International Series in
Operations Research & Management Science

Rajan Batta
Changhyun Kwon *Editors*

Handbook of OR/MS Models in Hazardous Materials Transportation

Springer

International Series in Operations Research & Management Science

Volume 193

For further volumes:
http://www.springer.com/series/6161

Rajan Batta · Changhyun Kwon
Editors

Handbook of OR/MS Models in Hazardous Materials Transportation

Springer

Editors
Rajan Batta
Department of Industrial
and Systems Engineering
University at Buffalo, The State University
of New York
Buffalo, New York, USA

Changhyun Kwon
Department of Industrial
and Systems Engineering
University at Buffalo, The State University
of New York
Buffalo, New York, USA

ISSN 0884-8289
ISBN 978-1-4614-6793-9 ISBN 978-1-4614-6794-6 (eBook)
DOI 10.1007/978-1-4614-6794-6
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013938602

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Introduction 1
Rajan Batta and Changhyun Kwon

Railroad Transportation of Hazardous Materials: Models for Risk Assessment and Management 9
Manish Verma and Vedat Verter

Operations Research Models for Global Route Planning in Hazardous Material Transportation 49
Lucio Bianco, Massimiliano Caramia, Stefano Giordani, and Veronica Piccialli

The Effect of Weather Systems in Hazmat Transportation Modeling 103
Mohsen Golalikhani and Mark H. Karwan

Value-at-Risk and Conditional Value-at-Risk Minimization for Hazardous Materials Routing 127
Iakovos Toumazis, Changhyun Kwon, and Rajan Batta

Hazardous Facility Location Models on Networks 155
Marcos Colebrook and Joaquín Sicilia

Network Interdiction Methods and Approximations in a Hazmat Transportation Setting 187
Justin Yates

Optimal Emergency Resources Deployment Under a Terrorist Threat: The Hazmat Case and Beyond 245
Rodrigo A. Garrido

The Role of OR in Emergency Evacuation from Hazmat Incidents 269
Brian Wolshon and Pamela Murray-Tuite

Index 293

Introduction

Rajan Batta and Changhyun Kwon

The Pipeline and Hazardous Materials Safety Administration of the U.S. Department of Transportation defines hazardous materials (hazmat) as a substance or material capable of posing an unreasonable risk to health, safety, or property when transported in commerce[1]. Hazmat accidents can result in significant impact to the population (death, injuries) and damage to the environment (destroyed or damaged buildings and infrastructure). Further, hazmat, especially explosive materials, can potentially be used by terrorists to attack civilians or to destroy critical infrastructure. This handbook provides models from Operations Research and Management Science that study various activities involving hazmat transportation: risk assessment, route planning, location decisions, evacuation planning, and emergency planning for terrorist attacks.

There are two important research areas in hazmat transportation that are widely studied in the literature: risk assessment and shipment planning. In the risk assessment area, important issues include measurement of accident probabilities and consequences in hazmat transport. Example works in the risk assessment area include modeling risk probability distribution over given areas, considering hazmat types and transport modes, and environmental conditions.

We can classify most scholarly contributions in the shipment planning into two subfields: local route planning and global route planning. When only a single

[1] http://phmsa.dot.gov/hazmat/glossary

R. Batta
Department of Industrial and Systems Engineering, University at Buffalo, The State University of New York, Bell Hall 410, Buffalo, NY 14260, USA
e-mail: batta@buffalo.edu

C. Kwon (✉)
Department of Industrial and Systems Engineering, University at Buffalo, The State University of New York, Bell Hall 318, Buffalo, NY 14260, USA
e-mail: chkwon@buffalo.edu

R. Batta and C. Kwon (eds.), *Handbook of OR/MS Models in Hazardous Materials Transportation*, International Series in Operations Research & Management Science 193,
DOI 10.1007/978-1-4614-6794-6_1, © Springer Science+Business Media New York 2013

origin–destination pair is concerned, the routing problem is called local route planning. When multiple origin–destination pairs and/or multiple types of hazmat are involved, it is called global route planning.

In global route planning problems, we often face two important groups of decision makers in hazmat transportation: network regulators and hazmat carriers. While hazmat carriers usually choose an economic path, e.g. the least cost path, network regulators attempt to minimize the potential impact of hazmat accidents to the surrounding communities. For this reason, OR/MS models in global routing planning often involve bi-level optimization for network design, location planning, and regulations.

The second half (chapters "Hazardous Facility Location Models on Networks" to "The Role of OR in Emergency Evacuation from Hazmat Incidents") of this handbook provide useful models and insights on other important issues, while the first half (chapters "Railroad Transportation of Hazardous Materials: Models for Risk Assessment and Management" to "Value-at-Risk and Conditional Value-at-Risk Minimization for Hazardous Materials Routing") cover the two fields of risk assessment and shipment planning. Chapter "Hazardous Facility Location Models on Networks" discusses location problems for undesirable facilities, while chapters "Network Interdiction Methods and Approximations in a Hazmat Setting" and "Optimal Emergency Resources Deployment Under a Terrorist Threat: The Hazmat Case and Beyond" discuss network interdiction and terrorist attack. Chapter "The Role of OR in Emergency Evacuation from Hazmat Incidents" provides an overview of evacuation planning in cases of accidents involving hazmat. In what follows, we provide a brief summary for each chapter.

In chapter "Railroad Transportation of Hazardous Materials: Models for Risk Assessment and Management," Verma and Verter provide risk assessment methods for a railroad mode of hazmat transportation with a routing problem in a south-east US railway network system. The chapter begins with describing typical freight rail transportation systems and defining key notions to understand rail systems: physical networks, service networks, itinerary, and blocking.

Chapter "Railroad Transportation of Hazardous Materials: Models for Risk Assessment and Management" also reviews three risk assessment methods in railway hazmat transportation: expected consequence, incident probability and population exposure. The expected consequence measure considers the accident probability at each link with the population exposure considering conditional probabilities of derailing of a railcar, derailing of a hazmat railcar, and release of hazmat. To obtain the population exposure, a notion of the exposure band is introduced.

While a simple version of the expected consequence measure is introduced in this chapter, the authors bring in-depth discussion to each component of the expected consequence measure: accident probability and accident consequence. First, the incident probability measure focuses on the likelihood of a hazmat incident considering train derailment, point of derailment, number of railcars derailed, and number of railcars releasing hazmat. These factors are modeled by conditional probabilities and constitute the incident probability at any given point of a path.

On the other hand, the population exposure measure emphasizes the total number of people who may be exposed to an undesirable consequence. While the average population exposure is used in the expected consequence measure, a worst-case approach is employed in the population exposure measure. The Gaussian plume model is used to describe the air dispersion of toxic gases at given wind speeds and directions. This model provides the region and radius of evacuation and consequently the number of people exposed.

In this chapter, an optimization model is proposed for routing. The presented hazmat railcar routing problem is a global route planning problem involving multiple numbers of commodities, transfer yards, train services, and itineraries. This chapter also suggests potential research questions in railway transportation of hazmat. As a next reading, readers who are interested in other popular risk measures in truck transportation of hazmat may refer to chapter "Value-at-Risk and Conditional Value-at-Risk Minimization for Hazardous Materials Routing," and those who are interested in in-depth discussion of weather consideration may refer to chapter "The Effect of Weather Systems in Hazmat Transportation Modeling."

Chapter "Operations Research Models for Global Route Planning in Hazardous Material Transportation," by Bianco, Caramia, Giordani and Piccialli, provides an extensive review of recent advances in the global route planning field. This chapter first reviews local route planning problems, classifying key features considered in each paper: security, decision support system, geographical information system, multi-objectives, stochastic, and survey.

The global route planning literature is classified into three sub-classes. First, research articles with equity considerations, dissimilar paths, and multi-objective models are reviewed. In this sub-class of problems, one of the main objectives is to find least risk routes between origins and destinations. Second, hazmat transportation network design problems are reviewed. In a network design problem, the central authority prohibits carriers from traveling certain parts of the road network, while the carriers find the minimum distance routes given the restrictions imposed by the central authority. Third, problems of toll setting policies are reviewed. In these problems, the central authority discourages carriers to travel certain links by charging tolls, rather than completely prohibiting travel (which is the special case of an infinite toll). In most network design problems and toll setting problems, bi-level mathematical optimization and game-theoretic approaches are used.

In a subsequent section, the authors provide formal descriptions of key hazmat transportation network design models. They first begin with models with an authority and several carriers, and then review a model with local and regional authorities. The latter model differs from the former models in the sense that the role of carriers is not considered and government authorities at different levels are considered. While the local authority at the upper level sets network link capacities, the regional authority at the lower level assigns hazmat flows on the capacitated network.

Three toll setting problems are also introduced with formal descriptions. The main advantage of toll setting policy over network design is flexibility of controlling hazmat flow. The first model reviewed is a bi-level optimization problem with binary

variables where the central authority sets the toll for each link and the carriers decide the least cost path considering the toll and the travel time. In the second model, risk equity is considered in the context of a toll setting policy as an extension to the first model, in a mathematical programming with equilibrium constraints (MPEC) framework. The toll on each link in the second model is assumed to be a quadratic function of the total risk induced on that link. In the third model, the notion of dual toll setting policy is introduced to charge tolls for both regular vehicles and hazmat trucks.

Golalikhani and Karwan present changes in risk assessment by weather, in chapter "The Effect of Weather Systems in Hazmat Transportation Modeling." Weather systems may affect both accident probabilities and accident consequences. Because weather affects road surface frictions and visibility of drivers, changes in weather also affect accident probability, hence the risk. This chapter begins with a model to estimate the accident probability given various factors, including weather conditions like fine weather, rain/fog, and snow/ice.

For the effects on the accident consequences, the chapter considers two types of models: threshold distance approaches and air pollution dispersion models. The threshold distance approaches are the same as what were called the fixed bandwidth approaches in chapter "Railroad Transportation of Hazardous Materials: Models for Risk Assessment and Management." The popular Gaussian plume model to describe dispersion of air pollutants is introduced and discussed in depth. The Gaussian model is mostly useful to predict the dispersion of continuous, buoyant air pollution plumes that begin at either the ground level or an elevated level. The authors also review two non-Gaussian dispersion models: dense gas models and Largrangean models. Dense gas models simulate the dispersion of pollution plumes that are heavier than air, and Lagrangean models mathematically describe the dispersion of pollution plume parcels by modeling their move in the air by a random walk.

The dynamic nature of weather systems is discussed and two examples of hazmat transportation systems considering weather systems are provided. In each model, detailed model formulations are provided and followed by a simple, but still practically useful, numerical example that illustrate how one may use the models in practice.

At the end of the chapter, potential research topics are provided in three directions. First, one may study more realistic and complicated weather models using GIS. Existing GIS-relevant studies do not consider either the weather system or the dynamic nature of it. Second, one may consider different dispersion models for different types of hazmat and different release conditions. Although the Gaussian plume model is useful, many hazmat types and accident conditions are not suitable to be considered by the Gaussian model. Third, one may model the shape and the movement of the weather system more realistically. The only such existing hazmat transportation model describes the dynamic nature of the weather system by a simple circle along a straight line.

In chapter "Value-at-Risk and Conditional Value-at-Risk Minimization for Hazardous Materials Routing," Toumazis, Kwon and Batta provide a summary of recent advances in local route planning. Popular risk measures for hazmat routing include

expected risk, population exposure, incident probability, perceived risk, maximum risk, mean-variance, disutility, and conditional probability measures. This chapter begins with a brief review of those popular risk measures and comparison.

Two recently proposed risk measures in hazmat transportation are introduced: Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). These two risk measures are popularly, especially in financial portfolio optimization to make investment decisions. In hazmat routing, VaR is a threshold accident consequence value such that the probability at which the accident consequence is greater than VaR is less than or equals to a certain probability (confidence level). VaR in general is computationally intractable; however, in hazmat the underlying risk variable is a discrete random variable. An algorithm that solves a finite number of shortest-path problems is proposed for finding the minimal VaR path. This chapter shows that VaR becomes risk-indifferent when the confidence level is small enough, that is, at all paths VaR becomes zero. When the confidence level is large enough, the VaR model becomes identical to the maximum risk model, which minimizes the maximum link accident consequence in a path.

Risk measures must satisfy the conditions for translation-invariance, subadditivity, positive homogeneity, and monotonicity. These four properties are desired properties for any risk measure. While VaR is useful and popular, unfortunately it is not coherent. As a coherent extension to VaR, CVaR is proposed. Roughly, CVaR is the expected value of consequences that is beyond VaR. Finding a minimal CVaR path is a better-structured optimization problem than finding a minimal VaR path. However, to solve a CVaR problem, we also need to solve a finite number of shortest-path problems. While the CVaR model becomes identical to the expected risk model when the confidence level is small enough, the CVaR model becomes identical to the maximum risk model when the confidence level is large enough.

A simple numerical example is provided to illustrate the shape of the probabilistic distribution and how to determine the VaR and CVaR values for various confidence levels. Later, a case study on a realistic road network around Albany, NY, USA is provided and key findings are discussed.

Chapter "Hazardous Facility Location Models on Networks," by Colebrook and Sicilia, provides a review of modeling methods and algorithmic approaches for determining locations of undesirable—both noxious and obnoxious—facilities such as hazmat waste dump sites, incinerators, gas stations, electrical plants, etc. The chapter begins with an extensive literature review, which consists of three parts. First the chapter provides a list of survey and review papers, and useful books so that readers can refer to them if interested. The chapter then provides a literature review on undesirable facility location problems on networks. It is followed by a literature review on multicriteria undesirable facility location problems on networks.

After introducing basic definitions and notations, the chapter provides three models: the uncenter (undesirable center) model, the maxian problem, and the anti-cent-dian model. The uncenter problem is to maximize the minimum distance between the undesirable facility and all other nodes. The maxian problem is to maximize the weighted sum of distances from the undesirable facility to all other nodes. The objective of the anti-cent-dian problem is to maximize a combination

of the minimum distance and the weighted average of distances between the undesirable facility and all other nodes. An algorithm is proposed for solving each problem.

This chapter also covers a location problem on multicriteria networks. There are some facilities that are semi-desirable or semi-undesirable, such as airports and train stations that provide services but also generate noise. In such a case, on the same network, two different "distances" may be defined for service and discomfort. The authors provide the multicriteria λ-anti-cent-dian problem with a formulation and an algorithm.

Chapter "Network Interdiction Methods and Approximations in a Hazmat Transportation Setting," by Yates, focuses on the application of interdiction models in a hazmat transportation setting. The chapter offers a detailed literature review on the past and current hazardous materials transportation research. The first topic of focus is on the quantification of risk, potential pitfalls, and estimation as a tool to measure risk. It also discusses the use of Geographic Information Systems in supporting risk analysis. The third topic of focus is on network interdiction problems.

After the literature review section, the chapter formulates interdiction models. Specifically, it explores the shortest path network interdiction problem and discusses multiple variations that can be derived from it. It then motivates development of alternative heuristic and approximation techniques to solve network interdiction problems.

After the formulation and initial exploration of interdiction models, the chapter highlights the computational performance and spatial characteristics of shortest path network interdiction problem variants. Computational results show that sensor allocation strategies with increased fidelity come at a steep computational price. Overall, this chapter succeeds in illustrating the importance of network interdiction modeling as it relates to hazardous materials transportation.

Chapter "Optimal Emergency Resources Deployment Under a Terrorist Threat: The Hazmat Case and Beyond," by Garrido, considers the logistics of emergency systems in the context of human-made catastrophes. In particular, the interest is in terrorist attacks that use a hazardous materials transport vehicles as a weapon. It is argued by the author that the infamous attacks of September 11, 2001, were indeed the case of a hazmat vehicle (in this case an airplane with an almost full fuel tank) being used a terrorist weapon and being exploded at a target of interest.

The author works on the assumption that a hazmat threat has been positively identified and consequently the resource allocation question becomes relevant. In many cases this information provides a basis more accurate than other methods to estimate attack probabilities. An example cited is that of a hazmat truck that deviates from its original path. Different cases of terrorist attack are analyzed. The first case is when resources are allocated on the basis of an expected event. The second case examined is where a given percentage of demand must be satisfied. The third case is a Stackelberg game approach where the probability of attack is a function of the defenders resources allocated to each zone. The fourth case is that of using a variational inequality approach, when the probability of attack is a function of all resources allocated to each zone. Overall, this chapter presents many different viable approaches

for estimating the probability of a hazmat terrorist attack and also presents many strategies for allocating resources to defend against such a potential attack.

In chapter "The Role of OR in Emergency Evacuation from Hazmat Incidents," Wolshon and Murray-Tuite focus on evacuation operations, guidance, and plans for hazmat incident responses. The authors first characterize hazmat incidents by pointing out that hazmat incidents are usually generated as a result of man-made disasters, but sometimes are caused by natural disasters. Such natural disasters can result in technological disasters, called natech disasters. Recent examples of natech disasters include nuclear radiation leaks from the Fukushima Daiichi Nuclear Power Plan in Japan at an earthquake. Hazmat disasters have a unique characteristic that an incident can happen virtually anywhere, while locations or approach directions of other disasters are usually fixed or predictable.

The authors introduce and define temporal parameters and spatial parameters in evacuation transportation processes. Simulation and analysis techniques are discussed in three modeling scales: macro-level, micro-level, and meso-level. Mixed modeling approaches are also briefly discussed.

After characterizing hazmat incidents and categorizing modeling and simulation approaches, the authors provide reviews of various OR models and applications for evacuation management. They cover ramp management, crossing elimination, staged/phased evacuation, destination assignment, combined departure-time/destination/route optimization, shelter location and assignment, and transit operations. For each topic, a literature review is provided with useful insights, and for some topics, a representative model is present and discussed. The authors close the chapter discussing the key conditions of OR models to be successful and widely used in evacuation planning for hazmat incidents.

It is our sincere hope that this handbook will be of value to both hazmat practitioners and academic researchers. Our objective was to seek contributions that covered a wide facet of hazmat logistics and planning using OR/MS methods. We believe that the collection of chapters indeed provides such a comprehensive treatment. We end by gratefully acknowledging the Department of Industrial and Systems Engineering at the University at Buffalo (SUNY) for providing a hospitable environment to complete this project.

Railroad Transportation of Hazardous Materials: Models for Risk Assessment and Management

Manish Verma and Vedat Verter

Introduction

Hazardous materials (hazmat) are harmful to humans and the environment because of their toxic ingredients, but their transportation is essential to sustain our industrial lifestyle. A significant majority of hazmat shipments are moved via the highway and railroad networks. For example, in the United States, railroad carries approximately 1.8 million carloads of hazmat annually, which translates into 5% of rail freight traffic (AAR 2006). On the other hand, in Canada, approximately 500,000 carloads of hazmat—equivalent to 12% of total traffic—are shipped by railroad (TSB 2004). The quantity of hazmat traffic on the railroad network is expected to increase significantly over the next decade, given the phenomenal growth of intermodal transportation and the growing use of rail-truck combinations to move chemicals. It is true that railroads have a favourable safety statistic (Oggero et al. 2006), but the possibility of spectacular events resulting from multicar incidents, however small, does exist. The derailment of the BNSF train in Lafayette (Louisiana, United States), spilling 10,000 gallons on hydrochloric acid and forcing more than 3,000 residents out of their homes, is an example of such low probability–high consequence events. In fact in the United States, between 1995 and 2009, around 120 train accidents resulted in release from multiple tank cars, which translates into an average of eight accidents every year (FRA 2010).

Over the past four decades, the railroad industry has spent considerable effort in reducing the frequency of tank car accidents as well as the likelihood of releases

M. Verma (✉)
DeGroote School of Business, McMaster University, Hamilton, Canada
e-mail: mverma@mcmaster.ca

V. Verter
Desautels Faculty of Management, McGill University, Montreal, Canada
e-mail: vedat.verter@mcgill.ca

R. Batta and C. Kwon (eds.), *Handbook of OR/MS Models in Hazardous Materials Transportation*, International Series in Operations Research & Management Science 193,
DOI 10.1007/978-1-4614-6794-6_2,

in the event of an accident. To this end, the Association of American Railroads, the Chemical Manufacturers Association and the Railway Progress Institute formed an inter-industry taskforce in the early 1970s (Conlon 1999). Unfortunately, the activities of this voluntary task force largely ceased in about 1994, and most of their internal reports were never publicized and considered proprietary to the sponsoring organizations (Barkan 2004; Conlon 2004). The more recent academic and industry initiatives have focused on analyzing past accident data in an effort to increase railroad safety by improving rail-tracks or railcar tank designs; and, on risk assessment and management.

In an effort to communicate the nature of railroad risk stemming from hazmat shipments, this book chapter will first describe the workings of a (freight) rail transportation system. An understanding of the characteristics of the transportation system will facilitate the discussion of different risk assessment methodologies and risk management techniques developed for railroad transportation of hazmat. Finally, a brief outline of some open problems is provided before concluding.

This book chapter is organized as follows: section "Rail Transportation System" sketches the workings of freight rail transportation, and introduces the integral operational elements. Section "Risk Assessment Methodologies" contains an extensive discussion of the three most popular measures of transport risk, i.e., *expected consequence*, *incident probability*, and *population exposure*. Though these measures were developed for highway transportation, however in recent years, they have been adapted to capture the dynamics of railroad accident and hence could be used to assess hazmat transport risk from railroad shipments. Section "Risk Management Techniques" outlines techniques to manage and/or mitigate hazmat transport risk, including design of rail tank cars, placement of hazmat cars in a train, and network routing. Finally, issues presenting potential for research are indicated in section "Potential Research Questions," and conclusions are given in section "Summary."

Rail Transportation System

The most common approach to represent rail transportation system is via a network, whose nodes represent *yards* (or stations) and whose arcs represent *tracks* on which trains carry freight (or passengers). Freight demand is usually expressed in terms of tonnage or number of railcars of certain commodities to be moved from an origin to a destination. For every origin-destination pair of demand, the corresponding freight may be shipped either directly or indirectly. When demand is important enough delivery delays are minimized by using direct trains, or else freight is routed through intermediate nodes in order to accumulate enough tonnage (or volume) to justify dispatching a train. As a result, one could view the rail operating policies as a sequence of decisions intended to bring about as close a match between demand (i.e., freight traffic) and available resources (i.e., feasible train routes, train itineraries, crew and motive power, and yard capabilities) (Verma 2005). To facilitate discussion on risk assessment and management in the following sections we next

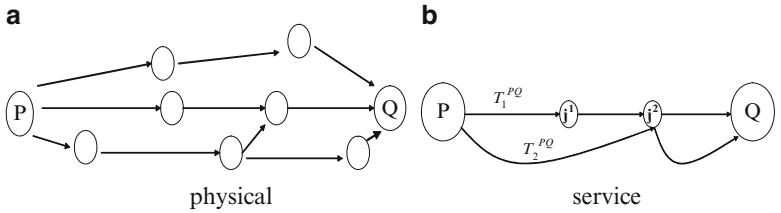


Fig. 1 Railroad Network (*Source*: Verma 2005). (**a**) Physical, (**b**) service

outline the relevant components of a rail transportation system, and invite the reader refer to Cordeau et al. (1998) for an excellent review of the different works in the railroad domain.

Figure 1a represents a physical network of a railroad by a directed graph $G_{ph} = (N, A_{ph})$ where N is the set of nodes and A_{ph} is the set of rail segments representing tracks between yards. Based on the physical network G_{ph}, the service network $G = (N, A)$ specifies the set of feasible routes on which train services may be operated. In Fig. 1b, the service network of two train services T_1^{PQ} and T_2^{PQ} is graphed. The service of a freight train is characterized by an origin yard, a destination yard, a sequence of rail track segments from the origin to destination yard, and a set of intermediate stops. The track segment between two consecutive stops of a train service is called a *service leg*. Although both train services have the same origin and destination, the set of intermediate stops and the service legs are different and hence the two are distinct. This is because while T_1^{PQ} goes through j^1 and j^2 with possible pick-up/drop-off operation at these intermediate yards, T_2^{PQ} has only j^2 as the intermediate yard. Furthermore, T_1^{PQ} has three service-legs (i.e., P–j^1, j^1–j^2, and j^2–Q), T_2^{PQ} has two (i.e., P–j^2 and j^2–Q).

In the railroad industry, demand (or traffic-class) is characterized by unique origin and destination yards. For example in Fig. 1b, P–j^2 would constitute a demand, since it has a unique origin and destination yard. It is easy to see that this demand can be met using either of the two train services in the available network. A feasible journey of a traffic-class from the origin to the destination yard, including the train services and yard operations, is called its *itinerary*. As indicated earlier, train service-legs are composed of track sections between two consecutive stops, whereas yard operations can entail grouping and/or transfer of railcars. For the simple scenario in Fig. 1, traffic-class P–j^2 has two itineraries: *first*, on train service T_2^{PQ} via intermediate yard j^1, where the traffic-class in consideration is not touched since the train performs just the pick-up/drop-off operation; and *second*, using T_2^{PQ}—the direct non-stop service. It is important to note that the indicated movement was a simple illustration of a railcar and did not involve any classification (i.e., grouping of railcars) and/or transfer at intermediate yards, an aspect that is integral to the operation of the railroad industry, and explained using Fig. 2.

Assume that we intend to track traffic-class (or demand) $A \rightarrow F$, where A is the origin and F the destination yard, on the given network. There are four train services:

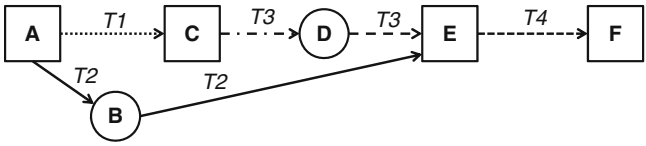


Fig. 2 Service network, itinerary, and blocking

T1 operates between yards A and C; ***T2*** originates in yard A and terminates at E, with B as the intermediate stop; ***T3*** starts at C, services D before terminating in yard E; and, ***T4*** operates between yards E and F. It is important that origin and destination yards of freight trains possess the capability to classify railcars, which are represented by square nodes in Fig. 2, and are distinct from nodes (i.e., circle) with only transfer capabilities. It may not be either evident or well known, but classification (also called *blocking*) is the major determinant that enables railroads to realize economies of scale. Blocking of railcars is done to prevent handling of each railcar at every intermediate yard on its journey. A group of railcars with common handling points are grouped together at the start of its journey, and this group is not disbanded until it reaches the specified handling yard, which is the destination for that block. On reaching the destination for that specific block, further classification and blocking operations may have to be performed on individual railcars depending on their destinations. The process continues till a railcar reaches its destination.

To make this more explicit, suppose ***T1*** brings railcars associated with traffic-class $A \rightarrow F$ to yard C, where some operation would be performed in preparation for the onward journey. At C, one of the two activities will be performed on the traffic-class in question. If the number of railcars in this traffic class is sufficient to form a *block* (i.e., a group of railcars with common handling point), then the entire set will be placed on the departure track at yard C waiting to be connected to the outbound train ***T3***. On the other hand if the number of railcars is not enough to be deemed a block, i.e., the typical case, then some yard intensive activities would have to be performed. The incoming railcars will be sorted according to the final (or an intermediate) destination and grouped together in a block, and placed on a departure track for outbound movement. In some instances, a classification yard may have dedicated tracks for building blocks for different destinations. Assume that marshalling yard C has a dedicated track for yard E, and hence all the railcars belonging to traffic-class $A \rightarrow F$ will be blocked with other railcars with E being either the intermediate handling point or the final destination.

Now train service ***T3*** leaves C with the blocks on its take-list, stops at D—which is only a service yard and hence only block-swap and/or drop-off can happen. The train terminates at yard E, where the incoming blocks would be sorted and then classified based on the next handling point. Finally, traffic class $A \rightarrow F$ will be placed on ***T4***, together with other traffic with F as the handling point, for the final leg of its journey. Given the network structure of the railroad, it is important to understand that the type of operations to be performed on incoming traffic at any yard depends on the operations performed on the preceding yards, and the anticipated activities

at the yards to follow. Typically these decisions are made at a strategic level, and hence each yard master knows the type of operations to be performed.

The sequence of blocks to which a railcar is assigned along its route from origin to destination is called the *blocking path* (Barnhart et al. 2000; Newton et al. 1998). It is worth noting that the blocking path may be different from the physical route. For example, traffic-class $A \rightarrow F$ has two routes: *A–**T2**–E–**T4**–F*; and, *A–**T1**–C–**T3**–E–**T4**–F*. Note that yards *B* and *D* are not pertinent in the discussion of blocking paths, since blocks cannot be built at those locations. The first route has two blocking paths: **A-F**, and **A-E-F**. On the other hand the second route has four blocking paths: **A-C-E-F**; **A-C-F**; **A-E-F**; and **A-F**. Hence, even for this simple instance, i.e., one traffic-class and sparse rail network, there are six possible blocking paths on two physical routes. For realistic problem instances, the number of blocking paths could grow exponentially and hence some intelligent enumeration techniques and/or rule of thumb should be implemented to generate reasonable paths. For example, one could enumerate only the direct paths, which involves no circular connections and/or paths that are within x% of the shortest paths between the given OD pair. Alternatively, one could limit the number of intermediate handlings.

To recap, the objective of this section was to provide a brief outline of a rail transportation system, which will facilitate development of risk assessment methodologies and discussion of the risk management (mitigation) techniques in the following sections.

Risk Assessment Methodologies

Although railroads move a significant quantity of hazmat both in the United States and in Canada, which has translated into increased research over the past decade, an overwhelming majority of academic initiatives in the preceding periods focused on road shipments (Erkut et al. 2007). The tremendous strides made in the highway domain, unfortunately, could not be extended to railroads because of differences between the two modes. For example, a train usually carries both regular and hazmat cargo together, whereas these two are almost never mixed in a truck shipment. Secondly, a rail tank car has roughly three times the capacity of a truck-tanker, and the number of hazmat railcars varies significantly among different trains. The resulting variability in the total amount of hazmat needs to be taken into account in assessing the rail transport risk, wherein railroads typically have much less routing flexibility compared to trucks. Finally, hazmat incidents involving freight trains could entail content-release from multiple railcars. For example, in the United States, between 1995 and 2009, around 120 train accidents resulted in release from multiple tank cars, which translates into an average of eight accidents every year (FRA 2010). In December 1999, Canadian National Ultratrain released 2.7 million liters of petroleum products due to the derailment of 35 tank cars just outside Montreal. Thirty cars were seriously punctured and had to be demolished at the accident site (Railway Investigation 2002). Another well-known accident

took place near Toronto in 1979, in which chlorine leaking from damaged tank cars forced the evacuation of 200,000 people (Swoveland 1987). Thus, a train accident can have more severe consequences than those involving trucks, mainly due to the higher volumes of hazmat being shipped and the interaction between railcars.

Though the last few years have witnessed the development of risk assessment methodologies that incorporate the specific nature of railroad shipments, they were preceded mostly by works focusing on accident rate analysis. Glickman and Rosenfield used past train derailment data to derive three forms of risk: the probability distribution of the number of fatalities in a single accident; the probability distribution of the total number of fatalities from all the accidents in a year; and, the frequency of accidents that result in any given number of fatalities (Glickman and Rosenfield 1984). On the other hand, Barkan et al. (2003) conducted a statistical analysis of the railroad accident data to conclude that the speed of derailment and the number of derailed cars are highly correlated with hazmat release, and then proposed estimating direct and conditional probabilities in conducting risk analysis (Anderson and Barkan 2004a). While these engagements made use of empirical data for insights and conclusions, the recent efforts geared towards developing assessment methodologies can be listed under the three most popular measures of hazmat transport risk: *expected consequence*; *incident probability*; and, *population exposure*. Each of the three measures has been extensively used to study highway transportation of hazmat, and, until recently, similar activities were not witnessed in the railroad domain. Though expected *consequence*—defined as the product of probability and consequence of an undesirable event—is perhaps the most popular measure, dearth of data and/or the limitations associated with this measure have resulted in the development of other risk measures. For example, the *incident probability* measure focuses just on the probability of undesirable event (Saccomanno and Chan 1985; Abkowitz et al. 1992), whereas the *population exposure* measure considers the total number of individuals exposed to hazmat shipments (Batta and Chiu 1988; ReVelle et al. 1991). We outline the developmental works associated with each transport risk measure in the subsections to follow.

Expected Consequence

Interestingly, until very recently, the most popular measure of hazmat transport risk was not adapted for railroad transportation, which in part could have resulted from data limitations and/or complexity of the transportation system. Subsequently, in an effort to develop a risk assessment methodology that incorporates the characteristics of railroad accidents, FRA (Federal Railroad Administration) freight derailment information from 1995 to 2009 was collected and analyzed to reveal the five major causes of railroad accidents: track, roadbed and structures; human factors; mechanical and electrical failures; signal and communication; and, miscellaneous—with track and human factors accounting for 70% of the accidents (Verma 2011). In addition, it was also noticed that derailment probability of each position in a train is

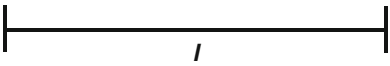


Fig. 3 Depiction of a rail-link

not constant and dependent on the train length. The aforementioned implies that an appropriate risk assessment methodology should consider: (a) the probability that a train carrying hazmat will be involved in an accident, which should be represented as a composite of the five causes of rail accidents; (b) the conditional probability that a hazmat railcar will derail, which would depend on both the train-length and railcar position in the train-consist; (c) the conditional probability that the railcar will rupture and release its contents; and, (d) the consequence as a result of hazmat release from multiple sources.

Although the 25,000 derailment instances collected for the above insights could have been further processed to reveal position-specific derailment probabilities, such numbers would not have much use since freight-train lengths vary. This is why a study commissioned by the FRA and the Department of Transportation (DOT) computed derailment probabilities based on the four quarters of the train. Approximately 165,000 derailed railcars were analyzed to conclude that derailment probabilities are highest in the first and lowest in the fourth quarter of the train (Thompson et al. 1992). Since the preliminary analysis in Verma (2011) was not entirely consistent with the conclusion in Thompson et al. (1992), the former proposed risk assessment at a higher resolution, i.e., by focusing on a unit smaller than a quarter, such as *deciles* (viz. ten equal parts). Conceivably a decile-based approach should result in better analysis, but only if train-lengths are similar (or constant). For example, a train with 12 railcars is completely contained within the first decile of a train with 240 railcars, and hence their derailment probabilities should be different. In the absence of any peer-reviewed work suggesting length-based categorization of freight-trains, it was suggested that freight-trains with up to 40 railcars be called *short*; between 41 and 120, *medium*; and, the rest *long* (Verma 2011). It is important to note that the expected consequence methodology, to be developed next, is similar to the ones presented for road shipments, except the adaptations necessary to capture the characteristics of railroad accidents.

Link Risk

Consider a rail-link l of unit length (Fig. 3). If the probability that a train meets with an accident on this rail-link is given by A_l and the resulting consequence by C_l, then the expected consequence (or traditional risk) over l can be represented by:

$$ER_l = A_l \times C_l \tag{1}$$

where C_l would be determined as:

$$C_l = P(D^i|l) \times P(H|D^i, l) \times P(R|H, D^i, l) \times Exposure_l \tag{2}$$

where $P(D^i|l)$ is the probability of derailment of a railcar in the ith decile of the train given the accident on link l; $P(H|D^i, l)$ is the probability that a hazmat railcar derailed in the ith decile of the train given the accident on link l; $P(R|H, D^i, l)$ is the probability of release from a hazmat railcar derailed in the ith decile of the train given the accident on link l; and, $Exposure_l$ is the population exposure due to the release from hazmat railcars given the accident on link l. While (1) is the well-known definition of traditional risk, (2) is the relevant adaptation to incorporate the characteristics of railroad accidents.

Although the FRA website provides comprehensive rail data, accident rates for each and every rail-link in the network may not be available given that hazmat accidents are rare events, and hence it would not be unreasonable to use network-wide accident rates. Furthermore, if the aggregated accident rates do not provide sufficient resolution for direct estimation of conditional probabilities, then logical diagrams, such as fault trees and event trees, should be used to estimate the probability of an event based on the historical probabilities of a set of basic event for which sufficient data or expert judgements is available.

It is important to note that railroads generally transport multiple hazmat, whose impact on human life and their interactions are unknown, and hence any risk assessment exercise would require development of an approximation technique. Verma (2011) proposes adopting a conservative approach by basing evaluation on the hazmat likely to be most detrimental, and provides the rationale by outlining three reasons: *first*, it precludes underestimation of risk; *second*, it facilitates better emergency response preparedness since the hazmat being considered is likely to cause maximum damage; and *third*, it offsets the adverse impact of not keeping track of individual hazmat.

Exposure

The population exposure approach, as proposed for road shipments (Batta and Chiu 1988; ReVelle et al. 1991), is modified to incorporate the possibility and volume of hazmat released from multiple sources, and the resulting relationship can be represented as:

$$Exposure_l = f(V_l,\ \rho(V_l)) \tag{3}$$

where V_l is the volume of hazmat released due to the accident on rail-link l, and $\rho(V_l)$ is the population density of centers exposed due to V_l. If complete information on hazmat interaction is not known, one could make use of the approximation technique and work with total volume of hazmat released from all sources. But if such information is available, then one should use the exact expression to capture the interaction. If v_l^n is the quantity of hazmat released from railcar n due to the accident on link l, the total volume of hazmat released from all the sources due to the accident on link l can be determined by:

$$V_l = \sum\nolimits_n v_l^n \tag{4}$$

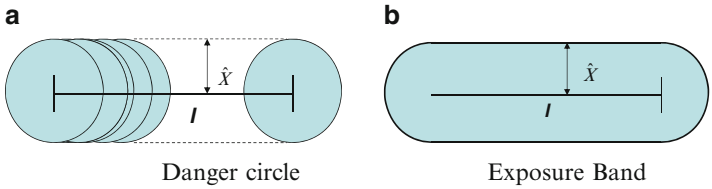


Fig. 4 Region of impact (*Source*: Verma 2011). (**a**) Danger circle, (**b**) exposure band

For *a priori* risk assessment, Verma (2011) proposes loss of entire lading from a rail tank car when no information on volume is available. This suggestion was defended by running a number of scenarios in ALOHA—an atmospheric dispersion model used for evaluating release of hazardous chemicals, where even a 4-in. rupture diameter resulted in the loss of the entire lading within 7 min, which is considerably less than the anticipated emergency response times. Now if the impact area can be represented as a danger circle (of radius $\hat{X}$), then the hazmat transport activity can be visualized as the movement of this circle along the rail-link (of length l), which carves out a band as the region of possible impact (Fig. 4). The number of people living in the band is the *population exposure* (Batta and Chiu 1988; ReVelle et al. 1991).

Evacuation Distance

If the quantity of hazmat released is not significant (i.e., minor spill), then the evacuation distance as specified in the Emergency Response Guidebook should be used for determining population exposure (ERG 2012). On the other hand, if the quantity of hazmat released is outside the specified guidelines (i.e., major and larger spills), as would typically be associated with train accidents, then evacuation distance should be determined by: aggregating release at various downwind locations (from multiple sources); and, then comparing the aggregate hazmat level with the immediately dangerous life and health (IDLH) levels to ascertain the threat zone. Indeed this is an application of the population exposure based methodology developed to tide over the limitations associated with collecting comprehensive data to estimate traditional risk (Verma and Verter 2007), and will be detailed in section "Population Exposure."

Route Risk

For railroads, a route is a collection of rail-links and intermediate yards connecting the origin and destination for the specific freight-train (Figs. 1 and 2). Assume that a rail route is comprised of only two links: l and $l + 1$. It is easy to see that travel on this is a probabilistic experiment, since the expected consequence on $l + 1$ depends

on whether the train meets with an accident on l. Hence, the expected consequence associated with this route is:

$$A_l C_l + [1 - A_l] A_{l+1} C_{l+1} \tag{5}$$

where, A_{l+1} is the probability of meeting with an accident on $l+1$, and C_{l+1} the resulting consequence. To generalize, if there are m rail-links in the route under consideration, the expected consequence associated with this route would be expressed as follows:

$$\begin{aligned} &A_l C_l + [1 - A_l]\, A_{l+1} C_{l+1} + [1 - A_l][1 - A_{l+1}]\, A_{l+2} C_{l+2} + \ldots \\ &+ \left\{ \prod_{i=1}^{m-1} [1 - A_i] \right\} A_m C_m \end{aligned} \tag{6}$$

where, $[1 - A_1]\,[1 - A_2]\,[1 - A_{k-1}] A_k C_k$ is the expected consequence for the kth rail-link.

Ideally appropriate accident rate probabilities, for rail-links, should be used to determine corresponding expected consequence, though in reality such numbers may not be available for each and every link in the network, and hence a different estimation technique may be required. In the absence of detailed accident data, (6) can be simplified given the infinitesimally small incident probabilities of the order 10^{-6}. Such approximation, made to tide over data limitations, have been shown to introduce negligible errors for a number of instances based on a road network (Erkut and Verter 1998). It is possible to make similar approximations for railroads, since incident probabilities are also of the order 10^{-6}, and hence one can assume $A_l = A$ for each rail-link l of similar track-type. If index i is used to denote the rail-links forming route P, then the expected consequence of transporting hazmat along route P would be:

$$ER_P = A \sum\nolimits_{i \in P} C_i \tag{7}$$

where C_i is the consequence due to train accident on rail-link i. Note that consequence will be different for each rail-link since it depends on the population density of the exposed population centers, which is not uniform along a train route.

Although Verma (2011) was a useful effort in getting an understanding of the nature of railroad risk, it made use of an approximate approach (i.e., using train-deciles) to compute risk. For example, for a 100 railcar train, the conditional derailment and release probability for the 81st and the 89th railcars are identical. However, in reality the conditional probabilities for the two positions would be different since they are subject to differing forces. Hence, there was a need for an assessment methodology that could not only reconcile various train lengths but also took into consideration—the derailment probabilities of every position in the train-consist, initial point of derailment, number of railcars derailed, and the resulting

consequence from multiple sources. The most recent effort by Bagheri et al. (2012) attempts to fill this gap, and is briefly outlined next.

Comprehensive Risk Assessment Framework

Consider again a rail-segment l, and a hazmat railcar in that is in the ith position along the train-length. Transport risk stemming from the ith hazmat railcar on the given rail-segment l can be defined as the product of the derailment probability for position i, P_l^i, and consequence, C_l^i.

$$Risk_l^i = P_l^i \times C_l^i \tag{8}$$

where, the probability of derailment for position i can be calculated as the product of the probability of train derailment on segment l, i.e., $P(TD)_l^i$, and the conditional probability that position i is the point where derailment starts (POD) given that the train has derailed, i.e., $P(POD)_l^i$. Note that the proposed framework focuses on railcars after the point of derailment in the computation of risk.

$$P_l^i = P(TD)_l^i \times P(POD)_l^i \tag{9}$$

and, the consequence of derailment can be calculated as the product of the conditional probabilities that m railcars derail as a result of derailment beginning at position i on the given rail-segment, i.e., $P(m|i)$, and j hazmat railcars among the m derailed will release, i.e., $P(j|m)$, together with the population exposure associated with j railcars, PE_j.

$$C_l^i = \sum_{m=1}^{n-i+1} P(m|i) \left[\sum_{j=0}^{m} P(j|m) PE_j \right] \tag{10}$$

Substituting (9) and (10) in (8), and summing over all the rail-segments of a route R, the complete expression for determining transport-risk is:

$$Total\ Risk_R = \sum_{l \in R} \sum_{i=1}^{n} \left\{ P(TD)_l^i \times P(POD)_l^i \times \sum_{m=1}^{n-i+1} P(m|i) \left[\sum_{j=0}^{m} P(j|m) PE_j \right] \right\} \tag{11}$$

Please note that (11) is, in part, motivated by the incident probability work of Bagheri et al. (2011), and hence we outline the relevant parameter estimation technique in the next section. We conclude this section with a comment that *population exposure* was estimated by adopting the worst-case approach by assuming least favorable weather conditions and focusing on maximum concentrate levels, and we provided the pertinent details in section "Population Exposure."

Incident Probability

Although the expected consequence approach is simple to use and justify, as indicated earlier, a dearth of detailed data necessitated development of alternate risk measures. The first alternate measure focused on the likelihood of a hazmat incident, i.e., *incident probability*. This measure, also developed in the highway context, was appropriate for hazmat with relatively small danger zones (Saccomanno and Chan 1985; Abkowitz et al. 1992). Consequently the assessment methodology adapted for railroads incorporated the pertinent operational characteristics, and endeavored to determine the best position to place hazmat railcars such that in-transit risk is minimized (Bagheri et al. 2011). To that end, the proposed framework recommends assigning hazmat railcars to those positions along the train that have the lowest probability of derailing along the different route segments, which could be done by making use of service engines at the rail marshaling yards. The proposed model requires two types of inputs: hazmat and regular shipment volume by intermediate and final destination along a given route; and, route attributes (i.e., design features of a rail segment).

To elaborate, again assume a rail-segment l. Now the incident probability risk from a railcar in the ith position along the train length can be determined by:

$$IP_l^i = P(TD)_l \times P(i|TD) \times Y_l^i \tag{12}$$

where, $P(TD)_l$ is the probability of train derailment on rail-segment l; $P(i|TD)$ is the conditional probability of a railcar in the ith position derailing; and, Y_l^i is a binary variable that assumes a value of 1 if the railcar in the ith position contains hazmat, and 0 otherwise. It should be clear that the introduction of binary variables ensures that only hazmat railcars are considered in impact assessment, whereas railcars with regular freight would not contribute to transport risk. The conditional probability of derailment, i.e., $P(i|TD)$, can be determined as:

$$P(i|TD) = \sum\nolimits_{j=1}^{i} \left[P_j^{POD} \times \sum\nolimits_{x=i-j+1}^{n-j+1} P(x|j) \right] \tag{13}$$

where, P_j^{POD} is the probability that the point of derailment starts at position j; and $P(x|j)$ is the probability of x railcars from a total of n derailing given the point of derailment starting at position j. Note that several factors can result in train derailment, and hence it is important to consider all of them when estimating the probability of derailment involving any position (Bagheri et al. 2011). Substituting (13) in (12), we get the elaborate expression for incident probability for the ith position on rail-segment l, which is:

$$IP_l^i = \left\{ P(TD)_l \times \sum\nolimits_{j=1}^{i} \left[P_j^{POD} \times \sum\nolimits_{x=i-j+1}^{n-j+1} P(x|j) \right] \right\} \times Y_l^i \tag{14}$$

and for all positions along a train of length n, and over all rail-segments on route R is:

$$IP_R = \sum_{l \in R} \sum_{i=1}^{n} IP_l^i \tag{15}$$

It should be evident that since (15) looks just at the hazmat incident, consequence is completely ignored—unlike in (11) where population exposure is crucial to the transport risk. Clearly the proposed model takes into consideration the specific route and rolling-stock attributes to determine the minimum risk slots for hazmat railcars for a given rail-segment, additional investigation is necessary to ascertain the impact of this model at a network level (i.e., multiple yards and trains). It is our belief that such studies will provide further insights into the feasibility of implementing a similar plan at different marshalling yards, and perhaps yield more generalizable results that are not specific to a particular transportation corridor.

We next discuss the techniques for estimating various parameters in both (11) and (14), and invite the reader to refer to the indicated references for additional details. As indicated earlier, population exposure in (11) is determined by using the framework of Verma and Verter (2007), and will be discussed in section "Population Exposure."

Train Derailment

Anderson and Barkan (2004b) showed that the probability of freight train derailment is a function of distance traveled, train length, and track class; and proposed the following expression for calculating the probability of derailment:

$$P(TD)_l^i = 1 - e^{\{-D[RC(TL)+RT]\}} \tag{16}$$

where, D is the travel distance; TL is the train length; RC is the derailment rate per billion freight car-miles; and, RT is the derailment rate per million freight train-miles. The model was based on aggregate data for accident rates for different types of track class in terms of the number of derailments per billion freight car-miles, and the number of derailments per million freight train-miles (Table 1). It is important to note that the probability of derailment will vary with rail segment and number of blocks in the train-consist.

Point of Derailment (POD)

A railcar can be involved in a derailment either by initiating the derailment, or by being amongst the units derailed following the POD—and the latter is assumed in (Bagheri et al. 2011).

Table 1 Derailment rates (*Source*: Anderson and Barkan 2004b)

	FRA track class[a]				
Derailments per	1	2	3	4	5 and 6
Million freight train miles	48.54	6.06	2.04	0.53	0.32
Billion freight car miles	720.10	92.70	31.50	7.80	4.90

[a]The FRA classified tracks based on various quality and speed considerations. Class 1 represents the poorest tracks and wherein speed limit is 16 km/h, whereas Class 6 is the best with permissible speed of 177 km/h

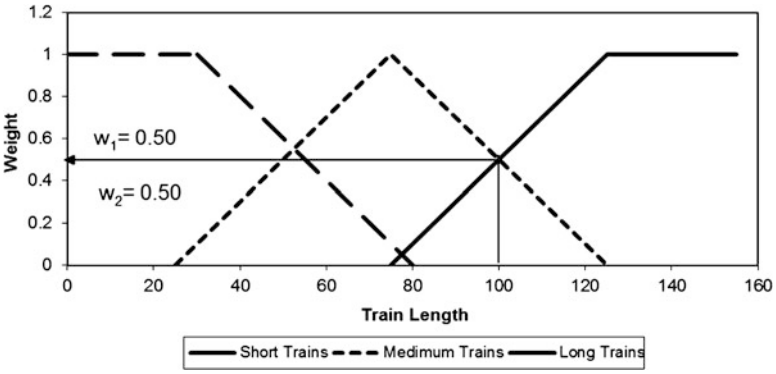


Fig. 5 Determining weights for different types of trains (*Source*: Bagheri et al. 2012)

It was suggested that derailment causes can be grouped under three classes depending on the part of the train likely to derail first. More specifically, causes which affect the front of the train (C_F), the rear of the train (C_R), and the middle of the train (C_M). In addition, trains were categorized into three types: *short* (up to 40 railcars); *medium* (between 41 and 120 railcars); and, *long* (more than 120 railcars). Finally, a nonparametric Kruskal-Wallis test was applied to show that *train-length* and *causes* provide a statistically significant explanation for median point of derailment (POD). Best fit POD distributions were obtained for all nine length-cause combinations, and different weights were used to account for uncertainty caused by overlap in train-length at the boundaries (i.e., around 40 and 120 railcars). Figure 5 can be used to determine the two types of trains under consideration and the appropriate weights.

For example, for a 100-railcar train over a rail-segment subject to cause C_F (causes that affect the front of the train), the respective weights are $w_1 = 0.50$ for a medium-train classification, and $w_2 = 0.50$ for a long-train classification, and zero for short-train classification. Now the probability of derailment for the tenth position (in a train with 100 railcars) with cause C_F is obtained by applying the following steps: (1) determine the weights for relevant train-lengths (i.e., w_1 and w_2); (2) compute the normalized point of derailment, which for this example is: $10/100 = 0.10$; (3) obtain the derailment probabilities for the position of interest (for this example, $i = 10$) the values are 0.021 and 0.023 from (Bagheri et al. 2011);

and, (4) substitute values in (17) to obtain $P(POD)_l^i$, which for this example is 0.022.

$$P(POD)_l^i = \frac{f_1(j) \times w_1(e) + f_2(j) \times w_2(e)}{w_1(e) + w_2(e)} \tag{17}$$

where, $f_1(j)$ and $f_2(j)$ are estimated for position j, and the membership values (weights) as explained above.

Number of Railcars Derailed

The number of railcars derailing is affected by the dissipation of kinetic energy following a train derailment, which implies that train speed and distance from the point of derailment are important. Note that the latter point is relevant because as the distance to the POD increases, forces of instability acting on the remaining railcars decrease. Given that risk is posed by a derailed hazmat railcar, all hazmat railcars placed before the POD or beyond the derailed block do not pose any risk on the given rail-segment. A truncated geometric expression, to estimate the probability that m railcars will derail given that derailment started at position i (Saccomanno and El-Hage 1989), is proposed.

$$P(m|i) = \frac{p(1-p)^{m-1}}{1-(1-p)^{TL_r}} \tag{18}$$

where, $m = 1, 2, \ldots, TL_r$, and TL_r is the number of railcars in the train past the point of derailment, and $(1-p)$ is the probability of derailment for a position after the point of derailment.

In an effort to evaluate the significance of other causal factors on the probability of derailment (i.e., p in 18), summary statistics for the FRA database were generated at the 95% significance level. From Table 2, it is possible to conclude that the number of railcars derailing beyond the POD has a strong association with all the primary causes. In addition, Saccomanno and El-Hage (1991) suggested using a logistic regression of the following form to estimate the probability of derailment beyond the POD.

$$p = \frac{e^{\{\beta_0 + \sum_k \beta_k X_k\}}}{1 + e^{\{\beta_0 + \sum_k \beta_k X_k\}}} \tag{19}$$

where, X_k indicates the impact of the kth independent factor on the probability of derailment beyond the POD. Table 2 depicts the summary statistics for the various causal factors for the 1997–2006 FRA dataset (Bagheri et al. 2011).

Table 2 shows that at 95% confidence level, each of the eight factors is significant, and that increasing the speed will increase the probability of derailment when using

Table 2 FRA data causal factor statistical summary (*Source*: Bagheri et al. 2011)

Parameters	Estimates	Std. error	Z statistics	Lower 95%	Upper 95%
Intercept	−2.013	0.082	−24.46	−1.85	−2.17
Residual length (R_L)	0.001	0.001	1.266	0.002	−0.001
Speed effect (S_E)	−0.032	0.002	−17.07	−0.029	−0.036
Roadbed (R_B)	0.419	0.018	2.367	0.766	0.072
Track geometry (T_G)	0.171	0.089	1.921	0.346	−0.003
Switches (S)	0.715	0.119	6.013	0.949	0.482
General car (G_C)	0.841	0.085	10.132	1.03	0.697
Axles/wheels (A_W)	1.108	0.077	14.404	1.26	0.958
All other (O)	0.444	0.073	6.056	0.587	0.30

Table 3 Number of derailments resulting in hazmat release (*Source*: FRA 2010)

Year	Number of accidents	Hazmat involved	Hazmat derailed	Hazmat release
1997	545	152	64	11
1998	620	201	95	23
1999	547	206	102	20
2000	598	202	91	22
2001	627	223	97	20
2002	541	192	95	16
2003	582	196	95	17
2004	594	174	74	14
2005	580	171	69	13
2006	563	204	103	11
Total	5, 797	1, 921	885	167

a logistic regression function. In addition, the impact of the factor A_W, representing axles/wheels, on derailment is less compared to that of track geometry.

Number of Railcars Releasing Hazmat

The conditional probability that *j* hazmat railcars release, given that *m* have derailed, is estimated from empirical data. For example, Bagheri et al. (2012) looked at the period from 1997 to 2006. Approximately 5,800 freight train accidents of which one-third included hazmat were recorded (Table 3). A total of 885 hazmat railcars derailed resulting in release from 167 (or, 18.8%). It was interesting to note that a number of hazmat release episodes involved multiple railcars, which necessitates taking into consideration potential volume released. For instance, in eight of the eleven accidents in 2006, multi railcar release episodes were recorded. In fact, the worst incident involved release from 20 hazmat railcars.

In an effort to take advantage of the empirical dataset, it was proposed that the conditional probability of release from a derailed hazmat car (q) be independent of each other, such that:

$$P(j|m) = q \times q \times \cdots \times q = q^j \tag{20}$$

According to the FRA database (1997–2006), the conditional probability of release from a derailed hazmat railcar (q) is 0.0903, obtained at an aggregate level by dividing the number of railcars releasing by the total number of hazmat railcar derailed.

Note that we have outlined the estimation technique for every parameter introduced in (11), except *population exposure*, which we do next.

Population Exposure

The second alternate measure, also developed to tide over the limitations associated with the expected consequence approach, represented transport risk as the total number of people exposed to the possibility of an undesirable consequence due to the shipment. For example, according to the Emergency Response Guidebook (ERG 2012), 800 m around a fire that involves a chlorine tank, railcar or tank-truck must be isolated and evacuated. Therefore, the people within the predefined threshold distance from the railroad are exposed to the risk of evacuation. This fixed bandwidth approach was originally suggested by Batta and Chiu (1988), and ReVelle et al. (1991). It is important that, in contrast with the traditional "average" risk measure, population exposure constitutes a "worst-case" approach to transport risk. Therefore, it is particularly suitable for assessing risk as perceived by the public as well as for estimating the required emergency response capability.

Verma and Verter (2007) contended that the use of the bandwidth approach, that implicitly assumes a standard hazmat volume, is inappropriate for estimating the number of people put at risk due to railroad shipments. This is because both the number and location of hazmat railcars vary considerably among trains, and hence it was important to define the boundary of impact area as a function of the hazmat volume on the train. A train with one propane tank-car, for example, exposes an individual living 1 km from the rail track to minor injury risk, whereas the same individual would be exposed to fatality risk due to a train with 21 propane tank cars. This can be explained by the considerable increase in the toxic concentration level at the individual's location due to the additional 20 railcars. In an effort to develop the population exposure framework for railroads, the authors focus on hazmat that becomes airborne on release (such as chlorine, propane and ammonia). Furthermore, they define exposures in terms of the level of toxic material concentration, which are estimated using the most popular air dispersion model, i.e., Gaussian plume model (GPM) (Arya 1999). For example, an individual is considered "exposed" to a certain undesirable consequence, if the imposed toxicity is at (or higher than) the associated threshold level for the given hazmat (also called the immediately dangerous to life and health—IDLH level).

The standard GPM is first adapted to represent a single railcar (release source), which is then extended to represent train shipments, which typically involve multiple release-sources. Assuming that the release-source and the impact point are at zero elevation, the single railcar model is as follows:

$$C(x, y) = \frac{Q}{\pi u \sigma_y \sigma_z} exp\left(-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2\right) \tag{21}$$

where $C(x, y)$ is the concentration level (ppm) at impact point (x, y) in *steady-state*, Q is the release rate of pollutant (mg/s), u is the average wind-speed (m/s), σ_y the horizontal dispersion coefficient (m), $\sigma_y = ax^b$, σ_z the vertical dispersion coefficient (m), $\sigma_z = cx^d$, x the downwind distance from the source (m), and y the crosswind (perpendicular) distance from the source (m).

In estimating the steady-state concentration level at point (x, y), the model assumes that the release rate and atmospheric conditions remain constant over the period of dispersion. Although the steady-state conditions are rarely reached, this is a common assumption—particularly reasonable during the first hour of release. The release rate, Q, depends on container volume, hazmat type and rupture diameter. The authors used ALOHA (US Environmental Protection Agency 2011), a popular software among North American regulatory agencies including EPA, US Department of Transportation and Transport Canada, to calculate the release rate. Although ALOHA can also be used for estimating the concentration level, $C(x, y)$, its results are only reliable within 1 h of the release event, and 10 km from the release source. In order to assess the population exposure under worst-case conditions, the highest release rate was incorporated by assuming a 24 in. rupture at the bottom of the railcar—though a sensitivity analysis on different rupture sizes was also performed. Dispersion coefficients σ_y and σ_z are determined by atmospheric stability category and the downwind distance, x, to the release source. Pasquill and Smith (1983) and more recently Arya (1999) provide the values of dispersion parameters a, b, c and d based on atmospheric stability category. Since minimum wind speed, under any atmospheric category, resulted in maximum concentration at all points in the plane—the authors assumed minimum possible wind speed of 2.5 m/s under the neutral atmospheric conditions. Given the objective of developing the most conservative scenarios, the authors focused on downwind points (i.e., where crosswind distance $y = 0$), which yields the following expression for determining concentrate level at downwind distance, x, from the release source:

$$C(x) = \frac{Q}{\pi u \sigma_y \sigma_z} \tag{22}$$

The two shaded areas in Fig. 6 depict the two zones of a Gaussian plume footprint from a single release source, when the wind is blowing east. The two zones represent areas where toxicity is higher than a pre-specified concentration level (i.e., IDLH level for the given hazmat). The inner zone corresponds to higher exposure to

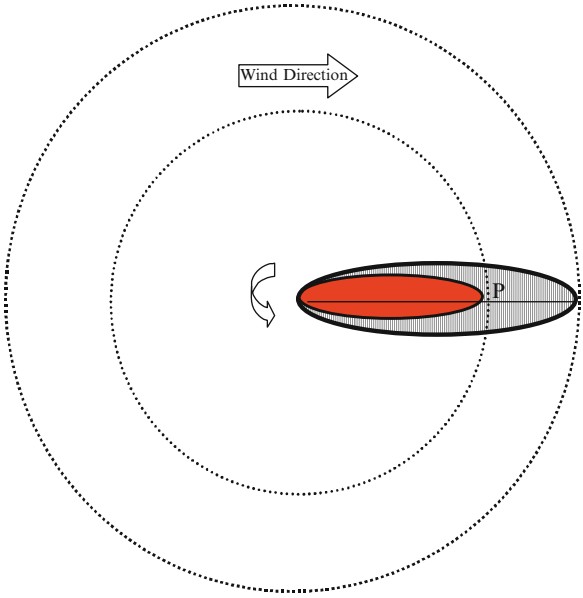


Fig. 6 The footprint of a Gaussian plume at a given wind speed (*Source*: Verma and Verter 2007).

hazmat transport risk, while the outer to non-severe exposure. The authors address the uncertainty in wind direction by rotating the footprint around the release source, which constructs two concentric circles. Consequently, the furthest point from the release source, where threshold concentration level is attained, defines the radius of the circle (e.g., *P* in Fig. 6). Of course this constitutes the most conservative approach to transport risk, since the concentrate level at any point on the circle cannot be higher under any plausible wind direction. Conceivably, there may be prevailing winds along some segments of the train's route. If, for example, wind blows within the east and north directions along a track segment, then only the upper-right quarters of the danger circles need to be used for estimating population exposure. Note that, in contrast with the fixed bandwidth approach, the radius of each impact area in Fig. 6 varies with the release rate.

The authors next extended the basic GPM to incorporate multiple release-sources as follows. Assume that the 11-railcar train in Fig. 7 is travelling east, and that *F* and *L* denote the first and last hazmat railcar, respectively. *M* is the point with equal amount of hazmat on both sides, which is referred to as the *hazmat-median* of the train. For trains with an even number of hazmat railcars, *M* is midpoint of the two hazmat railcars at the center of hazardous cargo. Note that *M* and *D*, the middle of the train, do not necessarily refer to the same point. Finally, *P1, P2* and *P3* are equidistant from *M*; and the five hazmat railcars are *blocked* at the back of the train (Verma and Verter 2007).

Both (Arya 1999) and (Pasquill and Smith 1983) suggest that pollution from an array of sources with an arbitrary distribution of position and strength of emission can be modeled by superposing the patterns of pollution from these sources, and

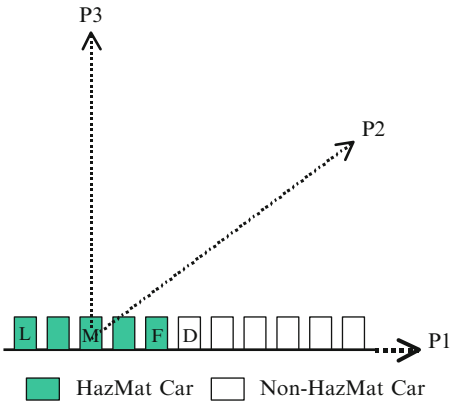


Fig. 7 Schematic representation of an 11-railcar train (*Source*: Verma and Verter 2007)

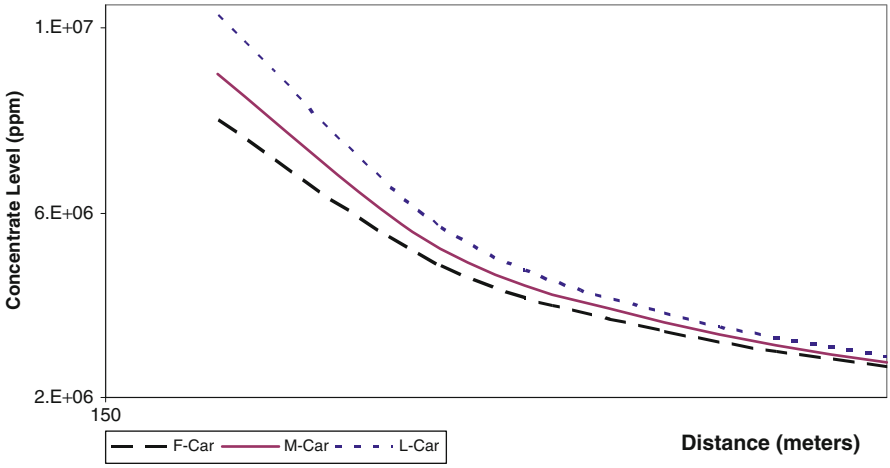


Fig. 8 Impact of a five hazmat railcar block (*Source*: Verma and Verter 2007)

hence aggregating the resulting contamination at each impact point. In Fig. 7, when the wind is blowing east, *P1* constitutes a downwind location where crosswind distance $y = 0$ for all railcars. In the event of a major incident that ruptures all five railcars with hazardous cargo, the total concentrate level at *P1* would be sum of the concentrate levels associated with each railcar, which can be estimated via (22). The three curves in Fig. 8 depict total concentration at *P1* as a function of the distance to *F, M,* and *L*, respectively. Consider a fixed reference point at $x = 0$. As the train travels east, *F, M,* and then *L* pass by the reference point. Thus, Fig. 8 also shows the upward shift in concentrate level as a result of the train's movement. Note that contaminant toxicity increases much faster at impact points closer to the train (i.e., non-linear).

