Network Models in Evacuation Planning

Hussein Ali Tarhini

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

Ebru K. Bish

Raghu Pasupathy

Hesham A. Rakha

June 25, 2014

Blacksburg, Virginia

Keywords: Cell Transmission Model (CTM), dynamic traffic assignment, evacuation planning, hospital evacuation

Network Models in Evacuation Planning

Hussein Ali Tarhini

(ABSTRACT)

This dissertation addresses the development and analysis of optimization models for evacuation planning. Specifically we consider the cases of large-scale regional evacuation using household vehicles and hospital evacuation.

Since it is difficult to estimate the exact number of people evacuating, we first consider the case where the population size is uncertain. We review the methods studied in the literature, mainly the strategy of using a deterministic counterpart, i.e., a single deterministic parameter to represent the uncertain population, and we show that these methods are not very effective in generating a good traffic management strategy. We provide alternatives, where we describe some networks where an optimal policy exist independent of the demand realization, and we propose some simple heuristics for more complex ones.

Next we consider the traffic management tools that can be generated from an evacuation plan. We start by introducing the cell transmission model with flow reduction proposed by Bish et al. (2013). This model captures the flow reduction after the onset of congestion. We then discuss the management tools that can be extracted from this model. We also propose some simplification to the model formulation to enhance its tractability. A heuristic for generating a solution is also proposed, and its solution quality is analyzed.

Finally, we discuss the hospital evacuation problem where we develop an integer programming

model that integrates the building evacuation with the transportation of patients. The impact of building evacuation capabilities on the transportation plan is investigated through the case of a large regional Hospital case study. We also propose a decomposition scheme to improve the tractability of the integer program.

Acknowledgments

I thank my advisor, Dr Douglas Bish, for his outstanding support and guidance. His patience, work ethic and encouragement has been outstanding on both professional and personal levels.

I thank my committee members, Dr. Ebru Bish, Dr. Hesham Rakha and Dr. Raghu Pasupathy for their helpful suggestions that has enhanced the quality of this dissertation. I am particularly grateful for Dr. Ebru Bish for her support throughout my graduate studies. I have also greatly benefited from my previous research with her.

I thank my former advisor, Dr. Bacel Maddah for introducing me to Operations Research, and encouraging me to come to the US. to continue my graduate studies.

I thank my friends at Virginia Tech, especially Hadi El-amine, for making my graduate studies worthwhile.

I thank my family, especially my mom and dad, for their love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Over the past several decades, there is an apparent trend of increasing frequency and intensity of disasters (Newkirk, 2001). As reported by the Federal Emergency Management Agency (FEMA, 2013), the number of declared disasters has risen both in the U.S. and worldwide. Examples of large scale disasters include Hurricane Sandy in 2012, Hurricane Karina in 2005, the Haiti earthquake in 2010, and the Tohuku earthquake in 2012. The discrepancy between the population growth and the infrastructure capacity to safely evacuate the population is a major challenge facing evacuation management. As a result, disaster management problems such as mass evacuation of threatened area, supply chain for relief efforts and hospital evacuations are relevant problems that are being studied in the literature.

A mass evacuation is mainly accomplished by the use household vehicles. Most dynamic flow models assume full control of vehicles in order to route them to safer areas. Optimization in this context can be used to produce routing strategies in an evacuation plan. An optimization can determine the best use of the available resources to generate solutions. However, the solution generated by an optimization model has to be a feasible set of decision rules that can be implemented in practice, based on the behavior of the drivers, existing traffic controls and within the operating budget. The optimization models should also be solved in a reasonable amount of time so they can be solved in real time before an event of disaster occurs. The model used should also model accurately the traffic flow to appropriately reflect the conditions during the evacuation process; specifically congestion during an evacuation can severely reduce the performance of an evacuation plan, the model used should be able to model the reduction of flow due to congestion. Also determining the exact number of people that are evacuating is not possible in some scenarios. In these cases, the model should be able to capture the uncertainty of demand in the evacuation process, and some tools need to be developed to tackle this problem.

Another evacuation planning problem is the special case of evacuating of hospital facilities. There were 275 reported evacuation incidents from 1971-1999, where more than 50% of evacuations occurred because of hazards originating in the hospital itself or from human intruders (Sternberg, 2004). This problem is different than the regional evacuation problem described earlier. Each type of patients require a different evacuation process, and may require different type of equipments. In addition, the evacuation process does not end when the hospital is clear, but must also include transportation of patients to alternative hospital facilities. Mathematical programming formulations for hospital evacuations has been developed to minimize the risk associated with the hospital evacuation process (Bish et al., 2011). However, this model does not integrate the building evacuation with the transportation process. A new formulation should include both processes in order for it to be applicable in practice. For this new model, computational tractability needs to be ensured for practical use. Also, special structures of this model should be utilized to potentially improve the solution performance.

1.2 Summary of research contributions

The research contributions of this dissertation are in two main areas, the first is the improvement of methodologies and models appropriate for large-scale regional evacuation planning. The second is an improved model and appropriate solution methodology for hospital evacuation planning.

The cell transmission model (CTM), a popular macroscopic traffic flow model introduced by Daganzo (1994, 1995) is widely used in transportation applications related to traffic flow management. In this dissertation, I make the following research contributions related to the CTM:

1. In the case where the demand is uncertain, we show that using deterministic counterpart strategies do not produce good solutions; We show this by studying a network structure studied in the literature where there exists an optimal evacuation policy that is independent on the demand realization.

2. For more complex networks, we show that simple heuristics can outperform the deterministic counterpart strategies.

3. Based on the general cell transmission model developed by Bish et al. (2013), and the System Optimal Dynamic Traffic Assignment by Ziliaskopoulos (2000), we discuss the traffic management strategies under the Cell Transmission Model with flow reductions and we compare it to that of the CTM.

4. The cell transmission model with flow reduction is a complex model. Bish et al. (2013) proposed incorporated this model within an integer linear program however the formulation is not efficient and cannot be used for bigger sized networks. We propose some simplifications to the formulation to improve its tractability. We also develop a heuristic that can generate a feasible high quality solution in short time.

In the second part of the dissertation, the following contributions has been made to the hospital evacuation problem:

5. Based on a realistic problem statement, we develop a hosptial evacuation model. The two important steps in a hospital evacuation are the building evacuation and transportation of patients. The resources available are the transportation staff to move the patients from their wards to the staging area and the ambulance fleet size.

6. Since this model is difficult to solve, we provide a decomposition scheme to reduce the decrease the memory requirements of the model.

7. We apply this model on a realistic case study of an evacuation of a large regional hospital, and we propose some performance measures for the solution.

1.3 Organization of the Dissertation

The remain of the dissertation is organized as follows. Chapter 2 presents an overview of the literature on mass evacuation planning and hospital evacuation; some of the research areas will be reviewed in more details within the manuscript chapters. In Chapter 3, we study the System Optimal Dynamic Traffic Assignment model where the demand is assumed uncertain. We review the methods studied in the literature, mainly the Deterministic Counterpart Strategies, and we show that these methods are not effective to generate a good routing plan. We provide alternatives, where we describe some networks where an optimal policy exist independent of the demand realization, and we propose some simple heuristics for more complex ones. In Chapter 4, we discuss the traffic management strategies under the Cell Transmission Model with flow reductions. We start by making some improvements on the model formulation proposed in Bish et al. (2013) to improve its tractability. We then describe the tools that can be employed in traffic management decisions. In Chapter 5, we discuss the hospital evacuation problem where we develop an integer programming model that integrates the building evacuation with the transportation of patients. The impact of building evacuation capabilities on the transportation plan is investigated through the Caiser Hospital case study. We also propose a decomposition scheme to improve the tractability of the integer program. Chapter 6 summarizes the main conclusions of the research and recommends some future research directions.

Chapter 2

Litterature Review

We begin by reviewing the literature related to Large-scale regional evacuation planning for car-based evacuations.: We start by reviewing by the different traffic flow models and discuss the strength and weakness of each approach. We then discus the dynamic traffic assignment model (DTA) and how it could be incorporated in an optimization model. In the second part we review the literature on hospital evacuation planning and emergency response. Specifically we focus on the mathematical modeling studies. The mathematical modeling studies can be categorized based on their focus on building evacuation and hospital evacuation transportation problems.

2.1 Traffic Flow Modeling

Traffic models can be either classified as microscopic or macroscopic. Microscopic models take into consideration the individual behavior of each vehicle in response to the current traffic conditions whereas macroscopic traffic modeling considers the movement of a stream of vehicles rather than tracking individual ones. In microscopic models, a system of differential equations of continuity and stream motion to obtain the equivalent state. The macroscopic models are based on the hydrodynamic theory of traffic flow (Richards, 1956). The individual vehicles are equivalent to a continuous fluid with a given density and a relation between speed and density is developed. In this section I will describe these approaches in more detail, and we will describe the strengths and weaknesses of each approach.

2.1.1 Microscopic Traffic models

Microscopic are based on the interaction between individual cars. One example is the car following models which describe how each vehicle responds to a change on the relative motion of the vehicle ahead by accelerating or breaking (Addison and Low, 1998). This is done by the use of differential equations that calculates the vehicle position and velocity. Examples of car following models include the intelligent driver model (IDM) and the Gipps model (Gipps, 1981; Treiber et al., 2000).

Microscopic traffic models are mostly used in traffic simulation models. Simulation software can accurately predict how actual traffic flow will behave. 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, computational time and effort required to execute the simulations and the lack of ability to develop plans. These models are better for testing plans against a baseline strategy.

2.1.2 Macroscopic Traffic models

Many macroscopic models of traffic flows are based on the hydrodynamic theory of traffic flow (Richards, 1956). Individual vehicles are not tracked, but traffic flows are modeled as a to continuous fluid with a given density and a relation between speed and density are developed. In this model, we denote $q(x,t)$ to be the average number of vehicles passing per time unit. A systematic procedure could be used to take into account cars completely in a given region at a fixed time. Using these measurements, we can calculate $k(x,t)$ the density of cars per distance unit (mean speed). There is a close relationship between velocity, density and flow, it can be described by the following:

$$q(x,t) = k(x,t)v(x,t) \tag{2.1}$$

Which means that as the density increases, the velocity of cars diminishes. So it is assumed that the velocity is a strictly decreasing function in density. The entity $c = \frac{dq}{dk}$ can be interpreted as the speed of a shock waves carrying continuous changes of flow through the stream of vehicles. One major drawback of this method is that evaluation of the shockwave is tedious even for a single link. So the models based on this approach tried to simplify it while staying consistent with it. In this dissertation,The models are based on the cell transmission model which we will present next.

The cell transmission model (CTM) presented by Daganzo (1994, 1995) predicts the traffic behaviour by evaluating flow and density at a finite number of intermediate points on the network. This is done by decomposing the network into small segments, and the flow is calculated at the links connecting these segments.

Daganzo (1994) showed that the relationship between the traffic flow q and density k is of the form depicted in Equation (2.1):

$$q = \min(vk, q, q_{max}, v(k_j - k)), \text{ for } 0 \leq k \leq k_j \tag{2.2}$$

Where v is the free flow speed, k_j is the jam density, and q_max is the maximum flow.

This method assumes that the road has been divided into homogeneous sections called as cells whose lengths equal the distance traveled by free flowing traffic in one clock interval. The state of the system at instant t is then given by the number of vehicles contained in each cell i, n_i^t. The following parameters are defined for each cell:

- N_i^t: Maximum number of vehicles that can be present in cell i at time t

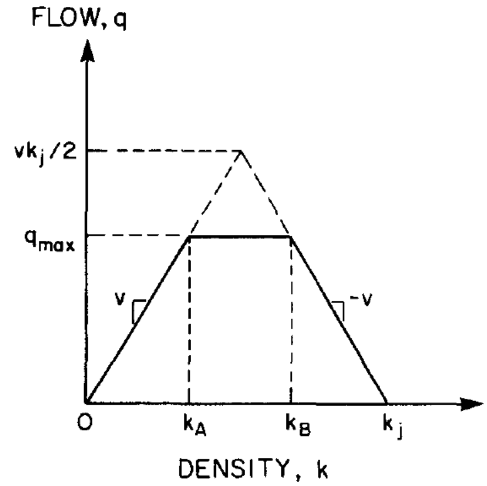


Figure 2.1: Flow density relationship for the CTM.

- Q_i^t: Maximum number of vehicles that can flow into a cell i when the clock advances from t to $t+1$

N_i^t is the product of the cell's length and its jam density. Q_i^t is the product of the clock of the clock interval and the cell's capacity. If cells are numbered consecutively from 1 to I, then we have the following relationship:

$$n_i^{t+1} = n_i^t + y_i^t - y_{i+1}^t \tag{2.3}$$

Where

$$y_i^t = min[n_{i-1}^t, Q_i^t, \delta(N_i^t - n_i^t)] \tag{2.4}$$

where $\delta = w/v$ as shown in Figure 2.1.

This model can be extended to model a more complex network. Daganzo (1995) used this model to study networks where the maximum number of arcs (links) entering/leaving is 3. So cells can be classified in 3 types: diverge if 1 link enters and two leave it, merge if two links enter and one leaves and ordinary if 1 link enters and 1 link leaves. Merge and Diverge links are shown in Figure 2.2.

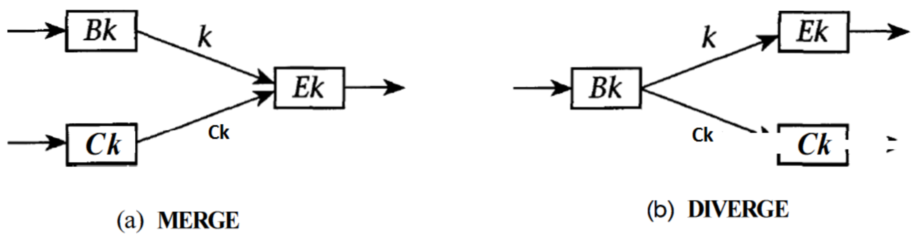


Figure 2.2: Representation of a merge and a diverge cell.

Daganzo (1995) then derived the flow equations for the three type of links. At ordinary links, let y_k^t denote the flow on link k from time t to t+1.

$$y_k^t = min[n_{BK}, min[Q_{BK}, Q_{EQ}], \delta_{EK}[N_{EK} - n_{EK}]] \tag{2.5}$$

To simplify the equation we define:

$$S_I^t = min[Q_I, n_I] \tag{2.6}$$

$$R_I^t = min[Q_I, \delta_I[N_I - n_I] \tag{2.7}$$

Equation 2.5 can be rewritten as:

$$y_k^t = min[S_{BK}, R_{EK}] \tag{2.8}$$

At the merge links, we assume that a fraction p_k of the vehicles come from BK and the remainder p_{ck} from CK, where $p_k + p_{ck} = 1$. If $R_{EK} > S_{BK} + S_{CK}$, we have:

$$y_k^t = S_{BK} \tag{2.9}$$

$$y_{ck}^t = S_{CK} \tag{2.10}$$

Otherwise we have:

$$y_k^t = mid[S_{BK}, R_{EK} - S_{CK}, p_k R_{EK}] \tag{2.11}$$

$$y_{ck}^t = mid[S_{CK}, R_{EK} - S_{BK}, p_{ck} R_{EK}] \tag{2.12}$$

Finally at diverge links, we assume that the proprtions of S_{BK} are β_{CK} and β_{EK} such that $\beta_{CK} + \beta_{EK} = 1$. The flow equations at the diverge links are:

$$y_k^t = \beta_{EK} y_{BK}^t \tag{2.13}$$

$$y_{ck}^t = \beta_{CK} y_{BK}^t \tag{2.14}$$

$$y_{BK}^t = min[S_{BK}^t, R_{EK}^t/\beta_{EK}^t, R_{CK}^t/\beta_{CK}] \tag{2.15}$$

The primary advantages of using the CTM, and other macroscopic models, are its ability to properly model the spillback propagation of congestion and its tractability within mathematical programs.

One drawback of using the CTM is that the flow through a bottleneck is not reduced after the onset of congestion at the bottleneck. Bish et al. (2013) modified the flow density in Figure 2.1 to account for congestion. The traffic flow is now divided into 2 regimes: In the free flow mode, the flow is bounded by the number of vehicles present, until the traffic density reaches a certain level where the congestion mode kicks in: after the onset of congestion, the traffic flow will decrease as the density of the traffic increases. We will review this model in more detail later on.

2.2 Evacuation planning and optimizing network flows

Evacuation planning deals with generating and evaluating evacuation plans through the use of traffic flow models. In that setting, the use of simulation alone is not enough. Optimization and mathematical programs should be used as a decision making tool to generate evacuation plans. The objective of the evacuation plan is typically minimizing the total evacuation time, total travel time or network clearance time. Some examples of the decisions generated by the mathematical programs include:

- Times to issue evacuation orders
- Routing plans
- Time dependent merge priorities and diverge percentages
- Evacuation shelter locations
- Evacuation time estimates

Regardless of the traffic flow model used, a method of assigning traffic to routes overtime needs to be used. Dynamic traffic assignment (DTA) is a broad area of research that addresses this method. Ziliaskopoulos (2000) introduced a linear program for DTA that assigned traffic to routes using the CTM. The LP used a system optimal objective function that minimized the total travel time through the network. Li et al. (2003) used a Dantzig-Wolfe decomposition to reduce the DTA LP to a minimum cost flow problems that are easier to solve. Zheng and Chiu (2011) argued that earliest arrival solutions exist for single destination networks and used this property to generate an optimal solution for the DTA problem.

In Liu et al. (2006), an optimization model 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. 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.

Chiu et al. (2007) presented an optimization model (LP), called the Joint Evacuation Destination- Route-Flow-Departure 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.

2.2.1 Evacuation and demand uncertainty

In evacuation planning, if there is insufficient information about the number of people that are evacuating, the demand level is assumed to be stochastic.

Waller and Ziliaskopoulos (2006) considers a known cumulative distribution function (CDF) of the demand at each origin, and set the deterministic demand parameter based on the probability that the realized demand will be lower, e.g., the demand parameter can be set such that there is a 70% chance that the realized demand will be less than or equal to the parameter. Using this demand parameter an LP is used to minimize the total system time (TST). Using this framework, various probability levels are empirically tested to assess their performance under different demand realizations. To transform the LP solution into an actionable plan the following rules are used: 1) if the realized demand is less than the demand parameter, a randomly chosen subset of flows will be implemented, and 2) if the realized demand is more than the demand parameter, the traffic flows of the solution are implemented, and the “extra” demand is routed over randomly selected routes from from the LP solution.

Yao et al. (2009); Chung et al. (2011, 2012) study the routing problem in an evacuation setting using a robust optimization approach based on a deterministic LP where a range is provided for the uncertain demand. Here the objective is to minimize TST with an additional large penalty for evacuees that have not reached a destination by the end of the given time horizon. To transform the LP solution into an actionable plan the following rules are used: 1) if the realized demand is less than the demand parameter, a randomly chosen subset of flows will be implemented, and 2) if the realized demand is more than the demand parameter, the ”extra” demand will remain at the destination (and thus incur the penalty). Because of the second rule, the large penalty for not evacuating the system, and the assumption of a known range for the demand, the deterministic demand parameter is set to the high-end of the demand range. This is considered a robust approach.

When demand is uncertain, there are various ways of measuring the quality of a solution. For instance, important qualities of a solution include the robustness of a solution (the likelihood that a solution is feasible under various realizations), the expected objective function value (i.e., expected TST), or the solutions worst-case performance, given the possible demand realizations (i.e., the worst possible TST). The solution approaches studied in Waller and Ziliaskopoulos (2006); Yao et al. (2009); Chung et al. (2011, 2012) consider various of these aspects, but these solutions all have problems directly related to the modeling framework used.

2.3 Hospital Evacuation

Planning the evacuation of a hospital is more complicated than the evacuation of most other types of buildings because of the special needs of the patients, which includes assistance in leaving the building, medical treatment during the evacuation, and transport to an alternate care hospital. The importance of hospital evacuation is highlighted by the hospital accreditation standards requiring hospitals to develop evacuation plans. Hospitals may need to be evacuated for several reasons; These include hurricanes, fires, floods, chemical leaks, bomb threats and loss of functionality.

The main purpose of a hospital evacuation is to minimize the risk of patients and staff (Bish et al., 2011). The two main sources of evacuation are the threat risk (reason for evacuation) and the transportation risk. The threat risk characteristics that affect the evacuation plan are the impact of the threat on the various patient types and how threat evolves over time. For instance, a hurricane would pose no immediate threat risk to the patients (since it can be forecasted), but eventually the threat risk might be considerable; a long term power outage would affect critical patients dependent on lifesaving equipment more than patients in the hospital for observation, while a fire for instance could pose a serious risk for all patient types. In fact, in certain situations,a partial evacuation might be appropriate to reduce risk; certain patients might be safer if evacuated whereas other might not tolerate the transportation risk. The transportation risk is a function of the patient type, the vehicle, and the time required to transport the patient to the selected receiving hospital.

The evacuation of a hospital requires the movement of patients from the building to the staging area and the transportation of patients to appropriate receiving facilities by available vehicles. Bish et al. (2011) studied the transportation of patients from evacuating hospitals to alternative receiving hospitals with the assumption that the building can be evacuated such that patients of the appropriate type are available to satisfy the transportation plan withing the physical loading capacity. However, the movement of patients out of the building of the staging area is likely to impose a bottleneck on the evacuation in practice.

General building evacuations have been extensively studied in the operations research literature. However, most of these building evacuation models rely on the mobility of the evacuee population. The prevalent objective of general building evacuation problems is minimizing the total evacuation time (or building clearance time) and these problems are generally formulated as dynamic (time-expanded) network flow problems and solved by applying the corresponding network flow algorithms.

Chalmet et al. (1982) develops deterministic network models for building evacuation assuming constant capacity and travel time for each arc. Hamacher and Tufekci (1987) solves building evacuation models with multiple objectives (such as minimizing the total evacuation time and avoiding cycling of evacuees or evacuation with multiple priority levels for different parts of the building). Choi et al. (1988) incorporates flow dependent capacities into building evacuation network which adds side constraints to the problem. Chen and Miller-Hooks (2008) formulates the building evacuation problem with shared information on the changes in evacuation routes as a mixed integer linear program that minimizes the total evacuation time.

The majority of evacuation models minimize some function of the evacuation time. Han et al. (2007) discuss various evacuation objectives minimizing measures of effectiveness including individual travel (or exposure) time, time-based risk exposure, and time and space based risk exposure. A combination of these measures of effectiveness can be implemented in multi objective evacuation optimization problems. Løvås (1995) studies a building evacuation network with stochastic variables and discusses performance measures related to accident effects, evacuation time, queuing and waiting, network distances, and network redundancy.

The building evacuation studies reviewed above do not specifically address the evacuation problems of healthcare facilities. Hospital evacuations, unlike general building evacuations, involve evacuees that require extensive assistance; therefore, these problems need to be handled in a distinctive manner. Furthermore, a hospital evacuation involves more than the safe and efficient clearance of the building. An equally important aspect of the evacuation is to transport patients to appropriate alternative care facilities. The literature is quite scarce in hospital evacuation modeling pertaining to both the hospital building evacuation and the subsequent transportation of patients. The Hospital Evacuation Decision Guide (see http://archive.ahrq.gov/prep/hospevacguide), prepared for the Agency for Healthcare Research and Quality (AHRQ), discusses the importance of estimating evacuation time, among other things, to support the decision to evacuate and the timing for this decision. The evacuation time (i.e., how long it takes to evacuate the hospital) is dependent on the resources available and how efficiently they are used. This metric is difficult to estimate.

Duanmu et al. (2010) focuses on the routing of hospital vehicles during a hurricane evacuation where the ambulances and general traffic compete for space in the regional traffic flow network. The ambulance trip times are estimated using a simulation model based on various hospital evacuation start times and multiple strategies that minimize the transportation time for patients are produced. Duanmu et al. (2010) does not consider any patient-specific attributes or requirements. Golmohammadi and Shimshak (2011) estimate the evacuation time for the hospital building evacuation using a predictive model that takes patient population and available resources as input and calculates the total evacuation time. Three patient types are defined based on mobility and the patients who are the fastest to evacuate are given the first priority. This patient prioritization rule is analogous to the shortest processing time rule in scheduling theory and can significantly increase the waiting time of the most critical patients. Bish et al. (2011) studies the allocation of patients, categorized by criticality and care requirements, to a limited fleet of vehicles of various capacities and medical capabilities, to be transported to appropriate receiving hospitals considering the current available space in each hospital for each category of patient. The objective is to minimize the expected risk, both the threat risk that is forcing the evacuation, and the risk inherent in transporting patients, some in critical condition.

Chapter 3

Routing strategies under demand uncertainty

Abstract

In this paper, we study network routing and traffic controls under demand uncertainty. Specifically, we examine the strategy of using a deterministic parameter in an optimization setting (a strategy employed in the literature) to represent the demand uncertainty, where traffic flows are modeled using the Cell Transmission Model (CTM). For a special class of networks, for which instances have been previously analyzed in the literature, we provide an optimal policy (i.e., a policy whose solution is optimal for any realization of the demand). Using this optimal policy we show the problems inherent using a deterministic parameter to represent uncertainty. We then show that, for other types of networks, for which optimal policies do not exist, simple heuristics can outperform the use of optimization with a deterministic parameter that represents the demand uncertainty.

Keywords: Cell Transmission Model (CTM); demand uncertainty; robust optimization; stochastic optimization.

