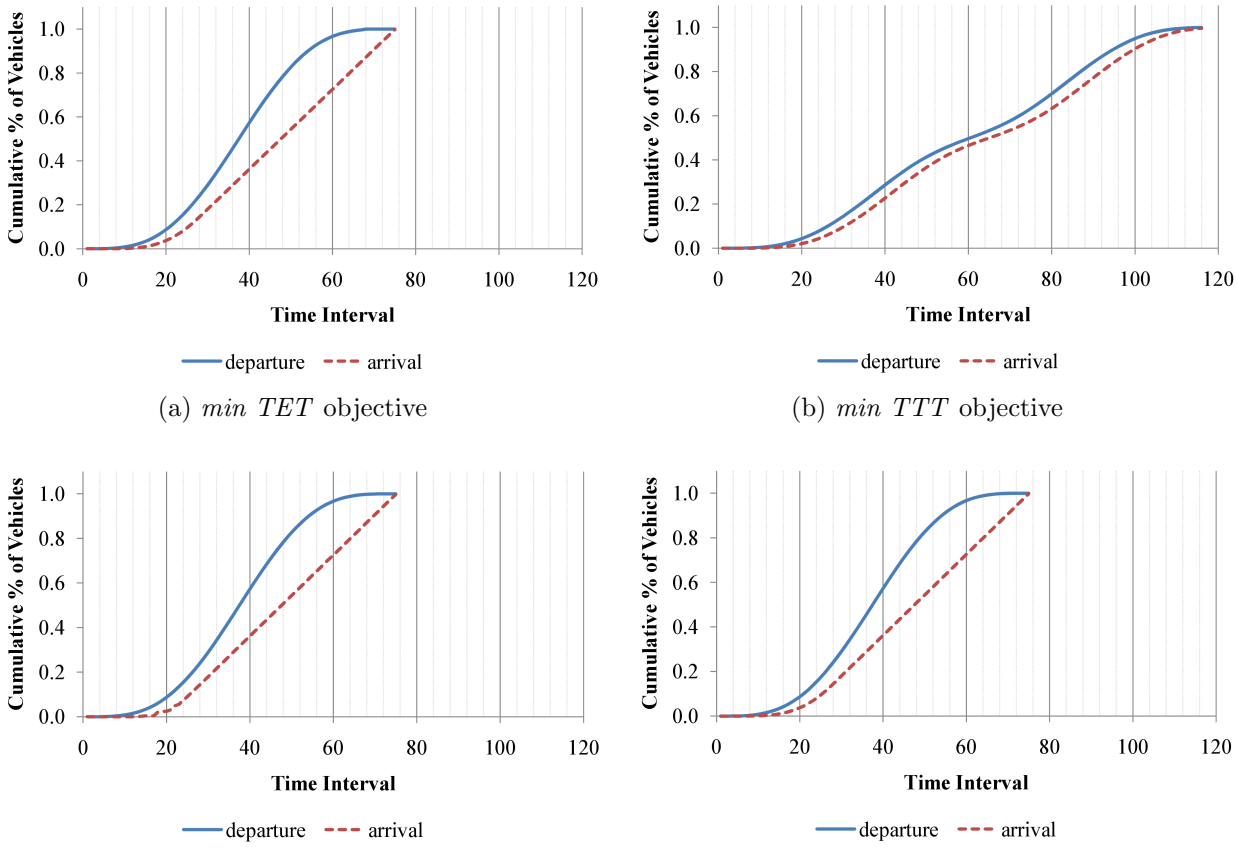
For the purposes of comparison, the network flows from the FRM and from the CTM-MIP will be compared using the three objective functions discussed so far. In the CTM-MIP formulation, traffic is modeled using the CTM. However, we use fixed loading curves at each of the four origins.

Table 5.2: Network 1 results given by the CTM-MIP

| | **Objective Functions** | | | |
|---|---|---|---|---|
| *CTM-MIP* | *min TET* | *min TTT* | *min NCT* | *min NCT + TTT* |
| **TET** | 105,319 | 144,200 | 105,563 | 105,319 |
| **TTT** | 21,719 | 10,000 | 21,963 | 21,719 |
| **NCT** | 80 | 125 | 80 | 80 |
| | | | | |
| **Staging** | | | | |
| *O11* | 1 | 1 | 1 | 1 |
| *O12* | 1 | 47 | 1 | 1 |
| *O2* | 1 | 1 | 1 | 1 |
| *O3* | 1 | 47 | 1 | 1 |

When the network flows are calculated using the CTM-MIP with the *min TET* and the *min NCT* objective functions, the two objective functions yield equivalent network clearance time results, see Table 5.2. Neither objective function stages any of the origins because there isn't a penalty for congestion within the network. Both objectives give a similar TET and TTT values and both give identical NCT values; the TET and TTT values only differ slightly due to different merge priorities determined by each objective function. Additionally, minimizing TTT increases the TET and NCT. If this objective was used to build an evacuation plan, it would ensure the shortest travel time across the network but would hold evacuees

at their respective origins much too long and would probably create poor compliance with the plan.




(a) *min TET* objective

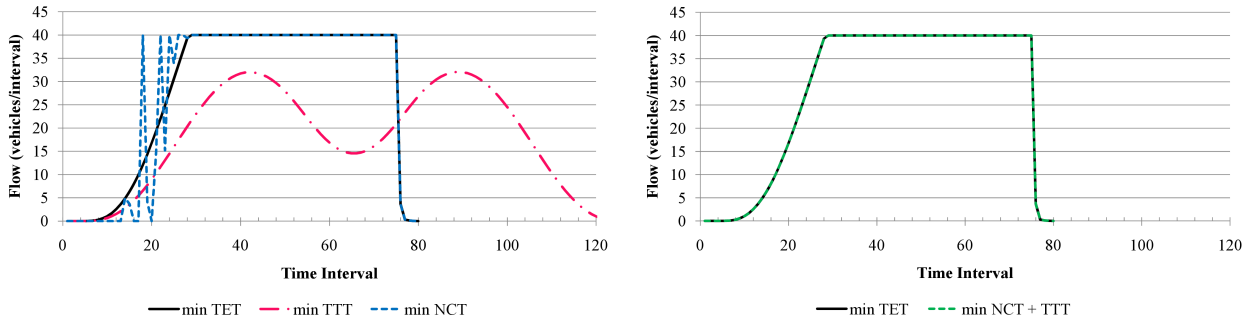
(b) *min TTT* objective

(c) *min NCT* objective

(d) *min NCT + TTT* objective

Figure 5.5: Cumulative Curves for Four Objective Functions using the CTM-MIP

Graphically, these comparisons can be seen in Figures 5.5 and 5.6; note, these figures, along with Tables 5.2 and 5.3, also include results which will be discussed next in Section 5.4.2. In the cumulative curves in Figure 5.5, the departure curve combines the loading curves at each origin based on the derived staging plan; the arrival curve represents the flows into the sink. The area between these curves represents the TTT and the area between the vertical axis and the arrival curve represents TET. The *min TET* and *min NCT* objective functions are also graphically equivalent as Figures 5.5a and 5.5c are virtually identical. Additionally, the min TTT cumulative curves in Figure 5.5b displays a much longer loading curve and the absence of congestion; the departure and arrival curves are only separated by the average free-flow travel time between the origins and the destination.




(a) *min TET*, *min TTT*, and *min NCT* objective func(b) *min TET* and *min NCT + TTT* objective functions tions

Figure 5.6: Comparisons of flows over time of from four objective functions using the CTM-MIP

In Figure 5.6a, both the *min TET* and *min NCT* objectives produce maximum flows over time ($Q = 40$) and visibly clear the network at the same time interval (t=80). Interestingly, only this figure shows the difference in how flows are calculated over time between the two objective functions. Although the two objectives yield similar TET / TTT solutions and identical NCT solutions, the *min NCT* objective uses much more traffic holding during the first 30 time intervals than the *min TET* objective. This displays some of the artificial flexibility that traffic holding gives the CTM-MIP.

Conversely, as illustrated in Table 5.3, minimizing TET does not yield a solution that minimizes NCT. The *min NCT* objective gives the optimal NCT value but at the detriment of the TET and TTT values. The *min NCT* objective also stages the origins more than the *min TET* objective which accumulates more wait time at the origins and is responsible for the 12% increase in the TET; this staging also creates more congestion on the network that increases the TTT by 50% but also allows the network to clear in 93 intervals. The results from the *min TTT* objective are similar to the results from the CTM-MIP; this objective yields the same minimum TTT value by staging the origins the most and returning the maximum values for TET and NCT.

Again, these differences can be been seen graphically in the cumulative curves. By comparing Figures 5.7a and 5.7c, the amount of congestion is much larger in Figure 5.7c than in Figure 5.7a. Also, the NCT values are visibly different; the departure and arrival curves meet much earlier in Figure 5.7c than in Figure 5.7a. The min TTT objective yields very similar results between the CTM-MIP and the FRM; Figure 5.7b is virtually identical

Table 5.3: Network 1 results given by the FRM

| | **Objective Functions** | | | |
|---|---|---|---|---|
| *FRM* | *min TET* | *min TTT* | *min NCT* | *min NCT + TTT* |
| **TET** | 110,182 | 139,600 | 123,444 | 121,689 |
| **TTT** | 16,082 | 10,000 | 25,844 | 24,089 |
| **NCT** | 101 | 125 | 93 | 93 |
| | | | | |
| **Staging** | | | | |
| *O11* | 1 | 1 | 1 | 1 |
| *O12* | 1 | 1 | 1 | 1 |
| *O2* | 1 | 47 | 17 | 16 |
| *O3* | 22 | 47 | 13 | 14 |

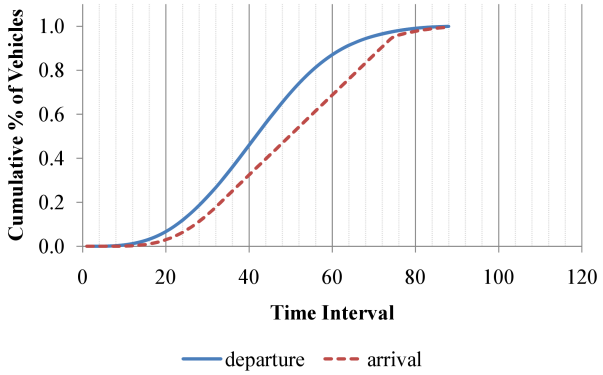
to Figure 5.5b.

The flows over time from the two objective functions are displayed in Figure 5.8a. The flows from the min NCT objective are much different than those from the min TET objective during the entire time horizon within the FRM; the flows are reduced by the effects of congestion at the network restriction and are erratic due to the limited amount of traffic holding still allowed only at the merge intersections. The difference in NCT values can be seen as well within Figure 5.8a.

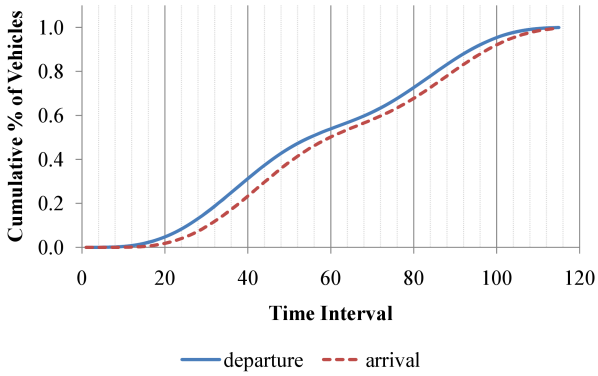
This small example illustrates that the *min TET* and *min NCT* objective functions are not equivalent. The principal reasons for this are the use of loading curves and the modeling the effects of congestion at network restrictions. Additionally, the limits on traffic holding within the model also contribute the differences between the two objective functions. All of these reasons make the FRM a more realistic model than the CTM-MIP and a better model to use in the development of an evacuation plan. While using the FRM, special consideration should be given when considering an objective function to use depending on what metric is more important for a particular situation.