When the wind is blowing northeast in Fig. 7, *P2* is downwind from *M* and has positive crosswind distances (i.e., $y > 0$) to the other four railcars. Therefore, the

maximum concentration at *P2* cannot exceed that of *P1*. Similarly, the maximum concentration at *P3,* which is attained when the wind is blowing north, is less than that of P1. Thus, *P1* is the *maximum concentration point* among all the locations equidistant from *M*. When the distance from hazmat-median of an 11-railcar propane block is 1,500 m, for example, the concentrate levels at *P2* and *P3* are 95.8% and 95.5%, of the maximum level, respectively. This difference decreases with distance and increases with the number of hazmat railcars.

Analogous to the single release-source case, it is possible to estimate the exposure around the train by rotating the maximum concentration point, *P1,* around the hazmat-median, *M*. Therefore, the authors use the hazmat-median as the reference point for the train—since this would assure consistency among the maximum concentrate levels under opposite wind directions, when hazmat railcars are blocked. If another point were used as reference, the concentrate levels at the opposite downwind locations from the hazmat railcar block would be different. Take, for example, *F* as an alternative reference point. Since all the railcars are behind *F*, the total concentrate level at a certain downwind distance will be higher when the train is moving upwind. Because the amount of hazardous cargo on both sides of *M* is the same, it constitutes the best option for a reference point.

Thus, the maximum concentrate level at distance x from the hazmat-median of an n-railcar hazmat block is:

$$C_n(x) = \frac{Q}{\pi uacx^b x^d} + \frac{Q}{\pi uac(x-s)^b(x-s)^d} + \frac{Q}{\pi uac(x+s)^b(x+s)^d} + \\ \ldots + \frac{Q}{\pi uac(x-ns/2)^b(x-ns/2)^d} + \frac{Q}{\pi uac(x+ns/2)^b(x+ns/2)^d} \tag{23}$$

where, s denotes the length of each railcar. Although (23) can be used to compute the aggregate contaminant level from multiple sources, (Verma and Verter 2007) makes use of the relative size difference between the cross-length of a Gaussian plume and length of a railcar to outline the following approximation instead of (23).

$$\bar{C}_n(x) = n \times \frac{Q}{\pi uacx^b x^d} \tag{24}$$

where n is the number of identical release sources with rate Q. This amounts to assuming that all the hazmat cargo is located at the hazmat-median of the train. Finally, in order to demonstrate the reliability of (24) for estimating evacuation distances, and also its robustness in terms of train make-up, two separate analyses were performed.

Three different trains with 30, 68 and 120 hazmat railcars were considered. For each train, evacuation distances were computed using both (23) and (24), and error—defined as the percentage deviation from the base value as determined using the exact approach—was computed. The approximate model estimated all the evacuation distances within an error margin of 1.26%. As expected, the aggregate

Table 4 Random positioning of hazmat railcars

	Number of hazmat railcars		
Evacuation distance (m)	5	11	21
Approximate	1,341	2,204	3,417
Average	1,366	2,213	3,422
Std. dev.	60	41	33

concentrate level shifts upward proportionately to the number of hazmat railcars, although the non-linearity in (23) implies that the aggregate concentrate level will increase at a faster pace in relation to the increase in the value of n. Consequently, the accuracy of (24) increases with distance and decreases with the number of hazmat railcars. At 1,000 m, for example, the approximation errors associated with the severe zone for 30 and 68 hazmat railcars are 1.14% and 7.11%, respectively. These errors reduce to 0.45% and 3.64% at 1,600 m from the hazmat-median of the train. Nevertheless, the approximation errors near the train are inconsequential since the concentrate levels are very high, making a severe consequence almost certain.

In order to analyze robustness of the approximate model with respect to positioning of the hazmat railcars in a train, three cases: 5; 11; and 21 hazmat tank-cars in a train with a total of 68 railcars were considered. Transport Canada regulations stipulate that the first and last five railcars in a non-unit train cannot carry hazardous cargo. Thus, 100 train make-ups were generated for each case by randomly positioning the hazmat tank-cars among the 6th and 63rd railcars. Table 4 shows the (statistics associated with) non-severe threshold distances as well as the approximations. Given n, (24) estimates the same distance for all random train make-ups, since all hazmat is aggregated at the hazmat-median. The accurate calculation of total concentrate level, however, needs to incorporate the actual distance to each hazmat railcar. To illustrate this, consider a train make-up with hazardous cargo in the 6th, 11th, 15th, 37th and 63rd railcars. The hazmat-median, in this example, is the 15th railcar. Note that the hazmat median will remain the same if the hazardous cargo in the 63rd railcar was moved to the 38th railcar, whereas the actual toxicity level at downwind distances from the train will change as a result. Thus, the average threshold distances of 100 random train make-ups for each of the three cases are depicted in Table 4.

Since the approximation error is within 2% for all three cases, the authors surmised that the approximate model remains effective under uncertainty regarding the positioning of hazmat railcars in the train. Also, the approximate model performs better as the number of hazmat railcars increases. This can be explained by the reduction in variance of the threshold distance as hazardous content of the train increases. The distances in Table 4 are calculated at downwind locations assuming that the train is traveling east, as in Fig. 7. If the train is traveling in the opposite direction, the average threshold distances will be slightly different, whereas the approximate distance will remain the same.

To summarize, in this section we have outlined the three most popular measures of transport risk. Each of three measures was developed in the highway domain,

and hence had to be adapted to capture the dynamics of railroad accident. Having outlined the comprehensive risk assessment methodologies, we turn our attention to elaborate on how to manage and/or mitigate hazmat transport risk.

Risk Management Techniques

A number of industry and academic studies have been undertaken over the past three decades to investigate the issue of risk mitigation. As alluded to earlier, the railroad industry has spent considerable effort in reducing the frequency of tank car accidents as well the likelihood of releases in the case of an accident. To that end, the Association of American Railroads, the Chemical Manufacturers Association and the Railway Progress Institute formed an inter-industry taskforce in the early 1970s (Conlon 1999). Unfortunately, the activities of this voluntary task force largely ceased in about 1994, and most of their internal reports were never publicized and considered proprietary to the sponsoring organizations (Barkan 2004; Conlon 2004). In this section, we will briefly outline the more recent initiatives focused on improving tank car safety at the design stage, draw from the insights of some earlier works dealing with strategic placement of hazmat railcars to reduce risk, and then develop an optimization program to solve the hazmat transportation problem faced by a railroad company.

Tank-Car Design

The tank car and railroad industries, along with the U.S. and Canadian governments, have conducted extensive research on tank car safety over the past three decades (Barkan 2008). This research, initiated in the wake of a series of catastrophic hazmat accidents in the early 1970s, has led to a number of improvements in the design of tank cars to make them more resistant to damage if they are involved in an accident. The first step in that direction was the formation of the Railroad Tank Car Safety Research and Test Project in 1970, which was a cooperative research program—between the Railway Progress Institute and the Association of American Railroads—on the causes of tank car failures.

Although subsequent analysis of failure modes revealed the need for several tank car design changes, perhaps far more important was the recognition of the potential value of detailed statistical understanding of tank car failure modes in accidents. Consequently, the project instituted a long-term data collection effort on tank cars involved in accidents—and has provided extensive information about component performance and conditions of the accident. Although we list just some of the notable academic works in part responsible for the current understanding of tank car safety and design, we invite the reader to consult Barkan (2008) for a review of all relevant works.

By studying the risks associated with non-pressurized materials, Raj and Pritchard (2000) reported that DOT-105 tank car design constitutes a safer option than DOT-111. On the other hand, Barkan et al. (2000) showed that tank cars equipped with surge pressure reduction devices experienced lower release rates than those without this technology. Finally, Saat and Barkan (2005) developed a metric to assess the performance of a tank car in an accident, and subsequently analyzed the trade-off between increased damage resistance and greater exposure to accidents (Barkan et al. 2007).

Placement of Hazmat Railcars

A sub-stream within risk mitigation focused on reducing the probability that a hazmat railcar gets involved in a train derailment. To that end, the early work of Fang and Reed (1979) suggested that the front of the train is more prone to derailment under loaded conditions, and hence hazmat railcars should be placed in the rear of the train. A later study commissioned by the Federal Railroad Administration and the U.S. Department of Transportation concluded that derailment probabilities are highest in the first and lowest in the fourth quarter of the train (Thompson et al. 1992). The same report also explored the possibility of implementing commodity-based blocking, as opposed to the destination-based blocking practice of the railroad industry (as explained in section "Rail Transportation System"), but ruled it out due to time and hence cost considerations.

The destination-based blocking approach was also visited in the study by Woodword, who suggested that separating hazmat railcars in a train decreases the probability of multiple railcars being derailed for small accidents involving relatively few railcars (Woodword 1989). On the other hand, the issue of additional handling and the resulting increase in time and cost was recently visited by Bagheri (2009), who subsequently outlined a placement strategy that could be implemented during the railcar blocking process that in turn will minimize hazmat risk on pre-defined transportation corridors (Bagheri et al. 2011). The latter was discussed in section "Incident Probability," where we voiced the need to conduct additional investigation to ascertain the impact of the incident probability model using multiple yards and train services. Doing so, we believe, will provide further insights into the feasibility of implementing a similar plan at different marshalling yards, and perhaps yield more generalizable results that are not specific to a particular transportation corridor.

The issue of hazmat railcar placement was also touched upon in Verma (2011). To that end, approximately 25,000 derailment instances were analyzed by grouping the derailment records into train-deciles. As indicated earlier, a decile-based approach may result in better analysis than a quarter-based approach, but only if train-lengths are similar (or constant). Unfortunately there was no-peer reviewed work categorizing freight-trains on length, and hence both Bagheri (2009) and Verma (2011) suggested that freight-trains with up to 40 railcars be called *short*; between

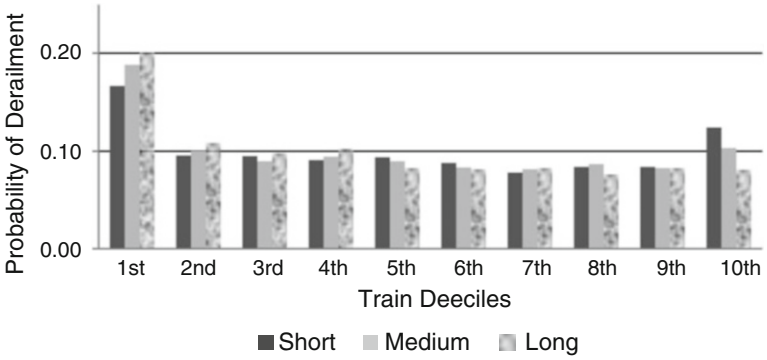


Fig. 9 Decile-based derailment probabilities (*Source*: Verma 2011)

41 and 120, *medium*; and, the rest *long*. Based on this categorization, of the 25,000 accident instances in the database, 93% resulted from freight-trains with less than 121 railcars, wherein *medium* trains accounted for 56%.

Figure 9 depicts the decile-based derailment probabilities for the three categories of trains. Given the level of resolution, two important observations were made for *all* train-lengths: *first*, the first decile has the highest derailment probability; and *second*, the front-half of the train is riskier. It is possible to conclude that the seventh decile is the safest for placing hazmat railcars if the train length is less than 121, and the eighth decile for longer train-lengths. This implies that decile-based derailment probabilities for *short* and *medium* trains are very similar, but marginally different than those for *long* trains while the critical length is 120 railcars. In fact, this analysis motivated the development of the risk assessment framework that took into consideration the position of hazmat railcar in the train-consist (i.e., expected consequence framework in section "Expected Consequence"). Subsequently, the risk assessment methodology was used to analyze a problem instance involving rail transportation of hazmat in the eastern United States, which in turn led to the following conclusions. *First*, transport risk is a function of train length, train-decile position of the hazmat railcar, and the number of intermediate handling. *Second*, the front of the train is riskier, and that seventh to ninth train-deciles are most appropriate for moving hazmat railcars for freight-trains of any length. *Finally*, rail-track risk can be reduced by strategically distributing hazmat railcars in the train-consist, although this will require increased handling at the marshalling yards. Such increased handling at the marshalling yards not only increase cost but also the probability of hazmat release. Verma (2009) proposed a bi-objective optimization model to solve a railroad transportation problem involving hazmat and regular freight, and pointed out that blocking paths with minimal intermediate handlings were preferred irrespective of whether the cost or the risk objective was emphasized. We provide more details on this work in the next section.

Network Routing

Although network routing and scheduling issues are very important to freight train operations, Glickman (1983), as far back as the early 1980s, had pointed out that railroads did not consider exposure risk when dealing with hazmat shipments. He argued that by considering population exposure, the railroads could diminish the likelihood of impacting the individuals working or living in the vicinity of rail tracks, which could be further reduced by upgrading rail tracks—a measure which would enhance railroad safety and performance. Thus, he argued that re-routing (with or without track upgrades) can improve hazmat safety. The issue of re-routing was revisited following the terrorist attacks on 9/11, when railroads in North America came under increasing pressure to avoid transporting certain hazmat through (or close to) large population centers. Consequently, the railroad industry and the government agencies conducted a network wide assessment to identify risk, and update emergency preparedness, but neither the study nor the results are publicly available (Plant 2004; Citizen's for Rail Safety 2007). In an effort to evaluate transport risk of different routes, in recent years, the issue of re-routing and appropriate tank car design was investigated by Saat and Barkan (2006), while Glickman et al. (2007) combined transportation modeling with risk assessment to study a sample of intercity routes and identify opportunities for improving safety at a reasonable cost. Both (Saat and Barkan 2006) and (Glickman et al. 2007) looked at a sample of transportation corridors for insights, and hence their results could not be generalized.

Given the involvement of various stakeholders, it should be evident that hazmat transportation problems should have multiple objectives—and hence a set of non-dominated solutions (or *Pareto*) should be determined. A Pareto-optimal solution is one in which one cannot improve on one objective without worsening at least one other objective. As indicated earlier, although routing and scheduling of regular freight is well studied in the railroad transportation domain (Cordeau et al. 1998), multi-objective routing of hazmat shipments is not. To the best of our knowledge, Verma et al. (2011) and Verma (2009) are the only refereed publications dealing with risk and cost objectives to route hazmat shipments, and form the basis of this section.

Figure 10 depicts the railroad infrastructure—comprised of rail-yards and track-sections—in south-east US. The solid-square nodes indicate fully equipped yards, i.e., classification and transfer operations are possible, while hollow-circular nodes can just perform the block-swap (transfer) operation. As outlined in section "Rail Transportation System" (using Figs. 1 and 2), any two nodes are connected by tracks, which are the service-legs of a train travelling non-stop between them. A sequence of service-legs and intermediate yards constitutes an itinerary available to a railcar for its journey. The managerial problem can be to determine the best routing plan for railcars, with hazmat and regular freight, and the number of trains of each type required to meet the given set of demand.

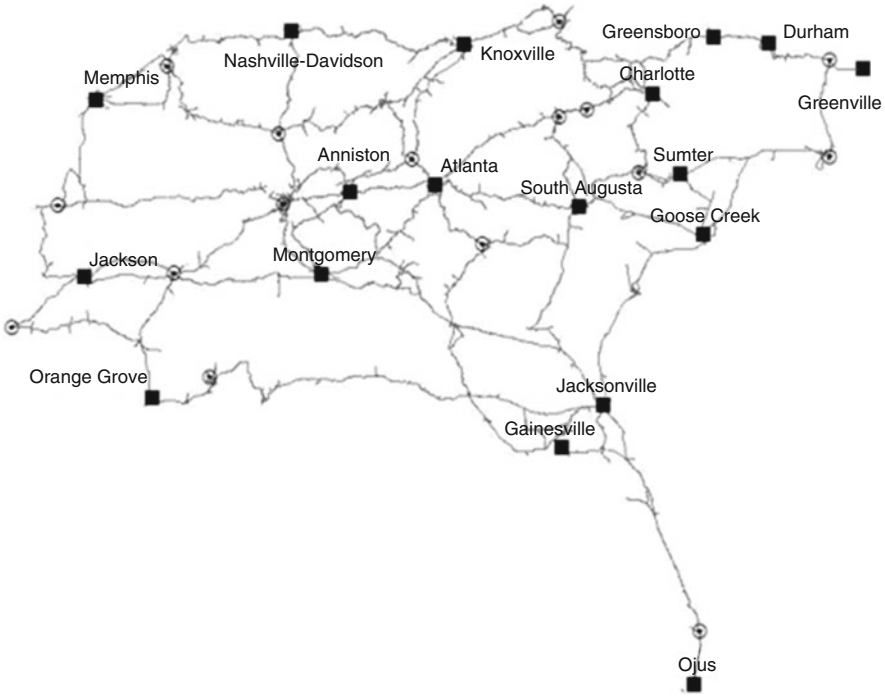


Fig. 10 Freight railroad network (*Source*: Verma 2009)

Both (Verma 2009) and (Verma et al. 2011) have proposed a bi-objective optimization model that would address the interests of two stakeholders, i.e., the regulatory agencies and the railroad companies. In assessing transport risk, which is the primary concern for the regulator, Verma (2009) makes use of the expected consequence approach while Verma et al. (2011) employ population exposure as the measure of risk—i.e., the assessment methodologies outlined in section "Expected Consequence" and "Population Exposure," respectively. On the other hand, the perspective of the railroad company is taken into consideration via transport cost. Since both models focus on the tactical planning problem of a railroad company that regularly transports a predetermined amount of mixed freight (i.e., hazmat and regular) across a railroad network, certain pertinent assumptions have been made. *First*, tactical planning is conducted on a weekly basis, and hence demand is expressed in terms of the number of railcars (hazmat and regular) to be shipped per week. *Second*, operational level details such as congestion and connections between train services are ignored. This amounts to assuming that all railcars to be moved are available at the pertinent locations on time. *Finally*, the hazmat being shipped on the train possesses similar chemical properties and the undesirable consequences of their interactions in case of an accident can be ignored. This assumption is important

since, as indicated earlier, there are no peer-reviewed publications demonstrating the interaction effects among all types of hazmat shipments.

Since expected consequence is the most popular measure of transport risk, we will first develop the optimization framework as proposed in (Verma 2009), and then contrast it with the one proposed in (Verma et al. 2011). We introduce the sets and notations, and then outline the model.

Sets and indices	
M	Set of hazmat to be moved in the network, indexed by m
I	Set of origin yards, indexed by i
J	Set of destination yards, indexed by j
C	Set of classification yards in the network, indexed by c
T	Set of transfer yards in the network, indexed by t
Z	Set of train services in the network, indexed by z
K_{ij}	Set of itineraries connecting yards i and j, and indexed by k
K_z	Set of itineraries using train service z
K_c	Set of itineraries using classification yard c
K_t	Set of itineraries using transfer yard t
Decision variables	
$H_{ij}^{k,m}$	Number of railcars with hazmat m using itinerary k to travel between yard i–j
R_{ij}^{k}	Number of railcars with regular freight using itinerary k to travel between yard i–j
N^z	Number of trains of type z needed in the network
Parameters	
$ERisk_{ij}^{k,m}$	Expected risk due to a railcar with hazmat m using itinerary k to travel between yard i–j
$C_{ij}^{k,m}$	Cost of moving a railcar with hazmat m using itinerary k between yard i–j
C_{ij}^{k}	Cost of moving a railcar with regular freight using itinerary k between yard i–j
FC^z	Fixed cost to operate train service of type z
h_{ij}^{m}	Number of railcars with hazmat m demanded at yard j from yard i
r_{ij}	Number of railcars with regular freight demanded at yard j from yard i
U^z	Capacity of train service of type z
U_c	Capacity of classification yard c
U_t	Capacity of transfer yard t

Minimize

$$\sum_{m,i,j,k} ERisk_{ij}^{k,m} H_{ij}^{k,m}$$

$$\sum_{m,i,j,k} C_{ij}^{k,m} H_{ij}^{k,m} + \sum_{i,j,k} C_{ij}^{k} R_{ij}^{k} + \sum_{z} FC^{z} N^{z} \tag{25}$$

Subject to:

$$\sum_{k} H_{ij}^{k,m} = h_{ij}^{m} \qquad \forall i, j, m \tag{26}$$

$$\sum_{k} R_{ij}^{k} = r_{ij} \qquad \forall i, j \tag{27}$$

$$\sum_{i,j,k \in K_{ij} \cap K_z} \left(\sum_{m} H_{ij}^{k,m} + R_{ij}^{k} \right) \leq U^{z} N^{z} \qquad \forall z \tag{28}$$

$$\sum_{i,j,k \in K_{ij} \cap K_c} \left(\sum_{m} H_{ij}^{k,m} + R_{ij}^{k} \right) \leq U_c \qquad \forall c \tag{29}$$

$$\sum_{i,j,k \in K_{ij} \cap K_t} \left(\sum_{m} H_{ij}^{k,m} + R_{ij}^{k} \right) \leq U_t \qquad \forall t \tag{30}$$

$$H_{ij}^{k,m} \geq 0 \ \ integer \tag{31}$$

$$R_{ij}^{k} \geq 0 \quad integer \tag{32}$$

$$N^{z} \geq 0 \quad integer \tag{33}$$

The first objective in (25) contains the transport risk stemming from routing hazmat railcars. The second objective involves cost to transport railcars, and the fixed cost to operate each type of train service. Constraints (26) and (27) ensure that weekly demand for both hazardous and regular freight are met. Constraint (28) states that the frequency of each train type is determined by the number of railcars, moving on different itineraries between origin and destination yards, but using that particular train service. Constraints (29) and (30) specify that classification and transfer operations at yards cannot exceed the corresponding capacities. Finally, constraints (31)–(33) specify sign restriction on the variables.

The transport cost parameters can be estimated using either publicly available information or some of the recent works, and (Verma 2009) assumed $0.50 to move a railcar one mile and $50 per intermediate handling. The fixed cost of a train is based on the number of hours it would take to provide a service, and an hourly rate of $500 was assumed. This amount included the hourly employment rate for a driver, an engineer, a brakeman, and an engine at $100, $100, $100 and $200,

respectively. Finally, it was assumed that the average speed of freight trains in the U.S. is around 22 miles per hour (Railroad Performance 2008).

For transport risk, as indicated above, the expected consequence approach outlined in section "Expected Consequence" was used. For instance, risk associated with each possible itinerary for a hazmat railcar was determined by: first identifying the train services and intermediate yards traversed; and then, incorporating the appropriate derailment and conditional probabilities, together with the exposure for different tracks and yards, in (1) and (2). The resulting number yielded the expected risk from a single hazmat railcar for a specific itinerary, with the process being repeated for other hazmat railcars, and for all available itineraries. It should be evident from the risk objective in (25) that transport risk is linear in the number of hazmat railcars, and that all hazmat possess similar chemical properties. For example, aggregate transport risk from five hazmat railcars is exactly 5-times that from a single hazmat railcar. Although we invite the reader to consult (Verma 2009) for additional details, we note that the indicated "linearity" was not observed when *population exposure* was used as the measure for transport risk in (Verma et al. 2011).

One of the major differences between (Verma 2009) and (Verma et al. 2011) was the expression for transport risk. Recall that the aggregate contaminant level at any point in the network can be determined using (23), which in turn exhibited non-linear curves. This is important since it implies that the population exposure for a particular service leg (or at a yard) of a freight train is a function of the total number of hazmat railcars involved, which is not known *a priori*. To make this more explicit, consider that n hazmat railcars are using service leg s of train type z. By making use of GPM and the methodology developed in Verma and Verter (2007), and outlined in section "Population Exposure," the aggregate concentrate level at downwind distance x can be determined using either (23) or (24). Further, assume that the IDLH level of the most lethal hazmat being transported is $\tilde{C}$, then the evacuation distance can be determined using (34).

$$\tilde{x} = \sqrt[b+d]{\frac{n \times Q}{\pi u a c \tilde{C}}} \times n \tag{34}$$

The movement of a danger circle, of radius $\tilde{x}$, along a rail link will carve out a band, and the number of people within the band is the resulting population exposure. For example, transport risk because of hazmat release from $Y_{z,s}$ on service leg s of train type z can be calculated by:

$$PE\left(Y_{z,s}\right) = \tilde{x}\left(Y_{z,s}\right) \times \textit{length of service leg } s \times \rho(\tilde{x}\left(Y_{z,s}\right)) \tag{35}$$

where ρ is the population density of the center exposed because of the transportation of hazmat on service leg s. The population center exposed depends on the threshold distance, which, in turn, depends on the hazmat volume being transported on a particular service leg. It should be clear from (35) that the function for calculating pop-

ulation exposure is non-linear with a rather complicated form, and without a closed-form expression. This implies that the population exposure risk objective will be without a closed form expression, since there is no a priori information on the number of hazmat railcars on different trains and at various yards in the network. Hence, the following expression will replace the one for expected consequence in (25).

$$Exposure\ Risk = \sum_z \sum_s PE\ (Y_{z,s}) + \sum_y PE\ \left(Y_y\right) \tag{36}$$

where, the first term refers to the cumulative population exposure resulting from all service-legs of different freight-trains, while the second will capture the aggregate exposure at various yards in the network. Next, we briefly outline the solution techniques and then some managerial insights that could be expected for rail hazmat transportation.

Solution Method

Two of the most common techniques for solving multi-objective models are pre-emptive optimization and weighted sums (Rardin 1998). The former calls for a sequential process, and makes use of the actual known preferences of the stakeholders. Meanwhile, the latter attaches weights to different objectives, and hence is preferred if stakeholder preferences are unknown. Since one is dealing with competing objectives, a singular solution may not be enough—and hence a number of scenarios should be solved. Furthermore, it is also important to use a surrogate measure, if one objective consistently dominates the other. For example, in (Verma 2009) the expected consequence values were consistently dominated by the cost numbers, which was skewing the results. Consequently, in addition to the instances with expected consequence as the measure of transport risk, equivalent problem instances with population exposure as the measure of risk were also solved.

If the network routing formulation can be expressed in closed form [such as in (Verma 2009)], then any optimization package can be used to solve the problem instances. On the other hand, in the absence of closed-form expression [such as (Verma et al. 2011)], solving realistic-size problem instances through the use of general-purpose optimization software would be rather inefficient. This is because of the difficulties associated with *a priori* determination of transport risk, as explained in relation to (35) and (36). However, Verma et al. (2011) contended that since such problems, typically, would contain a huge number of variables and relatively fewer constraints, a genetic algorithm (GA)-based solution methodology would be more effective and efficient (Holland 1975). They replaced the traditional mutation operator in GA with a local search heuristic, which was intentioned to ensure a more effective neighbourhood search (i.e., intensification). Consequently, their solution methodology is a Memetic Algorithm that combines global and local searches (Moscato 1989). We invite the reader to refer to (Verma et al. 2011) for all pertinent technical and methodological details.

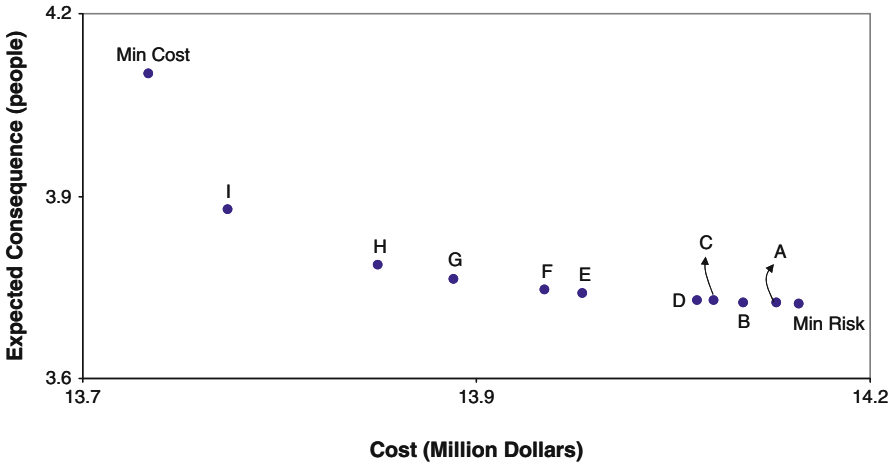


Fig. 11 Cost v/s Expected Consequence (*Source*: Verma 2009)

Insights

Since the hazmat risk and transport cost may not be equally important to the two stakeholders, a number of scenarios should be analyzed. For example, a total of 11 different instances of the given managerial problem were solved in (Verma 2009), with *min cost* and *min risk* solutions at the two extremes, and nine others resulting from attaching complementary weights on the coefficients of the two objectives. It is important to note that the above eleven non-dominated solutions constitute just a portion of the Pareto frontier, which ideally should list all possible non-dominated solutions. In an effort to further explain the utility of such a curve, we paraphrase the pertinent discussions from (Verma 2009) and some highlights from (Verma et al. 2011).

Each point in Fig. 11 represents a non-dominated solution, with *min cost* solution the most risky while the *min risk* solution the most expensive. Points *A* through *I* are the intermediate solutions arrived at by attaching different weights to the two coefficients. *A*, with 10% weight on the cost coefficient and 90% on the risk coefficient, is closest to the *min risk* solution. With higher emphasis on cost (and corresponding reduction in risk), the resulting solutions started moving towards the *min cost* point on the partial frontier, with *I* being the closest. It is important to note that solution *E* results when both objectives receive equal importance, and hence, ideally, should not be dominated by either risk or cost.

For the indicated problem instance, the *min cost* solution had a cost of just under \$13.7 million and an expected consequence of 4.1008 people, whereas the min risk solution had a cost of just over \$14.1 million and an expected consequence of 3.7231 people. This is an excellent instance of a problem in which the difference in magnitude between the objectives is masking the significance of the results, and should warrant re-running the scenarios with a surrogate for expected consequence

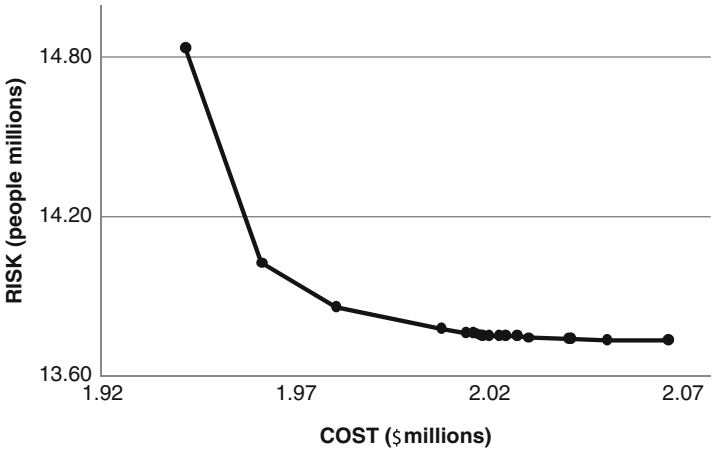


Fig. 12 Partial Pareto-Frontier (*Source*: Verma et al. 2011)

(such as population exposure). Doing so reveals that population exposure risk can be reduced by 2.6 million people by spending an extra $4.12 million, which may be a worthwhile trade-off for the regulators to pursue. Perhaps a more important observation via Fig. 11 is the significant increase in population exposure risk (or expected consequence) when the weight attached to the risk coefficient is decreased from 10 to 0% (i.e., from *I* to *Min Cost*). This weight allocation results in a saving of just under $50 thousand but increases exposure risk by 1.5 million individuals, which implies that every saved dollar exposes 30,000 additional individuals to hazmat risk.

Finally, as alluded to earlier, one should endeavour to generate the complete frontier of non-dominated solutions, but doing so is extremely cumbersome and computationally impossible in some instances. For example, Verma et al. (2011) generated over 7,500 different solutions of which only 56 were non-dominated (Fig. 12). In any event, because quantification of risk is one of the most challenging and contentious issues in hazmat transportation, such a (partial) frontier could be used by the primary stakeholders to conduct judicious evaluation of the monetary and societal implications of hazmat transportation.

To conclude, in this section, we have summarized various initiatives undertaken by stakeholders to devise techniques to manage and/or mitigate hazmat risk from railroad shipments. Design of tank-car, and the investigation of all relevant elements, has been a very popular research area—that has received significant support and attention from the railroad industry. On the other hand, the few works dealing with placement of hazmat railcars and train-makeup resulted primarily from analysis of empirical data, wherein the insights have been used primarily for academic purposes. Finally, tactical planning problems involving hazmat is a rather nascent area with potential for more research.

Potential Research Questions

We believe that the risk assessment methodologies outlined in chapter are sophisticated enough to capture the characteristics of railroad transportation of hazmat, although there is tremendous untapped research potential in the risk management domain. We make the case for the latter using three headings: blocking & train make-up; routing & scheduling; and, security aspects.

Blocking and Train Make-Up

As indicated in section "Rail Transportation System," *blocking* is the major determinant that enables railroads to realize economies of scale. This is done to prevent railcars from being handled at every yard on its journey, which in turn not only saves money but also reduces delay. Note that a block is associated with an OD pair, and railcars placed in this block are not reclassified till the block reaches the destination yard. In other words, railroads practice destination-based blocking. Interestingly, the movement of hazmat railcars through the marshalling yards has not received much attention, perhaps due to the perception that yards do not pose significant risk. This is in contrast to the empirical data presented in the literature. For example, a Transportation Safety Board study, involving accident records from 1996 to 2000, revealed that roughly 45% of railroad accidents in Canada occurred in the marshalling yards (Transportation Safety Board: Durham Regional 2008). This is supported by the empirical analysis of accident rates from 1994 to 2009, where the number of accident per million miles was significantly higher for switching yards than for all types of tracks (Verma 2011).

Although Thompson et al. (1992) ruled out the possibility of implementing commodity-based blocking due to time and cost considerations, there is a need for developing analytical approaches to solve the blocking-problem that incorporate not just cost but also the potential risk from hazmat shipments. Such a framework will be tremendously useful in not only facilitating additional insights into the workings of Bagheri (2009) by looking at multiple yards and train services, but also developing a network-wide blocking and train make-up model that would be driven by both transport-cost and hazmat risk.

Routing and Scheduling

It may have been evident that the routing models developed in (Verma 2009) and (Verma et al. 2011) would select routes based on the objective being emphasized. For example, if transport risk is more important, then routes closer to the minimum risk solution would be selected. Such a solution may not be acceptable to the resi-

dents alongside the chosen routes, since they would be subject to frequent exposure and hence hazmat risk, whereas there would be a zero chance of any incident on the other routes. This so called in-equitability can be addressed only by the involvement of regulators (Erkut et al. 2007), possibly by adapting methodologies introduced in the context of highway transportation (Gopalan et al. 1990a, b; Lindner-Dutton et al. 1991) for railroad transportation of hazmat. Developing such methodologies is challenging since railroads do not present as many routing options as highways, and also because hazmat railcars—unlike truck tankers—cannot travel by themselves. In any event, there is a room and potential need for developing analytical techniques that could yield more equitable solutions.

Note that the attributes of rail-links such as travel time, incident probabilities and population exposure could vary with time, and hence proper scheduling of trains could potentially reduce hazmat risk. For example, residential areas are less populated during the day, and some rail links are riskier to travel during night and in winter. The issue of scheduling of hazmat trucks has been studied by Nozick et al. (1997) and Miller-Hooks and Mahmassani (1998), and there is a need for similar engagement in the railroad domain. More specifically, scheduling of hazmat unit-trains should be investigated such that both the spatial and temporal link attributes could be exploited to mitigate hazmat transport risk.

Security Aspects

It should be noted that pre-9/11, the guiding policy of the federal program and the railroad industry in the United States was "safety," wherein the mission was to prevent fatalities, injuries, and property damage related to railroad operations and releases of hazmat from railcars. But post-9/11, the two agencies realized the "security" dimension of hazmat railcars, i.e., a terrorist could either target them or use them as a potential weapon, and inflict spectacular damage on the population, environment and the infrastructure (Milazzo et al. 2009). Conceivably, every yard and each track section could be a possible entry point for an individual with evil intentions, and hence protecting the railroad infrastructure would be daunting. This is not only because of the geographic dispersion of the railroad infrastructure, but also due to the nature of thinking required to assess the impact of newer aspects of rail operations (such as intermodalism, just-in-time deliveries), and the absence of enough empirical data to understand the mindset of terrorists (Verma and Verter 2011).

Although absence of enough empirical dataset makes tackling such problems extremely difficult, one could conceivably work backwards from the end objective of any terrorist attack, viz. catastrophic consequence. Translated to a threat involving railroad transportation of hazmat, this would seem to be equivalent to using one or more hazmat railcars as a weapon (or a target) with devastating results, near or close to a major urban center(s). For instance, every year over eight thousand tank cars of chlorine move by rail within two blocks of the U.S. capital, and successful

targeting of even one of these could potentially kill or injure hundreds of thousands. Similarly, an attacker who gains control of a hazmat tank car—in a rail-yard or in geographically dispersed and often lightly guarded industrial sidings—could use it as a weapon.

The vast railroad infrastructure and the existence of multiple railroad companies, each with its own network, calls for a co-operative development of a structured and systematic methodology to manage terrorism risk. Although it was reported that the railroad industry and the government agencies conducted a network wide assessment to identify risk, and update emergency preparedness, neither the study nor the results are publicly available (Plant 2004; Citizen's for Rail Safety 2007).

This is a rather open area of research, which does present the potential for more academic and cutting edge research within the "security" framework. *First*, there is no established assessment methodology that can capture the mindset of terrorist and/or be developed in the absence of empirical dataset (or demonstrated risk profile). This implies that even basic risk assessment would be extremely challenging, and perhaps contentious. *Second*, tackling such complex problems would necessitate a truly multidisciplinary approach. For example, one needs to draw upon game theory to both anticipate possible terrorist moves and identify potentially vulnerable sites in a railroad network, and recommend ways to mitigate and counter those risks. In the interest of space, we have tried to just sketch the possibilities in this area, and invite the reader to consult Verma and Verter (2011) for a comprehensive discussion of this topic.

To conclude, we have attempted to outline possible research areas within railroad transportation of hazmat. It is our belief that appropriate, and at times further, investigation of some or all of the indicated topics would help better manage or mitigate hazmat transport risk from railroad shipments. Clearly the listed areas (topics) are by no means exhaustive, but do reflect the research interests of the co-authors.

Summary

The railroad industry, crucial to the United States and Canadian economies, has long considered itself to be the safest and the most secure mode of transportation. In the United States, railroads move around 1.8 million hazmat carloads annually, and this number is expected to increase significantly over the next decade—in part driven by the phenomenal growth of intermodal transportation and the growing use of rail-truck combinations for moving chemicals. Fortunately a host of industry initiatives, such as the formation of an inter-industry task force in the 1970s, and the emphasis on reducing the frequency of tank car accidents, as well as the likelihood of a release, are collectively responsible for making railroads one of the safest modes for transporting hazmat. In spite of the favourable safety statistic of railroads, the possibility of spectacular events resulting from multi railcar incidents, however small, do exist.

This book chapter was a first attempt to introduce and discuss different models developed for both assessment and management of hazmat risk for railroad transport. In an effort to introduce the reader to the pertinent building blocks, the chapter started with a description of a typical freight rail operation. Three of the most popular measures for hazmat transport risk, i.e., *expected consequence*, *incident probability*, and *population exposure*—originally developed for highway and very recently adapted for railroads—were discussed extensively. Risk management and/or mitigation strategies were discussed under tank-car design, placement of hazmat railcars in a train, and network routing. Finally, issues necessitating further investigation were outlined.

Acknowledgements This research has been supported in part by two grants from the National Sciences and Engineering Research Council of Canada (Grants # 312936 and #183631). Both authors are member of the Interuniversity Research Center on Enterprise, Network Logistics and Transportation (CIRRELT) and acknowledge the research infrastructure provided by the Center. Also the comments and suggestions of an referee helped improve the paper significantly.

References

AAR (2006) Association of American railroads, current rail hazmat conditions called "Untenable," AAR News, June. http://www.aar.org/ViewContent.asp?Content_ID=3763. Accessed July 2011

Abkowitz M, Lepofsky M, Cheng P (1992) Selecting criteria for designating hazardous materials highway routes. Transp Res Rec 1333:30–35

Anderson RT, Barkan CPL (2004) Railroad accident rates for use in transportation risk analysis. Transp Res Rec 1863:88–98

Arya SP (1999) Air pollution meteorology and dispersion. Oxford University Press, Cambridge

Bagheri M (2009) Risk-based model for effective marshaling of dangerous goods railways cars. PhD Dissertation, University of Waterloo, Canada

Bagheri M, Saccomanno FF, Chenouri S, Fu L (2011) Reducing the threat of in-transit derailments involving dangerous goods through effective placement along the train consist. J Accid Anal Prev 43(3):613–620

Bagheri M, Verma M, Verter V (2012) An expected risk model for rail transport of hazardous materials. In: Garbolino E, Tkiouat M, Yankevich N, Lachtar D (eds) Transport of dangerous goods, NATO Science for Peace and Security Series C: Environmental Security. Springer, Dordrecht, pp 207–226

Barkan CPL (2008) Improving the design of higher-capacity railway tank cars for hazardous materials transport: optimizing the trade-off between weight and safety. J Hazard Mater 160:122–134

Barkan CPL (2004) Personal communication. February

Barkan CPL, Treichel TT, Widell GW (2000) Reducing hazardous materials releases from railroad tank car safety vents. Transp Res Rec 1707:27–34

Barkan CPL, Tyler Dick C, Anderson R (2003) Railroad derailment factors affecting hazardous materials transportation risk. Transp Res Rec 1825:64–74

Barkan CPL, Ukkusuri S, Waller ST (2007) Optimizing railroad tank cars for safety: the tradeoff between damage resistance and probability of accident involvement. Comput Oper Res 34:1266–1286

Barnhart C, Jin H, Vance PH (2000) Railroad blocking: a network design application. Oper Res 48:603–614

Batta R, Chiu SS (1988) Optimal obnoxious paths on a network: transportation of hazardous materials. Oper Res 36(1):84–92
Citizen's for Rail Safety (2007) Securing and Protecting America's Railroad System: U.S. Railroad and Opportunities for Terrorist Threats: prepared by Plant JF and Young RR. The Pennsylvania State University, Harrisburg
Conlon PCL (1999) Rail transportation of hazardous materials in the United States. Rail International—English Edition. June, pp. 8–17
Conlon PCL (2004) Personal communication. February
Cordeau J-F, Toth P, Vigo D (1998) A survey of optimization models for train routing and scheduling. Transp Sci 32(4):380–404
ERG (2012) 2012 Emergency Response Guidebook. http://www.phmsa.dot.gov/hazmat/library/erg, Accessed on May 9, 2013
Erkut E, Verter V (1998) Modeling of transport risk for hazardous materials. Oper Res 46(5):625–642
Erkut E, Tjandra S, Verter V (2007) Hazardous materials transportation. In: Barnhart C, Laporte G (eds) Handbooks in operations research and management science: transportation. Elsevier, Amsterdam, pp 589–601
Fang P, Reed HD (1979) Strategic positioning of railroad cars to reduce their risk of derailment. Technical report, Volpe Transportation Systems Center (DOTITSC) 7:67
FRA (2010) Federal Railroad Administration, accident data on demand, Federal Railroad Administration Office of Safety Analysis. http://safetydata.fra.dot.gov/officeofsafety. Accessed 20 July 2010
Glickman TS (1983) Rerouting railroad shipments of hazardous materials to avoid populated areas. Accid Anal Prev 15(5):329–335
Glickman TS, Rosenfield DB (1984) Risks of catastrophic derailments involving the release of hazardous materials. Manag Sci 30(4):503–511
Glickman TS, Erkut E, Zschocke MS (2007) The cost and risk impacts of rerouting railroad shipments of hazardous materials. Accid Anal Prev 39:1015–1025
Gopalan R, Batta R, Karwan MH (1990a) The equity constrained shortest path problem. Comput Oper Res 17(3):297–307
Gopalan R, Kolluri KS, Batta R, Karwan MH (1990b) Modeling equity of risk in the transportation of hazardous materials. Oper Res 38(6):961–975
Holland JH (1975) Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. University of Michigan Press, Ann Arbor, MI
Lindner-Dutton L, Batta R, Karwan MH (1991) Equitable sequencing of a given set of hazardous materials shipments. Transp Sci 25:124–137
Milazzo MF, Ancione G, Lisi R, Vianello C, Maschio G (2009) Risk management of terrorist attacks in the transport of hazardous materials using dynamic geoevents. J Loss Prev Process Ind 22:625–633
Miller-Hooks ED, Mahmassani HS (1998) Optimal routing of hazardous materials in stochastic time-varying transportation networks. Transp Res Rec 1645:143–151
Moscato P (1989) On evolution, search, optimization, genetic algorithms and martial arts: towards memetic algorithms. Caltech Concurrent Comput. Program. Technical Report C3P826. California Institute of Technology, Pasadena, CA, p 158–179
Newton HN, Barnhart C, Vance PH (1998) Constructing blocking plan to minimize handling costs. Trans Sci 32:330–345
Nozick LK, List GF, Turnquist MA (1997) Integrated routing and scheduling in hazardous materials transportation. Transp Sci 31(3):200–215
Oggero A, Darbra RM, Munoz M, Planas E, Casal J (2006) A survey of accidents occurring during the transport of hazardous substances by road and rail. J Hazard Mater 133A:1–7
Pasquill F, Smith FB (1983) Atmospheric diffusion, 3rd edn. Ellis Horwood, Chichester, UK
Plant JF (2004) Terrorism and the Railroads: Redefining Security in the Wake of 9/11. Rev Policy Res 21(3):293–305

Railroad Performance Measures (2008) http://www.railroadpm.org. Accessed 18 June 2008
Railway Investigation Report (2002) Transportation Safety Board of Canada (TSB): report number R 99H0010, September
Raj PK, Pritchard EW (2000) Hazardous materials transportation on US railroads. Transp Res Rec 1707:22–26
Rardin RR (1998) Optimization in operations research. Prentice Hall, New Jersey
ReVelle C, Cohon J, Shobrys D (1991) Simultaneous siting and routing in the disposal of hazardous wastes. Transp Sci 25:138–145
Saat MR, Barkan CPL (2005) Release risk and optimization of railroad tank car safety design. Transp Res Rec 1916:78–87
Saat MR, Barkan CPL (2006) The effect of rerouting and tank car safety design on the risk of rail transport of hazardous materials. In: Proceedings of the 7th World Congress on Railway Research, Montreal
Saccomanno FF, Chan AYW (1985) Economic evaluation of routing strategies for hazardous road shipments. Transp Res Rec 1020:12–18
Saccomanno FF, El-Hage S (1989) Minimizing derailments of railway cars carrying dangerous commodities through effective marshaling strategies. Transp Res Rec 1245:34–51
Saccomanno FF, El-Hage S (1991) Establishing derailment profiles by position for corridor shipments of dangerous goods. Can J Civ Eng 18(1):67–75
Swoveland C (1987) Risk analysis of regulatory options for the transport of dangerous commodities by rail. Interfaces (Providence) 17(4):90–107
Thompson RE, Zamejc ER, Ahlbeck DR (1992) Hazardous materials car placement in a train consist. US Department of Transportation Report DOT/FRA/ORD-92/18.1
Transportation Safety Board: Durham Regional Municipality (2008) Hazard Analysis and Risk Assessment: Final Report. Prepared by Stevanato & Associates, and John Newton Associates
TSB (2004) Transportation Safety Board of Canada, statistical summary railway occurrences 2004. http://www.tsb.gc.ca/en/stats/rail/2004/statssummaryrail04.pdf. Accessed 18 August 2008
US Environmental Protection Agency, Technology transfer network support center for regulatory air models. http://www.epa.gov/scram001/u22.htm. Accessed 15 March 2011
Verma M (2005) Analytical approaches to railroad and rail-truck intermodal transportation of hazardous materials. PhD Dissertation, McGill University, Canada
Verma M (2009) A cost and expected consequence approach to planning and managing railroad transportation of hazardous materials. Transp Res D 14(5):300–308
Verma M (2011) Rail transportation of dangerous goods: a conditional exposure approach to minimize transport risk. Transp Res C 19(5):790–802
Verma M, Verter V (2007) Railroad transportation of dangerous goods: population exposure to airborne toxins. Comput Oper Res 34:1287–1303
Verma M, Verter V (2011) Security aspects of hazmat transportation using railroads. In: Reniers G, Zamparini L (eds) Security aspects of uni- and multi-modal hazmat transportation systems. Wiley-VCH, Weinheim
Verma M, Verter V, Gendreau M (2011) A tactical planning model for railroad transportation of dangerous goods. Transp Sci 45(2):163–174
Woodword JL (1989) Does separation of hazmat cars in a railroad train improve safety from derailments? J Loss Prev Process Ind 2:176–178

Operations Research Models for Global Route Planning in Hazardous Material Transportation

Lucio Bianco, Massimiliano Caramia, Stefano Giordani, and Veronica Piccialli

Routing Issues in Hazmat Transportation

The distribution of goods from some depots to a set of customers by using a set of vehicles has received a lot of attention due to its impact on the final cost of the goods. The effective management of those vehicles gives rise to a variety of problems known as vehicle routing and scheduling problems. From a general point of view, the Vehicle Routing Problem (VRP) is an operational problem that concerns the service, in a given time period, of a set of customers by a set of vehicles which are located in one or more depots and are operated by a set of drivers on an appropriate road network.

Typically, the VRP involves decisions on route planning (i.e., selecting paths between pairs of nodes of the transportation network) and vehicle routing (i.e., assigning customers to vehicles and finding the order in which clients are visited by each vehicle).

The VRP is one of the most important and studied combinatorial optimization problems with several direct real-world applications, mainly in the industrial distribution arena. It was firstly introduced by Dantzig and Ramser (1959) to describe a real-world application related to the delivery of gasoline to service stations. They propose the first mathematical programming formulation and algorithmic approach for the VRP. Successively, Clarke and Wright (1964) propose an effective greedy heuristic and improve the Dantzig-Ramser's method. In the successive decades,

L. Bianco • M. Caramia • S. Giordani
Dipartimento di Ingegneria dell'Impresa, Università degli Studi di Roma Tor Vergata, via del Politecnico 1, 00133 Roma, Italy
e-mail: bianco@dii.uniroma2.it; caramia@dii.uniroma2.it; giordani@dii.uniroma2.it

V. Piccialli (✉)
Dipartimento di Ingegneria Civile e Ingegneria Informatica, Università degli Studi di Roma Tor Vergata, via del Politecnico 1, 00133 Roma, Italy
e-mail: piccialli@disp.uniroma2.it

R. Batta and C. Kwon (eds.), *Handbook of OR/MS Models in Hazardous Materials Transportation*, International Series in Operations Research & Management Science 193, DOI 10.1007/978-1-4614-6794-6_3,

hundreds of models and algorithms were proposed for the optimal and approximate solution of the VRP (see e.g. Toth and Vigo 2002). Today, dozens of packages to solve the VRP are available on the market. This interest in VRP is motivated by both its practical relevance and its high solution difficulty.

The VRP has many variants based on the constraints and the nature of the commodities to be delivered. Referring in particular to the hazardous material (hazmat) transportation, the VRP presents extremely typical characteristics. In fact, while in common freight transportation the objective of the VRP is in general the minimization of the transportation cost, in hazmat transportation the objectives to optimize depend on the subjects involved in the decision process and may be quite different.

What distinguishes hazmat transportation from the more general freight transport issues is basically the risk associated with an accidental release of such materials during their transportation. Even if the number of accidents is very small compared to the number of shipments of such materials, both at world level and at national level, this chance imposes a particular attention to the safety management in order to reduce the occurrence of dangerous events. More in general, the potential dangers on both the population and the environment render people very sensitive to this kind of transport. For this reason, hazmat transportation has stimulated a relevant research investigation with particular emphasis on risk assessment and route planning (routing) for hazmat shipments.

Therefore, while in general freight transportation the route planning phase is simple and assumed to be done a priori by selecting the shortest (cheapest) path between network nodes with the consequence that in practice VRP studies concentrate only on the vehicle routing decisions, in the context of hazmat transportation route planning decisions are very important and cannot be ignored.

The goal of our contribution is to analyze hazmat route planning models on a road network, while the analysis of the different models developed to evaluate the risk induced by hazmat shipments on the population and the environment is out of the scope of this chapter of the handbook.

Hazmat route planning is one of the main issues in hazmat transportation and deals with the selection among the alternative paths (routes) between origin–destination pairs on a given road network where to route hazmat shipment orders. The choice of a path depends on the different objectives that the distinct actors involved in the decision process want to pursue.

In fact, we assume that, besides the carriers, there are regional and local government authorities that want to regulate the hazmat transportation by imposing restrictions on the amount of hazmat traffic over the network links.

From a carriers' perspective, shipment contracts can be considered individually and a route decision needs to be made for each shipment. That is, each carrier selects a route between a given origin–destination pair for a given hazmat, transport mode, and vehicle type, with the objective of minimizing the transportation costs. Thus, for each shipment order, this problem focuses on a single-commodity and a single origin–destination plan. We call this problem "local route planning problem."

The interested reader can find the most important references on this problem in the next section.

Local route planning suffers from some limitations. Indeed, since the plans of the carriers are typically made without taking into account the general context, it may happen that certain links of the transportation network tend to be overloaded with hazmat traffic. This may result in a considerable increase of the accident probability on such road links, leading to inequity on the spatial distribution of the risk.

A chance to overcome this difficulty is to consider a government authority charged with the management of all hazmat shipments within and through its jurisdiction area. Therefore, this authority, in contrast with the carrier, has to consider a global transportation problem.

Although the transportation industry has been deregulated in many countries, hazmat transportation usually remains part of the governments' mandate mainly due to the associated societal and environmental risks.

This aspect leads to a harder class of problems that involve multi-commodities and multiple origin–destination routing decisions. We refer to this class of problems as "global route planning problem."

The main concern for a government authority is to control the risk induced by hazmat transportation over the population and the environment. Besides the minimization of the total risk, the authority should also promote equity in the spatial distribution of risk. This becomes crucial when certain populated zones are exposed to an intolerable level of risk resulting from the carriers' routing decisions.

Therefore, in the global route planning problem, the main problem (from the authority's point of view) is to find minimum risk routes, while limiting and equitably spreading the risk in any zone where the transportation network is embedded.

As a matter of fact, risk equity has to take into account also whenever it is necessary to carry out several hazmat shipments from a given origin to a given destination. In this situation, the planning effort has to be devoted to distribute risk uniformly among the zones of the traversed region.

Since, typically, the government does not have the authority to impose specific routes to hazmat carriers, it can only mitigate hazmat transportation risk by means of policies regulating the use of roads for hazmat shipments. The scenarios are essentially two. In the first one the government has the right either to close certain road segments to hazmat vehicles or to limit the amount of hazmat traffic flow on those network links. In the second scenario, the government uses link tolls to deter the carriers from using certain road segments and induce them consequently to route the shipments on less populated (or risky) links of the network.

In the context of global route planning the first policy falls in the field of "hazmat transportation network design" while we refer to the second one as "toll setting policies."

In section "State of the Art" we describe the state of the art on hazmat route planning, and in sections "Hazmat Transportation Network Design Models" and "Toll Setting in Hazmat Transportation" we focus our attention on the main mathematical models developed for the hazmat transportation network design and toll setting policies, respectively.

State of the Art

Most of the prevailing studies on hazmat transportation focuses on two related problems: (1) assessment of the transport risk associated with a shipment, and (2) identifying the route that minimizes transport risk. This chapter is devoted to the second aspect. The reader can refer to the survey of Erkut et al. (2007) and other contributions published in this handbook for a review of papers on the first topic and in particular on the methods for measuring the risk on the links of a given transportation network.

Hazmat Local Route Planning Problem

Hazmat local route planning problems are typically modeled with shortest-path models where a certain risk measure is considered as arc impedance. These kind of problems have been addressed by many operations research scientists. Their works cover road, rail, maritime transportation, proposing deterministic, stochastic, time-dependent models, as well as single and multiple objective models. Even if hazmat local route planning is not the main focus of this paper, for the sake of completeness, in Table 1 we list most of the contributions on local route planning for hazmat transportation by road: the table reports the paper list appearing in the survey of Erkut et al. (2007), with the addition of more recent contributions. For a detailed review of these and other works considering other transport mode, time-dependent models, or other specific issues such as integration of location and routing decisions, integration of routing and scheduling decisions, the reader is referred to the survey of Erkut et al. (2007).

Nonetheless the integration of routing and scheduling decisions in hazmat transportation merits to be recalled here with some details. Indeed, recent studies claim that risk attributes, like the population exposure and the accident probability, may vary considerably during the day. For instance, it has been observed that the accident probabilities are higher at night. Moreover, the expected population being exposed to the risk also varies with time since the population density in the urban areas changes with the daily activities and the associated mobility of the residents (Erkut and Alp 2007b).

Typically the modeling framework adopted to appropriately represent this phenomenon is a stochastic, time-varying network, where the link attributes (e.g., travel times, incident probabilities, and population exposure) are represented as random variables with probability distributions that vary with time. Almost all the contributions in hazmat transportation consider a priori optimization (i.e., the optimal routes are chosen before the travel begins), while there is a lack of papers considering both adaptive routing decisions and data updates based on real-time information.

Table 1 Contributions on local route planning for hazmat transportation by road

Road hazmat local route planning
Akgün et al. (2007), Duque et al. (2007), Huang (2006)[C,G,M], Erkut and Ingolfsson (2005), Huang and Cheu (2004)[C,G], Kara et al. (2003), Luedtke and White (2002)[C,U], Marianov et al. (2002), Erkut and Ingolfsson (2000), Frank et al. (2000), Leonelli et al. (2000), Zografos et al. (2000)[DSS], Tayi et al. (1999)[M], Bonvicini et al. (1998), Erkut and Verter (1998), Marianov and ReVelle (1998)[M], Erkut and Glickman (1997), Jin and Batta (1997), Nembhard and White (1997)[M], Sherali et al. (1997)[M], Verter and Erkut (1997), Ashtakala and Eno (1996)[S], Erkut (1996), Jin et al. (1996), Beroggi and Wallace (1995), Boffey and Karkazis (1995), Erkut (1995), Glickman and Sontag (1995)[M], Karkazis and Boffey (1995), McCord and Leu (1995)[M], Moore et al. (1995)[G], Sivakumar et al. (1995), Beroggi (1994), Beroggi and Wallace (1994), Ferrada and Michelhaugh (1994), Patel and Horowitz (1994)[G], Lepofsky et al. (1993)[G], Lassarre et al. (1993)[G], Sivakumar et al. (1993), Turnquist (1993)[M,S], Wijeratne et al. (1993)[M], Beroggi and Wallace (1991), Miaou and Chin (1991), Gopalan et al. (1990b), Chin and Cheng (1989)[M], Zografos and Davis (1989)[M], Abkowitz and Cheng (1988)[M], Batta and Chiu (1988), Vansteen (1987), Cox and Turnquist (1986), Belardo et al. (1985), Saccomanno and Chan (1985), Urbanek and Barber (1980), Kalelkar and Brinks (1978)[M]

[C]With security consideration
[DSS]Decision Support System model
[G]Using GIS
[M]Multi-objective
[S]Stochastic
[U]Survey/Annotated Bibliography

The combined routing and scheduling problem for hazmat shipments has been considered for the first time by Cox and Turnquist (1986): they propose a model that considers vehicle departure times in order to minimize the delay due to possible time-dependent restrictions over the links of the network. Other early contributions are proposed by Bowler and Mahmassani (1998), Miller-Hooks and Mahmassani (1998, 2000), and Miller-Hooks (2001). Assuming time-dependent link travel times, Nozick et al. (1997) formulate the hazmat transportation integrated routing and scheduling problem as a multi-objective shortest path problem in a network with time-dependent cost attributes on the network links: they propose a label-setting algorithm that heuristically spans the set of non-dominated (efficient) solutions; some additional heuristics, that can be used to improve the solution generated or reduce the computational burden, are proposed by Sulijoadikusumo and Nozick (1998). Chang et al. (2005) describe a method for finding non-dominated paths when the link attributes are uncertain, and the probability distributions that describe those attributes vary with the time of the day; therefore, their approach can be applied to hazmat transportation. Meng et al. (2005) propose a dynamic programming method for identifying non-dominated time-varying paths with fixed departure times at the origin and fixed waiting times at intermediate nodes of the paths between the origin and the destination. Erkut and Alp (2007b) extend the study on the integrated routing and scheduling problem by assuming en-route stops. They formulate the problem as a constrained shortest path problem in a network with time-dependent

link travel time and risk values and propose a dynamic programming based solution algorithm. The objective of their formulation is to determine the minimum risk path with the constraint that the total path travel time does not exceed a given threshold. More recently, Androutsopoulos and Zografos (2010) improve the solutions by excluding equivalent paths, i.e. solutions with identical paths, departure time, total travel time and risk, which only differ in the waiting time at the nodes. In particular, the authors analyze a bi-criteria routing and scheduling problem consisting in determining the non-dominated (in terms of travel time and risk) time-dependent paths for servicing a given and fixed sequence of customers (intermediate stops) within given time windows. The authors present an algorithm for determining the non-dominated scheduled paths; alternatively, a label setting algorithm is proposed for determining the k-shortest non-equivalent scheduled paths, thus approximating the non-dominated solutions close to the minimum travel time solution.

Papers studying the vehicle routing problem in the context of hazmat transportation are very few. One of the former work is by Tarantilis and Kiranoudis (2001) who consider a variant of the classical vehicle routing problem focusing on the minimization of population exposure risk. Zografos and Androutsopoulos (2002a,b) study a bi-objective vehicle routing problem with time windows (VRPTW) that minimizes both the risk and the cost of hazmat shipments. A heuristic algorithm for solving the problem is provided and tested on several VRPTW benchmark instances. Typically, route planning (i.e., selecting paths between pairs of network nodes) and vehicle routing (i.e., finding the order in which clients are visited by a vehicle) decisions are addressed in cascade with the latter being carried out beforehand determining a single path between each customer pair. More recently, Pradhananga et al. (2010) present a new ant colony system based meta-heuristic algorithm for a VRPTW with multiple attributes (risk, schedule time, and number of vehicles), where route choice and vehicle routing process are carried out in a single step.

Hazmat Global Route Planning Problem

On the contrary of hazmat local route planning, the global route planning problem has attained relatively little attention in the literature. We may classify most of the papers in three main research streams: route planning with equity considerations, hazmat transportation network design, toll setting approach. In Table 2 we list the main contributions.

Equity Considerations, Dissimilar Paths and Multi-objective Models

As said before, in global route planning for hazmat shipments, one of the main problems is finding minimum risk routes, assuring, at the same time, an equitable distribution of the risk on the interested area. This concept is well defined in the

Table 2 Contributions on global route planning for hazmat transportation by road

Road hazmat global route planning	
Generic models addressing equity	Caramia et al. (2010)[M], Caramia and Giordani (2009)[M], Martí et al. (2009)[M], Zografos and Androutsopoulos (2008)[DSS], Dadkar et al. (2008)[S,T], Carotenuto et al. (2007a,b), Bell (2007, 2006), Dell'Olmo et al. (2005)[M], Akgün et al. (2000), Marianov and ReVelle (1998), Current and Ratick (1995), Lindner-Dutton et al. (1991), Gopalan et al. (1990a,b), Zografos and Davis (1989)
Hazmat transportation network design	Reilly et al. (2012)[C], Dadkar et al. (2010)[C,S,T], Bianco et al. (2009), Verter and Kara (2008), Erkut and Gzara (2008), Erkut and Alp (2007a), Kara and Verter (2004)
Toll setting policies	Bianco et al. (2012), Wang et al. (2011), Marcotte et al. (2009)

[C]With security consideration
[DSS]Decision Support System model
[M]Multi-objective
[S]Stochastic
[T]Time-varying

works of Keeney (1980), and Keeney and Winkler (1985), where a measure of the collective risk is determined with explicit reference to the equity.