3.1 Introduction

In this paper, we study the problem of routing and controlling traffic flows through a network, from origins to destination, under demand uncertainty. The objective is to minimize the sum of the time that each vehicle (from the realized demand) remains in the network before reaching a destination; we refer to this as the Total System Time (TST) which is known as the system optimal problem (as opposed to the other common objective of obtaining a user equilibrium). This important problem is related to dynamic traffic assignment (DTA) and regional evacuation planning problems. Specifically, we study this problem using the cell transmission model (CTM) (Daganzo, 1994, 1995) to approximate traffic flows. One potential drawback of CTM is that it has a nonlinear flow-density relationship. Ziliaskopoulos (2000) proposed a Linear Program (LP) where traffic flows are governed by a linear version of CTM. This framework optimizes routing and traffic controls (e.g., flow priorities at network merges) to minimize TST, and has been extensively studied in the literature under deterministic demand for both DTA problems (e.g., Lo, 2001; Lin and Wang, 2004; Nie, 2011) and similar evacuation planning problems (e.g., Bish and Sherali, 2013; Bish et al., 2014; Chiu et al., 2007; Liu et al., 2006; Chiu et al., 2007; Tuydes and Ziliaskopoulos, 2006). This framework has also been studied under demand uncertainty in Waller and Ziliaskopoulos (2006); Yao et al.

(2009); Chung et al. (2011, 2012); these papers propose the use of a deterministic demand parameter to represent the stochastic demand in an optimization setting. The optimal traffic flows produced from this strategy (optimal, that is, for the given deterministic problem) are then subject to policies to produce a solution (i.e., a plan) for any realization of the demand.

Waller and Ziliaskopoulos (2006) considers a known cumulative distribution function (CDF) of the demand at each origin, and set the deterministic demand parameter based on the probability that the realized demand will be lower, e.g., the demand parameter can be set such that there is a 70% chance that the realized demand will be less than or equal to the parameter. Using this demand parameter an LP is used to minimize TST. Using this framework, various probability levels are empirically tested to assess their performance under different demand realizations. To transform the LP solution into an actionable plan the following rules are used: 1) if the realized demand is less than the demand parameter, a randomly chosen subset of flows will be implemented, and 2) if the realized demand is more than the demand parameter, the traffic flows of the solution are implemented, and the "extra" demand is routed over randomly selected routes from from the LP solution.

Yao et al. (2009); Chung et al. (2011, 2012) study the routing problem in an evacuation setting using a robust optimization approach based on a deterministic LP where a range is provided for the uncertain demand. Here the objective is to minimize TST with an additional large penalty for evacuees that have not reached a destination by the end of the given time horizon. To transform the LP solution into an actionable plan the following rules are used: 1) if the realized demand is less than the demand parameter, a randomly chosen subset of flows will be implemented, and 2) if the realized demand is more than the demand parameter, the "extra" demand will remain at the destination (and thus incur the penalty). Because of the second rule, the large penalty for not evacuating the system, and the assumption of a known range for the demand, the deterministic demand parameter is set to the high-end of the demand range. This is considered a robust approach.

When demand is uncertain, there are various ways of measuring the quality of a solution. For instance, important qualities of a solution include the robustness of a solution (the likelihood that a solution is feasible under various realizations), the expected objective function value (i.e., expected TST), or the solutions worst-case performance, given the possible demand realizations (i.e., the worst possible TST). The solution approaches studied in Waller and Ziliaskopoulos (2006); Yao et al. (2009); Chung et al. (2011, 2012) consider various of these aspects, but these solutions all have problems directly related to the modeling framework used. One of the main goals of this paper is to illustrate some of the problems. To do so, we develop an optimal policy for problems that have a special network structure (the network used in Waller and Ziliaskopoulos (2006); Chung et al. (2012), which we also use for illustrative purposes, has this structure). This policy is easily implemented and directly leads to an optimal solution for any realization of the demand (and is thus robust). This allows us to evaluate the solution approaches studied in Waller and Ziliaskopoulos (2006); Yao et al. (2009); Chung et al. (2011, 2012). Furthermore, for more general networks we provide a heuristic that also outperforms the solution deterministic counterpart approaches from the literature.

The remainder of the paper is structured as follows. Section 3.2 presents the model that we are studying, we then review the approaches used by Waller and Ziliaskopoulos (2006); Yao et al. (2009); Chung et al. (2011, 2012) to solve this model. In Section 3.3, we present a class of networks that are studied in the literature in which the optimal solution can be characterized irrespective of the demand level and we will use it to illustrate some of the disadvantages of the methods introduced in Section 3.2 to solve the model with demand uncertainty. In Section 3.4, we describe a heuristic that can be used in any network structure, and we compare it to the deterministic counterpart strategies. And finally, Section 3.5 is the conclusion.

3.2 Model

In this section, we present the cell transition model (CTM), which will be used to model traffic flows, and describe the problem of interest. Next we present two models, the first (Model 1) is a nonlinear (and nonconvex) program based on the CTM. As this model can be difficult to solve, we then present a second model (Model 2), which is a commonly used linearized version of CTM. We then discuss the implications of demand uncertainty for the problem.

The CTM (see Daganzo, 1994, 1995) utilizes a discrete, time-expanded network of cells and links (C, L) to represent a roadway system. Cells can be either source cells (set S_o) (demand is generated in the source cells, which have only a single, outgoing link, and no incoming links), sink cells (set S_e) (which have only a single incoming link, and no outgoing links), or roadway cells (set R). Links represent allowable, directed, movements between cells. When a roadway cell has two outgoing links, these links are diverge links (L_d), likewise when a roadway cell has two incoming links, these links are merge links (L_m). We will use the convention of calling a roadway cell having outgoing diverge links a diverge cell, and a roadway cell having incoming merge links a merge cell. All other links are ordinary (L_o); an ordinary link is the only outgoing and incoming link for two adjacent cells. Each source cell has only one outgoing link, while each sink cell has only one incoming link. Figure 3.1 illustrates these network components; Cell 1 is a source, Cell 8 a sink, and cells 2-7 are roadway cells, links (2,3) and (2,5) are diverge links, links (4,7) and (6,7) are merge links, and links (1,2), (3,4), (5,6) and (7,8) are ordinary links. The planning horizon is divided into T time intervals of length τ, and a roadway cell represents a section of roadway of length ℓ such that vehicles traveling at free-flow speed (u_f) traverse the section in one time interval; that is $\ell = u_f \times \tau$.

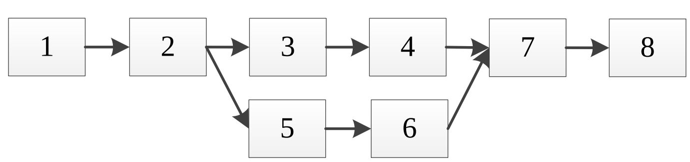


Figure 3.1: A small network example to illustrate the network components used in CTM.

Using this framework, our problem is to send flows (vehicles) through a network from source cells to sink cells, by selecting routes (decisions at diverge cells) and setting traffic controls (priority decisions at merge cells) to minimize the total system time (TST), i.e., the sum of the time each vehicle spends in the network before reaching a sink, under demand uncertainty. For a deterministic problem, measuring the performance of a solution is a straightforward task, but when the demand is uncertain, evaluating a solution is more complicated. Since the demand is a random variable the objective function value is also a random variable. One way of evaluating a potential solution is by using the expected value of the objective function value. Another is to use the worst case performance of a potential solution considering all the possible demand realizations. In addition, the feasibility of the solution (i.e., how robust the solution is), over the potential demand realizations, is also important. Waller and Ziliaskopoulos (2006); Yao et al. (2009); Chung et al. (2011, 2012) explore an approach to solving this problem under demand uncertain using a deterministic model and *deterministic counterparts*, i.e., deterministic parameters, to represent the random demand variables. We denote this general deterministic counterpart strategy as the *DC strategy*. Before we examine the various deterministic counterparts proposed in the literature, we present some additional notation and two related deterministic models.

Parameters:

d_i^t	:	demand generated in cell i in time interval t, $i \in S_o$
N_i	:	maximum number of vehicles that cell i can hold, $\forall\ i \in \mathrm{R}$
Q_i	:	inflow/outflow capacity of cell i, $\forall\ i \in \mathrm{R}$
δ_i	:	traffic flow parameter for cell i, $\forall\ i \in \mathrm{R}$

The CTM parameters have the following relationships for any roadway cell i: $0 < Q_i < N_i$, while $0 < \delta_i \leq 1$. By definition, Q_i represents the maximum flow into or out of cell i. This flow might not be achievable, if for instance, the adjacent downstream cell j had a lower Q-value. It does imply that the parameters be set such that $\delta_i[N_i - x_i^t] > Q_i$ when $x_i^t < Q_i$, else the definition of Q_i would be contradicted, that is, cell i itself would prohibit a flow of Q_i from ever entering cell i. As a result, $N_i > (1 + 1/\delta_i)Q_i$.

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 flowing from cell i to cell j during time interval t, $\forall\ (i,j) \in L$, $t = 1, \cdots, T$

p_j^t : proportion of vehicles flowing on a diverge link (i,j) from cell i, $\forall\ (i,j) \in L_d$, $t = 1, \cdots, T$

All the decision variables are restricted to be non-negative; furthermore we set $x_i^1 = 0$, $i \in R \cup S_e$. We now present the Model 1 formulation.

$$\textbf{Model 1}: \quad \text{Minimize} \sum_{t=1}^{T} \sum_{i \in C/S_e} x_i^t \tag{3.1}$$

subject to :

$$x_i^t = x_i^{t-1} - \sum_{k:(i,k) \in L} y_{ik}^{t-1} + d_i^{t-1}, \quad \forall i \in S_o, \; t = 2, \cdots, T \tag{3.2}$$

$$x_i^t = x_i^{t-1} + \sum_{k:(k,i) \in L} y_{ki}^{t-1} - \sum_{k:(i,k) \in L} y_{ik}^{t-1}, \quad \forall i \in C/S_o, \; t = 2, \cdots, T \tag{3.3}$$

$$S_i^t = min(x_i^t, Q_i), \quad \forall i \in R, \; t = 1, \cdots, T \tag{3.4}$$

$$R_i^t = min(Q_i, \delta_i[N_i - x_i^t]) \quad \forall i \in R, \; t = 1, \cdots, T \tag{3.5}$$

$$y_{ij}^t = min(S_i^t, R_j^t), \quad \forall (i,j) \in L_o, \quad t =, \cdots, T \tag{3.6}$$

$$y_{ij}^t \leq min(S_i^t, R_j^t), \quad \forall j \in R, \; (i,j) \in L_m, \; t = 1, \cdots, T \tag{3.7}$$

$$y_{ij}^t + y_{kj}^t = min(S_i^t + S_k^t, R_j^t), \quad \forall j \in R, \; (i,j), (k,j) \in L_m, \; i \neq k, \; t = 1, \cdots, T \tag{3.8}$$

$$y_{ij}^t = min(p_j^t S_i^t, R_j^t, \frac{p_j^t}{p_k^t} R_k^t), \quad \forall i \in R, \; (i,j), (i,k) \in L_d, \; j \neq k, \; t = 1, \cdots, T \tag{3.9}$$

$$p_j^t + p_k^t = 1, \quad \forall (i,j), \; (i,k) \in L_d, \; i \in R, \; j \neq k. \tag{3.10}$$

Objective Function (3.1) minimizes the total system time (TST), i.e., the sum of the time each vehicle spends in the network before reaching a sink. Constraints (3.2) and (3.3) are the flow conservation constraints. Constraints (3.4) defines S_i^t, the maximum allowable flow out of a cell i in interval t; Roadway cell i is considered to be in the congested state (i.e., congested) when $x_i \geq Q_i$. Constraint (3.5) is the maximum allowable flow into a cell. Constraint (3.6) is the equation for the flow at ordinary links. Constraints (3.7) and (3.8) determine the flows on the merge links. Constraint (3.7) places a limit on each individual flow, while (3.8) (a simplification of the original constraints proposed in Daganzo (1995)) determines the flows (i.e., priorities, which represent traffic controls) for the merge links and ensures the maximum possible flow into the merge cell. Constraints (3.9) and (3.10) determine the flow on the diverge links (i.e., the routing decisions). Constraints (3.10) for the set of diverge links (i, j) and (i, k) are equivalent to $\max\{y_{ij}^t + y_{ik} : y_{ij}^t + y_{ik} \leq S_i^t,\ p_j^t(y_{ij}^t + y_{ik}) \leq R_j^t,\ p_k^t(y_{ij}^t + y_{ik}) \leq R_k^t\}$ given the turn percentages p_j^t and p_k^t, where $y_{ij}^t = p_j^t(y_{ij}^t + y_{ik})$ and $y_{ik}^t = p_k^t(y_{ij}^t + y_{ik})$. In other words, either the diverge cell must be sending flow out at its maximum rate, or one of the downstream adjacent cells must be receiving flow at its maximum rate. The formulation is complete given the logical nonnegativity restrictions on the decision variables.

Next, we present two important observations about the CTM.

Observation 3.1 *Congestion on any roadway cell i (cell i is congested when $x_i \geq Q_i$) does not limit the flow out of i.*

Observation 3.1 is directly obtained from Equation (3.4), which defines the maximum flow out of any roadway cell, which considers the number of vehicles a limiting factor, which, is a limit on flow based on conservation, not congestion.

Observation 3.2 *For any route through the network, the roadway cell i having the smallest Q-value sets the maximum flow over the route, which cannot be reduced due to network congestion, and can be achieved, in isolation, given sufficient demand.*

Observation 3.2 implies that, if used in isolation, any route can send a flow equivalent to the minimum Q-value on that route given sufficient demand, no matter how congested any of the cells are. Of course, if, for example, two routes shave roadway cells, this can reduce the flow on one or both routes, but not because of congestion.

Model 1 can be difficult to solve since it is nonlinear and nonconvex. An alternative approach, that is commonly used, is a linearized CTM (see Ziliaskopoulos, 2000), which is the approach used in Model 2; where we replace Constraints (3.4)-(3.10) with the following:

$$\sum_{j:(i,j)\in L} {y_{ij}}^t \leq x_i^t, \quad \forall i \in C,\ t = 1, \cdots, T \tag{3.11}$$

$$\sum_{j:(i,j)\in L} y_{ij}^t \leq Q_i, \quad \forall i \in R,\ t = 1, \cdots, T \tag{3.12}$$

$$\sum_{j:(j,i)\in L} y_{ji}^t \leq Q_i, \quad \forall i \in R,\ t = 1, \cdots, T \tag{3.13}$$

$$\sum_{j:(j,i)\in L} y_{ji}^t \leq \delta_i^t(N_i - x_i^t), \quad \forall i \in R,\ t = 1, \cdots, T. \tag{3.14}$$

Constraints (3.11)-(3.14) are linearized replacements for (3.6)-(3.10) (see Ziliaskopoulos, 2000). The flow over any ordinary link, merging link, or set of diverging links is limited by the occupancy of the upstream cell (3.11) and its flow capacity (3.12). Furthermore the flow over any ordinary link, diverge link, or set of merging links is limited by the flow capacity of the downstream cell (3.13) and its remaining capacity (3.14). A consequence of linearizing the constraints in Model 2 is an expansion of the feasible region, e.g., the flow of an ordinary link y_{ij}^t has to be equal to $\min(S_i^t, R_j^t)$ in Model 1, whereas it can be in the interval

$[0, \min(S_i^t, R_j^t)]$ in Model 2. This (well known) problem is referred to as *traffic holding*, which occurs when a solution to Model 2 is infeasible for Model 1, in other words, solutions with traffic holding do not conform to the flow-density relationship underpinning CTM. Liu et al. (2006) mentions traffic holding as an unrealistic behavior, while Bish and Sherali (2013); Nie (2011) shows that traffic holding is not required in an optimal solution for the deterministic problem. Furthermore, we find in our initial studies that it is harmful under uncertainty. Unfortunately, Model 2 usually produces solutions having traffic holding. Since traffic holding is not a necessary feature of an optimal solution and potentially detrimental under demand uncertainty, we remove it using a lexicographic objective function that minimizes TST, and then minimizes $\sum_{(i,j)\in L}\sum_{t=1}^{T} t y_{ij}$ (Lin and Wang, 2004). This secondary objective forces the traffic flows to advance as much as possible, i.e., until one of the CTM flow constraints is tight. This alternative objective function gives us solutions that minimize TST and are feasible for Model 2. Another option is to solve Model 2 with objective function (3.1), obtain the optimal TST, and then use this as an additional constraint in Model 2, which we solve again, but now minimizing $\sum_{(i,j)\in L}\sum_{t=1}^{T} t y_{ij}$. An alternative objective, minimizing the network clearance time (NCT), i.e., finding the earliest time interval in which all the demand can reach the sinks, is often used in evacuation planning. For the deterministic problem a solution to Model 1 or 2 that minimizes TST also minimizes NCT (see Bish et al., 2014). Later, we will consider the NCT objective under demand uncertainty.

We now address an obvious obstacle when using the solution from Models 1 or 2 under demand uncertainty, and an important step in implementing a *DC strategy*, which is that the demand realization is unlikely to match the deterministic counterpart used, and thus the solution from these models must be modified. Two *implementation methods* for *DC strategies* are as follows:

Priority Method The flows from the deterministic solution (from either Models 1 or 2) are used to calculate turn percentages for diverge links and priorities for merge links for each interval. These are, in turn, applied to the realized demand. If, for a particular interval, the deterministic solution does not supply the appropriate flows to calculate these values, average values for the particular cell (merge or diverge) are used.

Flow Method The flows from the deterministic solution are used directly. If the demand realization for a source is less than the deterministic counterpart in any interval, the flows for the various routes selected for that intervals demand are proportionally reduced. If the demand realization is greater than specified by the deterministic counterpart in any interval, the flow from the solutions are used, and additional flows (in accordance with the demand realization) are sent on randomly selected routes used in the solution in that interval. If, for a particular interval, the deterministic solution does not supply appropriate flows for a source node's demand, then routes are randomly selected from all routes used in deterministic solution for that particular source node. The priority decisions on merge links are used as necessary.

Next, we discuss the literature, specifically the deterministic counterparts used and the implementation method. Waller and Ziliaskopoulos (2006) assumes that the demand has a known cumulative distribution function (cdf) $F_{d_i^t}(x)$ and the deterministic counterpart $\hat{d}_i^t$ is set based on a parameter α that is a probability measure of the extent to which constraint violations are permitted, thus $F_{d_i^t}(\hat{d}_i^t) = \alpha$. Waller and Ziliaskopoulos (2006) essentially studies the problem using Model 2 and a given deterministic counterpart based on various α-parameters. To evaluate the various solutions, the paper uses the *Flow Method* and simulation.

Yao et al. (2009); Chung et al. (2011, 2012) all use a similar framework to study demand uncertainty in an evacuation context. This framework is built on the assumption that there is a known demand range, and that if the demand realization is larger than the the deterministic counterpart, then the excess demand will remain at the source. This is coupled with an objective function that minimizes a modified TST. Specifically, this objective function is $\sum_{t=1}^{T} \sum_{i \in C/C_s} c^t x_i^t$, where $c^t = 1$ for $t \neq T$, $c^T = M$ and M is a large number representing the infeasibility cost resulting from demand uncertainty. Thus M is a penalty for evacuees that do not "clear" the network by time T. Due to this large penalty, and the above assumptions, they find the best approach is to set the deterministic counterpart to the high end of the range (similar to using an α-parameter of 1 in (Waller and Ziliaskopoulos, 2006)). To evaluate this solution, these papers assume if the demand realization is less than the deterministic counterpart, the flows on each route are reduced proportionally, which is similar to the *Flow Method* described above.

3.3 An optimal policy under uncertainty

In this section, we introduce a class of networks that have a special structure such that there exists an optimal policy for Models 1 and 2 under demand uncertainty. This optimal policy is easy to implement, and produces the optimal solution for any demand realization. We denote these networks as *single source shortest path dominated* (SSSPD) networks. We note that SSSPD networks are common in the literature, for instance, the SSSPD network we use in our analysis (see Figure 3.2) is used in Waller and Ziliaskopoulos (2006); Chung et al. (2012); Li et al. (2003). Using these networks, we examine the use of the DC strategy to derive solutions for Model 1 under uncertainty as was done in the literature. We introduce additional set notation and a set operator to define SSSPD networks.

$C(i,j)$	:	set of cells on the shortest route connecting cell i and cell j, i.e., the route with the fewest number of cells.
$Q_{min}(A)$	:	$\min(Q_i\colon\ i \in (A \cup R))$, i.e., the smallest Q-value of the roadway cell in set A. If A represents a route, then this represents the *flow capacity* of the route.

An SSSPD network has a single sink cell s. We denote the last merge cell before the sink as m (the sink only has a single incoming link) and the first diverge cell after source i as d_i (each source cell only has a single outgoing link); there is only one route connecting i to d_i and also only one route connecting m to s. Given this, we define an SSSPD networks as a single sink network where the following two conditions hold:

Condition 1. If there exists more than one route between source cell i and the sink cell s, then the shortest route between cell i and s must have either $Q_{min}(C(d_i, m)) \geq Q_{min}(C(m, s))$ or $Q_{min}(C(d_i, m)) \geq Q_{min}(C(i, d_i))$.

Condition 2. For each merge cell k used in the solution (including m) the following must hold: $Q_i \geq Q_k$ and/or $Q_j \geq Q_k$, where cells i and j are the two direct predecessors of k.

For SSSPD networks, we can define the optimal policy as follows:

Proposition 3.1. *The optimal policy for Models 1 and 2, under demand uncertainly, for a problem instance on an SSSPD network, has the following structure:*

1. *The demand at each source cell $i \in S_o$ uses the shortest route from source i to sink s, i.e., route $C(i, s)$. If there are multiple shortest routes, pick a shortest route satisfying Condition 1.*

2. *At every merge cell k, having i and j as predecessors, the cell with the lower Q-value has priority. Thus, without loss of generality, if $Q_i \geq Q_j$, then the flows onto k are $y^t_{jk} = min(S^t_j, R^t_k)$, $y^t_{ik} = min(S^t_i, R^t_k - y^t_{jk})$.*

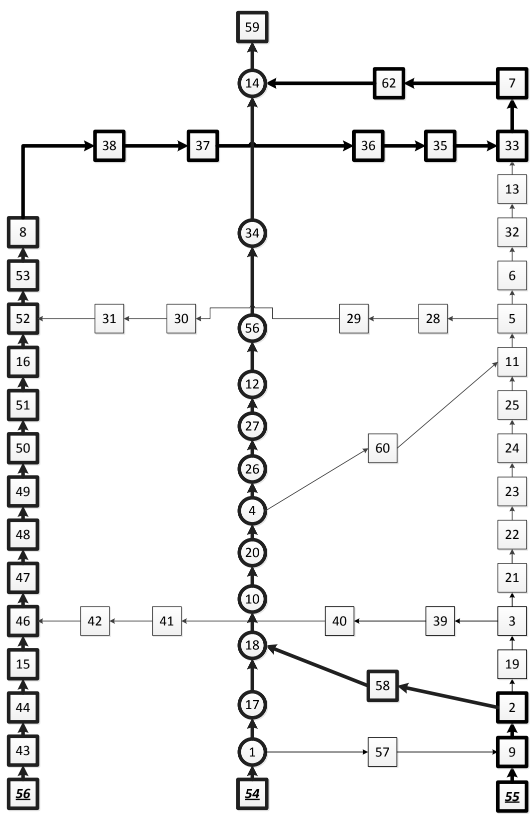
Proof. Every route between the source cells and the sink on an SSSPD network must go through merge cell m and the single route between merge cell m and the sink, thus $Q_{min}(C(m, s))$ is an upper bound on the flow into the sink in any interval. Under CTM, see Observation 3.1 and 3.2, the flow out of a roadway cell i is not reduced because of congestion on i, nor will congestion reduce the flow capacity of any route. Because congestion cannot limit the flow into the sink, the only way to reduce this flow, given ample demand, is to use the network inefficiently, i.e., send flows less than $Q_{min}(C(m, s))$ over route $C(m, s)$, when there is sufficient demand to utilize route $C(m, s)$ to its full extent. Inefficacy can be occur through routing (i.e., decisions at diverge cells) or traffic controls (i.e., priority decisions at merge cells).

We first consider routing. In a SSSPD network, when there are multiple routes between a source i and the sink s, the shortest route selected by the given policy, by Condition 1, has the roadway cell with the limiting flow capacity (i.e., the smallest Q-value) in the set of cells $C(i, d_i)$ or $C(m, s)$; in either case flows cannot be augmented by using multiple or alternate routes. Route $C(d_i, m)$ can send sufficient flow to fully utilize route $C(m, s)$, even if multiple routes share some of these cells (which does not diminish the flow into merge cell m), even in the presence of congestion, based on Observations 1 and 2 (the lack of flow reduction due to congestion for CTM). Having multiple routes share some of these cells would limit the flow from (some of) the sources sharing the roadway cells, but not the flow into the sink. This is independent of demand uncertainty, because it is based on the known parameters of the SSSPD network (i.e., flow capacity, Q_i for cell i).