### 5.4.2 Lexicographic Objective Function

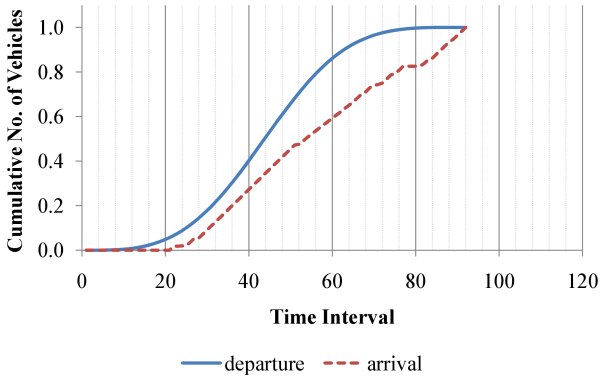
Instead of having to choose between the *min TET* and the *min NCT* objective functions, it would be ideal to combine the benefits of both objectives. A lexicographic objective function



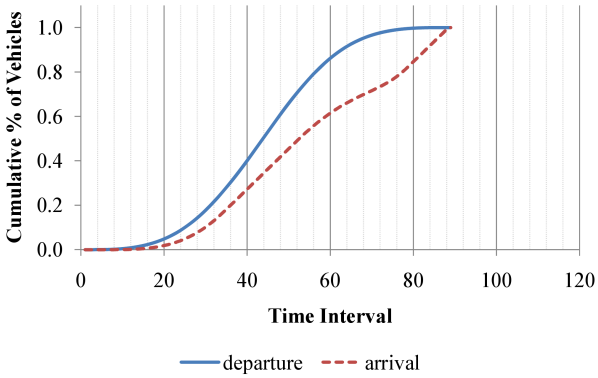

(a) *min TET* objective




(b) *min TTT* objective

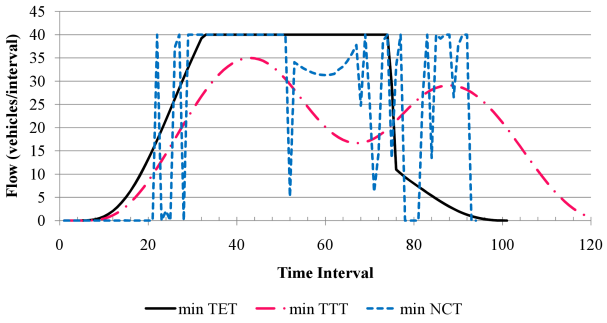



(c) *min NCT* objective

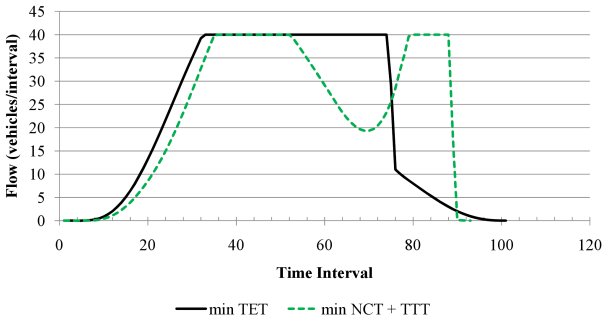



(d) *min NCT + TTT* objective

Figure 5.7: Cumulative curves for four objective functions using the FRM




(a) *min TET*, *min TTT*, and *min NCT* objectives




(b) *min TET* and *min NCT + TTT* objectives

Figure 5.8: Flows over time from four objective functions using the FRM

can be used to do this; it combines the *min NCT* objective with the *min TTT* objective and/or the *min TET* objective functions. This objective function was first developed in Bish and Sherali (2011). Objective Function (5.32) shows the combination, in priority, of the *min*

*NCT* and the *min TTT* objective functions.

$$\text{Minimize} \quad \sum_{t=1}^{T} E^t + \frac{1}{TD} \sum_{t=1}^{T} \sum_{i \in R} \tau x_i^t \tag{5.32}$$

The *min TTT* term of the objective function is scaled by an upper bound on TTT in order to give priority to minimizing NCT. Note, Constraint (5.31) is required when using this objective function as well. The results from using the lexicographic objective functions are also displayed in Tables 5.2 and 5.3. Note, the results from using a lexicographic objective function that combines *min NCT* and *min TET* were very similar and are not shown.

Within the CTM-MIP, the *min NCT + TTT* objective yields the same results as the *min NCT* and the *min TET* objectives (see Table 5.2). The cumulative curves and flows out of the network are also identical as shown in Figures 5.5d and 5.6b. However, within the FRM, the lexicographic objective yields better TET and TTT values than the *min NCT* objective while resulting in the same NCT value (see Table 5.3). This is accomplished by manipulating the merge priorities differently. The flows out of the network are also much less erratic than those from the *min NCT* objective, as shown in Figure 5.8b, and display traffic flow behavior expected as congestion builds and dissipates within the transportation network. This is promising since it shows that by modeling the effects of congestion, the lexicographic objective can combine the benefits of both objective functions when the *min TTT* objective, by itself, did not prove very useful.

In this small example, the *min TET* and *min TTT* objective models solve relatively quickly. However, the *min NCT* and lexicographic objective models require much more time to solve (20-30 times longer) when using the same planning horizon value, $T$. An alternate method is to use the min TET and min TTT objective functions using a bisection search of the planning horizon. For instance, by starting with $T = 200$, the bisection search would solve either objective function with $T = 100$. If the solution is infeasible, then the search would move to $T = 150$. If feasible, the search would move to $T = 125$. The bisection search continues until the smallest value for $T$ is achieved with a feasible solution. The $T$-value represents the optimal NCT value with an equivalent solutions for the min NCT or lexicographic objective functions. The principal advantage of this solution technique is that it will solve the FRM in much less time (over all iterations of the search) than when using the min NCT or the lexicographic objective functions.

## 5.5 Control Groups for Staging

In this section, we will use the FRM formulation with both the *min TET* objective function (5.28) and the lexicographic objective function (5.32) to explore the use of control groups for staging. The demand management strategy used in this paper for an evacuation network is to stage or sequence the evacuation of different origins, as a group, in order to achieve an optimal TET or NCT (depending on what objective is most important). In this strategy, we ***control*** the evacuation of different ***groups*** entering the network. This assumes a prior understanding of the size and number of control groups available to the network planner. However, we can use the FRM to determine a near-optimal number of control groups for use in staging for the evacuation plan.

In Constraint (5.17), the $\theta$ binary decision variable is used to determine the exact time interval when a loading curve begins to send traffic out of a specific origin. As introduced in Bish and Sherali (2011), this binary variable can be relaxed (between 0 and 1) in order to better understand the optimal number of control groups necessary to achieve the best results in regards to the evacuation metrics.

In order to illustrate this concept, we use an example network similar to Network 1 (see Section 5.4.1.1) but with only one origin control group for each roadway segment; O1 for the freeway and O2 and O3 for the two on-ramps that join the freeway. This network, Network 2, is shown in Figure 5.9. We will next evaluate the network flows produced by the FRM by relaxing the $\theta$-variable for controlling staging.

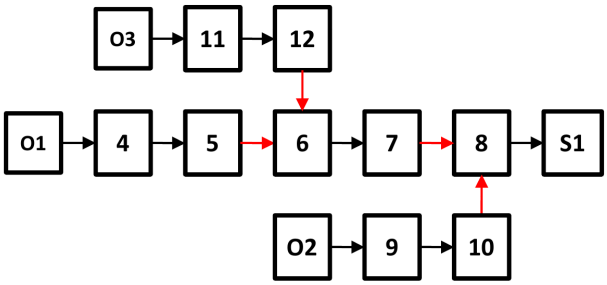



Figure 5.9: Network 2: 3-origin control group network

We used the same network parameter inputs as in Section 5.4.1.1, however, we stressed the system with additional demand. A total demand of 2350 vehicles was used: O1 was loaded with 1350 vehicles, O2 with 500 vehicles, and O3 with 500 vehicles. Table 5.4 displays the results from the runs of the FRM using the *min TET* objective function; the first column

displays the results when the $\theta$-variable is relaxed and the second column displays the results when the $\theta$-variable is restricted between 0 and 1 (binary).

Table 5.4: Results for Networks 2 and 3 using the *min TET* objective

| | **Network 2** | | | | **Network 3** | |
|---|---|---|---|---|---|---|
| | *relaxed $\theta$-variable* | | | *binary $\theta$-variable* | | *binary $\theta$-variable* |
| **obj value** | 121,360.93 | | | 122,384.73 | | 121,367.61 |
| **TET** | 121,361 | | | 122,385 | | 121,368 |
| **TTT** | 17,191 | | | 17,585 | | 17,187 |
| **NCT** | 109 | | | 110 | | 109 |
| **Staging** | *interval* | *fraction* | **Staging** | *interval* | **Staging** | *interval* |
| *O1* | 1 | 0.98 | *O1* | 1 | *O11* | 1 |
| *O1* | 32 | 0.02 | | | *O12* | 29 |
| *O2* | 1 | 1 | *O2* | 1 | *O2* | 1 |
| *O3* | 1 | 0.05 | *O3* | 32 | *O31* | 1 |
| *O3* | 31 | 0.41 | | | *O32* | 31 |
| *O3* | 30 | 0.55 | | | *O33* | 30 |

In Table 5.4, under the Network 2 column, we see that by relaxing the staging variable, we can achieve a lower bound on the objective function value. For some of the origins, there are multiple staging intervals listed due to the relaxed $\theta$-variable. The "fraction" column gives the proportion of each origin that is staged at the specific interval given in the "interval" column. For this example, the relaxation indicates that six control groups would be an efficient method of staging demand. By restricting the $\theta$-variable and using only three control groups, the objective value worsens along with the other evacuation metrics of TET, TTT, and NCT. Based on these results, Network 3 is a modification of Network 2 and consists of six control groups instead of three; see Figure 5.10.

Table 5.5: Total demand for Network 3

| | *O11* | *O12* | *O2* | *O31* | *O32* | *O33* |
|---|---|---|---|---|---|---|
| demand | 1323 | 27 | 500 | 20 | 275 | 205 |

The origins of Network 3 were loaded with same total demand as used in Network 2, however, the origin demand was divided into groups according to the fractional solutions for the $\theta$-variable. The results for Network 3 are also listed in Table 5.4, under the Network 3

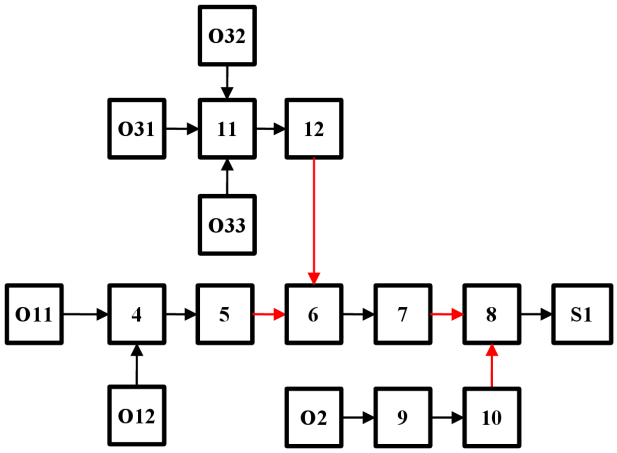



Figure 5.10: Network 3: 6-origin control group network

column. The objective function value improves, along with other evacuation metrics, when compared to the results from Network 2 when using three control groups. The staging solution is also very similar to the solution given when the $\theta$-variable is relaxed. The effects of improving the network flows by using different origin control groups can also be displayed by graphing the network flows over time at the network destination, or cell **S1** in our example.

Figure 5.11 displays these network flows for both Networks 2 and 3. When using the same number of origins as indicated by the $\theta$-variable relaxation, the network flows are virtually identical (the Network 2 (relaxed) and Network 3 (binary) curves in the figure). However, the network flows for Network 2 using three origin control groups and binary $\theta$-variables do not load demand onto the network as efficiently and there are flow reductions due to congestion are observed (i.e., the periods of flow below 40 vehicles/interval) which increase the TET, TTT, and NCT values.

Similarly, for the lexicographic objective function, we compared the network flows for the Network 2 for both the relaxed and binary $\theta$-variable model runs. These results are shown in Table 5.6.

The relaxed $\theta$-variable solution recommends using five origin control groups with the O1 demand split into two groups (85% and 15%) and the O3 demand split into two groups (5% and 95%). The objective solution for the three origin group network is worse than the relaxed $\theta$-variable solution, as expected. A new network, Network 4, was developed based on the five control group recommendation; see Figure 5.12.