Zografos and Davis (1989) are the first authors to propose a multi-objective model that explicitly incorporates equity considerations in hazmat global route planning. Their objectives are to minimize the total risk, the risk imposed on special population categories, the travel time, and the property damage. Equity distribution of the risk is achieved by constraining the capacity of the road links. The authors adopt a goal programming approach to solve the problem, and demonstrate (using hypothetical data) that forcing equity could increase the total risk up to 35 %. Other earlier models are proposed by Gopalan et al. (1990a,b), Lindner-Dutton et al. (1991), Current and Ratick (1995), and Marianov and ReVelle (1998); for a review of them the reader is referred to the survey of Erkut et al. (2007). More recently, Bell (2006, 2007) proposes a min-max model which minimizes the maximum link risk, and hence looks for risk equity by balancing the risk through the links of the network.

In the following, we review the most important contributions adopting the concept of dissimilar paths and/or multi-objective models.

Classical approaches to tackle with risk equity are those based on the generation of different paths in order to alternate the route among them and hence distribute the risk. Finding different paths from an origin to a destination in a given network is a classical optimization problem that extends the well known shortest path problem see, e.g., Ahuja et al. (1993). The generation of different paths may be obtained by solving the k-shortest path problem in which the shortest, second shortest, ..., kth shortest paths from an origin to a destination are found in the network. However, many of these alternative paths are likely to share a large number of links, while in the context of hazmat transportation the aim is to find spatially dissimilar paths that minimize the risk, and over which we can distribute as uniformly as possible

hazmat shipments in order to spread the risk with equity. Specifically, we need to consider the path dissimilarity problem, which consists in determining a set of paths of minimum length (and/or risk) and maximum dissimilarity.

Different methods have been proposed in the past to solve the path dissimilarity problem, also in other application contexts. One of these methods is the Iterative Penalty Method (Johnson et al. 1992), which is based on the iterative application of shortest path algorithms: at each iteration, a penalty is associated to each selected link, so as to discourage the repeated selection of the same link in the forthcoming iterations; therefore, in such a way, the method generates dissimilar paths. The main advantage of this method is that it only requires a shortest-path algorithm to generate paths. A drawback is that its effectiveness heavily depends on the penalization parameter, meaning that a small penalty may not achieve the goal of dissimilarity, while a large penalty may eliminate a great number of viable paths from consideration; moreover, the method has no way for measuring the quality of the produced dissimilar paths, in terms of spatial differences and lengths.

Another proposed method is the Gateway Shortest Path (Lombard and Church 1993), based on the generation of shortest paths from an origin to a destination that should go through given nodes called "gateways." At each step, this method first selects a gateway node, then computes two paths, one from the origin to this node and another from this node to the destination, and finally connects the two paths for obtaining the final path between the origin and the destination. The concept of "area under a path" is introduced to evaluate the dissimilarity between two paths. An advantage of this method is that a large number of alternative paths may be generated by simply using a shortest-path algorithm twice. Drawbacks are that the final paths may contain loops, and that the quality of the generated paths is strongly affected by the selected gateways.

A third approach is the Minimax method (Kuby et al. 1997), which selects routes starting from k assigned paths using some dissimilarity indices. The algorithm starts by generating k-shortest paths between the origin and the destination using a k-shortest path algorithm (see, e.g., Yen 1971). Then, a dissimilar subset of the k-shortest paths is constructed iteratively. An index paying attention to the similarity between the selected paths and to path lengths is defined to measure the desirability of a candidate path for inclusion in the dissimilar subset. The first path included is the shortest path, then the second path is the one that minimizes the similarity index. The third path is chosen minimizing the maximum (hence, the name of the method follows) of the similarity indices between the candidates and the first two paths, and so on until the desired number of paths is reached.

Successively, Akgün et al. (2000) review the above classical methods for generating dissimilar paths, and propose another dissimilar path model that makes use of a p-dispersion location model (Erkut 1990; Erkut et al. 1994); the latter consists in selecting p points in some space so as to maximize the minimum distance between any couple of selected points. In the method of Akgün et al. (2000), a large set of candidate paths is generated, using the k-shortest (in terms of risk) path algorithm and the iterative penalty method; these paths are then considered as a set of points, with the distance between two points representing the dissimilarity

between the two relative paths. Indeed, the authors observe that the minimax method is just a greedy algorithm for solving the p-dispersion problem, and therefore they propose an improvement employing a constructive semi-greedy algorithm with a local search method, previously used by Erkut (1990), to select a subset of p dissimilar paths from a set of shortest paths. According to one of the four indices proposed by Erkut and Verter (1998) to measure the dissimilarity among two paths, Akgün et al. (2000) measure the similarity between two paths as the relative total length of the common links with respect to the length of each single path, and then the dissimilarity by subtracting the similarity measure from 1. Experimentally, the authors show that their method is superior in comparison with the classical ones.

The dissimilarity index used by Akgün et al. (2000) is based on the evaluation of the lengths of common links between two paths, and hence it may happen that highly dissimilar paths (i.e. with very few common links) are indeed geographically very close, with the consequence that if these two paths are selected for routing hazmat vehicles, the people living in the path neighborhood will be influenced by the risk of both paths, implying a low degree of risk equity.

Starting from such a drawback in the definition of the dissimilarity index, Carotenuto et al. (2007a) propose a model for generating dissimilar paths that takes into account also the risk induced on the links in the neighborhood of a selected path; that is, they take into account the risk coming from incident effect propagation. The authors propose a measure of link risk that includes this aspect by extending the concept of buffer zone around a link introduced by Batta and Chiu (1988), and model the equity of risk by bounding the maximum risk sustained by each link of the network. They consider the selection of p distinct simple paths, so as to minimize the total path risk while satisfying a risk threshold constraint in the traversed links; in this way the risk equity is addressed. The authors provide a mathematical formulation for the proposed model and use it to get a Lagrangian relaxation in order to achieve an effective lower bound on the optimal solution value. The obtained lower bound is used to evaluate the effectiveness of two proposed constructive algorithms: a greedy and a greedy randomized algorithms. Besides the total risk, different indices are also considered for evaluating the generated paths.

In another paper, Carotenuto et al. (2007b) also consider the need to distribute the risk in an equitable way with respect to both the space and the time, avoiding as much as possible the presence of more than one hazmat vehicle at the same time on the same zone. They propose a routing and scheduling approach operating in two phases. For each given hazmat shipment request a set of routes from the origin node to the destination node is determined using the approach of Carotenuto et al. (2007a); then the shipment requests are scheduled and routed by assigning one of the selected routes and a starting time to each shipment in order to avoid any pair of hazmat vehicles being too much close at any given time. The authors model the problem as a flow-shop scheduling problem with alternative routes and heuristically solve it with a tabu-search algorithm.

Dell'Olmo et al. (2005) extend the similarity index used by Akgün et al. (2000) with the concept of buffer zone to include spatial information in the measure of dissimilarity. The similarity function they propose is essentially the same but with

the length of the links of a path replaced by the area of the buffer zone around the link: that is, the similarity is computed by evaluating the size of the common area inside the buffer zones around the links of the two paths, instead of simply measuring the lengths of common links.

Successively, Martí et al. (2009) review all of these previous methods and adapt them to a bi-objective routing problem where a set of paths from an origin to a destination must be generated with minimum length and maximum dissimilarity. They introduce a new dissimilarity function. They also propose a new greedy randomized local search procedure and test it against the reviewed methods, showing experimentally that it is able to give better approximations of the efficient frontier of the considered bi-objective problem than existing methods.

Dadkar et al. (2008) develop an extension of the k-shortest path algorithm of Yen (1971) for which the performance of each highway facility, with respect to each objective, can be stochastic and can vary over time. They extend the method of Chang et al. (2005) to construct an estimate of the travel time probability distribution. They also propose a mixed integer program to identify a subset of paths which represents an acceptable trade-off between geographic diversity and performance and solve the problem heuristically with a genetic algorithm. These models and algorithms are then applied to a realistic case study.

The main limit of the approaches reviewed above is that they do not directly take into account the representativeness of the selected paths, but consider path dissimilarity as the main goal. Indeed, selecting paths for routing hazmat shipments is intrinsically a multi-objective problem, since there are inherently conflicting objectives, e.g., risk minimization, cost/length minimization, equity maximization, etc. In particular, there are cases where low risk-routes lead to more expensive transportation costs because such routes may be much larger in terms of length than the shortest route.

Nonetheless, only few papers address the route planning problem by means of multi-objective optimization approaches. Cox (1984) develops a multi-objective algorithm in order to find the set of efficient paths for the hazmat transportation problem using different attributes associated to the network links, such as travel time, population density, etc.; successively, Wijeratne et al. (1993) propose a model considering stochastic attributes for the network links.

The multi-objective shortest path problem has been widely studied (see, e.g., Hansen 1980 and Martins 1984). Since the number of efficient paths may increase exponentially with the number of the network nodes (Hansen 1980), the algorithms proposed in the literature face the difficulty to manage the large number of efficient outcoming paths and the considerable computational time, even in case of small instances. Most of the algorithms are either label setting or correcting algorithms. Martins and Santos (1999) propose multi-objective labeling algorithms generalizing the optimality principle of the (single-objective) shortest path problem; they provide two implementations based on label setting and label correcting, respectively.

Zografos and Androutsopoulos (2008) propose a simple heuristic for determining alternative non-dominated hazardous materials distribution routes in terms of cost and risk minimization. They use the heuristic within a decision support system for

assessing alternative distribution routes in terms of travel time, risk and evacuation implications, while coordinating the emergency response deployment decisions with the hazardous materials routes.

Dell'Olmo et al. (2005) use multi-objective optimization approaches proposing a two phase approach. In the first phase, the whole set of efficient paths between the origin and the destination is determined by means of the label correcting algorithm of Martins and Santos (1999). In the second one, k efficient paths are selected maximizing their dissimilarity, by heuristically solve the p-dispersion problem, assuming the dissimilarity between each pair of paths found in the first phase as distance measure. The heuristic algorithm is a constructive heuristic followed by a simple local search. The path selection is done only with respect to path dissimilarity and does not consider the representativeness of the selected paths with respect to the cost vectors of all the efficient paths.

Successively, Caramia and Giordani (2009) and Caramia et al. (2010) try to overcome this drawback, proposing new approaches that selects k efficient paths with respect to several measures, e.g., length, travel time (cost), and risk; path selection is made by choosing k paths maximizing their representativeness with respect to the set of all the efficient paths, and with high spatial dissimilarity. The approach proposed by Caramia et al. (2010) first exploits the algorithm of Martins and Santos (1999) to find the set of efficient paths, and then uses a k-means algorithm to partition the latter set into k classes of paths, minimizing the total variance of the objective vector values of the paths in the same class. Next, one path from each one of the k classes is chosen by heuristically solving the problem of selecting paths maximizing the total spatial dissimilarity. Caramia and Giordani (2009) devise a similar but more sophisticated method, adopting a clustering-based approach for selecting k efficient paths maximizing their representativeness with respect to the cost vectors of all the efficient paths or the dissimilarity among the k selected paths. The approach works in three phases: in the first one, the set of efficient paths is determined e.g., with the use of the algorithm of Martins and Santos (1999). In the second phase, a fuzzy k-means based routine is used to compute fuzzy path-class memberships representing a fuzzy k-class partition of the efficient paths. In the third phase, a Montecarlo method, repeated for a certain number of times that exploits fuzzy memberships as path-class assignment probabilities, generates a k-class partition of the efficient paths, and from each one of the k path classes it selects the path with the closest cost vector to the class centroid. The k-class partition of the efficient paths (along with the related selection of k paths) is chosen by minimizing the sum over all the classes of the total square distance between the cost vector values of the paths of a class and the class centroid (i.e., maximizing path representativeness), or maximizing the dissimilarity among the k selected paths. Computational results are presented on random graphs and show a promising behavior with particular emphasis on the capability of producing representative paths with a considerably high dissimilarity.

We close this subsection mentioning the model proposed by Bonvicini and Spadoni (2008) which does not provide specific routing plans, but a hazmat flow plan on a given network. Differently from the above approaches, they

consider both multi-commodity and multi-origin destination, proposing a simple multi-commodity linear network flow model with global link capacities, with flow decision variables representing yearly hazmat vehicle flows; the model is solved using a commercial solver. In particular, global link capacities are considered to reduce risk overloading on specific links, and, hence, to look for risk equity. Different linear objective functions are introduced based on different link impedances (e.g., risk, length, cost, etc.); the model is capable to optimize one objective function at a time with the others implicitly considered as specific flow constraints in the model formulation.

Hazmat Transportation Network Design

The models reviewed above may be useful in finding a global routing plan for a major hazmat producer/carrier that takes into account the equitable distribution of transport risk in a region. However, these models are of little use in identifying a comprehensive global transportation plan in a jurisdiction with multiple carriers since governments, in general, have no authority to impose routes on individual carriers.

Indeed, many governments have the authority to close certain road segments to hazmat vehicles or to limit the amount of hazmat traffic flow on those links. These kinds of policies are usually categorized as Hazmat Transportation Network Design (HTND) policies, and equity concerns can be incorporated into the design objectives.

Network design has been widely studied in the past (see, e.g., the surveys of Magnanti and Wong 1984, Balakrishnan et al. 1997, and Yang and Bell 1998, for reviews of network design problems for road transportation). Differently from the classical network design problem where the aim is to find the most appropriate ways to expand a given infrastructure, in HTND the problem is identifying which road segments should be partially or entirely closed to hazmat transport in an existing network, for example in order to minimize the total risk induced by hazmat shipments. HTND has received little attention from researchers and only recently. Next, we review the papers that fall in this research field.

Most of network design literature models the problem as either a bilevel optimization problem (Bard 2006), or a Mathematical Program with Equilibrium Constraints (MPEC) (see e.g. Luo et al. 1996 and Outrata et al. 1998).

Kara and Verter (2004) were the first to study the HTND proposing a bilevel integer programming model by considering the roles of carriers and of a government authority. They assume that the carriers, represented by the follower (second level) decision makers in the bilevel model, will always use the cheapest routes on the hazmat transportation network designed by the government authority. The authority plays the role of the leader (first level) in the bilevel model, and has the objective to select the minimum total risk network to open to hazmat shipments, taking into account the cost-minimizing behavior of the carriers. In their model, hazmats are grouped into categories based on risk impact, and a network is designed for each

group, without considering the interactions among shipments of hazmats of different categories. Since the followers' problem is linear, the bilevel integer programming problem is reformulated as a single-level Mixed Integer Programming (MIP) problem by replacing the followers' problem by its Karush–Kuhn–Tucker (KKT) conditions and by linearizing the complementary slackness constraints. Then, the latter MIP problem is solved using a commercial optimization software. Marcotte et al. (2009) propose a more efficient single-level MIP reformulation replacing the complementary slackness constraints with equality constraints between the objective function values of the followers' primal and dual problems. However, both the two single-level reformulations may fail to find an optimal stable solution for the bilevel model. In fact, in general, there are multiple minimum-cost routing solutions for the followers over the designed network established by the leader, which may induce different total risk values over the network. We note that Kara and Verter (2004) do not take into account such an issue.

Erkut and Alp (2007a) consider a single-level HTND model, restricting the network to a tree, so that there is a single path between each couple of origin–destination pair; with this restriction, the carriers have no alternative paths on the tree, hence the carriers have no freedom in route selection, with the result that the structure of the proposed model has a single level. They formulate the tree design problem as an integer programming problem with the objective of minimizing the total risk, which is solved using a commercial optimization software. However, the solutions may result in circuitous and expensive routes. To avoid an economically infeasible solution for the carriers, the authors also propose a greedy heuristic that adds shortest paths to the tree so as to keep the risk increase to a minimum and allow the carriers to select cheaper paths.

Erkut and Gzara (2008) generalize the model of Kara and Verter (2004) by considering the undirected network case and designing the same network for all the shipments. They consider the possible lack of stability of the solution of the bilevel model obtained by solving the single-level MIP model, and propose a heuristic solution method that always finds a stable solution. Moreover, they extend the bilevel model to account for the cost/risk trade-off by including cost in the objective function of the leader problem.

All the above papers adopt a link-based formulation for the carriers' problem, while Verter and Kara (2008) provide a new path-based formulation for the HTND problem studied by Kara and Verter (2004), where the open links in the given road network chosen by the regulator determine the set of paths that are available to the carriers. This facilitates the incorporation of carriers' cost concerns in regulator's risk reduction decision, and allows to formulate the problem with a single-level integer programming formulation assuring that the cheapest path among the available ones is used by each carrier.

All the above HTND models consider the government and the carriers points of view, trying to mitigate the risk only from a macroscopic point of view but without considering the need to distribute the risk in an equitable way over the region in which the transportation network is embedded; in fact, the choices of the carriers, that are related to the cost, could overload, in terms of risk, some links of

the network, implying a lack of risk equity. This could be inadequate when there are multiple layers of government authorities being involved in the regulation of dangerous goods shipments (as, for example, it is common in Europe and North America), that are responsible at different geographical levels, e.g., regional area authorities and local area authorities. In such a scenario, a regional area authority aims to minimize the total risk over its controlled area, while a local area authority wants the risk over the local populated links of its jurisdiction to be the lowest possible.

Bianco et al. (2009) analyze such a scenario, and study new models for HTND addressing also risk equity. The problem they consider is the following: a set of hazmat shipments has to be shipped over a road transportation network in order to transport a given amount of hazardous materials from specific origin points to specific destination points. They assume that there are government authorities (at different levels, e.g., regional and local) that want to regulate the hazmat traffic by restricting the use of network links to the hazmat shipments with the aim, on the one hand, of minimizing the total risk induced by the shipments and, on the other hand, of spreading the risk equitably over the geographical region in which the transportation network is embedded. The former aim concerns the interests of a regional area authority, while the latter one goes in the direction of the interests of local area authorities (that are responsible to regulate the hazmat traffic inside their local area contained in the regional area) that would like to avoid local populated links in their jurisdictions to be overloaded in terms of induced risk by hazmat shipments. To the best of our knowledge, this is the first work that pays particular attention to the jurisdictional differences within different authorities. Bianco et al. (2009) also formulate their hazmat network design problem with a linear bilevel model, where at the higher (leader) level there is a meta-local authority (acting on behalf of all the involved local area authorities) that aims to minimize the maximum link risk over populated links of the whole network, that is, risk equity, and at the lower (follower) level there is the regional area authority that aims to minimize the total risk over the network. This corresponds to the existence of two decision makers, one (the regional authority) willing to define a feasible hazmat flow assignment on the network that induces the minimum total risk over the population, and the other (representing the local authorities) that, interpreting the optimal flow assignment of the previous (lower level) decision maker as a flow vector, minimizes the maximum link risk on the network, i.e., aiming at risk equity, by defining capacities over the network links that restrict the possible choices of the regional authority. They reduce the bilevel problem to a single-level integer linear program by replacing the follower problem with the KKT conditions and by linearizing the complementary slackness constraints. The resulting problem can be optimally solved through commercial integer programming software only for very small networks; therefore, the authors also provide heuristic based approaches experimented on a realistic regional network.

Reilly et al. (2012) and Dadkar et al. (2010) present a new hazmat network design problem that takes into account security issues by modeling the possible role of a terrorist. They provide a bilevel model, where at the lower level there are the carrier

and the terrorist whose decision making is modeled as a two-player game, and at the higher level there is a government authority willing to minimize the expected consequence of a terrorist attack by closing some of the network links, and therefore constraining the choices that the shipper/carrier can make. Dadkar et al. (2010) develop a non-zero sum game-theoretic model of the interactions among carriers and terrorist, by considering a single repetitive hazmat shipment from a single origin to a single destination. A heuristic is used to find optimal sets of government imposed facility restrictions that produce a Nash equilibria where the carrier's expected payoff is as high as possible given that the terrorist's expected payoff is below a prescribed threshold. This work assumes only one hazmat shipment and a single origin–destination pair. Reilly et al. (2012) extend the approach considering multiple origin–destination pairs but with a single carrier and a single hazmat type.

Toll Setting Policies

HTND policies can effectively restrict hazmat shipments in order to induce carriers to route shipments on low-risk paths. However, such a restriction could be too much since it does not consider the carriers' priorities, possibly wasting the usability of certain road segments. On the other hand, only restricting certain road segments could not rationally adjust hazmat flows to less-risk areas. Recently, an alternative policy tool was proposed by Marcotte et al. (2009) to discourage (but not prevent) hazmat carriers from using certain road segments via Toll Setting (TS). They also show that the toll setting policy is a more flexible regulation tool for hazmat than network design policy. By imposing tolls on certain road segments, the hazmat shipments are expected to be directed on less-populated roads according to the carriers' own selection (due to economic considerations) rather than by governors' restriction. This model may result in a more attractive policy to regulators since provide more flexible solutions, and at the same time more acceptable to carriers that maintain the freedom of using any link of the network.

Marcotte et al. (2009) propose a bilevel model that minimizes the risks and costs including both transportation costs and tolls on the links. They reduce the bilevel problem to a single-level MIP by replacing the followers' problem with optimality conditions and by linearizing the complementary slackness constraints; they also provide an alternative single-level MIP reformulation where complementary slackness conditions are replaced with the equality between the primal and dual objective functions of the followers' problem. Moreover, they show that the authority can easily finds a toll setting inducing a minimum risk solution by inverse optimization, that is, determining the link tolls that induce the carriers to choose the minimum risk route plan; hence, in this case, the problem is not a bilevel problem anymore, but reduces to a single-level problem. However, if the authority wants to keep into account also carriers' cost, the problem is a true bilevel optimization problem.

Wang et al. (2011) assume that both hazmat traffic and regular traffic affect population safety, since congestion increases delay and then accident probabilities. The idea is to control both regular and hazmat traffic via toll setting. The authors

make the following assumptions: (1) congestion induced by the traffic flow of hazmat trucks can be ignored; (2) to simplify the model, the users have perfect information of the current status; (3) all the model parameters are deterministic; (4) a single type of hazmat is considered; (5) travel delay is a linear function of traffic congestion; (6) risk is linearly affected by travel delay. The authors provide a bilevel formulation and an equivalent two-stage problem formulation, where the first-stage problem is a non-convex quadratic programming problem and the latter is a linear programming problem.

We conclude this section mentioning a new toll setting model proposed by Bianco et al. (2012), that extends the one proposed in Marcotte et al. (2009). This new model not only aims to minimize the total risk on the network but also to discourage the concentration of high level of risk on some links by minimizing the maximum link total risk, and in such a way to keep into account risk equity. To achieve this result, it is assumed that the toll paid by a carrier on a link depends on the total risk on that link. This implies that the routing choice of a carrier depends on the other carriers' choices, leading to a Nash game. The resulting model is a MPEC where the inner problem is a Nash game having the carriers as players, and the outer problem is faced by the authority that sets the tolls on the links in order to minimize both the network total risk and the maximum link total risk induced by the carriers' choices. In section "Toll Setting in Hazmat Transportation," we give a formal description of this model.

Hazmat Transportation Network Design Models

Let $G = (N, A)$ be a directed network with N being the set of n nodes (intersections in the road network) and A being the set of m directed links or arcs (road segments) between pairs of nodes. The network is assumed undirected when all the links are undirected (i.e., the road segments can be traversed in both directions); in this case, let E be the set of undirected links or edges of the network, and $A = \{(i, j), (j, i) | \langle i, j \rangle \in E\}$ be the set of pairs of arcs related to the edges traversed forth and backward, with arc $(i, j) \in A$ representing edge $\langle i, j \rangle \in E$ traversed from node i to node j.

Let c_{ij} be the transportation cost (length) of arc $(i, j) \in A$. If the network is undirected we assume that the traveling cost of edge $\langle i, j \rangle \in E$ does not depend on the direction in which the edge is traversed, and hence the costs of the arcs $(i, j), (j, i) \in A$ related to traversing that edge forth and backward, respectively, are assumed to be $c_{ij} = c_{ji}$ and equal to the edge traveling cost.

Given a set H of hazmat types, we consider a set K of p carriers with carrier $k \in K$ having to satisfy a single shipment order (commodity) of hazmat of type $h(k) \in H$ from origin node s^k to destination node t^k. The amount of hazmat to be shipped by carrier k (i.e., the demand of k) is denoted with b^k and with n^k we denote the number of trucks used for the shipment. For the sake of simplicity, we assume that carrier k uses a fleet of homogeneous truck (vehicles) each one of capacity q^k

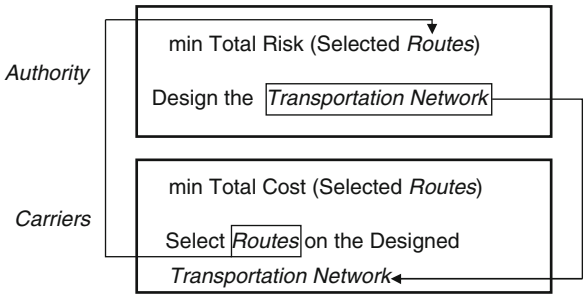


Fig. 1 The hazmat transportation network design problem

and traveling at full load (hence, we may assume that $b^k = n^k \cdot q^k$); indeed, for the sake of simplicity, all the carriers shipping the same hazmat type are assumed to use the same type of vehicles.

Let ρ_{ij}^h be the risk (e.g., number of people exposed) induced by a truck carrying hazmat of type h through link $(i, j) \in A$; typically, the risk does not depend on the direction in which a road segment is traversed, therefore it is assumed that $\rho_{ij}^h = \rho_{ji}^h$.

HTND: Models with an Authority and Several Carriers

In the generic Hazmat Transportation Network Design (HTND) problem it is assumed that an authority has the power to decide which links can be used for transportation of hazmat type $h \in H$, e.g., in order to force the carriers to route shipments on less risky links and hence minimizing the total risk induced by the hazmat shipments. Then the carriers select the routes on the available subnetwork, e.g., minimizing their transportation cost.

The modeling framework used to represent the HTND problem is a bilevel optimization problem where the authority plays the role of the leader (first level) decision maker, and has the objective to select the minimum total risk network to open to hazmat shipments, taking into account the cost-minimizing behavior of the carriers, that play the role of the follower (second level) decision maker. Figure 1 shows this bilevel framework introduced by Kara and Verter (2004).

In the HTND problem two set of variables are considered for modeling the decisions of the authority and carriers, respectively:

- y_{ij}^h, being the binary variables representing the decision of the authority and equal to 1 if arc $(i, j) \in A$ is open for the transportation of hazmat of type h, 0 otherwise.
- x_{ij}^k, being the binary variables modeling the decisions of the carriers and equal to 1 if arc $(i, j) \in A$ is used by carrier k for the shipment, 0 otherwise.

In fact, since the authority can open or close the links and no restriction (flow constraint) is assumed on open links, the decision of each carrier k results in

selecting a route (e.g., of minimum cost) on the available subnetwork from the origin s^k to the destination t^k with all the amount of hazmat being shipped along the selected route (i.e, unsplittable flow).

Although the main concern of the authority is the minimization of the risk while the carriers mainly address the minimization of the transportation cost, it is reasonable to assume that both the authority and the carriers take also into account at least in part the objective of the other party. Given the above decision variables, this more general version of the HTND problem can be formulated with the following bilevel program:

$$
\begin{array}{lll}
\min\limits_{y_{ij}^h} & \displaystyle\sum_{k\in K}\sum_{(i,j)\in A} n^k(\rho_{ij}^{h(k)} + w_1 c_{ij})x_{ij}^k & \\
\text{s.t.} & y_{ij}^h \in \{0,1\} & \forall\ (i,j)\in A,\ h\in H \\
\text{where } x_{ij}^k & \text{solve:} & \\
\min\limits_{x_{ij}^k} & \displaystyle\sum_{k\in K}\sum_{(i,j)\in A} n^k(c_{ij} + w_2 \rho_{ij}^{h(k)})x_{ij}^k & \\
\text{s.t.} & & \\
& \displaystyle\sum_{\{j:(i,j)\in A\}} x_{ij}^k - \sum_{\{j:(j,i)\in A\}} x_{ji}^k = e_i^k & \forall\ i\in N,\ k\in K \\
& x_{ij}^k \le y_{ij}^{h(k)} & \forall\ (i,j)\in A,\ k\in K \\
& x_{ij}^k \in \{0,1\} & \forall\ (i,j)\in A,\ k\in K,
\end{array}
\tag{1}
$$

where e_i^k is equal to $1, -1$ or 0 depending on if node i is the origin, the destination or a transshipment node for carrier k. For the sake of generality, both the objective functions of the authority (the leader) and the carriers (the followers) consider the minimization of total risk (assuming additivity of impacts) and total transportation cost, with parameters w_1 and w_2 that allow the comparison between total risk and carriers' total transportation cost in the objective functions of the leader and the followers, respectively.

The leader (outer) problem has no constraints (except for the binary constraints on the variables). The followers' (inner) problem presents the flow balance requirements on each network node and for each carrier (shipment), and the constraints on the carriers' decision variables implying that only the available links (decided by the leader) can be used by the carriers. One may control possible overload in terms of risk on the population residing in the neighborhood of a link by limiting the number of vehicle traversing any link (i.e., by adding to the leader problem the constraints $\sum_{k\in K} n^k x_{ij}^k \le \omega_{ij}$ for each link (i,j), with ω_{ij} being the maximum total number of hazmat trucks allowed on link (i,j)). Note that with the addition of these constraints the problem is not guaranteed to have a feasible solution even on a connected network. Without the addition of such kind of constraints, there is no interaction between shipments of different hazmat types, and the followers' problem decomposes into $|K|$ constrained shortest path problems.

The Model of Kara and Verter (2004)

Kara and Verter (2004) model the HTND problem on a directed network $G = (N, A)$ and propose a formulation that can be reduced to the general one above described with both parameters w_1 and w_2 set to 0, that is, the authority objective is the minimization of the total risk only and the objective of the carriers is the minimization of the transportation cost only. They estimate the risk ρ_{ij}^h by counting the number of people exposed by a truck transporting hazmat of type h. In particular, they consider the population located in a set P of population centers, and they assume to know for each population center $p \in P$ the number $PE_{ij}^{p,h}$ of people exposed by a truck transporting hazmat of type h on link $(i, j) \in A$; therefore, we have $\rho_{ij}^h = \sum_{p \in P} PE_{ij}^{p,h}$.

Once the leader variables y_{ij}^h are given and assumed as parameters, the followers' problem is an integer linear program with the "unimodularity" property, that is, the constraints' matrix is totally unimodular and the constraints' right-hand-side vector is integer; in this case, every extreme point, or vertex, of the feasible region of the linear relaxation is integral and thus the feasible region is an integral polyhedron (Papadimitriou and Steiglitz 1982). Hence, the integrality requirements on the followers' variables x_{ij}^k can be relaxed and replaced with $x_{ij}^k \geq 0$ without losing optimality. Therefore the bilevel problem is discrete-linear and in particular since the followers' problem is linear we can represent it with its primal-dual optimality (or KKT) conditions, reducing the bilevel problem to a pure single-level optimization problem by replacing the followers' problem with the equivalent optimality conditions. This is the strategy followed by Kara and Verter (2004), where the optimality conditions of the followers' problem are

$$
\begin{array}{ll}
\displaystyle\sum_{\{j:(i,j)\in A\}} x_{ij}^k - \sum_{\{j:(j,i)\in A\}} x_{ji}^k = e_i^k & \forall\ i \in N,\ k \in K \\
x_{ij}^k \leq y_{ij}^{h(k)} & \forall\ (i,j) \in A,\ k \in K \\
n^k c_{ij} - \pi_i^k + \pi_j^k - v_{ij}^k + \lambda_{ij}^k = 0 & \forall\ (i,j) \in A,\ k \in K \\
x_{ij}^k\ v_{ij}^k = 0 & \forall\ (i,j) \in A,\ k \in K \\
\lambda_{ij}^k (y_{ij}^{h(k)} - x_{ij}^k) = 0 & \forall\ (i,j) \in A,\ k \in K \\
x_{ij}^k \geq 0, v_{ij}^k \geq 0, \lambda_{ij}^k \geq 0 & \forall\ (i,j) \in A,\ k \in K \\
\pi_i^k \text{ free} & \forall\ i \in N,\ k \in K,
\end{array}
\tag{2}
$$

where:

- π_i^k are free dual variables associated with the first set of primal constraints of the follower's problem of model (1), with $i \in N$, and $k \in K$.
- λ_{ij} are non-negative dual variables associated with the second set of primal constraints of the follower's problem, with $(i, j) \in A$.
- v_{ij}^k are non-negative slack variables of the dual constraints of the follower's problem, with $(i, j) \in A$, and $k \in K$.

The first two set of constraints are the primal constraints that together with the non-negativity conditions on the primal variables x_{ij}^k form the primal feasibility conditions; the third set represents the dual constraints that together with the restrictions in sign on the dual variables π_i^k and λ_{ij}^k and the slack variables v_{ij}^k of the dual constraints define the dual feasibility conditions; finally, the fourth and fifth sets of constraints are the complementary slackness conditions.

Note that once the above conditions are added to the leader problem reducing the bilevel problem to a single-level one, we lose the total unimodularity property; therefore we again have to force integrality on variables x_{ij}^k in the single-level model, with the consequence that the latter is a MIP problem with the following bilinear constraints

$$\begin{aligned} & v_{ij}^k x_{ij}^k = 0 && \forall\ (i,j) \in A,\ k \in K \\ & \lambda_{ij}^k (y_{ij}^{h(k)} - x_{ij}^k) = 0 && \forall\ (i,j) \in A,\ k \in K. \end{aligned} \tag{3}$$

Kara and Verter (2004) propose to linearize these constraints as follows, taking advantage of the binary nature of the variables x_{ij}^k and y_{ij}^h, and with M being a sufficiently large positive scalar:

$$\begin{aligned} & v_{ij}^k \le M(1 - x_{ij}^k) && \forall\ (i,j) \in A,\ k \in K \\ & \lambda_{ij}^k \le M[1 - (y_{ij}^{h(k)} - x_{ij}^k)] && \forall\ (i,j) \in A,\ k \in K. \end{aligned} \tag{4}$$

Finally, the single-level MIP formulation for the HTND problem proposed by Kara and Verter (2004) is:

$$\begin{aligned} \min\ & \sum_{k \in K} \sum_{(i,j) \in A} n^k \rho_{ij}^{h(k)} x_{ij}^k \\ \text{s.t.}\ & \sum_{\{j:(i,j) \in A\}} x_{ij}^k - \sum_{\{j:(j,i) \in A\}} x_{ji}^k = e_i^k && \forall\ i \in N,\ k \in K \\ & x_{ij}^k \le y_{ij}^{h(k)} && \forall\ (i,j) \in A,\ k \in K \\ & n^k c_{ij} - \pi_i^k + \pi_j^k - v_{ij}^k + \lambda_{ij}^k = 0 && \forall\ (i,j) \in A,\ k \in K \\ & v_{ij}^k \le M(1 - x_{ij}^k) && \forall\ (i,j) \in A,\ k \in K \\ & \lambda_{ij}^k \le M[1 - (y_{ij}^{h(k)} - x_{ij}^k)] && \forall\ (i,j) \in A,\ k \in K \\ & v_{ij}^k \ge 0, \lambda_{ij}^k \ge 0 && \forall\ (i,j) \in A,\ k \in K \\ & \pi_i^k \text{ free} && \forall\ i \in N,\ k \in K \\ & x_{ij}^k \in \{0,1\} && \forall\ (i,j) \in A,\ k \in K \\ & y_{ij}^h \in \{0,1\} && \forall\ (i,j) \in A,\ h \in H. \end{aligned} \tag{5}$$

Kara and Verter (2004) observe that the proposed approach can be applied also when additional constraints are added to the leader problem as outlined at the beginning of this section since the unimodulairty of the followers' problem is not affected by the addition of some other constraints to the leader problem. The authors

solve the proposed single-level MIP model using a commercial solver (CPLEX 6.0) on a network with 48 nodes and 57 links representing the Western Ontario road network and by considering up to 22 shipments (carriers) and 2 hazmat types. No details are given for the type of computer used for the experimentation and the CPU time needed to find the optimal solution.

Marcotte et al. (2009) propose a more efficient single-level MIP reformulation replacing the complementary slackness constraints (3) with the following equality constraints between the objective function values of the followers' primal and dual problems

$$\sum_{(i,j)\in A} n^k (c_{ij} + w_2 \rho_{ij}^{h(k)}) x_{ij}^k = \sum_{i\in N} e_i^k \pi_i^k - \sum_{(i,j)\in A} \lambda_{ij}^k y_{ij}^{h(k)} \ \forall\, k \in K, \tag{6}$$

with $w_2 = 0$ according to the model of Kara and Verter (2004).

The bilinear terms $\lambda_{ij}^k y_{ij}^{h(k)}$ in Constraints (6) are linearized by replacing them with new variables τ_{ij}^k, and adding the following linear constraints to the model to ensure that $\tau_{ij}^k = 0$ when $y_{ij}^{h(k)} = 0$, and $\tau_{ij}^k = \lambda_{ij}^k$ when $y_{ij}^{h(k)} = 1$:

$$\begin{aligned}
&\tau_{ij}^k \geq 0 && \forall (i,j) \in A,\ k \in K \\
&\tau_{ij}^k - M_{ij}^k y_{ij}^{h(k)} \leq 0 && \forall (i,j) \in A,\ k \in K \\
&\lambda_{ij}^k - \tau_{ij}^k \geq 0 && \forall (i,j) \in A,\ k \in K \\
&\lambda_{ij}^k - \tau_{ij}^k + M_{ij}^k y_{ij}^{h(k)} \leq M_{ij}^k && \forall (i,j) \in A,\ k \in K,
\end{aligned} \tag{7}$$

with M_{ij}^k being sufficiently large positive scalars.

In this way, an alternative MIP formulation is obtained solely based on the integrality of the y variables. Marcotte et al. (2009) show that their alternative formulation is more efficient than that one proposed by Kara and Verter (2004), and also provide an improved solution methodology where the TS policy is used to construct an initial solution.

However, both the two single-level MIP reformulations may fail to find an optimal stable solution for the bilevel model. In fact, in general, there are multiple minimum-cost routing solutions for the followers over the designed network established by the leader, which may induce different total risk values over the network. We note that Kara and Verter (2004) do not take into account such an issue.

The Model of Erkut and Gzara (2008)

As recalled in the previous section both the single-level MIP reformulations may fail to find an optimal stable solution for the bilevel model, because they assume that, if there are multiple minimum-cost routing solutions for the carriers over the designed network established by the authority, each carrier will follow the minimum-cost route that induces also the minimum risk; therefore, the two single-level reformulations model the *optimistic* case. Unfortunately, there is no certainty

that the carriers will select these (optimistic) routes, neither a way to impose these specific route to the carriers, that instead have the freedom to select one of the minimum-cost route with the result that the total risk may be larger than the one coming from the optimistic solution produced by the single-level model.

Erkut and Alp (2007a) consider such an issue and bypass the stability problem by restricting the network to a tree, so that there is a single path between each couple of origin–destination pair; with this restriction, the carriers have no alternative paths on the tree, hence the carriers have no freedom in route selection, with the result that the structure of the proposed model has a single level.

Erkut and Gzara (2008) take into account the possible lack of stability of the single-level model and note that for the authority it may be preferable to design a network with a small deviation from the best one but with the property that there is a single minimum-cost routing solution for the carriers over the designed network, that is a "stable" network. Accordingly, they propose a heuristic solution method that always finds a stable solution. They also generalize the model of Kara and Verter (2004) by considering the undirected network case and designing the same network for all the shipments. Given an undirected network $G = (N, E)$, Erkut and Gzara (2008) therefore assume that if an undirected link is open it is available for being traversed in both directions. Moreover, they extend the bilevel model to account for the cost/risk trade-off by including cost in the objective function of the leader problem.

Below, we report the mathematical model proposed by Erkut and Gzara (2008) for the HTND problem, where, differently from the general model (1), y_{ij} are the binary variables representing the decision of the authority, with y_{ij} being equal to 1 if arc $(i, j) \in A$ is open for hazmat transportation (independently from the hazmat type). As for the general model (1), the carriers' decisions are modeled with the binary variables x_{ij}^k, where x_{ij}^k takes the value 1 if arc $(i, j) \in A$ is used by carrier k for the shipment, 0 otherwise.

$$
\begin{aligned}
&\min_{y_{ij}} \quad \sum_{k \in K} \sum_{(i,j) \in A} n^k (\rho_{ij}^{h(k)} + w_1 c_{ij}) x_{ij}^k \\
&\text{s.t.} \\
&\qquad y_{ij} = y_{ji} && \forall\ \langle i, j \rangle \in E \\
&\qquad y_{ij} \in \{0, 1\} && \forall\ (i, j) \in A,\ h \in H \\
&\text{where } x_{ij}^k \text{ solve:} \\
&\min_{x_{ij}^k} \quad \sum_{k \in K} \sum_{(i,j) \in A} n^k c_{ij} x_{ij}^k \qquad (8) \\
&\text{s.t.} \\
&\qquad \sum_{\{j:(i,j) \in A\}} x_{ij}^k - \sum_{\{j:(j,i) \in A\}} x_{ji}^k = e_i^k \ \ \forall\ i \in N,\ k \in K \\
&\qquad x_{ij}^k \le y_{ij} && \forall\ (i, j) \in A,\ k \in K \\
&\qquad x_{ij}^k \in \{0, 1\} && \forall\ (i, j) \in A,\ k \in K,
\end{aligned}
$$

Note that since the network is assumed to be undirected the closing or opening decisions are defined on the set E of undirected links of the given network. Therefore, in the model it is assumed $y_{ij} = y_{ji}$ for each $\langle i, j \rangle \in E$.

The heuristic method, originally developed by Erkut and Gzara (2008) for the case with $w_1 = 0$, works as follows. First, it computes the subnetwork that minimizes the total risk objective (i.e., that optimizes the leader's objective function). Note that this simply corresponds to determining a minimum risk path from origin node s^k to destination node t^k, for each carrier $k \in K$. Indeed, if this subnetwork is passed to the carriers they may select paths different from the ones that the authority has in mind, with the consequence that the resulting total risk may be different and, hence, larger than the minimum value. Of course, in this case we are sure that the subnetwork is not stable. The total risk values being the same is a necessary but not sufficient condition for stability; in order to check stability of the current subnetwork a multi-commodity integer network flow problem with an upper bound constraint on the cost (equal to the transportation cost of the carriers solution) and with the maximization of the total risk as objective is solved. Then, at each iteration, if the current subnetwork is not stable, the heuristic selects and eliminates an edge from the network, iterating the process until the resulting restricted network is stable. Two selection rules are proposed: the first one either selects the high-risk edge or the one selected by the carriers which is not in the minimum risk subnetwork; the other one selects the edge with the highest reduced risk, where the edge reduced risk is the increase in total risk when the edge is included in the solution network.

The heuristic algorithm was implemented in Matlab 6.5 and the network flow problem solved using CPLEX 8.1, on a Sun Ultra Sparc 450 workstation. The method is evaluated on a network with 105 nodes and 134 undirected links representing the city of Ravenna, Italy, and by considering 35 carriers, capable to find a stable heuristic solution within 30 s of CPU time.

The Model of Verter and Kara (2008)

The HTND models above discussed are link-based, since both the regulator's and carriers' decision variables are defined on the links of the network. Verter and Kara (2008) provide an alternative path-based formulation for the HTND problem, where some of the decision variables are defined on the paths of the network. The authors give two motivations for proposing such a model. The first one is methodological: the resulting model is a single-level optimization problem, which, therefore, does not suffer from the stability problem. The other reason is related to the application: according to the proposed model, hazmat carriers are engaged by the authority in the route planning decision making process. In particular, the carriers are asked to provide to the authority an ordered list of (best) paths that are acceptable from the carriers' point of view in order to identify compromise solutions between the two parties.

Assuming the road transportation network being represented by a directed network $G(V, A)$, let P^k be the ordered set of the best or acceptable paths for carrier k, with the paths listed in non-increasing order with respect to the carrier preferences (e.g., the minimization of transportation cost). Let $p_\ell^k \in P^k$ be the ℓth best (e.g., shortest) path of carrier k. In the following, each path p_ℓ^k is defined as a sequence of m_ℓ^k links of the network.

The authority decides which is the subset of links of the network to be opened to the transportation of hazmat of type h, for each $h \in H$. The open links determines which paths in P^k are available to carrier k, that, therefore, will select the first available path among the ordered paths of P^k.

Three sets of binary variables are defined in the model of Verter and Kara (2008):

- y_{ij}^h, being equal to 1 if link $(i, j) \in A$ is open for transport of hazmat of type h, 0 otherwise.
- z_ℓ^k, being equal to 1 if path $p_\ell^k \in P^k$ is available for carrier k, 0 otherwise.
- χ_ℓ^k, being equal to 1 if path $p_\ell^k \in P^k$ is used by carrier k, 0 otherwise.

The mathematical model is as follows.

$$
\begin{aligned}
\min \quad & \sum_{k \in K} \sum_{p_\ell^k \in P^k} \sum_{(i,j) \in p_\ell^k} n^k \rho_{ij}^{h(k)} \chi_\ell^k \\
\text{s.t.} \quad & \\
& \sum_{p_\ell^k \in P^k} \chi_\ell^k = 1 && \forall\, k \in K \\
& \chi_\ell^k \le z_\ell^k && \forall\, k \in K,\ p_\ell^k \in P^k \\
& z_\ell^k \le y_{ij}^{h(k)} && \forall\, k \in K,\ p_\ell^k \in P^k,\ (i,j) \in p_\ell^k \\
& z_\ell^k \ge \sum_{(i,j) \in p_\ell^k} y_{ij}^{h(k)} - m_\ell^k + 1 && \forall\, k \in K,\ p_\ell^k \in P^k \\
& \chi_\ell^k \ge z_\ell^k - \sum_{v=1}^{\ell-1} z_v^k && \forall\, k \in K,\ p_\ell^k \in P^k \\
& \chi_\ell^k, z_\ell^k \in \{0, 1\} && \forall\, k \in K,\ p_\ell^k \in P^k \\
& y_{ij}^h \in \{0, 1\} && \forall\, (i, j) \in A,\ h \in H.
\end{aligned}
\tag{9}
$$

Note that the model is a single-level Integer Programming (IP) problem. In fact, since carriers' behavior in path selection as response to the decision of the authority is explicitly represented by the ordered set of acceptable paths, it can be directly integrated into the decision model of the authority. Therefore, the model structure is no longer bilevel but single-level. The objective function to be minimized is the total risk due to the decisions of the carriers (that the authority knows in advance on the basis of the preferences that the carriers have submitted to it) in response to those of the authority. The first set of constraints guarantees that each carrier k will select a single path among the preferred paths of P^k. The second, third, and fourth sets of constraints couple variables y_{ij}^h with χ_ℓ^k through variables z_ℓ^k. In particular, the second set of constraints ensures that a path can be used only if it is available,

while the third and fourth sets of constraints identify the available paths on the basis of the decisions of the authority (the former force a path to be unavailable if one of its link is not open, while the latter declare available each path whose links are all open). The fifth constraint set ensures that the path with the smallest index among the available ones is used by each carrier. The remaining set of constraints is the binary constraints on the values of the variables.

Moreover, note that if $|P^k| = 1$, for each carrier k, the authority is unable to intervene: this corresponds to the unregulated scenario. On the contrary, if all the paths from s^k to t^k are acceptable for each carrier k, we have the ideal scenario for the authority. All other intermediate cases produce compromise solutions between the above two extremes. For example, one may think to consider, for each carrier, κ best (shortest) paths or all the paths that are within a maximum allowable percent detour from the shortest path. Verter and Kara (2008) observe that for the case of a single carrier per hazmat type, the increase in the number of available paths will lead to inferior routes for the carrier, since the authority has more chances to mitigate the risk inducing the carrier to use a less risky path. This is not the case with multiple carriers shipping the same hazmat type. In fact, a link that may be closed by the authority in a certain scenario, could on the opposite result be opened in a less constrained one (i.e., when a larger set of acceptable paths are considered for each carrier) with the consequence of a possible reduction of not only the total risk but also of the carriers' total cost.

Note that as for the link-based model (1) there is no interaction among shipments of different hazmat types, and, hence, the path-based model (9) is separable in as many independent problems as the number of hazmat types.

The authors experiment the proposed model with the use of a commercial IP solver (CPLEX 6.0) on a network of 48 nodes and 57 links representing the road network of Western Ontario, Canada (the same considered in Kara and Verter 2004), and with 53 carriers and 2 hazmat types. The solver requires more than 2.5 h of CPU time (the authors does not give any detail about the type of computer used for the experimentation) to return the optimal solution when the shortest 100 distinct paths for each carrier are considered, and the CPU time does not show a monotone trend as the number of paths increases. The authors also consider a large network with 176 nodes and 205 links, representing the highways of Quebec and Ontario, Canada, 84 carriers and 2 hazmat types, solving the instance with much larger path sets within 5 h.

HTND: A Model with Local and Regional Authorities

All the modeling approaches discussed above consider the role of the authority (the regulator of the infrastructure, i.e, the road network) and of the carriers (the users of the infrastructure), without taking into account equity in risk distribution over the population.

The first paper dealing with the equity issue is by Bianco et al. (2009). Differently from the other models, they do not consider the role of the carriers, and assume that

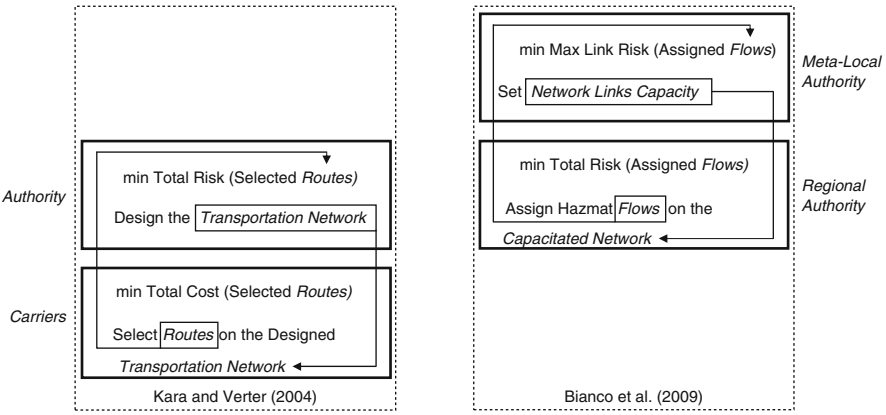


Fig. 2 Comparison between the models of Bianco et al. (2009) and that of Kara and Verter (2004)

there are distinct government authorities (at different levels, e.g., regional and local) that regulate the hazmat traffic by restricting the use of network links to the hazmat shipments. The aim is twofold: on the one hand the minimization of the total risk of the shipments, on the other hand the spreading of the risk equitably over the geographical region in which the transportation network is embedded.

The reason of such a modeling approach is that the local and the regional authorities act as multiple decision makers and in many cases they do not cooperate (e.g., because they are controlled by different parties, as often happens in Italy). Hence, a simple multi-objective single-level model does not adequately represent such a case, while the bilevel model better represents the scenario where there is a hierarchy of decision makers: the leader (i.e., the meta-local authority) and the follower (i.e., the regional authority). The leader tries to minimize the maximum link total risk (hence balancing the risk on the links of the network due to hazmat shipments, and then achieving in this way risk equity), imposing some restrictions on the amount of hazmat traffic over the links of the network in terms of link capacities. In such a way, the leader balances the link total risk on the links of the network due to hazmat shipments, achieving the risk equity. The follower decision maker instead has the freedom to choose the specific amounts of hazmat traffic to be routed over the links of the capacitated network in order to pursue the minimization of the network total risk.

In Fig. 2, we show the conceptual comparison between the model of Bianco et al. (2009) and that of Kara and Verter (2004). In particular, the model of Bianco et al. (2009) corresponds to a particular hazmat network design problem, where the goal is not determining a subgraph of the whole transportation road network, but finding (link) capacities leading to a balanced risk over the population as evenly as possible.

Let the road transportation network be represented by an undirected network $G = (N, E)$, with E being the set of undirected links or edges, and $A =$

$\{(i,j),(j,i)|\langle i,j\rangle \in E\}$ being the corresponding set of arcs considering each edge traversed in both directions. Two sets of variables are defined:

- u_{ij}, being the (bundle) capacity of arc $(i,j) \in A$ that limits the total amount of hazmat traversing edge $\langle i,j\rangle \in E$ from node i to node j.
- x_{ij}^k, being the fraction of the amount of hazmat of the shipment of carrier k traversing edge $\langle i,j\rangle \in E$ from node i to node j.

Let

$$\eta_{ij} = \sum_{k\in K} \frac{b^k}{q^k}\left(\rho_{ij}^{h(k)} x_{ij}^k + \rho_{ji}^{h(k)} x_{ji}^k\right)$$

be the link total risk over the population located in the neighbor of link $\langle i,j\rangle \in E$, and let η be the maximum link total risk among the η_{ij} values of each link $\langle i,j\rangle \in E$; moreover, let

$$R_{tot} = \sum_{k\in K} \frac{b^k}{q^k} \sum_{(i,j)\in A} \rho_{ij}^{h(k)} x_{ij}^k$$

be the network total risk over the population residing in the area embedding the road network represented by G.

According to the conceptual scheme of the model, the bundle capacities u_{ij}, with $(i,j) \in A$, are the variables controlled by the leader decision maker, who wants to minimize the value of η by imposing specific limits on the sum of hazmat flows $(b^k x_{ij}^k)$ on the arcs of the network. The follower controls variables x_{ij}^k in such a way to minimize the value of R_{tot}.

The mathematical bilevel model is

$$\eta^* = \min_{u_{ij}} \ \eta$$

s.t.

$$\sum_{k\in K} \frac{b^k}{q^k}\left(\rho_{ij}^{h(k)} x_{ij}^k + \rho_{ji}^{h(k)} x_{ji}^k\right) \le \eta \ \forall \ \langle i,j\rangle \in E$$

$$u_{ij} \ge 0 \qquad \forall \ (i,j) \in A$$

where x_{ij}^k solve:

$$R_{tot}^*(\mathbf{u}) = \min_{x_{ij}^k} \sum_{k\in K} \frac{b^k}{q^k} \sum_{(i,j)\in A} \rho_{ij}^{h(k)} x_{ij}^k \tag{10}$$

s.t.

$$\sum_{\{j:(i,j)\in A\}} x_{ij}^k - \sum_{\{j:(j,i)\in A\}} x_{ji}^k = e_i^k \ \forall \ i \in N,\ k \in K$$

$$\sum_{k\in K} b^k x_{ij}^k \le u_{ij} \qquad \forall \ (i,j) \in A,\ k \in K$$

$$x_{ij}^k \ge 0 \qquad \forall \ (i,j) \in A,\ k \in K.$$

Let η^* be the optimal solution value of the leader problem, that is the minimum possible value for η. The first set of constraints forces the link total risk over each

link $\langle i, j \rangle \in E$ to be not greater than η, and the second set is composed of the non-negativity constraints on the arc bundle capacities.

Let $R^*_{tot}(\mathbf{u})$ be the optimal solution value of the follower problem, that is the minimum network total risk value given the bundle capacity vector $\mathbf{u}$. The first set of constraints of this problem states the flow balance requirements on each network node and for each carrier, and the second set forces the total flow on each arc $(i, j) \in A$ to be not greater than the arc capacity u_{ij}. Finally, the third set contains the non-negativity constraints on the follower's variables.

Note that both the leader's and follower's problems are linear. Taking advantage of the linearity of the follower's problem, Bianco et al. (2009) reduce the bilevel problem to a single-level optimization problem by adding to the leader's problem the KKT optimality conditions of the follower's problem.

More in detail, consider the following additional variables:

- π_i^k, free dual variables associated with the first set of primal constraints of the follower's problem, with $i \in N$, and $k \in K$.
- λ_{ij}, non-negative dual variables associated with the second set of primal constraints of the follower's problem, with $(i, j) \in A$.
- γ_{ij}, non-negative slack variables of the second set of primal constraints of the follower's problem, with $(i, j) \in A$.
- v_{ij}^k, non-negative slack variables of the dual constraints of the follower's problem, with $(i, j) \in A$, and $k \in K$.

The KKT optimality conditions of the follower's problem are:

$$
\begin{array}{ll}
\displaystyle\sum_{\{j:(i,j)\in A\}} x_{ij}^k - \sum_{\{j:(j,i)\in A\}} x_{ji}^k = e_i^k & \forall\, i \in N,\ k \in K \\
\displaystyle\sum_{k\in K} b^k x_{ij}^k + \gamma_{ij} = u_{ij} & \forall\, (i,j) \in A \\
\displaystyle\frac{b^k}{q^k}\rho_{ij}^{h(k)} - \pi_i^k + \pi_j^k - v_{ij}^k + \lambda_{ij} = 0 & \forall\, (i,j) \in A,\ k \in K \\
x_{ij}^k\ v_{ij}^k = 0 & \forall\, (i,j) \in A,\ k \in K \\
\lambda_{ij}\ \gamma_{ij} = 0 & \forall\, (i,j) \in A \\
x_{ij}^k \geq 0, v_{ij}^k \geq 0 & \forall\, (i,j) \in A,\ k \in K \\
\lambda_{ij} \geq 0, \gamma_{ij} \geq 0 & \forall\, (i,j) \in A \\
\pi_i^k \text{ free} & \forall\, i \in N,\ k \in K.
\end{array}
\tag{11}
$$

The first two set of constraints are the primal constraints that together with the non-negativity conditions on the primal variables x_{ij}^k and slack variables γ_{ij} form the primal feasibility conditions; the third set represents the dual constraints that together with the restrictions in sign on the dual variables λ_{ij} and the slack variables v_{ij}^k of the dual constraints define the dual feasibility conditions; finally, the fourth and fifth sets of constraints are the complementary slackness conditions.

Note that adding the above conditions to the leader problem reduces the bilevel problem to a single-level one, producing a single-level non-linear optimization

problem, where the non-linearity is due to the following bilinear constraints

$$\begin{aligned} x_{ij}^k \ v_{ij}^k &= 0 \ \forall \ (i,j) \in A, \ k \in K \\ \lambda_{ij} \ \gamma_{ij} &= 0 \ \forall \ (i,j) \in A. \end{aligned} \tag{12}$$

Bianco et al. (2009) propose to linearize these constraints as follows, by introducing binary variables δ_{1ij}^k, δ_{2ij}^k, δ_{3ij}, and δ_{4ij}, and large scalars M_1, M_2, M_3, and M_4:

$$\begin{aligned} x_{ij}^k &\leq M_1 \ \delta_{1ij}^k \quad \forall \ (i,j) \in A, \ k \in K \\ v_{ij}^k &\leq M_2 \ \delta_{2ij}^k \quad \forall \ (i,j) \in A, \ k \in K \\ \lambda_{ij} &\leq M_3 \ \delta_{3ij} \quad \forall \ (i,j) \in A \\ \gamma_{ij} &\leq M_4 \ \delta_{4ij} \quad \forall \ (i,j) \in A \\ \delta_{1ij}^k + \delta_{2ij}^k &\leq 1 \ \forall \ (i,j) \in A, \ k \in K \\ \delta_{3ij} + \delta_{4ij} &\leq 1 \ \forall \ (i,j) \in A. \end{aligned} \tag{13}$$

The values of M_2 and M_3 should be sufficiently large, while we can set $M_1 = 1$, since $x_{ij}^k \leq 1$, and $M_4 = \sum_{k \in K} b^k$, since $\gamma_{ij} \leq u_{ij}$ and the latter can be bounded above by $\sum_{k \in K} b^k$.

Finally, the single-level MIP formulation for the model proposed by Bianco et al. (2009) is:

$$\eta^* = \min \eta$$

s.t.

$$\begin{aligned} &\sum_{k \in K} \frac{b^k}{q^k} \left(\rho_{ij}^{h(k)} x_{ij}^k + \rho_{ji}^{h(k)} x_{ji}^k \right) \leq \eta && \forall \ \langle i,j \rangle \in E \\ &\sum_{\{j:(i,j) \in A\}} x_{ij}^k - \sum_{\{j:(j,i) \in A\}} x_{ji}^k = e_i^k && \forall \ i \in N, \ k \in K \\ &\sum_{k \in K} b^k x_{ij}^k + \gamma_{ij} = u_{ij} && \forall \ (i,j) \in A, \ k \in K \\ &\frac{b^k}{q^k} \rho_{ij}^{h(k)} - \pi_i^k + \pi_j^k - v_{ij}^k + \lambda_{ij} = 0 && \forall \ (i,j) \in A, \ k \in K \\ &x_{ij}^k \leq M_1 \ \delta_{1ij}^k && \forall \ (i,j) \in A, \ k \in K \\ &v_{ij}^k \leq M_2 \ \delta_{2ij}^k && \forall \ (i,j) \in A, \ k \in K \\ &\lambda_{ij} \leq M_3 \ \delta_{3ij} && \forall \ (i,j) \in A \\ &\gamma_{ij} \leq M_4 \ \delta_{4ij} && \forall \ (i,j) \in A \\ &\delta_{1ij}^k + \delta_{2ij}^k \leq 1 && \forall \ (i,j) \in A, \ k \in K \\ &\delta_{3ij} + \delta_{4ij} \leq 1 && \forall \ (i,j) \in A \\ &x_{ij}^k \geq 0, v_{ij}^k \geq 0 && \forall \ (i,j) \in A, \ k \in K \\ &u_{ij} \geq 0, \lambda_{ij} \geq 0, \gamma_{ij} \geq 0 && \forall \ (i,j) \in A \\ &\pi_i^k \text{ free} && \forall \ i \in N, \ k \in K \\ &\delta_{1ij}^k \in \{0,1\}, \delta_{2ij}^k \in \{0,1\} && \forall \ (i,j) \in A, \ k \in K \\ &\delta_{3ij} \in \{0,1\}, \delta_{4ij} \in \{0,1\} && \forall \ (i,j) \in A. \end{aligned} \tag{14}$$

The authors use a commercial solver (CPLEX 8.0) to solve the MIP (14) capable to find the optimal solution within 3 h of CPU time on a PC with Pentium IV processor and 2 GB of RAM considering a network with 311 nodes and 441 undirected links (882 arcs), representing the road network of the Lazio region, Italy, and with 3 shipments (carriers).

Solving problem (14) returns, besides the optimal solution $\mathbf{u}^* = \{u_{ij}^*\}$ of the bilevel model which is the best leader choice, also the best (from the leader point of view) multi-commodity flow assignment $b^k\mathbf{x}^B = \{b^k x_{ij}^{*k}\}$ that the follower may choose among the minimum network total risk flow assignments given bundle link capacities $\{u_{ij}^*\}$, that is, among all his/her indifferent choices over the capacitated network established by the leader. The optimal solution value η^* of the bilevel model is equal to the maximum link total risk induced by (let us say, *optimistic*) flow assignment $b^k\mathbf{x}^B$, i.e, the best one in terms of the leader's objective. Nevertheless, there is no guarantee that the follower will adopt the flow assignment $b^k\mathbf{x}^B$ if there are multiple minimum network total risk flow assignments on the capacitated network. In this case, the optimal solution $\mathbf{u}^*$ of the bilevel model might be unstable. This situation occurs when there is another (optimal in terms of network total risk) flow assignment $b^k\mathbf{x}'$ different from $b^k\mathbf{x}^B$, inducing a maximum link total risk of value η' greater than η^*, and equal at most to the *pessimistic* solution value $\eta^W(\mathbf{u}^*)$ obtained when the follower among his multiple optimal solutions chooses the worst one in terms of the leader's objective.

The pessimistic value $\eta^W(\mathbf{u}^*)$ can be computed by solving the multi-commodity network flow problem (15) with bundle arc capacity vector $\mathbf{u}$, an upper bound $\widehat{R}_{tot}$ on the network total risk, where the objective to be maximized is link total risk η_{ij} over the population in the neighbor of link $\langle i, j\rangle \in E$. This is a Linear Programming (LP) problem that can be formulated as follows

$$\eta_{ij}^W(\mathbf{u}, \widehat{R}_{tot}) = \max \sum_{k\in K} \frac{b^k}{q^k}\left(\rho_{ij}^{h(k)} x_{ij}^k + \rho_{ji}^{h(k)} x_{ji}^k\right)$$

s.t.

$$\begin{aligned}
&\sum_{k\in K} \frac{b^k}{q^k} \sum_{(i,j)\in A} \rho_{ij}^{h(k)} x_{ij}^k \le \widehat{R}_{tot} \\
&\sum_{\{j:(i,j)\in A\}} x_{ij}^k - \sum_{\{j:(j,i)\in A\}} x_{ji}^k = e_i^k \quad \forall\, i \in N,\ k \in K \\
&\sum_{k\in K} b^k x_{ij}^k \le u_{ij} \qquad\qquad \forall\, (i,j) \in A,\ k \in K \\
&x_{ij}^k \ge 0 \qquad\qquad \forall\, (i,j) \in A,\ k \in K.
\end{aligned} \tag{15}$$

For each link $\langle i, j\rangle \in E$, problem (15) is solved with $\mathbf{u} = \mathbf{u}^*$ and $\widehat{R}_{tot} = R_{tot}^*(\mathbf{u}^*)$, with $R_{tot}^*(\mathbf{u}^*)$ being the network total risk value of the minimum-total risk multi-commodity network flow on the capacitated network G with bundle arc capacities vector $\mathbf{u}^*$; that is, $R_{tot}^*(\mathbf{u}^*) = \sum_{k\in K} \frac{b^k}{q^k} \sum_{(i,j)\in A} \rho_{ij}^{h(k)} x_{ij}^{*k}$ is the optimal solution value of the follower problem given the arc capacities vector $\mathbf{u}^*$

fixed by the leader (see the original bilevel model (10)). Let $\eta_{ij}^W(\mathbf{u}^*, R_{tot}^*(\mathbf{u}^*))$ be the optimal solution value of problem (15), for link $\langle i, j\rangle \in E$, $\eta^W(\mathbf{u}^*) = \max_{\langle i,j\rangle \in E}\{\eta_{ij}^W(\mathbf{u}^*, R_{tot}^*(\mathbf{u}^*))\}$.