Next we consider traffic controls. At a merge cell, by Equation (3.8) the solution will always send as much flow as possible, *given the available demand*, which allows for inefficiencies based on demand. Consider two merge links (i, k) and (j, k), and suppose without loss of generalization that $Q_i \leq Q_k \leq Q_j$, which adheres to Condition 2 for SSSPD networks. We are only interested in situation where there is sufficient demand to form a traffic queue (in this case Equation (3.8) allows a decision, which is where the queue forms, either on i, j, or both). If the queue for merge link (j, k) is exhausted before the queue for link (i, k), then cell k can be used inefficiently when $Q_i < Q_k$. Thus giving link (i, k) full priority ensures the most efficient use of the merge by exhausting the queue on the lower capacity link as early as possible (full priority, based on Equations (3.7) and (3.8) uses the merge cell to its full extent when demand is sufficient, and does not imply zero flow over link link (j, k)). When demand is exhausted on cell i, cell j can fully supply the flow to k since $Q_i \geq Q_j$. Of course, this is dependent on Observation 1, i.e., that congestion does not reduce flow. ∎

The optimal policy described in Proposition 3.1 provides an optimal solution to every possible realization of the demand, given an SSSPD network, and is thus a robust policy, even when the upper range of the demand distribution is not known. The optimal policy does not directly provide the values of the flows (the y-variables), which are dependent on the demand realization. This policy is quite easily implemented, since all vehicles from the same source are given the same route (such an easily implementable strategy was studied in Bish and Sherali (2013), but for networks where this is not necessarily the optimal policy). Conversely, the optimal solutions generated using the *DC strategy* are not simple to implement given the multiple routing and the solution that does not match the realized demand (thus the implementations discussed in Section 3.2). We note that there are other network configurations where this policy is optimal, and other network configurations that have similar optimal policies. Our purpose here is to use the optimal policy to evaluate and demonstrate the problems with using Models 1 and 2 and a deterministic counterpart to solve the problem under demand uncertainty.

To illustrate the shortcomings of the *DC strategy*, we use an example on the SSSPD network displayed in Figure 3.2 (including a table of cell parameters) and the optimal policy described in Proposition 3.1. This network was introduced in (Li et al., 2003) in a deterministic setting, and under demand uncertainty (along with a *DC strategy*) in Waller and Ziliaskopoulos (2006) and Yao et al. (2009). Figure 3.2 displays the shortest route from each source to the



Cell	freeway cells	Arterial cells
N_i	20	10
Q_i	12	3
δ_i^t	1	1

Figure 3.2: Test network and the shortest path subnetwork (see Waller and Ziliaskopoulos, 2006; Yao et al., 2009).

sink by showing the appropriate cells in bold. Source cells are square, with their identifier underlined, and cell 59 (also displayed as a square) is the sink. Square roadway cells represent arterial roadways, while circular roadway cells represent freeways. There is only one path from source cell 56 to the sink, thus there are no diverge cells between this source and sink, and this route does not have to adhere to the Condition 1 (which only applies if there are multiple routes). Merge cell 14 is the final merge before the sink (i.e., m), and by Condition 2, cell 62 has priority over cell 34 (cell 62 sends its maximum possible flow in any interval, and cell 34 can then make use of any excess flow capacity in that interval). The only other merge used in the solution is cell 18, where, again by Condition 2, cell 58 has priority over cell 17. These merges illustrate why Condition 2 is required under demand uncertainty, consider merge cell 14, here the route from source cell 54 can utilize the full flow capacity of cell 14 (i.e., Q_{14}), while the route from source cell 56 cannot. In the solution from the optimal policy, congestion forms before cell 14 (the bottleneck and last merge cell) and before cell 18 (a merge cell between the first diverges of source cells 54 and 55 and the last merge cell). Solutions from the deterministic counterpart strategies can have congestion form at other points in the network.

As in Waller and Ziliaskopoulos (2006); Yao et al. (2009); Chung et al. (2011, 2012), we will assume that the demand is only generated at the three source cells (cell 54, 55 and 56) at time 0. For this example the demand is assumed to be uniformly distributed between 0 and 400 for all the source cells. The planning horizon (T) is set to 200 time intervals. In this study, the demand leaves the source cells at the highest possible rate, until it is exhausted.

We generate 500 demand scenarios, which are then evaluated for the different *DC strategies* and the optimal policy, the results are summarized in Table 3.1.

		DC strategy						
		Priority Method			Flow Method			
		$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 1$	Optimal policy
TST	mean	33,459	34,181	35,212	34,866	34,306	35,212	33,341
	std	17,504	17,447	17,160	19,207	17,442	17,160	15,561
	max	71,885	73,447	76,966	95,380	85,917	76,966	70,994
NCT	mean	112.32	115.34	119.36	123.2	119.12	119.36	111.7
	std	33.3	32.4	30.4	40.19	34.06	30.4	29.5
	max	165	167	171	200	186	171	158

Table 3.1: Performance measures of the two DC strategies and the optimal policy.

As Table 3.1 illustrates, the two DC strategies converge for $\alpha = 1$, because they differ in how demand in excess of the deterministic counterpart is handled. Overall the Priority Method tends to do better for lower α-levels than the Flow Method. This is because the Flow Method randomly selects routes for any demand over the deterministic counterpart, unlike the Priority Method, where the excess" demand is routed in a less random manner. The better performance of the Priority Method also stems from the use of an SSSPD network. This is because, as α decreases the solution to Models 1 and 2 tend to use the shortest paths. This is because there is less congestion, which, at least for the deterministic setting, makes the use of the longer paths unlikely in an optimal solution. As a thought exercise, we can see that if we solve this problem for a single vehicle from each source, the model would always choose the shortest path, and important feature of the optimal policy. For example, Figure 3.3 shows the optimality gap for the two DC strategy implementations for $\alpha = 0.5$; here we see some of the biggest optimality gaps at high demand levels. We note that even at low demand levels, any particular source might require many random paths when using the Flow Method).

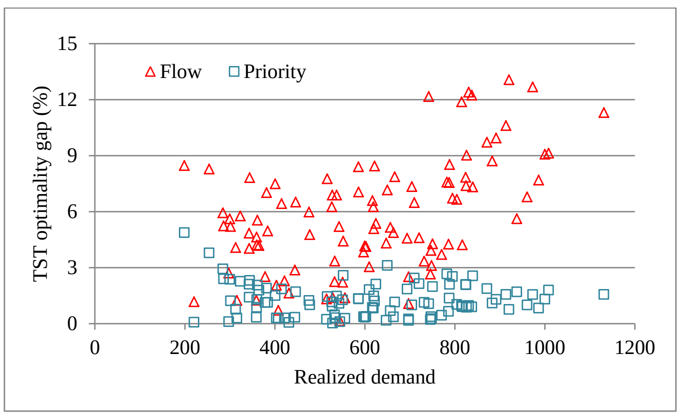


Figure 3.3: The TST optimality gap for $\alpha = 0.5$ for the Flow and Priority Methods.

In Waller and Ziliaskopoulos (2006) the DC strategy is implemented using the Flow Method. Our findings are similar in that we also observed that an α-value of 0.7 tends to yield the best results for the Flow Method. It does so by balancing the selection of a large number of routes randomly selected from our solutions route set, which occurs when the demand realization is larger than the deterministic counterpart, with the problems associated with proportionality reducing the flow compared to the solution found using the deterministic counterpart (which includes taking flow from the shortest paths, part of the described optimal policy), which occurs when the deterministic counterpart is larger than the realized demand.

The concept of efficiency used in the proof of Proposition 3.1 is illustrated in Figure 3.4, which shows the flow out of cell 4. To use the network efficiently, we must utilize the merge m (cell 14) as efficiently as possible, as this is the network's upper bound on flow. To do so, the shortest path from any source to the sink, which is the freeway path from source 54, must be used. In this network, the freeway is able to send enough flow to fully utilize the merge, which accounts for the quick jump from zero flow to a flow of 12 (the maximum flow capacity). Here we see both the DC strategy, and the optimal policy are similar for the initial intervals. It is at the later time intervals that we see inefficiencies. The optimal policy uses the merge as efficiently as the realized demand allows, and reduces flows in a step fashion as demand from the various sources is exhausted. The DC strategy produces a solution that uses the merge cell efficiently only if the demand realization matches the α-level . When it does not, we get the inefficiencies displayed in the figure. These can be directly attributed to controlling merge 14 and 18 inefficiently, and sending some portion of the flow over longer routes.

The optimal policy provides a solution that is optimal for every demand realization, and because of that, the solution is optimal for the TST (which the objective function minimizes), but also for the NCT. This is not the case with the DC strategy; when demand realizations do not conform to the deterministic counterpart used, this can impact the NCT in a different manner, as the NCT is a function of when the last vehicle clears the network. We can see this in the optimality gaps displayed in Figure 3.5, where the largest optimality gaps are for the

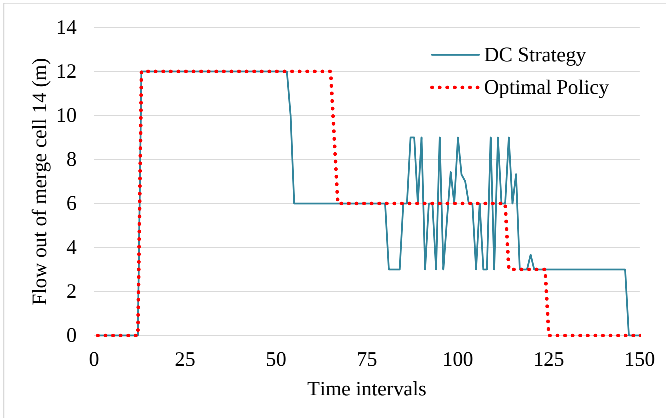


Figure 3.4: The flow out of merge cell 14 (the last merge m and a network bottleneck) for an instance using $\alpha = 0.5$.

NCT. For example the average optimality gap for the TST at $\alpha = 0.5$ for the Flow Method is around 5%. But for the NCT it is greater than 26%. This is because if the DC strategy efficiently routes most of the vehicles, the TST objective value will be close to optimal, but the small portion of the vehicles routed inefficiently can greatly increase the NCT because the NCT is only dependent on the last vehicle in the system.

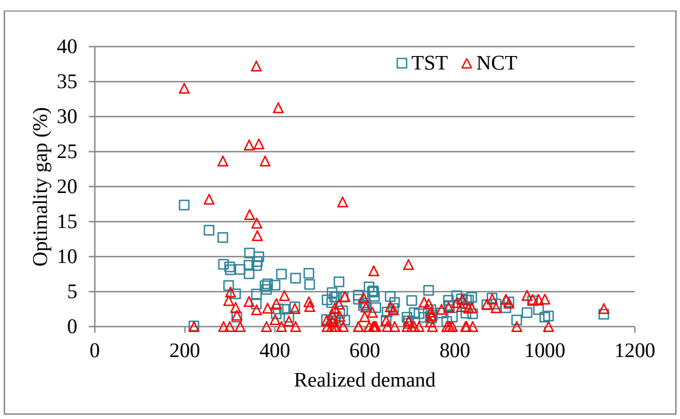


Figure 3.5: The optimality gap for $\alpha = 1$ for TST and NCT.

Yao et al. (2009) and Chung et al. (2011, 2012) use an objective function that is based on minimizing the TST, but modified as follows: $\sum_{t=1}^{T}\sum_{i\in C/S_e} c^t x_i^t$ where $c^t = 1$ for $t \leq t_1$ and $c^t = M$ for $t > t_1$, and M is a large number (a penalty). These papers assume that the uncertain demand falls within a given range, and that any demand realization in excess of the deterministic counterpart is not able to leave its source. Because of these assumption, along with the penalty M for any demand that does not reach the sink by time t_1 (which includes the demand not able to leave its source due to the deterministic counterpart being less than the realized demand), the best solution obtainable by the *DC strategy* from α is set to 1, that is, set the deterministic counterpart to the upper range of demand. The penalty for not evacuating by interval t_1 is related to the NCT. We see from Table 3.1, and Figure 3.5 that setting α to 1 is not optimal for the TST, nor is it optimal for the NCT. If we consider the penalty, the optimality gap between the deterministic counterpart and the optimal policy increases, for instance, if $t_1 = 160$, the optimal policy plan's objective function value will remain unchanged since the NCT is always less than 158, whereas the plans generated from the DC strategy incurs the penalty in some cases. We note that the penalty in the alternative objective function does not alter the DC strategy's plans because for the deterministic problem minimizing TST also minimizes NCT.

In summary, the *DC strategy* does not provide an optimal solution for Models 1 or 2 under demand uncertainty, which we illustrated by using SSSPD networks having optimal policies. The DC strategies have the following disadvantages:

1. Models 1 and 2 often have multiple optimal solutions in the deterministic setting, each

of which can have different results under uncertainty, yielding a high variance on the objective function value.

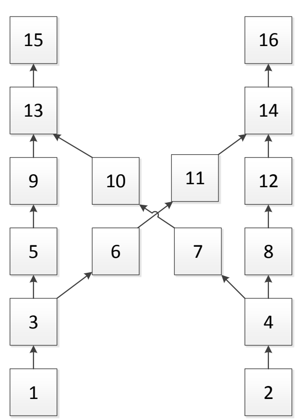
2. A solution generated by Models 1 or 2 require an implementation method, i.e. a set of rules to modify the solution when the demand realization does match the deterministic counterpart, which is not considered in the optimization.

3. The DC strategy solution is difficult to implement since the routing and traffic controls can be complex; solutions often have different priorities and turn percentages at merge and diverge links at each time interval. Conversely the optimal policy is easy to implement

4. Model 1 is difficult to solve, and Model 2, if not used carefully, can generate solutions that have traffic holding, which are not feasible for Model 1, and introduce further sub-optimalities under uncertainty.

5. In the optimal policy, minimizing the TST is equivalent to minimizing the NCT since we are maximizing the flow to the sink at each time interval. This is not the case in the DC strategies, two plans can have relatively close TST, but their NCT can be very different.

For more complex networks there are potentially even more disadvantages in the DC strategy. Consider the network displayed in Figure 3.6, which has two sources and two sinks. We assume that the demand for each source cell is uniform between 0 and 200. The flows are completely characterized by the turn percentages and priorities at cells 3,4,13,14. Each source has two shortest paths to a sink. Due to symmetry, the optimal policy has the following turn percentages: $p_5^t = 0.5$, $p_6^t = 0.5$, $p_7^t = 0.5$, $p_8^t = 0.5$, $\forall t = 1, \cdots, T$ irrespective of the realized demand because it will guarantee that the flow to the sinks will be maximized at all time intervals. Now suppose we use the DC strategy with any α-level. One possible solution has the following turn percentages $p_5^t = 1$, $p_6^t = 0$, $p_7^t = 0$, $p_8^t = 1$ $\forall t = 1, \cdots, T$. This plan is optimal whenever the demands at the source cells are equal. If we apply the first optimal plan to the scenario where the realized demand is 200 at cell 1 and 0 at cell 2, we get a TST of 1900 and a NCT of 15. In this solution the flow rate to the sinks is 20 each time interval. In the case of the DC strategy solution, since we are forcing $y_{36}^t = 0$, the flow rate to the sinks will be reduced to 10 each time interval which is the minimum possible flow. This will be the worst solution in terms of objective function value. The TST in this case is 2900 and the NCT is 25. The optimality gap is more than 50% for the TST and 66% for the NCT. In the next section, we examine networks without an optimal policy under uncertainty.



Cell	diverge cells	other cells
N_i	40	40
Q_i	20	10
δ_i^t	1	1

Figure 3.6: A network example with two sinks having an optimal policy.

3.4 A heuristic for more complex networks

Next we present a heuristic that can be used in more complex networks that do not have optimal policies. In this heuristic, the traffic management tools are turn percentages at diverge cells, which determines routing and priorities at merge cells. We will use the shortest paths subnetwork at its fullest potential, and we will evaluate whether we use the other paths or not.

Control at diverge links

Given two diverge links $(m, n), (m, o) \in L_d$, where n is on the shortest path subnetwork, o is not. Since n is in the shortest path to the sink cells, we will use the route to its maximum potential that is $y^t_{mn} = min(R^t_n, S^t_m)$. To determine y^t_{mo}, we define $\beta_o = |C(o, S_e)|/Q_{min}(C(o, S_e))$. As β_o decreases it will become more useful to send the vehicles through y^t_{mo} since either the distance between o and the shortest sink is small or the flow per time interval $Q_{min}(C(o, S_e))$ is high. So a comparison to β_n seems logical to decide whether we use the other path. Specifically if $\beta_o >> \beta_n$ then it would be better to set $y^t_{mo} = 0$. To determine how much β_o should be bigger than β_n, the comparison should depend on the average total demand connected to cell m: as the average total demand increases we should allow more flow in y^t_{mo}. Thus we came up with the following heuristic: Let

$\bar{d}_m$ be the average total demand connected to cell m. y^t_{mn} and y^t_{mo} are defined as follows:

$$y^t_{mn} = min(R^t_m, S^t_n) \tag{3.15}$$

$$y^t_{mo} = \begin{cases} \min(x^t_m - y^t_{mn}, Q_m - y^t_{mn}, R^t_n) & \text{if } \beta_o < \frac{\bar{d}_m}{Q_{min}(C(n,S_e))^2}\beta_n \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

Equation (3.15) sets the flow on the shortest path to its maximum level. Equation (3.16) describes the flow on the longer path; this flow depend on the values of β_o and β_n. Deciding on whether using the shortest path only or not depends on the demand level, the length of the paths connected to the sink, and the inflow/outflow parameter of these cells: As the demand increases, it is more likely to use more paths than using only shortest paths. Similarly as the length of the shortest path decreases and its flow increases it is more likely to use the shortest path only.

Control at merge links

Given two merge links, $(i, k), (j, k) \in L_m$ and assuming $R^t_k > Q_i$ and $R^t_k > Q_j$, then the flow equations are as follows:

$$y^t_{ik} = \begin{cases} \min(S^t_i, R^t_k) & \text{if } \min[\frac{x^t_i}{Q_i}, \frac{x^t_j}{R^t_k - Q_i}] \geq \min[\frac{x^t_i}{R^t_k - Q_j}, \frac{x^t_j}{Q_j}] \\ \min(S^t_i, R^t_k - y^t_{jk}) & \text{otherwise} \end{cases} \tag{3.17}$$

$$y_{jk}^t = \begin{cases} \min(S_j^t, R_k^t - y_{ik}^t) & \text{if } \min[\frac{x_i^t}{Q_i}, \frac{x_j^t}{R_k^t - Q_i}] \geq \min[\frac{x_i^t}{R_k^t - Q_j}, \frac{x_j^t}{Q_j}] \\ \min(S_j^t, R_k^t) & \text{otherwise} \end{cases} \tag{3.18}$$

If $R_k^t \leq Q_i$ or $R_k^t \leq Q_j$, then the flow equations are as follows:

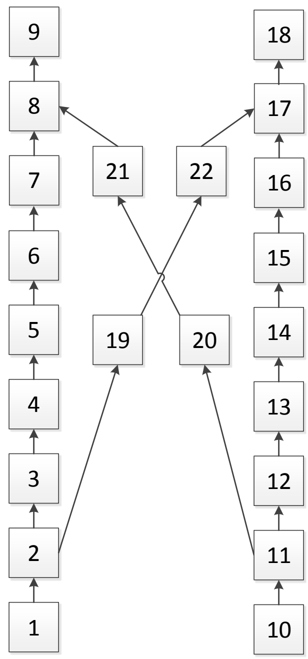
$$y_{ik}^t = \begin{cases} \min(S_i^t, R_k^t) & \text{if } Q_i < Q_j \\ \min(S_i^t, R_k^t - y_{jk}^t) & \text{otherwise} \end{cases} \tag{3.19}$$

$$y_{jk}^t = \begin{cases} \min(S_j^t, R_k^t - y_{ik}^t) & \text{if } Q_i < Q_j \\ \min(S_j^t, R_k^t) & \text{otherwise} \end{cases} \tag{3.20}$$

Through these equations, we are trying to maximize the time we use the merge cell k at full potential Q_k. Equations (3.19)-(3.20) are similar to Proposition 1 in the previous section. In Equations (3.17)-(3.18), $min[\frac{x_i^t}{R_k^t - Q_j}, \frac{x_j^t}{Q_j}]$ is the time merge cell k can be supplied a flow of R_k^t if cell i were given full priority. We evaluate the same ratio for cell j and we give full priority to the cell with the highest ratio.

Testing the heuristic

In this section, we will test the heuristics on several networks. First, considering the network displayed in Figure 3.2, the heuristic had the same performance as the optimal policy. In the network in Figure 3.6, since we have two identical paths at each source cells, the heuristic



Cell	Diverge cells	Other cells
N_i	40	40
Q_i	20	10
δ_i^t	1	1

Figure 3.7: A network example with two sinks without an optimal policy.

uses both paths to their fullest potentials and hence the heuristic also provides the optimal policy.

Now we test the heuristic on a more complex network as in Figure 3.7. We assume that demand at source cells 1 and 10 are uniform between 0 and 200. Each source has a shortest path to a sink and another longer path to another sink. As a result, the optimal solution is dependent on the demand realized at the source cells, and thus there is no optimal policy. For example if the demand is high at a given source cell, it would be better to use the both paths, whereas if the demand is low, using the shortest path only will yield a better solution. The results for the heuristic and the DC strategy (using the priority method) are shown in

Table 3.2.

		DC strategy			
		$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$	Heuristic
TST	mean	26,149	26,529	28,579	24,192
	std	14,559	14,368	15,271	13,854
	max	58,258	58,309	62,876	54,703
NCT	mean	22.69	23.25	23.85	19.68
	std	4.9	4	3.3	4.8
	max	29	29	29	26

Table 3.2: Performance measures of the DC strategy and the heuristic

We next test the heuristic on the Sioux Falls network, which is heavily studied in the transportation literature (see Abdulaal and LeBlanc, 1979). We did not find a CTM implementation of this network in the literature. Our CTM network version of the Sioux Falls network is presented in Figure 3.8. The network parameters are obtained from A. Chakirov (2014). We chose the time interval length to be 0.5 min and the jam density is assumed to be 198 vehicles/mile/lane. Two types of roadways is the Sioux Falls network, highways, which have three lanes and a maximal flow rate of 1800 vehicles/lane/hour, and urban roadways, which have two lanes with maximal flow of 900 vehicles/lane/hour. This yields two types of roadway cells in the CTM network: highway cells (marked as H in the network) having parameters of $Q_h = 45$ and $N_h = 119$, and urban cells (marked as U in the network) having $Q_u = 15$ and

$N_u = 79$. We assume that the demand at all source cells are uniformly distributed between 0 and 2000. We compare the performance of the heuristic and the deterministic counterpart strategies in Table 3.3.

		DC strategy			
		$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$	Heuristic
TST	mean	2.53×10^5	2.64×10^5	2.71×10^5	1.99×10^5
	std	9.44×10^4	8.75×10^4	8.11×10^4	8.38×10^4
	max	4.41×10^5	4.56×10^5	5.19×10^5	4.28×10^5
NCT	mean	120.34	122.13	125.56	110.15
	std	30.01	27.21	26.23	29.21
	max	170	173	175	156

Table 3.3: Performance measures of the DC strategy and the heuristic on the Sioux Falls network

The DC strategy finds a solution with all the issues discussed above, and the heuristic performed better than the DC strategy for all values of α. For the mean TST we find that $\alpha = 1$ performs the worst, and has a mean TST that is 37% larger than that found be the heuristic, and $\alpha = 0.5$ performs the best, but it still has a mean TST that is 27% larger than the heuristic. The heuristic also had better worst-case performance for all values of α and better network clearance time performance.

3.5 Conclusion

In this paper, we illustrate some of the problems in using the deterministic counterpart parameter to replace the stochastic demand in the CTM model: we started with a survey of the deterministic counterpart strategies in the literature, specifically we talk about the chance constrained optimization formulation proposed by Waller and Ziliaskopoulos (2006), and the robust optimization formulation proposed by Yao et al. (2009), along with the assumptions needed to implement it. We next present a class of networks, which are used in the literature and we show that there exists an optimal routing and traffic control policy that is independent of demand, so we can implement the optimal evacuation plan for any demand realization. We use an example of such networks to highlight some of the problems associated with the deterministic counterpart approaches. We also discuss the drawbacks of using the deterministic counterpart strategies in more complex networks where an optimal policy independent of demand does not exist, and we provide a simple heuristic that can outperform the DC strategies.

In the CTM model proposed by Daganzo (1994, 1995), the flow out of a roadway cell is not reduced by congestion on that cell; this property can be exploited when developing a good routing plan. In future research, we should find efficient methods to replace the deterministic counterpart strategies in dealing with demand uncertainty by exploiting that property. We could find more general structures of network that have known optimal solutions independent on demand, or develop heuristics that have known optimal gaps with respect to the optimal solution.

Another research direction is to study demand uncertainty in models where congestion does have an effect on the flow in the network. The model studied in Bish et al. (2013) is a generalization of the CTM where the flow out of a cell is reduced in the presence of congestion. In that setting, an optimal policy independent on demand will probably not exist for many network structures. The complexity of this problem makes it challenging due to the dependency of the flow rate on the level of the congestion in the network, but it can contribute to this area of research, especially in extreme cases where a very high demand level is expected like for example evacuation planning where the demand level is stochastic, and congestion plays a big role in the decision making process.