As in the *min TET* model runs, the total demand for Network 4 is the same as

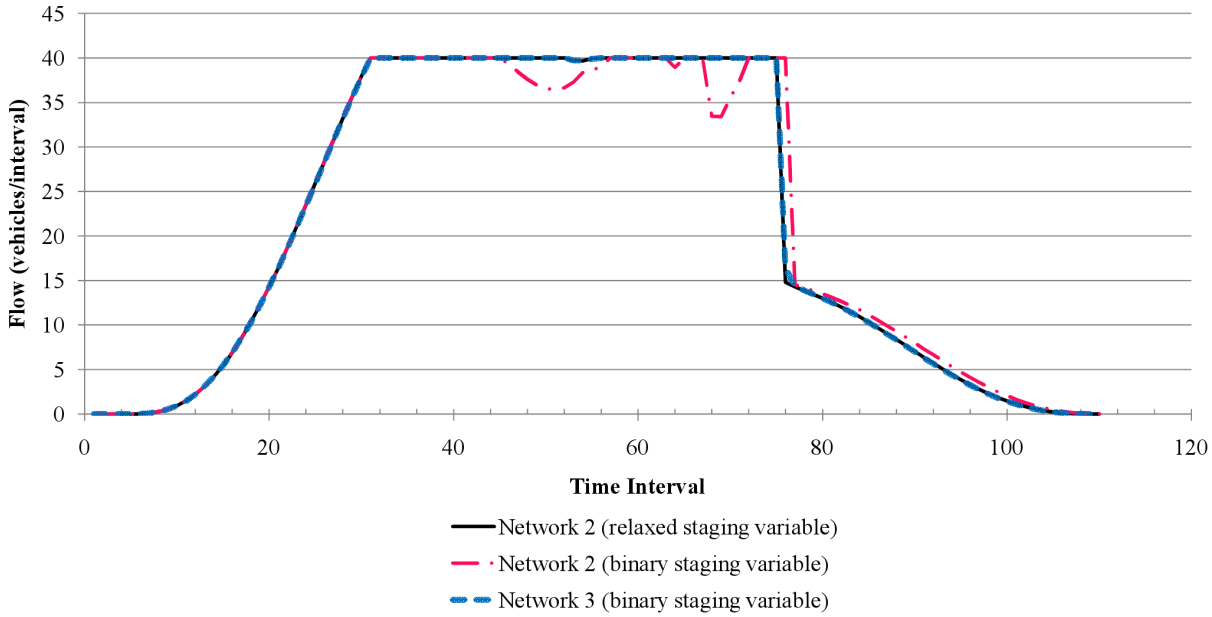



Figure 5.11: Network flows over time for Networks 2 and 3 using the *min TET* objective

used in Network 2 but the demand is split into groups based on the fractional values of the $\theta$-variable. See Table 5.7 for the exact demand loading used.

By using five origin control groups, the objective function value and the NCT is improved to that of the relaxed $\theta$-variable model run. Additionally, the staging solution is exactly the same as suggested by $\theta$-variable relaxation. Figure 5.13 displays the network flows for Networks 2 and 4. In order to achieve an optimal objective value using the lexicographic objective function, some congestion is formed in the network in order to obtain a minimum NCT (note, that in Network 4, the NCT is 98 intervals as compared to 109 intervals in Network 3). By using five control groups, instead of three, the network flows more closely match the flows modeled when using the relaxed $\theta$-variable to control staging. Interestingly, five control groups are required to obtain an optimal NCT value but six control groups are required to obtain an optimal TET value. Again, the decision on how many control groups to use would depend on what metric is most important.

These results show how the $\theta$-variable can be used to stage demand (via fixed loading curves) onto the network at exact time intervals and to help determine the optimal number of origin control groups. In both examples, whether using the *min TET* or the lexicographic objective functions, we were able to match the objective value of the relaxed $\theta$-variable

Table 5.6: Results for Networks 2 and 4 using the lexicographic objective

| | **Network 2** | | | | **Network 4** | |
|---|---|---|---|---|---|---|
| | *relaxed θ-variable* | | *binary θ-variable* | | *binary θ-variable* | |
| **obj value** | 97.07 | | 98.06 | | 97.07 | |
| **TET** | 132,524 | | 128,545 | | 132,154 | |
| **TTT** | 20,939 | | 19,245 | | 20,659 | |
| **NCT** | 98 | | 99 | | 98 | |
| **Staging** | *interval* | *fraction* | **Staging** | *interval* | **Staging** | *interval* |
| *O1* | 1 | 0.85 | *O1* | 1 | *O11* | 1 |
| *O1* | 19 | 0.15 | | | *O12* | 19 |
| *O2* | 21 | 1 | *O2* | 22 | *O2* | 21 |
| *O3* | 1 | 0.05 | *O3* | 20 | *O31* | 1 |
| *O3* | 19 | 0.95 | | | *O32* | 19 |

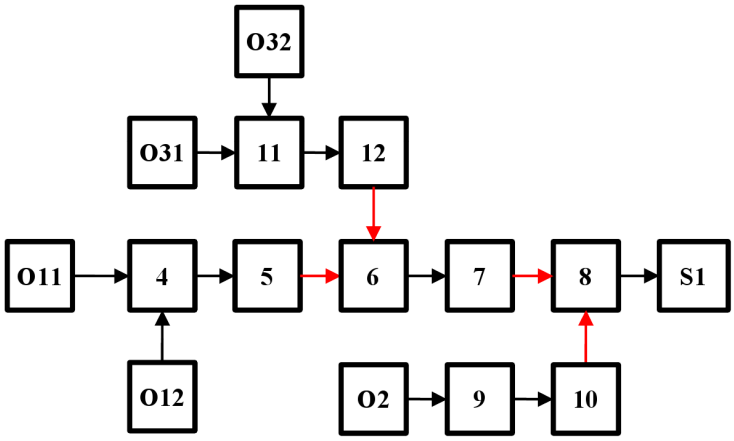



Figure 5.12: Network 4: 5-origin control group network

solutions. Admittedly, in an actual implementation, a network planner would not be able to breakdown neighborhoods into exact sub-groups as indicated by the $\theta$-variable relaxation. However, the factional values could be used as a guide to recommend additional control groups for use during an evacuation.

## 5.6 Modeling Intersection Controls

The FRM, as introduced in Bish et al. (2011a) and earlier in Section 5.3, was an initial attempt in formulating a model that would capture the effects of congestion and remove

Table 5.7: Total demand for Network 4

| | *O11* | *O12* | *O2* | *O31* | *O32* |
|---|---|---|---|---|---|
| demand | 1147.5 | 202.5 | 500 | 25 | 475 |

traffic holding from *ordinary links* within a transportation network. However, the model can be modified to better model intersection control measures. The following paragraphs will explore the effects of traffic holding at merge intersections and will present an improved formulation of the FRM for modeling intersection controls.

The FRM does not allow traffic holding along ordinary links and only allows traffic holding at intersections (or merging and diverging links). At these intersections, it could be reasonably argued that traffic holding could be duplicated by traffic control measures at the merge on-ramp. However, along the freeway link going into the merge and at a diverge intersection, there are not traffic controls that could duplicate the traffic holding behavior.

To analyze this behavior, we will study three test networks. All three networks will use four origins to load demand into the network and will have one destination. However, the networks will be distinct in the number of merge intersections: Network 5 - one merge,

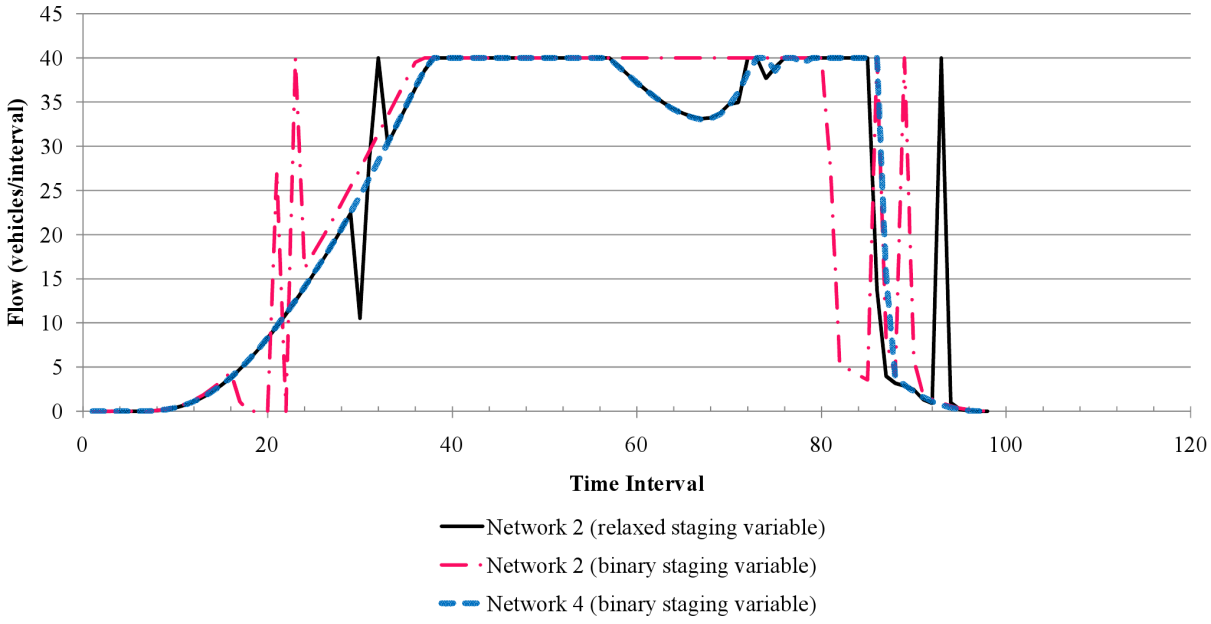



Figure 5.13: Network flows over time for Networks 2 and 4 using the lexicographic Objective

Network 6 - two merges, and Network 7 - three merges. See Figures 5.14 - 5.16 for the network structure. The total population used at each origin, $D_i$, is also shown in the figure.

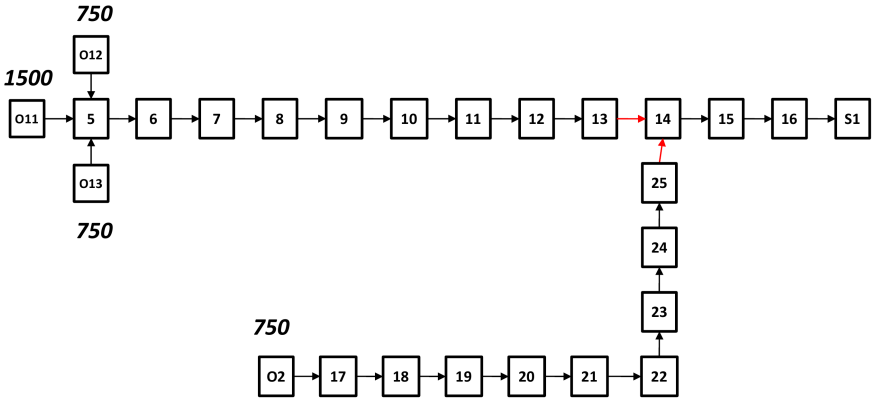



Figure 5.14: Network 5

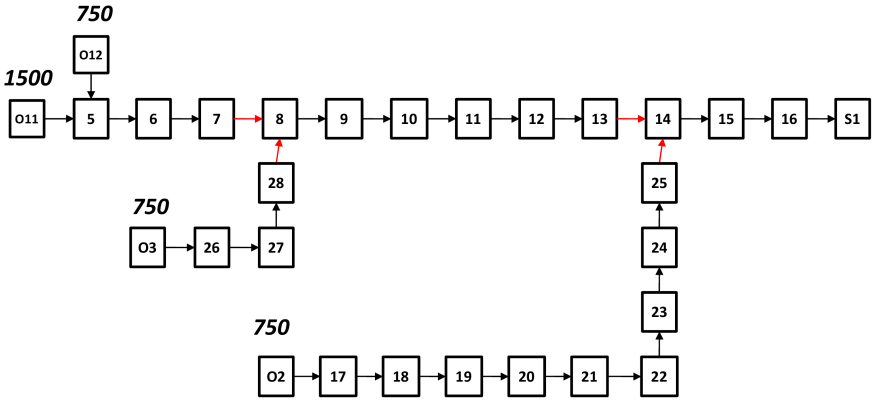



Figure 5.15: Network 6

The same macroscopic parameters were used to derive the model input parameters as were used in Network 1 in Section 5.4. See Table 5.8 for the input parameters for Networks 5 - 7.

In regard to network demand loading, we used a fixed evacuation time window, $W_i$, of 125 intervals (or approximately 1 hour). We set the evacuation response parameters, $(\alpha_i, \beta_i)$, to 4. For simplicity, the same demand loading curve parameters was used at each origin.

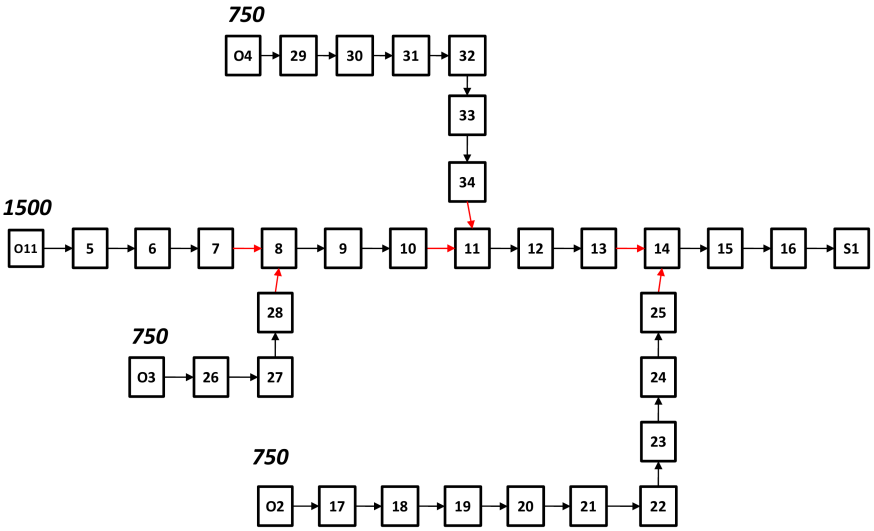



Figure 5.16: Network 7

Table 5.8: Networks 5 - 7 input parameters

| | Q | N | $\ell$ (ft) | $\tau$ (sec) |
|---|---|---|---|---|
| Freeway [Cells 5-16] | 40 | 225 | 2250 | 30 |
| Arterials [Cells 17 - 34] | 15 | 56.25 | 1125 | 30 |

### 5.6.1 Results

The path length (in time) for each route between origin and destination is equal in all three test networks; there are 12 cells between every origin and destination. Additionally, the total demand loaded onto each network is also equal - 3750 vehicles. This ensures that we can make equivalent comparisons across the networks.