If the gap between $\eta^W(\mathbf{u}^*)$ and η^* is too large, the leader may prefer to (heuristically) find a stable feasible bundle arc capacity assignment. This pushes Bianco et al. (2009) to devise a heuristic approach for finding a stable feasible solution of the bilevel model, inspired by the heuristic proposed by Erkut and Gzara (2008).

Heuristic Approach

Bianco et al. (2009) propose an iterative algorithm that at each iteration constructs a feasible solution $\mathbf{u}^H = \{u_{ij}\}$ of the bilevel model, and tests its stability. The latter is checked by comparing the (optimistic) heuristic solution value $\eta(\mathbf{u}^H)$ with the pessimistic value $\eta^W(\mathbf{u}^H)$ computed following the strategy described above.

If the current solution $\mathbf{u}^H$ is stable (i.e., $\eta(\mathbf{u}^H) = \eta^W(\mathbf{u}^H)$), the algorithm stops; otherwise, the algorithm removes a link from the network and starts a new iteration on the residual network. When the algorithm stops, the stable heuristic solution $\mathbf{u}^H$ on the original network is obtained by setting $u_{ij} = u_{ji} = 0$, for each removed link $\langle i, j\rangle \in E$ from the original network.

At each iteration the heuristic solution is computed by solving the following uncapacitated multi-commodity network flow problem

$$\eta^L(\omega) = \min\left(\eta + \omega \sum_{k\in K} \frac{b^k}{q^k} \sum_{(i,j)\in A} \rho_{ij}^{h(k)} x_{ij}^k\right)$$

s.t.

$$\begin{aligned}
&\sum_{k\in K} \frac{b^k}{q^k}\left(\rho_{ij}^{h(k)} x_{ij}^k + \rho_{ji}^{h(k)} x_{ji}^k\right) \le \eta \;\; \forall\, \langle i, j\rangle \in E \\
&\sum_{\{j:(i,j)\in A\}} x_{ij}^k - \sum_{\{j:(j,i)\in A\}} x_{ji}^k = e_i^k \;\; \forall\, i \in N,\; k \in K \\
&x_{ij}^k \ge 0 \qquad\qquad \forall\, (i, j) \in A,\; k \in K,
\end{aligned} \tag{16}$$

with $\omega = 1/\sum_{k\in K} \frac{b^k}{q^k} \sum_{(i,j)\in A} \rho_{ij}^{h(k)}$.

Note that, with this value of ω, the objective function η and R_{tot} are minimized in lexicographical order (i.e., among the solutions with minimum η value, the one of minimum R_{tot} value is found), and the optimal solution corresponds to the best hazmat flow assignment that the leader may obtain, assuming that he/she can directly control the decisions of the follower. In particular, let $\widehat{\mathbf{x}} = \{\widehat{x}_{ij}^k\}$ be the optimal solution, and $\widehat{\eta} = \max_{\langle i,j\rangle \in E}\{\sum_{k\in K} \frac{b^k}{q^k}(\rho_{ij}^{h(k)}\widehat{x}_{ij}^k + \rho_{ji}^{h(k)}\widehat{x}_{ji}^k)\}$ be the related maximum link total risk. The heuristic solution $\mathbf{u}^H = \{u_{ij}^H\}$ is constructed from

solution $\widehat{\mathbf{x}}$ by setting the bundle capacity value u_{ij}^H of arc $(i, j) \in A$ equal to the total hazmat flow on that arc, that is, $u_{ij}^H = \sum_{k \in K} b^k \widehat{x}_{ij}^k$. Let $\eta^H = \widehat{\eta}$ be the solution value related to solution $\mathbf{u}^H$.

In the following, it is shown the criterion for the selection of the link to be removed, and proved that after a number of iterations less than the number of network links the algorithms stops with a feasible stable solution for the residual network.

Consider the current heuristic solution $\mathbf{u}^H$. Assume, therefore, that the leader imposes the bundle capacities $\{u_{ij}\} = \mathbf{u}^H$ on the arcs of the network. The follower finds an optimal flow assignment on the capacitated network that minimizes the network total risk; this corresponds to solve the follower problem. Let $R_{tot}^*(\mathbf{u}^H)$ be the optimal solution value of this latter problem, that is, the value of the minimum network total risk on the capacitated network with bundle link capacities vector $\mathbf{u}^H$.

For each edge $\langle i, j \rangle \in E$, let $\eta_{ij}^W(\mathbf{u}^H, R_{tot}^*(\mathbf{u}^H))$ be the maximum link total risk over link $\langle i, j \rangle$ among all the optimal solutions of the follower problem on the capacitated network with bundle arc capacity vector $\mathbf{u}^H$; recall that $\eta_{ij}^W(\mathbf{u}^H, R_{tot}^*(\mathbf{u}^H))$ is the optimal solution value of problem (15), with $\mathbf{u} = \mathbf{u}^H$ and $R_{tot} = R_{tot}^*(\mathbf{u}^H)$.

If solution $\mathbf{u}^H$ is not stable, there is a link $\langle i, j \rangle \in E$ for which $\eta_{ij}^W(\mathbf{u}^H, R_{tot}^*(\mathbf{u}^H))$ is greater than η^H; let us assume that $\langle i', j' \rangle$ is the link for which $\eta_{i'j'}^W(\mathbf{u}^H, R_{tot}^*(\mathbf{u}^H)) = \max_{\langle i,j \rangle \in E}\{\eta_{ij}^W(\mathbf{u}^H, R_{tot}^*(\mathbf{u}^H))\}$ and that $\eta_{i'j'}^W(\mathbf{u}^H, R_{tot}^*(\mathbf{u}^H)) > \eta^H$, with $\eta_{ij}^W(\mathbf{u}^H, R_{tot}^*(\mathbf{u}^H))$ being induced by the follower optimal flow assignment over the capacitated network with bundle arc capacity vector $\mathbf{u}^H$.

In order to eliminate the difference between $\eta_{i'j'}^W(\mathbf{u}^H, R_{tot}^*(\mathbf{u}^H))$ and η^H, the algorithm removes link $\langle i', j' \rangle$ from the network and starts a new iteration where it searches for a new feasible solution of the bilevel model on the residual network. Note that the following theorem holds in the residual network.

Theorem 1 (Bianco et al. 2009). *Given a heuristic solution $\mathbf{u}^H$ of value η^H, if there is a link $\langle i, j \rangle \in E$ such that $\eta_{ij}^W(\mathbf{u}^H, R_{tot}^*(\mathbf{u}^H)) > \eta^H$, in the residual network obtained after the removal of link $\langle i, j \rangle \in E$ from the network, there is at least one path connecting the origin node s^k to the destination node t^k, for each carrier $k \in K$.*

Theorem 2 (Bianco et al. 2009). *The heuristic algorithm always stops with a stable heuristic solution $\mathbf{u}^H$.*

Also the heuristic algorithm (implemented in the C language) is experimented by Bianco et al. (2009) running the algorithm on the same PC and experimenting on the same network of the Lazio region by considering realistic instances with at most 30 carriers. Heuristic algorithm running times are quite limited: indeed, they are never greater than 7 min over all the instances tested. Note that, for each iteration, the heuristic algorithm solves $O(m)$ linear programs. In order to evaluate the effectiveness of the heuristic, the authors compare the heuristic solution values with the optimal solution values of the linear relaxation of the single-level MIP

problem (14). In particular, the heuristic algorithm was able to give always an optimal stable solution in all the test cases, that is, with the best (minimum) possible maximum link total risk.

Toll Setting in Hazmat Transportation

Toll Setting (TS) is a policy tool that can also be applied to hazmat transportation to regulate hazmat shipments. In the network design model it is assumed that the regulator has the authority to forbid a link of the network to the carriers, in order to preserve the most populated roads. Toll setting is an alternative policy where the authority sets tolls on all (or on a subset of) the links of the network, in order to deter the carriers to use certain roads and encourage them to use the less populated ones.

The Model of Marcotte et al. (2009)

Marcotte et al. (2009) extend the toll setting modeling approach of Labbé et al. (1998) to deal with hazmat transportation. The proposed model is the following NP-hard bilevel problem:

$$
\begin{aligned}
&\min \quad \sum_{k\in K}\sum_{(i,j)\in A} n^k(\rho_{ij}^{h(k)} + w_1(c_{ij} + t_{ij}^{h(k)}))x_{ij}^k \\
&\text{s.t.} \quad t_{ij}^h \geq 0 \qquad\qquad \forall (i,j)\in A,\ h\in H \\
&\text{where } x_{ij}^k \text{ solve:} \\
&\min \quad \sum_{k\in K}\sum_{(i,j)\in A} n^k(c_{ij} + t_{ij}^{h(k)} + w_2\rho_{ij}^{h(k)})x_{ij}^k \qquad (17)\\
&\text{s.t.} \\
&\qquad \sum_{\{j:(i,j)\in A\}} x_{ij}^k - \sum_{\{j:(j,i)\in A\}} x_{ji}^k = e_i^k \qquad \forall i\in N,\ k\in K \\
&\qquad x_{ij}^k \in \{0,1\} \qquad\qquad \forall (i,j)\in A,\ k\in K,
\end{aligned}
$$

where t_{ij}^h and x_{ij}^k are the variables controlled by the leader and the followers, respectively, with variables x_{ij}^k and the other notations having the same meaning as in the HTND problem (1). In particular,

- t_{ij}^h are non-negative variables representing the tolls imposed by the authority on arc $(i,j)\in A$ for each truck transporting hazmat of type $h\in H$.
- x_{ij}^k are binary variables modeling the decisions of the carriers and equal to 1 if arc $(i,j)\in A$ is used by carrier k for the shipment, 0 otherwise.

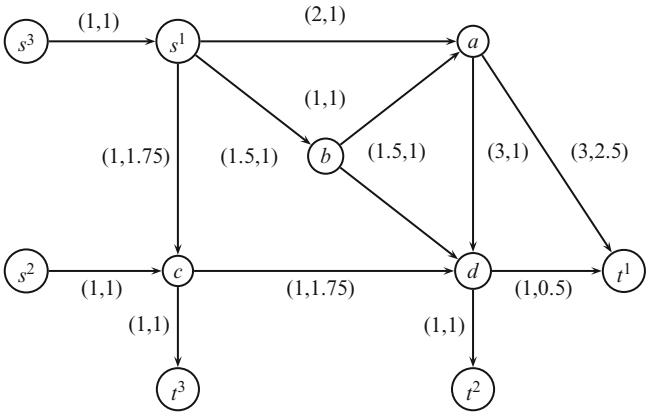


Fig. 3 An example of hazmat transportation network

In this model the authority sets the tolls on the arcs in such a way to minimize a weighted combination of population exposure and carriers' cost keeping into account that, for a certain value of the tolls, the carriers choose the minimum cost routes.

It can be show that HTND model (1) and TS model (17) are equivalent in case there is a single carrier or when each carrier transports a different type of hazmat, since the problem is separable by hazmat type. However, whenever a type of hazmat is transported by more than one carrier, model (17) is more flexible than model (1). Indeed, by setting high-enough tolls, model (17) gives the same solutions of model (1).

Consider as example the small network of Fig. 3, involving three carriers (i.e., $K = \{1, 2, 3\}$) all shipping the same hazmat type $h = 1$. Let $(s^k,\ t^k)$ be the origin-destination pair of the shipment of carrier k. Assume that carrier 1 needs two trucks for its shipment and each one of the others only one truck (i.e. $n^1 = 2, n^2 = n^3 = 1$). On each arc (i, j), the couple (c_{ij}, ρ^1_{ij}) is reported. Note that for both carriers 2 and 3 there is a single path connecting the origin to the destination of the shipment: that is, path $(s^2 \to c \to d \to t^2)$ and path $(s^3 \to s^1 \to c \to t^3)$ for carriers 2 and 3, respectively; these two paths pass through link (c, d) and (s^1, c), respectively, where there is the highest population exposure (risk). Assume that the authority's sole aim is the minimization of the risk (i.e. $w_1 = 0$), and the carriers only want to minimize their traveling cost (i.e. $w_2 = 0$). Then, for the pair (s^1, t^1) the most attractive path for the carrier is the minimum cost path, that is $s^1 \to c \to d \to t^1$. If carrier 1 follows this path (i.e., in the *unregulated* scenario) the total risk of the network will be 15.5. On the other hand (let us say in the *over-regulated* scenario), the authority would choose a minimum risk path for carrier 1, that is, equivalently either path $(s^1 \to a \to d \to t^1)$ or path $(s^1 \to b \to d \to t^1)$, getting the (minimum) network total risk of value 12.5. With a network design policy, the authority would close arc (d, t^1) in order to prevent carrier 1 from using the larger risky path $(s^1 \to c \to d \to t^1)$ (arcs (s^1, c) and (c, d) cannot be closed otherwise the other two carriers would

not have any path for their shipments). With this closure, carrier 1 would be forced to choose path ($s^1 \to a \to t^1$). Therefore with the network design policy the total risk on the network would be 14.5. With the toll setting strategy, the authority may set the tolls (e.g., sufficiently large on arcs (s^1, c) and/or (c, d)) so that carrier 1 chooses the minimum (transportation + toll) cost path ($s^1 \to b \to d \to t^1$) getting a network total risk equal to the minimum possible value, i.e. 12.5.

On this example it is shown that TS model (17) induces a lower risk than HTND model (1), while the opposite behavior cannot be obtained. This difference is due to the flexibility of model (17) that allows to differentiate among carriers: indeed it may happen that a carrier chooses a certain arc despite the toll, while if an arc is closed for the hazmat transportation of a given hazmat type, it becomes forbidden for all carriers transporting that type of hazmat.

As for the stability of solutions, in both models (1) and (17) ties may occur among inner-level solutions, and in this case the (optimistic) assumption is that the carriers choose the routes that are better for the authority. However, differently from the network design model (1), in model (17) ties can be broken by small perturbations of the tolls.

Another advantage of model (17) is that it allows the authority to easily find a toll setting inducing a minimum risk solution. In particular, a minimum risk solution can be obtained by solving the network flow problem

$$\begin{aligned} \min \ & \sum_{k \in K} \sum_{(i,j) \in A} n^k \rho_{ij}^{h(k)} x_{ij}^k \\ \text{s.t.} \ & \sum_{\{j:(i,j) \in A\}} x_{ij}^k - \sum_{\{j:(j,i) \in A\}} x_{ji}^k = e_i^k \quad \forall i \in N,\ k \in K \\ & x_{ij}^k \geq 0 \qquad\qquad \forall (i,j) \in A,\ k \in K. \end{aligned} \tag{18}$$

Given an optimal solution $\bar{x}$ of problem (18), the toll setting inducing $\bar{x}$ as flows chosen by the carriers can be found by inverse optimization. The idea is to choose the tolls in order that the values of the decision variables of the lower level are exactly the carriers' flows $\bar{x}$. This can be achieved by choosing any toll setting satisfying the following primal-dual optimality conditions of the inner problem of bilevel model (17):

$$\begin{aligned} & \sum_{\{j:(i,j) \in A\}} x_{ij}^k - \sum_{\{j:(j,i) \in A\}} x_{ji}^k = e_i^k && \forall\ i \in N,\ k \in K \\ & \pi_i^k - \pi_j^k - n^k t_{ij}^{h(k)} \leq n^k (c_{ij} + w_2 \rho_{ij}^{h(k)}) && \forall\ (i,j) \in A,\ k \in K \\ & x_{ij}^k (\pi_i^k - \pi_j^k - n^k t_{ij}^{h(k)} - n^k (c_{ij} + w_2 \rho_{ij}^{h(k)})) = 0 && \forall\ (i,j) \in A,\ k \in K \\ & x_{ij}^k \geq 0 && \forall\ (i,j) \in A,\ k \in K \\ & t_{ij}^h \geq 0 && \forall\ (i,j) \in A,\ h \in H \\ & \pi_i^k \text{ free} && \forall\ i \in N,\ k \in K, \end{aligned} \tag{19}$$

where π_i^k are the free dual variables related to the equality constraints of the inner problem, and variables x_{ij}^k instead of being binary are assumed to be continuous since the inner problem is a network flow problem.

Therefore, any toll setting satisfying conditions (19) with the flows equal to $\bar{x}$ is good for the authority. Among these feasible toll settings it can be chosen that one minimizing, for example, the sum of all the tolls paid by the carriers, that is the solution of the following LP problem:

$$
\begin{aligned}
&\min_{t,\pi} \sum_{k\in K}\sum_{(i,j)\in A} n^k \bar{x}_{ij}^k t_{ij}^{h(k)} \\
&\text{s.t.} \\
&\quad \pi_i^k - \pi_j^k - n^k t_{ij}^{h(k)} \le n^k (c_{ij} + w_2 \rho_{ij}^{h(k)}) && \forall (i,j) \in A,\ k \in K \\
&\quad \bar{x}_{ij}^k (\pi_i^k - \pi_j^k - n^k t_{ij}^{h(k)} - n^k (c_{ij} + w_2 \rho_{ij}^{h(k)})) = 0 && \forall (i,j) \in A,\ k \in K \\
&\quad t_{ij}^h \ge 0 && \forall (i,j) \in A,\ h \in H \\
&\quad \pi_i^k \text{ free} && \forall\ i \in N,\ k \in K.
\end{aligned}
\tag{20}
$$

Marcotte et al. (2009) prove that when all the arcs are subject to tolls, problem (20) is always feasible, implying that whenever the authority aims only at the minimization of the risk (i.e. $w_1 = 0$ in (17)) problem (17) is not a bilevel problem anymore, but reduces to an LP problem by inverse optimization. However, if the authority wants to keep into account also carriers' cost, problem (17) is a true bilevel optimization problem. In this case, inverse optimization can still be used to produce a "good" solution. The idea is to solve problem (18) where the objective function keeps into account also the carriers' costs:

$$
\begin{aligned}
&\min \sum_{k\in K}\sum_{(i,j)\in A} n^k (\rho_{ij}^{h(k)} + w_1 c_{ij}) x_{ij}^k \\
&\text{s.t.} \\
&\quad \sum_{\{j:(i,j)\in A\}} x_{ij}^k - \sum_{\{j:(j,i)\in A\}} x_{ji}^k = e_i^k \ \ \forall i \in N,\ k \in K \\
&\quad x_{ij}^k \ge 0 \qquad\qquad \forall (i,j) \in A,\ k \in K
\end{aligned}
\tag{21}
$$

and then use the produced solution $\bar{x}$ in (20). However, this strategy may lead to higher tolls on the carriers than the ones found by solving problem (17).

The strategy used in Marcotte et al. (2009) in order to solve problem (17) is to cast it as a single level problem by replacing the inner level problem with its primal-dual optimality conditions (19).

Then the single level reformulation of problem (17) is

$$\begin{aligned}
&\min \sum_{k\in K}\sum_{(i,j)\in A} n^k(\rho_{ij}^{h(k)} + w_1(c_{ij} + t_{ij}^{h(k)}))x_{ij}^k \\
&\text{s.t.} \\
&\sum_{\{j:(i,j)\in A\}} x_{ij}^k - \sum_{\{j:(j,i)\in A\}} x_{ji}^k = e_i^k && \forall i \in N,\ k \in K \\
&\pi_i^k - \pi_j^k - n^k t_{ij}^{h(k)} \le n^k(c_{ij} + w_2\rho_{ij}^{h(k)}) && \forall (i,j) \in A,\ k \in K \\
&x_{ij}^k(\pi_i^k - \pi_j^k - n^k t_{ij}^{h(k)} - n^k(c_{ij} + w_2\rho_{ij}^{h(k)})) = 0 && \forall (i,j) \in A,\ k \in K \\
&x_{ij}^k \ge 0 && \forall (i,j) \in A,\ k \in K \\
&t_{ij}^h \ge 0 && \forall (i,j) \in A,\ h \in H \\
&\pi_i^k \text{ free} && \forall\ i \in N,\ k \in K.
\end{aligned} \tag{22}$$

This problem can be solved by resetting the binary constraints on x, and linearizing the complementary constraints

$$x_{ij}^k(\pi_i^k - \pi_j^k - n^k t_{ij}^{h(k)} - n^k(c_{ij} + w_2\rho_{ij}^{h(k)})) = 0 \quad \forall (i,j) \in A,\ k \in K$$

by replacing them with the following set of constraints:

$$\begin{aligned}
&\pi_i^k - \pi_j^k - n^k t_{ij}^{h(k)} \ge n^k(c_{ij} + w_2\rho_{ij}^{h(k)}) - M_{ij}^k(1 - x_{ij}^k) && \forall (i,j) \in A,\ k \in K \\
&x_{ij}^k \in \{0,1\} && \forall (i,j) \in A,\ k \in K.
\end{aligned} \tag{23}$$

Moreover, bilinear terms $t_{ij}^{h(k)} x_{ij}^k$ of the leader's objective are linearized by introducing (new) variables τ_{ij}^k replacing the bilinear terms and adding the following set of constraints

$$\begin{aligned}
&\tau_{ij}^k \ge 0 && \forall (i,j) \in A,\ k \in K \\
&\tau_{ij}^k - M_{ij}^k x_{ij}^k \le 0 && \forall (i,j) \in A,\ k \in K \\
&\tau_{ij}^k - t_{ij}^{h(k)} \le 0 && \forall (i,j) \in A,\ k \in K \\
&\tau_{ij}^k - t_{ij}^{h(k)} - M_{ij}^k x_{ij}^k \ge -M_{ij}^k && \forall (i,j) \in A,\ k \in K,
\end{aligned} \tag{24}$$

with M_{ij}^k being sufficiently large positive scalars, to yield a MIP formulation.

An alternative MIP reformulation can be obtained by replacing the complementary slackness conditions, that is, the first set of constraints (23), with the following constraints imposing the equality on the values of the objective functions of the follower's primal and dual problems:

$$\sum_{(i,j)\in A} n^k\left(c_{ij} + w_2\rho_{ij}^{h(k)}\right)x_{ij}^k + \sum_{(i,j)\in A} n^k\tau_{ij}^k - \sum_{i\in N} e_i^k\pi_i^k = 0 \quad \forall k \in K \tag{25}$$

where the value of the variables τ_{ij}^k is properly set by constraints (24) to ensure that $\tau_{ij}^k = 0$ when $x_{ij}^k = 0$ and $\tau_{ij}^k = t_{ij}^{h(k)}$ when $x_{ij}^k = 1$.

The two MIP reformulations, requiring the same number of integer variables, are solved in Marcotte et al. (2009) by standard software (CPLEX 10.0). For the MIP formulation with equality of primal and dual objective values, constants M_{ij}^k are computed by solving for each carrier an LP relaxation of a maximum cost flow problem with cycle elimination constraints. For the complementary slackness based MIP reformulation, the bound on the constants M_{ij}^k are based on the ones computed for the other one, but their values are larger.

The two MIP problems are tested considering the same road network of the highway system of Western Ontario considered by Kara and Verter (2004), and considering either 53 shipments (carriers) neglecting shipments of less than 500 trucks, or the whole set of 287 shipments. The shipments involve two different hazmat types. The authors solve the two MIP reformulations of both HTND problem (1) and TS problem (17), and use the TS problem or its proxy (21) in order to warmly start the HTND problem. Indeed, they find a feasible solution of the HTND problem by solving a minimum-cost flow problem on a reduced network where all the arcs that are tolled in the optimal solution of TS (or of its proxy), but unused by a carrier, are removed. The extensive experiments carried out in Marcotte et al. (2009) lead to the following conclusions:

- The MIP reformulation based on the equality of primal and dual objective functions is better from a computational point of view both for HTND and TS problems.
- If the leader wants to minimize only the population exposure, it may help to set w_1 to a small positive value in order to encourage among equivalent minimum risk solutions the ones with lower cost for the carriers. If w_1 is set to zero, then TS problem (17) is not a true bilevel problem, but can be solved by inverse optimization. In this case the solution time required for solving problem (17) is much smaller and the warm starting of the MIP formulation of HTND problem (1) with the solution of problem (17) greatly reduces its solution time. Furthermore, comparing the solutions obtained by the two problems it turns out that most of the carriers use the same path in both solutions and the increase of the carriers' cost in the tolled solution is not significant, whereas the total risk is slightly reduced.
- In the general case, where the leader's objective involves also a carriers' term, if parameter w_1 is gradually increased, the population exposure slowly increases at first, while the carriers' cost decreases quickly, and then more rapidly. As a comparison to the true bilevel formulation a proxy of problem (17) is solved by inverse optimization, and this solution is also used as a warm start for HTND problem (1). The proxy is much simpler to solve than the true bilevel problem, and the solution is close to the true one when the small dataset of commodities is considered. When all the shipments are considered, the solution of the true bilevel is better and not significantly harder than problem (1) (both problems are solved in less than 10 min).

A weakness of model (17) is that tolls are not necessarily set on risky arcs. Similarly, for model (1) it may happen that arcs that are not risky are closed. To prevent this

situation, a constrained version of both problems is considered in Marcotte et al. (2009): in particular, an arc may be tolled or closed only if the population exposure exceeds a certain amount $R_{\min}$, and the two problems are solved for increasing values of this parameter. It turns out that when $R_{\min}$ increases HTND problem (1) becomes easier to solve, whereas the time needed to solve TS problem (17) is more stable. However, the solutions of problem (17) keep the population exposure low, while decreasing the number of tolled arcs. When $R_{\min}$ exceeds a certain threshold, the two models become equivalent.

All the experiments are carried out on an AMD Opteron Processor 248, 2,19 MHz computer, using two processors.

Summarizing, the toll setting approach may result in a lower risk than the network design one, and in some cases the resulting model can be solved very efficiently. The flexibility arises from the ability to differentiate among carriers.

The Model of Bianco et al. (2012)

A limit of the model of Marcotte et al. (2009) is that it does not keep into account at all risk equity. Indeed, the flow on the arcs is unsplittable (i.e., $x_{ij}^k \in \{0, 1\}$) so that it is impossible to further reduce the risk on the arcs of the route of a carrier transporting a big amount of hazmat. Similarly, it may happen that a certain arc is chosen by many carriers, resulting in a very high risk on this arc even if the total risk on the network is relatively low. This drawback is overcome by Bianco et al. (2012) by using a different toll setting where the toll paid by the carriers on each arc depends also on the total risk on that arc. Therefore the choices of each carrier depend on the other carriers' choices, and the tolls deter the carriers from choosing links with high total risk. The obtained model is a Mathematical Program with Equilibrium Constraint (MPEC) where the inner problem is a Nash game having the carriers as players.

In this scenario, the authority aims at minimizing the total risk over the network of its jurisdiction zone, and also aims at pursuing the risk equity by minimizing the maximum link total risk among the links of the network. The instrument available to the authority in order to control the hazmat flows of the carriers is again toll setting. Differently from the model of Marcotte et al. (2009), the amount of the toll on each link is assumed to be a quadratic function of the total amount of risk that all the carriers induce on that link. The idea is to induce the carriers to transport the hazmat along routes where the risk is lower, and at the same time to limit the amount of total risk on each single link. Therefore, the additional (tax) cost faced by each carrier depends on the hazmat flows of the other carriers since the amount of tax they have to pay on each arc depends on the total hazmat flow on that arc.

More in detail, in Bianco et al. (2012), for each arc $(i, j) \in A$ a tax $T_{ij}^{h(k)}$ is assumed to be paid by carrier k for each unit of risk induced by its shipment on arc (i, j). As the idea is to discourage carrier k from using a link with high total risk,

with the latter being induced by carrier k and all the other carriers that use that link, tax $T_{ij}^{h(k)}$ per unit of risk is assumed to be the sum of the following two terms

$$T_{ij}^{h(k)} = t_{ij}^{h(k)} + d_{ij}^{h(k)} \sum_{\ell \in K} \rho_{ij}^{h(\ell)} \frac{b^\ell}{q^\ell} x_{ij}^\ell,$$

where variables x_{ij}^ℓ, with $0 \le x_{ij}^\ell \le 1$, represents the fraction of shipment of carrier ℓ along arc (i, j), and $t_{ij}^{h(k)}$ and $d_{ij}^{h(k)}$ are parameters used to weight the risk contribution of carrier k with respect to the risk caused by the whole set of carriers. In particular, the former term is associated to the risk induced by carrier k on arc (i, j) and the latter is related to the total risk induced by all the carriers on the same link.

Therefore the tax paid by carrier k for using arc (i, j) is:

$$t_{ij}^{h(k)} \rho_{ij}^{h(k)} \frac{b^k}{q^k} x_{ij}^k + d_{ij}^{h(k)} \left(\sum_{\ell \in K} \rho_{ij}^{h(\ell)} \frac{b^\ell}{q^\ell} x_{ij}^\ell \right) \rho_{ij}^{h(k)} \frac{b^k}{q^k} x_{ij}^k.$$

The linear component (weighted by parameter $t_{ij}^{h(k)}$) is used to regulate the total risk on the network similarly to the TS model of Marcotte et al. (2009), while the quadratic component (weighted by parameter $d_{ij}^{h(k)}$) is used to control the maximum link total risk, possibly forcing carrier k to split his demand b^k along different routes. Thus the objective function of the subproblem of each carrier k becomes

$$\begin{aligned} \theta_k(\mathbf{x}; \mathbf{t}, \mathbf{d}) = & \sum_{(i,j) \in A} c_{ij} \frac{b^k}{q^k} x_{ij}^k + \sum_{(i,j) \in A} t_{ij}^{h(k)} \rho_{ij}^{h(k)} \frac{b^k}{q^k} x_{ij}^k \\ & + \sum_{(i,j) \in A} d_{ij}^{h(k)} \left(\sum_{\ell \in K} \rho_{ij}^{h(\ell)} \frac{b^\ell}{q^\ell} x_{ij}^\ell \right) \rho_{ij}^{h(k)} \frac{b^k}{q^k} x_{ij}^k, \end{aligned} \tag{26}$$

where $\mathbf{x}$ denotes the vector containing all the carriers' variables x_{ij}^k, and $\mathbf{t}, \mathbf{d}$ denote the vectors of toll parameters t_{ij}^h, d_{ij}^h, respectively. Therefore the carriers' problem constitutes a Nash Equilibrium Problem (NEP) where each carrier k is a player and his subproblem is:

$$\begin{aligned} & \min_{x_{ij}^k} \ \theta_k(\mathbf{x}; \mathbf{t}, \mathbf{d}) \\ & \text{s.t.} \\ & \qquad \sum_{\{j:(i,j) \in A\}} x_{ij}^k - \sum_{\{j:(j,i) \in A\}} x_{ji}^k = e_i^k \qquad \forall i \in N \\ & \qquad x_{ij}^k \ge 0 \quad \forall (i, j) \in A, \end{aligned} \tag{27}$$

Due to the compactness of the feasible set of each players' subproblem, the Nash game admits a solution for every value of parameters t_{ij}^h and d_{ij}^h. Unfortunately, in general, uniqueness of the equilibrium is not guaranteed.

Taking into account the role of the authority and of the carriers the whole model is a bilevel optimization problem, where a leader (the authority) sets tolls on the network links by choosing the values of parameters t_{ij}^h and d_{ij}^h in order to minimize a combination of risk magnitude and carrier travel cost. The followers (carriers) are the players of the game, where each player k (with $k \in K$) aims at solving subproblem (27) on the tolled network and hence selects the flow of his shipments controlling variables x_{ij}^k. In particular the leader wants to minimize the risk magnitude by minimizing the total risk of the network, and (secondly) the maximum link total risk on the arcs of the network. In this situation, we get the following MPEC:

$$
\begin{aligned}
&\min_{\mathbf{t},\mathbf{d}} \sum_{k\in K}\sum_{(i,j)\in A} \rho_{ij}^{h(k)} \frac{b^k}{q^k} x_{ij}^k + w_0 \cdot \Phi + w_1 \sum_{k\in K} \theta_k(\mathbf{x};\mathbf{t},\mathbf{d}) \\
&\quad \text{s.t.} \\
&\qquad \sum_{l\in K} \rho_{ij}^{h(l)} \frac{b^l}{q^l} x_{ij}^l \le \Phi, \quad (i,j) \in A \\
&\qquad t_{ij}^h \ge 0, d_{ij}^h \ge 0, \quad \forall (i,j) \in A,\ h \in H \\
&\qquad \text{where } \mathbf{x} \text{ is a Nash Equilibrium (NE) of} \\
&\qquad\quad \min_{x_{ij}^k} \ \theta_k(\mathbf{x};\mathbf{t},\mathbf{d}) \\
&\quad \text{s.t.} \\
&\qquad \sum_{\{j:(i,j)\in A\}} x_{ij}^k - \sum_{\{j:(j,i)\in A\}} x_{ji}^k = e_i^k \qquad \forall i \in N \\
&\qquad x_{ij}^k \ge 0 \quad \forall (i,j) \in A \\
&\qquad \text{for all } k \in K,
\end{aligned}
\tag{28}
$$

where Φ is a continuous variable representing the maximum link total risk, while w_0 and w_1 are some weight factors that allow to control the preference of the leader with respect to the different aspects he wants to control.

In order to understand the advantages of this model, consider again the small network shown in Fig. 3 and described in the previous section. The solution of model (17) provides a total risk equal to the (minimum) value 12.5, but it is determined without taking into account the risk equity. The maximum link total risk is equal to 2 and obtained on arcs (s^1, b) and (b, d). By using model (28), since the carriers' variables x_{ij}^k are assumed to be continuous, it is possible to set the tolls in order to induce carrier 1 to split its flow along the two minimum risk paths $(s^1 \to a \to t^1)$ and $(s^1 \to b \to d \to t^1)$, getting again the (minimum) total risk of value 12.5, but a lower maximum link total risk of value 1.75 (obtained on arcs (s^1, c) and (c, d)). Finally, the example shows that model (28) is able to achieve the same network total risk and a better distribution of the risk with respect to model (17). Therefore, the model of Bianco et al. (2012) can provide solutions dominating those of the model

of Marcotte et al. (2009). Note that the opposite cannot occur because the feasible set of model (28) strictly contains the feasible set of model (17).

Problem (28), being a MPEC, is NP-hard and there are few algorithms available for solving it. A favorable scenario is when all the carriers transport the same hazardous material, i.e. $h(k) = \tilde{h}$ for all $k \in K$. In this case, some extra properties hold:

1. The Nash equilibrium is unique, namely for any value of toll parameters $t_{ij}^{\tilde{h}}$ and $d_{ij}^{\tilde{h}}$, provided that $d_{ij}^{\tilde{h}} > 0$ and $\rho_{ij}^{\tilde{h}} > 0$ for all $(i, j) \in A$, there exists a unique equilibrium of problem (27).
2. The NEP is a *potential game*. Potential game are defined in Monderer and Shapley (1996) as follows: a game where each player $k = 1, \ldots, p$ has to solve the problem

$$\begin{aligned} &\min_{x^k} \ \theta_k(\mathbf{x}) \\ &\text{s.t.} \ \ x^k \in \mathscr{D}_k, \end{aligned}$$

with $\mathscr{D}_k$ being the feasible set of player k, is an exact potential game if there exist a function $P(\mathbf{x})$ such that

$$\nabla_{x^k} P(\mathbf{x}) = \nabla_{x^k} \theta_k(\mathbf{x}), \ \text{for every } k = 1, \ldots, p.$$

Whenever a game is potential, any solution of the problem

$$\begin{aligned} &\min \ P(\mathbf{x}) \\ &\text{s.t.} \ \ x^k \in \mathscr{D}_k, \quad k = 1, \ldots, p \end{aligned}$$

is a Nash equilibrium. In the single hazmat case, the NEP (27) is an exact potential game and its potential function is

$$P(\mathbf{x}) = \boldsymbol{\kappa}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T JF(\mathbf{x})\, \mathbf{x} \tag{29}$$

where

$$\boldsymbol{\kappa}_{ij}^k = \left(c_{ij} + t_{ij}^{\tilde{h}} \rho_{ij}^{\tilde{h}}\right) \frac{b^k}{q^k}$$

and JF is the matrix

$$JF(\mathbf{x}) = \begin{pmatrix} 2\left(\frac{b^1}{q^1}\right)^2 H & \frac{b^1 b^2}{q^1 q^2} H & \ldots & \frac{b^1 b^p}{q^1 q^p} H \\ \frac{b^2 b^1}{q^2 q^1} H & 2\left(\frac{b^2}{q^2}\right)^2 H & \ldots & \frac{b^2 b^p}{q^2 q^p} H \\ \vdots & \vdots & \ddots & \vdots \\ \frac{b^p b^1}{q^p q^1} H & \frac{b^p b^2}{q^p q^2} H & \ldots & 2\left(\frac{b^p}{q^p}\right)^2 H \end{pmatrix}$$

where $p = |K|$ and, by denoting with Diag($\mathbf{v}$) the diagonal matrix having as elements on the diagonal the elements of vector $\mathbf{v}$,

$$H = \text{Diag}\left((\rho_{ij}^{\tilde{h}})^2 d_{ij}^{\tilde{h}}\right)_{(i,j)\in A}.$$

Therefore, the NEP (27) is equivalent to the following strictly convex optimization problem:

$$\begin{aligned} &\min_{\mathbf{x}} \boldsymbol{\kappa}^T \mathbf{x} + \frac{1}{2}\mathbf{x}^T JF(\mathbf{x})\, \mathbf{x} \\ &\text{s.t.} \\ &\sum_{\{j:(i,j)\in A\}} x_{ij}^k - \sum_{\{j:(j,i)\in A\}} x_{ji}^k = e_i^k \qquad \forall i \in N,\ k \in K \\ &x_{ij}^k \geq 0 \quad \forall (i,j) \in A,\ \forall k \in K. \end{aligned} \tag{30}$$

3. Given a Nash equilibrium it is possible to define an LP problem that gives the optimal (with respect to a given criterion) toll setting forcing the carriers to choose the routing corresponding to that Nash equilibrium.

The first property ensures stability of the solutions of the bilevel problem (28). The second one implies that, for a given value of $t_{ij}^{\tilde{h}}$ and $d_{ij}^{\tilde{h}}$ the NEP (27) can be solved by means of a distributed algorithm, where each player implements his own best response function. In particular, each player in turn, given the choices of the other carriers, solves his own optimization problem (27). For potential games, the process converges to an equilibrium. The algorithm can be seen also as a distributed algorithm (a Gauss-Seidel type algorithm) for problem (30) whose convergence is guaranteed by strict convexity.

This algorithmic possibility implies that the authority does not need complete information on the carriers and it is not necessary to have a centralized system choosing the flows of the carriers.

Finally, the third property allows the use of inverse optimization as in Marcotte et al. (2009). In particular, let $\bar{\mathbf{x}}$ be a Nash equilibrium computed by solving problem (30) for a given set of toll parameters $\bar{t}_{ij}^{\tilde{h}}, \bar{d}_{ij}^{\tilde{h}}$. A reasonable question is whether there exists a tax assignment that is better with respect to some criterion (as for example the total cost payed by the carriers) that gives $\bar{\mathbf{x}}$ as the unique corresponding Nash equilibrium.

This is equivalent to solving a suitable LP problem, having as constraints the necessary and sufficient Karush–Kuhn–Tucker (KKT) optimality conditions of problem (30) with $\mathbf{x} = \bar{\mathbf{x}}$. The KKT optimality conditions for each carrier $k \in K$ can be rewritten as

$$\begin{aligned} &c_{ij}\tfrac{b^k}{q^k} + \rho_{ij}^{\tilde{h}}\tfrac{b^k}{q^k} t_{ij}^{\tilde{h}} + \gamma_{ij}^k(\bar{\mathbf{x}}) d_{ij}^{\tilde{h}} - \pi_i^k + \pi_j^k \geq 0,\ \forall (i,j) \in A \backslash \bar{A}^k \\ &c_{ij}\tfrac{b^k}{q^k} + \rho_{ij}^{\tilde{h}}\tfrac{b^k}{q^k} t_{ij}^{\tilde{h}} + \gamma_{ij}^k(\bar{\mathbf{x}}) d_{ij}^{\tilde{h}} - \pi_i^k + \pi_j^k = 0,\ \forall (i,j) \in \bar{A}^k, \end{aligned} \tag{31}$$

where $\bar{A}^k = \{(i,j) \in A : \bar{x}_{ij}^k > 0\}$, π_i^k are the free dual variables associated to the flow conservation constraints of subproblem (27) of player k, and

$$\gamma_{ij}^k(\bar{\mathbf{x}}) = \rho_{ij}^{\tilde{h}} \frac{b^k}{q^k}(\phi_{ij}(\bar{\mathbf{x}}) + \rho_{ij}^{\tilde{h}} \frac{b^k}{q^k} \bar{x}_{ij}^k)$$

with

$$\phi_{ij}(\bar{\mathbf{x}}) = \sum_{k \in K} \rho_{ij}^{\tilde{h}} \frac{b^k}{q^k} \bar{x}_{ij}^k$$

denoting the total risk on arc (i,j) corresponding to the solution $\bar{\mathbf{x}}$. Note that the flow conservation constraints and the non-negativity constraints on the values of the flows $\mathbf{x}$ are omitted since they are obviously satisfied at $\bar{\mathbf{x}}$.

Given constraints (31), and assuming that the aim is to find the toll setting that minimizes the sum of the cost of the carriers, we get the LP:

$$\begin{aligned}
\min_{t_{ij}^{\tilde{h}}, d_{ij}^{\tilde{h}}, \pi_i^k} \zeta(\bar{\mathbf{x}}; \mathbf{t}, \mathbf{d}) = & \sum_{k \in K} \sum_{(i,j) \in A} c_{ij} \frac{b^k}{q^k} \bar{x}_{ij}^k + \sum_{k \in K} \sum_{(i,j) \in A} t_{ij}^{\tilde{h}} \rho_{ij}^{\tilde{h}} \frac{b^k}{q^k} \bar{x}_{ij}^k \\
& + \sum_{k \in K} \sum_{(i,j) \in A} d_{ij}^{\tilde{h}} \left(\sum_{\ell=1}^{p} \rho_{ij}^{\tilde{h}} \frac{b^\ell}{q^\ell} \bar{x}_{ij}^\ell \right) \rho_{ij}^{\tilde{h}} \frac{b^k}{q^k} \bar{x}_{ij}^k \\
= & \sum_{k \in K} \sum_{(i,j) \in A} c_{ij} \frac{b^k}{q^k} \bar{x}_{ij}^k + \sum_{(i,j) \in A} \phi_{ij}(\mathbf{x}) t_{ij}^{\tilde{h}} + \sum_{(i,j) \in A} \left(\phi_{ij}(\mathbf{x})\right)^2 d_{ij}^{\tilde{h}}.
\end{aligned}$$

s.t.

$$\begin{aligned}
& c_{ij} \frac{b^k}{q^k} + \rho_{ij}^{\tilde{h}} \frac{b^k}{q^k} t_{ij}^{\tilde{h}} + \gamma_{ij}^k(\bar{\mathbf{x}}) d_{ij}^{\tilde{h}} - \pi_i^k + \pi_j^k \geq 0, \\
& \quad \forall (i,j) \in A \backslash \bar{A}^k, \ k \in K \\
& c_{ij} \frac{b^k}{q^k} + \rho_{ij}^{\tilde{h}} \frac{b^k}{q^k} t_{ij}^{\tilde{h}} + \gamma_{ij}^k(\bar{\mathbf{x}}) d_{ij}^{\tilde{h}} - \pi_i^k + \pi_j^k = 0, \\
& \quad \forall a \in \bar{A}^k, \ k \in K \\
& d_{ij}^{\tilde{h}} \geq \epsilon, \quad \forall (i,j) \in A, \ h \in H \\
& t_{ij}^{\tilde{h}}, d_{ij}^{\tilde{h}} \geq 0, \quad \forall (i,j) \in A, \ h \in H \\
& \pi_i^k \text{ free}, \quad \forall \ i \in N, \ k \in K,
\end{aligned} \tag{32}$$

where ϵ is a small value greater than 0.

This problem is well posed. Indeed, it has a nonempty feasible set since the toll parameter values $\bar{t}_{ij}^{\tilde{h}}, \bar{d}_{ij}^{\tilde{h}}$ are feasible, and the objective function is bounded from below, since it assumes non-negative values.

The Model of Wang et al. (2011)

An MPEC is proposed also by Wang et al. (2011), where a dual toll pricing method is introduced in order to mitigate the risk of hazardous material transportation by regulating both hazmat and regular traffic. The idea is to simultaneously control both hazmat and regular vehicles aiming at minimizing the total risk and the total cost on the whole network. It is assumed that all the road segments can be tolled and the user equilibrium objective is to channel traffic flows along the routes generated by the first-best model via toll setting. In order to introduce the mathematical model proposed by Wang et al. (2011), some notation is needed. Given the network $G = (N, A)$, hazmat and regular Origin–Destination (O–D) pairs are denoted by K and Q respectively, and for each O–D pair $k \in K$ and $q \in Q$ the corresponding demand is denoted by b^k and d^q, whereas the arc flow vectors by $x^k \in \Re^{|A|}$ and $y^q \in \Re^{|A|}$. The aggregated arc flow vectors given by the sum of all the arc flows are denoted by $u \in \Re^{|A|}$ for the hazmat vehicles and $v \in \Re^{|A|}$ for the regular vehicles. The toll vector for regular vehicles is denoted by $\alpha \in \Re^{|A|}$, whereas the one for hazmat vehicles is denoted by $\beta \in \Re^{|A|}$. The feasible flows are the ones belonging to the sets:

$$\begin{aligned} V &= \{v \in \Re^{|A|} : v = \textstyle\sum_{q \in Q} y^q, \mathcal{N} y^q = d^q, y^q \geq 0 \ \forall q \in Q\} \\ U &= \{u \in \Re^{|A|} : u = \textstyle\sum_{k \in K} x^k, \mathcal{N} x^k = b^k, x^k \geq 0 \ \forall k \in K\} \end{aligned} \tag{33}$$

where $\mathcal{N}$ is the node-arc incidence matrix of the network. The two feasible sets can be rewritten using matrix-vector notation as

$$\begin{aligned} V &= \{v \in \Re^{|A|} : v = Z_1 y, M_1 y = d, y \geq 0\} \\ U &= \{u \in \Re^{|A|} : u = Z_2 x, M_2 x = b, x \geq 0\} \end{aligned} \tag{34}$$

The following assumptions are made:

- Since the number of hazmat vehicles is significantly smaller than the number of regular ones, congestion caused by hazmat vehicles can be ignored.
- Network users are assumed to have perfect information on the status of the network.
- The model is assumed to be deterministic, namely there is no uncertainty on the travel and the behavior of the users.
- A single type of hazmat is considered.

Let $s(v+u) \in \Re^{|A|}$ be the vector of travel cost functions. Since it is assumed that no congestion effect is due to hazmat trucks, it is reasonable to approximate $s(v+u)$ with $s(v)$. Let $R_{ij}(v_{ij}, u_{ij})$ be the risk function on arc $(i, j) \in A$. The risk R_{ij} on each arc $(i, j) \in A$ is induced by the hazmat vehicles, but it is influenced also by the regular traffic, since a big amount of regular vehicles may increase the probability of accidents involving hazmat trucks. The MPEC proposed by Wang et al. (2011) is

$$\begin{aligned}
&\min_{v,u,\alpha,\beta} \; w_1 \sum_{(i,j)\in A} R_{ij}(v_{ij}, u_{ij}) + w_2 s(v)^T v + w_3 s(v)^T u \\
&\text{s.t.} \\
&\qquad (s(v)+\alpha)^T (t - v) \geq 0 \quad \forall t \in V \\
&\qquad (s(v)+\beta)^T (r - u) \geq 0 \quad \forall r \in U \\
&\qquad v \in V \\
&\qquad u \in U \\
&\qquad \alpha, \beta \geq 0.
\end{aligned} \tag{35}$$

The objective function is a weighted combination (with weights w_1, w_2, and w_3) of the network total risk and the network total cost for the regular and hazmat vehicles, respectively. The first two groups of constraints represent the network user equilibrium conditions for the regular and hazmat vehicles, respectively, according to the Wardrop's first principle.

The risk function used in Wang et al. (2011) is a duration–population–frequency risk function

$$R_{ij}(v_{ij}, u_{ij}) = s_{ij}(v_{ij}) \rho_{ij} u_{ij}, \tag{36}$$

where ρ_{ij} is the population exposure along the arc (i, j).

As a cost function it is chosen

$$s_{ij}(v_{ij}) = \tau_{ij} \left(1 + \left(\frac{v_{ij}}{C_{ij}}\right)^{\gamma}\right), \tag{37}$$

where τ_{ij} is the travel time on arc (i, j) when there is no traffic flow and C_{ij} is the capacity of the arc (i, j), and γ is assumed equal to 1.

Wang et al. (2011) reformulate the MPEC problem (35) as a two stage problem, with some assumptions on the objective function. In particular, the first stage problem determines the optimal flow that minimizes the total risk without imposing the equilibrium constraints:

$$\begin{aligned}
&\min_{v,u} \; w_1 \sum_{(i,j)\in E} R_{ij}(v_{ij}, u_{ij}) + w_2 s(v)^T v + w_3 s(v)^T u \\
&\text{s.t.} \quad v \in N \\
&\qquad\; u \in U
\end{aligned} \tag{38}$$

Choosing (36) and (37) as risk and cost functions respectively, problem (38) is a non convex Quadratic Programming (QP) problem, that is NP-hard as the original MPEC. Wang et al. (2011) consider a branch and bound method and an active set method for attacking the problem.

Given a solution $(\bar{u}, \bar{v})$ of problem (38), a toll setting (α, β) makes this solution a user equilibrium if it satisfies the following conditions:

$$\begin{aligned}
&(s(\bar{v})+\alpha)^T (v - \bar{v}) \geq 0 \quad \forall v \in V \\
&(s(\bar{v})+\beta)^T (u - \bar{u}) \geq 0 \quad \forall u \in U
\end{aligned}$$

that are shown by Wang et al. (2011) to hold if and only if there exist vectors ϕ and ω such that

$$\begin{aligned} &Z_1^T(s(\bar{v})+\alpha)^{\geq} M_1\phi \\ &(s(\bar{v})+\alpha)^T\bar{v} = d^T\phi \\ &Z_2^T(s(\bar{v})+\beta)^{\geq} M_2\omega \\ &(s(\bar{v})+\beta)^T\bar{u} = b^T\omega. \end{aligned}$$

Among all the possible tolls (α, β), Wang et al. (2011) consider a revenue minimizing toll vector pair, in order to keep users' costs as low as possible, obtained by solving the following linear problem:

$$\begin{aligned} &\min_{\alpha,\beta,\phi,\omega} \ \alpha^T\bar{v} + \beta^T\bar{u} \\ &\text{s.t.} \\ &\qquad Z_1^T(s(\bar{v})+\alpha)^{\geq} M_1\phi \\ &\qquad (s(\bar{v})+\alpha)^T\bar{v} = d^T\phi \\ &\qquad Z_2^T(s(\bar{v})+\beta)^{\geq} M_2\omega \\ &\qquad (s(\bar{v})+\beta)^T\bar{u} = b^T\omega. \end{aligned} \tag{39}$$

The model is tested in Wang et al. (2011) on the road network of Albany in the New York State. This network has 46 nodes and 70 arcs, considering two pairs of regular demand nodes and three pairs of hazmat demand nodes. Problem (38) is solved by a QP solver `bqpd` based on a null space active set method with a technique for resolving degeneracy. The dual-tolled traffic flows are compared with the minimum travel cost flow of regular traffic and the corresponding minimum risk flow of hazmat traffic. In the experiments, the weights w_1, w_2 and w_3 are increased one at the time. Keeping fixed w_2 and w_3, the risk decreases as w_1 increases, but the regular traffic delay increases. With large enough w_1 no tolls are set on the arcs. On the other hand, if w_1 and w_3 are fixed and w_2 is high enough, the risk is still decreased while the delays are slightly increased. In these solutions, no tolls are imposed on the hazmat trucks. No information is reported on CPU time needed in order to get these solutions.

Conclusions

In this chapter, the hazmat transportation problem on road networks has been presented in terms of multi-commodities and multiple origin–destination problem. This approach involves routing decisions by each carrier but also by the government authority due to the intrinsic risk (societal and environmental) associated with hazmat transportation.

The main concern for the authority (regulator) is to define restrictions to the routes selected by of all the carriers in order to control and minimize the total risk induced by hazmat transportation and to achieve equity in the spatial distribution of

the risk. On the contrary, the main objective of each carrier is the transportation cost minimization.

For this problem, called "global route planning problem," approaches considering a single decision maker choosing the routes are of little use because, in general, the regulator has no authority to impose routes on individual carriers. For this reason, two alternative policies have been considered for the regulators.

The first policy deals with the possibility of the authority to close certain road segments to hazmat vehicles (or to limit the amount of hazmat traffic flow on those links) to avoid, for example, the use of too high risky road sections. The second policy deals with the use of (link) tolls by the authority to induce carriers to route their shipments on less risky roads.

In this context, the first strategy falls in the field of "network design," while the second one falls in the field of "toll setting policies."

For both of these two kind of problems the main mathematical models present in the literature, so as the related applications, have been illustrated. The advantages and drawbacks of these models have been also discussed in comparison of each one with respect to the others.

References

Abkowitz M, Cheng P (1988) Developing a risk/cost framework for routing truck movements of hazardous materials. Accid Anal Prev 20(1):39–51

Ahuja R, Magnanti T, Orlin J (1993) Network flows: theory, algorithms and applications. Prentice-Hall, Englewood Cliffs, NJ

Akgün V, Erkut E, Batta R (2000) On finding dissimilar paths. Eur J Oper Res 121(2):232–246

Akgün V, Parekh A, Batta R, Rump C (2007) Routing of a hazmat truck in the presence of weather systems. Comput Oper Res 34(5):1351–1373

Androutsopoulos K, Zografos K (2010) Solving the bicriterion routing and scheduling problem for hazardous materials distribution. Transport Res C 18(5):713–726

Ashtakala B, Eno L (1996) Minimum risk route model for hazardous materials. J Transport Eng ASCE 22(5):350–357

Balakrishnan A, Magnanti T, Mirchandani P (1997) Annotated bibliographies in combinatorial optimization, Chap. Network design. Wiley, New York, pp 311–334

Bard J (2006) Practical bilevel optimization: algorithms and applications (nonconvex optimization and its applications). Springer, New York

Batta R, Chiu S (1988) Optimal obnoxious paths on a network: transportation of hazardous materials. Oper Res 36(1):84–92

Belardo S, Pipkin J, Seagle J (1985) Information support for hazardous materials movement. J Hazard Mater 10(1):13–32

Bell M (2006) Mixed route strategies for the risk-averse shipment of hazardous materials. Netw Spat Econ 6:253–265

Bell M (2007) Mixed routing strategies for hazardous materials: decision-making under complete uncertainty. Int J Sustain Transport 1(2):133–142

Beroggi G (1994) A real-time routing model for hazardous materials. Eur J Oper Res 75(3): 508–520

Beroggi G, Wallace W (1991) Closing the gap-transit control for hazardous material flow. J Hazard Mater 27(1):61–75

Beroggi G, Wallace W (1994) Operational risk management: a new paradigm for decision-making. IEEE Trans Syst Man Cybern 24(10):1450–1457
Beroggi G, Wallace W (1995) Operational control of the transportation of hazardous materials: an assessment of alternative decision models. Manage Sci 41(12):1962–1977
Bianco L, Caramia M, Giordani S (2009) A bilevel flow model for hazmat transportation network design. Transport Res C 17:175–196
Bianco L, Caramia M, Giordani S, Piccialli V (2012) A game theory approach for regulating hazmat transportation. Tech. Rep. RR-21.12, Dipartimento di Ingegneria dell'Impresa, University of Rome "Tor Vergata", Italy
Boffey B, Karkazis J (1995) Linear versus nonlinear models for hazardous materials routing. INFOR 33:114–117
Bonvicini S, Spadoni G (2008) A hazmat routing model satisfying risk criteria. In: P.O. Inveldi (ed.), Transportation research trends, Nova Science Publishers. Inc., New York, pp 129–169
Bonvicini S, Leonelli P, Spadoni G (1998) Risk analysis of hazardous materials transportation: evaluating uncertainty by means of fuzzy logic. J Hazard Mater 62(1):59–74
Bowler L, Mahmassani H (1998) Routing of radioactive shipments in networks with timevarying costs and curfews. Tech. Rep. ANRCP-1998-11, Amarillo National Resource Center for Plutonium, TX, USA
Caramia M, Giordani S (2009) On the selection of k efficient paths by clustering techniques. Int J Data Min Model Manag 1(3):237–260
Caramia M, Giordani S, Iovanella A (2010) On the selection of k routes in multi-objective hazmat route planning. IMA J Manag Math 21(3):239–251
Carotenuto P, Giordani S, Ricciardelli S (2007a) Finding minimum and equitable risk routes for hazmat shipments. Comput Oper Res 34(5):1304–1327
Carotenuto P, Giordani S, Ricciardelli S, Rismondo S (2007b) A tabu search approach for scheduling hazmat shipments. Comput Oper Res 34(5):1328–1350
Chang T, Nozick L, Turnquist M (2005) Multi-objective path finding in stochastic dynamic networks, with application to routing hazardous materials shipments. Transport Sci 39(3):383–399
Chin SM, Cheng PM (1989) Bicriterion routing scheme for nuclear spent fuel transportation. Transport Res Rec 1245:60–64
Clarke G, Wright J (1964) Scheduling of vehicles from a central depot to a number of delivery points. Oper Res 12(4):568–581
Cox R (1984) Routing and scheduling of hazardous materials shipments: algorithmic approaches to managing spent nuclear fuel transport. Tech. rep., Ph.D. Thesis, Cornell University, Ithaca, NY
Cox R, Turnquist M (1986) Scheduling truck shipments of hazardous materials in the presence of curfews. Transport Res Rec 1063:21–26
Current J, Ratick S (1995) A model to assess risk, equity, and efficiency in facility location and transportation of hazardous materials. Locat Sci 3:187–202
Dadkar Y, Jones D, Nozick L (2008) Identifying geographically diverse routes for the transportation of hazardous materials. Transport Res E 44:333–349
Dadkar Y, Nozick L, Jones D (2010) Optimizing facility use restrictions for the movement of hazardous materials. Transport Res B 44:267–281
Dantzig G, Ramser X (1959) The truck dispatching problem. Manag Sci 6(1):80–91
Dell'Olmo P, Gentili M, Scozzari A (2005) On finding dissimilar pareto-optimal paths. Eur J Oper Res 162(1):70–82
Duque J, Barbosa-Pvoa A, Novais A (2007) Synthesis and optimization of the recovery route for residual products under uncertain product demand. Comput Oper Res 34(5):1463–1490
Erkut E (1990) The discrete p-dispersion problem. Eur J Oper Res 46:48–60
Erkut E (1995) On the credibility of the conditional risk model for routing hazardous materials. Oper Res Lett 18:49–52
Erkut E (1996) The road not taken. OR/MS Today 23:22–28
Erkut E, Alp O (2007a) Designing a road network for dangerous goods shipments. Comput Oper Res 34(5):1389–1405

Erkut E, Alp O (2007b) Integrated routing and scheduling of hazmat trucks with stops en-route. Transport Sci 41(1):107–122
Erkut E, Glickman T (1997) Minimax population exposure in routing highway shipments of hazardous materials. Transport Res Rec 1602:93–100
Erkut E, Gzara F (2008) Solving the hazmat transport network design problem. Comput Oper Res 35(7):2234–2247
Erkut E, Ingolfsson A (2000) Catastrophe avoidance models for hazardous materials route planning. Transport Sci 34(2):165–179
Erkut E, Ingolfsson A (2005) Transport risk models for hazardous materials: revisited. Oper Res Lett 33(1):81–89
Erkut E, Verter V (1998) Modeling of transport risk for hazardous materials. Oper Res 46(5):625–642
Erkut E, Ülküsal Y, Yeniçerioğlu O (1994) A comparison of p-dispersion heuristic. Comput Oper Res 21(10):1103–1113
Erkut E, Tjandra S, Verter V (2007) Hazardous materials transportation. In: C. Barnhart, G. Laporte (eds.), Transportation, Handbooks in operations research & management science, vol 14, Elsevier, Amsterdam, pp 539–621
Ferrada J, Michelhaugh R (1994) Development of an expert system for transportation of hazardous and radioactive materials. In: Proceedings of international topical meeting on nuclear and hazardous waste management, Spectrum '94, vol 1.2, American Nuclear Society Inc., Atlanta, pp 997–1002
Frank W, Thill J, Batta R (2000) Spatial decision support system for hazardous material truck routing. Transport Res C 8:337–359
Glickman T, Sontag M (1995) The tradeoffs associated with rerouting highway shipments of hazardous materials to minimize risk. Risk Anal 15(1):61–67
Gopalan R, Batta R, Karwan M (1990a) The equity constrained shortest path problem. Comput Oper Res 17(3):297–307
Gopalan R, Kolluri K, Batta R, Karwan M (1990b) Modeling equity of risk in the transportation of hazardous materials. Oper Res 38(6):961–975
Hansen P (1980) Bicriterion path problems. In: G. Fandel et al. (eds.), Multiple criteria decision making: theory and applications, vol 177, Springer, Berlin, pp 109–127
Huang B (2006) Gis-based route planning for hazardous material transportation. J Environ Inform 8(1):49–57
Huang B, Cheu R (2004) Gis and genetic algorithms for hazmat route planning with security considerations. Int J Geogr Inform Sci 18(18):769–787
Jin H, Batta R (1997) Objectives derived from viewing hazmat shipments as a sequence of independent bernoulli trials. Transport Sci 31(3):252–261
Jin H, Batta R, Karwan M (1996) On the analysis of two new models for transporting hazardous materials. Oper Res 44(5):710–723
Johnson P, Joy D, Clarke D, Jacobi J (1992) Highway 3.01, an enhanced highway routing model: program, description, methodology, and revised user's manual. Tech. Rep. ORNL/TM-12124, Oak Ridge National Laboratory, Oak Ridge, TN
Kalelkar A, Brinks R (1978) Use of multidimensional utility functions in hazardous shipment decisions. Accid Anal Prev 10:251–265
Kara B, Verter V (2004) Designing a road network for hazardous materials transportation. Transport Sci 38(2):188–196
Kara B, Erkut E, Verter V (2003) Accurate calculation of hazardous materials transport risks. Oper Res Lett 31(4):285–292
Karkazis J, Boffey T (1995) Optimal location of routes for vehicles transporting hazardous materials. Eur J Oper Res 86(2):201–215
Keeney R (1980) Equity and public risk. Oper Res 28:527–534
Keeney R, Winkler L (1985) Evaluating decision strategies for equity of public risks. Oper Res 33(5):955–970

Kuby M, Zhongyi X, Xiaodong X (1997) A minimax method for finding the k best "differentiated" paths. Geogr Anal 29(4):298–313
Labbé M, Marcotte P, Savard G (1998) A bilevel model of taxation and its application to optimal highway pricing. Manag Sci 44(1):1608–1622
Lassarre S, Fedra K, Weigkricht E (1993) Computer-assisted routing of dangerous goods for haute-normandie. J Transport Eng 119(2):200–210
Leonelli P, Bonvicini S, Spadoni G (2000) Hazardous materials transportation: a risk-analysis-based routing methodology. J Hazard Mater 71:283–300
Lepofsky M, Abkowitz M, Cheng P (1993) Transportation hazard analysis in integrated gis environment. J Transport Eng ASCE 119(2):239–254
Lindner-Dutton L, Batta R, Karwan M (1991) Equitable sequencing of a given set of hazardous materials shipments. Transport Sci 25(2):124–137
Lombard K, Church R (1993) The gateway shortest path problem: generating alternative routes for a corridor location problem. Geogr Syst 1:25–45
Luedtke J, White C (2002) Hazmat transportation and security: survey and directions for future research. Tech. rep., Department of Industrial & Systems Engineering, Georgia Institute of Technology
Luo ZQ, Pang JS, Ralph D (1996) Mathematical programs with equilibrium constraints. Cambridge University Press, Cambridge, New York
Magnanti T, Wong R (1984) Network design and transportation planning: models and algorithms. Transport Sci 18(1):1–55
Marcotte P, Mercier A, Savard G, Verter V (2009) Toll policies for mitigating hazardous materials transport risk. Transport Sci 43(2):228–243
Marianov V, ReVelle C (1998) Linear non-approximated models for optimal routing in hazardous environments. J Oper Res Soc 49(2):157–164
Marianov V, ReVelle C, Shih S (2002) Anticoverage models for obnoxious material transportation. Environ Plann Plann Des 29(1):141–150
Martí J, Gonzalez Velardi JL, Duarte A (2009) Heuristics for the bi-objective path dissimilarity problem. Comput Oper Res 36:2905–2912
Martins E (1984) On a multicriteria shortest path problem. Eur J Oper Res 16:236–245
Martins E, Santos J (1999) The labeling algorithm for the multiobjective shortest path problem. Tech. rep. http://www.mat.uc.pt/~eqvm/cientificos
McCord M, Leu A (1995) Sensitivity of optimal hazmat routes to limited preference specification. INFOR 33(2):68–83
Meng Q, Lee D, Cheu R (2005) Multiobjective vehicle routing and scheduling problem with time window constraints in hazardous material transportation. J Transport Eng 131(9):699–707
Miaou S, Chin S (1991) Computing k-shortest path for nuclear spent fuel highway transportation. Eur J Oper Res 53(1):64–80
Miller-Hooks E (2001) Adaptive least-expected time paths in stochastic, time-varying transportation and data networks. Networks 37(1):35–52
Miller-Hooks E, Mahmassani H (1998) Optimal routing of hazardous materials in stochastic, time-varying transportation networks. Transport Res Rec 1645:143–151
Miller-Hooks E, Mahmassani H (2000) Least expected time paths in stochastic, time-varying transportation networks. Transport Sci 34(2):198–215
Monderer D, Shapley L (1996) Potential games. Game Econ Behav 14:124–143
Moore J, Sandquist G, Slaughter D (1995) A route-specific system for risk assessment of radioactive materials transportation accidents. Nucl Technol 112(1):63–78
Nembhard D, White C (1997) Applications of non-order-preserving path selection to hazmat routing. Transport Sci 31(3):262–271
Nozick L, List G, Turnquist M (1997) Integrated routing and scheduling in hazardous materials transportation. Transport Sci 31(3):200–215

Outrata J, Kocvara M, Zowe J (1998) Nonsmooth approach to optimization problems with equilibrium constraints. Kluwer Academic Publishers, Dordrecht, The Netherlands
Papadimitriou C, Steiglitz K (1982) Combinatorial optimization: algorithms and complexity. Prentice-Hall, Inc., Englewood Cliffs, NJ, USA
Patel M, Horowitz A (1994) Optimal routing of hazardous materials considering risk of spill. Transport Res A 28(2):119–132
Pradhananga R, Taniguchi E, Yamada T (2010) Ant colony system based routing and scheduling for hazardous material transportation. Procedia Soc Behav Sci 2(3):6097–6108
Reilly A, Nozick L, Xu N, Jones D (2012) Game theory-based identification of facility use restrictions for the movement of hazardous materials under terrorist threat. Transport Res E 48(1):115–131
Saccomanno F, Chan A (1985) Economic evaluation of routing strategies for hazardous road shipments. Transport Res Rec 1020:12–18
Sherali H, Brizendine L, Glickman T, Subramanian S (1997) Low probability— high consequence considerations in routing hazardous materials shipments. Transport Sci 31(3):237–251
Sivakumar R, Batta R, Karwan M (1993) A network-based model for transporting extremely hazardous materials. Oper Res Lett 13(2):85–93
Sivakumar R, Batta R, Karwan M (1995) A multiple route conditional risk model for transporting hazardous materials. INFOR 33(1):20–33
Sulijoadikusumo G, Nozick L (1998) Multiobjective routing and scheduling of hazardous materials shipments. Transport Res Rec 1613:96–104
Tarantilis C, Kiranoudis C (2001) Using the vehicle routing problem for the transportation of hazardous materials. Oper Res 1(1):67–78
Tayi G, Rosenkrantz D, Ravi S (1999) Path problems in networks with vector-valued edge weights. Networks 34(1):19–35
Toth P, Vigo De (2002) The vehicle routing problem. SIAM monographs on discrete mathematics and applications, Philadelphia, PA, USA
Turnquist M (1993) Multiple objectives, uncertainty and routing decisions for hazardous materials shipments. Proceedings of the 5th international conference on computing in civil and building engineering, ASCE, New York, pp 357–364
Urbanek G, Barber E (1980) Development of criteria to designate routes for transporting hazardous materials. Final report. Tech. Rep. FHWA-RD-80-105, Federal Highway Administration, Washington, DC
Vansteen J (1987) A methodology for aiding hazardous materials transportation decisions. Eur J Oper Res 32(2):231–244
Verter V, Erkut E (1997) Incorporating insurance costs in hazardous materials routing models. Transport Sci 31(3):227–236
Verter V, Kara B (2008) A path-based approach for the hazardous network design problem. Manag Sci 54(1):29–40
Wang J, Kang Y, Kwon C, Batta R (2011) Dual toll pricing for hazardous materials transport with linear delay. Netw Spat Econ 1–19, doi:10.1007/s11,067–011–9156–9
Wijeratne A, Turnquist M, Mirchandani P (1993) Multiobjective routing of hazardous materials in stochastic networks. Eur J Oper Res 65:33–43
Yang H, Bell M (1998) Models and algorithms for road network design: a review and some new developments. Transport Rev 18(3):257–278
Yen J (1971) Finding the k shortest loopless paths in a network. Manag Sci 17(11):712–716
Zografos K, Androutsopoulos K (2002a) A heuristic algorithm for solving hazardous materials distribution problems. Eur J Oper Res 152(2):507–519
Zografos K, Androutsopoulos K (2002b) Heuristic algorithms for solving hazardous materials logistical problems. Transport Res Rec 1783:158–166
Zografos K, Androutsopoulos K (2008) A decision support system for integrated hazardous materials routing and emergency response decisions. Transport Res C 16(6):684–703

Zografos K, Davis C (1989) Multi-objective programming approach for routing hazardous materials. J Transport Eng 115:661–673

Zografos K, Vasilakis M, Giannouli M (2000) Methodological framework for developing decision support systems for hazardous material emergency response operations. J Hazard Mater 71:503–521

The Effect of Weather Systems in Hazmat Transportation Modeling

Mohsen Golalikhani and Mark H. Karwan

Introduction

In most transportation models, the objective is to find minimal cost routes from origins to destinations. However, for a hazmat transportation model, a cost minimizing objective is generally not sufficient, because one needs to consider the risk associated with hazardous materials. This makes the hazmat models more complicated when compared to many other transportation models. There are different approaches for considering risk in a hazmat transportation model. For example, some models consider the risk as one of the objectives in a multi-objective model in which other objectives can be cost, distance, travel time, etc. (e.g., Chin and Paul 1989). Some models consider the risk as one component of a single composite cost function and try to find the minimum cost route (e.g., Abkowitz and Cheng 1988). Finally, other models only consider risk as the single objective and try to find a route with minimum risk (e.g., Akgun et al. 2007).