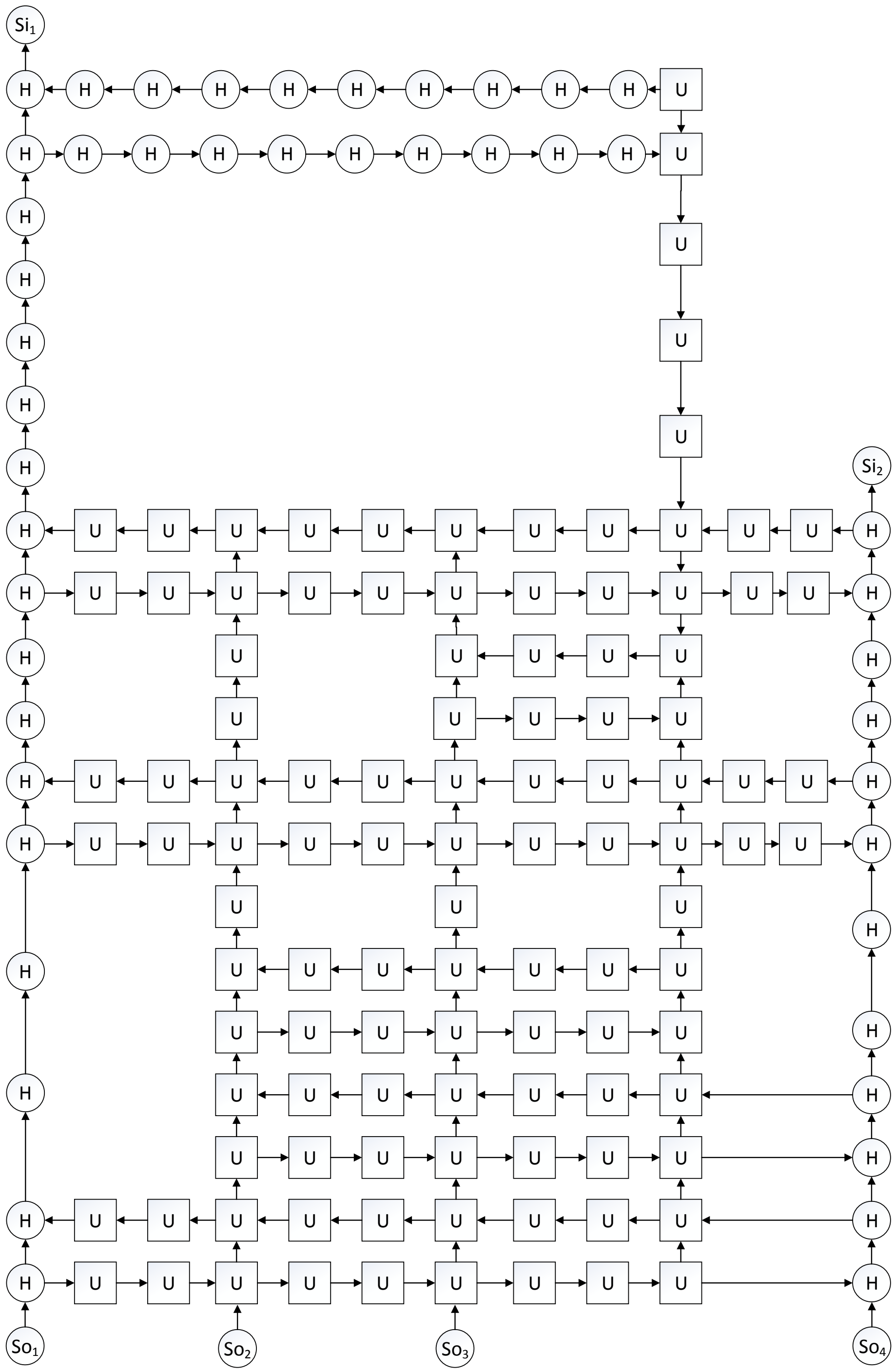


Figure 3.8: Sioux Falls network.

Chapter 4

Traffic management strategies under the Cell Transmission Model with flow reductions

Abstract

The cell transmission model (CTM), implemented in a linear program, is a widely used modeling framework, based on network flows, for studying optimal management strategies for traffic systems. The CTM does not properly model the how congestion can reduce flow rates, and thus we study a generalization of the CTM that does allow for congestion-based flow reductions. Under this modeling framework, we describe five management tools used in optimal traffic strategies, and show how flow reduction changes how these tools are used.

Furthermore, because the tools have various levels of realism (three of the tools involve traffic holding, a consequence of implementing the CTM in a linear program, which are not part of the original CTM framework), we discuss ways to limit the tools that are used to produce a strategy. This requires an integer programming approach, which is much less tractable than the linear programming framework. Because of this we provide a reformulation that reduces the size of the problem, as well as a heuristic.

Keywords: Cell Transmission Model; network flows; dynamic traffic assignments; traffic holding; traffic management strategies

4.1 Introduction

The cell transmission model (CTM) was introduced by Daganzo (1994) to study traffic flow over time. Ziliaskopoulos (2000) implemented a simplified form of the CTM into a linear program with the objective of minimizing the total system time. The model has many applications such as dynamic traffic assignment and scheduling for mass evacuation events (Waller and Ziliaskopoulos, 2006). Solutions obtained from this model however cannot avoid holding vehicles. When vehicle holding occurs, it implies that vehicles stop in the middle of the road even when there is capacity in the downstream cell to move forward. Ziliaskopoulos (2000) briefly discussed traffic holding as a possible set of traffic controls that could be implemented to optimize flow, however Liu et al. (2006) describes traffic holding as an unrealistic behavior for evacuation. Lo (2001) applied a mixed-integer approach to eliminate vehicle-holding. Such an approach however increases the complexity of the problem. Shen et al. (2007) suggests that a post-processing method of the optimal solution provided by the LP model, however it is difficult to obtain path-based flows from the solution. Zheng and Chiu (2011) proposes a network ow algorithm which can efficiently solve the SO-DTA problem with no traffic holding. Nevertheless, this approach is limited to single origin, single sinks networks.

Another modeling problem comes from using the CTM itself: the CTM fails to capture the reduction in the flow discharge rate after the onset of congestion at a bottleneck. Traffic models typically have flows that initially increase with density until a critical density is reached, after which the flow decreases (May, 1990). This has been shown in various empirical studies (Banks, 1990; Hall and Agyemang-Duah, 1991; Cassidy and Bertini, 1999; Chung et al., 2007). Bish et al. (2013) discussed this phenomena and proposed a generalization of the CTM that models the reduction in flow discharge rates at a bottleneck after the onset of congestion. This new model was incorporated into a mixed integer program to eliminate traffic holding. However the formulation is highly inefficient because of its exponential spacial(roadway segments needs to be divided into smaller cells) and time expansion of the problem size.

Merchant and Nemhauser (1978) proposed a non convex and nonlinear model with discretized time steps in which the outflow of an arc in each time only depends on the amount of flow on that arc at the beginning of the time period. Carey (1987) revised this model and transformed this model into a convex one. Nie (2011) proposed an algorithm that can solve this model.

Carey and Eswaran (2000) introduced a time expanded network for flow dependent transit time. For each time period there are several copies of an arc of the underlying network corresponding to different transit times. Capacity constraints are introduced to model the dependency between the flow on all copies of an arc corresponding to different transit times.

Köhler and Skutella (2005) proposed a different model of flow dependent transit time. In this model, flow on an arc depends on the amount of vehicles traversing this arc . An algorithm has been developed to compute a near optimal solution for the problem of sending an amount of flow within minimal time through the network.

One major factor contributing to the problem size for both models is modeling freeway segments: the freeway segments needs to be divided into smaller cells to model it as per the CTM. Although there is no decision making in those segments, decision variables has to be created to model the flow at these cells, which makes the formulation inefficient. In this paper, we address this problem and make the following contributions: 1) The development a better formulation for freeway segments in CTM where only the flow in, and out of the freeway are taken as decision variables; 2) The simplification of the generalized CTM, where the new formulation can be used to solve bigger problems previously not possible with the original formulation; 3) Numerically demonstrating the new formulations and the algorithm, using a bigger size network, which cannot be solved using the original formulation of the generalized CTM and 4) discuss and compare the management strategies of both models.

The remainder of the paper will be structured as follows. Section 4.2 reviews the general CTM proposed by Bish et al. (2013) and the SO-DTA proposed by Ziliaskopoulos (2000), and describe the traffic management strategies that we can extract from these models. In Section 4.3, we propose a simplification that significantly improve the tractability of the model. Section 4.4 illustrates the previous sections by a numerical example. Section 4.5 presents the research conclusions.

4.2 The modeling framework

In this section, we review the cell transmission model (CTM) (Daganzo, 1994, 1995), including a generalization (see, Bish et al., 2013) that provides a more realistic response to congestion, namely a reduction in flow. Then we review a linear program (LP) (Ziliaskopoulos, 2000) that is much used for applications such as dynamic traffic assignment and evacuation planning, and modify it for the more generalized CTM. For this LP, we detail the *tools* used to model to optimize traffic flows, and discuss some properties of this model.

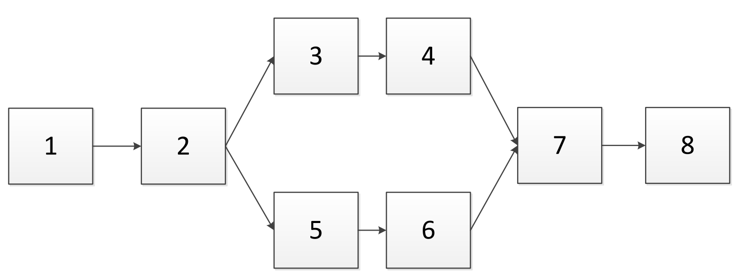


Figure 4.1: A small network example to illustrate the network components used in the CTM.

The CTM utilizes a discrete time-expanded network of cells and links (C, L) to represent the roadway system of interest. Cells can be either source cells (set S_o), sink cells (set S_e), or roadway cells (set R). In this paper we will study a single commodity flow problem, and thus if there are multiple sinks, the model determines the flow sent to each sink. Links represent allowable movements between cells. When a roadway cell has two incoming links, these links are merge links (L_m), likewise when a roadway cell has two outgoing links, these links are diverge links (L_d). We will use the convention of calling a roadway cell having incoming merge links a merge cell, and a roadway cell having outgoing diverge links a diverge cell. All other links are ordinary (L_o); an ordinary link is the only outgoing and incoming link for two adjacent cells. Each source cell has only one outgoing link, while each sink cell has only one incoming link. Figure 4.1 illustrates these network components; Cell 1 is a source, Cell 8 a sink, and cells 2-7 are roadway cells, likewise, links (4,7) and (6,7) are merge links, links (2,3) and (2,5) are diverge links, and links (1,2), (3,4), (5,6) and (7,8) are ordinary. The planning horizon is divided into T time intervals of length τ, and a roadway cell represents a section of roadway of length ℓ such that vehicles traveling at free-flow speed (u_f) traverse the section in one time interval; that is $\ell = u_f \times \tau$. Additional notation follows:

Parameters:

- N_i : maximum number of vehicles that cell i can hold, which is related to the concept of jam density, $\forall\ i \in$ R
- Q_i : maximum attainable flow into or out of cell i, $\forall\ i \in$ R
- Ω_i : flow out of cell i at the maximum traffic density ($x_i^t = N_i$), $\Omega_i \in (0, Q_i)$

δ_i : traffic flow parameter for cell i, $\forall\ i \in$ R

d_i^t : demand generated in cell i in time interval t, $i \in S_o$

We will assume that the parameters Q_i, N_i, Ω_i and δ_i do not vary with respect to time, however it is easy to include time varying parameters. These parameters have the following relationships for any roadway cell i: $Q_i \le N_i$, $0 < \delta_i \le 1$ and $\Omega \le Q$, furthermore, because Q_i is the maximum *attainable* flow into or out of cell i, the parameters must be set such that when $x_i^t \le Q_i$ we have $\delta_i(N_i - x_i^t) \ge Q_i$, otherwise a flow of Q_i would not be attainable. As a result we have $N_i \ge Q_i(1 + \frac{1}{\delta_i})$.

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 flowing from cell i to cell j during time interval t, $\forall\ (i, j) \in L$, $t = 1, \cdots, T$

The maximum possible flow out of cell i in time interval t (S_i^t) and the maximum possible flow into cell j in time interval t (R_j^t) are calculated by (4.1) and (4.2), respectively.

$$S_i^t = \min\{x_i^t, Q_i - (x_i^t - Q_i)(Q_i - \Omega_i)/(N_i - Q_i)\} \tag{4.1}$$

$$R_j^t = \min\{Q_j, \delta_j(N_j - x_j^t)\} \tag{4.2}$$

When $x_i^t \le Q_i$, cell i is in free-flow state and $S_i^t = x_i^t$; when $x_i^t > Q_i$ cell i is in the congested state and $S_i^t = Q_i - (x_i^t - Q_i)(Q_i - \Omega_i)/(N_i - Q_i)$. This linear expression for the flow out of cell i, which is studied in Bish et al. (2013), equals Q_i when $x_i^t = Q_i$ and decreases to Ω_i when $x_i^t = N_i$. By setting $Q_i = \Omega_i$ this expression simplifies to Q_i, the expression used in the original CTM (see, Daganzo, 1994, 1995). Henceforth, we refer to original CTM as $CTM_{\Omega=Q}$, generalized CTM as $CTM_{\Omega\leq Q}$, and CTM with congestion based flow reductions as $CTM_{\Omega<Q}$.

Under $CTM_{\Omega=Q}$ the flow out of cell i is not reduced when i is in the congested state ($x_i > Q_i$), even if the number of vehicles in i reaches N_i, the holding capacity of the cell (i.e., the cells jam density). Conversely, $CTM_{\Omega<Q}$ reduces the flow below Q_i when cell i is that cell is congested state, a much more realistic traffic flow behavior. This observation follows directly from Equation (4.1). This simple modification to the CTM has profound effects on the *realism of the traffic flow*, which we discuss in more detail later on.

Using the flow limits from (4.1) and (4.2) Ziliaskopoulos (2000) formulated an LP that minimizes the total time that vehicles are in the system (i.e., before they reach a sink node) for single commodity flows (i.e., for a networks having a single sink, or multiple sinks where flows are not specified for a particular sinks) for the $CTM_{\Omega=Q}$. We extend this LP formulation for $CTM_{\Omega\leq Q}$ as follows:

$$\text{Minimize} \sum_{t=1}^{T} \sum_{i \in C/S_e} x_i^t \tag{4.3}$$

subject to :

$$x_j^t = x_j^{t-1} + \sum_{i:(i,j)\in L} y_{ij}^{t-1} - \sum_{k:(j,k)\in L} y_{jk}^{t-1}, \quad \forall j \in C/S_e,\ t = 2, \cdots, T \tag{4.4}$$

$$\sum_{j:(i,j)\in L} {y_{ij}}^t \leq x_i^t, \quad \forall i \in C, \ t = 1, \cdots, T \tag{4.5}$$

$$\sum_{j:(i,j)\in L} y_{ij}^t \leq Q_i - (x_i^t - Q_i)(Q_i - \Omega_i)/(N_i - Q_i), \quad \forall i \in R, \ t = 1, \cdots, T \tag{4.6}$$

$$\sum_{i:(i,j)\in L} y_{ij}^t \leq Q_j, \quad \forall j \in R, \ t = 1, \cdots, T \tag{4.7}$$

$$\sum_{j:(j,i)\in L} y_{ji}^t \leq \delta_i^t(N_i - x_i^t), \quad \forall i \in R, \ t = 1, \cdots, T. \tag{4.8}$$

Objective function (4.3) minimizes the *total system time*, that is, the sum of the total time each vehicle spends in the network before reaching a sink. Constraint (4.4) is the flow conservation constraints; here x_i^1 is the total demand at source $i \in S_o$ or the initial number of vehicles on roadway $i \in R$. We note that we could easily modify (4.4) using a time indexed parameter to represent a time variant demand. For each source cell i, we add an additional constraint to ensure that flow is equivalent to $Q_j, \ j : (i, j) \in L$ (source cells have only one outgoing link) to eliminate any staging of vehicles (see Bish and Sherali, 2013; Bish et al., 2014). Constraints (4.5) and (4.6) are the limits for flow out of cell i specified by (4.1), while (4.7) and (4.8) are the limits for flow into cell j specified by (4.2). In addition, the logical nonnegativity constraints on the x and y variables are required.

While this LP models the flow restrictions of (4.1) and (4.2), CTM specifies further constraints on the flows that are nonlinear in nature. The flow over an ordinary link (i, j) during time interval t is calculated by (4.9) and the flow over a pair of merge links (i, k) and (j, k) must adhere to (4.10); when $S_i^t + S_j^t > R_k^t$, (4.10) allows merge priorities to be made for the flows on these links. The flow over a pair of diverge links (i, j) and (i, k) must adhere to (4.11). These disjunctive functions produce a feasible region that is a nonconvex set.

$$y_{ij}^t = \min\{S_i^t, R_j^t\} = \min\{x_i^t, Q_i - (x_i^t - Q_i)(Q_i - \Omega_i)/(N_i - Q_i), Q_j, \delta_j(N_j - x_j^t)\} \quad (4.9)$$

$$y_{ik}^t + y_{jk}^t = \min\{S_i^t + S_j^t, R_k^t\} \quad (4.10)$$

$$y_{ij}^t + y_{ik}^t \leq \min\{S_i^t, R_j^t + R_k^t\} \wedge (y_{ij}^t = R_j^t \vee y_{ik}^t = R_k^t \vee y_{ij}^t + y_{ik}^t = S_i^t) \quad (4.11)$$

Neglecting these CTM flow restrictions for the sake of a linear model results in *traffic holding* (Ziliaskopoulos, 2000), that is, flows that are less than the flows required to satisfy (4.9)-(4.11). Observing that (4.10)-(4.11) allow for traffic management decisions, and including *traffic holding* on various cell types, we describe five *traffic management tools* (TMTs) that can be employed by the LP to minimize the objective function.

TMT 1: Prioritizing at merge links. For merge links (i, k) and (j, k) when flows y_{ik}^t and y_{jk}^t are constrained by (4.10), the model can make flow priority decisions for flows into cell k.

TMT 2: Routing at diverge links. For diverge links (i, j) and (i, k), when flows y_{ij}^t and y_{ik}^t are constrained by (4.11), the model can make routing decisions. All that (4.11) requires is that the flow out of cell i be at its maximum (S_i^t) or the flow into either cell j or k be at its maximum (R_j^t and R_k^t, respectively).

TMT 3: Traffic holding at merge links. For merge links (i, k) and (j, k) when $y_{ik}^t + y_{jk}^t < \min(S_i^t + S_j^t, R_k^t)$ then traffic holding before the merge is used.

TMT 4: Traffic holding at diverge links. For diverge links (i, j) and (i, k) if $y_{ij}^t + y_{ik}^t < \min(S_i^t, R_j^t + R_k^t)$ then traffic holding at the diverge is used.

TMT 5: Traffic holding at ordinary links. For ordinary link (i, j) when $y_{ij}^t < \min(S_i^t, R_j^t)$ then traffic holding on the roadway cell i is used.

Traffic holding at ordinary links is problematic. Ziliaskopoulos (2000) justified traffic holding as a possible set of traffic controls that could be implemented to optimize flows while (Liu et al., 2006) mentions traffic holding as an unrealistic behavior. Ordinary links are a modeling construct, and there might be no analogous traffic controls available, while merges and diverges are natural places for traffic controls.

Traffic management tools 1 and 2 are allowed under CTM equations (4.10) and (4.11), while the TMTs 3, 4, and 5, which involve traffic holding, are not. Of the traffic holding tools, traffic holding at ordinary cells is problematic as specified earlier. On the other hand, merges and diverges are more likely locations for traffic management tools to be implemented. One of our main objectives is to show how these tools are used to optimally manage traffic flows, under both $CTM_{\Omega=Q}$ and $CTM_{\Omega<Q}$.

Proposition 4.1 *In the absence of traffic holding, a roadway cell i in the free-flow state, i.e., $x_i^t \leq Q_i$, will remain in the free-flow state unless the flow out of i is limited by downstream adjacent cell(s).*

Proof. Consider flow in the absence of traffic holding. The flow out of cell i is limited by cell i and the adjacent downstream cell(s) j, $j : (i, j) \in L$. By definition, $x_i^t \leq Q_i$ in the free-flow state, so by (4.1) we have $S_i^t = x_i^t$, and thus the flow out of cell i in interval t will be equal to the number of vehicles in cell i at the beginning of interval t (i.e., x_i^t) unless flow is inhibited by the cell(s) j. Furthermore, by (4.2) at most Q_i vehicles can enter cell i in interval t, hence cell i cannot enter the congested state unless the flow is inhibited by either $y_{ij}^t = Q_j < Q_i$ or $y_{ij}^t = \delta_j(N_j^t - x_j^t) < Q_i$. ∎

Of the two expressions that inhibit the flow out of a cell in the free-flow state (see Proposition 4.1), the first, $y_{ij}^t = Q_j < Q_i$, is based on the structure of the network (e.g., going from a three-lane road to a two-lane road), while the second, $y_{ij}^t = \delta_j(N_j^t - x_j^t) < Q_i$, is based on congestion (the expression $\delta_j(N_j^t - x_j^t)$ can only be less than Q_j if cell j is in the congested state). Consider a CTM network, whenever we have a link $(i,j) \in L$ where $Q_i > Q_j$ we denote cell j as a local bottleneck. We note that this can be more complex, for instance, for a merge cell j even if $Q_i < Q_j$ for both the incoming links, cell j can still be a local bottleneck if $\sum_{i:(i,j)\in L_m} Q_i > Q_j$. Without loss of generality, we simplify the general discussion of local bottlenecks assuming the simple $Q_i > Q_j$ condition. The next proposition describes congestion before a local bottleneck under $CTM_{\Omega=Q}$.

Proposition 4.2 *Consider roadway cell j, a local bottleneck (i.e., $Q_i > Q_j,\ i:(i,j) \in L$), that is initially in the free-flow state. Because $Q_i > Q_j$, with sufficient traffic flow, cell i can enter the congested state, but under $CTM_{\Omega=Q}$ this congestion in cell i cannot reduce the flow through cell j.*

Proof. Under $CTM_{\Omega=Q}$ by (4.9) and $Q_i > Q_j$ we have $y_{ij}^t = \min\{x_i^t, Q_j, \delta_j(N_j - x_j^t)\}$.

Because cell j in initially in the free-flow state, this expression is further simplified to $y_{ij}^t = \min\{x_i^t, Q_j\}$. Since at most a flow of Q_j can enter cell j, cell j will remain in the free-flow state (in the absence of traffic holding) unless the flow out of cell j is restricted to be less than Q_j, which by (4.9) can only occur if downstream cell $k,\ k : (j,k) \in L$ has $Q_j > Q_k$ (which implies that cell k is also a local bottleneck) or is congested enough that $Q_j > \delta_k(N_k - x_k^t)$. ∎

Proposition 4.2 (see Bish et al., 2014, for a related discussion) has some important implications, and is the rational for studying $CTM_{\Omega \leq Q}$, because we want to properly penalize congestion in the sense that congestion can inhibits flow (see Bish et al., 2013), which is consistent with traffic flow theory. Many papers that use CTM (Ziliaskopoulos, 2000) display a trapezoidal (or sometimes triangular) flow density diagram for CTM, which has congestion ($x_i^t > Q_i$) reducing flow. This occurs under $CTM_{\Omega=Q}$ because of the adjacent downstream cell. This lack of flow reduction under $CTM_{\Omega=Q}$ leads naturally to Proposition 4.3.

Furthermore, in the absence of traffic holding, only congestion in cell k ($k : (j,k) \in L$) can reduce the flow through cell j below $S_j^t = \min\{x_j^t, Q_j\}$

Proposition 4.3 *For the LP under $CTM_{\Omega=Q}$ there exists an optimal solution that does not have traffic holding.*

Proof. See Bish and Sherali (2013); Nie (2011); Shen et al. (2007) ∎

We observe that the optimal solution found by solving the LP almost always has traffic holding, by Proposition 4.3 this is just one of the alternate optimal solutions. Proposition 4.3 is intuitive, traffic holding can help improve the solution when traffic congestion reduces the flow rate because it can be used to mitigate congestion, but when congestion has no negative affect on flows, traffic holding does not improve the solution. Bish et al. (2013) provides examples where traffic holding is required for an optimal solution under $CTM_{\Omega<Q}$.

To illustrate the differences between the LP results under $CTM_{\Omega=Q}$ and $CTM_{\Omega<Q}$, as well as some traffic holding issues, consider the following example:

Example 1 *Consider the network in Figure (4.2); All roadway cells have Q-values of 30, N-values of 210, and δ-values of 1. Cells S_o1 and S_o2 are source cells, having* $x^1_{S_o1} = 750$, $x^1_{S_o2} = 750$. *Cell S_e is the sink cell. This network is derived from an example in Bish et al. (2013).*

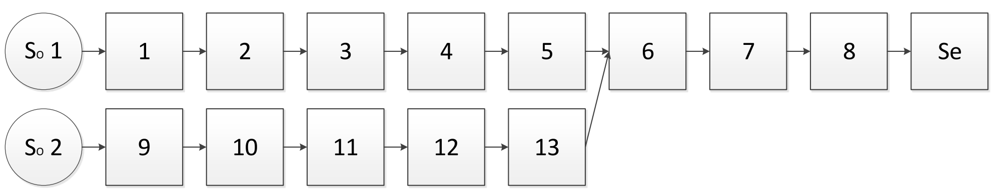


Figure 4.2: Small tree test network.