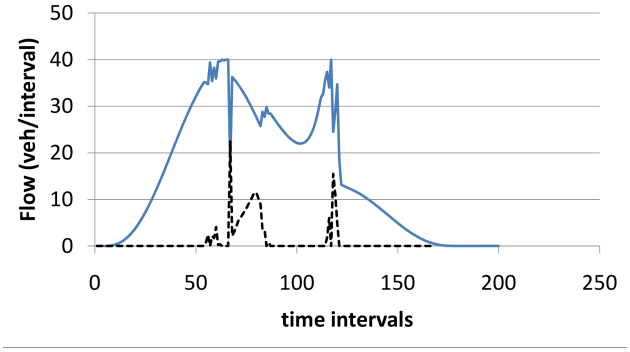
In order to demonstrate a weakness in the FRM, consider the solutions for the FRM formulation [Objective Function (5.28) and Constraints (5.2) - (5.25)] across the three test networks shown in Table 5.9.

Table 5.9: FRM results for Networks 5-7

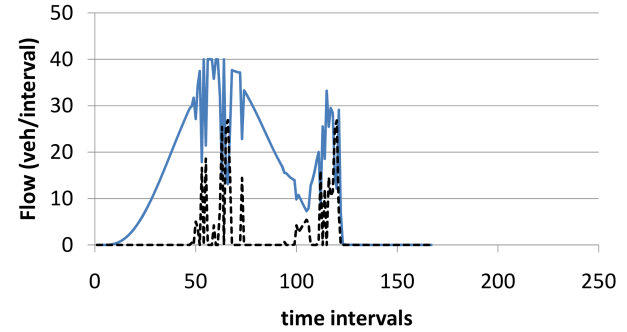
| | **Network 5** | **Network 6** | **Network 7** |
|---|---|---|---|
| **TET** | 353,421 | 346, 802 | 330, 210 |
| | | | |
| **Staging** | | | |
| *O11* | 1 | 1 | 1 |
| *O12* | 47 | 1 | - |
| *O13* | 48 | - | - |
| *O2* | 1 | 23 | 1 |
| *O3* | - | 52 | 1 |
| *O4* | - | - | 40 |
| | | | |
| **Traffic Holding (vehicles)** | n/a | 194 | 552 |
| **Bottleneck Utilization** | 55.8% | 54.5% | 60.1% |

The three objective values in Table 5.9 should be equal, however, this is clearly not the case. The reason why is due the decisions made by the FRM. The model can use three strategies in order to influence the objective function: 1) modify the staging of the four origins through the use the $\theta$-variable, 2) adjust the merge priorities between the freeway and on-ramp at each merge intersection, and 3) use traffic holding at some of the upstream restrictions in order to ensure that the system bottleneck flows at full capacity. Although the first two strategies seem reasonable, the last strategy is not realistic since traffic control measures cannot always duplicate the behavior. Since the goal of the FRM initially was remove traffic holding from ordinary links, traffic holding can now only occur in the cells immediately upstream of each merge. Therefore, traffic holding can be used in 2 cells in Network 5, 4 cells in Network 6, and 6 cells in Network 7. This difference in storage space

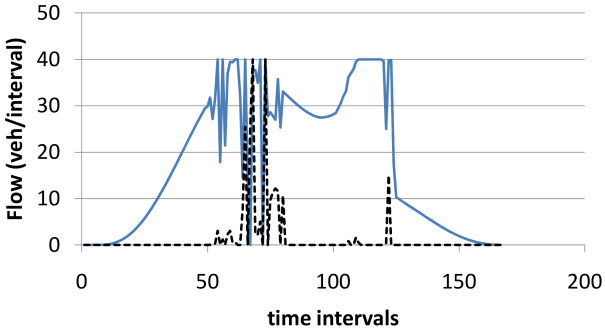
between Networks 7 and 5 accounts for the reduction in objective value from 353,421 to 330,210.




(a) Cell 8 - Network 6




(b) Cell 8 - Network 7




(c) Cell 11 - Network 7

Figure 5.17: Restriction flows over time for Networks 6 and 7

By plotting the flows over time out of each restriction, we can graphically show traffic holding can improve the capacity of the system bottleneck, see Figure 5.17. In each figure, the solid line displays the flows over time and the dashed line displays the amount of traffic holding occurring over time at each restriction. Table 5.9 totals the cumulative amount of traffic holding (in vehicles) in Networks 6 and 7; in Network 7, by holding over 550 vehicles, the FRM is able to improve the objective function solution by 7% and improve the utilization of the bottleneck by 5%. The bottleneck utilization is an average, over time, of flows out of the bottleneck expressed as a percentage of the bottleneck capacity, $Q_i$. The increased bottleneck utilization is also displayed in Figure 5.18; note that the Network 7 flows clear the network sooner than the Network 5 flows due to the improvement in bottleneck utilization. Additionally, traffic holding allows the model to stage the origins less; in Network 5, the model delays origins through staging by 93 time intervals as compared to 39 time intervals in Network 7.

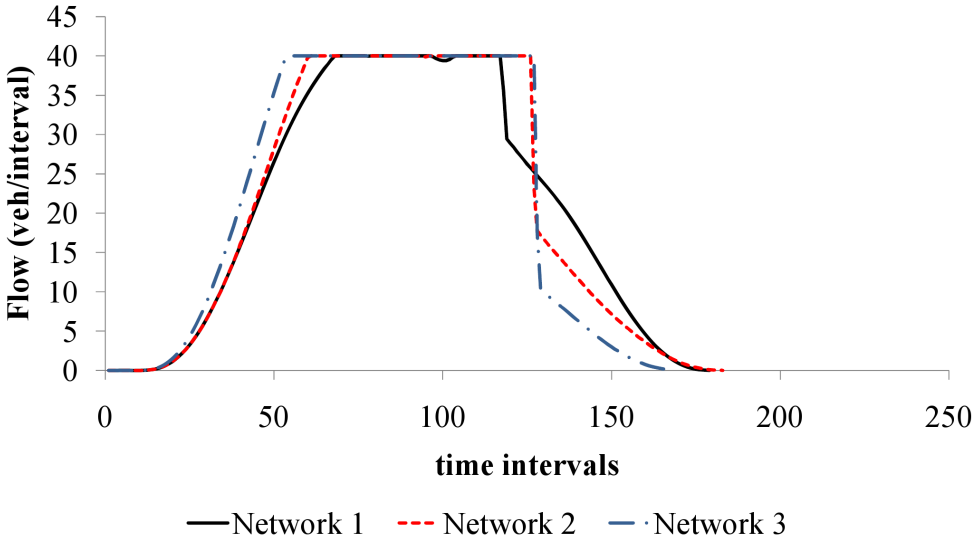



Figure 5.18: Bottleneck flows over time for Networks 5-7

### 5.6.2 Intersection Control Constraints

In order, to remove traffic holding from the entire network, the FRM formulation must be modified. We will remove the inequality constraints that allow traffic holding to occur, Constraints (5.3) - (5.6), and expand our mixed integer constraints and introduce some additional notation.

Sets:

$L_{M,f}$ : a subset of $L_M$; a set of freeway merge links between cells within the network
$L_{M,o}$ : a subset of $L_M$; a set of on-ramp merge links between cells within the network

Variables:

$e2f_i^{t+}$ : the freeway component of the $e2_i^{t+}$ variable; this variable controls the priority given to the freeway link at a merge intersection, $\forall (i,k) \in L_M,\ t = 1, \cdots, T$
$e2o_i^{t+}$ : the on-ramp component of the $e2_i^{t+}$ variable; this variable controls the priority given to the on-ramp link at a merge intersection, $\forall (i,k) \in L_M,\ t = 1, \cdots, T$

$$\sum_{(i,k) \in L_M} (x_i^t - e1_i^{t+}) - (Q_k - e3_k^{t+}) = e2_i^{t+} - e2_i^{t-}, \quad \forall (i,k) \in L_{M,f},\ t = 1, \cdots, T \tag{5.33}$$

$$e2_i^{t+} = e2f_i^{t+} + e2o_i^{t+}, \quad \forall (i,k) \in L_{M,f},\ t = 1, \cdots, T \tag{5.34}$$

$$y_{ik}^t = x_i^t - e1_i^{t+} - e2f_i^{t+}, \quad \forall (i,k) \in L_{M,f},\ t = 1, \cdots, T \tag{5.35}$$

$$y_{jk}^t = x_j^t - e1_j^{t+} - e2o_i^{t+}, \quad \forall (j,k) \in L_{M,o},\ t = 1, \cdots, T \tag{5.36}$$

Constraint (5.33) expands Equation (5.1) to merge links and calculates the $e2_i^{t\pm}$ variables specifically for the freeway cells immediately upstream of each merge. Constraint (5.34) divides the $e2_i^{t+}$ term into freeway and on-ramp link components. Note, the $e2_i^{t+}$, $e2f_i^{t+}$, and $e2o_i^{t+}$ terms apply to the entire intersection and are attached to the upstream freeway cell, $i$ in this case, for modeling simplicity. Constraints (5.35) and (5.36) calculate the flows on each link leading into the merge bottleneck ensuring that the bottleneck flows at capacity when total demand meet or exceeds the capacity of the bottleneck.

Additionally, we may want to dictate how merge priorities are controlled at the merge. Our constraint set does not dictate the merge priorities but just that traffic holding should not exist. If traffic control measures exist at the on-ramp that would manage the merge priority (through traffic signals, police officers, etc.), additional constraints may be necessary.

$$e2o_i^{t+} \geq 0.2e2_i^{t+}, \quad \forall (i,k) \in L_{M,f}, \ t = 1, \cdots, T \tag{5.37}$$

$$e2o_i^{t+} \leq 0.4e2_i^{t+}, \quad \forall (i,k) \in L_{M,f}, \ t = 1, \cdots, T \tag{5.38}$$

$$e2o_i^{t+} = 0.25e2_i^{t+}, \quad \forall (i,k) \in L_{M,f}, \ t = 1, \cdots, T \tag{5.39}$$

Constraints (5.37) and (5.38) define a range of priority given to the on-ramp link of the merge (a range of 20-40% was selected as a reasonable range but clearly any range could be used). Constraint (5.39) defines an exact priority for the on-ramp link of the merge (again, 25% is just an example priority).

In order to compare the results of the FRM using these different constraints, four model runs were developed. All runs used the *min TET* Objective Function (5.28). See Table 5.10 for the FRM results across the four model runs using Network 6 (two merge network example).

- **Run 1:** FRM with traffic holding at merge intersections
- **Run 2:** FRM without traffic holding [with Constraints (5.33) - (5.36) and without Constraints (5.3) - (5.6)]
- **Run 3:** FRM without traffic holding and a merge priority range assigned to the on-ramp link (with Constraints (5.33) - (5.38), and without Constraints (5.3) - (5.6)]

- **Run 4:** FRM without traffic holding and an exact merge priority assigned to the on-ramp link [with Constraints (5.33) - (5.36), (5.39), and without Constraints (5.3) - (5.6)]

Table 5.10: FRM results for Network 6

| | **Run 1** | **Run 2** | **Run 3** | **Run 4** |
|---|---|---|---|---|
| **TET** | 346,802 | 361,514 | 362,253 | 363,000 |
| **Staging** | | | | |
| *O11* | 1 | 1 | 1 | 1 |
| *O12* | 1 | 55 | 1 | 1 |
| *O2* | 23 | 54 | 55 | 55 |
| *O3* | 52 | 1 | 55 | 56 |
| **NCT** | 186 | 189 | 189 | 190 |
| **Traffic Holding** | 194 | 0 | 0 | 0 |
| **Bottleneck Utilization** | 54.5% | 51.2% | 52.2% | 51.9% |

The objective function value in Run 1 is artificially low due to the traffic holding, as previously discussed. The results from Runs 2-4 are similar but clearly larger than the value in Run 1. Additionally, more staging is required now in order to make-up for loss of traffic holding in the model. Therefore, although the values for TET and staging are increased, these results are more realistic and would form the basis for a more solid evacuation plan. Run 4 has the greatest value for TET, staging (or total delay at the origins), and NCT since it is the most restrictive set of constraints that remove traffic holding and set an exact merge priority for the on-ramp link.