Regardless of how we consider and implement risk via objectives or constraints in a hazmat transportation model, an equally critical issue is how we calculate the risk for a given route. There are several ways of quantifying risk (see Erkut and Verter 1998). However, the most common way is the expected consequence approach, whereby risk is defined as the product of two factors: (a) the probability of an accident with an undesirable consequence (such as injury, illness, or death); and (b) the population affected by the consequence. The affected population depends on two other factors; the area impacted by a hazmat accident, and the number of persons within the impact area.

M. Golalikhani • M.H. Karwan (✉)
Department of Industrial and Systems Engineering, University at Buffalo,
The State University of New York, Buffalo, NY, 14260 USA
e-mail: mkarwan@buffalo.edu

R. Batta and C. Kwon (eds.), *Handbook of OR/MS Models in Hazardous Materials Transportation*, International Series in Operations Research & Management Science 193,
DOI 10.1007/978-1-4614-6794-6_4,

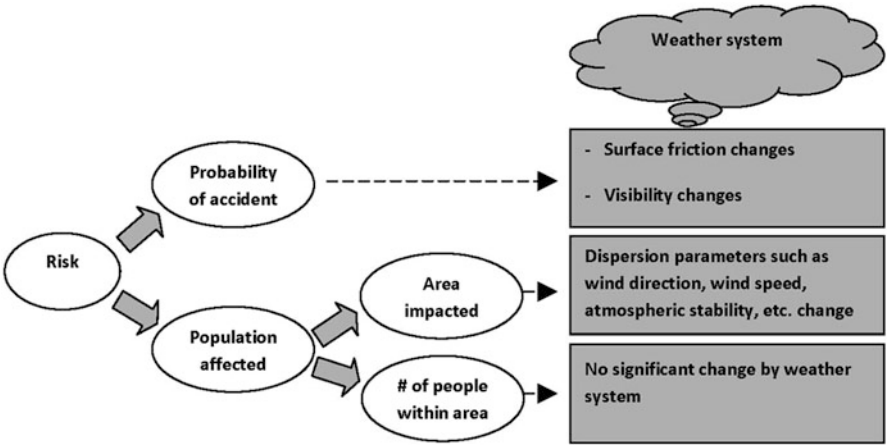


Fig. 1 The effect of a weather system on risk components

Weather systems can affect the risk by changing both the probability of an accident and the affected population. The probability of an accident changes for two reasons. First, the weather system affects the surface friction of a road. For example in rain or snow, the road will be more slippery and more accidents will occur. Second, the weather system changes visibility and the probability of an accident increases as visibility decreases. The weather system also changes the affected population by changing the area impacted by a hazmat accident. This happens because the dispersion of pollutants within the accident area is a function of weather parameters such as wind direction, wind speed, atmospheric stability, etc. Figure 1 illustrates how the weather system changes the risk of a route.

Therefore, a main challenge of considering the weather system in a hazmat transportation model is the way we modify the risk values in the presence of such a system. However, this is not the only challenge. Another important challenge of considering the weather system in any transportation model (i.e., not only hazmat ones) is the dynamic nature of the weather system. In most transportation models, the cost of traveling a link is fixed. Therefore, one can apply a shortest path algorithm to the network and find the optimal route. However, when we consider weather systems in a transportation model, the cost of traveling a link may change as time advances. This may happen because factors such as travel speed change as the weather changes. This dynamic nature of the weather system is even more significant in hazmat models, because as we discussed earlier, the risk changes significantly with changes in weather conditions. Therefore, for considering the dynamic aspect of the weather system, one needs to consider a time dependent shortest path problem instead of a simple shortest path problem. This makes the problem more difficult to solve especially for large transportation networks.

The rest of this chapter is organized as follows. Sections "Effect of Weather Systems on Accident Probabilities" and "Effect of Weather Systems on Dispersion

of Pollutants" explain the necessary modifications in risk factors when one considers a weather system in a hazmat transportation model. This includes the effect of the weather system on the probability of an accident (section "Effect of Weather Systems on Accident Probabilities") and the effect of weather systems on the area impacted by the pollutant (section "Effect of Weather Systems on Dispersion of Pollutants"). Section "The Dynamic Nature of the Weather System" explains the dynamic nature of a weather system and the algorithms that can be used to handle this dynamic nature in a transportation network. Section "Example of Hazmat Transportation Models Which Consider Weather Effects" reviews two models in the literature which successfully incorporated the weather system in a hazmat transportation study. And finally section "Conclusions and Future Research Directions" provides conclusions and future research directions. We note that since most of the related studies in the literature focus on the effect of weather systems on highway transportation, the main focus of this chapter is also on this mode of Hazmat transportation.

Effect of Weather Systems on Accident Probabilities

The presence of a weather system affects both visibility and surface friction thus increasing the probability of an accident. Saccomanno and Chan (1985) analyzed truck accident data for Toronto and estimated the probability of an accident based on visibility and surface friction for different kinds of roads. The results from this analysis are provided in Table 1. The analysis shows that the probability of an accident increases significantly in the presence of a weather system. For example, the probability of an accident increases from 1.478×10^{-6}/mile for dry pavement and unrestricted visibility to 2.740×10^{-6} for wet pavement with restricted visibility. This amounts to an 85% increase due to night travel or weather systems.

The results provided in Table 1 have been used in some other hazmat transportation models. For example, Akgun et al. (2007) modeled the affect of weather

Table 1 Probability of accident based on visibility and surface friction

	Probability ($\times 10^{-6}$/mile)					
Conditions	A	B	C	D	E	F
Dry pavement						
Unrestricted visibility	3.715	2.744	0.876	1.478	2.215	0.830
Restricted visibility	3.957	1.779	1.672	2.519	5.218	0.935
Wet pavement						
Unrestricted visibility	1.816	1.895	0.956	1.728	2.580	0.878
Restricted visibility	0.957	1.737	0.531	2.740	8.279	0.593

Note: A = arterial/collectors with speed = 50 km/h; B = arterial/collectors with speed >50 km/h; C = expressway with speed <100 km/h; D = expressway with speed =100 km/h; E = ramps; and F = major intersections

systems on finding the minimum risk route in a hazmat transportation network. In this model, they assumed that the adverse weather conditions would increase accident rates (and also increase travel times and the affected population). In their numerical experiments, based on the results in Table 1, they assumed the amount of such an increase to be equal to the average value of 60% for interstate highways, US highways, and state highways. We summarize the complete model by Akgun et al. (2007), in section "The Dynamic Nature of the Weather System" of this chapter.

A recent hazmat transportation model that uses the results in Table 1 is provided by Kim et al. (2011). They presented a framework for geographical information system (GIS)-based decision support system (DSS) for vehicle transportation of hazardous materials. Their framework is designed such that it can provide online routing instructions in response to updated information of traffic and weather conditions. The available weather data that are updated online include visibility and atmospheric conditions that affect surface friction. These updated weather data are used to recalculate the accident probabilities of different routes using the results of Table 1. The accident probabilities are then used to calculate the disutility of each link or arc of a transportation network. This probability is the weighted summation of the normalized travel time and the normalized risk of a given arc. This study uses the following formula to calculate the risk of each arc:

$$risk = P(A|W) \times P(R|A) \times Pop$$

Here, $P(A|W)$ is the conditional probability of a truck accident given the weather conditions as explained above (according to Table 1), $P(R|A)$ is the conditional probability of hazmat release given an accident, and Pop is the population exposure to a hazmat release.

Kim et al. (2011) also provide a heuristic method for finding a minimum disutility route and then use the transportation network between Washington, D.C., and Baltimore Maryland in order to demonstrate the ability of the heuristic in dynamically updating the routes in response to real time weather data. An important note for this study is that the weather conditions and wind direction are assumed to not affect the population exposure in calculating risk. Instead, the paper simply assumes that all populations within 5 miles of release points will be exposed to the hazmat risk. However, a more comprehensive approach is to consider the weather conditions including wind direction in calculating the population exposed to a hazmat risk. We study this approach in the next section.

A more recent study that can be used instead of the analysis by Saccomanno and Chan (1985) in order to consider weather conditions in the calculation of accident probabilities is provided by Fabiano et al. (2002). Their study proposes equations (1) and (2) below to calculate the frequency of an accident on a given road i.

$$f_i = \gamma_i L_i n_i \tag{1}$$

$$\gamma_i = \gamma_0 \sum\nolimits_{j=1}^{6} h_j \tag{2}$$

Table 2 Factors correlated to intrinsic road characteristics

Intrinsic characteristics	h_1	h_2	h_3	h_6
Direct road	1			
Road bend (radius >200 m)	1.3			
Road bend (radius <200 m)	2.2			
Plane road		1		
Slope road (gradient <5%)		1.1		
Steep slope road (gradient >5%)		1.2		
Downhill road (gradient <5%)		1.3		
Steep downhill road (gradient >5%)		1.5		
Two lanes for each carriageway			1.8	
Two lanes and emergency lane for each carriageway			1.2	
Three lanes and emergency lane for each carriageway			0.8	
Tunnel				0.8
Bridge				1.2

Table 3 Factors correlated to weather condition

Weather condition	h_4
Fine weather	1
Rain/fog	1.5
Snow/ice	2.5

Table 4 Factors correlated to traffic characteristics

Traffic characteristics	h_5
Low intensity <500 vehicle/h	0.8
Medium intensity <1,250 vehicle/h with heavy traffic <125 truck per day	1
High intensity >1,250 vehicle/h	1.4
High intensity >1,250 vehicle/h with heavy traffic >250 truck per day	2.4

where γ_i is the expected frequency on the ith road stretch (accident km^{-1} per vehicle), L_i the road length (km), n_i is the vehicle number (vehicle), γ_0 the basic frequency (accident km^{-1} per vehicle), and h_j is the local enhancing/mitigating parameters.

Fabiano et al. (2002) also introduce enhancing/mitigating parameters which are subdivided into six categories: h_1 and h_2 refer to geometric characteristics, of the road, h_3 to the type of the roadway, h_4 to the weather condition, h_5 to the type and intensity of the traffic, and h_6 to the presence or absence of a tunnel or bridge. These six parameters have been estimated using historical data from a transportation network starting from the Genoa port area towards the industrialized North Italian and Central Europe districts (see Tables 2, 3 and 4).

Effect of Weather Systems on Dispersion of Pollutants

In order to calculate the risk of a particular route in a hazmat transportation system, one needs to estimate the population which may be affected near the area of an accident. To capture this, there are two types of approaches found in the literature: The first approach assumes that the population within a predefined threshold distance from an accident is exposed to the pollutants. The second approach uses different dispersion models to determine the area affected by an airborne pollutant. In the following sections we discuss these two approaches and how they are affected by the weather system.

Threshold Distance Approach

The Threshold distance approach was first employed by ReVelle et al. (1991) and Batta and Chiu (1988). It assumes that people within a fixed radius of an accident will be affected by the pollutant. For example, according to the North American Emergency Response Handbook (2004), an area with a radius of 800 m around a fire that involves a chlorine tank, railcar or tank-truck must be isolated and evacuated. While considering a weather system in a hazmat transportation model, one needs to make adjustments to this fixed threshold approach. For example, Akgun et al. (2007) assumes that a hazmat release at a specific location poses a threat to a neighboring point if it is within a threshold distance, λ, of that neighboring point. When the release location is affected by a weather system, the threshold distance changes to λ', which may be greater or smaller than λ depending on the type of weather system and the type of hazardous material. However, Akgun et al. (2007) do not provide a detailed procedure for calculating λ' thus leaving an important unanswered question of how to change λ to λ' based on the parameters of a weather system. A paper that gives an answer to this question is provided by Brown and Dunn (2007). They present a Monte Carlo simulation based method to evaluate distances over which the public should be protected in the event of a hazmat release involving an airborne hazard. Their approach explicitly considers weather conditions, including wind velocity, in order to determine a 90th-percentile safe distance from the incident for given release characteristics. Unlike Akgun et al. (2007), this 90th-percentile safe distance is not a circular region. However, as shown in Fig. 2, this region forms a square region having a side dimension equal to and lying downwind and symmetrically about the accident location. A square shape is chosen in this paper in order to provide a simple, familiar way of defining the zone and, more importantly, to account for uncertainty in the plume trajectory owing to wind direction variability and possible effects of topographical features. In Fig. 2, the circular region surrounding the accident site is the zone from which persons not involved with the response should be kept clear.

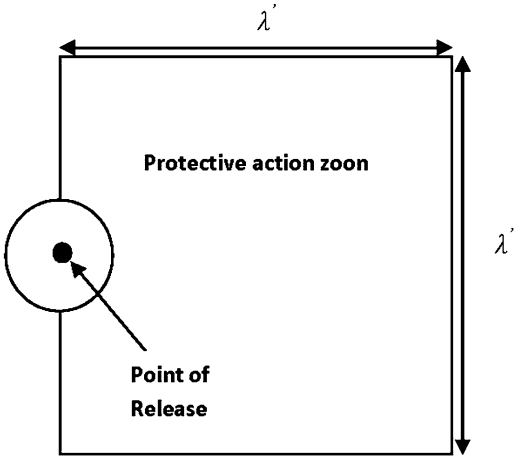


Fig. 2 The protective distance in Dunn (2007)

Air Pollution Dispersion Models

Using air pollution dispersion models is another approach in which we can directly consider weather parameters in determining the area affected by a pollutant. There are various air pollution dispersion models found in the literature. However, the most common hazmat transportation studies use a Gaussian model which is very popular due to its simplicity and ease of use. Below we summarize the application of dispersion models in determining the area affected by hazmat release.

Gaussian Model

The Gaussian model is perhaps the most commonly used air pollutant dispersion model. It assumes that the air pollutant dispersion has a Gaussian (i.e., normal) distribution. The earliest study that used the Gaussian distribution for modeling air pollutant dispersion is developed by Sir Graham Sutton (1947). Further advances in this model were made by Briggs (1965) in model refinement and validation. The Gaussian model is most often used for predicting the dispersion of continuous, buoyant air pollution plumes originating from ground-level or elevated sources. This model may also be used for predicting the dispersion of non-continuous air pollution plumes (called puff models).

The Gaussian plume model is based on several limiting assumptions. Zhang et al. (2000) characterize these assumptions and their validity for hazmat transportation models as follows:

1. The gas does not change its chemical properties during dispersion: this assumption restricts the applicability of the Gaussian plume model to stable chemicals and to accidents which do not result in explosion.

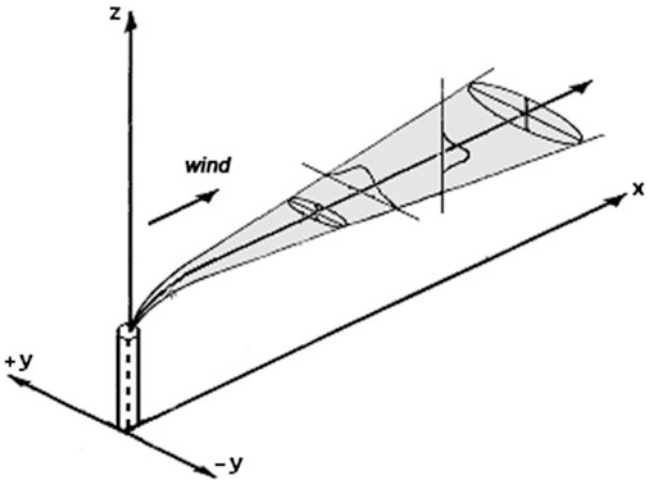


Fig. 3 The Gaussian plum model

2. Atmospheric conditions are homogeneous in the study area and constant over the period of dispersion: this assumption is fairly realistic for hazmat risk analysis because the impact area of accidents is usually small (within 10 km), and because accidental release of hazmat usually occurs within a short period of time.
3. The terrain is gentle or flat, and the ground surface does not absorb the gas: the validity of this assumption depends on the nature (paved, agricultural) of the release site. For the cases that this assumption is not valid, there are some modified versions for Gaussian plume models in the literature that take terrain into account. To see an example of such models, readers are referred to Strimaitis et al. (1986).
4. The rate of emission is continuous and steady: this assumption is violated by transportation related releases which usually experience a quick and short-lived emission. The steady-state concentrations estimated by Gaussian plume models assume a continuous release source, and are likely to overestimate location-specific concentrations. This assumption is particularly important, and it limits the applicability of Gaussian plume models to gaseous dangerous goods accidents.

The Gaussian plume model is shown in Fig. 3. We employ the following notation and present the model in (3):

- x: The distance downwind from the source.
- y: The distance crosswind (perpendicular) from the source.
- z: The elevation of the destination point.
- h_e: The elevation of the source.
- $C(x,y,z,h_e)$: The concentration level at point (x,y,z).
- Q: The release rate of pollutant.

- u: The average wind speed.
- σ_y and σ_z: Horizontal and vertical plume standard deviations as a function of x (also known as dispersion coefficients).

$$C\,(x, y, z, h_e) = \frac{Q}{2\pi u \sigma_y \sigma_z} \exp\left(-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2\right)\left[exp\left(-\frac{1}{2}\left(\frac{z-h_e}{\sigma_z}\right)^2\right) + exp\left(-\frac{1}{2}\left(\frac{z+h_e}{\sigma_z}\right)^2\right)\right] \tag{3}$$

In most hazmat transportation models, we can assume the release source to be near the ground. In this case one can assume $h_e = 0$, $z = 0$, and obtain the concentration level on the ground using the following equation (4):

$$C\,(x, y, z, h_e) = \frac{Q}{\pi u \sigma_y \sigma_z} \exp\left(-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2\right) \tag{4}$$

In this case the concentration on the plume's center line ($y = 0$) is represented by (5):

$$C\,(x, y, z, h_e) = \frac{Q}{\pi u \sigma_y \sigma_z} \tag{5}$$

The dispersion coefficients σ_y and σ_z are dependent on the stability of the atmosphere and the downwind distance x to the release source. There are different methods for determining these two parameters. The most commonly used method in the literature is power law functions (Turner 1969) provided in (6) and (7). In this method, popular due to simplicity and ease of use, the dispersion parameters (a, b, c, and d) are determined based on different stability categories as defined by Pasquill (1962). These categories include: very unstable (A), moderately unstable (B), slightly unstable (C), neutral (D), slightly stable (E) and moderately stable (F). The value of these parameters for each of these six stability categories can be found in Turner (1969).

$$\sigma_z = \mathrm{a}.x_b \tag{6}$$

$$\sigma_z = c.x_d \tag{7}$$

The first studies that incorporated Gaussian plume models into route selection in a hazmat transportation network are provided by Patel and Horowitz (1994) and Karkazis and Boffey (1995). In particular, these two papers use Gaussian plume models to determine the population affected by a hazmat release when calculating the risk of each route. Then they provide methods for selecting a route

with minimum risk in the network. This is conceptually simple but computationally intensive because, in order to calculate the risk associated with each link one needs to integrate the Gaussian plume model over the length of the link (see section "The Model by Patel and Horowitz (1994)"). This integration is computationally prohibitive if it is calculated numerically.

To solve this problem of computational complexity, Patel and Horowitz (1994) made simplifying assumptions about the dispersion parameters and achieved a reasonable level of computational efficiency. They assumed the dispersion coefficients σ_y and σ_z to be linear functions by fixing the dispersion parameters b and d to be equal to 1.0. In particular, they simplified (6) and (7) to the linear equations $\sigma_z = a.x$ and $\sigma_z = c.x$.

Zhang et al. (2000) explained that considering dispersion parameters b and d to be equal to 1.0 may be a damaging assumption since these parameters differ by atmospheric stability category and assuming them to be 1.0 entirely ignores the matter of atmospheric stability. Therefore, in order to overcome the computational complexity of using the Gaussian plume model in calculating the risk of travel on links, they adopted a raster GIS framework. This raster framework transforms a continuous space into a discrete one by modeling it as a tessellation of square grid cells called pixels. Working in discrete space, enables the authors to compute the concentrate levels more efficiently without having to make the linearity assumption of Patel and Horowitz (1994). If the resolution (pixel size) of the raster representation is lower, the computation of concentrate levels will be more efficient but the accuracy of computations will be worse. Therefore, the choice of the resolution level in this raster GIS framework is a tradeoff between accuracy and computational efficiency.

Although a significant portion of hazmat transportation is done via railroad, all of the above mentioned Gaussian models have been developed for highway transportation. Verma and Verter (2007) extended the Gaussian plume model for the case which involves multiple release sources as happens when a train carrying multiple hazmat cargoes is involved in an accident. They used their proposed Gaussian model to develop a risk assessment method for railroad transportation of hazardous materials. This method works with the most conservative estimate of weather conditions, which will arguably enable preparedness for any weather situation.

Other Dispersion Models

As we discussed earlier, even the Gaussian plume model (which is a very simple model) is computationally intensive when used in a route selection model for a hazmat transportation network. Therefore, using other dispersion models that are more complicated (compared to the Gaussian plume model) makes the resulting routing model even more inefficient in terms of computations. Hence the application of other models in hazmat transportation studies is limited and to the best of our

knowledge does not include routing models for a large transportation network. Below we review two studies that use non-Gaussian dispersion models:

- *Dense gas models*: Dense gas models are a class of air pollution dispersion models that simulate the dispersion of pollution plumes that are heavier than air (also known as dense gas). Leeming and Saccomanno (1994) used the dense gas model in a case study involving the risk assessment of two routes in transportation of chlorine from a specific supplier to an industrial facility. The routing options include a road and a rail route each of which is divided into 14 and 18 subsections respectively. The risk along each subsection is assumed to be constant. This assumption makes the computations much easier because it avoids integrating the dispersion model over the length of the route (as is done in some Gaussian plume models such as Patel and Horowitz 1994). To find the risk of these two routing options, the authors considered different release scenarios based on weather condition, stability, wind speed and wind rise, and estimated the area affected by a hazmat release in each scenario (using a dense gas dispersion model).
- *Lagrangean models*: Lagrangean dispersion models consider the movement of air pollutants as a random walk process. They calculate the dispersion of pollutants by first defining a box that contains an initial concentration of pollutants, and then following the trajectory of the box as it moves downwind. These models use a moving frame of reference to follow the pollution plume as it moves in the atmosphere. The Lagrangean model has been used by Hwang et al. (2001) to estimate impact zones for six toxic-by-inhalation materials. These impact zones are then used for risk assessment for the national transportation of these hazardous materials.

The Dynamic Nature of the Weather System

Weather conditions change dynamically. Therefore, those metrics for a route that are dependent on the weather condition (such as risk and travel time) will also change as weather changes. This makes the problem of finding an optimal route more difficult. There are different approaches toward this complicating aspect of the weather system as outlined below:

Some studies do not consider the dynamic nature of the weather system and assume that the weather parameters to be fixed during the vehicle movement from origin to destination (e.g., Zhang et al. 2000). This may be a valid assumption if the travel time is short enough. However, for large transportation networks in which the travel time from the origin to the destination may take a significant part of a day or several days, we cannot ignore the changes in the weather parameters.

Some studies do not model the dynamic nature of the weather system and simply use the historical data for a given region and assign a probability distribution for each weather scenario (e.g., the wind direction is from the north with a probability

of 30%, from the east with a probability of 70%, and has a negligible probability for other directions). This probability distribution can then be used to calculate an average affected population or average accident probability. An example of this approach can be seen in Karkazis and Boffey (1995). This approach may cause large error in risk calculations for any particular route/day because it doesn't consider the real time weather data.

Another approach that is more popular and more realistic is based on the assumption that those regions that are close to the current location of a vehicle will be traveled to in the near future. Therefore, it is reasonable to use current weather data for those points. However, for other regions that are relatively far from the current location of the vehicle, we cannot use the current weather parameters because the weather is likely to change by the time the vehicle reaches that region. Therefore, this approach uses real time weather data for those regions within some neighborhood area of the current location of the vehicle and uses historical or predicted weather data for other regions. Examples of this approach can be found in Patel and Horowitz (1994) and Kim et al. (2011).

Another method (that is also more difficult to implement) is more fully modeling the dynamic nature of the weather conditions. In this approach, the weather condition on different regions of the transportation network is to be predicted for different time intervals (e.g., prediction by modeling the movement of the weather system as done in Akgun et al. 2007). Then these predictions are used to calculate the risk (or other factors such as travel speed) as a function of time. Since in this approach, risk is dependent on time, one needs to solve a *time dependent shortest path problem* to calculate the route with lowest risk. This problem can be defined on either discrete or continuous time scales depending on whether time is discretized and thus integer-valued or real-valued. The general time dependent shortest path problem is at least NP-Hard since it may be used to solve a variety of NP-Hard optimization problems such as the knapsack problem. However, depending on how one defines the problem, it may not be in NP since its output is not polynomially bounded. Different methods have been developed for solving this problem. Here we summarize some of these methods:

- Cooke and Halsey (1966): This discrete time algorithm is the first algorithm developed for the time dependent shortest path problem. The algorithm solves the problem from every node to the destination using a finite number of iterations, and for any finite initial starting time. It has theoretical computational complexity $O(V^3M)$, where V is the number of nodes and M is the number of time steps.
- Dreyfus (1969): This algorithm is a label setting procedure that is a generalization of Djikstra's (1959) static shortest path algorithm. The algorithm calculates the time dependent shortest path between two nodes for one departure time step in $O(V^2)$.
- Orda and Rom (1990): This algorithm finds the time dependent shortest path under various waiting constraints. Unlike Dreyfus (1969), this algorithm is not limited to FIFO (First-in-First-out) links.

- Ziliaskopoulos and Mahmassani (1993): This algorithm calculates the time dependent shortest paths from all nodes to the destination for every time step over a given time horizon.
- Cai et al. (1997): This set of algorithms is proposed for solving three cases of time dependent shortest path problems. In the first case, waiting at a vertex without any restriction is allowed. In the second one, waiting at any vertex is strictly prohibited; and in the third case, there is a vertex-dependent upper bound on the waiting time at each vertex.
- Chabini (1998): This work suggests an efficient algorithm for all-to-one (i.e., all nodes to one destination node) discrete dynamic shortest path problem. This algorithm is proved to have an optimal run time complexity that equals the complexity of the problem.

In section "Example of Hazmat Transportation Models Which Consider Weather Effects," we study an example of the dynamic model of the weather system that leads us to a time dependent shortest path problem. This model is provided by Akgun et al. (2007) and studies the movement of the weather system along the transportation network in order to predict the weather parameters as a function of time. This leads to time dependent risk values for different links and nodes. Therefore, one can find the least risk path by using the above discussed algorithms.

Finally we note that a more general form of the time dependent shortest path problem occurs when the parameters associated with each link (e.g., travel time, risk, etc.) are both time dependent and probabilistic (rather than deterministic). In this situation each arc of the transportation network is assigned travel time and travel risk random variables with time-varying probability distribution functions. Solution procedures for finding optimal routs in such situations can be found in Miller-Hooks and Mahmassani (1998, 2000).

Example of Hazmat Transportation Models Which Consider Weather Effects

Reviewing the literature of hazmat transportation models, one can find that current studies which consider the effect of weather systems have at least one of the three following components: first, the accident probabilities are adjusted based on the weather condition; second, the area affected by a hazmat release is calculated using a Gaussian plume model (in rare cases other dispersion models have been used); and third, the dynamic nature of the weather system may be considered by modeling the problem as a time dependent shortest path problem. To the best of our knowledge there is no model in the literature that incorporates all three of these components at the same time (due to the difficulty of considering all of these together). In this section we review two models in the literature that cover these three components. The first model by Akgun et al. (2007) considers the dynamic nature of weather systems and updates the accident probabilities and speed of the

vehicle as the weather changes (first and third components). The second model by Patel and Horowitz (1994) consider the Gaussian plume model in calculating the area affected by a hazmat release (second component).

The Model by Akgun et al. (2007)

Akgun et al. (2007) deals with the problem of finding a least risk path for hazmat transportation on a network exposed to a weather system. The weather system is assumed to have circular shape and move in a linear direction along the transportation network. The paper first uses the following notation and models the dynamic variations of the attributes of a single link (including travel time and risk) as the weather system moves along the network.

- $G = (N,E)$: Connected, undirected and planar transportation network with node set N and link set E.
- (x_s,y_s): Cartesian coordinates of a given point s on the plane.
- w_i: Positive weigh of node i. This weigh typically signifies the population at this node.
- $g_{(i,j)}$: Positive weigh of the link (i,j).
- $l_{(i,j)}$: The length of the link (i,j). For a single link (i,j), it is assumed (without loss of generality) that this link is horizontal to the Cartesian plane with the coordinates of node j being $(0, l_{(i,j)})$.
- $f_{(i,j)}(z)$: Population density function of the link (i,j) where z ranges from 0 to $l_{(i,j)}$. This density function is normalized such that $\int_0^{l_{(i,j)}} f_{(i,j)}(z) = 1$.
- $d(a,b)$: Euclidean distance between two points a and b on the plane that is: $d\,(a,b) = \sqrt{[x_a - x_b]^2 + [y_a - y_b]^2}$.
- $\mathscr{W}$: Weather system that is assumed to be circular and travels on a straight line with constant speed.
- r: Radius of $\mathscr{W}$.
- $(x_w(t), y_w(t))$: Center of $\mathscr{W}$ at time t.
- $(x_w, y_w) = (x_w(0), y_w(0))$: Center of $\mathscr{W}$ at time $t = 0$.
- θ: Angle between x axis and travel direction of $\mathscr{W}$.
- v_w: Constant speed of $\mathscr{W}$.
- $(x(t), y(t))$: Location of hazmat carrying vehicle.
- v and v': Speed of hazmat carrying vehicle inside and outside of $\mathscr{W}$ respectively.
- h_i, h_j and h'_i, h'_j: Probability of accident in node i and j when the vehicle is traveling link (i,j) inside and outside of $\mathscr{W}$ respectively.
- $q_{(i,j)}$ and $q'_{(i,j)}$: Probability of accident per unit length of movement when the vehicle is traveling link (i,j) inside and outside of $\mathscr{W}$ respectively.
- $d(t)$: The distance between the hazmat carrying vehicle and the center of $\mathscr{W}$.
- t_1^*: The first time the vehicle enters $\mathscr{W}$ (vehicle traveling a single link).
- t_w^*: The duration the vehicle stays in $\mathscr{W}$ (vehicle traveling a single link).

- $t_2^* = t_1^* + t_w^*$: The time the vehicle leaves $\mathscr{W}$ (vehicle traveling a single link).
- $\delta(z,c)$ and $\delta'(z,c)$: Indicator function to show whether a hazmat release at point c poses a threat to point z when point c is outside and inside of the weather system respectively.
- $R_{(i,j)}$: The risk of traveling on link (i,j).
- $F(z;i,j)$: The threat that traveling on link (i,j) poses on point z.

Using the above notation, one can find the center of $\mathscr{W}$ at time t using the following equations (8) and (9):

$$x_w(t) = x_w + (v_w \cos\theta)t \tag{8}$$

$$y_w(t) = y_w + (v_w \sin\theta)t \tag{9}$$

Assuming that the vehicle starts at node i at time $t = 0$ and travels towards node j along the horizontal link (i, j), the location of hazmat carrying vehicle at time t is:

$$(x(t),\ y(t)) = (vt, o) \tag{10}$$

Therefore, the distance between the vehicle and the center of $\mathscr{W}$ at time t can be obtained by the following equation (11):

$$d^2(t) = [x(t) - x_w(t)]^2 + [y(t) - y_w(t)]^2 = At^2 + Bt + C \tag{11}$$

where,

$$A = v^2 - 2vv_w\cos\theta + v_w^2,\ B = 2x_w v_w \cos\theta + 2y_w v_w \sin\theta - 2vx_w,\ \text{and } C = x_w^2 + y_w^2.$$

Using (8)–(11), the paper studies the effect of the weather system on travel time by considering the following three possible cases:

1. The vehicle is never in $\mathscr{W}$.
2. The vehicle starts outside $\mathscr{W}$, and then enters $\mathscr{W}$.
3. The vehicle starts in $\mathscr{W}$.

In the latter two cases the vehicle either:

(a) Stays in $\mathscr{W}$ until node j is reached or
(b) Exits $\mathscr{W}$ before reaching node j.

For finding t_1^*, Akgun et al. solve the equation $d^2(t_1) = r^2$ using the vehicle location of (10) and the coordination of the weather system following equations (8) and (9) and finds the roots of this quadratic equation to be:

$$t_1^{\pm} = \frac{-B \pm \sqrt{B^2 - 4A(C - r^2)}}{2A} \tag{12}$$

The smaller root, t_1^-, corresponds to an entrance time (the larger root, t_1^+, corresponds to an exit time only under the assumption that the vehicle maintains a speed $v' = v$ inside $\mathscr{W}$). Suppose the roots are real (if not, then we are in Case 1 and assign $t_1^* = t_2^* = -\infty$). If $t_1^- \leq 0$, then $t_1^* = 0$ and we are in Case 3; else if $t_1^- \leq l(i,j)/v$ (that means t_1^- is less than the amount of time to traverse the link (*i*,*j*) in normal conditions), then $t_1^* = t_1^-$ and we are in Case 2; else we are in Case 1 and assign $t_1^* = t_2^* = -\infty$.

After finding t_1^* as explained above, one can find t_w^* for Cases 2 and 3 by solving the equation $d^2(t_w) = r^2$ and using the coordinates $x(t_w) = vt_1^* + v't_w$, $y(t_w) = 0$, $x_w(t_w) = x_w t_1^* + (v_w \cos\theta)t_w$, and $y_w(t_w) = y_w t_1^* + (v_w \sin\theta)t_w$.

The paper then generalizes the results for the above traveling time (in which it was assumed that the vehicle starts traveling at time 0) to the case that it is possible to start travel at any time s in the time window $[0, T_s]$. This is done by changing (11) to the following equation (13):

$$d^2(t) = At^2 + B(s)t + C(s) \tag{13}$$

where,

$$B(s) = B + 2v_w(v_w - v\cos\theta)s, \text{ and } C(s) = C + v_w^2 s^2 + 2(x_w v_w \cos\theta + y_w v_w \sin\theta)s.$$

After this generalization, the Akgun et al. prove a theorem which says that the time that the vehicle spends in the weather system while traversing a link is concave in the starting time $s \in [0, T_s]$. This theorem suggests that in order to minimize exposure to the weather system, a vehicle should begin travel on a link at one of the endpoints of the permissible travel window, $[0, T_s]$.

After analyzing the weather effects on travel time, the paper studies the weather effects on risk as discussed below.

Akgun et al. first use the threshold distance approach explained in section "Threshold Distance Approach" and define the indicators, $\delta(z,c)$ and $\delta'(z,c)$ to show whether a hazmat release at point *c* poses a threat to point *z* when point *c* is outside and inside of the weather system respectively (λ and λ' are the threshold distances explained in section "Threshold Distance Approach"):

- $\delta(z,c) = 1$ if $d(z,c) \leq \lambda$; 0 otherwise.
- $\delta'(z,c) = 1$ if $d(z,c) \leq \lambda'$; 0 otherwise.

The risk of traveling on link (*i*,*j*) is then defined in (14) in which the first term accounts for the potential damage that the vehicle could do to nodal population centers on the network, and the second term does the same for population that resides on links of the network (which is continuously distributed).

$$R_{(i,j)} = \sum\nolimits_{z \in N} w_z F(z;i,j) + \sum\nolimits_{l,k \in A} \int_0^{l_{(l,k)}} g_{(l,k)} f_{(l,k)}(z) F(z;i,j)\, dz \tag{14}$$

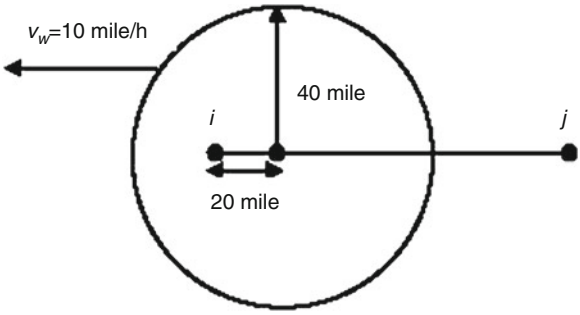


Fig. 4 Numerical example 1

For calculating $F(z;i,j)$ in the above equation (14), one needs to consider the different cases of $x(t_1^*)$ and $x(t_2^*)$ when the vehicle is traveling on link (i,j). For example consider the case $x(t_1^*) \in (0, l(i,j))$ and $x(t_2^*) \in (0, l(i,j))$. For this case $F(z;i,j)$ can be obtained as provided in (15) (other cases will be handled similarly). In this equation the first and last terms account for the possibility of an accident at node i and node j, respectively, which in this case occur with the vehicle outside the weather system $\mathscr{W}$. The second and fourth terms are to account for the possibility of an accident along the link before entering and after exiting $\mathscr{W}$, respectively. The third term is for an accident along the link while in $\mathscr{W}$.

$$F(z;i,j) = h_i \delta(z,i) + \int_0^{x(t_1^*)} q_{(i,j)} \delta(z,(y,0))\, dy + \int_{x(t_1^*)}^{x(t_2^*)} q'_{(i,j)} \delta'(z,(y,0))\, dy + \int_{x(t_2^*)}^{l_{(i,j)}} q_{(i,j)} \delta(z,(y,0))\, dy + h_j \delta(z,j) \tag{15}$$

After calculating the time dependent risk values using (14) and (15), Akgun et al. largely focus on finding the minimum risk route in the obtained time dependent shortest path problem. This can be done using the exact methods explained in section "The Dynamic Nature of the Weather System," or alternatively by using heuristic methods provided in the paper.

Example We now demonstrate how to find the risk of traveling a single link (i,j) for the problem illustrated in Fig. 4. Note that the length of link (i,j) is equal to 100 miles and the speed of the vehicle is 20 and 40 mile/h inside and outside of the weather system respectively. We assume that the accident probability ($\times 10^{-6}$/mile) for a vehicle which is not affected by a weather system is equal to 2 for the nodes and 3 for the links and that these rates will be doubled in the presence of the weather system. We also assume that in case of a hazmat release the affected area will have a radius of 2 miles and this value is the same inside and outside of the weather system. And finally, we assume that the population is uniformly distributed along the link without any significant extra population on Nodes i and j. We consider all other weights to be equal to one.

Solution Since the movement of the vehicle and the weather system are both only in the horizontal direction, we only need to consider the x direction (y is always

equal to zero). We know that $t_1^* = 0$ (Since the vehicle starts from inside the weather system). We find t_w^* in the following (note that we consider node i as the origin of our coordinate sysytem):

$$d\,(t_w) = r = 40 \rightarrow \sqrt{[x\,(t_w) - x_w\,(t_w)]^2 - 0} = 40$$

$$\rightarrow [vt_w] - [x_w + (v_w \cos\theta) t_w] = 40$$

$$\rightarrow [20t_w] - [20 + (10)\,(-1)\,t_w] = 40 \rightarrow\ t_w^* = 2$$

Therefore we can obtain $t_2^* = t_1^* + t_w^* = 0 + 2 = 2$ resulting in $x\left(t_2^*\right) = 2 * 20 = 40$ miles. This means that the vehicle travels the first 40 miles inside the weather system, and then it exits the weather system and travels the remaining 60 miles outside the weather system.

Now we can calculate the threat of traveling on this link at a given point z as follows (note that in this example $\delta = \delta'$):

$$F\,(z; i, j) = 4\delta\,(z, i) + \int_0^{40} 6\delta\,(z, x)\,dx + \int_{40}^{100} 3\delta\,(z, x)\,dx + 2\delta\,(z, j)$$

To evaluate $F(z;i,j)$ we consider the following possible cases for z:

- $0 \le z \le 2:\ F\,(z; i, j) = 4 + \int_0^{z+2} 6dx = 6z + 16$
- $2 \le z \le 40:\ F\,(z; i, j) = \int_{z-2}^{z+2} 6dx = 24$
- $40 \le z \le 98:\ F\,(z; i, j) = \int_{z-2}^{z+2} 3dx = 12$
- $98 \le z \le 100:\ F\,(z; i, j) = 2 + \int_{z-2}^{100} 3dx = 308 - 3z$

Therefore, the total risk is:

$$R_{(i,j)} = F\,(z = i; i, j) + F\,(z = j; i, j) + \int_0^{100} \frac{1}{100} F\,(z; i, j)\,dz$$

$$= 0 + 0 + \frac{1}{100}\left[\int_0^2 F\,(z; i, j)\,dz + \int_2^{40} F\,(z; i, j)\,dz + \int_{40}^{98} F\,(z; i, j)\,dz + \int_{98}^{100} F\,(z; i, j)\,dz\right]$$

$$= \frac{1}{100}\left[\int_0^2 (6z + 16)dz + \int_2^{40} 24dz + \int_{40}^{98} 12dz + \int_{98}^{100} (308 - 3z)dz\right]$$

$$= \frac{1}{100}\,[44 + 912 + 696 + 22] = 16.74$$

Note that if we ignore the impact of the weather system in the above example, the total risk can be obtained as follows:

$$R_{(i,j)} = 0 + 0 + \frac{1}{100}\left[\int_0^2 (3z+8)dz + \int_2^{98} 12dz + \int_{98}^{100} (308-3z)dz\right]$$
$$= 11.96$$

The above value is 29% lower than the previous case in which the effect of weather is not ignored.

The Model by Patel and Horowitz (1994)

The study by Patel and Horowitz (1994) is the first paper in the literature that uses the Gaussian plume model for estimating the affected population in a hazmat transportation model. The paper tries to find the minimum risk route for a hazmat transporting vehicle on a network embedded on a (uv) Euclidean plane. The optimal route refers to the path P that minimizes the total risk posed to the entire area in the plane containing the transportation network. In this paper the total risk is considered as total expected concentration of hazardous gas and is defined according to the following equation (16):

$$R(P) = \sum_{(i,j)\in P} \int_u \int_v R_{ij}\,(u,v)\,du\,dv + \sum_{n\in P} \int_u \int_v R\,(n;u,v)\,du\,dv \quad (16)$$

where

- $R(P)$: Total risk of a given path P.
- $R_{ij}(u,v)$: The risk posed by link (i,j) to point (u,v).
- $R(n;u,v)$: The risk posed by node n to point (u,v).

For computational purpose, the paper describes population areas as mesh points that are distributed throughout the study area and evaluates (16), by using mesh points (u_k,v_k) on the uv plane. This enables the authors to approximate (16) with the following equation (17):

$$R(P) = \sum_{(i,j)\in P} \sum_{u_k} \sum_{v_k} R_{ij}\,(u_k,v_k) + \sum_{n\in P} \sum_{u_k} \sum_{v_k} R\,(n;u_k,v_k) \quad (17)$$

It is obvious that the larger the number of these mesh points, the better (17) can approximation (16). To simplify the notation, the paper denotes the mesh point

(u_k,v_k) by k and defines the risk posed by a link and a node to a particular mesh point according to the following equations:

$$R_{ij}(k) = I_{ij} \int_{\lambda} w(\lambda)\, d\lambda \tag{18}$$

$$R(n;k) = I_n w(\lambda) \tag{19}$$

where

- λ: Measures the Euclidian distance from a potential accident to point k.
- $w(\lambda)$: Weight-function based on the Gaussian plume model of (3).
- I_{ij}: The risk intensity of link (i,j) in terms of the probability of a crash.
- I_n: The risk intensity of node n (i.e., the probability of an accident in that node).

To calculate $w(\lambda)$, one needs to integrate the Gaussian plume model over the length of a link. Therefore, for moderate size problems, evaluation of (16) may be too difficult if $w(\lambda)$ is evaluated numerically. However, if $w(\lambda)$ can be expressed in closed form, a reasonable computational efficiency can be achieved. For this purpose, as explained in section "Air Pollution Dispersion Models," Patel and Horowitz first make a simplifying linear assumption in dispersion parameters of the Gaussian plume model and then calculate closed form expressions for $w(\lambda)$ considering different cases of wind direction including:

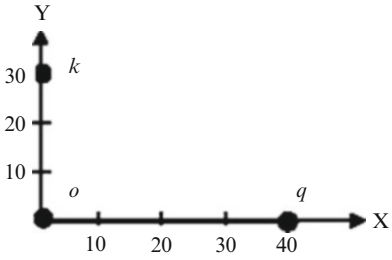
1. Uniform average wind direction
2. Maximum concentration wind direction
3. Specific wind direction

Since the detailed calculations for obtaining the closed form expression of $w(\lambda)$ for each of the above mentioned cases are very technical, we invite the reader to refer to Patel and Horowitz (1994) for all details. Here we summarize some results obtained from the calculation of these closed form expressions of $w(\lambda)$:

- The closed form expression of $w(\lambda)$ for uniform average wind direction (Case 1), differ from the one for maximum concentration wind direction (Case 2) by only a constant. That constant is the same for links and nodes and does not depend on environmental considerations such as wind speed or population distribution. Therefore, any optimal route would be invariant with the choice of Case 2 or Case 3 wind direction.
- The optimal route will not vary with the type of hazardous material, the size of the spill, the type of vehicle or container, and the assumed wind speed.
- Several case 3 (specific wind direction) can be averaged to cover non-uniform wind direction.

Example Here we calculate the risk imposed by a vehicle traveling link (o, q) to the single mesh point k shown in Fig. 5. Assume the risk intensity of an accident to be 0.6 for the link and negligible for the nodes and the emission rate in case of

Fig. 5 Numerical example 2



an accident to be 1,500 mg/s. Also consider uniform average wind direction with average wind speed of 5 m/s and dispersion parameters $a = 0.105$ and $c = 0.060$.

Note: According to Appendix B of Patel and Horowitz (1994), the closed form function of $w(\lambda)$ for a vehicle traveling link (o, q) under uniform average wind direction is $\frac{Q}{\pi uac}\frac{1}{\sqrt{2\pi}}\frac{a}{\lambda\sqrt{\lambda^2-\lambda_o^2}}$.

Solution Using the given parameters and knowing that the distance from node o to mesh point k is equal 30 (i.e., $\lambda_o = 30$), the $w(\lambda)$ function becomes: $w(\lambda) = \frac{634.9}{\lambda\sqrt{\lambda^2-30}}$.

Therefore, the total risk imposed by link (o, q) on the single mesh point k is:

$$R_{oq}(k) = I_{oq}\int_{\lambda} w(\lambda)\, d\lambda = 0.6\int_{30}^{50}\frac{634.9}{\lambda\sqrt{\lambda^2-30}}d\lambda = 8.559$$

Conclusions and Future Research Directions

We have discussed the major effects of a weather system on mathematical models involving the search of for an optimal route in hazmat transportation networks. We explained how the presence of the weather system affects the risk of the routes in such models and we showed the necessary modifications in risk calculations. We also explained the dynamic nature of the weather system and the way it affects the modeling and particularly the solution approaches for a hazmat transportation problem. We reviewed the literature of hazmat transportation models that consider the effect of weather systems in their studies and briefly summarized two such models. Below, we summarize some of the future research directions:

1. *Using GIS to study more realistic and complicated models*: GIS has already been used in by a few researchers as a tool to study the effect of weather systems in hazmat transportation networks (e.g., Zhang et al. 2000 and Kim et al. 2011). However, advances in computer capabilities make it possible to use GIS for more realistic and complicated models (comparing those that are already in the literature). For example the model by Zhang et al. (2000) does not consider the effect of a weather system in calculating the population affected by

an accident and the model by Kim et al. (2011) does not consider the dynamic nature of the weather system. Extending these models to more realistic cases can be considered an interesting area for future research.

2. *Using different dispersion models based on the type of hazmat and release condition*: Since the Gaussian plume model is the most convenient dispersion models in terms of computational efficiency, most of the hazmat studies use this model when modeling the dispersion of the pollutant. However, considering the assumptions of the Gaussian Plume model, this model is not a suitable tool for many kinds of hazmat and many accident scenarios. For example, consider an accident releasing hazardous vapors that are heavier than air (such as toluene). Since the Gaussian plume model is developed for buoyant gases, it cannot cover such cases. To study such cases, the use of other dispersion models should be considered. Currently there are a few papers in the literature that use other dispersion models for some specific applications as explained in section "Air Pollution Dispersion Models." However, more general applications and modeling components (as explained in the beginning of section "Example of Hazmat Transportation Models Which Consider Weather Effects") for these other dispersion models are still an open area for future research.
3. *More realistic modeling for the shape and the movement of the weather system*. The only hazmat transportation study that models the shape and the movement of the weather system in order to capture the dynamic nature of the weather system is provided by Akgun et al. (2007). That study models the weather system as a circle that moves in a straight line. However, this simple shape and movement of the weather system seems to be too simple to capture the real weather conditions in a network. Therefore, more realistic modeling of a weather system can be another area for future research. For example the weather system may have any shape other than a circular one or it can move in a non-straight direction. Some initial unpublished work by Wang et al. (2008) was presented at the INFORMS Annual Meeting using one or more ellipses in place of a single circle.

References

Abkowitz M, Cheng PD (1988) Developing a risk-cost framework for routing truck movements of hazardous materials. Accid Anal Prev 20:39–51

Akgun V, Parekh A, Batta R, Rump C (2007) Routing of a hazmat track in the presence of weather systems. Comput Oper Res 34:1351–1373

Batta R, Chiu SS (1988) Optimal obnoxious paths on a network: transportation of hazardous materials. Oper Res 36(1):84–92

Briggs GA (1965) A plume rise model compared with observations. J Air Pollut Central Assoc 15:433–438

Brown D, Dunn W (2007) Application of a quantitative risk assessment method to emergency response planning. Comput Oper Res 34:1243–1265

Cai X, Kloks T, Wong CK (1997) Time-varying shortest path problems with constraints. Networks 29:141–149

Chabini I (1998) Discrete dynamic shortest path problems in transportation applications: complexity and algorithms with optimal run time. Transp Res Rec 1645:170–175

Chin S, Paul DC (1989) Bicriterion routing scheme for nuclear spent fuel transportation. Transp Res Rec 1245:60–64
Cooke L, Halsey E (1966) The shortest route through a network with time-dependent internodal transit times. J Math Anal Appl 14:492–498
Dreyfus SE (1969) An appraisal of some shortest-path algorithms. Oper Res 17:395–412
Djikstra EW (1959) A note on two problems in connexion with graphs. Numer Math 1:269–271
Erkut E, Verter V (1998) Modeling of transport risk for hazardous materials. Oper Res 46(5): 625–642
Fabiano B, Curro F, Palazzi E, Pastorino R (2002) A framework for risk assessment and decision-making strategies in dangerous good transportation. J Hazard Mater 93:1–15
Hwang ST, Brown DF, O'Steen JK, Policastro AJ, Dunn W (2001) Risk assessment for national transportation of selected hazardous materials. Transp Res Rec 1763:114–124
Jagger S (1983) Development of CRUNCH: a dispersion model for continuous release of denser than air vapor into atmosphere. UK Atomic Energy Authority, London
Kim M, Miller-Hooks E, Nair R (2011) A geographical information system-based real-time decision support framework for routing vehicle carrying hazardous materials. J Intell Transp Syst 15(1):28–41
Karkazis J, Boffey TB (1995) Optimal location of routes for vehicles transporting hazardous materials. Eur J Oper Res 86:201–215
Leeming DG, Saccomanno FF (1994) Use of quantified risk assessment in evaluating the risks of transporting chlorine by road and rail. Transp Res Rec 1430:27–35
Miller-Hooks E, Mahmassani HS (1998) Optimal routing of hazardous materials in stochastic, time-varying transportation networks. Transp Res Rec 1645:143–151
Miller-Hooks E, Mahmassani HS (2000) Least expected time paths in stochastic, time-varying transportation networks. Transp Sci 34(2):198–215
Orda A, Rom R (1990) Shortest-path and minimum-delay algorithms in networks with time-dependent edge-lengths. J ACM 37:607–625
Pasquill F (1962) Atmospheric diffusion. D. Van Nostrand C., London
Patel MH, Horowitz AJ (1994) Optimal routing of hazardous materials considering risk of spill. Transp Res A 28A(2):119–132
ReVelle C, Cohon J, Shobrys D (1991) Simultaneous siting and routing in the disposal of hazardous wastes. Transp Sci 25(2):138–145
Saccomanno FF, Chan AY-W (1985) Economic evaluation of routing strategies for hazardous road shipments. Transp Res Rec 1020:12–18
Spicer T, Havens J (1989) User's guide for dense gas dispersion model. US Environmental Protection Agency, Research triangle park, NC
Strimaitis DG, Dicristofaro DC, Lavery TF (1986) The complex terrain dispersion model. EPA document EPA-450/2-78-027R, Research triangle park, North Carolina
Sutton OG (1947) The theoretical distribution of airborne pollution from factory chimneys. Q J Royal Meteorol Soc 73:426–436
Turner DB (1969) Workbook of atmospheric dispersion estimates. Public health service, 999-AP-26 (NTIS PB191842). Environmental Protection Agency, Research Triangle Park, NC
US Department of Transportation, Transport Canada, Secretariat of Transport and Communications of Mexico (2004) 2004 emergency response guidebook. J.J. Keller & Associates, Neenah, WI
Verma M, Verter V (2007) Railroad transportation of dangerous goods: population exposure to airborne toxins. Comput Oper Res 34:1287–1303
Wang X, Batta R, Karwan MH (2008) Flexible routing of a hazmat vehicle in a weather system. Presentation at informs annual meeting, Washington DC, 12–15 October 2008
Zhang J, Hodgson J, Erkut E (2000) Using GIS to assess the risks of hazardous materials transport in networks. Eur J Oper Res 121:316–329
Ziliaskopoulos AK, Mahmassani HS (1993) Time-dependent, shortest-path algorithm for real-time intelligent vehicle highway system applications. Transp Res Rec 1408:94–100

Value-at-Risk and Conditional Value-at-Risk Minimization for Hazardous Materials Routing

Iakovos Toumazis, Changhyun Kwon, and Rajan Batta

Introduction

Hazardous material (hazmat), as defined by the U.S. Department of Transportation Pipeline and Hazardous Materials Agency, is a substance or material capable of posing an unreasonable risk to health, safety, or property when transported in commerce[1]. There are various types of hazardous material transportation which range from movements of relatively harmless products, like hair spray and perfumes, to massive shipments of gasoline by highway cargo tanks, to transportation of poisonous, explosive, and radioactive materials. Accidents involving transporting hazmat are very rare. However when one does happen, the damages and consequences can be catastrophic in a means of both human casualties and residential environments.

During the year 2011, as shown in Table 1, there have been 13,908 hazmat incidents which resulted in 145 injuries, 10 deaths and damages of total worth $104,113,342. Note that the procedure of transporting these kinds of materials is divided in four phases: loading, in transit, in transit storage and unloading. The cost

[1] http://phmsa.dot.gov/hazmat/glossary

I. Toumazis
Department of Industrial and Systems Engineering, University at Buffalo, The State University of New York, Bell Hall 339A, Buffalo, NY 14260, USA
e-mail: iakovost@buffalo.edu

C. Kwon (✉)
Department of Industrial and Systems Engineering, University at Buffalo, The State University of New York, Bell Hall 318, Buffalo, NY 14260, USA
e-mail: chkwon@buffalo.edu

R. Batta
Department of Industrial and Systems Engineering, University at Buffalo, The State University of New York, Bell Hall 410, Buffalo, NY 14260, USA
e-mail: batta@buffalo.edu

R. Batta and C. Kwon (eds.), *Handbook of OR/MS Models in Hazardous Materials Transportation*, International Series in Operations Research & Management Science 193,
DOI 10.1007/978-1-4614-6794-6_5,

Table 1 2011 Hazmat summary by transportation phase (U.S. Department of Transportation Pipeline and Hazardous Materials Agency)

Transportation phase	Incidents	Injuries		Fatalities	Damages
		Hospitalized	Non-hospitalized		
Loading	2,633	0	22	1	$793,719
In transit	3,552	12	58	9	$84,687,976
In transit storage	530	1	6	0	$882,307
Unloading	7,193	10	36	0	$17,749,340
Grant total	13,908	23	122	10	$104,113,342

of the damages resulted by incidents during the transit phase had a total cost of $84,687,976 and the injuries during the same phase were 70, among which 12 needed hospitalization. From all ten deaths that occurred in 2011, nine of them took place in the process of transit and only one occurred during the other three phases, namely the loading phase.

On average, more than 800,000 shipments occur daily in the USA mostly using trucks as mean of transportation (Craft 2004) especially for relatively short distances. Using trucks for transportation is very popular since they are operationally flexible. In other words, trucks among other transportation modes, have the advantage of pick up and drop off hazmat materials very close from the origin and to the destination, respectively. However, the number of incidents involving the truck transport mode is the highest. In addition, trucks can be easily used for terrorist attacks against people and buildings (Federal Motor Carrier Safety Administration 2007, 2008; Murray-Tuite 2007; Huang et al. 2004). These statistics among with the possible threats emphasize the importance of efficient and effective regulative operation of urban traffic networks involving hazmat transportation.

Hazmat accidents rarely happen (low-probability incidents), but if they do occur, then the consequences can be disastrous (high-consequence incidents), reflecting on both the population and the environment. It is important to make a risk-averse route decision in hazmat transportation. In this book chapter, we will review the existing routing models and risk measures, and will introduce recent advancements with Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) models applied in hazmat transportation. VaR and CVaR are shown to be proper risk measures for flexible hazmat route decision making.