For this network only three of the traffic management tools can be used, namely TMT 1 (prioritizing at merge links), TMT 3 (traffic holding at merge links), and TMT 5 (traffic holding at ordinary links). Under $CTM_{\Omega=Q}$, from Proposition 4.3 we know that TMTs 3 and 5 are not required. In this example, TMT 3 is actually harmful, because traffic holding at the merge would reduce the flow through cell 6, the local (and only) bottleneck. Under

$CTM_{\Omega=Q}$ the optimal LP solution has an objective function value of 48750. Despite not improving the objective function value under $CTM_{\Omega=Q}$ the LP solution utilizes TMT 5 (traffic holding at ordinary links), for instance, both cells 1 and 9 reach densities of $N = 210$ (the maximum), and, we find intervals where the density of the adjacent downstream cell (cell 2 and 10, respectively) are zero despite maximum density in the preceding cell. Since this is clearly an undesirable, unrealistic, and unneeded deployment of traffic holding as tool, we would like to remove unnecessary traffic holding (i.e., traffic holding that does not improve system performance). Zheng and Chiu (2011) developed an algorithm that generate a solution with no traffic holding for single commodity problems under $CTM_{\Omega=Q}$, but this algorithm will not work under $CTM_{\Omega<Q}$. Lin and Wang (2004) suggested the use of the following alternate lexicographic objective function to remove traffic holding:

$$\text{Minimize} \sum_{t=1}^{T} \sum_{i \in C/C_s} x_i^t + \epsilon \sum_{t=1}^{T} \sum_{(i,j) \in L} t y_{ij}^t, \tag{4.12}$$

where ϵ is a number small enough to optimize the two expressions in preemptive order, but did not elaborate on that value. The secondary objective, $\sum_{t=1}^{T} \sum_{(i,j) \in L} t y_{ij}^t$, rewards advancement of the flow and penalizes traffic holding. We will show how to derive ϵ, but first we consider the following proposition to help us with this endeavour.

Proposition 4.4 *The objective function* $\min \sum_{C/S_e} \sum_{t=1}^{T} x_i^t$ *is equivalent to* $\min \sum_{t=1}^{T} \sum_{i:(i,j) \in S_e} t y_{ij}^t$.

Proof. Consider a unit of flow (i.e., a vehicle) that leaves the system (i.e., enters a sink) in interval t. The contribution to $\sum_{C/S_e} \sum_{t=1}^{T} x_i^t$ for this unit of flow is t (note, all demand is in the system in the first interval), which is also its contribution to $\sum_{t=1}^{T}\sum_{i:(i,j)\in S_e} ty_{ij}^t$. As this is true for every unit of flow, the two objective functions are equivalent. ∎

We assumed that the time horizon T is large enough to allow every vehicle in the system to sink, and note that we have to add the constraint $\sum_{t=1}^{T}\sum_{(i,j)\in L:j\in S_e} y_{ij}^t = \sum_{i\in C/S_e} x_i^1$. This proposition would still hold if the demand at the sources were a time-dependent parameter, but with a constant added to the second objective to equalize the objective function values.

Proposition 4.5 *The objective function (4.12) will minimize* $\sum_{t=1}^{T}\sum_{i\in C/C_s} x_i^t$ *and* $\sum_{t=1}^{T}\sum_{(i,j)\in L} ty_{ij}^t$ *in preemptive order if* $\epsilon < 1/(\max\sum_{t=1}^{T}\sum_{(i,j)\in L} ty_{ij}^t)$

Proof. Sherali (1982) shows that for two objective functions, say f_1 and f_2, to be minimized over a bounded region, for which optimal solutions exist, that $\min f_1 + \epsilon f_2$ will minimize $\{f_1, f_2\}$ in preemptive order if

$$\epsilon < \min(\frac{1}{f_{1max} - f_{1min}}, \frac{1}{f_{2max} - f_{2min}}).$$

Setting $f_1 = \sum_{t=1}^{T}\sum_{i\in C/C_s} x_i^t$ and $f_2 = \sum_{t=1}^{T}\sum_{(i,j)\in L} ty_{ij}^t$ and observing that these functions are bounded and always have non-negative objective function values, we have

$$\min(\frac{1}{f_{1max}}, \frac{1}{f_{2max}}) \leq \min(\frac{1}{f_{1max} - f_{1min}}, \frac{1}{f_{2max} - f_{2min}})$$

Using Proposition 4.4 we have

$$\sum_{C/S_e}\sum_{t=1}^{T} x_i^t \Leftrightarrow \sum_{t=1}^{T}\sum_{(i,j):j\in S_e} ty_{ij}^t < \sum_{t=1}^{T}\sum_{(i,j)\in L} ty_{ij}^t,$$

thus setting $\epsilon < 1/(\max\sum_{t=1}^{T}\sum_{(i,j)\in L} ty_{ij}^t)$ will minimize $\{f_1, f_2\}$ in preemptive order. ∎

Proposition 4.6 *The objective function (4.12) with an ϵ that ensures a lexicographic ordering of the two objectives produces a solution to the LP under $CTM_{\Omega \leq Q}$ that has the minimal amount of traffic holding required to optimize objective function (4.3).*

Proof. Based on the lexicographic property (see Proposition 4.5) objective function (4.12) will always give the optimal solution to (4.3). This objective function will then try to minimize the secondary objective $f_2 = \sum_{t=1}^{T} \sum_{(i,j) \in L} t y_{ij}^t$. This expression provides a time based penalty for movement on the links, if we consider a solution where all demand reaches the sinks on or before time interval T, then every unit of flow has a given path from its source cell to a sink cell, and thus set number of link traversals. To minimize f_2 flow will be advanced as much as possible on its given path considering the network configuration and the primary objective. ∎

We note that by Proposition 4.3 a solution with no traffic holding exists for LP under $CTM_{\Omega=Q}$. So this objective function will generate an optimal solution with no traffic holding for LP under $CTM_{\Omega=Q}$.

Removing traffic holding using objective function (4.12) still produces an optimal solution, and in this solution the number of vehicles in both cells 5 and 13 reach $N = 210$. Thus, in the LP solutions under $CTM_{\Omega=Q}$, either with or without traffic holding, certain roadway cells reach the maximum density (which for this case was the jam density, see Bish et al., 2013). We contend that both these solutions are not realistic. Consider the case without traffic holding, at the bottleneck, where two identical roadways merge, we have high levels of congestion, yet the flow through the bottleneck is not reduced. This is because $CTM_{\Omega=Q}$ has no mechanism to reduce the flow after the onset of congestion at a bottleneck.

Now we consider TMT 1, both LP under $CTM_{\Omega=Q}$ use TMT 1 (merge priorities) in an erratic manner, shifting between 100% priority for cell 5 and 13 (e.g., having six time intervals in a row where cell 5 has 100% priority, then shifting to 100% priority for cell 13 for one interval, before once more giving cell 5 full priority for another two intervals). Interestingly, for this example, under $CTM_{\Omega=Q}$, *any solution with no traffic holding is optimal* since both cells 5 and 13 can supply the bottleneck at its maximum flow rate. For example, we could give cell 5 100% priority until all traffic from source S_o1 is exhausted, and then allow cell 13 to send traffic to cell 6, or we could give cell 5 and 13 each 50% priority throughout; the only thing the optimal solution under $CTM_{\Omega=Q}$ requires is that cell 6 be used to its fullest flow potential as defined by Q_6. Thus the LP solution in this case produces just one of many solutions, and the solution produced is not necessarily very sensible.

Under $CTM_{\Omega<Q}$, specifically $CTM_{\Omega} = 0.2Q$, the LP solution also yields an objective function of 48750, but now the LP must use different management strategies (i.e., combinations of TNTs 1 and 5), as congestion can potentially reduce the cells ability to send out flow. Interestingly, we still observe very high traffic densities (once again up to 210), but the overall strategy changes because the flow reduction caused by these high densities must be considered. Because there is more traffic holding than is needed, we once again use objective function (4.12) to eliminate traffic holding that does not contribute to an optimal solution. Figure 4.3 displays the flow out of the bottleneck (i.e., $y^t_{6,7}$) and into the bottleneck

from cells 5 and 13 (i.e., $y^t_{5,6}$ and $y^t_{13,6}$, respectively). To enable these flows, the traffic

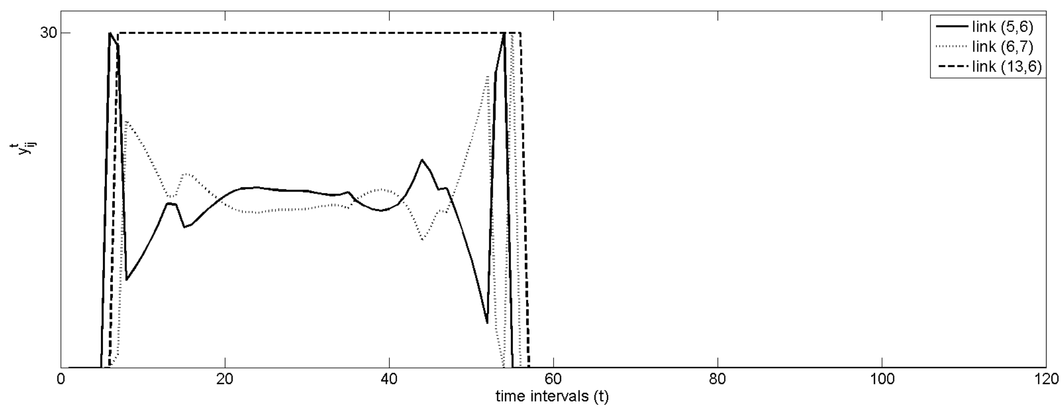


Figure 4.3: Merge and bottleneck flows obtained when the solving the LP under $CTM_{\Omega=0.2Q}$ and objective function (4.12) using TMTs 1 and 5.

holding produces the highest traffic densities on cells 4 and 12, protecting cells 5 and 13 from congestion. Also, the highest densities are 168 and 169 vehicles, respectively, for cells 4 and 12, higher densities would reduce flow to the bottleneck. When we eliminate traffic holding completely, the new objective function value is 67827. Figure 4.4 shows the flows for this solution on links $(5, 6)$,$(13, 6)$ and $(6, 7)$ (similar to Figure 4.3). So, without TMT 5, the total system time is much higher, as well as the *network clearance time* (NCT), which was 108 without traffic holding. To better understand TMT 1, consider a strategy that would be optimal under $CTM_{\Omega=Q}$, namely, giving each or the merge links 50% priority. Figure 4.5 displays the flows into and out of the bottleneck for this case; here the TST is 74549 and the NCT is 93 which is the optimal network clearance time. This result shows another interesting difference between solutions under $CTM_{\Omega=Q}$ and $CTM_{\Omega<Q}$. Zheng and Chiu (2011) show that for single commodity problems an Earliest Arrival Flow (EAF) solution

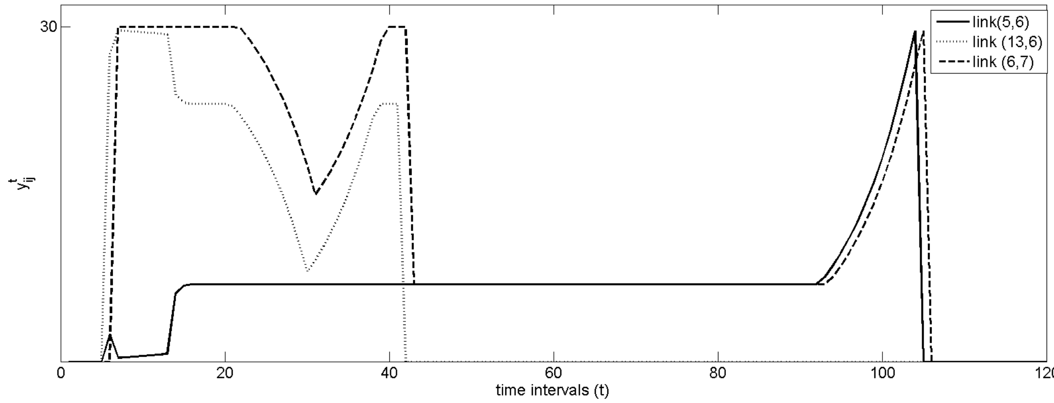


Figure 4.4: Merge and bottleneck flows obtained when the solving the LP under $CTM_{\Omega=0.2Q}$ using TMT 1 only (traffic holding at ordinary cells is eliminated).

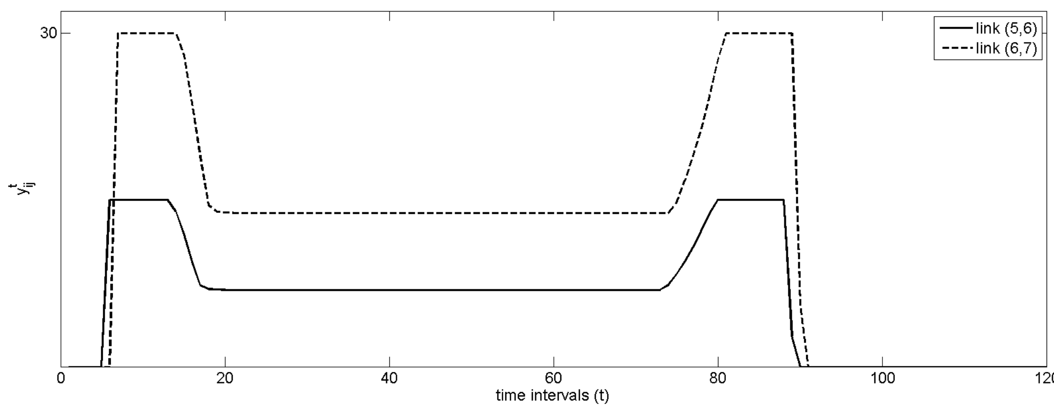


Figure 4.5: Merge and bottleneck flows obtained when the solving the LP under $CTM_{\Omega=0.2Q}$ using TMT 1 with 50% merge priority (no traffic holding).

exists under $CTM_{\Omega=Q}$. An EAF solution maximizes $\sum_{t=1}^{t'} \sum_{(j,s)\in L} y_{js}^t,\ \forall t' = 1, \cdots, T$ and it is equivalent to minimizing objective function (4.3). Likewise, Jarvis and Ratliff (1982) show that for dynamic networks without congestion-based flow reduction that EAF solutions also minimize the network clearance time, the time interval when the last flow enters a sink (note, an EAF solution in a multi-commodity may not exist Fleischer, 2001). As we have just demonstrated, under $CTM_{\Omega<Q}$ an EAF solution does not always exist.

For $CTM_{\Omega=Q}$, flow is not reduced after the onset of congestion, which is not consistent with traffic flow theory (see Bish et al., 2013). As a result, even if the number of vehicles on a roadway segment reaches the maximum number possible ("bumper-to-bumper traffic"), which might happen before a bottleneck, there is no reduction of flow from that segment. As a consequence of this, *traffic control strategies derived from an optimization framework under* $CTM_{\Omega=Q}$ *are insensitive to congestion.* Furthermore, *optimization frameworks that utilize* $CTM_{\Omega<Q}$ *are, under certain conditions, more difficult to solve.*

When solving the LP with $CTM_{\Omega\leq Q}$, we assume that we can use the 5 TMTs, however sometimes traffic holding at certain links is not desired. We next discuss how to get an optimal solution for the problem with $CTM_{\Omega\leq Q}$ while eliminating various traffic management tools.

4.3 Solution methodologies

In this section we provide solution methodologies for solving the LP under $CTM_{\Omega\leq Q}$ considering additional restrictions on the TMTs that can be used. If we allow all five TMTs, then the solution can be obtained directly from the LP, and by using the objective function (4.12), we can eliminate all traffic holding (TMTs 3, 4 and 5) under $CTM_{\Omega=Q}$ and limit unnecessary traffic holding under $CTM_{\Omega<Q}$. But to eliminate all the traffic holding under $CTM_{\Omega<Q}$ is

more difficult, as this requires the addition of nonlinear constraints to the basic LP, or the addition of binary variables. As mentioned earlier, the traffic holding TMTs violate the CTM equations (4.9)-(4.11), and thus it might not be desirable to use them, and TMT 5 is especially problematic, as ordinary cells are an artifact of the modeling framework and do not necessarily aligned with any controllable section of roadway.

Before we begin, we provide a network example to evaluate our solution methodologies, and later, to illustrate the optimal strategies.

Example 2 *Consider the network in Figure (4.6); The freeway cells (the cells represented using a circle) have Q-values of 12 and N-values of 24, while the arterial cells (roadway cells represented by squares) have a flow Q-values of 3 and N-values of 10. All roadway cells have δ-values of 1. Cells 54, 55, and 56 are source cells, having $x^1_{54} = 500, x^t_{56} = 150, x^t_{55} = 150$. Cell 59 is the sink cell.*

First we introduce a heuristic algorithm for solving this problem, where we iteratively solve the LP under $CTM_{\Omega \leq Q}$ using the lexicographic objective function (4.12) and adding additional linear constraints, depending on the traffic holding TMTs we wish to eliminate. Let CS_0 be the set of constraints defining the feasible region of the LP given by (4.4)-(4.8), and CS_n be the set of constraints for the n^{th} iteration of the following algorithm, where $n = 1, \cdots, T$. We note that Steps 2 and 3 of the algorithm can be omitted if traffic holding is to be allowed on merge and diverge links, respectively.

Step 0: Solve the LP under $CTM_{\Omega \leq Q}$ with objective function (4.12) and constraint set S_0 to

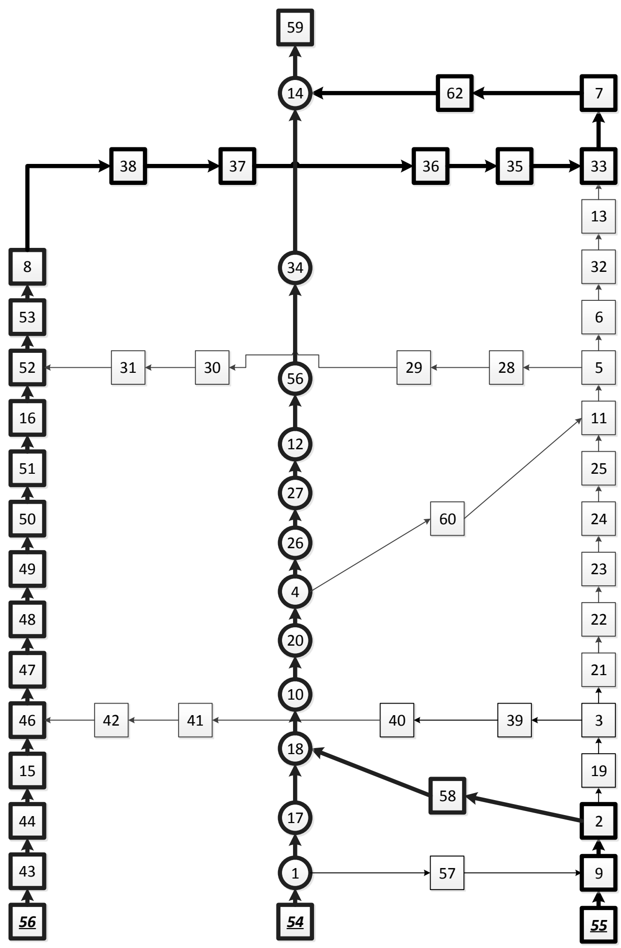


Figure 4.6: Large test network (see Waller and Ziliaskopoulos, 2006; Yao et al., 2009).

obtain the solution $(\hat{x}_i^t, \hat{y}_{ij}^t)$. Check for traffic holding, if no traffic holding exists, then stop, this solution is the optimal solution without traffic holding.

Step 1: Let t_{hold} be the first time interval that traffic holding is detected in solution $(\hat{x}_i^t, \hat{y}_{ij}^t)$. Add the following constraints to the set CS_{n-1} to begin constructing the new feasible region CS_n. To fix flows before time interval t_{hold} add constraints:

$$y_{ij}^t = \hat{y}_{ij}^t, \forall (i,j) \in L, \quad t = 1, \cdots, t_{hold} - 1$$

To eliminate TMT 1 in time interval t_{hold} add the following constraints:

$$y_{ij}^{t_{hold}} = \min(S_i^{t_{hold}}, R_j^{t_{hold}}), \quad \forall (i,j) \in L_o,$$

where the S and R parameters are calculated from the current solution $(\hat{x}_i^t, \hat{y}_{ij}^t)$.

Step 2: To eliminate TMT 3 at merge links $(i,k),(j,k) \in L_m$, add the following constraints to the definition of CS_n:

$$y_{ik}^{t_{hold}} + y_{jk}^{t_{hold}} = \min(S_i^{t_{hold}} + S_j^{t_{hold}}, R_k^{t_{hold}})$$

Step 3: To eliminate TMT 4 at diverge links $(i,j),(i,k) \in L_d$, add the following constraints to the definition of CS_n:

$$y_{ij}^{t_{hold}} + y_{ik}^{t_{hold}} \geq z_{1i} S_i^{t_{hold}}$$

$$y_{ij}^{t_{hold}} \geq z_{2i} R_j^{t_{hold}}$$

$$y_{ik}^{t_{hold}} \geq (1 - z_{1i} - z_{2i}) R_k^{t_{hold}}$$

$$z_{1i}, z_{2i} \ binary$$

We note that these binary constraints force at least one of the upper limits S_i^t, R_j^t, R_k^t to be tight. In many circumstances, these constraints can be simplified, potentially eliminating one or both binary variables, when a particular upper bound is not feasible. For instance, when $S_i^{t_{hold}} < R_j^{t_{hold}} + R_k^{t_{hold}}, S_i^{t_{hold}} \leq R_j^{t_{hold}}, S_i^{t_{hold}} \leq R_k^{t_{hold}}$ the flow cannot be at the limits specified by either R parameter, and we need only add the constraint $y_{ij}^{t_{hold}} + y_{ik}^{t_{hold}} = S_i^{t_{hold}}$.

Step 4: Solve the LP under $CTM_{\Omega \leq Q}$ with objective function (4.12) and constraint set CS_n, i.e., under a new more constrained feasible region, to obtain the new solution $(\hat{x}_i^t, \hat{y}_{ij}^t)$. Check for traffic holding, if no traffic holding exists, then stop, the solution is feasible. Otherwise go to Step 1.

Proposition 4.7 *The objective function (4.12) will minimize* $\sum_{t=1}^{T} \sum_{i \in C/C_s} x_i^t$, $\sum_{t=1}^{T} \sum_{(i,j) \in L} t y_{ij}^t$ *in preemptive order for all iterations in the algorithm if* $\epsilon < \frac{1}{max(f_2 : x, y \in CS_0)}$

Proof. By Proposition 4.5, this value for ϵ will work for the first iteration. Moreover

$$
\begin{aligned}
& max(f_2 : x, y \in CS_0) \geq max(f_2 : x, y \in CS_n) \\
\Rightarrow & \frac{1}{max(f_2 : x, y \in CS_0)} \leq \frac{1}{max(f_2 : x, y \in CS_n)}
\end{aligned}
$$

So any value for ϵ that minimizes $\{f_1, f_2\}$ in preemptive order over CS_0 will also minimize $\{f_1, f_2\}$ in preemptive order over CS_n $\forall n$. ∎

This algorithm iteratively solves the LP, adding additional constraints at each iteration. In $CTM_{\Omega < Q}$, Proposition 4.3 no longer holds, because in the case of $CTM_{\Omega < Q}$, traffic holding can improve the objective function by unrealistically avoiding the effects of congestion. However, the algorithm will generate a feasible solution with no traffic holding for $CTM_{\Omega < Q}$. We denote $\{x_{i(sol)}^t, y_{ij(sol)}^t\}$ to be the solution generated by the algorithm. We now have the following:

$$\min(\sum_{t=1}^{T}\sum_{i\in C/C_s} x_i^t, (x_i^t, y_{ij}^t) \in CS_0) \leq \sum_{t=1}^{T}\sum_{i\in C/C_s} x_{i(sol)}^t \tag{4.13}$$

In the set CS_0, we assume we have full traffic control over all sets. So now we have a lower bound which is the objective function value at the first iteration and a higher bound which is the objective function value for the solution generated by the algorithm. Hence whenever $\min(\sum_{t=1}^{T}\sum_{i\in C/C_s} x_i^t, (x_i^t, y_{ij}^t) \in CS_0) = \sum_{t=1}^{T}\sum_{i\in C/C_s} x_{i(sol)}^t$, the solution generated is optimal.

Of course, this algorithm is not required when minimizing the total system time under $CTM_{\Omega=Q}$ (see Proposition 4.6, but even without using the lexicographic objective function, (4.12), this algorithm will still produce an optimal solution under $CTM_{\Omega=Q}$ since we can easily show that the objective function value will remain the same after each iteration.

In the algorithm, the first iteration we solve an LP with $|L|T$ constraints where $|L|$ is the number of links in the network (the x variables can be completely removed from the LP by Equation (4.4)) and the size the LP will decrease as we are solving the different iterations, and as a result their corresponding solving time will be reduced. Next we discuss how to generate an optimal solution for $CTM_{\Omega\leq Q}$ in general.

In order to eliminate traffic holding at ordinary links, Bish et al. (2013) formulated the problem by adding binary variables that insure that the flow on the ordinary links adheres to (4.9). This requires three binary variables for each ordinary link for each time interval.

Despite valid inequalities that make this mixed binary program more tractable, it cannot be used to solve large networks.

We propose a simplification under $CTM_{\Omega \leq Q}$ for the special case where $\delta = 1$, which Dixit et al. (2008); Kalafatas and Peeta (2006) justify for freeway models. The network in Example 2 also had $\delta = 1$(Waller and Ziliaskopoulos, 2006; Yao et al., 2009). we first show that Constraint (4.6) is redundant for a series of identical ordinary cells, then using this insight, we show that these cells can be aggregated using variable substitutions, and finally, we provide an equivalent objective function formulation and attendant side constraint for minimizing total system time, which is useful given our variable substitution, which eliminates variables required for objective function (4.3).