## 5.7 FRM Solution Effort Improvement

As previously discussed, congestion is a significant problem during evacuations; it increases exposure to the evacuation threat, can further endanger evacuees, and encourage noncompliance with evacuation plans and directives. However, within a network flow optimization model, congestion can be beneficial – up to a certain limit. In order to maximize flow through a bottleneck, a network must store enough demand in order to send adequate flow

to maximize the utilization of the bottleneck. Congestion enables this storage upstream of network restrictions and bottleneck(s). Without congestion, there can be intervals of time where adequate demand does not exist to keep the bottleneck flowing to capacity. However, maintaining a fully congested network would be counterproductive since is has been proven that congestion reduces network flows. There must be an equilibrium struck by maintaining just enough congestion in order to efficiently utilize the network and maximize flow.

Mathematical models that utilize the CTM to model network flows do not maintain this equilibrium and do not reduce network flows due to congestion. When using a *min TET* objective function, the CTM will maximize congestion in order to achieve the optimal objective function value (as shown in Section 5.4). However, this logic is flawed. The FRM, by embedding the CTM-FR, penalizes network flows with congestion and therefore discourages group-level staging policies that build excessive amounts of congestion. This is performed through the binary decision variables that indicate the traffic stream conditions - the $z1_i^t, z2_i^t$, and, $z3_i^t$ binary variables, or the $z$-variables for brevity. They were defined in Section 5.3.

The FRM, as a mixed integer formulation, requires a lengthy branch-and-bound process to solve for these binary decision variables. In this process, each variable is relaxed between 0 and 1 and iteratively solved for until binary values are determined. This process can require millions of iterations and often a computer will run out of memory before solving the problem. A future research area would be to extend the problem solving to a cluster of connected to computers to speed this process.

However, if we can make some assumptions about what an evacuation plan should look like, we can set some of the binary variables ahead of time and drastically improve the solution effort time of the FRM. By setting the $z$-variables to zero where congestion is not desired during the evacuation, we can influence both the optimization effort and the evacuation plan. This gives a transportation network planner some flexibility in the plan development by developing a staging plan that controls network demand in order to avoid congestion at predetermined areas.

For the purposes of solving the example evacuation networks in this research, we chose to set the $z$-variables to 0 in ordinary cells along the freeway; these we also defined as the no-holding cells for the purposes of controlling traffic holding, set $noH$. However, any set of cells or roadway segments could be chosen. By prohibiting congestion from the $noH$

cells, we allow congestion to form only in the cell immediately upstream of the restrictions and network bottleneck. This stores sufficient demand and efficiently utilizes the bottleneck capacity. Constraint (5.40) was added to the FRM to establish this behavior and improve the solution effort.

$$z1_i^t = 0, \quad z2_i^t = 0, \quad z3_i^t = 0, \ \forall i \in noH, \ t = 1, \cdots, T \tag{5.40}$$

In Section 5.4.1.1, we solved the FRM with the *min TET* objective function for Network 1. To demonstrate the validity of Constraint (5.40), when it was included in the constraint set, the solution was obtained within 2 seconds. When the constraint was removed, the model solved within 6% optimality within 7 hours. In the solution, none of the $z$-variables in set $noH$ were set to 1. So by setting these variables to 0 with Constraint (5.40), we controlled the congestion behavior and reduced the solution time considerably, for this particular example.

## 5.8 Comparison of Group-level Staging to Simultaneous Evacuation

We close this paper with a demonstration of the value and benefit of using group-level staging to generate an evacuation plan. By using the FRM and the additional constraints introduced in Section 5.6, we can produce network flows that are most implementable and can drastically improve current evacuation plans.

By means of a comparison, we show the benefit of the FRM formulation; the results from a simultaneous evacuation are compared to those results from a staged evacuation as produced by the FRM. We return to Network 1 as introduced in Section 5.4.1.1, but with different network loading parameters. These new parameters are listed in Table 5.11. The rest of the input parameters are the same as used in Section 5.4.1.1.

From a demand management perspective, there are very few decisions to make once the evacuation order has been issued in a simultaneous evacuation. To create comparable results, we used MATLAB to simulate the traffic flow logic of the CTM-FR. The FRM was used to create an evacuation plan with group-level staging. We used fixed merge priority in both the MATLAB simulation and in the FRM – 75% priority to the freeway and 25% priority to the on-ramps at both merge intersections in Network 1. The formulation of the

Table 5.11: Network 1 input parameters for Section 5.8

| **Demand** | | | |
|---|---|---|---|
| *O1* | *O2* | *O3* | **Total** |
| 800 | 300 | 300 | 1400 |
| **Loading Curve Parameters** | | | |
| $T$ | $W$ | $\alpha$ | $\beta$ |
| 75 | 45 | 4 | 4 |

FRM used was Run 4 as discussed in Section 5.6 (with Constraints (5.33) - (5.36), (5.39), and without Constraints (5.3) - (5.6)). The *min TET* objective function was used.

The results of this comparison are listed in Table 5.12 and the network flows out of the network bottleneck are shown in Figure 5.19.

Table 5.12: Comparison of evacuation metrics between simultaneous and staged evacuation

| | *SE* | *FRM w/ Staging* |
|---|---|---|
| **TET** | 59,502 | 48,150 |
| **TTT** | 25,888 | 9,050 |
| **NCT** | 84 | 66 |
| **Staging** | | |
| *O1* | 1 | 1 |
| *O2* | 1 | 19 |
| *O3* | 1 | 6 |

By using group-level staging, the improvements over the status quo for demand management is impressive. Even for this small example, the objective function was improved by 24%. By staging the demand loaded from the two arterial origins (O2 and O3) by 23 time intervals, the NCT value improved by 27% and congestion was reduced dramatically.

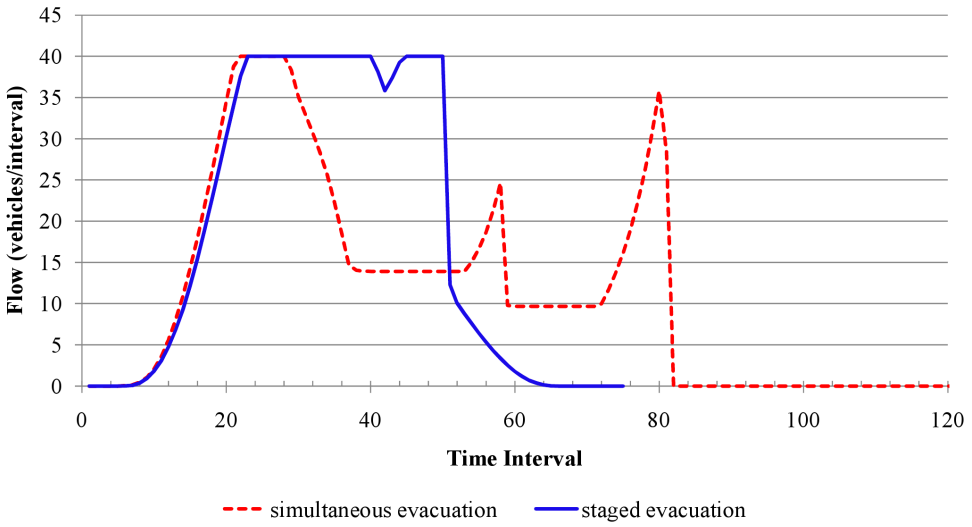



Figure 5.19: Comparison of network flows between simultaneous and staged evacuation

## 5.9 Conclusions and Future Research Directions

The use of both supply and demand management strategies can drastically improve evacuation plans and reduce the levels of congestion that further threaten evacuees. This paper has developed a network flow optimization model that employs a group-level staging strategy to dramatically reduce congestion and shorten the duration of an evacuation.

The optimization model, known as the Flow Reduction Model (FRM), was first introduced in Bish et al. (2011a). The model made two significant contributions: 1) the development of a macroscopic traffic flow model that reduces network flows at the onset of congestion (the CTM-FR), and 2) the restriction of the "traffic holding" modeling phenomenon to intersections (only) where the modeling behavior can be reproduced by intersection traffic control measures. In this paper, we introduced some additional demand management constraints to allow the FRM to develop an evacuation plan that is an improvement over simultaneous evacuation. The behavior of different objective functions within the FRM framework were explored and along with considerations to make when prioritizing certain evacuation metrics such as evacuation time, travel time, and network clearance time. An important conclusion made was that minimizing evacuation time and network clearance time are not equivalent when using the FRM; this is due to the realistic congestion modeling, the use of fixed demand loading curves, and limits placed on traffic holding within the model. Additionally, it was shown that a lexicographic objective function can be used to combine some of the benefits

of the objective functions.

We also demonstrated how the FRM can be used to efficiently size origin control groups. This is an important tool when attempting to understand how best to stage an evacuation. The limited ability of the FRM to use traffic holding to improve solutions was also shown and then completely removed with additional intersection constraints. Two sets of constraints were used to introduce either a fixed priority or a range of priorities that replicate potential traffic control measures. Although these constraints worsen the objective solution and increase computational effort, the results are more realistic and implementable.

The paper concluded with a comparison of two evacuation plans and displayed the clear benefits of using group-level staging within a realistic network flow optimization model. This research has inspired further research in the following areas:

**Time Dependent Demand Loading Curves:** Evacuation metrics could be further improved through the use of group-level staging with the FRM if the demand loading curves were dependent upon the staging solutions. For instance, if an origin control group was staged 30 time intervals after another origin control group, the former group could assumably use a smaller evacuation time window (parameter $W$ in this paper). This would create more prescriptive evacuation plans but could further reduce evacuation duration.

**Evacuation Plan Compliance:** The FRM framework could be altered to answer the following question, "What if only 50% of the population follow the group-level staging plan?" The assumption is that non-compliance would increase congestion and evacuation duration but should also out perform simultaneous evacuation plans; further research could help to show what range of compliance is required.

**Simulation-Optimization:** We need to further evaluate FRM group-level staging solutions within microscopic simulation, INTEGRATION, in order to ensure that the evacuation plan is an improvement over simultaneous evacuation. This approach could also inspire further enhancements to the FRM framework.

# Chapter 6

# Conclusions and Directions for Future Research

In this dissertation we explored the use of network flows over time for evacuation planning. The research was primarily focused on improving regional evacuations conducted by automobiles for short-notice evacuations although the modeling framework, insights, and conclusions can also be applied to no-notice evacuations. Past hurricane events that have hit the Atlantic and Gulf coasts of the United States have plagued transportation systems with heavy traffic congestion. Congestion amplifies the threat to evacuees - it can expose evacuees to the very threat they are trying to flee from and it can create other hazards along the evacuation route such as vehicle accidents and medical issues. Our research explored the effects of congestion within transportation networks and formulated an improved network flow model that incorporates these effects and creates outputs in the form of decision variables that can be implemented within an evacuation plan. In addition to realistic modeling, our model presents an evacuation demand management strategy that can reduce congestion within a transportation network as a direct result of an improved evacuation plan. This work, as it has been presented, is not yet in the form that can be used across an entire transportation network for a city or region but is intended to demonstrate the advantages of our modeling framework, the Flow Reduction Model (FRM).

This dissertation has explored three areas of interrelated research. In Chapter 3, we used a microscopic traffic simulation software package, INTEGRATION, to study of the contributing factors that form congestion. Through this study we made several important

contributions:

- We showed that INTEGRATION can model the capacity drop phenomenon and can be used to demonstrate the effects of congestion on queue discharge flows.
- The study showed that by using heavy vehicle percentages found in empirical studies, modeling specific vehicle dynamics, and using realistic acceleration rates that significant queue discharge flow reductions are produced as a result of congestion within the network. These flow reductions are consistent with several freeway empirical studies performed both in the United States and in Canada.
- The simulation study additionally proved that onset of congestion is probabilistic in nature and generally follows the normal distribution.

By validating INTEGRATION to reliably reproduce the capacity drop phenomenon and reductions in queue discharge flow at the onset of congestion, we were able to approximate these results with the development of a dynamic network flow optimization model in Chapter 4. The primary contributions of this chapter are:

- The Cell Transmission Model with Flow Reductions (CTM-FR) was developed as a macroscopic traffic flow model with a deterministic parameter, $\Omega_i$, that reduces flow as congestion builds within a network. The CTM-FR produces the results of the original CTM when $\Omega_i = Q_i$; the CTM-FR is a more general macroscopic flow model that can reproduce a wide range of results depending on the level of congestion where $\Omega_i \in (0, Q_i]$.
- The Flow Reduction Model (FRM) is a time-expanded network flow optimization model that incorporates the CTM-FR within a mixed integer framework to reduce the level of traffic holding common to many mathematical programs used for dynamic traffic assignment.
- Constraints were proposed to reduce the branch-and-bound computations required to solve the mixed integer formulation of the FRM. These constraints take advantage of the network relationships between the cells and links and the behavior of congestion.