VaR and CVaR in Various Applications

Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) have been broadly used in finance as a risk measure. VaR has mainly the following five applications: risk management, risk measurement, financial control, financial reporting and computing regulatory capital. It was initially designed to measure the overnight risk in certain highly diversified portfolios. Furthermore, VaR is a risk measure that measures the financial risk of an investment, portfolio, or exposure over some

specified period of time. Its easy interpretation as a summary measure of risk and consistent treatment of risk across different financial instruments and business activities is the reason for its popularity. In addition, VaR is also used to approximate the "maximum reasonable loss" a company can expect to realize from its total financial exposures.

CVaR is a risk measure that is a computationally tractable and coherent alternative to VaR. The notion of CVaR is closely related to the notions of Expected Tail Loss, Tail Conditional Expectation, Tail VaR, Average VaR, Worst Conditional Expectation, or Expected Shortfall (Rockafellar and Uryasev 2000; Dowd and Blake 2006). Unlike VaR, CVaR provides an optimization framework with convexity. In financial portfolio optimization problems, CVaR optimization can be solved as a linear programming problem with sampling for continuous random variable cases, or as it is without sampling for discrete random variable cases (Rockafellar and Uryasev 2000; Mansini et al. 2007).

Besides its wide use in finance and banking, VaR and CVaR have been also used in other areas. They were used as risk measures in agricultural enterprises. Pruzzo et al. (2003) introduced the methodology of VaR and CVaR (as expected shortfall) in order to make animal breeding decisions. Their goal was to examine the use of VaR and CVaR as means to affiliate risk into breeding decisions. Manfredo and Leuthold (1999) used VaR to forecast market risk and cattle feeding margins, and Dah-Nein Tzang (1990) used VaR model to generate minimum risk hedge ratios simultaneously and applied the model to the soybean complex. The concepts of VaR and CVaR were also utilized to generate electricity in deregulated market (Robert Dahlgren and Lawarre 2003), product selection and plant dimensioning arena (Sodhi 2005). In the former manuscript VaR and CVaR concepts are used to address the problem of risk assessment in a market environment from the power industry prospective. In the later paper the VaR and CVaR models are used to address the problem of minimizing the risk rising from short product lifestyles and high uncertainty concerning the demand in the electronics industry.

In addition, the concepts of VaR and CVaR have been recently applied in hazmat transportation (Kang et al. 2011, 2013; Kwon 2011). However the application of these models is significantly different than the models used in finance. The most notable difference is that the models addressing hazmat transportation, are focused on measuring the risk resulted by following a specified route in the network. Therefore, the investment (which in this case is the route) and the loss measured (that is the accident consequence) are totally different, and of course they cannot be compared. Note that in finance the measurement units of both the investment and the loss are same as, say dollars. Another difference between the two models is that in hazmat transportation the risk of each road segment in a path is non-additive to each other, while losses of portfolios in financial models are additive. Furthermore even though the problem of minimizing CVaR is convex for financial optimization problems (Rockafellar and Uryasev 2000) that is not the case for the minimization of CVaR in hazmat transportation. In the next sections, we propose an efficient exact algorithm for solving the problem of CVaR minimization in the concept of hazardous materials routing. Given the above differences, the models of VaR and CVaR in hazmat transportation require application-specific analysis and computational methods.

Table 2 Mathematical notation

Notation	Description
$G(\mathscr{N}, \mathscr{A})$	A graph of road network
$\mathscr{N}$	Set of nodes, $\lvert\mathscr{N}\rvert = n$
$\mathscr{A}$	Set of arcs, $\lvert\mathscr{A}\rvert = m$
p_{ij}	Accident probability on arc (i, j)
c_{ij}	Accident consequence on arc (i, j)
$\mathscr{P}$	Set of all available paths for given O–D pair
$\mathscr{C}$	Set of ascending-order sorted arc consequences in G
$\mathscr{A}^l$	Arc set for path l, and $\lvert\mathscr{A}^l\rvert = \lvert l\rvert$ is the number of arcs in l
$\mathscr{C}^l$	Set of ascending-order sorted arc consequences for path l
R^l	Discrete random variable for the risk along path $l \in \mathscr{P}$

Risk Measures in Hazardous Materials Transportation

In this section, we summarize the existing approaches for hazmat routing and compare them with the proposed VaR and CVaR approaches. In hazmat transportation there are two main research areas: risk assessment (e.g. Abkowitz et al. 1984; Patel and Horowitz 1994) and hazmat shipments planning (e.g. List et al. 1991; Erkut et al. 2007). The latter field, which is studied in this chapter, involves the determination of the "best" possible route to transport the shipment from the origin to the destination (Nozick et al. 1997; Helander and Melachrinoudis 1997; Nembhard and White 1997). In hazmat transportation there are two groups of decision makers, namely the network regulators and the hazmat carriers; each one focusing on separate problems. The first problem that is studied mainly by the hazmat carriers is the local route planning problem that contemplates each shipment independently. That is, the determination of the safest path, between alternative paths, for a single shipment from its origin to the desirable destination. The second problem that is mostly of interest to the network regulators, i.e. government authorities and environmentalists is the global route planning. The goal is to plan multiple shipments when many origin and destination pairs exist in the understudy network, in order to moderate end regularize the risk to the whole network.

We consider a directed and weighted network $G = (\mathscr{N}, \mathscr{A})$. For each arc $(i, j) \in \mathscr{A}$, there are two attributes: accident probability denoted by p_{ij} and accident consequence denoted c_{ij}. Accident consequences c_{ij} can be computed by a risk assessment method, like the λ-neighborhood concept proposed by Batta and Chiu (1988), and accident probabilities p_{ij} are collected from certain data sources. The mathematical notation that is used in this chapter is summarized in Table 2. Suppose a path l consists of an *ordered* set of arcs $\mathscr{A}^l = \{(i_k, j_k) \in \mathscr{A} : k = 1, 2, \ldots, |l|\}$ where (i_k, j_k) is the kth arc in the path. To compute the risk generated by this path, a variety of models (Table 3) that use different risk measures can be utilized.

Table 3 Classic path risk evaluation models (Erkut and Ingolfsson 2005)

Model	Risk measure	Objective
TR	Expected risk	$\min_{l\in\mathscr{P}} \sum_{(i,j)\in\mathscr{A}^l} p_{ij}c_{ij}$
PE	Population exposure	$\min_{l\in\mathscr{P}} \sum_{(i,j)\in\mathscr{A}^l} c_{ij}$
IP	Incident probability	$\min_{l\in\mathscr{P}} \sum_{(i,j)\in\mathscr{A}^l} p_{ij}$
PR	Perceived risk	$\min\limits_{l\in\mathscr{P}} \sum\limits_{(i,j)\in\mathscr{A}^l} p_{ij}(c_{ij})^q$
MM	Maximum risk	$\min\limits_{l\in\mathscr{P}} \max\limits_{(i,j)\in\mathscr{A}^l} c_{ij}$
MV	Mean-variance	$\min\limits_{l\in\mathscr{P}} \sum\limits_{(i,j)\in\mathscr{A}^l} (p_{ij}c_{ij} + kp_{ij}(c_{ij})^2)$
DU	Disutility	$\min\limits_{l\in\mathscr{P}} \sum\limits_{(i,j)\in\mathscr{A}^l} p_{ij}(\exp(kc_{ij}-1))$
CR	Conditional probability	$\min_{l\in\mathscr{P}} \left(\sum\limits_{(i,j)\in\mathscr{A}^l} p_{ij}c_{ij} \Big/ \sum\limits_{(i,j)\in\mathscr{A}^l} p_{ij} \right)$

The Traditional Risk model (TR) computes the expected value of the consequence along path l (Sherali et al. 1997):

$$\mathbb{E}\left[R^l\right] = \sum_{(i_k,j_k)\in\mathscr{A}^l} \prod_{(i_h,j_h)\in\mathscr{A}^l, h<k} (1-p_{i_hj_h})p_{i_kj_k}c_{i_kj_k} \tag{1}$$

and manipulate it as a risk measure. In this model the shipment is terminated the moment an accident occurs on an arc (i, j). Using this objective, the choice of the best route can be formulated as a nonlinear binary integer program. However, the function (1) can be approximated as an additive function that leads to a tractable shortest-path problem formulation (Erkut and Verter 1998). According to the North American data on hazmat transportation accident statistics, the probabilities of an accident to take place are extremely small, usually in the range of 10^{-8} and 10^{-6} per mile traveled (Abkowitz and Cheng 1988).Therefore we can use the approximation

$$\prod_{(i_h,j_h)\in\mathscr{A}^l, h<k} (1-p_{i_hj_h}) \approx 1$$

for all k. It follows that, the function (1) can be approximated by the following function (Jin and Batta 1997):

$$\mathbb{E}\left[R^l\right] \approx \sum_{(i,j)\in\mathscr{A}^l} p_{ij}c_{ij} \tag{2}$$

which is much simpler to optimize, since the resultant problem is a shortest-path problem with the cost of traversing an arc (i, j) being equal to $p_{ij}c_{ij}$.

Both the Population Exposure (PE) model (ReVelle et al. 1991) and the Incident Probability (IP) model (Saccomanno and Chan 1985) can be viewed as two extreme variations of the Traditional Risk (TR) model. The former focuses on reducing the

accident probability, and the latter focuses on the total consequence on the impacted region. Another similar model is the Perceived Risk (PR) model (Abkowitz et al. 1992), which uses a perception factor q for consequence to provide a risk-averse decision (Abkowitz et al. 1992). However, the value of q is hard to understand and difficult to quantify. The Conditional Risk (CR) model (Sivakumar et al. 1993) evaluates the expected consequence assuming the first accident surely happens.

A big disadvantage of the Traditional Risk (TR) model is its risk-neutral attitude. In other words, TR model fails to capture the public posture against hazmat transportation. Motivated by this, the PR model has taken under consideration this defect and includes a weight parameter on consequences to reflect the public reaction on the risk. Erkut and Ingolfsson (2000) analyses three additional catastrophe-avoidance modeling methods. The first model is the Maximum Risk (MM) model. The objective of this model is to minimize the maximum consequence of the path in order to avoid significant damages and casualties. The MM model measures how far the tail of the consequence distribution extends. The second model, the Mean-Variance (MV) model, is popularly used to appraise the trade-offs between return and risk of an investment portfolio. The third model brings into play utility theory on hazmat transportation to formulate the risk problem, and develops a Disutility (DU) model in the form of $U(c) = \exp(\lambda c)$ where $\lambda > 0$. The DU model is risk-averse in the sense that the $(i + 1)$st life lost costs more than the ith life lost.

These models depend on existing statistics for one or two risk parameters, and often result in a unique optimal route, regardless the risk preference of the decision maker. In PR, MV, and DU models, the decision maker may change a parameter (q or k) to reflect own risk preference; however, it is unclear how to decide such parameters in those models. VaR and CVaR models are motivated by such a weakness of the previous models and provide a solution to it. They produce a more flexible and reliable route modeling approach for hazmat transportation. Depending on the decision makers' attitude to risk, one can make multiple planning decisions according to each individual risk preferences. Instead of a single optimal route output, these models have a two dimensional framework which produces alternative route choices given different confidence levels. In addition, while most existing hazmat routing methods study the entire risk distribution (e.g. expected value as in TR model), CVaR obviously focuses more on the long tail to avoid extreme events. In financial investment problems, studying only the long tail may not result in an optimal solution in some cases, because high risk can mean high return. However, in hazmat transportation, high-risk (catastrophic hazmat accidents) should not be traded with high-return. In conclusion, being risk-averse by focusing on the long tail is more reasonable for hazmat transportation.

Value-at-Risk Minimization Model in Transportation

At this point, we will present the VaR risk model for hazardous material transportation and analyze its properties. Suppose we have a network $G(\mathcal{N}, \mathcal{A})$, with a single origin–destination (O–D) pair for the transportation of the hazmat shipment. Let $\mathcal{P}$

denote the set of alternative routes l, which can be used for the shipment. Given a confidence level $\alpha \in (0, 1)$, and a path $l \in \mathscr{P}$ VaR is defined as follows:

$$\mathsf{VaR}_\alpha^l = \min\left\{\beta : P(R^l > \beta) \leq 1 - \alpha\right\} \tag{3}$$

In other words, VaR is defined to be the minimal level β such that the value of the risk measure R^l exceeding that level β has probability less than or equal to $1 - \alpha$. Hence, our objective, given the set of alternative paths $\mathscr{P}$, is to determine the path $l \in \mathscr{P}$ which has the minimum VaR. That is,

$$\mathsf{VaR}_\alpha^* = \min\{\mathsf{VaR}_\alpha^l : l \in \mathscr{P}\} \tag{4}$$

Each path l consists of a set of arcs $\mathscr{A}^l$ in ascending order. In addition, let $C_{(k)}^l$ denote the kth smallest value in the set $\{c_{ij} : (i, j) \in \mathscr{A}_l\}$ and $p_{(k)}^l$ be the corresponding arc accident probability. Then the risk measure R^l takes the following values:

$$R^l = \begin{cases} 0 & \text{with probability} 1 - \sum_{i=1}^{m_l} p_{(i)}^l \\ C_{(1)}^l & \text{with probability} p_{(1)}^l \\ \vdots & \vdots \\ C_{(m_l)}^l & \text{with probability} p_{(m_l)}^l \end{cases}$$

where $m_l = |\mathscr{A}_l|$ (the cardinality of $\mathscr{A}_l$).

Given R^l as above the cumulative distribution function (CDF) of R^l is:

$$F_{R^l}(\beta) = \Pr(R^l \leq \beta) = \begin{cases} 1 - \sum_{i=1}^{m_l} \pi_{(i)}^l, & \text{if } \beta \leq 0 \\ 1 - \sum_{i=2}^{m_l} \pi_{(i)}^l, & \text{if } 0 \leq \beta \leq C_{(1)}^l \\ \vdots & \\ 1 - \sum_{i=k+1}^{m_l} \pi_{(i)}^l, & \text{if } C_{(k-1)}^l \leq \beta \leq C_{(k)}^l \\ \vdots & \\ 1, & \text{if } C_{(m_l)}^l \leq \beta \end{cases}$$

where $\pi_{(k)}^l = P(R^l = C_{(k)}^l)$. Then using the above CDF and the fact that $P(R^l \leq \mathsf{VaR}^l) > \alpha$ we obtain

$$\mathsf{VaR}_{\alpha}^{l} = \begin{cases} 0, & \text{if} 0 \le \alpha \le 1 - \sum_{i=1}^{\bar{m}^l} \pi_{(i)}^{l} \\ C_{(1)}^{l}, & \text{if } 1 - \sum_{i=1}^{\bar{m}^l} \pi_{(i)}^{l} < \alpha \le 1 - \sum_{i=2}^{\bar{m}^l} \pi_{(i)}^{l} \\ \vdots & \\ C_{(k)}^{l}, & \text{if } 1 - \sum_{i=k}^{\bar{m}^l} \pi_{(i)}^{l} < \alpha \le 1 - \sum_{i=k+1}^{\bar{m}^l} \pi_{(i)}^{l} \\ \vdots & \\ C_{(\bar{m}^l)}^{l}, & \text{if } 1 - \pi_{(\bar{m}^l)}^{l} < \alpha \le 1 \end{cases} \tag{5}$$

To illustrate the distribution of the risk, we provide a simple example in the "An Illustrative Numerical Example" section.

Based on (5) we obtain a set of probability segments $(0, \alpha_1^l], (\alpha_1^l, \alpha_2^l], ..., (\alpha_k^l, \alpha_k^l + 1], ..., (\alpha_{\bar{m}^l}^l, 1)$, where $\alpha_k^l = 1 - \sum_{i=k}^{\bar{m}^l} \pi_{(i)}^l$. Next, we define $\beta_\alpha^l = VaR$. Therefore from (5) we obtain that $\beta_\alpha^l = C_{(k)}^l$ if and only if

$$\sum_{i=k+1}^{\bar{m}^l} \pi_{(i)}^{l} < 1 - \alpha \le \sum_{i=k}^{\bar{m}^l} \pi_{(i)}^{l} \tag{6}$$

Hence, using the definition of $\pi_{(k)}^l$ we obtain

$$\sum_{i=k+1}^{\bar{m}^l} \pi_{(i)}^{l} = \sum_{i=k+1}^{\bar{m}^l} P\left(R^l = C_{(i)}^{l}\right) = \sum_{(i,j)\in\mathscr{A}^l, c_{ij} > C_{(k)}^l} p_{ij} \tag{7}$$

$$\sum_{i=k}^{\bar{m}^l} \pi_{(i)}^{l} = \sum_{i=k}^{\bar{m}^l} P\left(R^l = C_{(i)}^{l}\right) = \sum_{(i,j)\in\mathscr{A}^l, c_{ij} \ge C_{(k)}^l} p_{ij} \tag{8}$$

Consequently we conclude that β^l is VaR for path l as in (3) if and only if the following conditions are met:

$$\sum_{(i,j)\in\mathscr{A}^l, c_{ij} > \beta_\alpha^l} p_{ij} \le 1 - \alpha \tag{9}$$

$$\sum_{(i,j)\in\mathscr{A}^l, c_{ij} \ge \beta_\alpha^l} p_{ij} > 1 - \alpha \tag{10}$$

Hence, the VaR minimization problem (4) is equivalent to the following problem

$$\min_{l} \beta_{\alpha}^{l} \tag{11}$$

subject to

$$\sum_{(i,j)\in\mathscr{A}^l, c_{ij}>\beta_{\alpha}^{l}} p_{ij} \leq 1-\alpha \qquad \forall i,j \in \mathscr{N}, \quad \forall l \in \mathscr{P} \tag{12}$$

$$\sum_{(i,j)\in\mathscr{A}^l, c_{ij}\geq\beta_{\alpha}^{l}} p_{ij} > 1-\alpha \qquad \forall i,j \in \mathscr{N}, \quad \forall l \in \mathscr{P} \tag{13}$$

Clearly, the optimal VaR value β_{α}^{*}, which is the optimal objective function value of the problem (11), is one of c_{ij} values. One strategy to solve the problem (11) is to examine all c_{ij} values in ascending order starting from the smallest value until the two conditions (12) and (13) are met. The two summations of p_{ij} in (12) and (13) clearly decrease with β_{α}^{l}. Therefore the second condition (13) will be easily satisfied for smaller β_{α}^{l} values, while the first condition (12) will not. To satisfy the first condition (12), we have to find a path l with the smallest $\sum_{(i,j)\in\mathscr{A}^l, c_{ij}>C_{(k)}^{l}} p_{ij}$ for the given β_{α}^{l}. As we examine c_{ij}, if we reach the point where both conditions are met for the first time, the corresponding c_{ij} is the optimal solution β_{α}^{*}.

We elaborate this idea mathematically. Consider the set $C = \{0, C_{(1)}, \ldots, C_{(\bar{m})} : C_{(1)} < C_{(2)} < \cdots < C_{(\bar{m})}\}$ which consists of all arc consequences values sorted in ascending order. Then we know that $\beta_{\alpha}^{*} \in C$. We reformulate the VaR minimization problem as a bi-level formulation. First, we modify the arc probabilities for a given $\beta_{\alpha} \in C$ as follows:

$$\bar{p}_{ij} = \begin{cases} p_{ij} & \text{if } c_{ij} > \beta_{\alpha} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

for all $(i,j) \in \mathscr{A}$. This modification of $\bar{p}_{ij}$ certifies that arcs with consequences smaller than β_{α} are considered in the route choice with greater importance than arcs with consequences greater than β_{α}. We let the binary variable x_{ij} be equal to 1 if arc (i,j) belongs to the route used for the hazmat shipment and 0 otherwise.

Then for a given confidence level α, we have the following bilevel problem equivalent to the VaR model:

$$\min_{\beta_{\alpha}} \mathscr{X}(\bar{\alpha}, f)\beta_{\alpha} \tag{15}$$

subject to

$$\beta_{\alpha} \in C \tag{16}$$

$$f = \min_{x\in\Omega} \sum_{(i,j)\in\mathscr{A}} \bar{p}_{ij} x_{ij} \tag{17}$$

where

$$\Omega = \Big\{x : \sum_j x_{ij} = 1 \text{for } i = O, \qquad \sum_j x_{ji} = 1 \text{for } i = D,$$

$$\sum_j x_{ji} - \sum_j x_{ij} = 0, \qquad \forall i \notin \{O, D\}, \quad x_{ij} \in \{0, 1\} \quad \forall i, j \in \mathcal{N}\Big\} \tag{18}$$

where

$$\mathcal{X}(\bar{\alpha}, f) = \begin{cases} 0 & \text{if } f \leq \bar{\alpha} \equiv 1 - \alpha \\ 1 & \text{otherwise} \end{cases} \tag{19}$$

The outer problem is solved when $\mathcal{X}(\bar{\alpha}, f)\beta_\alpha$ will become zero, i.e., $f \leq 1-\alpha$, and therefore the first condition (12) is met with the smallest such β_α. We note that the inner problem can be easily solved as it is an instance of shortest-path problems.

The proposed algorithm is as follows:

- *Step 1*: Sort all the arc consequences in ascending order, $C = \{0, C_{(1)}, ..., C_{(\bar{m})}\}$, and set $n \leftarrow 0$.
- *Step 2*: Let $\beta_\alpha^n = C_{(n)}$ and update $\bar{p}_{ij}$ values. Then solve the following problem:

$$f = \min_{x \in \Omega} \sum_{(i,j) \in \mathcal{A}} \bar{p}_{ij} x_{ij}$$

  in order to obtain the path l^n and the objective function value f^n, using an efficient shortest path algorithm like Dijkstra's Algorithm.
- *Step 3*: If $f^n \leq \bar{\alpha}$ or $n = \bar{m}$, stop; the current path l^n is the optimal VaR path for confidence level α.
- *Step 4*: If $f^n > \bar{\alpha}$ and $n < \bar{m}$, then set $n \leftarrow n + 1$ and go to Step 2.

Now, we provide the properties of the VaR model for the extreme values of the confidence level α. Proofs are found in Kang et al. (2013).

Theorem 1. *A scalar* $\alpha_{\min}$ *exists, such that* $\mathsf{VaR}^* = 0$ *for all* $\alpha \in (0, \alpha_{\min}]$.

Theorem 1 shows that in the case where α is very small, then the value of VaR, for all the paths, is zero. For this reason, the decision maker can make his decisions based on other criteria like shortest distance or minimum cost.

Theorem 2. *A scalar* $\alpha_{\max}^{\mathrm{MM}}$ *exists such that* $l^*_{(\alpha-\mathsf{VaR})} \equiv l^*_{\mathrm{MM}}$ *for all* $\alpha \in (\alpha_{\max}^{\mathrm{MM}}, 1)$, *where* l^*_{MM} *is the optimal route determined by the MM model.*

In other words, Theorem 2 tells that for a sufficiently large value of the confidence level α close to 1, the optimal route determined by the VaR model is the same as the one determined by the Maximum Risk model. This result reveals the extreme risk-averse attitude of the MM model.

Theorem 3. *For any interval* $(\alpha_{k^{\mathscr{P}}}, \alpha_{k^{\mathscr{P}}+1}] \supset (0,1)$*, the optimal route proposed by the VaR model remains the same for all the values of the confidence level in this interval.*

Theorem 3 declares that no matter the changes in the confidence level, there are always ranges of confidence intervals in which there is a unique optimal solution. In addition, from (5) we know that every candidate path in $\mathscr{P}$ is possible to have different risk value when the confidence level belongs into different confidence intervals marked

$$\left(1 - \sum_{i=k}^{\bar{m}^{\mathscr{P}}} \pi_{(i)}^{\mathscr{P}}, 1 - \sum_{i=k+1}^{\bar{m}^{\mathscr{P}}} \pi_{(i)}^{\mathscr{P}}\right]$$

Thus, for different confidence intervals, it is possible a different path to exist which minimizes the risk. Therefore, in a realistic network (a large network, with many available alternative paths) it is likely that a shipment would have different optimal solutions whenever the confidence level α changes.

In general, there are different optimal paths under various confidence levels $\alpha \in (0,1)$. We can therefore generate a set of optimal VaR paths for confidence levels of interests, and use the set as a basis for route choices. We will illustrate how we generate such a set in "An Illustrative Numerical Example" and "Case Study" sections.

Conditional Value-at-Risk Minimization Model in Transportation

Despite its wide usage, VaR has been criticized as a risk measure due to the fact that it is not a coherent risk measure (Artzner et al. 1999; Dowd and Blake 2006) and it might lead to inaccurate perception of risk (Nocera 2009; Einhorn 2008). It is claimed that VaR cuts off and ignores what is happening in the tail of the distribution as shown in Fig. 1. A similar argument is also valid under the scope of hazmat transportation. Due to the fact that accident probabilities are very small in each road segment with high consequences at the same time, using VaR as a risk measure may result the cut off of that particular road segment.

Both VaR and CVaR are quantile-based risk measures (Dowd and Blake 2006). Their main difference is the fact that CVaR accounts for the distribution in the long tail whereas VaR does not. CVaR is also a coherent risk measure in the sense of Artzner et al. (1999) for general loss distributions, including discrete distributions (Pflug 2000; Rockafellar and Uryasev 2002). These additional properties possessed by CVaR gives a motivation for the development of the CVaR risk model presented below.

We assume a directed and weighted graph $G(\mathscr{N}, \mathscr{A})$ and a single origin–destination pair. Also suppose that some estimates of hazmat accident probabilities

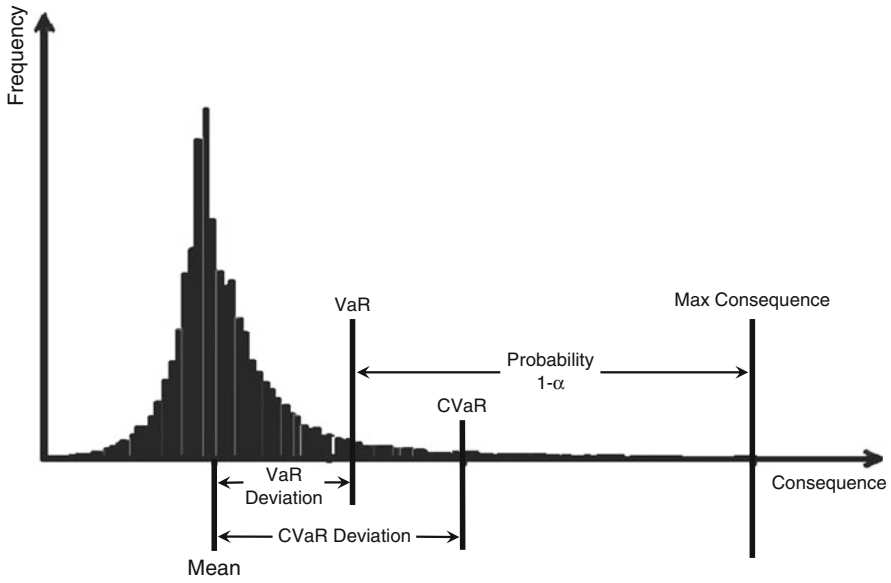


Fig. 1 VaR and CVaR deviations [*Source*: Sarykalin et al. (2008)]

and accident consequences, denoted by p_{ij} and c_{ij}, respectively, are available for each road segment (i, j). Our objective, given a confidence level α, is to choose a path l that minimizes CVaR as follows:

$$\min_{l \in \mathscr{P}} \mathsf{CVaR}_{\alpha}^{l} \tag{20}$$

where $\mathscr{P}$ is the set of all paths in the network. From the definition of CVaR, for a path $l \in \mathscr{P}$ at the confidence level α we have

$$\mathsf{CVaR}_{\alpha}^{l} = \frac{1}{1-\alpha} \int_{\alpha}^{1} \mathsf{VaR}_{\beta}^{l} \, \mathrm{d}\beta \tag{21}$$

which is in the form of the expected shortfall (Acerbi 2002). Since VaR_{β}^{l} and its integration are not available in an analytical form, CVaR in the form (21) is not used in an optimization format problem.

An alternative definition of CVaR for a path $l \in \mathscr{P}$ at the confidence level α is (Rockafellar and Uryasev 2002; Sarykalin et al. 2008):

$$\mathsf{CVaR}_{\alpha}^{l} = \lambda_{\alpha}^{l} \mathsf{VaR}_{\alpha}^{l} + (1 - \lambda_{\alpha}^{l}) \mathbb{E}[R^{l} : R^{l} > \mathsf{VaR}_{\alpha}^{l}] \tag{22}$$

where $\lambda_{\alpha}^{l} = \left(\Pr[R^{l} \leq \mathsf{VaR}_{\alpha}^{l}] - \alpha \right) \Big/ (1 - \alpha)$. While the second component, $\mathbb{E}[R^{l} : R^{l} > \mathsf{VaR}_{\alpha}^{l}]$, is incoherent (Acerbi 2002, 2004), the entire $\mathsf{CVaR}_{\alpha}^{l}$ in (22) is a

coherent risk measure (Rockafellar and Uryasev 2002; Pflug 2000). Note that the second component in the hazmat context can be approximated as follows:

$$\mathbb{E}[R^l : R^l > \mathsf{VaR}_\alpha^l] \approx \sum_{(i,j)\in\mathscr{A}^l : c_{ij} > \mathsf{VaR}_\alpha^l} p_{ij} c_{ij} \tag{23}$$

The measure CVaR_α^l given in the form (22) with an approximation (23) is hard to be considered as an objective function for the CVaR minimization problem due to the conditioning.

Rockafellar and Uryasev (2000) have shown that CVaR can be computed by minimizing the following function with respect to γ.

$$\Phi_\alpha^l(\gamma) = \gamma + \frac{1}{1-\alpha}\mathbb{E}[R^l - \gamma]^+ \tag{24}$$

$$\approx \gamma + \frac{1}{1-\alpha}\left\{\left(1 - \sum_{(i,j)\in\mathscr{A}^l} p_{ij}\right)[0-\gamma]^+ + \sum_{(i,j)\in\mathscr{A}^l} p_{ij}[c_{ij}-\gamma]^+\right\} \tag{25}$$

where $\mathscr{A}^l$ is the set of arcs in the path l and the notation $[x]^+ = \max(x, 0)$. We observe that a minimum of (25) always occur for $\gamma \geq 0$. We can show this by comparing the two cases of $\gamma = 0$ and $\gamma = -m$ where $m > 0$:

$$\Phi_\alpha^l(0) \approx \frac{1}{1-\alpha}\left\{\sum_{(i,j)\in\mathscr{A}^l} p_{ij} c_{ij}\right\}$$

$$\Phi_\alpha^l(-m) \approx -m + \frac{1}{1-\alpha}\left\{\left(1 - \sum_{(i,j)\in\mathscr{A}^l} p_{ij}\right)m + \sum_{(i,j)\in\mathscr{A}^l} p_{ij}(c_{ij}+m)\right\}$$

$$= -m + \frac{1}{1-\alpha}\left\{m + \sum_{(i,j)\in\mathscr{A}^l} p_{ij} c_{ij}\right\}$$

$$= \frac{\alpha m}{1-\alpha} + \frac{1}{1-\alpha}\left\{\sum_{(i,j)\in\mathscr{A}^l} p_{ij} c_{ij}\right\}$$

which indicates that $\Phi_\alpha^l(0) < \Phi_\alpha^l(-m)$ for all $m > 0$ and $\alpha \in (0, 1)$.

Therefore, we can write the CVaR minimization problem in the following form:

$$\min_{l\in\mathscr{P}} \mathsf{CVaR}_\alpha = \min_{l\in\mathscr{P},\gamma\geq 0}\left[\gamma + \frac{1}{1-\alpha}\sum_{(i,j)\in\mathscr{A}^l} p_{ij}[c_{ij}-\gamma]^+\right]$$

$$= \min_{x\in\Omega,\gamma\geq 0}\left[\gamma + \frac{1}{1-\alpha}\sum_{(i,j)\in\mathscr{A}} p_{ij}[c_{ij}-\gamma]^+ x_{ij}\right]$$

$$= \min_{\gamma\geq 0}\left[\gamma + \frac{1}{1-\alpha}\min_{x\in\Omega}\sum_{(i,j)\in\mathscr{A}} p_{ij}[c_{ij}-\gamma]^+ x_{ij}\right] \tag{26}$$

where Ω is defined in (18).

We can find an optimal solution of the CVaR minimization problem (26) by decomposing $\mathbb{R}^+$ into the following intervals :

$$[0, C_{(1)}],\ [C_{(1)}, C_{(2)}],\ \ldots,\ [C_{(\bar{m}-1)}, C_{(\bar{m})}] \text{ and } [C_{(\bar{m})}, \infty)$$

where $C_{(k)}$ is the kth smallest value in the set $\{c_{ij} : (i,j) \in \mathscr{A}\}$. Then defining $C_{(0)} = 0$, we obtain:

$$\sum_{(i,j)\in\mathscr{A}} p_{ij}[c_{ij}-\gamma]^+ x_{ij}$$

$$= \begin{cases} \displaystyle\sum_{(i,j)\in\mathscr{A}, c_{ij}>C_{(k)}} p_{ij}(c_{ij}-\gamma)x_{ij}, & \text{if } \gamma \in [C_{(k)}, C_{(k+1)}],\ k = 0,\ldots,\bar{m}-1 \\ 0, & \text{if } \gamma \in [C_{(\bar{m})}, \infty) \end{cases}$$

Therefore, $\mathsf{CVaR}^*_\alpha = \min_{k=0,\ldots,\bar{m}} \mathsf{CVaR}^k_\alpha$, where problem:

$$\mathsf{CVaR}^k_\alpha = \min_{\gamma\geq 0}\left[\gamma + \frac{1}{1-\alpha}\min_{x\in\Omega}\sum_{(i,j)\in\mathscr{A}, c_{ij}>C_{(k)}} p_{ij}(c_{ij}-\gamma)x_{ij}\right] \tag{27}$$

For each k, we are minimizing a linear function of γ in (27), over the interval $[C_{(k)}, C_{(k+1)}]$. Because of that, we know that the optimal γ value will be obtained either at $\gamma = C_{(k)}$ or at $\gamma = C_{(k+1)}$. Therefore, we can obtain an optimal solution γ^* by examining only the values in the set $\{0, C_{(1)}, C_{(2)}, \ldots, C_{(\bar{m})}\}$. That is:

$$\mathsf{CVaR}^*_\alpha = \min_{\gamma=0, C_{(1)},\ldots,C_{(\bar{m})}}\left[\gamma + \frac{1}{1-\alpha}\min_{x\in\Omega}\sum_{(i,j)\in\mathscr{A}} p_{ij}[c_{ij}-\gamma]^+ x_{ij}\right] \tag{28}$$

This analysis leads to the following algorithm for solving (28):

- *Step 1*: For $k = 0, 1, \ldots, \bar{m}$ solve the following nominal problems:

$$\mathsf{CVaR}^k_\alpha = C_{(k)} + \frac{1}{1-\alpha}\min_{x\in\Omega}\sum_{(i,j)\in\mathscr{A}, c_{ij}>C_{(k)}} p_{ij}(c_{ij}-C_{(k)})x_{ij}$$

- *Step 2*: Let $k^* = \arg \min_{k=0,1,\ldots,\bar{m}} \mathsf{CVaR}_\alpha^k$.
- *Step 3*: Then we obtain a solution: $\mathsf{CVaR}_\alpha^* = \mathsf{CVaR}_\alpha^{k^*}$ and $x^* = x^{k^*}$.

Next we provide some important properties of the CVaR model, similar to the ones given earlier for the VaR model. Proofs are found in Toumazis and Kwon (2012).

Theorem 4. *There exists a scalar $\alpha_{\min}$, such that $l^*_{\mathsf{CVaR}} = l^*_{\mathrm{TR}}$, $\forall \alpha \in (0, \alpha_{\min}]$, where l^*_{CVaR} and l^*_{TR} are the optimal paths determined by the CVaR model and the Traditional Risk models respectively.*

Theorem 5. *There exists a scalar $\alpha^{\mathrm{MM}}_{\max}$, such that $l^*_{\mathsf{CVaR}} = l^*_{\mathrm{MM}}$, $\forall \alpha \in (\alpha^{\mathrm{MM}}_{\max}, 1)$, where l^*_{CVaR} and l^*_{MM} are the optimal paths determined by the CVaR model and the Maximum Risk models respectively.*

These two theorems basically say that for sufficiently small α, the CVaR minimization is equivalent to the TR model, and for sufficiently large α, the CVaR minimization model is equivalent to the MM model. The property of the CVaR model for a small α value is an important improvement from the VaR model. In "Case Study" section we will observe that the optimal VaR values for all available paths are zero until the confidence level is as big as 0.999977 in a realistic hazmat network. In fact, the confidence level $\alpha = 0.999977$ may be regarded safe enough in many other situations, it is easy to misuse the VaR model, and in such cases the VaR model will give an arbitrary path chosen by the shortest-path problem solver. However, even for the same α value, the CVaR minimization model gives the optimal TR path; therefore it is at least a risk-neutral path.

An Illustrative Numerical Example

In this section, we demonstrate the computations of VaR and CVaR on a small example network shown in Fig. 2. For this network we defined node 1 as the origin and node 15 as the destination. The example network consists of 15 nodes and 33

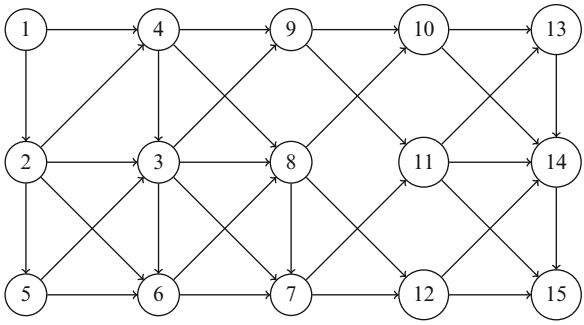


Fig. 2 A test network with 15 nodes and 33 arcs

Table 4 Cost coefficients used for the test network

(i, j)	p_{ij}	c_{ij}	(i, j)	p_{ij}	c_{ij}
(1, 4)	0.0007	7,670	(6, 8)	0.0005	5,248
(1, 2)	0.0002	7,800	(7, 11)	0.0063	5,670
(2, 4)	0.0009	920	(7, 12)	0.0007	6,576
(2, 3)	0.0004	7,177	(8, 7)	0.0021	8,754
(2, 5)	0.0008	4,112	(8, 10)	0.0003	5,329
(2, 6)	0.0004	9,894	(8, 12)	0.0005	7,656
(3, 9)	0.0008	8,553	(9, 10)	0.0009	5,714
(3, 8)	0.0008	7,474	(9, 11)	0.0004	3,210
(3, 6)	0.0004	1,534	(10, 13)	0.0025	1,452
(3, 7)	0.0006	960	(10, 14)	0.0003	5,606
(4, 3)	0.0003	4,542	(11, 14)	0.0001	9,220
(4, 9)	0.0001	4,540	(11, 15)	0.0009	5,202
(4, 8)	0.0008	3,724	(11, 13)	0.001	2,427
(5, 3)	0.0001	7,162	(12, 15)	0.0006	482
(5, 6)	0.0002	3,772	(12, 14)	0.0007	4,643
(6, 7)	0.0012	2,460	(13, 14)	0.0028	4,142
			(14, 15)	0.0059	1,615

Table 5 Optimal paths given by the VaR model for various confidence levels α

Confidence level α	Optimal VaR route	Optimal VaR value
[0, 0.997899]	$1 \to 4 \to 9 \to 11 \to 15$	0
[0.99790, 0.99840]	$1 \to 2 \to 6 \to 8 \to 12 \to 15$	482
[0.99841, 0.99849]	$1 \to 2 \to 4 \to 8 \to 12 \to 15$	920
[0.99850, 0.99879]	$1 \to 2 \to 4 \to 3 \to 7 \to 12 \to 15$	960
[0.99880, 0.99920]	$1 \to 2 \to 4 \to 9 \to 11 \to 14 \to 15$	1,615
[0.99921, 0.99959]	$1 \to 2 \to 4 \to 9 \to 11 \to 14 \to 15$	3,210
[0.99960, 0.99969]	$1 \to 2 \to 4 \to 9 \to 11 \to 13 \to 14 \to 15$	4,142
[0.99970, 0.99979]	$1 \to 2 \to 4 \to 9 \to 11 \to 13 \to 14 \to 15$	4,540
[0.99980, ≈ 1]	$1 \to 4 \to 3 \to 7 \to 11 \to 15$	7,670

links. Accident probabilities and accident consequences are randomly generated and are shown in Table 4.

We compare the optimal routes given by the VaR model with the optimal routes by the CVaR model. The optimal paths for various confidence levels by the VaR and CVaR models are given in Tables 5 and 6, respectively. Please note that we did not provide CVaR values in Table 6, while we provided VaR values in Table 5. That is because the value of the CVaR measure constantly changes within any interval of α. However, one can see the optimal CVaR-values from Fig. 3.

To understand VaR measure better, we give a closer look on how the risk is distributed in a path at a specific confidence level, for example, $\alpha = 0.99930$. At this confidence level the optimal path proposed by the VaR model is:

$$1 \to 2 \to 4 \to 9 \to 11 \to 14 \to 15$$

Table 6 Optimal paths given by the CVaR model for various confidence levels α

Confidence level α	Optimal CVaR route
$[0, 0.996359]$	$1 \to 2 \to 4 \to 9 \to 11 \to 15$
$[0.99636, 0.999146]$	$1 \to 2 \to 4 \to 9 \to 11 \to 14 \to 15$
$[0.999147, 0.99979]$	$1 \to 2 \to 4 \to 9 \to 11 \to 13 \to 14 \to 15$
$[0.99980, \approx 1]$	$1 \to 4 \to 3 \to 7 \to 11 \to 15$

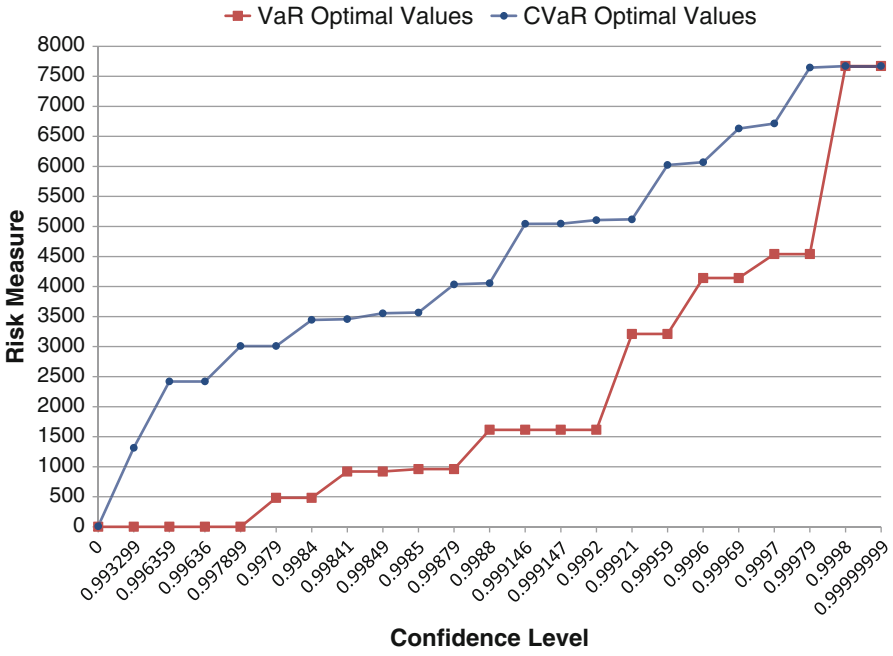


Fig. 3 VaR and CVaR optimal values for various confidence levels

Given this path, one can find the accident probabilities and consequences of the arcs that are included in the path. Namely, the path consists of the arcs $(1, 2)$, $(2, 4)$, $(4, 9)$, $(9, 11)$, $(11, 14)$, $(14, 15)$, which have respectively the following accident probabilities and consequences:

(i, j)	p_{ij}	c_{ij}
(1, 2)	0.0002	7,800
(2, 4)	0.0009	920
(4, 9)	0.0001	4,540
(9, 11)	0.0004	3,210
(11, 14)	0.0001	9,220
(14, 15)	0.0059	1,615

The risk from traversing the path has the following distribution:

$$R = \begin{cases} 0 & \text{with probability } 0.9896 \\ 920 & \text{with probability } 0.0009 \\ 1,615 & \text{with probability } 0.0059 \\ 3,210 & \text{with probability } 0.0004 \\ 4,142 & \text{with probability } 0.0028 \\ 4,540 & \text{with probability } 0.0001 \\ 7,800 & \text{with probability } 0.0002 \\ 9,220 & \text{with probability } 0.0001 \end{cases} \tag{29}$$

Given R as above, we then have the following cumulative distribution function (CDF) of R:

$$F_R = P(R \leq \beta) = \begin{cases} 0.9896 & \text{if } \beta \leq 0 \\ 0.9905 & \text{if } 0 < \beta \leq 920 \\ 0.9964 & \text{if } 920 < \beta \leq 1,615 \\ 0.9968 & \text{if } 1615 < \beta \leq 3,210 \\ 0.9996 & \text{if } 3210 < \beta \leq 4,142 \\ 0.9997 & \text{if } 4142 < \beta \leq 4,540 \\ 0.9999 & \text{if } 4540 < \beta \leq 7,800 \\ 1 & \text{if } 7800 < \beta \leq 9,220 \end{cases} \tag{30}$$

The value of VaR for this path can be computed as follows:

$$\begin{aligned} \mathsf{VaR} &= \min\{\beta : P(R > \beta) \leq 1 - \alpha\} \\ &= \min\{\beta : P(R > \beta) \leq 1 - 0.99930\} \\ &= \min\{\beta : P(R > \beta) \leq 0.0007\} \end{aligned}$$

Therefore, from (30), we can find the value of the minimum β such that $P(R > \beta) \leq 0.0007$. We can write

$$P(R > \beta) \leq 0.0007 \;\Rightarrow\; 1 - P(R \leq \beta) \leq 0.0007 \;\Rightarrow\; P(R \leq \beta) \leq 0.9993$$

Hence we obtain $\mathsf{VaR} = 3,210$.

Note that for $\alpha > 0.9998$, the value of VaR from traversing the above path should be 7,800, which is the maximum accident consequence (i.e. $\max c_{ij}$) in the path. However, for confidence levels greater than 0.9998, say 0.9999, we can see that by following a different route, namely $1 \to 4 \to 3 \to 7 \to 11 \to 15$, the value of

$\mathsf{VaR} = 7,670$. Therefore, since the VaR value for the latter path is smaller than the value of the risk measure resulting by the former path, the algorithm alters the proposed path to the one with the minimum VaR value.

Extending this idea, we can conclude the following: Since the smallest accident probability in the data set we used, is 0.0001 (see Table 4) for confidence levels greater than 0.9999 it is guaranteed that the value of VaR for the proposed route would be equal to the maximum accident consequences in the set of arcs forming that proposed optimal path. In other words, at confidence levels beyond that value, the VaR model is equivalent to the Maximum Risk (MM) model; see Theorem 2.

Similarly, one can make the same arguments for the CVaR model following the same exact procedure as demonstrated above for the VaR model.

For the remainder of this section, we would like to emphasize a number of interesting observations by studying Tables 5 and 6. First, we observe what the two models have in common:

- In both models, each optimal route maintains its optimality within a specific interval of confidence level values.
- It is clear that the proposed models generate various optimal routes.
- For confidence level values very close to 1, both models propose the same path as the Maximum Risk Model.

Now, we list what differences the two models have:

- The optimal paths proposed by the VaR and CVaR models are not necessarily the same for each confidence level. There are cases that the optimal routes given by the two models are the same for the same α-value, but this is not always the case.
- The optimal value of the risk measures for the same confidence level is not always the same for the two models. Specifically, as stated earlier, when VaR is the risk measure, then its optimal value remains constant for an interval of confidence levels. On the other hand, CVaR continuously alters its value depending on the value of the confidence level.
- The number of alternative optimal paths given by the VaR model is greater than the one from the CVaR model. This is because the VaR model covers from risk-indifferent to risk-averse paths, while CVaR model only covers from risk-neutral to risk-averse paths.

All the above observations are detailed discussed in the following section, which describes a case study we conducted on the Albany road network. In this way, the reader may better understand the importance of the above findings when they are applied on existing networks.

Case Study

In order to develop our model and obtain numerical results, we utilized the Albany, New York, USA transportation network. The Albany network used in this chapter consists of 90 nodes and 148 arcs and is demonstrated in Fig. 4.

We computed the nominal accident probabilities using the formula $p_{ij} = 10^{-6} \times$ (length of edge (i, j)) as in Abkowitz and Cheng (1988). We also used the λ-neighborhood concept developed by Batta and Chiu (1988) to compute the nominal accident consequences for every arc in the network. Both the road lengths and population statistics were obtained from the Department of Transportation and Department of Commerce Websites.

Computations were performed by using Matlab2010a and ran on a 2.8 GHz Intel Core 2 Duo computer system. We used the O–D pair $(1, 12)$ to illustrate the VaR and CVaR models. Both models were tested for the same data sets, under various confidence levels in order to authenticate their effectiveness. The computation times for VaR and CVaR models were less than 2 s.

Next we present our findings from the development of the proposed VaR model. In Table 7 we provide, for various confidence levels, the corresponding optimal paths and the value of the VaR, as a risk measure resulted from traversing the proposed route.

From the resulting VaR optimal routes, we highlight our findings. First, we can see that for confidence levels in the interval $[0, 0.999977]$ the VaR model is not effective since the route that recommends is risk-indifferent. That is, the VaR measure has zero value for all paths. Therefore the optimal path presented in the table may not be safe at all. In fact, it is simply an arbitrary path chosen by the algorithm among all available paths. In this case the decision maker should determine the shipment route based on other criteria. This result also testifies the validity of Theorem 1.

We wish to emphasize here the reason for which the optimal route given by the model for $\alpha \in [0, 0.999977]$ remains unchanged while for confidence levels greater than 0.999977 the suggested path regularly changes as Table 7 indicates. That occurs because the accident probabilities p_{ij} for every link in the network, are very small. As a result the confidence level value mentioned above is considered to be small and consequently the value of $1-\alpha$ (see Fig. 1) is quite big. That means that important data are not captured from the model at confidence levels in the interval $[0, 0.999977]$.

Secondly, as Theorem 3 indicates, we can also observe in Table 7 that every path proposed by the VaR model preserve its optimality for a certain confidence level interval. More importantly, it is obvious that the proposed model generates various routes depending on the value of the confidence level. Consequently, the proposed VaR model gives to the decisions makers the opportunity to alter their decisions depending on their confidence providing a flexible framework.

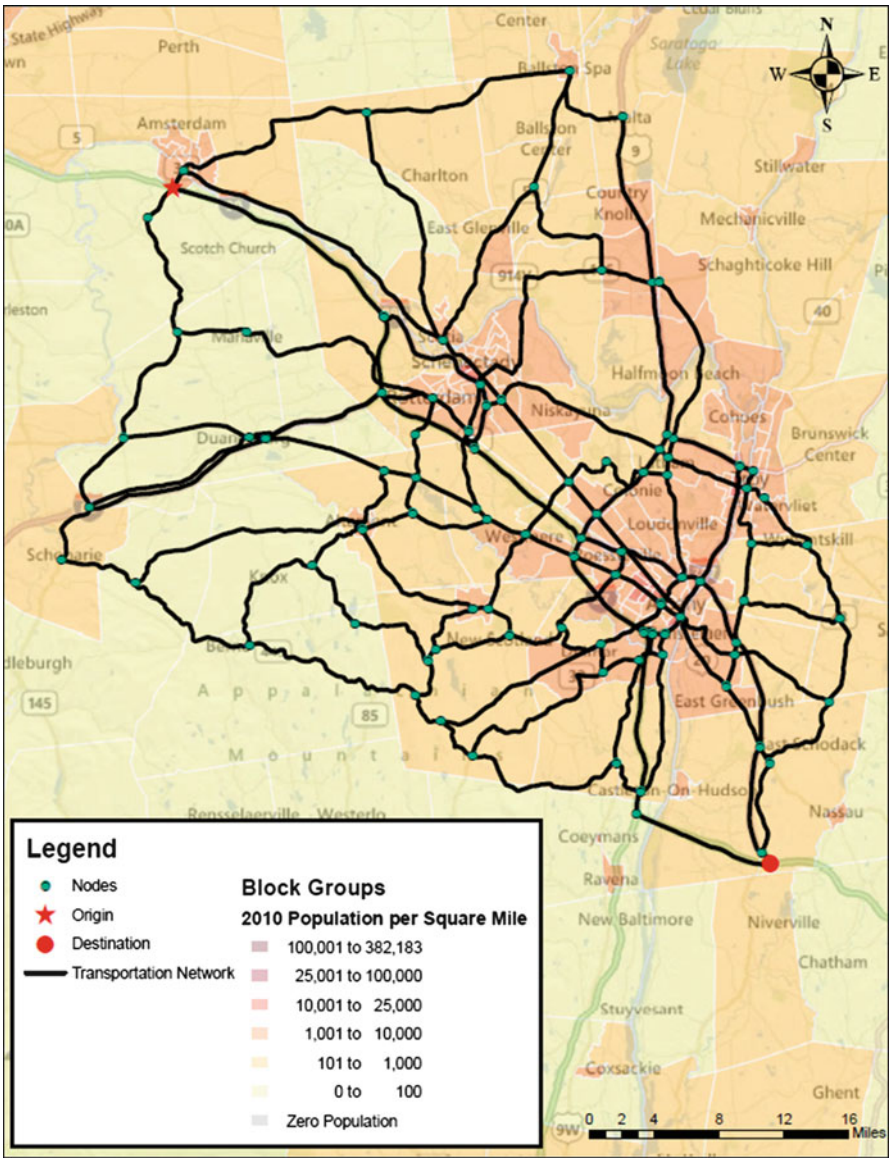


Fig. 4 Albany area highway network

The optimal paths recommended by the CVaR model are shown in Table 8. As one can easily see, the optimal routes resulted by the CVaR model for confidence levels in the interval $[0, 0.999991]$ are the same for the same reason as discussed earlier for the VaR model. Specifically for these range of values of the confidence level, the resulting optimal routes are the same as the one resulted by

Table 7 Optimal paths given by the VaR model for various confidence levels α

Confidence level α	Optimal VaR route	Optimal VaR value
0	1,74,78,42,82,27,20,21,10,11,12	0
0.75	1,74,78,42,82,27,20,21,10,11,12	0
0.999977	1,74,78,42,82,27,20,21,10,11,12	0
0.999978	1,70,45,71,58,59,4,5,17,18,19,20,21,10,11,12	824.10
0.999979	1,70,45,71,58,59,4,5,17,18,19,20,21,10,11,12	837.14
0.999980	1,70,45,71,58,59,4,5,17,18,19,20,21,10,11,12	837.14
0.999981	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	1,047.71
0.999982	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	1,047.71
0.999983	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	1,143.96
0.999986	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	1,143.96
0.999987	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	1,212.47
0.999988	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	1,212.47
0.999989	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	1,301.19
0.999991	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	1,301.19
0.999992	1,70,45,13,81,72,73,63,64,65,54,66,67,68,41,29,30,12	3,312.53
0.999994	1,70,45,13,81,72,73,63,64,65,54,66,67,68,41,29,30,12	3,312.53
0.999995	1,70,45,13,14,15,55,56,62,63,64,65,54,66,67,68,41,29,30,12	4,908.01
0.999998	1,70,45,13,14,15,55,56,62,63,64,65,54,66,67,68,41,29,30,12	4,908.01
0.999999	1,70,45,13,14,15,55,56,62,63,64,65,54,66,67,68,41,29,30,12	5,062.25

Table 8 Optimal paths given by the CVaR model for various confidence levels α

Confidence level α	Optimal CVaR route	Optimal CVaR value
0	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	0.058961
0.75	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	0.11792
0.999977	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	2,279.26598
0.999978	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	2,342.66171
0.999979	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	2,408.6241
0.999980	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	2,480.4196
0.999981	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	2,556.3757
0.999982	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	2,640.1904
0.999983	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	2,730.4686
0.999986	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	3,070.4346
0.999987	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	3,218.0979
0.999988	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	3,385.2331
0.999989	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	3,575.9012
0.999991	1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	4,081.3924
0.999992	1,70,45,13,81,72,73,63,64,65,54,66,67,68,41,29,30,12	4,339.9188
0.999994	1,70,45,13,81,72,73,63,64,65,54,66,67,68,41,29,30,12	4,682.3820
0.999995	1,70,45,13,14,15,55,56,62,63,64,65,54,66,67,68,41,29,30,12	4,940.3977
0.999998	1,70,45,13,14,15,55,56,62,63,64,65,54,66,67,68,41,29,30,12	4,988.9863
0.999999	1,70,45,13,14,15,55,56,62,63,64,65,54,66,67,68,41,29,30,12	5,062.2545

Table 9 Traditional Risk model

Optimal route	$E[R]$
1,70,45,13,81,72,73,69,66,67,68,41,29,30,12	0.058961

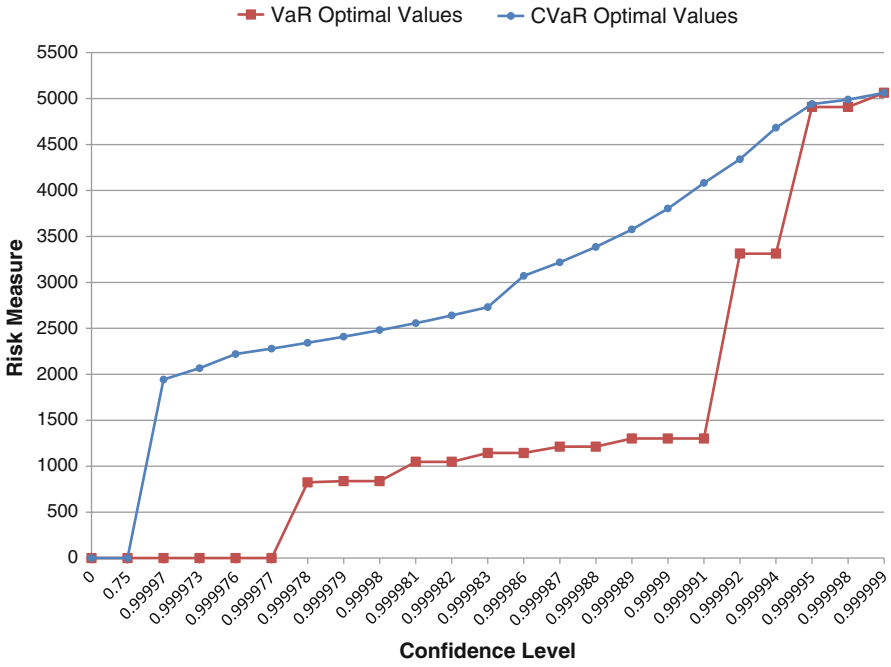


Fig. 5 Optimal VaR and CVaR values for various confidence levels

the Traditional Risk model as shown in Table 9. These observations numerically confirms Theorem 4.

We emphasize here the advantage CVaR has towards VaR, based on how the two risk measures exploit the long tail of the distribution. Even for relatively smaller confidence levels values, CVaR model propose as optimal path the same route as VaR propose for confidence levels greater than 0.999978. CVaR alters its proposed route less times than the VaR, not because of some defect, but because of a better starting point given to the model by the better account of the distribution's long tail.

Obviously, CVaR and VaR optimal routes for the same confidence level are not always the same. For example, for $\alpha = 0.999980$ we can see that the optimal routes provided by the two models differ from each other. Of course, the proposed optimal path for confidence levels very close to 1 is same in both models with the optimal VaR and CVaR values being equal.

Figure 5 shows how the optimal values of the two risk measures increase as the value of confidence level is approaching 1.

The remainder of this section focuses on the comparison of the two models under specific confidence levels. We proceed to compare two cases. First let us take a

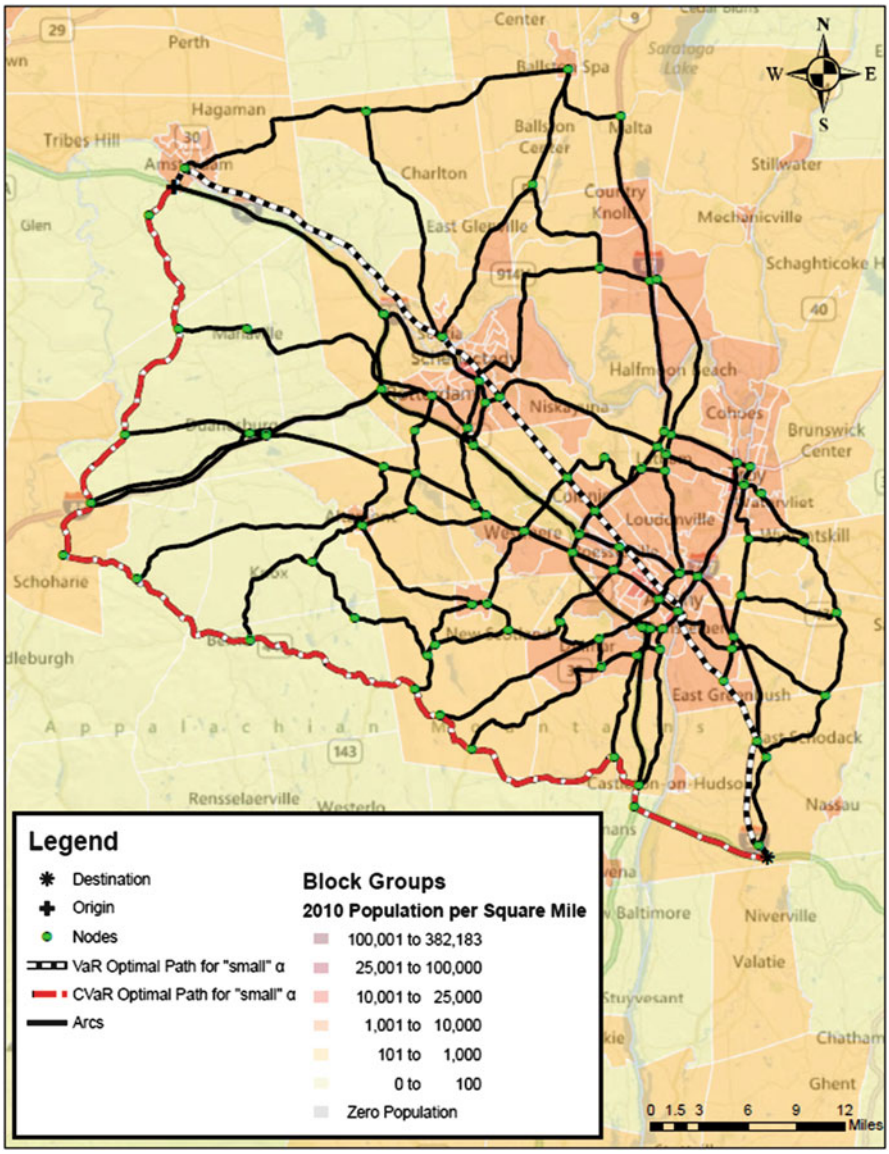


Fig. 6 Graphical representation of optimal paths when $\alpha = 0.999977$

small confidence level $\alpha = 0.999977$. This value of α is considered to be small since we compare it with the average hazmat accident probability level which, as mentioned before, is 10^{-6}. Figure 6 shows graphically the optimal paths for each model. Comparing the optimal routes, we see that the VaR model proposes an optimal path that starts along I-90 but immediately continues on NY State Route

30 heading north, and then follows NY State Route 5 all the way through Albany City where it reaches U.S. Route 20. Then it continues on U.S. Route 20 until it crosses I-90 on which is finished. We note here that the route given by the VaR model passes directly through the high population density areas of Albany City, Schenectady and Scotia.

In contrast, the route proposed by the CVaR model seems to be more reasonable with less cutoff risks compared to the one proposed by the VaR model. Note that this route is the same as the one given by the Traditional Risk model. It starts along I-90 and immediately detours and continuous along NY State Route 30 as well only this time heading south until it reaches NY State Route 443 on which it continues. Then it follows NY State Route 396 and then it detours on I-87 for a short period. Finally it finishes on I-90 crossing the NY State Thruway Berkshire Connector. With a more detailed study of the proposed route, we can see that CVaR's optimal route also passes through the populated area of Esperance. However a possible hazmat accident at that area will not have the same consequences as one in the center of Albany City from which the optimal path proposed by the VaR model passes through. At the same time, it avoids I-90 until the end, at which point there is no other choice but to take I-90 in order to reach the destination. Hence, we can conclude that the CVaR model provides a more risk-averse framework compared to the VaR model. In other words, CVaR model handles the risk in the long tail in a better way.