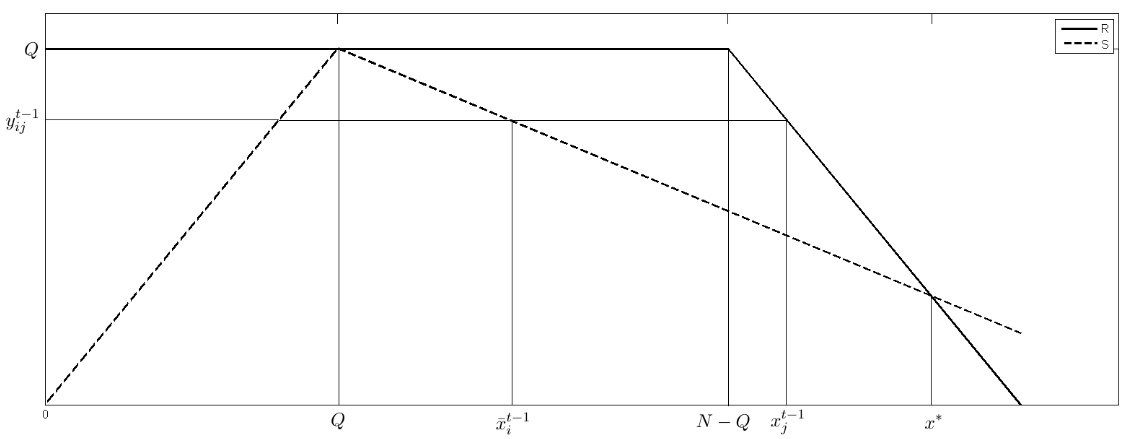


Figure 4.7: S and R versus x.

Proposition 4.8 *For a series of n identical ordinary cells ($Q_i = Q$, $N_i = N$, $\Omega_i = \Omega$, and $\delta_i = 1$, $i = 1, \cdots, n$) that are initially in the free-flow state, in the absence of traffic holding Constraint (4.6) does not restrict the flow over the $n-1$ links connecting these cells.*

Proof Since all cells are identical, the graphs for S and R across all cells should be the same. Figure 4.7 shows the plot of S and R versus density. The S and R curves intercept at two points, the first point is Q and we label the second one as x^*.

Consider an arbitrary link (i, j) between two consecutive nodes in the series of ordinary cells. By (4.9) if $x_i^t \leq Q$ or $x_j^t \geq x^*$, (4.6) is not the minimizer of (4.9). So, we must examine the case where $x_i^t, x_j^t \in (Q, x*)$. Suppose cell i becomes congested at time t'. By Proposition 4.1, cell i can only enter the congested state when $x_i^t \geq (N_j - x_j^t)$ and $N_j - x_j^t \leq Q_j$. Let us suppose that Constraint (4.6) does limit flow over arc (i, j) and let $t > t'$ be the first time interval that it does so, which implies the following:

$$y_{ij}^t = Q - \frac{(x_i^t - Q)(Q - \Omega)}{N - Q} < N_j - x_j^t \tag{4.14}$$

Because time interval t is the first interval where (4.14) holds, we know that in time interval $t-1$, $y_{ij}^{t-1} = N_j - x_j^{t-1}$. Furthermore we have $x_j^t = x_j^{t-1} + y_{ij}^{t-1} - y_{jk}^{t-1}$ from (4.4), and substituting $N_j - x_j^{t-1}$ for y_{ij}^{t-1} in this expression and simplifying yields $x_j^t = N_j - y_{jk}^{t-1}$ or $y_{jk}^{t-1} = N_j - x_j^t$ and using (4.14) we see that $y_{jk}^{t-1} < y_{ij}^t$.

On the other hand we have $N_j - Q_j < x_j^{t-1} < x_j^t < x^*$, since congestion at cell i cannot start unless this condition is satisfied.

However in interval $t-1$, $y_{ij}^{t-1} = N_j - x_j^{t-1}$ because t is the first interval that (4.6) determines the flow over arc (i, j). Also the flow into cell i is bounded by $N - x_i^{t-1}$ at time $t-1$.

$$x_i^t \ \leq \ x_i^{t-1} + (N - x_i^{t-1}) - y_{ij}^{t-1}$$

$$
\begin{aligned}
&= x_i^{t-1} + (N - x_i^{t-1}) - (N - x_j^{t-1}) \\
&= x_j^{t-1}
\end{aligned}
$$

So $x_i^t \leq x_j^{t-1}$. Using this inequality we get:

$$
\begin{aligned}
y_{i,j}^t &= Q - \frac{(x_i^t - Q)(Q - \Omega)}{N - Q} \\
&\geq Q - \frac{(x_j^{t-1} - Q)(Q - \Omega)}{N - Q} \\
&\geq y_{j,k}^{t-1}
\end{aligned}
$$

And this is a contradiction. So Constraint (4.6) is never binding and the proposition holds.∎

This proposition is in line with the observation from Bish et al. (2013) that under $CTM_{\Omega \leq Q}$ without traffic holding,as congestion builds, the density shockwave propagates in a backward wave from the bottleneck(upstream), and as the congestion dissipates, the density shockwave propagates in a forward recovery wave towards the bottleneck (downstream).

Proposition 4.9 *For a series of n identical ordinary cells under $CTM_{\Omega=Q}$ ($Q_i = Q$, $N_i = N$, and $\delta_i = 1$, $i = 1, \cdots, n$) that are initially in the free-flow state, in the absence of traffic holding we can aggregate the cells and calculate S_1^t and R_n^t as follows:*

$$X_1^t = X_1^{t-1} + y_{in}^{t-1} - y_{out}^{t-n} \tag{4.15}$$

$$X_n^t = X_n^{t-1} + y_{in}^{t-n} - y_{out}^{t-1} \tag{4.16}$$

$$R_1^t = \min(n(N) - X_1^t, Q) \tag{4.17}$$

$$S_n^t = \min(X_n^t, Q), \tag{4.18}$$

where y_{in}^t is the number of vehicles that enter cell 1 in interval t (which is constrained by (4.17) and the state of the cell(s) directly upstream of cell 1), y_{out}^t is the number of vehicles that exit cell n in time interval t (which is constrained by (4.18) and the state of the cell(s) directly downstream of cell n), X_1^t represents the number of vehicles that can reach cell 1 from the congestion that starts at cell n, while X_n^t represents the number of vehicles that can reach cell n by time t.

Proof. If $n = 1$, this system of equations (4.15)-(4.17) are equivalent to the original CTM equations.

Suppose Proposition 4.9 is true for n, i.e. that we can aggregate n identical cells. We will prove it for $n + 1$ cells.

Given we have a freeway consisting of $n + 1$ cells, by the induction hypothesis, we can aggregate cell 2 through $n + 1$:

$$S_{n+1}^t = \min(X_{n+1}^t, Q)$$

$$X_{n+1}^t = X_{n+1}^{t-1} + y_{12}^{t-n} - y_{out}^{t-1}$$

If cell $n+1$ is congested at time t then $S_{n+1}^t = Q_{n+1}$, and this is similar to the original CTM constraint for S_{n+1}^t.

Suppose cell $n + 1$ is in free flow state at time t. By Proposition 4.1, cell 1 will be in the free-flow state at $t - n - 1$.

$$y_{12}^{t-n} = x_1^{t-n}$$

$$
\begin{aligned}
&= x_1^{t-n-1} + y_{in}^{t-n-1} - y_{12}^{t-n-1} \\
&= x_1^{t-n-1} + y_{in}^{t-n-1} - x_1^{t-n-1} \\
&= y_{in}^{t-n-1}
\end{aligned}
$$

Replacing this value of y_{in}^{t-n-1}, the expression S_{n+1}^t is equal to the following:

$$
\begin{aligned}
S_{n+1}^t &= X_{n+1}^t \\
&= X_{n+1}^{t-1} + y_{12}^{t-n} - y_{out}^{t-1} \\
&= X_{n+1}^{t-1} + y_{in}^{t-n-1} - y_{out}^{t-1}
\end{aligned}
$$

So the proposition is now proven for S_{n+1}^t. We next prove the theorem for R_1^t.

$$R_1^t = \min(N_1 - x_1^t, Q_1)$$

If cell 1 is in the free-flow state at time t then $R_1^t = Q_1$, and this is the same as the original CTM constraint for R_1^t.

Suppose cell 1 is congested at t. By Proposition 4.1, cell 2 will be congested at time $t-1$. Since cell 2 through $n+1$ can be aggregated by the induction hypothesis, we have:

$$
\begin{aligned}
&x_1^t = x_1^{t-1} + y_{in}^{t-1} - y_{12}^{t-1} \\
&y_{12}^{t-1} = \sum_{i=2}^{n+1} N_i - X_2^{t-1} \\
&X_2^t = X_2^{t-1} + y_{12}^{t-1} - y_{out}^{t-n-1}
\end{aligned}
$$

As a result, R_1^t can be rewritten in the following manner:

$$\begin{aligned}
R_1^t &= N_1 - x_1^t \\
&= N_1 - (x_1^{t-1} + y_{in}^{t-1} - y_{12}^{t-1}) \\
&= \sum_{i=1}^{n+1} N_i - (x_1^{t-1} + y_{in}^{t-1} + X_2^{t-1})
\end{aligned}$$

The second term in the parenthesis can be simplified by the following:

$$x_1^{t-1} + X_2^{t-1} + y_{in}^{t-1} = x_1^{t-2} + y_{in}^{t-2} + X_2^{t-2} - y_{out}^{t-n-1} + y_{in}^{t-1} \tag{4.19}$$

Let $X_1^t = x_1^{t-1} + X_2^{t-1} + y_{in}^{t-1}$, by Equation (4.19):

$$X_1^t = X_1^{t-1} + y_{in}^{t-1} - y_{out}^{t-n-1} \tag{4.20}$$

R_1^t can be simplified as follows:

$$R_1^t = \sum_{i=1}^{n+1} N_i - X_1^t$$
$$X_1^t = X_1^{t-1} + y_{in}^{t-1} - y_{out}^{t-n-1}$$

So the proposition is now proven. ∎

Using Proposition 4.8 and 4.9 together under $CTM_{\Omega \leq Q}$ we can aggregate the first $n-1$ cells in a series of identical ordinary cells because Constraint (4.6) is redundant for the first $n-1$ links connecting these cells and this constraint is only binding for the link going out of the last cell of the freeway. To keep track of the flow reduction at the exit of the freeway, the density at the last cell should be known at all times, hence the last cell should not be aggregated. Because the flow into cell n is not included in the aggregation, the last link may suffer from traffic holding . To remedy this, and stop traffic holding, we introduce the following binary variables to between the aggregate cells and cell n.

$$X_1^t = X_1^{t-1} + y_{in}^{t-1} - y_{n-1,n}^{t-n+1} \tag{4.21}$$

$$X_{n-1}^t = X_{n-1}^{t-1} + y_{in}^{t-n+1} - y_{n-1,n}^{t-1} \tag{4.22}$$

$$x_n^t = x_n^{t-1} + y_{n-1,n}^{t-1} - y_{out}^{t-1} \tag{4.23}$$

$$y_{in}^t \leq Q \tag{4.24}$$

$$y_{in}^t \leq (n-1)N - X^t \tag{4.25}$$

$$q^t - Q = e1_n^{t+} - e1_n^{t-} \tag{4.26}$$

$$q^t - e1_n^{t+} - (N - x_n^t) = e2_n^{t+} - e2_n^{t-} \tag{4.27}$$

$$y_{n-1}^t = q^t - e1_n^{t+} - e2_n^{t+} \tag{4.28}$$

$$y_{out}^t \leq x_n^t \tag{4.29}$$

$$y_{out}^t \leq Q - \frac{(x_n^t - Q)(Q - \Omega)}{N - Q} \tag{4.30}$$

$$e1_n^{t+} \leq M(z1_n^t),\ e1_n^{t-} \leq M(1 - z1_n^t) \tag{4.31}$$

$$e2_n^{t+} \leq M(z2_n^t),\ e2_n^{t-} \leq M(1 - z2_n^t) \tag{4.32}$$

$$e1_n^{t\pm} \geq 0, \quad z1_n^t \in \{0, 1\} \tag{4.33}$$

$$e2_n^{t\pm} \geq 0, \quad z2_n^t \in \{0, 1\} \tag{4.34}$$

It is noteworthy that for $\delta < 1$, Constraint (4.6) is not redundant for the first few links, and as a result, the above simplification cannot be made if $\delta < 1$.

Due to the aggregation, we have to use the alternate objective function proposed in Proposition 4.4, since we cannot keep track of all the cell densities anymore.

Using Proposition 4.9, we can completely remove the flows $y_1^t, \cdots, y_n^t$ as decision variables from the problem with their corresponding constraints. As a result a total of nT decision variables and $4nT$ corresponding constraints can be removed from the SO-DTA problem.

We can also implement this simplification in the algorithm, and it will reduce its solving time. First it will reduce the total number of decision variables and a result it will reduce the solving time in each iteration. Secondly, it will reduce the number of iterations required, since at these freeway lanes, there will be no traffic holding.

Proposition 4.8 states that using $CTM_{\Omega=Q}$ or $CTM_{\Omega\leq Q}$ for cells not upstream adjacent to a bottleneck produces identical results. This happens because as the density of cell n increases until it will reach an equilibrium density x^*, where x^* satisfy the following equation:

$$N - x^* = Q - \frac{(x^* - Q)(Q - \Omega)}{N - Q}$$

And given enough demands all cells will attain the same equilibrium density regardless of the value of its Ω. This exchangeability of $CTM_{\Omega=Q}$ or $CTM_{\Omega\leq Q}$ for cells not upstream further illustrates the problems with $CTM_{\Omega=Q}$ at bottlenecks, because the flow reduction is solely based on the density of the adjacent cell which is ineffective at a bottleneck. So from Proposition 4.8, $CTM_{\Omega=Q}$ or $CTM_{\Omega\leq Q}$ behaves in a similar fashion for cells not upstream.

And this is consistent with Daganzo (1994, 1995) except for all cells at bottlenecks. Thus bottlenecks drive the overall performance of the network.

4.4 Numerical examples

The objective of this section is to illustrate the following by the use of numerical examples:

- Evaluate the performance of the heuristic
- Compare the tractability of the old and new formulations of $CTM_{\Omega \leq Q}$
- Compare the solutions generated by using two different values of Ω.

We start with the network in Example 1. The management strategies has been discussed in detail in section 4.2. We denote ILP as the original formulation proposed by Bish et al. (2013) to remove all traffic holding, and ILP_n the formulation where cells are aggregated. In ILP_n, the only decision variables are flows over links $(4,5)$,$(12,13)$,$(5,6)$,$(13,6)$ and $(6,S_e)$. Table 4.3 displays the solution time for different values of Ω_i. As Table 4.3 illustrates, the performance of the alternate formulation is much better than the original ILP.

We then test this formulation on a bigger size network where the original formulation could not solve due the high number of binary decision variables.

We set a 2 hour time limit to get a solution. Table 4.4 displays the solution time for different values of Ω_i for example 2.

Ω_i	$0.8Q_i$	$0.7Q_i$	$0.6Q_i$	$0.5Q_i$	$0.4Q_i$	$0.3Q_i$	$0.2Q_i$
ILP	3.2	66.4	84.5	79.8	140.9	186.3	319.6
ILP_n	1.01	2	6.2	12	13.1	30.5	31

Table 4.3: Run times (seconds) for the original ILP and ILP_n, for a range of Ω values.

Ω_i	$0.8Q_i$	$0.7Q_i$	$0.6Q_i$	$0.5Q_i$	$0.4Q_i$	$0.3Q_i$	$0.2Q_i$
ILP	limit	limit	limit	limit	limit	limit	limit
ILP_n	51	92	120	208	966	1293	2600

Table 4.4: Run times (seconds) for ILP and ILP_n, for a range of Ω values.

The objective function was the same for all values of Ω. When we ran the algorithm on this network for $\Omega = 0.2Q$, the running time was under 2 minutes and we had $z_{\Omega=0.2Q} = z_{\Omega=Q}$ which means that the solution obtained is optimal for $\Omega = 0.2Q$.

Proposition 4.10 *If $z_{\Omega<Q} = z_{\Omega=Q}$ then a policy generated by $CTM_{\Omega<Q}$ is always optimal for higher values of Ω.*

Proof Suppose we have $\Omega_1 < \Omega_2$ and that $z_{\Omega_1} = z_{\Omega=Q}$ and we denote z_ω to be the objective function when we apply the policy generated by CTM_{Ω_1} in the context of CTM_{Ω_2}. Then the following will hold:

$z_{\Omega=Q} \leq z_{\Omega_2} \leq z_\omega \leq z_{\Omega_1}$

$z_{\Omega_2} \leq z_\omega$ holds because we are applying the same policy for a higher value of Ω hence it will penalized less for congestion. Since $z_{\Omega_1} = z_{\Omega=Q}$ then $z_{\Omega_1} = z_\omega$ and as a result the policy generated by CTM_{Ω_1} is optimal for CTM_{Ω_2}, so the theorem will hold in general. ∎

So a solution generated by $CTM_{\Omega \leq Q}$ is always optimal for higher values of Ω if their objective function values are equal, however the converse is not true. To illustrate this idea, we study the network in Example 2. The objective function value was the same for all values of Ω_i, this is due to the location of merge and diverge links, where we assume there are traffic controls. The optimization model will hold the traffic artificially at these links to avoid congestion downstream. These links will be used as a way to control congestion in order to maximize the flow at the bottleneck (at (34,14), and (62,14)). Figure 4.8 illustrates that idea. In the case of $\Omega = Q$ the bottleneck cell is more congested, so this solution may be infeasible for $\Omega = 0.2Q$.

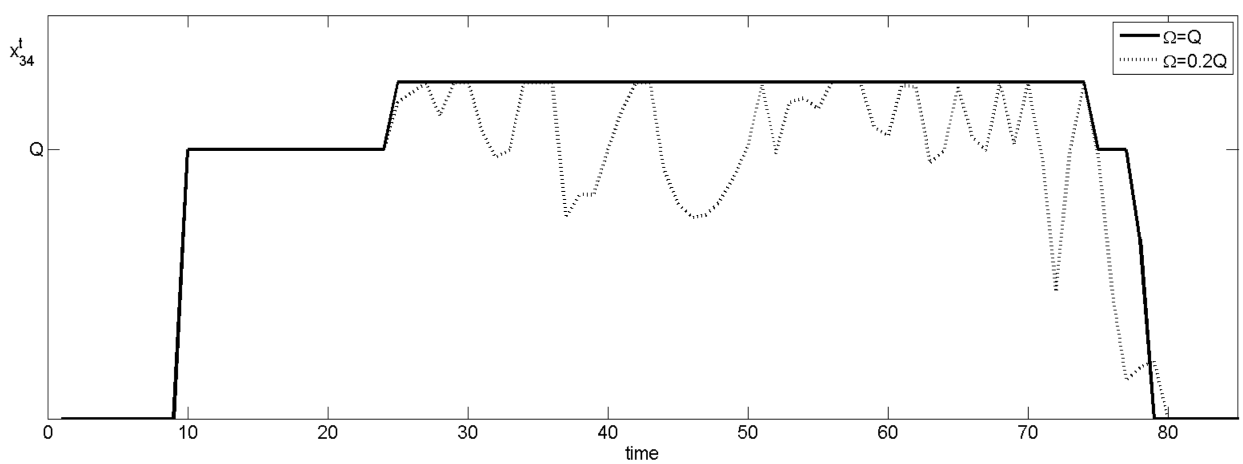


Figure 4.8: Density at cell 34 for $\Omega = 0.2Q$ and $\Omega = Q$.

Next we test the performance of the algorithm if no traffic holding is also required at merge and diverge links on the network in Example 2. We also compare the results on an optimal solution for $CTM_{\Omega=Q}$ on $CTM_{\Omega\leq Q}$, where we showed in Chapter 3 that a using the shortest paths will result in an optimal solution on this network.

TMTs	$1-4$	$1-2,4$	$1-3$	$1-2$	Shortest Path
$CTM_{\Omega=Q}$	22245	22245	22245	22245	22245
$CTM_{\Omega=0.4Q}$	22245	22361	22270	23017	69178
$CTM_{\Omega=0.2Q}$	22245	22715	22506	27774	123380

Table 4.5: Performance of the algorithm for the various Traffic control levels.

In the case of no traffic holding at merge links or diverge links only the maximal optimality gap was less than 3%, whereas when no traffic holding were allowed at any link the maximal optimality gap was less 25%. Finally, an optimal policy for $CTM_{\Omega=Q}$ was not a good solution for $\Omega \leq Q$, the objective function of the corresponding solution was more than three times the objective function of the solution generated from the algorithm.

4.5 Conclusions

The cell transmission model with flow reduction models the impact of congestion on traffic flow which is an important in deriving traffic management strategies. On the other hand, in $CTM_{\Omega=Q}$ the maximum flow can be achieved at the maximum traffic density. Therefore

$CTM_{\Omega<Q}$ offer better insights in coming up with a traffic management strategy. When used in an optimization framework, five tools can be derived from these models: Prioritizing at merge link, Routing at diverge links, Traffic holding at merge links, Traffic holding at diverge links and Traffic holding at ordinary links. For $CTM_{\Omega=Q}$, since flow is not affected by congestion, traffic holding can be omitted to obtain an optimal solution. We compare the traffic management strategies generated by these two models, and we show that for $CTM_{\Omega<Q}$ the solution has a better quality and structure than $CTM_{\Omega=Q}$. One major drawback of using $CTM_{\Omega<Q}$ is the model's tractability. To improve the model's tractability, we present a simplification to the CTM, where we reduce the number of variables at a given freeway segment. This simplification will allow us to solve for bigger networks, previously not possible with the original formulation. We also propose a heuristic that generates a feasible solution for $CTM_{\Omega<Q}$.

Although the simplification proposed significantly improves the tractability of the model, the computational effort required to obtain an optimal solution is still not polynomial with respect to the size of the network due to the presence of binary variables in the model. Future research will be build on the simplifications, and finding more efficient ways to solve the optimization problem. One direction to do it is to further reduce the number of binary variables required. Another direction is to use decomposition strategies to solve the program more efficiently.

Chapter 5

Decision Support for Hospital Evacuation

Abstract

This paper studies hospital evacuation planning which is a crucial part of a hospital's emergency management plan. In a hospital evacuation, patient must be moved from the building to a staging area, where they wait until they can be loaded onto a vehicle and transported to alternate care facility. Patients need assistance and medical care throughout this process, and the level of care and assistant is dependent on the patient's condition. Furthermore this process must be accomplished under limited resource, e.g., medical transport team and vehicle fleet size. We develop an evacuation model, in which we minimize the transportation and threat risks of the patients. In this model, the patients in the hospital would require aid from staff to be moved to the staging area, and then moved outside the building where they will be loaded to the available vehicles. So the hospital building evacuation and the patients transportation to the receiving hospitals are dependent. The resulting model is an integer program, in which the structure is complex. We exploit some of the structural properties of the feasible region and we relax some of the integrality constraints of the decision variables. We then propose a branch and price algorithm that reduce the memory requirements required to solve this model. We then demonstrate this model on a realistic case study based on the evacuation of a large regional hospital.

Keywords: Hospital evacuation planning, Emergency response, branch and price

5.1 Introduction

In this paper we introduce and study a model for hospital evacuations. Planning the evacuation of a hospital is more complicated than the evacuation of most other types of buildings because of the special needs of the patients, which includes assistance in leaving the building, medical treatment during the evacuation, and transport to an alternate care hospital. The importance of hospital evacuation is highlighted by the hospital accreditation standards requiring hospitals to develop evacuation plans. Hospitals may need to be evacuated for several reasons; These include hurricanes, fires, floods, chemical leaks, bomb threats and loss of functionality.

The main purpose of a hospital evacuation is to minimize the risk of patients and staff (Bish et al., 2011). The two main sources of evacuation risks are the threat risk (reason for evacuation) and the transportation risk. The threat risk characteristics that affect the evacuation plan are the impact of the threat on the various patient types and how threat evolves over time. For instance, a hurricane would pose no immediate threat risk to the patients (since it can be forecasted), but eventually the threat risk might be considerable; a long term power outage would affect critical patients dependent on lifesaving equipment more than patients in the hospital for observation, while a fire for instance could pose a serious risk for all patient types. In fact, in certain situations,a partial evacuation might be appropriate to reduce risk; certain patients might be safer if evacuated whereas other might not tolerate the transportation risk. The transportation risk is a function of the patient type, the vehicle, and the time required to transport the patient to the selected receiving hospital.

The evacuation of a hospital requires the movement of patients from the building to the staging area and the transportation of patients to appropriate receiving facilities by available vehicles. Bish et al. (2011) studied the transportation of patients from evacuating hospitals to alternative receiving hospitals with the assumption that the building can be evacuated such that patients of the appropriate type are available to satisfy the transportation plan withing the physical loading capacity. However, the movement of patients out of the building of the staging area is likely to impose a bottleneck on the evacuation in practice.

General building evacuations have been extensively studied in the operations research literature. However, most of these building evacuation models rely on the mobility of the evacuee population. The prevalent objective of general building evacuation problems is minimizing the total evacuation time (or building clearance time) and these problems are generally formulated as dynamic (time-expanded) network flow problems and solved by applying the corresponding network flow algorithms.

Chalmet et al. (1982) develops deterministic network models for building evacuation assuming constant capacity and travel time for each arc. Hamacher and Tufekci (1987) solves building evacuation models with multiple objectives (such as minimizing the total evacuation time and avoiding cycling of evacuees or evacuation with multiple priority levels for different parts of the building). Choi et al. (1988) incorporates flow dependent capacities into building evacuation network which adds side constraints to the problem. Chen and Miller-Hooks (2008) formulates the building evacuation problem with shared information on the changes in evacuation routes as a mixed integer linear program that minimizes the total evacuation time.