In Chapter 5, we augmented the constraints of the FRM for use in evacuation planning. A demand management strategy was used within the FRM formulation to create evacuation

plans that seek to avoid congestion and reduce the duration of an evacuation. The primary research contributions of the final chapter are:

- A group-level staging strategy was developed where specific groups within a region are issued separate evacuation orders to start their respective evacuations; i.e., we evacuate the groups in stages. The evacuations are modeled by a sigmoid curve determined by a predetermined evacuation window and evacuee response parameters. Group-level staging constraints were developed and were added to the FRM formulation.
- The FRM provides optimal evacuation plans that out perform common simultaneous evacuations when considering common metrics such as evacuation time, travel time, and network clearance time.
- The FRM can be further improved to completely remove the "traffic holding" modeling phenomenon and better model intersection flow controls.
- The objective functions to minimize total evacuation time (TET) and network clearance time (NCT) are not equivalent within the FRM due to modeling of congestion, the use of loading curves, and better intersection controls within the FRM.
- The FRM can also indicate the appropriate number of control groups needed to most efficiently stage and load demand onto an evacuation network.

## 6.1 Future Research Directions

These research contributions will aid in extending the current body of research concerning network flow optimization applied to evacuation planning. However, there are a few areas for improvement to continue this important research.

- The capacity drop phenomenon should be further investigated within INTEGRATION for larger networks consisting of merging, diverging, and weaving sections. This study would improve the approximation of the effects of congestion within the FRM.
- The FRM requires more computational effort to solve than linear programs that incorporate just the CTM for network flow optimization. A more efficient formulation of

the FRM is needed to reduce the computational effort and expand the practical use of the FRM. Additionally, other solution techniques could be employed to solve the FRM such as column generation or other decomposition techniques.

- Further development of group-level routing strategies is required. However, not until a more efficient formulation of the FRM is developed since adding additional routing binary decision variables would further increase the computational effort required to solve the FRM.
- The evaluation of FRM solutions should be conducted within INTEGRATION. Insights acquired through the analysis of the simulation results could be used to refine the FRM formulation.
- The inputs into the FRM are entirely deterministic. The performance of the FRM should be evaluated with stochastic model inputs and a robust optimization formulation developed.
- In order to solve the FRM and other dynamic network flow optimization models, future research efforts should explore the use of cluster computing, such as at the Network Dynamics and Simulation Science Laboratory (NDSSL) or similar research groups at Virginia Tech.

The use of demand management strategies will be crucial in the development of evacuation plans in the future. The southeastern coastal regions of the United States are growing, the ownership and use of automobiles is increasing, and severe weather seems to be increasing. These factors will create additional problems with congestion and could make future evacuations even more hazardous. However, our technology and communication infrastructure systems are also improving. Soon, we will be able to issue customized evacuation plans to regions, neighborhoods, and perhaps individual households. With increased capability to monitor conditions during an evacuation and better predict hazard impact locations and times, demand management strategies will be required to adjust and re-issue evacuation plans. It is our hope the modeling framework presented in this research can assist in the future development of these strategies to dramatically improve the state of evacuation planning.

# Bibliography

# Bibliography

AASHTO (1954). "A Policy on Geometric Design of Rural Highways." *Technical report*, AASHTO:American Association of State Highway and Transportation Official, Washington, DC.

Abkowitz, M. and Meyer, E. (2007). "Technological Advancements in Hazardous Materials Evacuation Planning." *Transportation Research Record: Journal of the Transportation Research Board*, **Vol 1522, Issue 1996**, pp. 116–121.

Ahn, K., Rakha, H., and Trani, A. (2004). "Microframework for Modeling of High-Emitting Vehicles." *Transportation Research Record: Journal of the Transportation Research Board*, **1880**, pp. 39–49.

Ahn, K., Rakha, H., Trani, A., and Aerde, M.V. (2001). "Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels." *Journal of Transportation Engineering*, **128(2)**, pp. 182–190.

Alsnih, R. and Stopher, P.R. (2004). "Review of Procedures Associated with Devising Emergency Evacuation Plans." *Transportation Research Record: Journal of the Transportation Research Board*, **1865**, pp. 89–97.

Avgoustis, A., Aerde, M.V., and Rakha, H. (2004). "Framework For Estimating Network-wide Safety Impacts Of Intelligent Transportation Systems." In *Intelligent Transportation Systems Safety and Security Conference*. Miami, FL.

Banks, J.H. (1991). "Two-capacity Phenomenon at Freeway Bottenecks: A Basis for Ramp Metering?" *Transportation Research Record*, **1320**, pp. 83–90.

Banks, J.H. (1990). "Flow Processes at a Freeway Bottleneck." *Transportation Research Record*, **1287**, pp. 20–28.

Bertini, R.L. and Leal, M.T. (2005). "Empirical Study of Traffic Features at a Freeway Lane Drop." *Journal of Transportation Engineering*, **131:6**, pp. 397–407.

Bish, D.R., Chamberlayne, E.P., and Rakha, H.A. (2011a). "Optimizing Network Flows with Congestion-based Flow Reductions." *manuscript*, pp. 1–30.

Bish, D.R. and Sherali, H.D. (2011). "Aggregate-level Demand Management in Evacuation Planning." *manuscript*, pp. 1–33. Virginia Polytechnic and State University; Grado Department of Industrial and Systems Engineering.

Bish, D.R., Sherali, H.D., and Hobeika, A.G. (2011b). "Optimal Evacuation Strategies using Staging and Routing." *manuscript*, pp. 1–21. Manuscript. Grado Department of Industrial and Systems Engineering and Charles E. Via Department of Civil Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Carey, M. (1987). "Optimal Time-Varying Flows on Congested Networks." *Operations Research*, **35 (1)**, pp. 58–69.

Cassidy, M.J. and Bertini, R.L. (1999). "Some Traffic Features at Freeway Bottenecks." *Transportation Research Part B: Methodological*, **3:1**, pp. 25–42.

Cassidy, M.J. and Rudjanakanoknad, J. (2005). "Increasing the capacity of an isolated merge by metering its on-ramp." *Transportation Research Part B: Methodological*, **39**, pp. 896 – 913.

Chamberlayne, E.P., Rakha, H.A., and Bish, D.R. (2011a). "Modeling the Capacity Drop Phenomenon at Freeway Bottlenecks using the INTEGRATION Software." *manuscript*, pp. 1–29. Manuscript. Grado Department of Industrial and Systems Engineering and Charles E. Via Department of Civil Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Chamberlayne, E.P., Rakha, H.A., El-Metwally, M., and Bish, D.R. (2011b). "Simulation Study of Freeway Bottleneck Flow Reductions at the onset of Congestion." In *Transportation Research Board 90th Annual Meeting, Jan 23 - 27 2011, [Paper 11-1188].*

Chen, X. and Zhan, F.B. (2004). "Agent-Based Modeling and Simulation of Urban Evacuation: Relative Effectiveness of Simultaneous and Staged Evacuation Strategies." In *Presented at the 83rd Annual Meeting of the Transportation Research Board*. Washington, DC.

Chen, X. and Zhan, F. (2008). "Agent-Based Modeling and Simulation of Urban Evacuation: Relative Effectiveness of Simultaneous and Staged Evacuation Strategies." *Journal of Operational Research Society*, **59**, pp. 25–33.

Chiu, Y., Zheng, H., Villalobos, J., and Gautam, B. (2007). "Modeling No-choice Mass Evacuation Using a Dynamic Traffic Flow Optimization Model." *IIE Transactions*, **39**, pp. 83–94.

Chung, K., Rudjanakanoknad, J., and Cassidy, M.J. (2007). "Relation between traffic density and capacity drop at three freeway bottlenecks." *Transportation Research Part B: Methodological*, **41**, pp. 82–95.

Church, R.L. and Cova, T.J. (2000). "Mapping Evacuation Risk on Transportation Networks Using a Spatial Optimization Model." *Transportation Research - C*, **8**, pp. 321–336.

Cova, T.J. and Johnson, J.P. (2003). "A Network Flow Model for Lane-Based Evacuation Routing." *Transportation Research - A*, **37**, pp. 579–604.

Daganzo, C.F. (1994). "The Cell Transmission Model - A Dynamic Representation Of Highway Traffic Consistent With The Hydrodynamic Theory." *Transportation Research Part B-Methodological*, **28 (4)**, pp. 269–287.

Daganzo, C.F. (1995). "The Cell Transmission Model, Part 2: Network Traffic." *Transportation Research Part B-Methodological*, **29 (2)**, pp. 79–93.

Demarchi, S. (2002). "A new formulation for Van Aerde's speed-flow-density relationship (in portuguese)." In *XVI Congresso De Pequisa e Ensino em Transportes*. Natal, RN, Brazil.

Dion, F., Rakha, H., and Kang, Y.S. (2004). "Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections." *Transportation Research Part B: Methodological*, **38(2)**, pp. 99–122.

Dixit, V.V., Ramasamy, S., and Radwan, E. (2008). "Assessment of I-4 Contraflow Plans: Microscopic Versus Mesoscopic Simulation." *Transportation Research Record: Journal of the Transportation Research Board*, **No. 2041**, pp. 89–97.

Drew, D.R. (1968). *Traffic Flow Theory and Control*. McGraw-Hill Book Company.

Edie, L.C. (1961). "Car-Following and Steady-State Theory for Noncongested Traffic." *Operations Research*, **9, No. 1**, pp. 66–76.

Farzaneh, M. and Rakha, H. (2006). "Impact of Differences in Driver-Desired Speed on Steady-State Traffic Stream Behavior." *Transportation Research Record: Journal of the Transportation Research Board*, **1965**, pp. 142–151.

Fu, H. and Wilmot, C.G. (2004). "A Sequential Logit Dynamic Travel Demand Model for Hurricane Evacuation." *Transportation Research Record*, **1882**, pp. 19—-26.

Fu, H., Wilmot, C., , and Baker, E. (2007). "Modeling the Hurricane Evacuation Response Curve." *Transportation Research Record: Journal of the Transportation Research Board*, **2022**, pp. 94–102.

Gomes, G. and Horowitz, R. (2006). "Optimal freeway ramp metering using the asymmetric cell transmission model." *Transportation Research Part C*, **14**, pp. 244–262.

Hall, F.L. and Agyemang-Duah, K. (1991). "Freeway Capacity Drop and the Definition of Capacity." *Transportation Research Record*, **1320**, pp. 91–98.

Han, L.D., Yuan, F., and II, T.U. (2007). "What is an Effective Evacuation Operation?" *Journal of Urban Planning and Development*, **133(1)**, pp. 3–8.

Hardy, M. and Wunderlich, K. (2007). "Evacuation Management Operations (EMO) Modeling Assessment: Transportation Modeling Inventory." *Technical report*, Research and Innovative Technology Administration, ITS Joint Program Office.

Hobeika, A. and Jamei, B. (1985). "MASSVAC: A model for calculating evacuation times under natural disasters." In J.M. Carroll, editor, *Proceedings of the Conference on Emergency Planning, Emergency Planning, Simulation Series*, volume 15. pp. 23–28.

Jarvis, J. and Ratliff, D. (1982). "Some Equivalent Objectives for Dynamic Network Flow Problems." Management Science. **28**, 106–109.

Kalafatas, G. and Peeta, S. (2006). "An Exact Graph Structure for Dynamic Traffic Assignment: Formulation, Properties, Computational Experience." In *86th Annual Meeting of the Transportation Research Board*.

Kerner, B.S. and Rehborn, H. (1996). "Experimental properties of complexity in traffic flow." *Physical Review E*, **53(5)**, pp. R4275–R4278.

Kerner, B.S. and Klenov, S.L. (2006). "Probabilistic Breakdown Phenomenon at On-Ramp Bottlenecks in Three-Phase Traffic Theory." *Transportation Research Record: Journal of the Transportation Research Board*, **1965**, pp. 70–78.

Kerner, B.S., Klenov, S.L., and Wolf, D.E. (2002). "Cellular automata approach to three-phase traffic theory." *Journal of Physics A: Mathematical and General*, **35:47**, pp. 9971–10,013.

Kerner, B. (2008). "On-Ramp Metering Based on Three-Phase Traffic Theory Downstream Off-Ramp and Upstream On-Ramp Bottlenecks." *Transportation Research Record: Journal of the Transportation Research Board*, **2088**, pp. 80–89.

Laval, J.A. and Daganzo, C.F. (2006). "Lane-changing in traffic streams." *Transportation Research Part B: Methodological*, **40**, pp. 251 – 264.