Similarly we provide the resulting optimal paths for $\alpha = 0.999999$ in Fig. 7. We observe that for confidence level values greater than 0.999981 both models suggest the same path for respectively same confidence levels (it is obvious from the comparison of Tables 7 and 8). This is something that we expected, since the value of $1 - \alpha$ is getting closest to zero (as the α value increases) and therefore the two models captures the same information from the tail of the distribution (as shown in Fig. 1).

Also note, that for confidence level approximately equal to 100 % both VaR and CVaR proposes the same route. In addition the values of their risk measures are almost equal as shown in Tables 7 and 8.

Extensions and Conclusions

This chapter demonstrates some new ideas that can be applied in hazardous materials transportation which are based on the risk measurements Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). The objective of the proposed models was to minimize the risk experienced by the transportation of a hazmat shipment along the Albany, NY road network under certain confidence level. Even though the definition of VaR model is easier to understand, we showed that the proposed CVaR model is easier to optimize and it captures the risk in the long tail in a better way. Both models provide a flexible framework to the decision maker resulting in different optimal route choices for different confidence intervals.

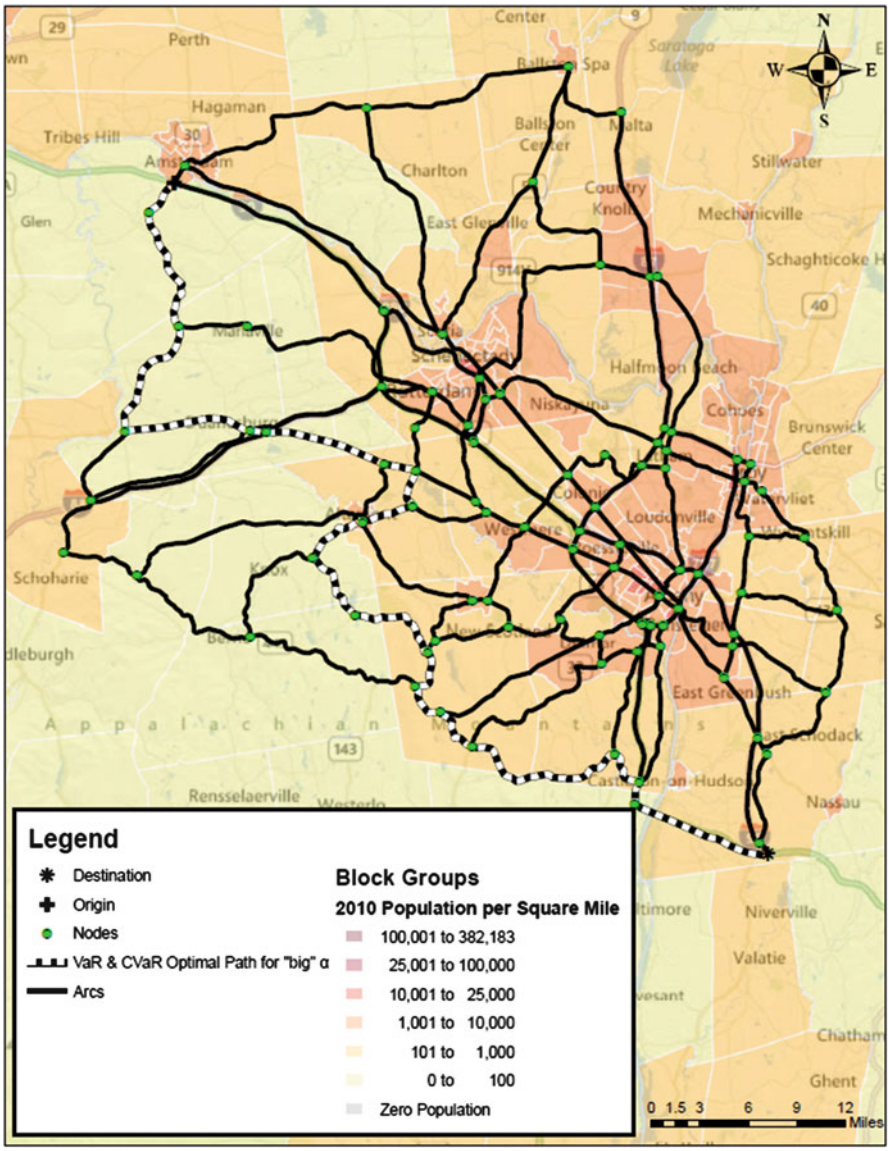


Fig. 7 Graphical representation of optimal paths when $\alpha = 0.999999$

References

Abkowitz M, Cheng P (1988) Developing a risk/cost framework for routing truck movements of hazardous materials. Accid Anal Prev 20(1):39

Abkowitz M, Eiger A, Srinivasan S (1984) Estimating the release rates and costs of transporting hazardous waste. Transport Res Rec 977:22–30

Abkowitz M, Lepofsky M, Cheng P (1992) Selecting criteria for designating hazardous materials highway routes. Transport Res Rec 1333:30–35

Acerbi C (2002) Spectral measures of risk: a coherent representation of subjective risk aversion. J Bank Finance 26:1505–1518

Acerbi C (2004) Coherent representations of subjective risk-aversion. In: Szeö G (ed) Risk measures for the 21st century. Wiley, New York, pp 147–207

Artzner P, Delbaen F, Eber J, Heath D (1999) Coherent measures of risk. Math Finance 9(3): 203–228

Batta R, Chiu S (1988) Optimal obnoxious paths on a network: transportation of hazardous materials. Oper Res 36(1):84–92. http://www.jstor.org/stable/171380

Craft R (2004) Crashes involving trucks carrying hazardous materials. Federal Motor Carrier Safety Administration. http://www.fmcsa.dot.gov/facts-research/research-technology/analysis/fmcsa-ri-04-024.htm

Dah-Nein Tzang RML (1990) Hedge ratios under inherent risk reduction in commodity complex. J Futures Markets 10:497–504

Dowd K, Blake D (2006) After var: the theory, estimation, and insurance applications of quantile-based risk measures. J Risk Insur 73(2):193–229

Einhorn D (2008) Private profits and socialized risk. Global Association of Risk Professionals Risk Review, June/July, No. 42, pp 10–18

Erkut E, Ingolfsson A (2000) Catastrophe avoidance models for hazardous materials route planning. Transport Sci 34(2):165

Erkut E, Ingolfsson A (2005) Transport risk models for hazardous materials: revisited. Oper Res Lett 33(1):81–89

Erkut E, Verter V (1998) Modeling of transport risk for hazardous materials. Oper Res 46(5): 625–642

Erkut E, Tjandra S, Verter V (2007) Hazardous materials transportation. In: Barnhart C, Laporte G (eds) Transportation, Handbooks in operations research & management science, vol 14, chap 9, North Holland, pp 539–611. http://www.sciencedirect.com/science/article/pii/S0927050706140098

Federal Motor Carrier Safety Administration (2007) Final report: designation of highway routes for hazardous materials shipments: literature review. U.S. Department of Transportation, prepared by Battelle

Federal Motor Carrier Safety Administration (2008) Guidance document: hazardous materials routing using safety and security criteria. U.S. Department of Transportation, prepared by Battelle

Helander M, Melachrinoudis E (1997) Facility location and reliable route planning in hazardous material transportation. Transport Sci 31(3):216–226

Huang B, Cheu R, Liew Y (2004) GIS and genetic algorithms for HAZMAT route planning with security considerations. Int J Geogr Inform Sci 18(8):769–787

Jin H, Batta R (1997) Objectives derived from viewing hazmat shipments as a sequence of independent Bernoulli trials. Transport Sci 31(3):252–261

Kang Y, Batta R, Kwon C (2011) Generalized route planning model for hazardous material transportation with var and equity considerations. Working paper

Kang Y, Batta R, Kwon C. Value-at-risk model for hazardous material transportation. Ann Oper Res. doi: 10.1007/s10479-012-1285-0

Kwon C (2011) Conditional value-at-risk model for hazardous materials transportation. In: Jain S, Creasey RR, Himmelspach J, White KP, Fu M (eds) Proceedings of the 2011 winter simulation conference. pp 1708–1714

Pruzzo L, Cantet RJ, Fioretti CC (2003) Risk-adjusted expected return for selection decisions. J Anim Sci 81:2984–2988

List G, Mirchandani P, Turnquist M, Zografos K (1991) Modeling and analysis for hazardous materials transportation: risk analysis, routing/scheduling and facility location. Transport Sci 25(2):100–114

Manfredo MR, Leuthold RM (1999) Market risk and the cattle feeding margin: an application of value-at-Risk. Agribusiness 17(3):333–353

Mansini R, Ogryczak W, Speranza M (2007) Conditional value at risk and related linear programming models for portfolio optimization. Ann Oper Res 152(1):227–256

Nune R, Murray-Tuite P (2007) Comparison of potential paths selected by a malicious entity with hazardous materials: minimization of time vs. minimization of distance. Proceedings of the 2007 winter simulation conference. Washington, DC. http://dl.acm.org/citation.cfm?doid=1351542.1351745

Nembhard D, White C III (1997) Applications of non-order-preserving path selection of hazmat routing. Transport Sci 31(3):262–271

Nocera J (2009) Risk mismanagement. The New York Times Magazine. http://www.nytimes.com/2009/01/04/magazine/04risk-t.html. Accessed 4 Jan 2009

Nozick L, List G, Turnquist M (1997) Integrated routing and scheduling in hazardous materials transportation. Transport Sci 31(3):200–215

Patel M, Horowitz A (1994) Optimal routing of hazardous materials considering risk of spill. Transport Res A 28(2):119–132

Pflug G (2000) Probabilistic constrained optimization: methodology and applications. Kluwer Academic Publishers, New York

ReVelle C, Cohon J, Shobrys D (1991) Simultaneous siting and routing in the disposal of hazardous wastes. Transport Sci 25(2):138

Robert Dahlgren CCL, Lawarre J (2003) Risk assessment in energy trading. IEEE Trans Power Syst 18(2):503–511

Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. J Risk 2(3):21–42. http://www.thejournalofrisk.com/public/showPage.html?page=1491

Rockafellar R, Uryasev S (2002) Conditional value-at-risk for general loss distributions. J Bank Finance 26(7):1443–1471

Saccomanno F, Chan A (1985) Economic evaluation of routing strategies for hazardous road shipments. Transport Res Rec 1020:12–18

Sarykalin S, Serraino G, Uryasev S (2008) Value-at-risk vs. conditional value-at-risk in risk management and optimization. Tutorials Oper Res 270-294. https://www.informs.org/content/download/132735/1131225/file/Tutorials-2008-CD-titlepage.pdf

Sherali H, Brizendine L, Glickman T, Subramanian S (1997) Low probability–high consequence considerations in routing hazardous material shipments. Transport Sci 31(3):237–251

Sivakumar RA, Rajan B, Karwan M (1993) A network-based model for transporting extremely hazardous materials. Oper Res Lett 13(2):85–93

Sodhi MS (2005) Managing demand risk in tactical supply chain planning for a global consumer electronics company. Prod Oper Manag 14:69–79

Toumazis I, Kwon C (2012) Routing hazardous materials on time-dependent networks using conditional value-at-risk. Working paper

Hazardous Facility Location Models on Networks

Marcos Colebrook and Joaquín Sicilia

Introduction

This chapter aims to be a comprehensive compilation of references and methods dealing with undesirable facility location on networks. In this sense, more than 90 papers have been briefly commented, along with several models on undesirable single facility location on networks with multiple criteria that have been analyzed and described.

Most of the papers most regarding location problems address the siting of facilities such as emergency services (police/fire stations), educational centers, medical facilities, etc., that are considered desirable by the surrounding population. However, there are some other facilities such as garbage dump sites, landfills, chemical plants, nuclear reactors, military installations and polluting (noise/gas) plants that turn out to be undesirable (repulsive) for the surrounding population, that avoids them and tries to stay away from them. In this sense, Erkut and Neuman (1989) distinguish between *noxious* (hazardous) and *obnoxious* (nuisance) facilities, although both can be simply regarded as *undesirable*.

Despite these undesirable facilities being necessary in general to the community, for instance, garbage dump sites, gas stations, electrical plants, etc., the location of such facilities might cause a certain disagreement among the population. Such a disagreement may result in a true opposition of people to the installation of undesirable facilities in their neighborhood. Moreover, in the last decade, a new nomenclature has been developed to define these oppositions: NIMBY (*Not In My Back Yard*), NIMNBY (*Not In My Neighbor's Back Yard*), NIABY (*Not In Anyone's Back Yard*), NIMTOO/NIMTOF (*Not In My Term of Office*), NOPE (*Not On Planet Earth*), LULU (*Locally Unwanted Land Use*), BANANA (*Build Absolutely Nothing Anywhere Near Anyone*).

M. Colebrook (✉) • J. Sicilia
Depto. Estadística, Investigación Operativa y Computación,
Universidad de La Laguna, Tenerife, Spain
e-mail: mcolesan@ull.es; jsicilia@ull.es

R. Batta and C. Kwon (eds.), *Handbook of OR/MS Models in Hazardous Materials Transportation*, International Series in Operations Research & Management Science 193,
DOI 10.1007/978-1-4614-6794-6_6,

Taking these concerns into account, and due to the great concern on environmental issues that has arisen in the last decades, this chapter aims to analyze some undesirable facility location models, preferably on networks.

On the other hand, network location models have usually dealt with single criterion problems, that is, concerning one weight per node and/or one length per edge. However, to properly model many real problems the decision maker requires placing parameters on both the nodes such as demand, importance, number of customers, etc., and on the edges such as length, time, travel cost, etc. Many researchers, in several excellent reviews and books, for instance, ReVelle, Cohon and Shobrys (1981a, b), Ross and Soland (1980), Krarup and Pruzan (1990), Current, Min and Schilling (1990), Daskin (1995), have deeply emphasized the importance of dealing with multiple objectives in Location Analysis.

Furthermore, many authors have deeply argued in the literature that a lot of multicriteria/multiobjective location problems have remained unresearched even though this topic has become quite relevant in the last three decades. In this sense, Erkut and Neuman (1989) emphasized on the need for multiobjective approaches to the siting of undesirable facilities when they stated that (p. 289): "Current models can be used to generate a small number of candidate sites, but the final selection of a site is a complex problem and should be approached using multiobjective decision making tools." Daskin (1995) and Zhang (1996) also pointed out not only the need to include multiple criteria in undesirable facility location problems, but also the fact that poor attention has been paid by researchers to these problems and hence, scarce research has been done in this promising field. Therefore, section "Undesirable Facility Location on Multicriteria Networks" of this chapter mainly focuses on network location models concerning multiple criteria, in terms of considering several node weights and several edge lengths.

In the remaining paragraphs we summarize the contents of the chapter. In section "Network Location Models Within Location Theory" we justify the importance of Network Location Models within the field of Location Theory. Section "Brief Historical Background and Review of the Literature" allows the reader to get acquainted with the definition and classical literature in Location Theory. In this respect, more than 90 references are reviewed, from surveys and books in general location problems, to more specialized papers on multicriteria undesirable location models on networks. Section "Basic Definitions and Notation" presents the basic definitions and notation used throughout the chapter for the standard networks and also for networks with multiple parameters on both nodes and edges. The problem of locating an undesirable facility on a network is addressed in section "Locating Undesirable Facilities on Simple Networks." Section "Undesirable Facility Location on Multicriteria Networks" is devoted to the location of undesirable facilities on multicriteria networks. In section "Conclusions and Directions for Further Research" we summarize the conclusions and describe some open problems that may be researched with regards to the location/transportation of hazardous materials. Finally, the last section lists all the bibliography referenced.

Network Location Models Within Location Theory

In a very wide sense, location problems deal with finding the right site where one or more new facilities (services) should be placed, in order to optimize (minimize or maximize) some specified criteria, which are usually related to the distance (performance measure) from the facilities to the demand points (customers).

The mathematical field that formulates location problems, builds up appropriate mathematical models and derives methods for solving them is called *Location Theory* or *Location Analysis*. Being a branch of the *Operational Research* framework, this subject provides decision-makers with qualitative tools to find good solutions for realistic location decision problems. Besides, modern Location Analysis has drawn the interest of practitioners such as economists, geographers, regional planners and architect researchers, as well as researchers in diverse fields like *Industrial Engineering*, *Management Science* and *Computer Science*.

Regarding location theory classification, location problems mostly fall in one of the following three categories:

- *Continuous location*: locations are allowed to be anywhere in a continuous d dimensional space.
- *Discrete location*: a finite number of possible locations on the space are specified in advance. Sometimes it is also called location-allocation.
- *Network location*: special kind of location problems which are modeled on networks or trees.

In this chapter we focus on *Network Location Problems*. This type of problems can model real location problems on river networks, air transport networks (flight corridors), ocean transport networks (shipping lanes); highways, roads, avenues and street networks; and communication and computer networks. The literature on network location is full of inherent real applications, some of which will be briefly mentioned in the next literature review.

Despite most of these location problems seeming to be close related to the contemporary world, they have been originally proposed centuries ago. This is described in the next section where we present a brief historical background, as well as a comprehensive review of the literature on Network Location Analysis. After this, we introduce a general notation and basic concepts in Location Theory. These concepts are used to describe the models developed in the following sections.

Brief Historical Background and Review of the Literature

The origin of modern location theory is credited to A. Weber (1909), who incorporated the original problem proposed by Fermat into Location Analysis in his influential essay on the theory of industrial location *"Über den Standort der Industrien"* (Theory of the location of industries), translated later by Friedrich (1929).

Jordan (1869) obtained a characterization of the median set of a tree. With regards to location problems on general networks, we must mention Hakimi (1964), who introduced both the median and the center on weighted networks, and thus, his principal paper set the foundations for the development of forthcoming network location problems.

Literature on Location Analysis is extremely huge and fairly interlaced. One of the first and most extensive compilations is due to Domschke and Drexl (1985), who compiled a bibliography of over 1,800 papers. Later on, Drezner (1995) provided more than 1,200 references. Trevor Hale (1998) keeps a web page with a list of over 3,400 location science, facility location and related references. And this number keeps counting!

Next, we cite some reviews, surveys and books on classical location problems.

Surveys, Reviews and Books on Location Problems

For a reader not quite acquainted with Location Analysis, we now cite some interesting bibliography on classical location problems and models.

A classical and state-of-the-art text on discrete location problems is due to Mirchandani and Francis (1990). Drezner (1995) presented a wide-ranging survey of location analysis. Drezner and Hamacher (2004) covered theory, methodology and selected applications of Location Analysis. Eiselt and Sandblom (2004) present a unified treatment of decision analysis, location theory and scheduling, with topics ranging from multicriteria decision-making to location and layout planning.

Chan (2005) describes procedures to perform site location, land-use planning, location-routing, competitive allocation of products/services and spatial forecasting. Nickel and Puerto (2005) address the flexible location problem called the Ordered Median Problem (OMP), presenting both structural properties and solution approaches of the OMP for continuous, network and discrete location problems.

Some of the latest books on Location Analysis are from Farahani and Hekmatfar (2009), who describe the four main parts (customers, facilities, space and metrics) for each specific location model exemplified by real-world cases; and from Eiselt and Marianov (2011), who compile several contributions written by eminent experts in the field of location analysis, surveying the original seminal papers and providing an up-to-date review of the latest references.

Since the main goal of this chapter is to describe the major achievements on network location regarding hazardous facilities, in the subsequent sections we review, in chronological order, the most outstanding references on location of undesirable facilities on networks considering both one single criterion and several criteria.

Undesirable Facility Location Problems on Networks

There are not many papers devoted to location of undesirable (sometimes called *obnoxious*) facilities on networks. This subject barely emerged in mid 1970s, and gradually drew the interest of researchers due to environmental issues. These types of problems are the opposite of the classical center (*minimax*) and median (*minisum*) problems, and hence, they are usually modeled using the *maximin* and the *maxisum* criteria. Other authors established alternative criteria which are not covered in this chapter. Slater (1975) defined the security center and security centroid of a graph using the criterion that a vertex u is "more central" than vertex v if there are more vertices closer to u than to v.

In the same way as Hakimi is the forerunner of Network Location Analysis, Church and Garfinkel (1978) are the precursors of the location of undesirable facilities on networks. They dealt with the problem of locating a point on a network so as to maximize the sum of its weighted distances (*maxisum*) to the nodes, and proposed an algorithm in $O(mn \log n)$ time. The optimal point was called *maxian*. Minieka (1983) characterized the *anticenter* and *antimedian* location models. The former is formulated as a *maxmax* problem, whereas the latter is a directed approach to that of Church and Garfinkel (1978).

Ting (1984) treated the problem of locating a single facility in a tree network considering the maxisum criterion, and provided a solution algorithm with computational effort $O(n)$. Kuby (1987) pointed out that the optimal maximin objective value could be used as a lower bound on the distances between selected facilities. Moon (1989) addressed the problem of finding a point in a tree network whose distance to the closest pendant vertex (incident to a single edge) is maximal. He presented a polynomial time algorithm in $O(n)$ time.

Tamir (1988) demonstrated that for some center and (obnoxious) location problems it is possible to take advantage of dynamic data structures to achieve better complexity bounds. Labbé (1990) dealt with the location of an obnoxious facility on a network using a voting procedure. She also defined the *anti-Condorcet* point as a point such that no other point is farther from a strict majority of users. Tamir (1991) discussed new complexity results for several models dealing with the location of obnoxious or undesirable facilities on graphs such as p-maximin and p-maxisum problems, which concern the location of p facilities under the maximin and maxisum objectives, respectively.

Regarding location and routing of hazardous wastes, Stowers and Palekar (1993) developed a combined model that quantifies the total exposure of the population during transportation as well as long term storage.

Kincaid and Berger (1994) studied the problem of selecting a subset of size p of the distance matrix column indices such that the smallest row sum in the resulting $n \times p$ submatrix is as large as possible. Drezner and Wesolowsky (1995) considered the problem of locating a point that should be as far as possible from arcs and nodes of a network. Berman et al. (1996) approached the location of a new facility on a network so that the total number (weight) of nodes within a prespecified distance is minimized.

Moreno-Pérez and Rodríguez-Martín (1999) addressed the problem of locating an undesirable facility on a network maximizing a convex combination of the average and minimum distance to the population. Since this is the opposite of the cent-dian model, they called it the *anti-cent-dian*. The same problem including distance constraints was previously pointed out by Moon and Chaudhry (1984) as the *anticenter-maxian* model. Colebrook and Sicilia (2006) improved the anti-cent-dian facility location problem on networks, providing an efficient $O(mn)$ time algorithm.

Although Tamir (1988, 2001) already presented a brief $O(mn)$ method for the maximin problem, Melachrinoudis and Zhang (1999) solved the location of a point on a network under the maximin criterion with the same computational effort. Soon after, Berman and Drezner (2000) developed the same problem from a linear programming viewpoint in $O(mn)$ time as well. Colebrook et al. (2002) presented a different model formulation and improved upper bounds for the location of an undesirable (obnoxious) center on general networks, which diminished the computational time required to get the solution.

Salhi et al. (2000) provided an alternative analytical approach to the Voronoi based method for the weighted 1-maximin location problem, which concerns the location of one facility under the maximin criterion. Their enhanced method was relying on two reduction tests and a suitable branch and bound scheme. Zhang et al. (2000) developed an algorithm to safely route hazardous materials on network, assessing the potential risks on human population by GIS techniques.

Burkard et al. (2001) derived algorithms with linear running time in the cases where the network is a path or a star, and improved previous results proposed by Tamir (1988, 1991). In a quite similar approach, Burkard and Dollani (2003) studied the pos/neg 1-center problem on networks, which asks to minimize a linear combination of the maximum weighted distance of the center to the positive and negative weighted vertices respectively. On networks, they provided an $O(mn \log n)$ algorithm, whereas on star graphs the problem can be solved in linear time. They also studied the extensions to the location of p facilities on trees.

López-de-los-Mozos and Mesa (2001) analyzed a new locational equity measure defined as the maximum absolute deviation. They investigated its properties and proposed an algorithm for locating a single facility on a network such that it minimizes this new criterion. Carrizosa and Conde (2002) addressed a p-facility location for semi-desirable facilities whose location was restricted to the edges of a planar network with rectilinear edges.

Cappanera et al. (2003) addressed the problem of simultaneously locating obnoxious facilities and routing obnoxious materials between a set of built-up areas and the facilities, defining a discrete combined location-routing model referred to as the Obnoxious Facility Location and Routing model (OFLR).

Berman and Wang (2004) considered, among others, the 1-antimedian and 1-maximin undesirable facility location problems on undirected networks with node weights as independent discrete random variables. Colebrook et al. (2005b) studied the problem of locating an undesirable facility on a network so as to maximize its total weighted distance to all nodes, giving a new upper bound and a new algorithm in $O(mn)$ time.

Berman and Wang (2006) studied the 1-median and 1-antimedian problems with probabilistic node demands, which are assumed to be independent continuous random variables, whereas Berman and Wang (2007) considered the problem of locating semi-obnoxious facilities provided that some demand points, within a certain distance from an open facility, are expropriated. In Berman and Wang (2008), the problem of locating a semi-obnoxious facility was considered assuming that close demands nodes could be expropriated by the developer.

Erkut and Alp (2007) considered the problem of selecting routes for hazardous material transportation, applying their model to the road network of Ravenna (Italy). Berman et al. (2007) presented a novel methodology based on arc-covering to determine the network optimal design so as to maximize the ability to respond to dangerous incidents. Their results assessed the emergency response capability to transport incidents in Quebec and Ontario (Canada).

Recently, Berman and Huang (2008) compared several mathematical formulations to locate undesirable facilities on a network so as to minimize the total demand covered subject to the condition that no two facilities are allowed to be closer than a pre-specified distance. Drezner et al. (2009) analyzed the location of a facility inside a planar network with nuisance/hazard created on its links, so the total nuisance should be minimized. Lately, Yamaguchi (2011) examined a line network model where individuals collectively choose the location of an undesirable public facility through bargaining with the unanimity rule.

Regarding surveys and reviews on undesirable location, Moon and Chaudhry (1984) discussed and surveyed uncapacitated distance constrained network location problems such as maxian, defense, anti-center, dispersion, anticenter-maxian and dispersion-defense models. A widely cited review on this subject was due to Erkut and Neuman (1989), who brilliantly surveyed over sixty papers on maximization location models and presented a synthesis of the solution methods. In the same sense, Erkut and Verter (1995), and later Verter and Erkut (1995), overviewed and treated logistics models involving hazardous materials.

In addition to the network models, it is also worth citing some papers due to their contribution and direct application. One of these papers is the overview on (semi-) undesirable facility location by Plastria (1996). A close related paper by Carrizosa and Plastria (1999) presented a critical overview of the mathematical models used in the field of semi-obnoxious facility location. Murray et al. (1998) reviewed several approaches for addressing equity and community impact in the location of undesirable facilities. In an excellent report, Cappanera (1999) surveyed mathematical models for undesirable location problems in the plane and particularly on networks.

Concerning straightforward applications, we must cite Cáceres et al. (2007), who considered the problem of locating a waste pipeline in a coastal region, taking into account the maximization of two criteria.

There is no book solely devoted to location of undesirable facilities yet. Daskin (1995) discussed dispersion models, outlined a maxisum problem and commented on some multiobjective location problems. In Puerto (1996), there is a chapter concerning location of undesirable centers on the plane as well as on networks.

The two latest book chapters on undesirable facility location are by Hosseini and Esfahani (2009), who reviewed obnoxious facility location problems from the point of view of their classification, diverse models, applications, solutions and techniques, and case studies; and by Melachrinoudis (2011), who surveyed and assessed the classical contributions on undesirable facility location from the late 1970s till nowadays.

In the next section we review the most relevant references on multiobjective/multicriteria undesirable facility location models on networks.

Multicriteria Undesirable Facility Location on Networks

Surprisingly, literature on multicriteria undesirable facility location starts in the late 1980s. It seems that the concern on the location of undesirable facilities has grown only in the last years, along with the use of multiobjective/multicriteria tools to model and solve such problems.

Ratick and White (1988) proposed a multiobjective model for the location of undesirable facilities considering three objectives: minimizing the facility location costs, minimizing the opposition to the siting plan, and maximizing equity. List and Mirchandani (1991) presented a combined routing/siting model that can be used not only for making routing decisions on waste shipments, but also for siting decisions of waste treatment facilities. Risk, cost and risk equity were considered jointly in a multiobjective framework. A simplified form of their model was applied to the Capital District of the State of New York. Erkut and Neuman (1992) developed a multiobjective model for the location of one or more undesirable facilities to service a region which minimizes the total cost of the facilities located, the total opposition to such facilities, and the number of power-generating stations.

By means of a multiobjective model, Rahman and Kuby (1995) examined the tradeoffs between minimizing costs (transshipment and fixed-charge problems) and public opposition (decreasing distance function from the facility) in the location of a solid waste transfer station. A case study was also accomplished in the City of Phoenix, Arizona.

Giannikos (1998) presented a multiobjective model for locating disposal facilities and transporting hazardous waste along the links of a network considering four objectives, namely, minimization of total operating cost, minimization of total perceived risk, equitable distribution of risk among population centers and equitable distribution of the disutility caused by the operation of the treatment facilities.

Zhang and Melachrinoudis (2001) considered the problem of locating an obnoxious facility on a general network using two objectives, maximizing the minimum weighted distance from the point to the vertices (maximin) and maximizing the sum of weighted distances between the point and the vertices (maximsum). Hamacher et al. (2002) presented a polynomial time algorithm for the location of a semi-obnoxious facility on networks, and generalized the results to include maximin and minimax objectives.

Skriver and Andersen (2003) modeled a semi-obnoxious facility location problem as a bicriterion problem in both the plane and the network case, applying these models to the location of a new international airport in the Jutland mainland, Denmark.

Colebrook and Sicilia (2007) analyzed several location problems of undesirable facilities on multicriteria networks establishing new properties to characterize the efficient solutions and rules to remove inefficient edges. Tuzkaya et al. (2008) addressed the problem of locating an undesirable facility in Istanbul (Turkey) using the multi-criteria decision making technique called Analytic Network Process (ANP). Lately, Zhao and Shuai (2010) proposed a new multiobjective 0-1 integer LP model for the location-allocation problem in response network design for hazardous materials transportation.

Once more, the ensuing papers are commented for their real life application, though they might not be addressed on networks. Melachrinoudis et al. (1995) developed a dynamic (multiperiod) multiobjective mixed integer programming model for locating landfills. Their objectives are: minimization of total cost during the planning horizon, minimization of total risk posed on population centers, minimization of total risk posed on ecosystem and minimization of risk inequity over all individuals and time periods in the planning horizon.

Hokkanen and Salminen (1990) described an application of multicriteria decision aid to the location of a waste treatment facility in eastern Finland. The alternative locations for the new facility were considered based on 14 criteria by 28 decision makers.

Rakas et al. (2004) developed a multiobjective model to determine the location of undesirable facilities using real-world data. Alumur and Kara (2007) proposed a new multiobjective hazardous waste location-routing model that minimizes the total cost and the transportation risk, and it was implemented in the Central Anatolian region of Turkey.

To the best of our knowledge, there is no published book on multicriteria undesirable facility location problems on networks. However, Daskin (1995) devoted a complete section of a chapter to emphasize the need of more multicriteria models on undesirable facility location.

Lastly, before presenting some basic definitions and the notation, we briefly comment four doctoral dissertations on multicriteria undesirable location. Saameño (1992) studied the problem of locating obnoxious facilities on a polygonal region with multiple objectives. Zhang (1996) mainly developed algorithms to solve the 1-maximin problem on a network, and the maximin-maxisum network location problem. Skriver (2001) investigated, among other models, the bicriterion semi-obnoxious location problem, the multicriteria semi-obnoxious network location problem with sum and center objectives and the bicriteria network location problem with criteria dependent lengths and minisum objectives. Finally, Colebrook (2003) devoted several chapters to analyze and develop some undesirable location models on networks.

Basic Definitions and Notation

In this section we introduce the concepts and basic definitions that are essential for the remaining sections. We begin with the notation on classical network models, followed by the definitions related to networks with multiple criteria.

Standard Networks

Mathematical networks can model innumerable real world problems such as aisle/road networks, river/air/ocean transport networks or communication/computer networks. All of these networks are barring exceptions, simple (no loops or multiple edges), connected and undirected.

Thus, let $N = (V, E)$ be a network with such features, where $V = \{v_1, v_2, \ldots, v_n\}$ denotes the set of vertices or nodes, and $E = \{(v_s, v_t) : v_s, v_t \in V\}$ the set of edges, with $n = |V|$ and $m = |E|$. The nodes represent demand, supply or junction points on which existing facilities or clients are already placed, whereas edges correspond to transportation lines, roadways, railways or communication channels.

Each node $v_i \in V$ is set with a positive weight w_i as follows:

$$\begin{aligned} w: \quad V \quad &\to \quad \mathbb{R}_+ \\ v_i \in V &\to w(v_i) = w_i > 0 \end{aligned}$$

This weight w_i stands for demand rates, time/cost/loss per unit distance, number of clients, probability that a demand occurs at node v_i, or even the importance of a potential damage. Obviously, the weights are positive because a weight $w_i = 0$ means null demand, time, etc, and hence it makes no sense.

On the other hand, each edge $e = (v_s, v_t)$ is labeled with a positive number l_e in terms of the following length function:

$$\begin{aligned} l: \quad E \quad &\to \quad \mathbb{R}_+ \\ e = (v_s, v_t) \in E &\to l(e) = l_e > 0 \end{aligned}$$

Thus, a point x inside edge e ranges in the interval $[0, l_e]$. This length represents travel time/cost, reliability or any other travel attribute. The lengths are positive since any $l_e = 0$ implies a null distance between v_s and v_t, and hence, it can be discarded. Figure 1 shows a network with $n = 5$ nodes and $m = 7$ edges. Weights w_i are in bold, whereas lengths l_e are in italic.

Besides, each edge is assumed to be rectifiable, in the sense that there is a one-to-one correspondence between each edge and the interval $[0, 1]$. Hence, given any edge $e = (v_s, v_t) \in E$ of length l_e and an inner point $x \in e$, then there is a unique

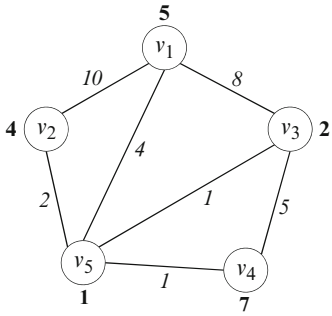


Fig. 1 Network with five nodes (weights in *bold*) and seven edges (lengths in *italic*)

number $t_e(x) \in [0, 1]$ such that $t_e(x)l_e$ and $(1 - t_e(x))l_e$ are the lengths along edge e between v_s and x, and x and v_t, respectively.

A *path* is a sequence of adjacent edges, with each of the adjacent edges sharing a common node. Then, for each pair of nodes $v_a, v_b \in V$ we define a *distance* $d(v_a, v_b)$ between these two nodes as the length of any shortest path in N joining v_a and v_b. Moreover, given any two points $x, y \in N$, the distance $d(x, y)$ is the length of the shortest path between x and y. Given a certain edge $e = (v_s, v_t)$, it is sometimes possible that $d(v_s, v_t) < l_e$ since the edge may not provide the shortest path between the nodes v_s and v_t. This distance function $d(\cdot, \cdot)$ satisfies the following *metric properties* for any $x, y \in N$:

1. *Nonnegativity*: $d(x, y) \geq 0$, with $d(x, y) = 0$ if $x = y$.
2. *Symmetry*: $d(x, y) = d(y, x)$.
3. *Triangle inequality*: $d(x, y) \leq d(x, z) + d(z, y)$, for any $z \in N$.

At this point, the principal issue to be emphasized is that network location models are usually based on the assumption that travel distances are lengths of shortest paths. In this sense, given any edge $e = (v_s, v_t) \in E$, a node $v_i \in V$ and an inner point $x \in e$, we define the distance between point x and node v_i as:

$$d(x, v_i) = \min\{x + d(v_s, v_i), l_e - x + d(v_t, v_i)\}$$

The point on e where $d(x, v_i)$ attains its equilibrium, i.e. $x + d(v_s, v_i) = l_e - x + d(v_t, v_i)$, is called a *bottleneck point* b_i, with

$$b_i = \frac{d(v_t, v_i) + l_e - d(v_s, v_i)}{2}$$

A fundamental property of network distances is the following *piecewise linearity and concavity property*. This property states that the function in $x \in e = (v_s, v_t)$ defined by $d(x, v_i)$:

1. Is continuous on e.
2. As x varies from node v_s to v_t in edge e, either

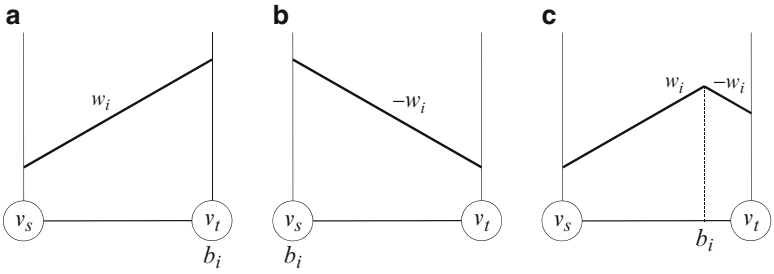


Fig. 2 The three possible plots of $d(x, v_i)$

- Increases linearly with slope w_i (see Fig. 2a), or
- Decreases linearly with slope $-w_i$ (see Fig. 2b), or
- First increases linearly and then decreases linearly, with a breakpoint at b_i (see Fig. 2c).

3. Is *concave*, in the sense that a line segment joining any two points on the graph of the function lies on or below such graph.

These are the basic concepts on standard networks. In the next section we introduce the basic notions on networks with multiple criteria, namely, considering several weights on each node as well as several lengths on each edge.

Networks with Multiple Parameters on Nodes and Edges

Most of the huge literature on network location problems deals with the optimization of one *single criterion*. This criterion is usually associated with the weighted distance from a certain point to the rest of the nodes, for example, the minimization of the total weighted distance from a facility to the customers.

However, there are many applications in which several parameters need to be considered on each node and on each edge. Thus, several weights on each node may represent different criteria to be considered by the decision-maker(s), namely, demand rate, importance, number of potential clients, etc. On the other hand, several lengths (travel costs) on each edge might deal with distance, travel time, traffic congestion, toll rate, travel cost, etc.

In this sense, on each node $v_i \in V$, the previous weight function is now replaced by the following:

$$\begin{array}{rccc} w: & V & \to & \mathbb{R}^p \\ & v_i \in V & \to & w(v_i) = w_i = (w_i^1, \ldots, w_i^p) \end{array}$$

where p is the number of weights per node. For any vector of weights w_i, each w_i^r is a nonnegative value for $r = 1, \ldots, p$, and we assume that not all are equal to zero.

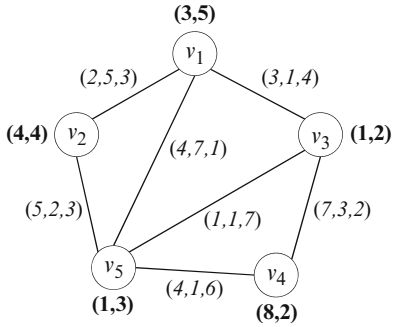


Fig. 3 Five-node and seven-edge network with several parameters

Likewise, each edge is set with a vector of lengths (costs), as follows:

$$\begin{array}{rcl} l: & E & \to & \mathbb{R}^q \\ & e = (v_s, v_t) \in E & \to & l(e) = l_e = (l_e^1, \ldots, l_e^q) \end{array}$$

in which q is the number of lengths. Again, we assume that each component l_e^r is nonnegative for any vector l_e, and not all $l_e^r = 0$, for $r = 1, \ldots, q$.

As an example of a network holding several parameters, Fig. 3 shows the same network as Fig. 1, but with two weights per node (in bold) and three lengths per edge (in italic).

Let r be a length index, with $1 \leq r \leq q$, and let $x \in e = (v_s, v_t)$ be a point inside edge e. Then, $c_e^r(x, v_s)$ is defined as the piece of line segment between x and v_s considering length r. Obviously, we have that $0 \leq c_e^r(x, v_s) \leq l_e^r$, with $c_e^r(x, v_t) = l_e^r - c_e^r(x, v_s)$.

For each pair of nodes $v_a, v_b \in V$ we can define the *distance* $d^r(v_a, v_b)$ between these two nodes as the length of any shortest path in N joining v_a and v_b considering length r. Likewise, given any two points $x, y \in N$, the distance $d^r(x, y)$ is the length of the shortest path between x and y. These q distance functions also comply with the metric properties stated in the preceding section.

Given any node $v_i \in V$, we have that

$$d^r(x, v_i) = \min\{c_e^r(x, v_s) + d^r(v_s, v_i), c_e^r(x, v_t) + d^r(v_t, v_i)\}$$

denotes the distance between a point and a node considering length r, with $b_i^r = (d^r(v_t, v_i) + l_e^r - d^r(v_s, v_i))/2$ being the bottleneck point concerned with node v_i. These q network distance functions fulfill the piecewise linearity and concavity property as well.

Finally, we introduce some basic theory on multicriteria/multiobjective optimization. Usually, *multicriteria* models are those which perform a simultaneous optimization of several incommensurable objectives, for instance, minimizing the maximal travel distance and minimizing the total travel cost. On the other hand, a closely related concept is that of *vector optimization*, which determines the nondominated solutions to a multicriteria problem.

In this sense, let $f = (f_1, f_2, \ldots, f_k)$ and $g = (g_1, g_2, \ldots, g_k)$ be two vectors belonging to $\mathbb{R}^k$. Vector f is said to *dominate* vector g, and it is denoted by $f \prec g$, if and only if:

$$f_i \leq g_i, \forall i = 1, \ldots, k \quad \text{and} \quad \exists j \in \{1, \ldots, k\} : f_j < g_j$$

Then, given the subset of vectors $U \subseteq \mathbb{R}^k$, a vector $f \in U$ is called *non-dominated*, *efficient* or *Pareto optimal* (Pareto 1896) with respect to subset U if there is no other vector $g \in U$ such that $g \prec f$. The set of all non-dominated vectors with respect to U is denoted by U_{ND}. For a further knowledge in multicriteria optimization, the reader is referred to Steuer (1986).

Having described the basic concepts and the notation used to model the location problems developed in this chapter, in the following sections we present the location models for undesirable facilities on networks.

Locating Undesirable Facilities on Simple Networks

In the following subsections we develop several models that can be used to locate hazardous facilities on networks considering one single criterion. These models comprise the undesirable center problem, the maxian problem, and the anti-centdian problem.

The Undesirable Center Problem

As we remarked in the literature review, there are not many papers devoted to undesirable location on networks. One of them is by Melachrinoudis and Zhang (1999), who proposed a O(mn) time algorithm based on three upper bounds and on a modified procedure of Dyer (1984). However, their upper bounds can be tightened, and the procedure can be improved by means of a more convenient formulation of the solution. The other paper by Berman and Drezner (2000) approaches the problem in a linear programming way. Though it has the same theoretical complexity, its running time is extremely high, since the algorithm has to process every single edge.

Now, we formulate the undesirable 1-center (maximin) problem on networks. Given any point $x \in N$ we define $f(x) = \min_{v_i \in V} w_i d(x, v_i)$.

Then, the problem consists of calculating

$$\max_{x \in N} \min_{v_i \in V} w_i d(x, v_i) = \max_{x \in N} f(x)$$

and a point $x_N \in N$ is an undesirable 1-center point iff $f(x_N) = \max_{x \in N} f(x)$.

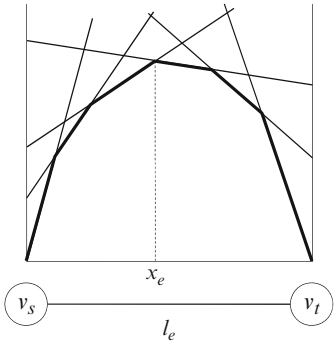


Fig. 4 Objective function $f(x)$, which is actually the lower envelope of all distance functions

This problem is the opposite of the 1-center problem (minimax), so it could be called the *anti-center*. Unfortunately, this term was already coined by Minieka (1983) to define the *maxmax* problem. We instead propose the term 1-*uncenter* (*un*desirable *center*) to define the optimal location point.

If there is at least one vertex v_i such that $w_i = 0$, then $f(x) = 0$, $\forall x \in N$ and obviously any point on network N would be a 1-uncenter. Therefore, we consider only $w_i > 0, \forall v_i \in V$.

When all the node weights are equal, $\forall v_i \in V, |w_i = w$, the local 1-uncenter x_e is sited at the central point of edge e. Therefore, the unweighted 1-uncenter x_N is located in the middle of the longest edge(s) (see Melachrinoudis and Zhang 1999; Berman and Drezner 2000). This is done in O(m) time.

However, when all node weights are not equal, we can reformulate the 1-uncenter problem over each edge $e = (v_s, v_t) \in E$ as follows: $x_N \in N$ is a 1-uncenter point iff $f(x_N) = \max_{e \in E} f(x_e)$.

Since the local 1-uncenter point is the maximum value of the concave objective function $f(x)$, as shown in Fig. 4, it should be located at the intersection of two distance functions lines with opposite sign slopes. Our goal is to find in the lower envelope of function $f(x)$ these two lines and the intersection point between them.

By introducing new tighter bounds that can significantly reduce the number of edges and the number of distance function lines over each edge, and by means of a more convenient problem formulation, we developed a new O(mn) time algorithm, which is briefly outlined in Algorithm 1.

This method has been applied to the following network depicted in Fig. 5, which has $n = 8$ nodes and $m = 18$ edges. The weights (in bold) on the nodes range randomly from 1 to 9, whereas the lengths (in italics) randomly vary from 1 to 49.

The solution to this example is $F_N = 50$, which is the 1-uncenter value at $S = \{(26, e_{36})\}$. Note that the algorithm processes only 6 out of 18 potential edges. Even though these numbers may not seem important, they will be quite relevant when the network size gets bigger, both in nodes and edges.

To test the computational effort of the new algorithm, several experiments were run for different sets of graph densities as well as for planar networks. These tests showed that the running times of the new algorithm are faster than both the

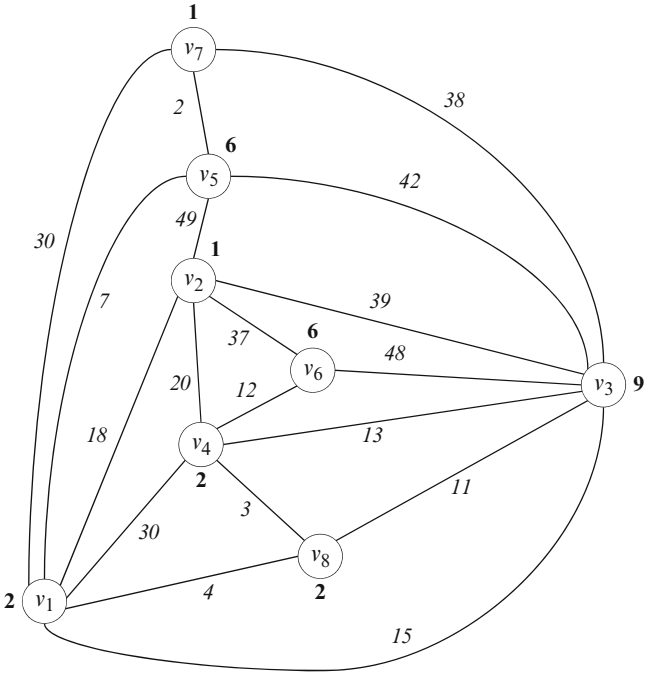


Fig. 5 Planar network with $n = 8$ and $m = 18$

approaches by Berman and Drezner (2000) and Melachrinoudis and Zhang (1999), since the number of edges processed are less, gaining in some instances a reduction of over 50%. As a consequence, the computing times of the new algorithm are better, achieving in some cases a reduction of 80%. Besides, the reduction augments as the number of nodes n increases.

For more details in the mathematical results, algorithm description, the example trace and the computational time experiment, the reader is referred to Colebrook *et al.* (2002).

Algorithm 1. The uncenter function.

function *UnCenter*(Network N, Distance Matrix d)
{ *// Current best value on network N*
$F_N := 0$
// Solution set
$S := \emptyset$
for all edges $e := (v_s, v_t) \in E$ **do**
{ *// Compute the upper bounds*
Compute upper bound $UB1$
if $F_N > F_{UB1}$ **then** continue to next edge

```
        Compute upper bound UB2
        if F_N > F_UB2 then continue to next edge
        Compute upper bound UB3
        if F_N > F_UB3 then continue to next edge
        // Set (x_e, F_e) to the best value found so far
        if F_UB2 ≤ F_UB3 then  (x_e, F_e) := (x_UB2, F_UB2)
        else                   (x_e, F_e) := (x_UB3, F_UB3)
        Create sets L and R. All lines must be below F_UB2.
        // Continue till the new value F_e cannot improve the current F_N,
        // or until one of the node sets becomes empty
        while F_e ≥ F_N and (L ≠ ∅ or R ≠ ∅) do
        {   Pair all nodes in L against R, using a max{|L|, |R|} matching
            (x_e, F_e):= Intersection point with minimal function value
            Project the value x_e on the lower envelope to get v_a and v_b
            x_e := Intersection point of distance lines v_a and v_b
            F_e := Distance value of point x_e
            Remove from L and R all lines above the new value F_e
        }
        if F_e ≥ F_N then
        {   F_N := F_e
            Store the pair (x_e, e) in S
        }
    }
    return (F_N, S)
}
```

The Maxian Problem

As stated in the review section, the literature on undesirable network location began in the mid 1970s with Church and Garfinkel (1978), who defined and solved the 1-maxisum (*maxian*) problem in $O(mn \log n)$ time, being n the number of nodes and m the number of edges.

Later on, Tamir (1991) briefly suggested that the 1-maxisum problem could be solved in $O(mn)$ time using an algorithm given by Zemel (1984). However, to the best of our knowledge, there is no reference in the literature directly describing such an algorithm for the network 1-maxisum problem thus far. Hence, in this section we provide an algorithm which solves this problem in $O(mn)$ time.

Given any point x on network N, we define

$$f(x) = \sum_{v_i \in V} w_i d(x, v_i)$$

as the sum of weighted distances from point x to all the nodes of the network.

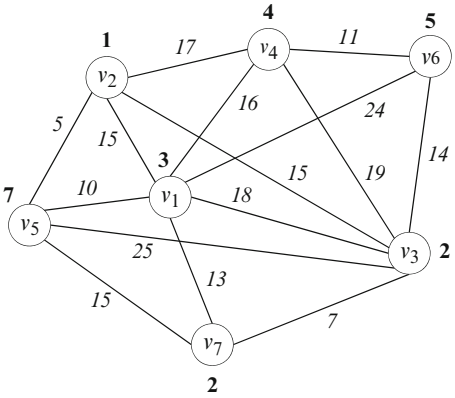


Fig. 6 Weighted network with 7 nodes and 15 edges

The undesirable *one-facility maximum* (maxian) *problem* is expressed as

$$\max_{x \in N} f(x)$$

and a point $x_N \in N$ is a *maxian* point iff $f(x_N) = \max_{x \in N} f(x)$. Several interesting properties arise for this problem.

From Church and Garfinkel (1978), an initial upper bound $UB(e)$ is derived, which is improved with a new upper bound. Likewise, this bound can be dynamically updated without increasing the total computational time. Hence, we have developed a new algorithm in $O(mn)$ to solve this problem. The procedure makes use of the new upper bound, and thus, allows skipping out from the search process as soon as the upper bound is less than the global optimum. The outline of the new procedure is showed in Algorithm 2.

To illustrate the method, consider the network in Fig. 6 with $n = 7$ nodes and $m = 15$ edges. The node weights (in bold) are integers randomly generated between 1 and 9, whereas the edge lengths range between 1 and 25.

The optimum value of $f(x)$ in this example is $f_N = 500$, which is attained in the interval $[8.5, 10.5]$ at edge (v_3, v_4). We finally remark that, due to the new upper bound, we have only processed 8 of the 15 total edges. Using the old upper bound the algorithm would have run over 13 edges. This fact is very important, since it speeds up the search for the optimal points once we are sure that the new upper bound is worse than the current best solution.

This new algorithm has been compared with the procedure proposed by Church and Garfinkel (1978), including the initial bound, on low and high dense networks, as well as on planar networks. In all cases, the new algorithm accomplishes a better performance in terms of processing times. The computational experiment was tested on complete networks with, respectively, a half, a quarter and an eighth of the total number of edges. In the three cases, the new algorithm is almost 50% faster

than Church and Garfinkel's. The same happens for planar networks. Besides, the reduction percentage in the number of edges to be processed is almost 25% less.

Again, we refer the reader to Colebrook et al. (2005b) for the details of this algorithm and the new upper bound.

The Anti-Cent-Dian Problem

In previous sections we have addressed the 1-uncenter (*maximin*) problem and the 1-maxian (*maxisum*) problem on networks. Now we are going to combine these two objectives to obtain a location criterion called the *anti-cent-dian*.

Algorithm 2. The new algorithm for the maxisum problem.

---

function *MaxianAlgorithm*(Network N, Distance Matrix d)
{ $f_N := 0$ *// Current best value on network N*
$S := \emptyset$ *// Solution set*
for all edges $e := (v_s, v_t) \in E$ **do**
{ Compute W_s and W_t
$X_e := \emptyset$
// Let X_e represent either a single point x or an interval $[x^1, x^2]$.
if (W_s and W_t yield a simple solution) **then** Store solution in X_e
else
{ $F_j := f(v_s)$, $W_j := W_s$
$F_k := f(v_t)$, $W_k := W_t$
// Compute initial value of the new upper bound $NUB(e)$
if $NUB(e) < f_N$ **then** continue to next edge
$l := 1, r := n$
while $X_e = \emptyset$ **and** $NUB(e) \geq f_N$ **do**
{ $d_q :=$ Median value of all d_i with $l \leq i \leq r$
$b_q := (d_q + l_e)/2$
Compute W_L and W_R
if (W_L, W_R and w_q yield a solution) **then** Store it in X_e
else
{ *// Search for the optimum to the left or right*
if $W_L + w_q < W_R$ **then**
$l := q + 1$, update F_j, W_j, W_L, $f(b_q)$
else $r := q - 1$, update F_k, W_k
Update the upper bound $NUB(e)$ at point b_q
}
}

```
    }
    if X_e ≠ ∅ and f(X_e) ≥ f_N then
    {   f_N := f(X_e)
        Store the pair (X_e, e) in S
    }
  }
  return (f_N, S)
}
```

The network anti-cent-dian model considers the convex combination of the maximin and the maxisum criteria. Moreno-Pérez and Rodríguez-Martín (1999) developed two algorithms that provide, respectively, the optimal location for a fixed λ that determines the convex combination, and the set of optimal locations for all convex combinations. Both of them run in $O(mn \log n)$ time. In this section we show that the complexity of the first algorithm can be reduced to $O(mn)$.

We now define the unweighted uncenter (maximin) function and the maxian (maxisum) function. Given any point x on network N, we define

$$f_{\min}(x) = \min_{v_i \in V} d(x, v_i)$$

as the minimum unweighted distance from point x to all nodes of the network. Recall that a point $y_N \in N$ is an uncenter point iff $f_{\min}(y_N) = \max_{x \in N} f_{\min}(x)$. When all node weights w_i are equal, the point y_N is located in the middle of the longest edge. Then, the uncenter point for any edge $e = (v_s, v_t)$ is $y_e = l_e/2$, and hence $f_{\min}(y_e) = l_e/2$. Thus, the local optimum can be obtained in O(1).

On the other hand, given $W = \sum_{v_i \in V} w_i$ and a point $x \in N$, we now define

$$f_{\text{sum}}(x) = \frac{1}{W} \sum_{v_i \in V} w_i d(x, v_i)$$

as the average sum of weighted distances from point x to all the nodes of the network. A point $z_N \in N$ is a maxian point iff $f_{sum}(z_N) = \max_{x \in N} f_{sum}(x)$. The local maxian point on edge e is denoted by z_e.

Finally, the *anti-cent-dian* function is defined as

$$f_{\text{acd}}(\lambda, x) = \lambda f_{\min}(x) + (1 - \lambda) f_{\text{sum}}(x)$$

and any point $x_N \in N$ maximizing $f_{\text{acd}}(\lambda, x)$ for a particular value of λ, $0 \le \lambda \le 1$, is called a λ-*anti-cent-dian* point. In particular, if $\lambda = 0$, the anti-cent-dian is equal to the maxian; whereas for $\lambda = 1$, we obtain the uncenter. Figure 7 shows a typical plot of function $f_{\text{acd}}(\lambda, x)$ over edge e. For $\lambda = 0$ the anti-cent-dian function is

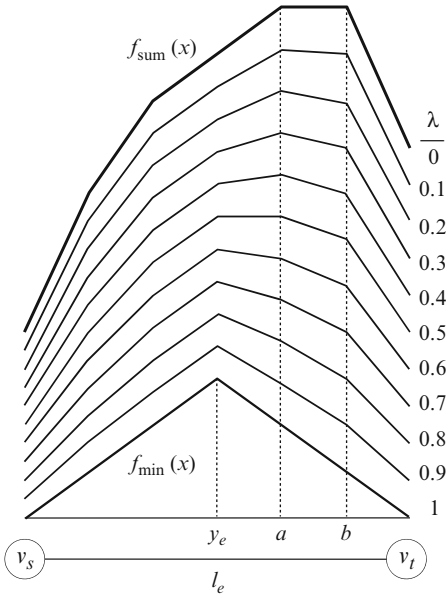


Fig. 7 Plots of $f_{\text{acd}}(\lambda, x)$ for different values of λ

$f_{\text{sum}}(x)$. As parameter λ grows, the anti-cent-dian function makes a *morphing* to the $f_{\text{min}}(x)$ function.

Taking into account some properties and given a value of λ, $0 \leq \lambda \leq 1$, the latter problem can be formulated over each edge e as follows:

$$f_{\text{acd}}(\lambda, x_e) = \max_{x \in e} f_{\text{acd}}(\lambda, x)$$

and a point $x_N \in N$ is a λ-anti-cent-dian point iff $f_{\text{acd}}(\lambda, x_N) = \max_{e \in E} f_{\text{acd}}(\lambda, x_e)$.

A method to determine all λ-anti-cent-dian points for any value of $\lambda \in [0, 1]$ in $O(mn \log n)$ time was proposed by Moreno-Pérez and Rodríguez-Martín (1999). It has derived from an $O(mn \log n)$ algorithm by Hansen et al. (1991). This complexity cannot be reduced since the algorithm is based on the computation of a convex hull of $O(mn)$ points, which is done in $O(mn \log n)$ time (see Hershberger 1989).

On the other hand, Moreno-Pérez and Rodríguez-Martín (1999) also presented an $O(mn \log n)$ procedure to obtain the anti-cent-dian point when λ is fixed to a particular value. Nevertheless, an $O(mn)$ time algorithm can be achieved, as shown in Colebrook and Sicilia (2006).

Since the following multicriteria location model generalizes the λ-anti-cent-dian problem described above, the algorithm scheme and the example for this model are shown in the next section.

Undesirable Facility Location on Multicriteria Networks

As we have stated in the preceding sections, the classical location criteria *minimax* (center) and *minisum* (median) are useless to locate an *obnoxious/noxious* (undesirable) facility. Thus, the *maximin/maxmax* and the *maxisum* criteria arose to model, respectively, the undesirable center problem and the undesirable median problem. By placing the new facility away from existing facilities, the maximin criterion reduces the effect on the worst impacted existing facility, whereas the maxisum criterion diminishes the collective effect (average) on the existing facilities.

Nevertheless, some facilities might be considered *semi-desirable* since they provide a main service to the community but they can also cause inconveniences to the neighboring areas, for instance, an airport, a train station, or any other noisy facility. These problems can be perfectly modeled combining the minimax/minisum criteria and the maximin/maxisum criteria.

In this sense, most of the undesirable facility location models analyzed in previous works are basically single-criterion. However, Erkut and Neuman (1989) emphasized on the need for multiobjective approaches to the siting of undesirable facilities. Daskin (1995) and Zhang (1996) also pointed out not only the need to include multiple criteria in undesirable facility location problems, but also the fact that poor attention has been paid by researchers to these problems and hence, little research has been done in this promising field.

Accordingly, in this section we present a multicriteria undesirable facility location model on networks with several weights on the nodes and several lengths on the edges, combining the maximin and maxisum criteria by a parameter λ. Such a model can be considered as the opposite to the multicriteria network λ-centdian problem presented in the last section and hence, it can be described as the *multicriteria λ-anti-cent-dian* problem on networks.

Given any point $x \in N$, any weight s $(1 \leq s \leq p)$ and any length r $(1 \leq r \leq q)$, let $f^{sr}_{\min}(x) = \min_{v_i \in V} w_i^s d^r(x, v_i)$ be the minimum weighted distance from x to the set of nodes. Besides, given any point $x \in N$, we define the function $f^{sr}_{\text{sum}}(x) = \sum_{v_i \in V} w_i^s d^r(x, v_i)$ as the sum of weighted distances from point x to the set of nodes, with $1 \leq s \leq p$ and $1 \leq r \leq q$.

Through a parameter λ, the convex combination of these two latter problems was addressed as the multicriteria λ-anti-cent-dian problem. Thus, given $\lambda \in [0, 1]$ and $x \in N$, the λ-anti-cent-dian function is defined as follows

$$f^{sr}_{\text{acd}}(\lambda, x) = \lambda f^{sr}_{\min}(x) + (1 - \lambda) f^{sr}_{\text{sum}}(x)$$

being $f^{sr}_{\min}(x) = \min_{v_i \in V} w_i^s d^r(x, v_i)$ and $f^{sr}_{\text{sum}}(x) = \sum_{v_i \in V} w_i^s d^r(x, v_i)$, with $s = 1, \ldots, p$ and $r = 1, \ldots, q$. This model was introduced in a previous section, though function $f_{\min}(x)$ was unweighted and $f_{\text{sum}}(x)$ was divided by the total sum of weights. Provided that both $f^{sr}_{\min}(x)$ and $f^{sr}_{\text{sum}}(x)$ are continuous, concave and

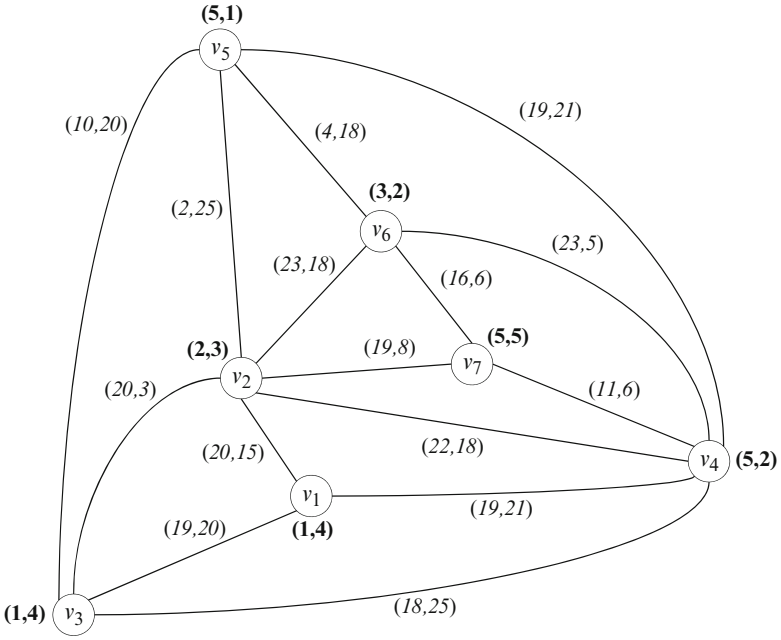


Fig. 8 A network with two lengths per edge and two weights per node

piecewise linear functions on x, the λ-anti-cent-dian function $f^{sr}_{\text{acd}}(\lambda, x)$, being a convex combination of the two latter functions, fulfills these characteristics as well.

As shown in Algorithm 3, we proposed a rule to delete inefficient edges and a polynomial algorithm in $O(k^3m^2n^2)$ time to solve this problem, being k the number of criteria. Besides, for $\lambda = 0$ we can solve the multicriteria maxian problem, whereas for $\lambda = 1$ we can obtain the solution for the multicriteria uncenter problem. Furthermore, when $p = q = 1$ this procedure can even solve the single criterion uncenter, maxian or anti-cent-dian problem. The computational experience strengthens the polynomial complexity of the algorithm as well as the effectiveness of the rule to eliminate the inefficient edges.