The majority of evacuation models minimize some function of the evacuation time. Han et al. (2007) discuss various evacuation objectives minimizing measures of effectiveness including individual travel (or exposure) time, time-based risk exposure, and time and space based risk exposure. A combination of these measures of effectiveness can be implemented in multi objective evacuation optimization problems. Løvås (1995) studies a building evacuation network with stochastic variables and discusses performance measures related to accident effects, evacuation time, queuing and waiting, network distances, and network redundancy.

The building evacuation studies reviewed above do not specifically address the evacuation problems of healthcare facilities. Hospital evacuations, unlike general building evacuations, involve evacuees that require extensive assistance; therefore, these problems need to be handled in a distinctive manner. Furthermore, a hospital evacuation involves more than the safe and efficient clearance of the building. An equally important aspect of the evacuation is to transport patients to appropriate alternative care facilities. The literature is quite scarce in hospital evacuation modeling pertaining to both the hospital building evacuation and the subsequent transportation of patients. The Hospital Evacuation Decision Guide (see http://archive.ahrq.gov/prep/hospevacguide), prepared for the Agency for Healthcare Research and Quality (AHRQ), discusses the importance of estimating evacuation time, among other things, to support the decision to evacuate and the timing for this decision. The evacuation time (i.e., how long it takes to evacuate the hospital) is dependent on the resources available and how efficiently they are used. This metric is difficult to estimate.

Duanmu et al. (2010) focuses on the routing of hospital vehicles during a hurricane evacuation where the ambulances and general traffic compete for space in the regional traffic flow network. The ambulance trip times are estimated using a simulation model based on various hospital evacuation start times and multiple strategies that minimize the transportation time for patients are produced. Duanmu et al. (2010) does not consider any patient-specific attributes or requirements. Golmohammadi and Shimshak (2011) estimate the evacuation time for the hospital building evacuation using a predictive model that takes patient population and available resources as input and calculates the total evacuation time. Three patient types are defined based on mobility and the patients who are the fastest to evacuate are given the first priority. This patient prioritization rule is analogous to the shortest processing time rule in scheduling theory and can significantly increase the waiting time of the most critical patients. Bish et al. (2011) studies the allocation of patients, categorized by criticality and care requirements, to a limited fleet of vehicles of various capacities and medical capabilities, to be transported to appropriate receiving hospitals considering the current available space in each hospital for each category of patient. The objective is to minimize the expected risk, both the threat risk that is forcing the evacuation, and the risk inherent in transporting patients, some in critical condition.

In this paper, we consider a hospital evacuation problem that integrates the movement of patients from the hospital building to the staging area and their subsequent transportation to receiving hospitals with adequate capacity. This model has a complex structure as a result of the assumptions on the capacity and the dependencies among variables. We propose a branch and price column generation method to improve the tractability of the model as well and we discuss some of the assumptions that can be relaxed or modified for the same purpose.

5.2 Model

We consider a single evacuating hospital in order to analyze the interaction between the building evacuation and the transportation phases in detail. The proposed model can be easily expanded to include multiple evacuating hospitals. The complete evacuation process studied is depicted in Figure 5.1. The first phase of a hospital evacuation is the process of moving patients from their location in the hospital to the staging area. The second phase is the process of loading the patients into ambulances and sending them to the receiving hospitals. The following assumptions has been made: vehicles do not stop at multiple receiving hospitals, loading time of a hospital is independent of patient type in it, loading time and unloading time of a vehicle are equal, travel times to receiving hospitals have known fixed lengths, each patient type should be assigned to the corresponding bed type and the time to move the patient from his ward to the staging area is only dependent on the patient type.

Let P the set of patient types (e.g., Intensive care unit, neonatal intensive care unit, ambulatory, etc.) where there are W_p patients of type $p \in P$ in the evacuating hospital. These patients must be transported to a set (J) of potential receiving hospitals, where hospital $j \in J$ has B_b^j beds available where B is the of beds and $b \in B$. Let H be the set of all hospitals. The study period is divided into T time intervals of equal length. The travel times between evacuating hospitals are known, and are independent of vehicle type.

Patients are moved to the staging area, to await transport to a receiving hospital, by medical transport teams. We assume that more advanced teams must be preferred for assisting critical patients and any team can assist the non-critical patients. For example, neonatal intensive care unit (NICU) patients need constant respiratory support and an experienced nurse must be holding and assisting the baby patient along with other staff members who move necessary medical equipment for the patient. Therefore, an upgrade team assignment

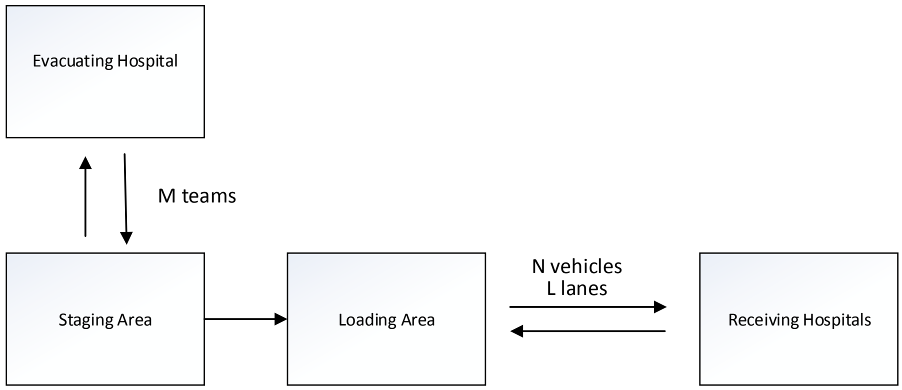


Figure 5.1: Building evacuation and transportation process.

strategy is implemented such that more critical patients can only be assisted by advanced teams, but less critical patients can be assisted by any team as along as there are enough teams available. Due to different levels of care required by patients of different types, the evacuation teams are categorized similar to the vehicles to represent their capabilities over set K. The set of teams that can assist patients of type p is denoted by K_p. All teams are assumed to be at the staging area at the beginning of the planning horizon and a team is assumed to move one patient at a time. The total number of idle and busy teams of type $k \in K$ at the evacuating hospital at time t is represented by M_k^t. The time it takes a team to move a patient of type P to the staging area is τ_p time intervals. Patients reaching the staging area can occupy space for multiple time intervals if they have to wait while other patients are loaded before them or if all vehicles are en route.

A set of vehicle types (V), e.g. Advanced life support (ALS) ambulances, Basic life support (BLS) ambulances is given. At time interval t, the total number of busy and idle vehicles of type $v \in V$ is N_{vt}. The time required to load a vehicle to type $v \in V$ is γ_v time intervals and it is assumed to be equal to the unloading time for that vehicle. Vehicles transport patients directly from the evacuating hospital 1 to hospital $j \in J$ without stopping at multiple hospitals hospitals and return to the hospital as needed.

The number of vehicles that can be loaded at the evacuating hospital in any time interval is based on the limitations of the evacuating hospital's loading area. This is represented by a set of parameter L_i that represent the total number of lanes available for a given set of vehicles. The loading area is considered as a separate area and patients occupy the loading area as long as the vehicle is loading. If the number of vehicles available to be loaded at a time interval exceeds L, the excess number of vehicles have to wait at the evacuating hospital before being loaded.

The objective is a building evacuation and transportation plan that minimizes the total evacuation risk where risk is defined as the probability that an undesired event occurs. The evacuation risk is defined as a combination of: 1) the threat risk that patients are exposed to while waiting to be transported, and 2) the transportation risk incurred during travel. The cumulative threat risk, Λ_{pt}, calculated in (5.1), is the probability of the undesired event for a patient of type p that remains in the evacuating hospital though time interval t. λ_{pt} is the probability of the undesired event for a patient of type p that remains in the hospital at time t. λ_{pt} are assumed to be independent.

$$\Lambda_{pt} = 1 - \prod_{f=1}^{t}(1 - \lambda_{pf}), \forall p \in P, t = 1, .., T. \tag{5.1}$$

The cumulative transportation risk, Θ^j_{pv}, is calculated in (5.2). θ_{pv} is the probability of the undesired event for a patient of type $p \in P$ transferred by a vehicle of type $v \in V$ for one time interval. When transporting patients to hospital j, a vehicle of type v that returns to the evacuating hospital is engaged for $2(\tau^j + \gamma_v)$ time units, however only $(\tau^j + 2\gamma_v)$ time units contribute to the risk expression since transportation risk is only incurred only as long as the patient is in the vehicle, including loading and unloading times.

$$\Theta_{pv}^{j} = 1 - (1 - \theta_{pv})^{(\tau^{j} + 2\gamma_{v})}, \forall j \in J, p \in P, v \in V, t = 1, .., T. \tag{5.2}$$

The evacuation risk, R_{pvt}^{j}, associated with the evacuation decision for a patient is calculated in (5.3) by combining the cumulative threat risk (Λ_{pt}) based on the time interval t when the patient is transported and the cumulative transportation risk (Θ_{pv}^{j}) based on the vehicle type and the receiving hospital selected.

$$R_{pvt}^{j} = 1 - (1 - \Lambda_{p(t-1)})(1 - \Theta_{pv}^{j}), \forall j \in J, p \in P, v \in V, t = 1, .., T. \tag{5.3}$$

Parameters:

T : number of time intervals in the study period

W_p : number of patients of type $p \in P$ initially at the evacuating hospital

M_{kt} : total number of teams of type $k \in K$ at the evacuating hospital at time t

τ_p : numbers of time intervals required to move a patient of type p from his ward to the staging area

B_b^j : number of beds of type $b \in B$ available at hospital $j \in J$

τ^j : number of time intervals required to travel from the evacuating hospital to hospital $j \in J$

N_{vt} : total number of busy and idle vehicles of type $v \in V$ at time t

C_{iv} : number of patients that can be loaded onto a vehicle to type $v \in V$, for a given set of patients $P_i \in P$

γ_v : number of time intervals to load/unload a vehicle of type $v \in V$

L_i : number of lanes available for vehicles in $V_i \subset V$

Λ_{pt} : cumulative threat risk for a patient of type $p \in P$ that remains in the evacuating hospital through time interval t

Θ^j_{pv} : cumulative transportation risk for a patient of type $p \in P$ transferred by a vehicle of type $v \in V$ to hospital $j \in J$

R^j_{pvt} : total evacuation risk for a patient of type $p \in P$ transferred by a vehicle of type $v \in V$ to hospital $j \in J$ starting at time t

Decision Variables:

s_{pkt} : number of patients of type p moved from their ward in the evacuating hospital to the hospital's staging area by a team of type k starting in time interval t, $\forall p \in P, k \in K, t = 1, \cdots, T$

x^j_{pvt} : number of patients of type p transported from the staging area of the evacuating hospital to the staging area of receiving hospital j by a vehicle of type v starting in time interval t, $\forall v \in V, t = 1, \cdots, T$

y^j_{vt} : number of vehicles of type v that move to hospital j starting at time interval t, $\forall i, j \in H, v \in V, t = 1, \cdots, T$.

We next present the hospital evacuation model.

Objective Function:

$$\text{Minimize } \sum_{j\in J}\sum_{p\in P}\sum_{v\in V}\sum_{t=1}^{T} R^j_{pvt} x^j_{pvt} \tag{5.4}$$

Objective function (5.4) minimizes the total evacuation risk where the summation represents the total risk for patients transported to alternative facilities.

Transportation Teams Constraints:

$$\sum_{t=1}^{T} s_{pkt} = 0, \forall p \in P, k \notin K_p \tag{5.5}$$

$$\sum_{p\in P}\sum_{f=0}^{\min(\tau_p-1,t)} s_{pk(t-f)} \leq M_{kt} \tag{5.6}$$

$$\sum_{k\in K}\sum_{t=1}^{T} s_{pkt} \leq W_p, \forall p \in P \tag{5.7}$$

$$\sum_{k\in K}\sum_{f=1:t>\tau_p}^{t-\tau_p} s_{pkf} - \sum_{j\in J}\sum_{v\in V}\sum_{f=1}^{t} x^j_{pvf} \geq 0, \forall p \in P, t = 1, .., T \tag{5.8}$$

Constraint (5.5) sets the restriction of upgrade team assignment strategy. Constraint (5.6) limits the number of busy teams to the total number of teams available. The total number of patients of each type moved out from the evacuating hospital is limited by the initial population size in constraint (5.7). Constraint (5.8) limits the number of patients transported to the number of patients available at the staging area.

Evacuating Constraints:

$$\sum_{j\in J}\sum_{v\in V}\sum_{t=1}^{T} x^j_{pvt} = W_p, \forall p \in P \tag{5.9}$$

$$\sum_{v\in V}\sum_{p\in P_i}\sum_{t=1}^{T} x^j_{pvt} \leq B^j_b, p \in P_b \subset P, \forall j \in J, b \in B \tag{5.10}$$

$$\sum_{p\in P_v} x^j_{pvt} \leq C_{P_v} y^j_{vt}, \forall j \in J, v \in V, P_v \subset P, t = 1, .., T \tag{5.11}$$

$$\sum_{j\in H} y_{vt}^j + \sum_{j\in J:t>\tau^j} y_{v(t-\tau^j)}^j \leq N_{vt}, \forall v \in V, t = 1, ..., T \tag{5.12}$$

$$\sum_{j\in J} \sum_{v\in V_i} \sum_{f=t-\gamma_v+1}^{t} y_{vf}^j \leq L_i, V_i \subset V, \forall t = 1..T \tag{5.13}$$

Every patient should be evacuated by Constraint 5.9. Constraint (5.10) defines the number of beds of each type available at each alternative care location. Constraint (5.11) represents the vehicle capacity restriction on the number of patients transferred at each time interval. Constraint (5.12) bounds the number of busy vehicles by the total number of vehicle available. Constraint (5.13) restricts the number of vehicles that can be loaded at each time interval due to a physical loading capacity.

Integrality Constraints:

$$s_{pkt} \geq 0 \text{ and integer}, \forall p \in P, k \in K, t = 1, ..., T \tag{5.14}$$

$$x_{pvt}^j \geq 0 \text{ and integer}, \forall p \in P, v \in V, j \in J, t = 1, ..., T \tag{5.15}$$

$$y_{vt}^j \geq 0 \text{ and integer}, \forall v \in V, t = 1, ..., T \tag{5.16}$$

Constraints (5.14)-(5.16) are the integrality and non-negativity constraints.

We next present a structural property of the above model that improve tractability.

Lemma 5.1 *If a matrix $\boldsymbol{A}$ is full rank, the total unimodularity of $\boldsymbol{A}$ is preserved under the following three elementary row (column) operations: (1) exchanging two rows (columns),(2) multiplying a row (column) by -1, and (3) adding a row (column) to another row (column).*

The proof of Lemma 5.1 is provided in Schrijver (1998).

Theorem 5.1 *For a feasible set of y-variables and s-variables,and a continuous relaxation of x-variables there exists an optimal solution in which the continuous variables have integral values.*

Proof Since the y-variables and s-variables are given, they can be considered as parameters. Therefore, constraints (5.5)-(5.7) and constraints (5.12)-(5.13) are redundant. Then a continuous relaxation of the x-variables transforms the constraint set as follows:

$$\sum_{j \in J} \sum_{v \in V} \sum_{t=1}^{T} x_{pvt}^{j} = W_p, \forall p \in P \tag{5.17}$$

$$\sum_{k \in K} \sum_{f=1:t>\tau_p}^{t-\tau_p} s_{pkf} - \sum_{j \in J} \sum_{v \in V} \sum_{f=1}^{t} x_{pvf}^{j} \geq 0, \forall p \in P, t = 1, .., T \tag{5.18}$$

$$\sum_{v \in V} \sum_{p \in P_i} \sum_{t=1}^{T} x_{pvt}^{j} \leq B_b^j, p \in P_b \subset P, \forall j \in J, b \in B \tag{5.19}$$

$$\sum_{p \in P_v} x_{pvt}^{j} \leq C_{P_v} y_{vt}^{j}, \forall j \in J, v \in V, P_v \subset P, t = 1, .., T \tag{5.20}$$

$$x_{pvt}^{j} \geq 0, \forall p \in P, v \in V, j \in J, t = 1, ..., T \tag{5.21}$$

We can sum constraints (5.20) to eliminate constraint (5.19) by elementary row operations. Similarly, we can eliminate constraint (5.17) by summing over constraint (5.18). The reduced echelon form of **A** consists of constraints (5.18) and (5.20). Therefore A in an $m \times n$ matrix where $m = |P|T + |J||P_v|T$, $n = |P||V||J|T$ and $\min\{m, n\} = m$. Since the constraints are linearly independent and $\min\{m, n\} = m$, **A** is full rank and this reduction would preserve total unimodularity by Lemma 5.1. A matrix is totally unimodular if the determinant of every square submatrix formed from it is -1, 0 or +1.

Camion (1965) proved that a matrix is totally unimodular if and only if every square Eulerian submatrix formed from it is singular, where a submatrix is Eulerian if both the sum of each row and the sum of each column is even, and equivalently, that a matrix is totally unimodular if and only if the sum of all elements in every Eulerian square submatrix formed from it is a multiple of 4. It can be shown that **A** in our case is totally unimodular by examining the characteristics of its submatrices. The entries of the coefficient matrix **A** are +1 or 0 and any one of the constraints can be negated to obtain -1 entries. Each x_{pvt}^{j} appears in $(T - t + 1)$ consecutive constraints in Constraint set (5.18) and once in constraint set (5.20).

The resulting matrix is of the form:

$$\begin{pmatrix} Z & \cdots & \cdots & \cdots \\ 0 & Z & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ Y & Y & \cdots & Y \end{pmatrix}$$

Where Y is the matrix from constraint (5.20) , and Z is of the form:

$$Z = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

The sum of each column of any submatrix extracted from Z is alternating between even and odd so every Eulerian submatrix has some elements from the Y matrices row since the sum of each column and row of this submatrix has to be even. If an Eulerian submatrix has row/column of zeros, it would be singular. The rest of the Eulerian submatrices has to have the following form:

$$\begin{pmatrix} R & 0 & \cdots & 0 \\ 0 & R & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R \\ X & X & \cdots & X \end{pmatrix}$$

In this matrix the sum of each column is equal to the corresponding row, and since this value is even, the resulting sum of all elements is a multiple of 4.

As a result $\mathbf{A}$ is totally unimodular. Since the right-hand side of the constraint set is integer valued and by Cramer's rule all extreme point solutions to the LP are integer valued. ∎

5.3 A branch and price method

In this section we propose a Branch and Price method for solving the above integer linear programming model. The method is a hybrid of branch and bound and column generation method. This approach is based on the observation that for big problems, most columns are non basic and hence are equal to zero in any given optimal solution. So in this method, we start with a reduced set of columns and the columns are added to the model as needed. This will reduce the memory and computational requirements.

We start by making the following observation about the formulation; The transportation team constraints and the evacuating constraints would be completely separable if we removed Constraint (5.8).

Let c be the matrix definition of the objective function, A to be the matrix definition of Constraint (5.8), X_1 the feasible set corresponding to constraints (5.5)-(5.7) and (5.14), X_2 the feasible set corresponding to Constraints (5.10)-(5.13) and (5.15)-(5.16).

The first step is to generate initial columns in order to be able to start the branch and price algorithm. The first column is $(\mathbf{x}, \mathbf{y}, \mathbf{s}) = \mathbf{0}$. We generate the second column by the following procedure; We first solve the following problem:

$$\text{Minimize} \quad \sum_{j\in J}\sum_{p\in P}\sum_{v\in V}\sum_{t=1}^{t_1} R^j_{pvt}x^j_{pvt} + \sum_{p\in P}\Lambda_{pt_1}(W_p - \sum_{j\in J}\sum_{v\in V}\sum_{t=1}^{t_1} x^j_{pvt}) \tag{5.22}$$

$$\text{subject to } \mathbf{s} \in X_1, (\mathbf{x}, \mathbf{y}) \in X_2, (\mathbf{x}, \mathbf{y}, \mathbf{s}) \in X_3 \tag{5.23}$$

Here we choose $t_1 < T$, if t_1 is small enough, the problem will be easy to solve. However it is not possible to evacuate all patients by time t_1, hence we add the second term in the objective function. We define $(\mathbf{x_1}, \mathbf{y_1}, \mathbf{s_1})$ to be the solution from 1 to t_1 and zero otherwise. We then solve the following

$$\begin{aligned}
\text{Minimize} \quad & \sum_{j \in J} \sum_{p \in P} \sum_{v \in V} \sum_{t=1}^{t_2} R^j_{pvt} x^j_{pvt} + \sum_{p \in P} \Lambda_{pt_2} (W_p - \sum_{j \in J} \sum_{v \in V} \sum_{t=1}^{t_2} x^j_{pvt}) && (5.24)\\
\text{subject to} \quad & \mathbf{s} \in X_1, (\mathbf{x}, \mathbf{y}) \in X_2, (\mathbf{x}, \mathbf{y}, \mathbf{s}) \in X_3 && (5.25)\\
& (\mathbf{x}, \mathbf{y}, \mathbf{s}) \geq (\mathbf{x_1}, \mathbf{y_1}, \mathbf{s_1}) && (5.26)
\end{aligned}$$

Here $t_1 < t_2 \leq T$. We keep doing this process until $t_n = T$. The resulting solution is feasible and we put it in the initial column set.

We further decompose A and c, into $\mathbf{A_x}, \mathbf{A_y}, \mathbf{A_s}$ and $\mathbf{c_x}, \mathbf{c_y}, \mathbf{c_s}$ where each matrix corresponds to its given decision variable set.

We define the reduced master problem (RMP) to be as follows:

$$\begin{aligned}
\text{Minimize} \quad & \sum_{i=1}^{n_x} \mathbf{c_x} \mathbf{x_i} \lambda_{xi} + \sum_{i=1}^{n_y} \mathbf{c_y} \mathbf{y_i} \lambda_{yi} + \sum_{i=1}^{n_s} \mathbf{c_s} \mathbf{s_i} \lambda_{si} && (5.27)\\
\text{subject to} \quad & \sum_{i=1}^{n_x} \mathbf{A_x} \mathbf{x_i} \lambda_{xi} + \sum_{i=1}^{n_y} \mathbf{A_y} \mathbf{y_i} \lambda_{yi} + \sum_{i=1}^{n_s} \mathbf{A_s} \mathbf{s_i} \lambda_{si} \geq \mathbf{0} && (5.28)\\
& \sum_{i=1}^{n_x} \lambda_{xi} = 1, \sum_{i=1}^{n_y} \lambda_{yi} = 1, \sum_{i=1}^{n_s} \lambda_{si} = 1 && (5.29)\\
& \lambda_{xi}, \lambda_{yj}, \lambda_{sk} \geq 0, i = 1, ..., n_x, j = 1, ..., n_y, k = 1, ..., n_s && (5.30)
\end{aligned}$$

Where $\mathbf{x_i}, \mathbf{y_j}, \mathbf{s_k}$ are the set of columns already generated and n_x, n_y, n_s to be their corresponding number. The RMP is a linear program since the integrality constraints are relaxed. Let $(\mathbf{w}, \alpha_\mathbf{s}, \alpha_\mathbf{x}, \alpha_\mathbf{y})$ be the corresponding dual solution of the RMP for Constraints (5.29) and (5.30).

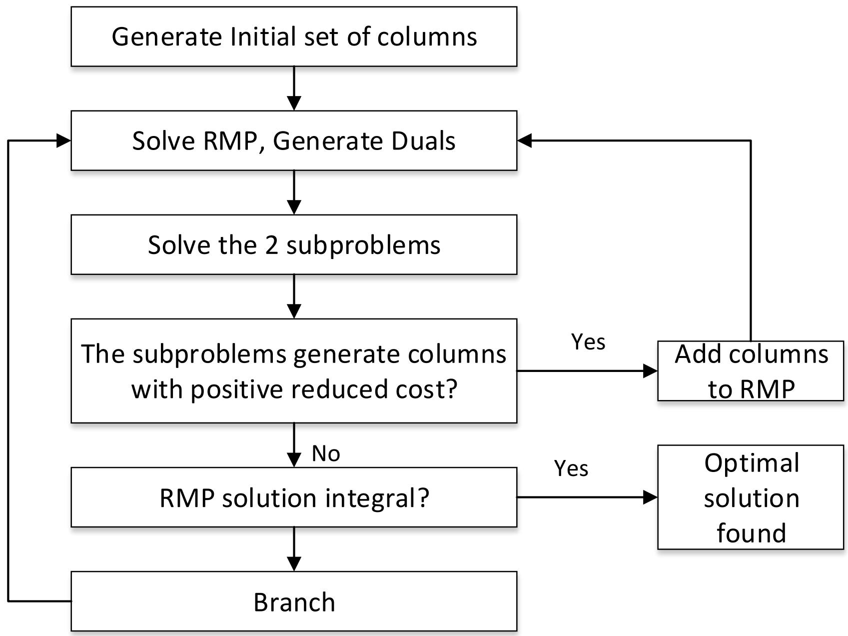


Figure 5.2: Branch and Price algorithm.

Since X_1 and X_2 are separable, we define two subproblems as follows:

$$\text{Maximize} \quad (\mathbf{w}A_s - c_s)\mathbf{s} + \alpha_s \tag{5.31}$$

$$\text{subject to } \mathbf{s} \in X_1 \tag{5.32}$$

$$\text{Maximize} \quad (\mathbf{w}A_x - c_x)\mathbf{x} + \alpha_x + (\mathbf{w}A_y - c_y)\mathbf{y} + \alpha_y \tag{5.33}$$

$$\text{subject to } (\mathbf{x}, \mathbf{y}) \in X_2 \tag{5.34}$$

Figure 5.2 illustrates the branch and price algorithm. The algorithm will converge to the optimal solution (Desaulniers et al., 2005).