Li, Y., Waller, S., and Ziliaskopoulos, A. (2003). "A Decomposition Scheme for System Optimal Dynamic Traffic Assignment Models." *Networks & Spatial Economics*, **3**, pp. 441–455.

Lighthill, M. and Whitham, G. (1955). "On Kinematic Waves. I. Flood Movement in Long Rivers." *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences (1934-1990)*, **229**, pp. 281–316.

Lim, Y.Y. (2003). *Modeling And Evaluating Evacuation Contraflow Termination Point Designs*. Master's thesis, Louisiana State University.

Lin, W.H. and Wang, C. (2004). "An Enhanced 0-1 Mixed-Integer LP Formulation for Traffic Signal Control." *IEEE Transactions on Intelligent Transportation Systems*, **5, No. 4**, pp. 238–245.

Litman, T. (2006). "Lessons from Katrina and Rita: What Major Disasters Can Teach Transportation Planners." *Journal of Transportation Engineering*, **132 (1)**, pp. 11–18.

Liu, Y., Lai, X., and Chang, G.L. (2006). "Cell-Based Network Optimization Model for Staged Evacuation Planning Under Emergencies." *Transportation Research Record*, **1964**, pp. 127–135.

Liu, Y., Chang, G.L., Liu, Y., and Lai, X. (2008). "Corridor-Based Emergency Evacuation System for Washington, D.C.: System Development and Case Study." *Transportation Research Record: Journal of the Transportation Research Board*, **2041:2008**, pp. 58–67.

Lo, H.K. (2001). "A Cell-Based Traffic Control Formulation: Strategies and Benefits of Dynamic Timing Plans." *Transportation Science*, **35, No. 2**, pp. 148–164.

May, A.D. (1990). *Traffic Flow Fundamentals*. Prentice Hall.

Merchant, D.L. and Nemhauser, G.L. (1978a). "Optimality Conditions for a Dynamic Traffic Assignment Model." *Transportation Science*, **12:3**, pp. 200–207.

Merchant, D. and Nemhauser, G. (1978b). "A Model and an Algorithm for the Dynamic Traffic Assignment Problem." *Transportation Science*, **12:3**, pp. 183–199.

Mitchell, S.W. and Radwan, E. (2006). "Heuristic Priority Ranking of Emergency Evacuation Staging to Reduce Clearance Time." *Transportation Research Record: Journal of the Transportation Research Board*, **1964**, pp. 219–228.

Nagel, K., Wagner, P., and Woesler, R. (2003). "Still Flowing: Approaches to Traffic Flow and Traffic Jam Modeling." *Operations Research*, **51(5)**, pp. 681–710.

Papageorgiou, M., Papamichail, I., Spiliopoulou, A., and Lentzakis, A. (2008). "Real-time merging traffic control with applications to toll plaza and work zone management." *Transportation Research Part C: Emerging Technologies*, **16**, pp. 535 – 553.

Papageorgiou, M. (1998). "Some Remarks on Macroscopic Traffic Flow Modelling." *Transportation Research - A*, **32:5**, pp. 323–329.

Peeta, S. and Ziliaskopoulos, A. (2001). "Foundations of Dynamic Traffic Assignment: The Past, the Present, and the Future." *Networks and Spatial Economics*, **1:2001**, pp. 233–265.

Persaud, B., Yagar, S., and Brownlee, R. (1998). "Exploration of the Breakdown Phenomenon in Freeway Traffic." *Transportation Research Record: Journal of the Transportation Research Board*, **1634**, pp. 64–69.

Rakha, H., Aerde, M.V., Bloomberg, L., and Huang, X. (1998). "Construction and Calibration of a Large-Scale Microsimulation Model of the Salt Lake Area." *Transportation Research Record: Journal of the Transportation Research Board*, **1644**, pp. 93–102.

Rakha, H. and Crowther, B. (2003). "Comparison and Calibration of FRESIM and INTEGRATION Steady-state Car-following Behavior." *Transportation Research - Part A: Policy and Practice*, **37:1**, pp. 1–27. **37**, 1–27.

Rakha, H. (2009a). "CEE5604: Traffic Characteristics & Flow: Car-Following Models." Lecture Notes.

Rakha, H. (2009b). "CEE5604: Traffic Characteristics & Flow: Macroscopic Traffic Flow Behavior - Traffic Stream Models." Lecture Notes.

Rakha, H., Aerde, M.V., Ahn, K., and Trani, A.A. (2000a). "Requirements for Evaluating Traffic Signal Control Impacts on Energy and Emissions Based on Instantaneous Speed and Acceleration Measurements." *Transportation Research Record: Journal of the Transportation Research Board*, **1738**, pp. 56–67.

Rakha, H., Ahn, K., and Trani, A. (2004). "Development of VT-Micro model for estimating hot stabilized light duty vehicle and truck emissions." *Transportation Research Part D: Transport and Environment*, **9(1)**, pp. 49–74.

Rakha, H. and Crowther, B. (2002). "Comparison of Greenshields, Pipes, and Van Aerde Car-Following and Traffic Stream Models." *Transportation Research Record*, **1802 (Paper No. 02-2143)**, pp. 248–262.

Rakha, H., Flintsch, A.M., Ahn, K., El-Shawarby, I., and Arafeh, M. (2005). "Evaluating Alternative Truck Management Strategies Along Interstate 81." *Transportation Research Record: Journal of the Transportation Research Board*, **1925**, pp. 76–86.

Rakha, H., Kang, Y.S., and Dion, F. (2001). "Estimating Vehicle Stops at Undersaturated and Oversaturated Fixed-Time Signalized Intersections." *Transportation Research Record: Journal of the Transportation Research Board*, **1776**, pp. 128–137.

Rakha, H., Pasumarthy, H., and Adjerid, S. (2009). "A Simplified Behavioral Vehicle Longitudinal Motion Model." *Transportation Letters: The International Journal of Transportation Research*, **1(2)**, pp. 95–110.

Rakha, H.A. (1991). *An Evaluation of the Benefits of User and System Optimised Route Guidance Strategies*. Master's thesis, Queen's University. M.Sc., Civil Engineering.

Rakha, H.A. and Lucic, I. (2002). "Variable Power Vehicle Dynamics Model for Estimating Truck Accelerations." *Journal of Transportation Engineering*, **128(5)**, pp. 412–419.

Rakha, H.A., Medina, A., Sin, H., Dion, F., Aerde, M.V., and Jenq, J. (2000b). "Traffic Signal Coordination Across Jurisdictional Boundaries: Field Evaluation of Efficiency, Energy, Environmental, and Safety Impacts." *Transportation Research Record: Journal of the Transportation Research Board*, **1727**, pp. 42–51.

Rakha, H.A. and Zhang, Y. (2004). "INTEGRATION 2.30 framework for modeling lane-changing behavior in weaving sections." *Transportation Research Record: Journal of the Transportation Research Board*, **1883**, pp. 140–149.

Rakha, H.A. and Zhang, Y. (2006). "Analytical Procedures for Estimating Capacity of Freeway Weaving, Merge, and Diverge Sections." *Journal of Transportation Engineering*, **132:8**, pp. 618–628.

Richards, P.I. (1956). "Shock Waves On The Highway." *Operations Research*, **Vol. 4 Issue 1**, pp. 42–51.

Sbayti, H. and Mahmassani, H.S. (2006). "Optimal Scheduling of Evacuation Operations." *Transportation Research Record: Journal of the Transportation Research Board*, **1964**, pp. 238–246. *TRB 2006 Annual Meeting*, Washington D.C.

Schönhof, M. and Helbing, D. (2007). "Empirical Features of Congested Traffic States and Their Implications for Traffic Modeling." *Transportation Science*, **41(2)**, pp. 135–166.

Sheffi, Y., Mahmassani, H., and Powell, W.B. (1982). "A Transportation Network Evacuation Model." *Transportation Research Part A: General*, **Vol 16, Issue 3**, pp. 209–218. **16**(3), 209–218.

Shen, W., Nie, Y., and Zhang, H.M. (2007). "Dynamic Network Simplex Method for Designing Emergency Evacuation Plans." *Transportation Research Record: Journal of the Transportation Research Board*, **2022**, pp. 83–93.

Sherali, H.D., Carter, T.B., and Hobeika, A.G. (1991). "A Location-Allocation Model and Algorithm for Evacuation Planning under Hurricane/Flood Conditions." *Transportation Research Part B: Methodological*, **25B:6**, pp. 439–452. **25**(6), 439–452.

Snare, M.C. (2002). *Dynamics Model for Prediction Maximum and Typical Acceleration Rates of Passenger Vehicles.* Master's thesis, Charles E. Via, Jr. Department of Civil and Environmental Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Treiber, M., Kesting, A., and Helbing, D. (2006). "Understanding widely scattered traffic flows, the capacity drop, and platoons as effects of variance-driven time gaps." *Phys. Rev. E*, **74 (1)**, pp. 1–10.

Tuydes, H. and Ziliaskopoulos, A. (2006). "Tabu-based Heuristic Approach for Optimization of Network Evacuation Contraflow." *Transportation Research Record*, **1964**, pp. 157–168.

Tuydes, H. and Ziliaskopoulos, A.K. (2004). "Network Redesign to Optimize Evacuation Contraflow." In *83rd Annual Meeting of the Transportation Research Board*.

Ukkusuri, S.V. and Waller, S.T. (2008). "Linear Programming Models for the User and System Optimal Dynamic Network Design Problem: Formulations, Comparisons and Extensions." *Networks and Spatial Economics*, **8:383**, p. 406.

USDOT (2008). "Highway Statistics 2008." *Technical report*, Federal Highway Administration.
URL http://www.fhwa.dot.gov/policyinformation/statistics/2008/ps1.cfm

USGS (2005). "Hurricane Hazards—A National Threat." www.usgs.gov. Accessed on 5/11/2011.
URL http://pubs.usgs.gov/fs/2005/3121/2005-3121.pdf

Van Aerde, M. (1995). "Single Regime Speed-Flow-Density Relationship for Congested and Uncongested Highways." In *74th TRB Annual Conference*, Paper No 950802. Washington, DC.

Van Aerde, M. and Rakha, H. (1995). "Multivariate Calibration of Single Regime Speed-Flow-Density Relationships." In *Proceedings of the 6th 1995 Vehicle Navigation and Information Systems Conference*. Seattle, WA, pp. 334–341.

Van Aerde, M. and Yagar, S. (1988a). "Dynamic Integrated Freeway/Traffic Signal Networks: A Routing-Based Modelling Approach." *Transportation Research Part A: General*, **22(6)**, pp. 445–453.

Van Aerde, M. and Yagar, S. (1988b). "Dynamic Integrated Freeway/Traffic Signal Networks: Problems and Proposed Solutions." *Transportation Research Part A: General*, **22(6)**, pp. 435–443.

Van Aerde, M. and Rakha, H. (2007a). *INTEGRATION © RELEASE 2.30 FOR WINDOWS: User's Guide - Volume II: Advanced Model Features.* M. Van Aerde & Assoc., Ltd., Blacksburg, VA.

Van Aerde, M. and Rakha, H. (2007b). *INTEGRATION © RELEASE 2.30 FOR WINDOWS: User's Guide – Volume I: Fundamental Model Features.* M. Van Aerde & Assoc., Ltd., Blacksburg, VA.

Williams, B.M., Tagliaferri, A.P., Meinhold, S.S., Hummer, J.E., and Rouphail, N.M. (2007). "Simulation and Analysis of Freeway Lane Reversal for Coastal Hurricane Evacuation." *Journal of Urban Planning and Development*, **133:1**, pp. 61–72.

Wilmot, C.G. and Mei, B. (2004). "Comparison of Alternative Trip Generation Models for Hurricane Evacuation." *Natural Hazards Review*, **Vol 5, Issue 4**, pp. 170–178.

Wolshon, B. (2001). "One-Way-Out: Contraflow Freeway Operation for Hurricane Evacuation." *Natural Hazards Review*, **2(3)**, pp. 105–112.

Wolshon, B., Urbina, E., Wilmot, C., and Levitan, M. (2005). "Review of Policies and Practices for Hurricane Evacuation. I: Transportation Planning, Preparedness, and Response." *Natural Hazards Review*, **Vol 6, Issue 3**, pp. 129–142.

Xie, C., Lin, D.Y., and Waller, S.T. (2009). "A dynamic evacuation network optimization problem with lane reversal and crossing elimination strategies." *Transportation Research Part E*, **article in press**, pp. 1–22.

Yao, T., Mandala, S.T., and Chung, B.D. (2009). "Evacuation Transportation Planning Under Uncertainty: A Robust Optimization Approach." *Networks & Spatial Economics*, **9**, pp. 171–189.

Yazici, M.A. and Ozbay, K. (2008). "Evacuation Modelling in the United States: Does the Demand Model Choice Matter?" *Transport Reviews*, **28**, pp. 757–779.

Ziliaskopoulos, A.K. (2000). "A Linear Programming Model for the Single Destination System Optimal Dynamic Traffic Assignment Problem." *Transportation Science*, **34 (1)**, pp. 37–49.

# Appendix A

# Model Formulations

## A.1 FRM with Group-level Staging

A choice of objective functions: minimize TET, TTT, NCT, or lexicographically NCT and TTT.

$$\text{Minimize} \quad \sum_{t=1}^{T} \sum_{j \in S_o \cup R} \tau x_j^t \tag{A.1}$$

$$\text{Minimize} \quad \sum_{t=1}^{T} \sum_{i \in R} \tau x_i^t \tag{A.2}$$

$$\text{Minimize} \quad \sum_{t=1}^{T} E^t \tag{A.3}$$

$$\text{Minimize} \quad \sum_{t=1}^{T} E^t + \frac{1}{TD} \sum_{t=1}^{T} \sum_{i \in R} \tau x_i^t \tag{A.4}$$

subject to:

$$x_j^t = x_j^{t-1} + \sum_{i:(i,j) \in L} y_{ij}^{t-1} - \sum_{k:(j,k) \in L} y_{jk}^{t-1}, \quad \forall\ j \in C,\ t = 2, \cdots, T \tag{A.5}$$

$$\sum_{j:(i,j) \in L} y_{ij}^t \leq x_i^t, \quad \forall i \in R,\ t = 1, \cdots, T \tag{A.6}$$

$$\sum_{j:(i,j) \in L} y_{ij}^t \leq (N_i - x_i^t) \frac{Q_i - \Omega_i}{N_i - Q_i} + \Omega_i, \quad \forall i \in R,\ t = 1, \cdots, T \tag{A.7}$$

$$\sum_{i:(i,j) \in L} y_{ij}^t \leq Q_j, \quad \forall j \in R,\ t = 1, \cdots, T \tag{A.8}$$

$$\sum_{i:(i,j) \in L} y_{ij}^t \leq \delta_j [N_j - x_j^t], \quad \forall j \in R,\ t = 1, \cdots, T \tag{A.9}$$

$$e1_i^{t+} \leq N_i (z1_i^t), \quad \forall i \in R,\ t = 1, \cdots, T \tag{A.10}$$

$$e1_i^{t-} \leq (N_i \frac{Q_i - \Omega_i}{N_i - Q_i} + \Omega_i)(1 - z1_i^t), \quad \forall i \in R,\ t = 1, \cdots, T \tag{A.11}$$

$$e2_i^{t+} \leq N_i (z2_i^t), \quad \forall i \in R,\ t = 1, \cdots, T \tag{A.12}$$

$$e2_i^{t-} \leq N_i (1 - z2_i^t), \quad \forall i \in R,\ t = 1, \cdots, T \tag{A.13}$$

$$e3_j^{t+} \leq N_i (z3_j^t), \quad \forall i \in R,\ t = 1, \cdots, T \tag{A.14}$$

$$e3_j^{t-} \leq N_i(1 - z3_j^t), \quad \forall i \in R,\ t = 1, \cdots, T \tag{A.15}$$

$$x_i^t - \left\{ (N_i - x_i^t)\frac{Q_i - \Omega_i}{N_i - Q_i} + \Omega_i \right\} = e1_i^{t+} - e1_i^{t-}, \quad \forall i \in R,\ t = 1, \cdots, T \tag{A.16}$$

$$Q_i - \delta_i(N_i - x_i^t) = e3_i^{t+} - e3_i^{t-}, \quad \forall i \in R,\ t = 1, \cdots, T \tag{A.17}$$

$$(x_i^t - e1_i^{t+}) - (Q_j - e3_j^{t+}) = e2_i^{t+} - e2_i^{t-}, \quad \forall (i,j) \in L_O,\ t = 1, \cdots, T \tag{A.18}$$

$$y_{ij}^t = x_i^t - e1_i^{t+} - e2_i^{t+}, \quad \forall (i,j) \in L_O,\ t = 1, \cdots, T \tag{A.19}$$

$$\sum_{j:(i,j)\in L} y_{ij}^t = \sum_{s=1:s\leq t}^{F} \theta_i^s d_i^{t-s+1}, \quad \forall\ i \in S_o,\ t = 1, \cdots, T \tag{A.20}$$

$$\sum_{s=1}^{F} \theta_i^s = 1, \quad \forall\ i \in S_o \tag{A.21}$$

$$z1_i^t = 0, \quad z2_i^t = 0, \quad z3_i^t = 0, \ \forall i \in noH,\ t = 1, \cdots, T \tag{A.22}$$

$$x_i^1 = 0, \quad \forall\ i \in R \cup S_i \tag{A.23}$$

$$x_i^1 = D_i, \quad \forall\ i \in S_o \tag{A.24}$$

$$x_j^t \geq 0, \quad \forall j \in C,\ t = 1, \cdots, T \tag{A.25}$$

$$y_{ij}^t \geq 0, \quad \forall (i,j) \in L,\ t = 1, \cdots, T \tag{A.26}$$

$$e1_i^{t\pm} \geq 0, \quad z1_i^t \in \{0,1\}, \quad \forall (i,j) \in L_O,\ t = 1, \cdots, T \tag{A.27}$$

$$e2_i^{t\pm} \geq 0, \quad z2_i^t \in \{0,1\}, \quad \forall (i,j) \in L_O,\ t = 1, \cdots, T \tag{A.28}$$

$$e3_j^{t\pm} \geq 0, \quad z3_j^t \in \{0,1\}, \quad \forall (i,j) \in L_O,\ t = 1, \cdots, T. \tag{A.29}$$

If Objective Function (A.3) or (A.4) is used, the following constraint is added to the formulation:

$$E^t \geq 1 - \frac{1}{D}\sum_{i\in S_i} x_i^t, \quad t = 1, \cdots, T \tag{A.30}$$

If modeling simultaneous evacuation, replace Constraints (A.20) and (A.21) with the following constraint:

$$\sum_{j:(i,j)\in L} y_{ij}^t = d_i^t, \quad \forall\ i \in S_o,\ t = 1, \cdots, T \tag{A.31}$$

In order to remove traffic holding from the merge intersections, remove Constraints (A.6) - (A.9) and add the following constraints:

$$\sum_{(i,k)\in L_M} (x_i^t - e1_i^{t+}) - (Q_k - e3_k^{t+}) = e2_i^{t+} - e2_i^{t-}, \quad \forall (i,k) \in L_{M,f}, \ t = 1, \cdots, T \tag{A.32}$$

$$e2_i^{t+} = e2f_i^{t+} + e2o_i^{t+}, \quad \forall (i,k) \in L_{M,f}, \ t = 1, \cdots, T \tag{A.33}$$

$$y_{ik}^t = x_i^t - e1_i^{t+} - e2f_i^{t+}, \quad \forall (i,k) \in L_{M,f}, \ t = 1, \cdots, T \tag{A.34}$$

$$y_{jk}^t = x_j^t - e1_j^{t+} - e2o_i^{t+}, \quad \forall (j,k) \in L_{M,o}, \ t = 1, \cdots, T \tag{A.35}$$

To set a range of priority at a merge intersection, add the following constraints:

$$e2o_i^{t+} \geq 0.2e2_i^{t+}, \quad \forall (i,k) \in L_{M,f}, \ t = 1, \cdots, T \tag{A.36}$$

$$e2o_i^{t+} \leq 0.4e2_i^{t+}, \quad \forall (i,k) \in L_{M,f}, \ t = 1, \cdots, T \tag{A.37}$$

Or to set an exact priority at a merge intersection, add the following constraint:

$$e2o_i^{t+} = 0.25e2_i^{t+}, \quad \forall (i,k) \in L_{M,f}, \ t = 1, \cdots, T \tag{A.38}$$

## A.2 System Optimal Linear Program (CTM-LP)

adapted from Ziliaskopoulos (2000)

$$\text{Minimize} \quad \sum_{t=1}^{T} \sum_{j \in R} \tau x_j^t \tag{A.39}$$

subject to:

$$x_j^t = x_j^{t-1} + \sum_{i:(i,j)\in L} y_{ij}^{t-1} - \sum_{k:(j,k)\in L} y_{jk}^{t-1}, \ \forall\ j \in C,\ t = 2, \cdots, T \tag{A.40}$$

$$x_i^1 = 0, \quad \forall\ i \in R \cup S_i \tag{A.41}$$

$$x_i^1 = \sum_i \sum_{t=1}^{T} d_i^t, \quad \forall\ i \in S_o \tag{A.42}$$

$$\sum_{j:(i,j)\in L} y_{ij}^t = d_i^t, \ \forall\ i \in S_o,\ t = 1, \cdots, T \tag{A.43}$$

$$\sum_{k:(j,k)\in L} y_{jk}^t \leq x_j^t, \quad \forall j \in R,\ t = 1, \cdots, T \tag{A.44}$$

$$\sum_{k:(j,k)\in L} y_{jk}^t \leq Q_j, \quad \forall j \in R,\ t = 1, \cdots, T \tag{A.45}$$

$$\sum_{i:(i,j)\in L} y_{ij}^t \leq Q_j, \quad \forall j \in R,\ t = 1, \cdots, T \tag{A.46}$$

$$\sum_{i:(i,j)\in L} y_{ij}^t \leq \delta_j [N_j - x_j^t], \quad \forall j \in R,\ t = 1, \cdots, T \tag{A.47}$$

$$x_j^t \geq 0, \quad \forall j \in C,\ t = 1, \cdots, T \tag{A.48}$$

$$y_{ij}^t \geq 0, \quad \forall (i,j) \in L,\ t = 1, \cdots, T. \tag{A.49}$$

## A.3 CTM-MIP with Group-level Staging

A choice of objective functions: minimize TET, TTT, NCT, or lexicographically NCT and TTT.

$$\text{Minimize} \quad \sum_{t=1}^{T} \sum_{j \in S_o \cup R} \tau x_j^t \tag{A.50}$$

$$\text{Minimize} \quad \sum_{t=1}^{T} \sum_{i \in R} \tau x_i^t \tag{A.51}$$

$$\text{Minimize} \quad \sum_{t=1}^{T} E^t \tag{A.52}$$

$$\text{Minimize} \quad \sum_{t=1}^{T} E^t + \frac{1}{TD} \sum_{t=1}^{T} \sum_{i \in R} \tau x_i^t \tag{A.53}$$

subject to:

$$x_j^t = x_j^{t-1} + \sum_{i:(i,j) \in L} y_{ij}^{t-1} - \sum_{k:(j,k) \in L} y_{jk}^{t-1}, \ \forall \ j \in C, \ t = 2, \cdots, T \tag{A.54}$$

$$\sum_{k:(j,k) \in L} y_{jk}^t \leq x_j^t, \quad \forall j \in R, \ t = 1, \cdots, T \tag{A.55}$$

$$\sum_{k:(j,k) \in L} y_{jk}^t \leq Q_j, \quad \forall j \in R, \ t = 1, \cdots, T \tag{A.56}$$

$$\sum_{i:(i,j) \in L} y_{ij}^t \leq Q_j, \quad \forall j \in R, \ t = 1, \cdots, T \tag{A.57}$$

$$\sum_{i:(i,j) \in L} y_{ij}^t \leq \delta_j [N_j - x_j^t], \quad \forall j \in R, \ t = 1, \cdots, T \tag{A.58}$$

$$x_i^1 = 0, \ \ \forall \ i \in R \cup S_i \tag{A.59}$$

$$x_i^1 = \sum_{i} \sum_{t=1}^{T} d_i^t, \ \ \forall \ i \in S_o \tag{A.60}$$

$$\sum_{j:(i,j) \in L} y_{ij}^t = \sum_{s=1:s \leq t}^{F} \theta_i^s d_i^{t-s+1}, \quad \forall \ i \in S_o, \ t = 1, \cdots, T \tag{A.61}$$

$$\sum_{f=1}^{F} \theta_i^f = 1, \quad \forall\ i \in S_o \tag{A.62}$$

$$\sum_{i \in S_o} \theta_i^1 \geq 1 \tag{A.63}$$

$$x_j^t \geq 0, \quad \forall j \in C,\ t = 1, \cdots, T \tag{A.64}$$

$$y_{ij}^t \geq 0, \quad \forall (i,j) \in L,\ t = 1, \cdots, T. \tag{A.65}$$

If Objective Function (A.52) or (A.53) is used, the following constraint is added to the formulation:

$$E^t \geq 1 - \frac{1}{D} \sum_{i \in S_i} x_i^t, \quad t = 1, \cdots, T \tag{A.66}$$

If modeling simultaneous evacuation, replace Constraints (A.61) and (A.62) with the following constraint:

$$\sum_{j:(i,j) \in L} y_{ij}^t = d_i^t, \quad \forall\ i \in S_o,\ t = 1, \cdots, T \tag{A.67}$$