5.4 Case study

Consider the problem of evacuating a single hospital to a set of potential receiving hospitals. The complete evacuation process studied is depicted in Figure 5.1. The first phase of a hospital evacuation is the process of moving patients from their location in the hospital to the staging area. The second phase is the process of Loading the patients into ambulances and sending them to the receiving hospitals. The following assumptions have been made:

There are nine possible patient types; Neonatal Intensive Care Unit (NICU), Neonatal Intensive Care Unit with ventilator (vNICU), Pediatric Intensive Care Unit (PICU), Pediatric Intensive Care Unit with ventilator (vPICU), Intensive Care Unit (ICU), Intensive Care Unit with ventilator (vICU), Bed-Bound (BB), Ambulatory Oxygen-Dependent (AOD) and other Ambulatory (OA). Patients must be moved to a staging area before transport to a receiving hospital by patient movement teams. There are two types of patient movement teams. The first team type, denoted as T_v, is especially trained to move patients on ventilators; this type of team is required to move vNICU, vPICU, and vICU patients, but can move any type of patient. The other team type, denoted as T_o, can move any patient type not requiring a ventilator (i.e., NICU, PICU, ICU, BB, AOD, and AO). The time required for a team to move a patient on a ventilator to the staging area is denoted as τ_v, while the time required to move other patients is τ_o. Once patients are moved to the staging area, they can be loaded onto vehicles for transport to receiving hospitals.

There are four vehicle types used to transport patients: Critical Care Transport (CCT), Advanced Life Support (ALS) and Basic Life support (BLS) ambulances and buses. The assignments of patients to vehicle types must follow the following rules:

- CCT transport one patient at a time, and only patients of types vNICU, vPICU, and vICU.
- ALS transport NICU, PICU, and ICU patients. An ALS can transport one NICU,

or one PICU with an optional additional AOD or OA patient, or one ICU with an optional additional AOD or OA patient, at a time.

- BLS transport one BB patient with an optional additional AOD or OA patient, or one AOD patients with an optional additional AOD or OA patient. BLS do not carry only OA patients.
- Buses transport up to 30 OA patients at a time; buses only transport OA patients.

Patients are loaded onto vehicles from the staging area using loading lanes; there are two types of lanes, ambulance lanes for the CCT, ALS, and BLS, and bus lanes. We assume that each vehicle type has a loading time, which is equal to the unloading time, and is independent of the number of patients or type of patients to be loaded/unloaded from the vehicle.

The travel time from each receiving hospital to the evacuating hospital is known, and each receiving hospital has known bed availability. Beds are categorized in a manner similar to patients, but simplified, there are NICU beds that can be used by vNICU and NICU patients, PICU beds that can be used by vPICU and PICU patients, and ICU beds, that can be used by vICU and ICU patients, and regular beds that can be used by all other patient types (i.e., BB, AOD, and AO patients).

5.4.1 Methodology and analysis

To solve this hospital evacuation problem, we will use an optimization methodology described earlier. First we divide the planning horizon into equal time intervals. Next we have to calculate the parameters that depend on time to match the model. We do that by diving the value of parameter by the length of the time interval and taking the ceiling of the new value. As a result if the ratio is not integer, we would be overestimating the parameter. Hence the generated plan is always overestimating the evacuation time of the hospital. So in choosing a smaller time interval length, we would be producing a better solution, however due to the complex nature of the problem it will become less tractable. Similarly, choosing a larger time interval will make the model easier to solve, however the quality of the solution will deteriorate. To illustrate this idea, we will solve this model for two time interval lengths, 5 minutes and 10 minutes. Since the second time is a multiple of the first, we would expect to get a better evacuation time with $t = 5$ minutes, however the time to solve will be bigger.

We use a risk based methodology, where patients of each type are given a threat risk (the reason for the evacuation) and a transportation risk. The objective is to minimize the risk, which is highly correlated to minimizing the duration. There are many advantages of using a risk based methodology (Bish et al., 2011). See Table 5.7 in the Appendix for the risk-values used in this study.

Upon solving this model we aim at generating an optimal evacuation plan that minimizes total risk, and to identify the limiting resources(resources are transport teams,vehicles and lanes).

The lane utilization can be calculated using the following formula:

$$\frac{\frac{1}{T}\sum_{t=1}^{T}\sum_{j\in J}\sum_{v\in V_i}\sum_{f=t-\gamma_v+1}^{t} y_{vf}^{j}}{L_i} \times 100 \tag{5.35}$$

where L_i is the number of lanes available, V_i is the subset of vehicles using Lanes of type i, γ_v is the loading time of vehicle v and y^j_{vf} corresponds to the movement of vehicles.

The transport teams utilization can be calculated in a similar fashion.

The minimum number of idle vehicles of each type can be calculated by:

$$N_{vt} - \max_t \sum_{f=\max(t-\tau^j,1)}^{t} y^j_{vf} \tag{5.36}$$

where N_{vt} is the total number of vehicles of type v at time t

Thus if we decrease the number of available vehicles by that amount, the optimal solution will still be feasible, and we have more assigned vehicles than needed for the evacuation process.

5.4.2 Problem instance

The problem instance that we examine has 386 patients in the evacuating hospital, and they are distributed among the patients types as shown in the following table:

There are 78 potential receiving hospitals in this instance; their travel times from the evacuating hospital and bed availability is provided in Figure 5.3 in the Appendix. To move the patients to the staging area, there are two teams having ventilator training (i.e., type T_v), and three other teams (i.e., type T_o). The time required to move patients on ventilators to the staging area is $\tau_v = 22$ minutes, while all other patients require $\tau_o = 10$ minutes.

Patient Type	Number
vNICU	4
NICU	14
vPICU	5
PICU	11
vICU	16
ICU	62
OBB	66
AOD	139
OA	69

Table 5.3: Number of patients of each type at the the evacuating hospital.

After 30 minutes, four CCT, seven ALS, and 27 BLS, and three buses are available. And after 60 minutes two CCT, three ALS, and 27 BLS additional vehicles are available. The vehicle loading times are 20 minutes for CCT and ALS ambulances, 15 minutes for BLS ambulances and 5 minutes for buses. We will solve this problem with time lengths of 5 minutes and 10 minutes. For example for the case of the 10 minutes time intervals, if the time to move a ventilator patient to the staging area is 22 minutes, then transport time will be three time intervals (i.e., 30 minutes). This makes the problem easy to solve, but adding eight minutes to the task can obviously inflate the required time to evacuate the hospital Finally, this instance has one ambulance lane and one bus lane for vehicle loading. Finally, we simplify the vehicle-patient rules given in the problem statement, by allowing BLS to carry only OA patients. This was done to improve solvability, but we note that this actually does not occur in our given solutions (OA patients are transported mainly by bus).

5.4.3 Solution

We solve this problem instance using 5 minutes and 10 minutes interval lengths. For the case of 10 minutes intervals, the solution for the one ambulance lane was found in 15 minutes. Solving the problem with only one ambulance lane is the most difficult case to solve because the number of time intervals required is large, and we note the optimal solution was found fairly quickly, and a majority of the 15 minutes was used to confirm optimality. On the other hand, for the case of 5 minutes intervals, it took more than 3 hours to converge to an optimal solution. This is because, the number of decisions variables has doubled since the number of decision variables is a multiple of T. The optimization model was run on ILOG CPLEX 12.5.1 with 2.53 Ghz quad Core and 16 GB of RAM. To evaluate the algorithm, we ran both the original formulation and the branch and price algorithm on CPLEX on MATLAB. For the 10 minute time interval length, the branch and price algorithm converged in 12 minutes whereas using the original formulation we could not find any solution.

Table 5.4 and 5.5 shows the results from our model for cases with one ambulance lane (original instance), two ambulance lanes, three ambulance lanes, and infinite number of lanes. As was expected, the evacuation for the 5 minutes was smaller for all the problem instances.

	Evacuation Time (hours)	Utilization %		
		Ambulance Lane	T_v-Teams	T_o-Teams
1 Lane	59.833	100	25	24
2 Lanes	29.833	100	49	79
3 Lanes	20.5	100	70	75
Infinite lanes	15.333	50	95	95

Table 5.4: Evacuation time and other performance measures with 10 minutes time intervals.

	Evacuation Time (hours)	Utilization % Ambulance Lane	T_v-Teams	T_o-Teams
1 Lane	54.5	100	28	24
2 Lanes	27.1	100	44	50
3 Lanes	18	100	46	51
Infinite Lanes	15	N/A	95	97

Table 5.5: Evacuation time and other performance measures with 5 minutes time intervals.

There are many interesting analysis that can be tested using this model, for instance Table 5.6 shows the minimum number of idle vehicles (waiting to be loaded at the evacuating hospital) for the various cases. From this, we can see that there are at least 4 ALS and 49 BLS idle in every time interval after the first 60 minutes, i.e., after all vehicles are in the system. This means that we can reduce the number of ALS by 4 and BLS by 49 and not impact the solution.

Also we ran this problem with unconstrained lanes and unconstrained staging. The evacuation time was 10 hours. The main bottleneck in this case was the number of ambulances available.

	CCT	ALS	BLS	Bus
1 Lane	0	4	49	0
2 Lanes	0	0	44	0
3 Lanes	0	0	42	0
12 lanes	0	0	18	0

Table 5.6: Minimum number of idle vehicles by type (after 60 minutes) for the 10 minutes time intervals solution.

	Beds				
Hospital	ICU	NICU	PICU	Regular	time
H01	0	2	0	0	8.63
H02	3	3	0	3	13.9
H03	1	1	3	1	20.6
H04	1	3	0	1	21.1
H05	0	0	0	0	21.7
H06	0	5	0	0	24.8
H07	0	0	3	0	0.54
H08	0	1	0	0	0.58
H09	0	0	0	0	2.26
H10	0	0	0	0	2.41
H11	1	0	0	1	3.71
H12	0	0	0	0	3.69
H13	0	4	0	0	4.06
H14	5	0	0	5	4.63
H15	0	0	0	2	5.58
H16	2	6	1	2	6.96
H17	2	0	0	0	6.87
H18	0	0	0	1	6.87
H19	1	2	0	2	6.19
H20	2	10	0	0	6.63
H21	0	13	3	0	7.37
H22	0	7	0	8	6.85
H23	8	0	0	0	10.2
H24	0	0	0	1	10.1
H25	1	3	0	0	11.4
H26	0	0	0	1	11.9
H27	0	0	0	1	12.9
H28	0	2	0	1	13.7
H29	1	0	0	5	12.1
H30	1	0	0	0	12.2
H31	1	1	1	1	14.3
H32	5	1	1	0	11.4
H33	0	0	0	2	14.2
H34	1	18	5	8	14.9
H35	0	0	0	1	14.2
H36	2	0	0	1	15.7
H37	8	4	0	2	14.5
H38	1	0	0	0	15.1
H39	1	0	0	2	13.8
H40	2	1	0	0	17.3
H41	0	0	0	1	16.6
H42	2	6	0	1	18.7
H43	0	2	0	0	18.9
H44	1	4	0	0	16
H45	1	0	0	2	18.4
H46	0	0	0	1	14.9
H47	0	1	1	1	20.2
H48	2	2	0	1	18.4
H49	1	0	0	1	20.9
H50	1	1	0	2	20.6
H51	1	8	1	8	21.5
H52	1	2	0	0	22.6
H53	2	19	0	1	24.5
H54	8	0	0	1	23.8
H55	0	1	0	0	24.7
H56	0	0	0	2	27.5
H57	0	0	0	1	25.2
H58	1	2	0	2	22.4
H59	1	55	9	8	25.1

Figure 5.3: Number of beds available at each receiving hospital and their travel time from the evacuating hospital in minutes.

Patient Type	Threat Risk	Transportation Risk
vNICU	$0.006e^{t/T}$	10^{-4}
NICU	$0.005e^{t/T}$	7.5×10^{-5}
PICU	$0.006e^{t/T}$	10^{-4}
vPICU	$0.005e^{t/T}$	7.5×10^{-5}
ICU	$0.006e^{t/T}$	10^{-4}
vICU	$0.005e^{t/T}$	7.5×10^{-5}
OBB	$0.004e^{t/T}$	5×10^{-5}
AODP	$0.004e^{t/T}$	5×10^{-5}
OA	$0.002e^{t/T}$	2.5×10^{-5}

Table 5.7: Number of patients and their corresponding risks at the evacuating hospital.

5.5 Conclusion

In this paper we propose a hospital evacuation model that include the building evacuation phase. The objective of this model is to minimize the risk (threat and transportation risks) of the patients. The limiting resources incorporated in this model are: staff assisting patients out of the building, loading area capacity, vehicle fleet size and number of beds in receiving hospitals. We increased the tractability of the model by relaxing some of the integrality constraints due to its totally unimodular matrix structure. We also proposed a branch and price algorithm to decrease the memory requirements for this model, and this is specially helpful for big problem instances. We also apply this model on a realistic case study and propose some performance measures for the solution.

This model can be easily extended to the case of multi evacuating hospitals with shared resources. One possible extension of this study is to relax some of the assumptions in the model while keeping the problem tractable.

Chapter 6

Conclusion and Directions for Future Research

The goal of this dissertation is to study the different models in evacuation planning. In particular, we focus on mass evacuation and hospital evacuation planning. This work is a branch of the humanitarian logistics which became of great significance in the last decade as the frequency and magnitude of natural disasters have increased significantly. Examples of large scale natural disasters include Hurricane Sandy in 2012, Hurricane Katrina in 2005, Haiti earthquake in 2012 and Tohuku earthquake in 2012.

In the first part, we studied the case of mass evacuation planning, and this is mainly accomplished by the use of house hold vehicles. In chapter 3, we consider the case where the evacuee population size is not known. We use the cell transmission model (CTM) as the traffic flow model. This model is a macroscopic flow model that model the traffic behavior by decomposing the network into small segments and evaluating the flow at these segments. This model can be embedded into an optimization problem, where the flow across the segments can be used as a decision making tool to generate an evacuation plan. Since the population size is assumed to be unknown, the optimization problem now has a parameter which is a random variable. To deal with the demand uncertainty, an approach studied in the literature is to use deterministic parameters to represent the random variables. This approach has several drawbacks: Since the demand realization does not match the deterministic counterpart, a solution requires an implementation method which not considered in the optimization. We discuss two implementation strategies and we show by the use of numerical examples that the resulting solution is difficult to implement since the routing and traffic controls can be complex. We propose a different approach for this problem. In certain networks, an optimal routing policy that is independent on the demand can be found. We study a class of networks which are present in the literature in which we characterize the optimal solution. For more complex networks, in which an optimal policy has to be dependent on the demand size, we propose a heuristic and we show that it outperforms the deterministic counterpart approaches. To show the benefits of using the CTM-FR, we compare the solution

One of the drawbacks of the CTM, is that congestion does not reduce the flow. In chapter 4, we discuss the cell transmission model with flow reduction (CTM-FR) in which flows will be reduced in the presence of congestion. The CTM-FR has more depths than the CTM because now traffic volume will have an impact on the solution. However the original formulation was not efficient and even for small networks, it would not converge in real time. We started by simplifying the formulation and making it more tractable. Next we discuss the traffic management tools that can be employed: Routing at diverge links, Prioritizing at merge links, and holding traffic at ordinary links. We also compared the solutions generated by CTM and CTM-FR. For example, the objective function to minimize the total system time (TST) and nwtwork clearance time (NCT) are equivalent in CTM, but this is not the case in CTM-FR. We alse showed that CTM-FR does have a better structure and will provide better insights that the CTM.

In the second part of the thesis, we address the problem of hospital evacuation. This is different than mass evacuation, since the patients are not homogeneous anymore, unable to move unaided, and each patient type require different evacuation procedure. In chapter 5, we developed a model that integrates the hospital building and the transportation between the evacuating hospital and receiving ones. This model is based on our discussion with hospital and regional emergency managers. The resulting model is complex since the movement of patients to the staging area and the transportation of patients to the receiving hospitals are dependent. We were able to remove some of the integral constraints due to the structural properties of the model. We also propose a branch and price decomposition algorithm for the model to decrease the memory requirements for the optimization problem. Finally we apply this model on a case study on the evacuation of the Caiser hospital in Los Angeles.

6.1 Suggestions For Future Research

The Research in dissertation can be extended into various directions. Here are some suggestions to expand the current research

- Study demand uncertainty in models where congestion does have an impact on the flow in the network. In the cell transmission model with flow reduction setting, an optimal policy independent on demand will probably not exist for many network structures. The complexity of this problem makes it challenging due to the dependency of the flow rate on the level of the congestion in the network, but it can contribute to this area of research, especially in extreme cases where a very high demand level is expected and congestion plays a big role in the decision making process.

- The cell transmission model with flow reduction still requires a lot of computational effort. A more tractable formulation is required to expand the practical use og this model.

- Improve the tractability of the proposed hospital evacuation problem. The current model can be easily expanded to the case of several evacuating hospitals, however due to the complexity of the current formulation it cannot be applied in real time to large problem instances.

- In the hospital problems, several inputs can be stochastic(e.g. number of patients, number of beds), therefore one way to extend to study is to assume some of these inputs to be random.

Chapter 7

Bibliography

A. Chakirov, F.P. (2014). "Enriched Sioux Falls Scenario with Dynamic And Disaggregate Demand."

Abdulaal, M. and LeBlanc, L.J. (1979). "Methods for combining modal split and equilibrium assignment models." *Transportation Science*, **13 (4)**, pp. 292–314.

Addison, P.S. and Low, D.J. (1998). "A novel nonlinear car-following model." *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **8 (4)**, pp. 791–799.

Banks, J.H. (1990). "Flow processes at a freeway bottleneck." *Transportation Research Record*, **1287**, pp. 20–28.

Bish, D.R., Chamberlayne, E.P., and Rakha, H.A. (2013). "Optimizing network flows with congestion-based flow reductions." *Networks and Spatial Economics*, **13 (3)**, pp. 283–306.

Bish, D.R. and Sherali, H.D. (2013). "Aggregate-level demand management in evacuation planning." *European Journal of Operational Research*, **224 (1)**, pp. 79–92.

Bish, D.R., Sherali, H.D., and Hobeika, A.G. (2014). "Optimal evacuation strategies with staging and routing." *Journal of Operational Research Society*, **65**, pp. 124–140.

Bish, D.R., Agca, E., and Glick, R. (2011). "Decision support for hospital evacuation and emergency response." *Annals of Operations Research*, pp. 1–18.

Camion, P. (1965). "Characterization of totally unimodular matrices." *Proceedings of the American Mathematical Society*, **16 (5)**, pp. 1068–1073.

Carey, M. (1987). "Optimal time-varying flows on congested networks." *Operations Research*, **35 (1)**, pp. 58–69.

Carey, M. and Eswaran, S. (2000). "An approach to modelling time-varying flows on congested networks." *Transportation Research Part B: Methodological*, **34 (3)**, pp. 157–183.

Cassidy, M.J. and Bertini, R.L. (1999). "Some traffic features at freeway bottlenecks." *Transportation Research - B*, **33 (1)**, pp. 25–42.

Chalmet, L., Francis, R., and Saunders, P. (1982). "Network models for building evacuation." *Fire Technology*, **18 (1)**, pp. 90–113.

Chen, L. and Miller-Hooks, E. (2008). "The building evacuation problem with shared information." *Naval Research Logistics (NRL)*, **55 (4)**, pp. 363–376.

Chiu, Y., Zheng, H., Villalobos, J., and Gautam, B. (2007). "Modeling no-notice mass evacuation using a dynamic traffic flow optimization model." *IIE Transactions*, **39 (1)**, pp. 83–94.

Choi, W., Hamacher, H.W., and Tufekci, S. (1988). "Modeling of building evacuation problems by network flows with side constraints." *European Journal of Operational Research*, **35 (1)**, pp. 98–110.

Chung, B.D., Yao, T., Xie, C., and Thorsen, A. (2011). "Robust optimization model for a dynamic network design problem under demand uncertainty." *Networks and Spatial Economics*, **11 (2)**, pp. 371–389.

Chung, B.D., Yao, T., and Zhang, B. (2012). "Dynamic traffic assignment under uncertainty: A distributional robust chance-constrained approach." *Networks and Spatial Economics*, **12 (1)**, pp. 167–181.

Chung, K., Rudjanakanoknad, J., and Cassidy, M.J. (2007). "Relation between traffic density and capacity drop at three freeway bottlenecks." *Transportation Research - B*, **41 (1)**, pp. 82–95.

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 II: network traffic." *Transportation Research Part B: Methodological*, **29 (2)**, pp. 79–93.

Desaulniers, G., Desrosiers, J., and Solomon, M.M. (2005). *Column generation*, volume 5. Springer.

Dixit, V.V., Ramasamy, S., and Radwan, E. (2008). "Assessment of I-4 contraflow plans: Microscopic versus mesoscopic simulation." *Transportation Research Record*, **2041**, pp. 89–97.

Duanmu, J., Taaffe, K., and Chowdhury, M. (2010). "Minimizing patient transport times during mass population evacuations." *Transportation Research Record: Journal of the Transportation Research Board*, **2196 (1)**, pp. 150–158.

FEMA (2013). "Disaster declarations by year."
URL `http://www.fema.gov/disasters/grid/year`

Fleischer, L.K. (2001). "Faster algorithms for the quickest transshipment problem." *SIAM Journal on Optimization*, **12 (1)**, pp. 18–35.

Gipps, P.G. (1981). "A behavioural car-following model for computer simulation." *Transportation Research Part B: Methodological*, **15 (2)**, pp. 105–111.

Golmohammadi, D. and Shimshak, D. (2011). "Estimation of the evacuation time in an emergency situation in hospitals." *Computers & Industrial Engineering*, **61 (4)**, pp. 1256–1267.

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

Hamacher, H. and Tufekci, S. (1987). "On the use of lexicographic min cost flows in evacuation modeling." *Naval Research Logistics (NRL)*, **34 (4)**, pp. 487–503.

Han, L.D., Yuan, F., and Urbanik, T. (2007). "What is an effective evacuation operation?" *Journal of Urban Planning and Development*, **133 (1)**, pp. 3–8.

Jarvis, J.R. and Ratliff, D.H. (1982). "Some equivalent objectives for dynamic network flow problems." *Management Science*, **28 (1)**, pp. 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*.

Köhler, E. and Skutella, M. (2005). "Flows over time with load-dependent transit times." *SIAM Journal on Optimization*, **15 (4)**, pp. 1185–1202.

Li, Y., Waller, S.T., and Ziliaskopoulos, A.K. (2003). "A decomposition scheme for system optimal dynamic traffic assignment models." *Networks & Spatial Economics*, **3**, pp. 441–455.

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 (4)**, pp. 238–245.

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.

Lo, H.K. (2001). "A cell-based traffic control formulation: strategies and benefits of dynamic timing plans." *Transportation Science*, **35 (2)**, pp. 148–164.

Løvås, G.G. (1995). "On performance measures for evacuation systems." *European Journal of Operational Research*, **85 (2)**, pp. 352–367.

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

Merchant, D.K. and Nemhauser, G.L. (1978). "A model and an algorithm for the dynamic traffic assignment problem." *Transportation Science*, **12**, pp. 183–199.

Newkirk, R.T. (2001). "The increasing cost of disasters in developed countries: A challenge to local planning and government." *Journal of Contingencies and Crisis Management*, **9 (3)**, pp. 159–170.

Nie, Y. (2011). "A cell-based Merchant-Nemhauser model for the system optimum dynamic traffic assignment problem." *Transportation Research - B*, **45**, pp. 329–342.

Richards, P.I. (1956). "Shock waves on the highway." *Operations research*, **4 (1)**, pp. 42–51.

Schrijver, A. (1998). *Theory of linear and integer programming.* Wiley. com.

Shen, W., Nie, Y., and Zhang, H.M. (2007). "Dynamic network simplex method for designing emergency evacuation plans." *Transportation Research Record*, **2022**, pp. 83–93.

Sherali, H.D. (1982). "Equivalent weights for lexicographic multiple objective programs: characterizations and computations." *European Journal of Operational Research*, **11 (4)**, pp. 367–379.

Sternberg, E. (2004). "Counting Crises: US Hospital Evacuations." *Prehospital Disaster Medicine*, **19 (2)**, pp. 150–157.

Treiber, M., Hennecke, A., and Helbing, D. (2000). "Congested traffic states in empirical observations and microscopic simulations." *Physical Review E*, **62 (2)**, p. 1805.

Tuydes, H. and Ziliaskopoulos, A. (2006). "Tabu-based heuristic approach for optimization of network evacuation contraflow." *Transportation Research Record*, **1964**, pp. 157–168.

Waller, S. and Ziliaskopoulos, A.K. (2006). "A chance-constrained based stochastic dynamic traffic assignment model: Analysis, formulation and solution algorithms." *Transportation Research Part C: Emerging Technologies*, **14 (6)**, pp. 418–427.

Yao, T., Mandala, S.R., and Chung, B.D. (2009). "Evacuation transportation planning under uncertainty: A robust optimization approach." *Networks and Spatial Economics*, **9 (2)**, pp. 171–189.

Zheng, H. and Chiu, Y.C. (2011). "A network flow algorithm for the cell-based single-destination system optimal dynamic traffic assignment problem." *Transportation Science*, **45 (1)**, pp. 121–137.