To illustrate the method, Fig. 8 shows a random planar network with $n = 7$ nodes, $m = 15$ edges, $p = 2$ weights per node and $q = 2$ lengths per edge. Thus, we have $k = 4$ criteria. Beside each node $v_i \in V$ we placed (in bold) two integer weights (w_i^1, w_i^2) randomly generated in the interval $[1, 5]$. Likewise, each edge $e = (v_s, v_t) \in E$ is labeled (in italics) with two integer lengths (l_e^1, l_e^2) randomly ranging in the interval $[1, 25]$. We set the parameter λ to 0.5.

The algorithm begins by removing all edges that contain no efficient point. For the example shown in Fig. 8, only 8 out of the 15 initial edges remain after the deletion, namely: (v_1, v_3), (v_1, v_4), (v_2, v_5), (v_2, v_6), (v_3, v_4), (v_3, v_5), (v_4, v_5) and (v_5, v_6). On this set of remaining edges we now proceed to compute, for each combination of weights and lengths, the functions $f^{sr}_{\min}(x)$ and $f^{sr}_{\text{sum}}(x)$.

Table 1 Set of efficient points of the network shown in Fig. 8

Edge	Efficient points
(v_1, v_3)	[3.8, 8.5]
(v_2, v_5)	[1.00923, 1.64]
(v_2, v_6)	[9.61111, 12.5]
(v_3, v_5)	[5.28571, 8.25]
(v_4, v_5)	[7.9418, 15.381]

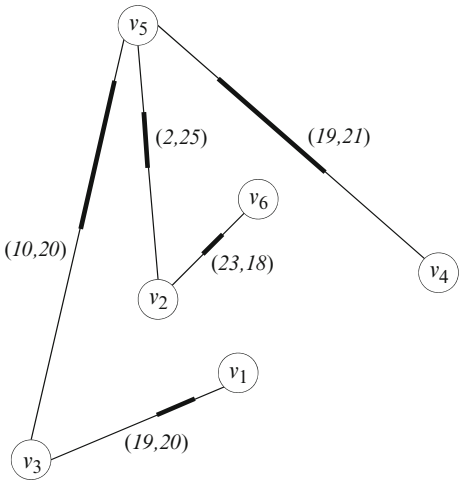


Fig. 9 Efficient points are drawn in *bold* on the partial network

Subsequently, given the parameter $\lambda = 0.5$ we calculate the λ-anti-cent-dian functions $f^{sr}_{\text{acd}}(\lambda, x)$.

Finally, the solution is the set of non-dominated segments, which are located on 5 edges only. The set of efficient points is shown in Table 1 and it is also drawn in bold on the partial network of Fig. 9.

Algorithm 3. The multicriteria λ-anti-cent-dian function (MACD).

function *MACD*(Network $N(V, E)$, DistanceMatrix d, Parameters p, q, λ)
{ Let $P := \emptyset$ be the set of candidate points to be non-dominated
Let $S := \emptyset$ be the set of possible non-dominated segments
Remove all edges containing no efficient points
for all remaining edges $e := (v_s, v_t) \in E$ **do**
{ **for** $s := 1$ to p **do**
for $r := 1$ to q **do**
{ **if** $\lambda \neq 0$ **then** Compute $f^{sr}_{\text{min}}(x)$
if $\lambda \neq 1$ **then** Compute $f^{sr}_{\text{sum}}(x)$
}
for $s := 1$ to p **do**
for $r := 1$ to q **do**

```
            Compute f^sr_acd(λ, x) = λ f^sr_min(x) + (1 − λ) f^sr_sum(x)
    Get the set of efficient points X_e
    Let x_1, ..., x_j be the sorted sequence of j breakpoints for the
        k = p × q λ-anti-cent-dian functions inside X_e
    if j = 1 then P := P ∪ {x_1}
     else
          for i := 1 to j − 1 do
          {   Let [x_i, x_{i+1}] be a segment of edge e within X_e
              S := S ∪ {[x_i, x_{i+1}]}
          }
    }
    // Let P_ND the set of non-dominated points and S_ND the
          set of non-dominated segments.
     P_ND := PointComparison(P)
     S_ND := SegmentComparison(S)
    (P_ND, S_ND) := PointAgainstSegmentComparison(P_ND, S_ND)
     return P_ND and S_ND
}
```

Algorithm 3 was programmed in C++ programming language using the class library LEDA 4.2.1, on a two 1.2 Ghz processor Pentium III with 1 Gb of RAM under Red Hat Linux.

Two kinds of experiments were performed. In both of them, random planar networks were generated with $m = 3n - 6$ edges using the generators developed by LEDA. Likewise, parameter λ varies from $\lambda = 0$ (maxian problem) to $\lambda = 1$ (uncenter problem) with a step of 0.5. Both the number of weights per node p and the number of lengths per edge q range between 1 and 3. Ten instances were generated for each combination of the latter parameters. The weight values are random integers uniformly distributed in the interval $[1, 10]$, whereas the edge lengths are random integers in the range $[1, 50]$. We remark that calculation of the distance matrix was not included in the total computing time.

In the first experiment, random planar networks were generated with $n = 10$ up to 100 in steps of 10 nodes. Table 2 shows the average times. Regardless of the number of nodes n, the computing time grows as both p and q increase. The average percentage of edges deleted is shown in Table 3. In most cases the number of removed edges is very high, achieving in some instances 99% of deletion. This issue becomes quite remarkable when $p = q = 1$ (single criterion). In this particular case, the bounds seem to be very tight, and thus, the removal rule becomes very effective since over 95% of the edges are deleted, leaving only those edges that contain the final optimal points.

Moreover, the performance of the new algorithm was also tested on bigger random planar networks with $n = 50$–500 nodes, with a step of 50 nodes. In the case of $p = q = 1$, the percentage of deletion in all cases is over 99%. However,

Table 2 Average computing time results for planar networks with $n = 10$ up to 100 nodes

$\lambda = 0$										
		$p = 1$			$p = 2$			$p = 3$		
n	m	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$
10	24	0.01	0.01	0.04	0.00	0.01	0.09	0.00	0.02	0.14
20	54	0.01	0.02	0.09	0.01	0.03	0.09	0.01	0.03	0.17
30	84	0.02	0.04	0.10	0.02	0.06	0.17	0.03	0.05	0.34
40	114	0.02	0.04	0.16	0.03	0.08	0.30	0.05	0.06	0.35
50	144	0.04	0.06	0.24	0.05	0.11	0.29	0.06	0.14	0.21
60	174	0.04	0.07	0.34	0.06	0.13	0.45	0.07	0.14	0.40
70	204	0.05	0.08	0.26	0.06	0.13	0.42	0.09	0.18	0.62
80	234	0.05	0.10	0.33	0.08	0.16	0.64	0.10	0.21	0.99
90	264	0.09	0.14	0.43	0.12	0.23	0.64	0.15	0.29	0.65
100	294	0.09	0.17	0.58	0.13	0.25	0.73	0.17	0.34	1.08
$\lambda = 0.5$										
		$p = 1$			$p = 2$			$p = 3$		
n	m	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$
10	24	0.01	0.02	0.03	0.01	0.01	0.07	0.01	0.03	0.05
20	54	0.01	0.03	0.08	0.02	0.03	0.26	0.02	0.08	0.27
30	84	0.02	0.03	0.22	0.03	0.05	0.47	0.03	0.12	0.17
40	114	0.03	0.08	0.31	0.04	0.12	0.44	0.05	0.09	0.65
50	144	0.05	0.08	0.39	0.06	0.12	0.50	0.07	0.22	0.55
60	174	0.06	0.12	0.54	0.07	0.17	0.78	0.08	0.27	1.14
70	204	0.05	0.18	0.50	0.08	0.26	0.91	0.11	0.27	1.30
80	234	0.07	0.16	0.65	0.10	0.23	0.83	0.13	0.45	1.44
90	264	0.10	0.19	0.76	0.14	0.35	1.47	0.18	0.54	1.81
100	294	0.12	0.23	0.62	0.16	0.39	1.14	0.20	0.55	1.73
$\lambda = 1$										
		$p = 1$			$p = 2$			$p = 3$		
n	m	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$
10	24	0.00	0.01	0.01	0.00	0.01	0.04	0.01	0.05	0.12
20	54	0.01	0.01	0.03	0.02	0.06	0.18	0.02	0.12	0.41
30	84	0.02	0.02	0.04	0.03	0.07	0.29	0.03	0.21	0.56
40	114	0.01	0.03	0.09	0.03	0.06	0.42	0.04	0.20	1.33
50	144	0.04	0.06	0.10	0.07	0.14	0.45	0.07	0.35	2.09
60	174	0.04	0.06	0.11	0.05	0.10	0.67	0.07	0.47	1.91
70	204	0.05	0.09	0.16	0.05	0.23	0.53	0.07	0.54	2.88
80	234	0.04	0.07	0.15	0.07	0.26	0.74	0.09	0.48	3.28
90	264	0.08	0.13	0.25	0.12	0.26	1.15	0.16	0.69	3.86
100	294	0.08	0.15	0.28	0.11	0.42	1.23	0.15	1.05	4.21

Table 3 Average percentage of edges removed for planar networks with $n = 10$ up to 100 nodes

$\lambda = 0$										
		$p = 1$			$p = 2$			$p = 3$		
n	m	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$
10	24	95.83	75.83	43.33	94.17	70.83	25.00	92.50	63.33	30.00
20	54	98.15	65.93	52.78	98.15	70.74	63.89	98.15	83.70	46.30
30	84	98.81	70.71	57.38	97.86	55.12	69.88	98.81	88.10	43.81
40	114	99.12	89.47	63.77	98.25	72.54	60.61	99.12	98.25	60.00
50	144	99.31	75.35	66.04	99.31	79.72	81.67	99.31	80.69	89.58
60	174	99.43	90.40	58.51	99.43	68.51	58.62	99.43	82.76	78.74
70	204	99.51	87.06	74.22	99.51	83.38	72.60	99.22	85.69	66.37
80	234	99.57	90.00	74.87	99.44	84.02	68.80	99.32	85.81	59.79
90	264	99.62	85.49	73.52	99.55	81.97	73.67	99.62	86.06	79.24
100	294	99.66	85.14	65.88	99.66	85.54	75.03	99.63	81.05	66.84
$\lambda = 0.5$										
		$p = 1$			$p = 2$			$p = 3$		
n	m	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$
10	24	95.83	45.00	59.17	95.83	75.00	36.67	89.17	70.00	69.17
20	54	98.15	67.22	62.96	98.15	87.04	46.85	95.93	65.74	56.67
30	84	98.81	92.26	52.86	98.81	91.19	57.62	98.81	75.95	88.93
40	114	99.12	67.54	60.70	99.12	76.14	68.68	98.86	95.44	66.58
50	144	99.31	87.78	59.93	98.75	88.26	70.49	98.06	82.64	78.54
60	174	99.43	78.39	56.26	99.25	87.18	66.26	99.25	80.06	65.00
70	204	99.51	73.24	66.13	99.31	77.94	61.52	99.36	90.20	68.58
80	234	99.57	85.98	70.38	99.44	91.03	68.63	99.49	78.08	71.07
90	264	99.62	86.70	60.68	99.51	83.86	60.49	99.62	80.98	68.94
100	294	99.66	86.46	76.53	99.66	87.07	73.91	99.56	87.45	74.01
$\lambda = 1$										
		$p = 1$			$p = 2$			$p = 3$		
n	m	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$	$q = 1$	$q = 2$	$q = 3$
10	24	95.83	45.42	64.17	95.83	56.25	22.92	87.08	33.75	32.08
20	54	98.15	79.26	54.07	85.19	40.93	23.15	81.85	50.00	38.89
30	84	98.57	92.86	72.02	67.74	65.48	31.90	94.05	47.86	38.45
40	114	99.12	84.91	45.00	85.44	78.51	32.54	89.30	58.77	31.32
50	144	99.31	77.50	67.64	72.57	68.89	48.54	89.86	56.11	25.35
60	174	99.43	83.74	75.46	94.77	88.28	41.90	89.25	45.63	38.10
70	204	99.51	75.34	66.76	96.67	67.25	57.60	94.85	53.92	37.65
80	234	99.57	94.36	79.96	94.06	66.97	57.01	94.10	71.97	36.50
90	264	99.62	85.34	70.23	92.61	79.09	45.23	88.41	60.83	35.11
100	294	99.66	82.62	72.55	97.55	65.10	49.93	93.88	47.21	39.86

when $p = q = 3$, the average edge removal percentage is greater for $\lambda = 0$ than for $\lambda = 1$, and hence, the average times in the latter are higher.

This model can be slightly modified to generalize other models studied in the literature. For instance, if we define a set of k parameters $\Lambda = \{\lambda^1, \ldots, \lambda^k\}$, then we can deal with each function $f^i_{\text{acd}}(\lambda^i, x)$ independently. Thus, the problem proposed by Zhang and Melachrinoudis (2001) might be denoted as $\max_{x \in N}(f^1_{\text{acd}}(\lambda^1, x), f^2_{\text{acd}}(\lambda^2, x))$, with $p = 2$, $q = 1$, $k = p \times q = 2$ and $\Lambda = \{\lambda^1 = 1, \lambda^2 = 0\}$. On the other hand, the multicriteria semi-obnoxious median problem presented by Hamacher et al. (2002) can be formulated as $\max_{x \in N}(f^i_{\text{acd}}(\lambda^i, x), -f^j_{\text{acd}}(\lambda^j, x))$, with $p > 1$, $q = 1$, $\lambda^i = \lambda^j = 0$ and $i \in Q_1$, $j \in Q_2$, $|Q_1 \cup Q_2| = p$, $Q_1 \cap Q_2 = \emptyset$, being Q_1 the set of obnoxious objective functions, and Q_2 the set of desirable objective functions. Obviously, if $Q_2 = \emptyset$ then we get the multicriteria maxian problem discussed in this chapter.

Finally, we remark that if $p > 1$ and $q = 1$ then the number of criteria matches the number of weights per node, i.e., $k = p$. Besides, if $\lambda = 0$ then the number of breakpoints for all the k objective functions of a given edge is $O(n)$, since all the $f^{s1}_{\text{sum}}(x)$ functions share the same breakpoints. Hence, the total number of segments to compare is $O(mn)$. Therefore, the overall complexity of the algorithm is reduced to $O(km^2n^2)$, which is the same complexity achieved by Hamacher et al. (2002) for the location of a semi-obnoxious facility on networks with sum objectives.

For more details, the reader is referred to Colebrook and Sicilia (2007).

Conclusions and Directions for Further Research

This chapter aimed to be a comprehensive compilation of references and methods dealing with undesirable facility location on networks. In this sense, more than 90 papers have been briefly commented, along with several models on undesirable single facility location on networks with multiple criteria that have been analyzed and described.

We first addressed the undesirable 1-center (uncenter) location problem on networks. By means of a more suitable problem formulation, a new $O(mn)$ algorithm can be developed, which is more straightforward and computationally faster than the ones already reported in the literature. Besides, we have also analyzed the problem of locating an undesirable median (maxian) on a network, obtaining a new and better upper bound. We have briefly presented the idea of a new algorithm in $O(mn)$ time to solve this problem.

Finally, we studied the uncenter and maxian problems on multicriteria networks, establishing new properties and rules to remove inefficient edges. We have also presented the multicriteria λ-anti-cent-dian model as a convex combination of the two latter problems through a parameter λ. An effective rule to remove edges containing inefficient points, as well as a polynomial algorithm in $O(m^2n^2k^3)$ time, being k the number of criteria. Besides, this procedure can solve both the

multicriteria uncenter problem and the multicriteria maxian problem. Moreover, when the network holds a single weight per node and a single length per edge, this algorithm can efficiently solve the single criterion uncenter, maxian and λ-anti-cent-dian problems. Lastly, this model might be slightly modified to generalize other models presented in the literature.

Some directions for future research could be:

- Try to apply the undesirable location problems to real world applications, or re-design them to acquire the real details that are not covered in the models. A direct use could be any application involving GIS (*Geographic Information System*) technologies.
- Compile in a single software application all the models described in this chapter, along with the classical algorithms for *desirable* facility location problems. A first attempt of this project was presented in Colebrook et al. (2005a).
- Expose all the algorithms developed so far as Web Services in the Internet, so they could be easily used from any computing device (PC, smartphone, tablet, etc). This is a nice project that we keep in mind a long time ago, and we hope to develop it shortly.

Acknowledgements This work has been partially supported by *Ministerio de Ciencia e Innovación*, Spanish Government, research projects MTM2010-18591 and MTM2009-08830.

References

Alumur S, Kara BY (2007) A new model for the hazardous waste location-routing problem. Comput Oper Res 34(5):1406–1423

Berman O, Drezner Z (2000) A note on the location of an obnoxious facility on a network. Euro J Oper Res 120(1):215–217

Berman O, Wang J (2004) Probabilistic location problems with discrete demand weights. Networks 44(1):47–57

Berman O, Wang J (2006) The 1-median and 1-antimedian problems with continuous probabilistic demand weights. INFOR 44(4):267–283

Berman O, Wang Q (2007) Locating semi-obnoxious facilities with expropriation: minisum criterion. J Oper Res Soc 58:378–390

Berman O, Wang Q (2008) Locating a semi-obnoxious facility with expropriation. Comput Oper Res 35(2):392–403

Berman O, Drezner Z, Wesolowsky GO (1996) Minimum covering criterion for obnoxious facility location on a network. Networks 28(1):1–5

Berman O, Huang R (2008) The minimum weighted covering location problem with distance constraints. Comput Oper Res 35(2):356–372

Berman O, Verter V, Kara B (2007) Designing emergency response networks for hazardous materials transportation. Comput Oper Res 34(5):1374–1388

Burkard RE, Dollani H (2003) Center problems with pos/neg weights on tree. Euro J Oper Res 145(3):483–495

Burkard RE, Dollani H, Lin Y, Rote G (2001) The obnoxious center problem on a tree. SIAM J Discrete Math 14(4):498–509

Cáceres T, Mesa JA, Ortega FA (2007) Locating waste pipelines to minimize their impact on marine environment. Euro J Oper Res 179(3):1143–1159
Cappanera P (1999) "A survey on obnoxious facility location problems", Technical Report 11, Dipartamento di Informatica, Università di Pisa
Cappanera P, Gallo G, Maffioli F (2003) Discrete facility location and routing of obnoxious activities. Discrete Appl Math 133(1–3):3–28
Carrizosa E, Conde E (2002) A fractional model for locating semi-desirable facilities on networks. Euro J Oper Res 136(1):67–80
Carrizosa E, Plastria F (1999) Location of semi-obnoxious facilities. Stud Locational Anal 12:1–27
Chan Y (2005) Location, transport and land-use: modelling spatial-temporal information. Springer, New York
Church RL, Garfinkel RS (1978) Locating an obnoxious facility on a network. Transport Sci 12(2):107–118
Colebrook M, Alonso S, Sicilia J (2005) "A new software tool to solve multicriteria facility location problems on networks" (invited session). In: Proceedings of the Triennial Conference of the International Federation of Operational Research Societies (IFORS 2005)
Colebrook M, Gutiérrez J, Sicilia J (2005b) A new bound and an O(*mn*) algorithm for the undesirable 1-median problem (maxian) on networks. Comput Oper Res 32(2):309–325
Colebrook M, Gutiérrez J, Alonso S, Sicilia J (2002) A new algorithm for the undesirable 1-center problem on networks. J Oper Res Soc 53(12):1357–1366
Colebrook M (2003) Desirable and undesirable single facility location on networks with multiple criteria. Ph.D. dissertation, Universidad de La Laguna, Tenerife, Spain
Colebrook M, Sicilia J (2006) An O(*mn*) algorithm for the anti-cent-dian problem. Appl Math Comput 183(1):350–364
Colebrook M, Sicilia J (2007) Undesirable facility location problems on multicriteria networks. Comput Oper Res 34(5):1491–1514
Current J, Min H, Schilling D (1990) Multiobjective analysis of facility location decisions. Eur J Oper Res 49(3):295–307
Daskin MS (1995) Network and discrete location: models, algorithms and applications. Wiley, New York
Domschke W, Drexl A (1985) Location and layout planning: an international bibliography. Springer, Berlin
Drezner Z (ed) (1995) Facility location: a survey of applications and methods. Springer, New York
Drezner T, Drezner Z, Scott CH (2009) Location of a facility minimizing nuisance to or from a planar network. Comput Oper Res 36(1):135–148
Drezner Z, Hamacher HW (eds) (2004) Facility location: applications and theory. Springer, New York
Drezner Z, Wesolowsky GO (1995) Obnoxious facility location in the interior of a planar network. J Regional Sci 35(4):675–688
Dyer ME (1984) Linear time algorithms for two- and three-variable linear programs. SIAM J Comput 13(1):31–45
Eiselt HA, Sandblom C-L (2004) Decision analysis, location models, and scheduling problems. Springer, New York
Eiselt HA, Marianov V (2011) Foundations of Location Analysis. Springer, New York
Erkut E, Alp O (2007) Designing a road network for hazardous materials shipments. Comput Oper Res 34(5):1389–1405
Erkut E, Neuman S (1989) Analytical models for locating undesirable facilities. Euro J Oper Res 40(3):275–291
Erkut E, Neuman S (1992) A multiobjective model for locating undesirable facilities. Ann Oper Res 40:209–227
Erkut E, Verter V (1995) Hazardous materials logistics. In: Drezner Z (ed) Facility locations: a survey of applications and methods. Springer, New York, pp 467–506
Farahani RZ, Hekmatfar M (2009) Facility location: concepts, models, algorithms and case studies. Springer, New York

Friedrich F (1929) Alfred Weber's theory of the location of industries. Chicago University Press, Chicago, IL
Giannikos I (1998) A multiobjective programming model for locating treatment sites and routing hazardous wastes. Euro J Oper Res 104(2):333–342
Hakimi SL (1964) Optimum locations of switching centers and the absolute centers and medians of a graph. Oper Res 12(3):450–459
Hale T (1998) Location science references. http://gator.dt.uh.edu/~halet/
Hamacher HW, Labbé M, Nickel S, Skriver AJV (2002) Multicriteria semi-obnoxious network location problems (MSNLP) with sum and center objectives. Ann Oper Res 110(1–4):33–53
Hansen P, Labbé M, Thisse J-F (1991) From the median to the generalized center. Recherche Operationnelle/Oper Res 25(1):73–86
Hershberger J (1989) Finding the upper envelope of n line segments in O(n log n) time. Inform Process Lett 33(4):169–174
Hosseini S, Esfahani AM (2009) Obnoxious Facility Location. In: Farahani RZ, Hekmatfar M (eds) Facility location: concepts, models, algorithms and case studies. Springer, New York
Hokkanen J, Salminen P (1997) Locating a waste treatment facility by multicriteria analysis. J Multi-Crit Decis Anal 6(3):175–184
Jordan C (1869) Sur les assamblages de lignes. Zeitschrift Reine Angew Math 70:185–190
Kincaid RK, Berger RT (1994) The maxminsum problem on trees. Location Sci 2(1):1–9
Krarup J, Pruzan PM (1990) Ingredients of locational analysis. In: Mirchandani PB, Francis RL (eds) Discrete location theory. Wiley, New York, pp 1–54
Kuby MJ (1987) Programming models for facility dispersion: the p-dispersion and maxisum dispersion problems. Geogr Anal 19(4):315–329
Labbé M (1990) Location of an obnoxious facility on a network—a voting approach. Networks 20(2):197–207
List GF, Mirchandani PB (1991) An integrated network/planar multiobjective model for routing and siting for hazardous materials and wastes. Transport Sci 25(2):146–156
López-de-los-Mozos MC, Mesa JA (2001) The maximum absolute deviation measure in location problems on networks. Euro J Oper Res 135(1):184–194
Melachrinoudis E, Zhang FG (1999) An O(mn) algorithm for the 1-maxmin problem on a network. Comput Oper Res 26(9):849–869
Melachrinoudis E (2011) The location of undesirable facilities. In: Eiselt HA, Marianov V (eds) Foundations of location analysis. Springer, New York
Melachrinoudis E, Min H, Wu X (1995) A multiobjective model for the dynamic location of landfills. Location Sci 3(3):143–166
Minieka E (1983) Anticenters and antimedians of a network. Networks 13(3):359–364
Mirchandani PB, Francis RL (eds) (1990) Discrete location theory. Wiley, New York
Moon ID (1989) Maximin center of pendant vertices in a tree network. Transport Sci 23(3): 213–216
Moon ID, Chaudhry SS (1984) An analysis of network location problems with distance constraints. Manag Sci 30(3):290–307
Moreno-Pérez JA, Rodríguez-Martín I (1999) Anti-cent-dian on networks. Stud Locational Anal 12:29–39
Murray AT, Church RL, Gerrard RA, Tsui WS (1998) Impact models for siting undesirable facilities. Papers Reg Sci 77(1):19–36
Nickel S, Puerto J (2005) Location theory: a unified approach, Springer, Berlin
Pareto V (1896) Cours d'economie politique, F. Rouge, Lausanne 1
Plastria F (1996) Optimal location of undesirable facilities: a selective overview. Belgian J Oper Res Stat Comput Sci 36(2–3):109–127
Puerto J (ed) (1996) Lecturas en teoría de localización. Universidad de Sevilla, Secretariado de Publicaciones
Rahman M, Kuby M (1995) A multiobjective model for locating solid waste transfer facilities using an empirical opposition function. INFOR 33(1):34–49

Rakas J, Teodorović D, Kim T (2004) Multi-objective modeling for determining location of undesirable facilities. Transport Res Part D Transport Environ 9(2):125–138

Ratick SJ, White AL (1988) A risk-sharing model for locating noxious facilities. Environ Plann B 15(2):165–179

ReVelle CS, Cohon JL, Shobrys D (1981a) Multiple objective facility location. Sistemi Urbani 3:319–343

ReVelle CS, Cohon JL, Shobrys D (1981b) Multiple objectives in facility location: a review. Lecture notes in economic and mathematical systems, organisations: multiple agents with multiple criteria, vol 190. Springer, pp 321–337

Ross GT, Soland RM (1980) A multicriteria approach to the location of public facilities. Eur J Oper Res 4(5):307–321

Saameño JJ (1992) Localización multicriterio de centros peligrosos. Ph.D. thesis, Universidad de Sevilla

Salhi S, Welch SB, Cunninghame-Green RA (2000) An enhancement of an analytical approach: the case of the weighted maximin network location problem. Math Algorith 1(4):315–329

Skriver AJV (2001) Multicriteria analysis on network and location problems. Ph.D. thesis, University of Aarhus

Skriver AJV, Andersen KA (2003) The bicriterion semi-obnoxious location (BSL) problem solved by an ε-approximation. Euro J Oper Res 146(3):517–528

Slater P (1975) Maximin facility location. J Res Natl Bur Stand 79B(3–4):107–115

Steuer RE (1986) Multiple criteria optimization: theory, computation and application. Wiley, New York

Stowers CL, Palekar US (1993) Location models with routing considerations for a single obnoxious facility. Transport Sci 27(4):350–362

Tamir A (1988) Improved complexity bounds for center location problems on networks by using dynamic data structures. SIAM J Discrete Math 1(3):377–396

Tamir A (1991) Obnoxious facility location on graphs. SIAM J Discrete Math 4(4):550–567

Tamir A (2001) "Comment on E. Melachrinoudis and F.G.-S. Zhang, An O(mn) algorithm for the 1-maximin problem on a network, Computers & Operations Research 26 (1999) 849–869, Comput Oper Res 28(2), 189

Ting SS (1984) A linear-time algorithm for maxisum facility location on tree networks. Transport Sci 18(1):76–84

Tuzkaya G, Önüt S, Tuzkaya UR, Gülsün B (2008) An analytic network process approach for locating undesirable facilities: an example from Istanbul, Turkey. J Environ Manage 88(4): 970–983

Verter V, Erkut E (1995) Hazardous materials logistics: an annotated bibliography. In: Haurie A, Carraro C (eds) Operations Research and Environmental Management. Kluwer, Dordecht, pp 221–267

Yamaguchi K (2011) Location of an undesirable facility on a network: a bargaining approach. Math Soc Sci 62(2):104–108

Zemel E (1984) An O(n) algorithm for the linear multiple choice knapsack and related problems. Informat Process Lett 18(3):123–128

Zhang FG (1996) Location on networks with multiple criteria. Ph.D. thesis, Northeastern University, Boston, MA

Zhang FG, Melachrinoudis E (2001) The maximin-maxisum network location problem. Comput Optimiz Appll 19(2):209–234

Zhang J, Hodgson J, Erkut E (2000) Using GIS to assess the risks of hazardous materials transport in networks. Euro J Oper Res 121(2):316–329

Zhao J, Shuai B (2010) A new multi-objective model of location-allocation in emergency response network design for hazardous materials transportation. In: Proceedings of the 2010 IEEE International Conference on Emergency Management and Management Sciences (ICEMMS 2010), 246–249

Network Interdiction Methods and Approximations in a Hazmat Transportation Setting

Justin Yates

Introduction

The United States transportation system is an extensive and integrated component in the eight key infrastructures upon which the livelihood of the U.S. is dependent (Department of Homeland Security 2009). The accessibility and mobility enabled through open use of the transportation system is a vital and necessary freedom which contributes to the fluidity of the American environment. The transportation system is expansive and heavily utilized with an average of over 2 billion daily vehicle-miles of travel (nearly twice as much travel since the early 1980s) on the roughly 4 million miles of paved roadway, nearly 47,000 miles of Interstate highway, 600,000 bridges and 366 U.S. highway tunnels over 100 m (Texas Transportation Institute 2011; Transportation Security 2012). Travelers and shippers may also choose to utilize more than 300,000 miles of freight rail, nearly 10,000 miles of urban and commuter rail systems, or connect between 500 commercial-service and 14,000 general aviation airports (Transportation Research 2002).

In this chapter, a general review of network-based hazardous material transportation models will be given. Specific attention will be given to the network interdiction model and its variants (e.g. shortest path network interdiction) as these models have recently become popular in the domain of homeland security. The chapter will focus on the application of network interdiction models to networks of various size and structure and the ensuing computational performance (including objective value, sensitivity to network properties, etc.) and spatial structure (e.g. resource allocation, network connectivity/density) of the interdiction solutions. A systematic experimental analysis will be designed to identify salient network and regional properties impacting interdiction solutions (e.g. proximity to origin points, initial arc

J. Yates (✉)
Industrial and Systems Engineering, Texas A&M University, College Station, TX, USA
e-mail: jtyates@tamu.edu

R. Batta and C. Kwon (eds.), *Handbook of OR/MS Models in Hazardous Materials Transportation*, International Series in Operations Research & Management Science 193,
DOI 10.1007/978-1-4614-6794-6_7,

metric values, etc.). Current approximations for the network interdiction model will also be analyzed against the obtained solutions, including alternate approximation formulations as well as alternate solution approaches. The evaluation of such techniques will lead to greater insight on the effect of network and problem structure to resource allocation in interdiction models.

Literature Review

This section will provide some supporting background in past and current hazardous materials transportation research. A history and survey of past research initiatives will be followed by identification of current research threads in network-based infrastructure protection whose roots and foundation can be traced back to the hazardous material literature. Attention will be given to the quantification of risk, potential pitfalls, and benefits/drawbacks of estimation as a tool to measure risk. Discussion and justification for a selection of related mathematical models such as the Vehicle Routing Problem with Time Windows, Discrete Fractional Programming, and Shortest Path Network Interdiction will be provided in addition to some brief detail on algorithm/heuristic modeling within hazardous materials transportation problems. In addition to optimization, this section will introduce past practices in the field of Geographic Information Science (GIS) geared towards supporting and augmenting risk analysis, routing and scheduling problems through spatial reasoning methods. This section will conclude with a detailed discussion on the Network Interdiction problem, which is used as the test-bed formulation throughout the remainder of this chapter.

In 2001, there were 41,527 active hazmat motor carriers in the United States driving an average of 800,000 truck shipments per day of hazardous materials (hazmat) over the nation's roadways (Field 2004). By 2011, the number of active hazmat motor carriers has grown to 61,000, transporting over 2 billion tons of hazmat annually (Transportation Security 2012). Similarly in 2011, there were 5.76 million hazmat inspections carried out by the U.S. Department of Transportation with 3.75 million vehicle inspections. In approximately 19,400 of 740,000 inspection cases on interstate and hazmat certified carriers, unsafe or fatigued driving conditions were reported (Transportation Security 2012).

The commingling of commercial, personal, and hazmat travel has fueled an emphasis on safety in the transportation industry, not merely from the perspective of individual harm, but also the durability and maintenance of the integrity and serviceability of the transportation systems themselves (as an example, the U.S. Department of Transportation (DOT) publishes a biennial report on the status of hazmat transportation) (National Highway 1996; US Department of Transportation 2006). Additionally, academic researchers have heightened focus on the problem of increasing safety and, more recently, security of hazmat shipments, especially through populated areas or near perceived targets (i.e. nuclear power plants, water resource plants, etc.).

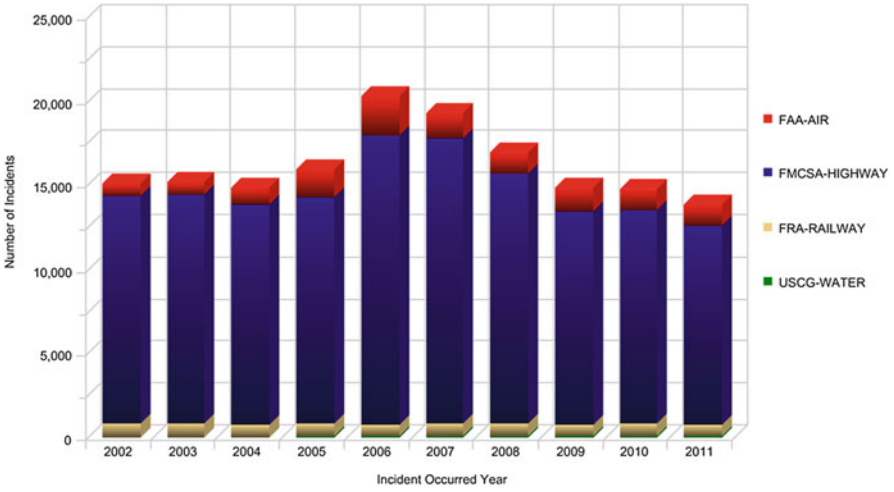


Fig. 1 All incidents by mode and incident year (US Department of Transportation 2012)

Figure 1 and Tables 1, 2, 3, and 4 illustrate the frequency of incidents occurring during hazmat transportation by air, highway, railway and waterway. Hazardous material transportation incidents in 2011 resulted in over $100 million in damages with a 10-year cumulative total of over $670 million. Additionally, the size and magnitude of hazmat transportation across the U.S. and the attractiveness of its cargo (which could be used by both domestic and international organizations to create situations of intentional public exposure or weaponization) generates support for an immense number of research opportunities geared towards creating safer, more stable and less vulnerable hazmat transportation.

A History in Hazardous Materials Transportation Research

The concept of risk, and its quantification, however ambiguous, has been the driving force behind many popular models related to infrastructure protection, transportation and, in recent years, homeland security. Risk in hazardous materials transportation was most succinctly measured as the product of incident probability and incident consequence. Incident probability implies occurrence of an accident that releases or exposes a region/population to hazardous material while incident consequence quantitatively measures the impact(s) of release. Ideally, these probabilities would be based on real-world data and historical statistics, leaving little room for interpretation. In practice, there are many pitfalls in quantifying risk, including, the lack of accurate and specific historical data and lack of a clear and agreed upon definition of risk head this list. With respect to data collection, a true calculation of risk would include meteorological and topological knowledge, accurate effects

Table 1 Incidents by mode and incident year

Mode of transport	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	Grand total
FAA-AIR	732	750	993	1,654	2,406	1,556	1,278	1,356	1,293	1,240	13, 258
FMCSA-HIGHWAY	13,502	13,594	13,068	13,460	17,159	16,930	14,805	12,730	12,645	11,857	139, 750
FRA-RAILWAY	870	802	765	745	703	753	748	642	751	692	7, 471
USCG-WATER	10	10	17	69	68	61	99	90	105	63	592
Grand total	15,114	15,156	14,843	15,928	20,336	19,300	16,930	14,818	14,794	13,852	161, 071

Table 2 Fatalities by mode and incident year

Mode of transport	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	Grand total
FAA-AIR	0	0	0	0	0	0	0	0	0	0	0
FMCSA-HIGHWAY	9	15	11	24	6	9	6	11	8	10	109
FRA-RAILWAY	1	0	3	10	0	0	1	1	0	0	16
USCG-WATER	0	0	0	0	0	0	3	0	0	0	3
Grand total	10	15	14	34	6	9	10	12	8	10	128

Table 3 Injuries by mode and incident year

Mode of transport	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	Grand total
FAA-AIR	4	1	11	44	2	8	7	10	2	7	96
FMCSA-HIGHWAY	118	105	155	178	192	160	153	153	153	109	1,476
FRA-RAILWAY	14	13	122	693	25	57	63	38	13	20	1,058
USCG-WATER	0	0	0	0	15	3	0	0	2	8	28
Grand total	136	119	288	915	234	228	223	201	170	144	2,658

Table 4 Damages by mode and incident year

2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	Grand total
$108,630	$100,483	$188,481	$198,316	$670,521	$88,030	$191,434	$707,939	$20,267	$171,467	$2,445,568
$48,075,528	$49,109,443	$47,157,765	$40,179,241	$59,502,295	$47,279,979	$42,889,714	$50,628,781	$63,841,312	$96,629,446	$545,293,504
$9,745,140	$4,126,165	$13,901,020	$15,454,556	$10,739,810	$27,305,219	$7,939,038	$17,557,034	$17,358,060	$7,013,438	$121,139,480
$247,802	$261,324	$1,654,664	$114,000	$58,828	$19,097	$138,350	$100,887	$574,103	$205,000	$3,374,055
$58,177,100	$53,597,415	$62,901,930	$55,946,113	$70,971,454	$74,692,325	$51,158,536	$68,994,641	$71,793,742	$104,019,351	$672,252,607

and dispersion of the spilled substance, location of individuals at the time of the incident, and human does response, to name a few. Note that this list does not begin to include components related to economic and environmental damage, as well as social and socio-economic implications which may be desired in quantifying incident consequence) (List et al. 1991; Erkut and Verter 1998). This impracticality has led hazmat modelers to adopt estimation metrics that avoid such extensive, time consuming data collection, with varying degrees of complexity and success.

The ultimate goal of risk measurement is to utilize estimation techniques that drive the optimization procedure and accurately replicate real-world scenarios without obtaining an overwhelming amount of data. List et al. (1991) refers to this as constructed risk or a constructed index. A constructed index, in its most simplistic form, decomposes the network, examining its individual arcs, assigning a pseudo cost to each arc, and implementing algorithms such as Yen's shortest path algorithm in an effort to succinctly yet accurately represent the true environment. Here, (1) illustrates the risk function, where R_{AB} is the risk (which would be used instead of link length as the cost in a shortest path algorithm) for link AB, p_{AB} is the probability of an incident occurring on link AB, and C_{AB} is the consequence for AB, which is nearly always contingent (either partially or exclusively) on the population density within a certain vicinity of the road segment (Erkut and Verter 1998).

$$R_{AB} = p_{AB} C_{AB} \tag{1}$$

Assisting in the determination of p_{AB}, incident probability studies examining variations in release rate by mode, carrier type, vehicle type, road classification, time of day and weather conditions may be used (List et al. 1991). Estimation tools for incident consequence typically take on the form of a "danger circle" (Erkut and Verter 1998) or "buffer zone" (Laefer and Pradhan 2006; Huang et al. 2004) such that all individuals within the zone are determined to be exposed to a fatal hazmat incident. It is important to note that, depending on the time of day and area, population estimates may vary significantly from static figures such as census counts (Erkut and Verter 1998). Total edge consequence may be derived under the assumption that an edge is composed of n unit segments, each with uniform parameters (Erkut and Verter 1998). Expected edge consequence may then be defined by (2) (variable interpretation is the same as above) and, since p (probability of a hazmat incident) is typically very small (on the order of one per one million miles), approximated as in (3) (Erkut and Verter 1998).

$$pC + (1-p)pC + (1-p)^2 pC + \ldots + (1-p)^{n-1} pC \tag{2}$$

$$(pn)C \tag{3}$$

It is this value that would be substituted in the objective function of a shortest path problem, creating an optimization problem that would return the path of minimum risk from an origin-destination (O-D) pair given the current network.

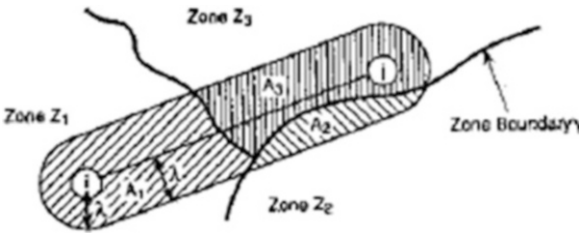


Fig. 2 Zone segmentation along a link

While this approach is eloquent in its relative simplicity, key assumptions reduce the realism of the output. Primarily, the assumption that each unit segment of an edge has uniform properties is limiting. Additional nodes that separate a link without uniform properties into its uniform components may be added, without influencing the optimal outcome. However, this approach is impractical for large networks and also increases the number of constraints in the optimization problem, potentially increasing computation time significantly. If this assumption cannot be made, then expected edge consequence may not be approximated as succinctly as in (3), preventing the shortest path approach (Erkut and Verter 1998).

In practice, multiple trips are necessary to effectively move material, and consideration of consequence over numerous shipments between an O-D pair is necessary to accurately reflect the repercussions of an incident. Viewing these shipments as a sequence of independent Bernoulli trials was discussed in detail by Jin et al. (1996) and Jin and Batta (1996) and continued in Batta and Chiu (1998). Underlying these works is the observation that multiple, but finite, hazmat trips are often needed to transport all of the material. Total shipments may be unrestricted (continuing until all material is shipped), or may be suspended or ceased after a critical threshold on the number of accidents is reached (Jin et al. 1996). Probability of link incident (p_i) and consequence of link incident (C_i) remain, while introduction of the variable t (threshold number of accidents) and T (total trips to be made) allow for new objective considerations (Jin et al. 1996). Varying the values of t and T, alternative objectives such as expected total consequence, expected consequence per trip, and expected number of trips between two successive accidents are considered (Jin and Batta 1996).

Risk equity, described as the fair dispersion of risk throughout a population, represents yet another way that hazmat transportation has been viewed and modeled. The objective function in a risk equity problem is to find a set T of routes (not necessarily distinct) that minimize total risk over a network/region while constraining the difference in total risk between every pair of zones within a specified threshold T_μ (Gopalan et al. 1990). Figure 2 illustrates a typical instance of network segmentation, where link (i, j) directly spans two zones and indirectly influences a third (Gopalan et al. 1990).

Instead of viewing the incident and its consequence separately, multi-criterion optimization problems can be formulated to consider individual hazmat transportation problem components individually within a system-optimized mathematical

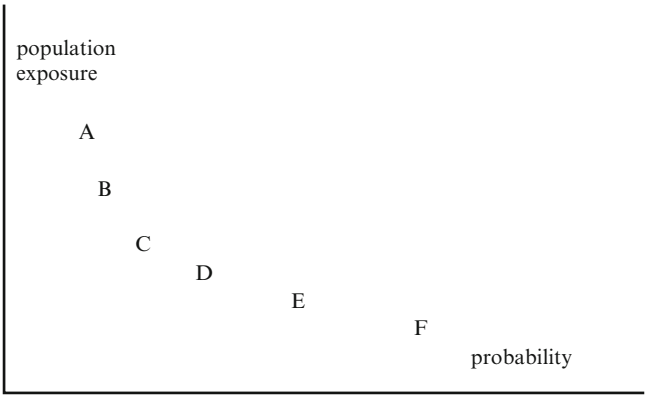


Fig. 3 Efficient frontier of a typical bi-criterion minimization problem

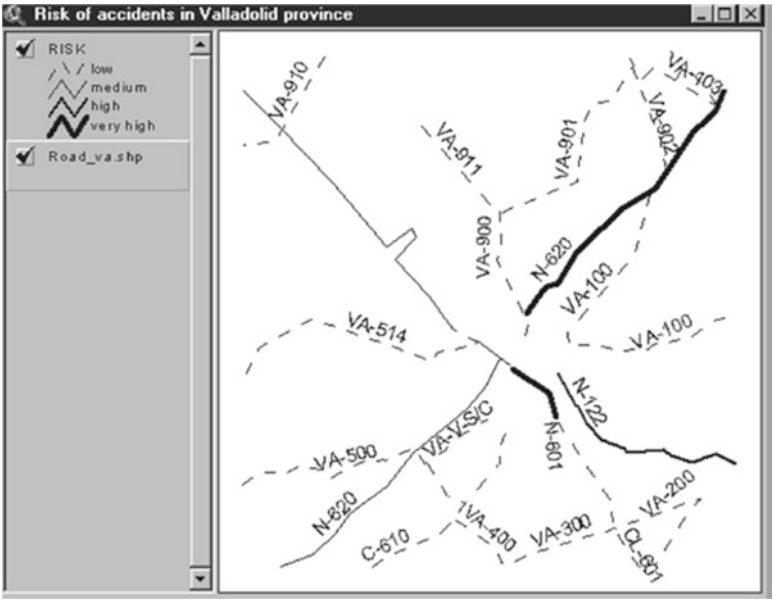


Fig. 4 Risk map for aniline transportation in Valladolid

model. Figure 3 shows a typical bi-criterion efficient frontier for two factors (incident probability and population exposure), with each letter representing a different optimal solution on the efficient frontier generated applying different weights to the objective function criteria (Erkut and Verter 1998). Huang et al. (2004) extends this approach, identifying five criteria (exposure, socio-economic impact, risk of hijack, traffic conditions, and emergency response) of potential interest in hazmat route choice (Fig. 4).

Shifting from the macro-scale view of equity and multiple criteria, approaches focusing on individual network components at a micro scale emerged as a natural complement. Arc vulnerability modeling considers extrinsic, tangible elements that lie in the vicinity of the link (i.e. hospitals, education centers, sports facilities, shopping centers, power stations, water treatment centers, etc.) to quantify an individual arc metric which is used to formulate the given optimization problem. The concept may be enhanced by incorporating non-vicinity related tangibles such as network redundancy, capacity, traffic demand and highway configuration, which may positively or negatively influence the importance of a link (Cova and Conger 2003). Also, natural disasters such as earthquakes, may cause fires, landslides, and facilitate dam failures (and consequently floods), all possible contributors to network disconnection and transportation disruption. The flexibility such approaches gave to modelers has made them strong favorites and many contemporary homeland security models can be traced back to these formulations.

Considering, for example, least-flood-risk as an additional criteria for routing in a multi-criterion model, link cost could be quantified as in (4), where the denominator models flood characteristics ($\alpha_h \in [0, 1]$ representing flood height and $\alpha_v \in [0, 1]$ representing flood velocity) such that if no flooding is present, the cost is simply the numerator (in this case, the length of link (i, j)) (Cova and Conger 2003).

$$c_{ij} = \frac{L_{ij}}{\alpha_h \alpha_v} \tag{4}$$

Simulation software may also be used to help quantify link vulnerability, especially in the area of natural disasters. The Federal Emergency Management Agency (FEMA) has created a software tool named HAZUS, which may be used to observe how transportation networks react to the adverse affects of natural disasters (Federal Emergency Management 2006). Similarly, the Federal Highway Administration recognizes REDARS, used exclusively to determine how seismic events impact road and highway systems (MCEER 2006).

Network Models and Interdiction

Network stability and the maintenance of serviceability have already been shown to be of great concern when considering the routing of hazmat. The functioning of links in a network may be negatively influenced by congestion and accidents, weather, seismic activity and natural disasters, the occurrence of hazmat incidents, intentional acts to disrupt the network, or by the combination of any two or more such instances.

One way to identify network vulnerability is through the identification of critical links. A link is deemed "weak" if incident probability is high, "important" if consequence of an incident is large, and "critical" if it is both weak and important (Jenelius et al. 2006). These values are derived through the observation of how

link absence affects path travel time, using multiple, predetermined, O-D paths to evaluate (Jenelius et al. 2006). Additionally, incorporation of key-infrastructure (i.e. proximity of the link to schools, hospitals, military, power or water facilities) could assist in more realistic modeling, where such vulnerable sites (potential targets) contribute to the link's importance (Luedtke and White 2002). Criticality of links is also dependent on the geographic features over which the network lies (i.e. coastline, mountain ridge). These features hinder the existence of nearby options available for re-routing without significant time delay. Therefore, it is useful to consider inherent network vulnerability through quantitative means and incorporate this into a risk or routing model. The generality of Jenelius et al. (2006), is applied to optimization of hazmat transportation, developing a mathematical model called the Hazardous-Network Design Problem (HDP).

Given a road network, HDP selects links that should be closed to hazmat transportation in order to minimize total risk (Kara and Verter 2004). The formulation of the HDP is bi-level, containing an outer and inner problem that more accurately represents the interaction between policy makers and hazmat carriers (there is often predominance in hazmat routing problems favoring the carriers' viewpoint (i.e. routing) and omitting regulator decision-making pertaining to link availability). The inner problem minimizes the combined travel distance of the trucks subject to flow conservation, and may be viewed as either a minimum cost network flow or a constrained shortest path problem, while the solution of the outer problem minimizes population exposure. The two problems interact with the binary decision variables of the outer problem becoming the parameters of the inner problem (Kara and Verter 2004). Success of the model prescribes the available road network and route choices for hazmat transportation.

The term interdict is defined by Merriam-Webster's dictionary as the adjective "to destroy, damage, or cut off (as an enemy line of supply) by firepower to stop or hamper an enemy." The problem of network interdiction may then be taken to mean the intentional destruction, by force, of a network to impede or cease enemy use. Within the realm of optimization, especially in the military community, the study of interdiction problems has been given significant attention.

Considering network flow, the interdiction problem may be represented as a multi-commodity problem with two players (Lim and Smith 2007). The first player, the follower, makes profit by delivering commodities to designated destinations. The leader attempts to minimize the followers profit by selectively destroying arcs, the destruction of which costs the leader by subtracting a link destruction amount from the leaders' interdiction budget. Arcs may either be destroyed discretely (either capacity flow is possible or no flow is possible) or continuously (partial flow over links is allowed). The Multi-Commodity Flow Network Interdiction Problem (MFNIP) is then modeled as the minimization of the maximum profit the follower may achieve, subject to conservation of flow constraints, the leaders budget constraint, and non-negativity (Lim and Smith 2007). From the followers' perspective, MFNIP quantifies a worst-case scenario that showcases the weakest (or most vulnerable) links in a network. This information may then be used in the strengthening of weak points or the enhancement of network connectivity to

better secure the system (Qiao et al. 2007). In addition to transportation networks, the multi-commodity flow problem may be applied to airline operations, supply chains, and telecommunications, as well as water supply networks, and power grid systems, all of which may then be examined through the MFNIP formulation to gauge network security and stability.

A review of early literature on the routing, schedule, location and risk analysis for hazardous materials transportation can be found in List et al. (1991). Specifically associated with risk analysis, Erkut and Verter (1998) provides model overviews and examines how the quantification of risk in these models affects model performance/accuracy. In the case where risk is taken to be the acceptable threshold of accidents in transport willing to be endured, Jin and Batta (1996) gives a nonlinear constrained shortest path approach and examines the effect of the accident threshold on routing decisions.

Beginning with Wood (1993) and continuing through Israeli and Wood (2002), Brown et al. (2006) and others, network interdiction began to emerge as a natural extension to hazardous materials transportation research. The problems are decidedly similar and deal with undesirable transportation through a network. In the hazardous materials transportation problems, risk and exposure were two of the quantifiable measures applied to each arc or node of the network and used as the basis for determining appropriate route selections. In the case of Israeli and Wood (2002) and Brown et al. (2006), the quantifiable arc measure is length, which is increased when an arc is interdicted. In these two-player network interdiction problems, these elongated arcs effectively deter the opposing player from using these arcs in composing their shortest path. Scaparra and Church (2008) and Church and Scaparra (2007) model interdiction at network nodes with interdiction removing the ability of a facility located at that specific node from satisfying demand to other nodes. Considering the minimization of weighted demand-distance as the objective, an optimal nodal interdiction strategy will increase the opposing player's cost to satisfy network demand.

As these and other recent mathematical models transition from an emphasis on risk assessment and hazardous materials transportation to problems of homeland security and extreme events, five major factors can be used to delineate and differentiate model focus, intent and capability. The five major factors offered in Yates and Sanjeevi (2012) are formulation, objective function, interdiction metric, component interdiction and the Origin–destination policy. Formulation refers to single versus bi/multi-level models. Objective function details whether the original objective function provided for a given problem is additive or multiplicative/probabilistic (note that this is the original objective function and does not refer to any transformations applied during the solution process). Interdiction metric refers to whether the individual metrics used are continuously or discretely interdicted while Component Interdiction dictates whether these metrics are arc-based, node-based, or network/spatially based. Lastly, Origin–destination policy refers to the existence of a single O-D pair or multiple O-D pairs in the problem. Table 5 is now introduced to provide an overview of some representative recent

Table 5 Summary of interdiction-related problem properties (Yates and Sanjeevi 2012)

	Formulation		Objective function		Interdiction metric			Component interdiction		Origin–destination	
	Single level	Multi-level	Summation	Product	>0	Z+	[0,1]	Arc	Node	Single	Multiple
Wein and Atkinson (2007)	x			x		Layers		N/A		N/A	
Israeli and Wood (2002)		x	x		Length			x		x	
Brown et al. (2006)		x	x		Length/value			x			x
Morton et al. (2007)		x		x			Detection	x		Probabilistic	
Scaparra and Church (2008)		x	x		Weighted distance				x		x
Church et al. (2004)	x		x		Weighted distance				x		x
Church and Scaparra (2007)		x	x		Weighted distance				x		x
Murray et al. (2007)	x		x		Flow			Connectivity			x
Matisziw et al. (2007)	x		x		Flow			Connectivity			x
Matisziw and Murray (2009)	x		x		Flow			Connectivity			x
Grubesic and Murray (2005)	x		x		Demand				x		x
Southworth (2009)	x		x		Population exposure			Spatial		No origins/destinations	

network interdiction literature published after 2004. For additional discussion of network interdiction problems published prior to 2004, refer to Church et al. (2004).

At its core, the shortest path network interdiction problem (SPNIP) is a two-player deterministic game being played over a network composed of arcs and nodes with a given arc/node metric (length, detection probability, etc.) and with identified origin and destination sets. In these models, any and all of these components could be completely known by both players (i.e. having perfect information) or could contain some element of mis-information, deception (imperfect information). The ability to interdict is also constrained by limited resources, which could be modeled as a finite number of arcs/nodes to interdict or a budget limitation where each interdiction comes with some associated interdiction cost.

Developed interdiction models can have objective functions that are a single level or multiple levels to reflect to degree of interaction and knowledge among the players being modeled. Examples of single-level models include Church et al. (2004) and Church and Scaparra (2007). In this case, these models are variants of traditional optimization models such as the p-Median and Maximal Covering problems that have been adapted to include interdiction concepts. In many instances, these single-layer models are solved for a variety of problem parameters and threshold values to determine a pareto front, or set of interdiction strategies to better gain situational knowledge. Such analysis is extremely useful when the capability/intent of an adversary is in question or when information is unreliable/imperfect.

Multi-level models (Morton et al. 2007; Israeli and Wood 2002; Brown et al. 2006; Church and Scaparra 2007), in contrast to single-level models, are often integer or mixed integer programs that model the decision making of players sequentially in the same formulation. Instead of solving under multiple parameter and threshold instances, the interaction between the interdictor and the defender is modeled simultaneously. The objective function in multi-level models is one that typically reflects pure competition, with the interdictor seeking to minimize an overall network metric (such as flow or satisfied demand) and the defender seeking to maximize this minimum metric. In other words, the defender's job is to minimize the effect of interdiction on their network operations. In Table 1, ">0" indicates that the interdiction metric is continuous (i.e. interdiction increases arc length by x with $x > 0$), "Z^+" indicates the metric is integral (i.e. interdiction is based upon the number of layers penetrated) and "[0, 1]" indicates that the metric is probabilistic. In the probabilistic case, interdiction can be modeled as the probability of path detection P where $P = \prod_i a_i x_i$, $i \varepsilon P$ models path probability as the product of all arc probabilities a_i on path P ($x_i = 1$ if $i \varepsilon P$, 0 otherwise).

Multi-level models, due to their structure, can often be decomposed and solved iteratively to optimality using standard optimization techniques. One of the most straight-forward and intuitive of these solution approaches is Bender's Decomposition (Bard 1998), where the interdictor and defender "trade" moves and with each move providing some degree of information to their counterpart (recall that these problems can be set up with perfect or imperfect information). In this way, information from these consecutive player movements is continually accrued and used within the next iteration of the decomposition. At some point, the interdictor

and defender reach a state of equilibrium where no new strategies are employed. In the worst case, this equilibrium occurs after all possible moves have been explored (i.e. complete enumeration), though in practice significantly less iterations are required. This certificate of optimality, in conjunction with its intuitive approach, is a major benefit to using Bender's Decomposition (Bard 1998).

Formulating Interdiction Models

In this section, we begin to explore the shortest path network interdiction problem (SPNIP) as defined by Israeli and Wood (2002) and discuss multiple variations which can be derived from it. We will begin examining solutions to the SPNIP and its variations by looking at their computational performance when solved using Bender's Decomposition, implementation of which will also be addressed. As patterns and trends emerge in the solutions, we will begin to motivate development of alternative heuristic and approximation techniques to solve network interdiction problems. These techniques will be discussed and compared in "Developing interdiction approximations and heuristics" of this chapter.

Mathematical Models and Notation

SPNIP and SPNIP-M

We begin by presenting the SPNIP formulation of Israeli and Wood (2002) and a modified shortest path network interdiction problem (SPNIP-M) formulation of Yates and Casas (2010). Each is a discrete bi-level optimization problem with an attacker and defender. The attacker considers all identified origins and targets and uses the network to find the path with, in this case, lowest detection probability (note that many network measures such as distance or cost, could be used in place of detection). The defender locates a limited number of resources which increase arc (and subsequently path) detection probabilities. Through the remainder of this work, we will refer to the defense resources as sensors, though this term is used relatively loosely. In our modeling, a sensor's properties include a predefined range (beyond which their influence is considered to be null) as well as an associated location cost and a parameter for sensor strength.

As a point of delineation, we note that the formulation of SPNIP assumes that sensors are located directly on network arcs in a 1-to-1 fashion. SPNIP-M, on the other hand, locates sensors geographically within the region at pre-specified locations. These locations, referred to as atoms, are point locations within the continuous region containing the network/infrastructure of interest. The atom set containing all possible sensor locations for a given problem is determined through a number of different methods which can include set distances, line-of-sight, proximity, or a function containing any combination of measurable geographic

and network properties. For our initial models, we assume a simple and uniform grid pattern for atoms, though section "Developing Interdiction Approximations and Heuristics" will discuss how more intelligent atom derivations can be derived and implemented within SPNIP-M. Overall, the geographic structure of SPNIP-M increases the complexity and realism of sensor location in the interdiction model, giving the modeler more flexibility.

Terminology

Atom: Potential sensor location point within the geographic region occupied by the network.

Attacker: Seeks the path of lowest detection (i.e. shortest path) on the network. The obtained path is simple and complete and will consider all possible origin and destination pairs (previously referred to as the follower).

Defender: Allocates sensors to increase detection probability. Sensors may be located directly on the network arcs in SPNIP or at designated geographic locations (atoms) in SPNIP-M and SPNIP-LB.

Detection: The probability that movement along a given arc (path) will be observed.

Sensor: Increases detection on arcs which fall within its given range. The degree to which detection is increased depends upon the sensor's power. Sensors are placed directly on arcs in the SPNIP and at designated geographic locations (atoms) in SPNIP-M and SPNIP-LB. Sensors have a known allocation cost.

Notation

A	Set of suitable sensor locations	c_s	Cost to locate a type s sensor
Λ	Set of network arcs	k_{ni}	$\{1, -1\}$ if node $n \varepsilon N$ is the $\{head, tail\}$ or arc $i \varepsilon \Lambda$, *else 0*
B	Total defense budget	η_s	Sensitivity of sensor type s, with $0 \leq \eta_s \leq 1$
N	Set of network nodes	$q_s - \{1, -1, 0\}$ if node $n \varepsilon N$ is $\{origin, target, intermediate\}$	
S	Set of sensor types	$r^{as}(i)$	1 if arc $i \varepsilon R^{as}$, else 0
τ	Overlapping coverage threshold	u_{ist}	Probability of non-detection for i covered by t type s sensors
R^{as}	Set of arcs within the influence range of a type s sensor located at atom $a \varepsilon A$		

Decision Variables

$w_i = 1$ if arc i is used in the attacker path, else 0.
$y_{as} = 1$ if sensor type s is located at atom a, else 0.
$x_{ist} = 1$if arc i is covered by t type s sensors, else 0.

[SPNIP]

$$z = \min \max \prod_{i,s,t} u_{ist}^{w_i y_{at}}$$

$$s.t. \quad \sum_i k_{ni} w_i = q_n \quad \forall$$

$$\sum_s y_{is} = 1 \quad \forall\, i$$

$$\sum_{i,s} c_s y_{is} \leq B$$

$$w, x, y \in \{0,\ 1\}$$

[SPNIP-M]

$$z^m = \min \max \prod_{i,s,t} u_{ist}^{w_i x_{ist}}$$

$$s.t. \quad \sum_i k_{ni} w_i = q_n \quad \forall\, n$$

$$x_{ist} - \tfrac{1}{t} \sum_a r^{as}(i) * y_{as} \leq 0 \quad \forall\, i, s, t$$

$$\sum_{s,t} x_{ist} = 1 \quad \forall\, i$$

$$\sum_{a,s} c_s y_{as} \leq B$$

$$w, x, y \in \{0,\ 1\}$$

Regardless of the formulation, we assume a non-zero detection probability for all arcs as an attacker can never realistically be guaranteed to reach his/her target. When dealing with network-based transportation, this detection probability, albeit potentially small, can be attributed to incidental traffic violations, accidents with other motorists, or a concerned citizen alerting local authorities to a suspicious individual or vehicle. We assume that these initial arc non-detection probabilities are known. We also assume that detection is equivalent to capture as a simplistic proxy for the more complicated case where detection and interception (i.e., capture) are separate factors.

As the SPNIP and SPNIP-M formulations show, the objective function yields a path detection probability calculated by multiplying arc non-detection values for all arcs comprising the optimal attacker simple path through the network (i.e. one that begins at a designated origin, terminates at a designated target and does not cycle) given the defender's optimal sensor location strategy. We calculate the impact of a sensor's coverage as $u_{ist} = u_{i01} \prod_t \eta_s$ where η_s indicates the sensor's strength. In SPNIP-M, there is a maximum threshold of coverage, τ, beyond which an arc's non-detection probability will not be affected by additional sensor coverage. In this way, all u_{ist} values may be calculated a-priori.

In both models, the first set of constraints imposes a conservation of flow within paths that the attacker identifies and is the only constraint which includes the attacker decision variable w_i. The second constraint set in SPNIP-M does not appear in the SPNIP model and functions as a relational constraint between sensor location and the corresponding arc influence upon the network. Simply stated, an arc cannot be influenced by t type s sensors unless the defender has allocated t type s sensors containing arc i in their respective ranges. The remaining constraints in both models

guarantee arc coverage (every arc is either covered, $x_{i1t} = 1$, or not covered, $x_{i01} = 1$) and limit the defender's available resources. All variables are modeled as binary decision variables.

SPNIP-LB

In both SPNIP and SPNIP-M, an arc is considered as covered by a sensor when any portion of that arc, no matter how large or small, falls within the sensor's range. Using this type of binary approach to coverage is highly restrictive and, one can argue, does not accurately reflect real-world sensor performance. Functions exist which define this type of behavior and have been used in past military models, where longer time spent in the range of enemy radar functionally increased one's detection probability (Przemieniecki 2000). Using a similar functional approach, we define a length-based approach to shortest path network interdiction (SPNIP-LB) to augment the binary SPNIP and SPNIP-M formulations.

[Length Based]:

u_i = initial probability of non-detection for arc i

l_{ias} = length of arc i within the range of a type s resource located at atom a

v_{ias} = proportion of non-detection reduction when arc i is influenced by a type s resource at atom a

$= e^{-\eta_s l_{ias}}$ where $\eta_s \geq 0 \quad \forall\, s \in S$

SPNIP-LB

$$\min_{y,x} \max_{w} \prod_{s,a,i} u_i^{w_i} v_{ias}^{w_i x_{ias}}$$

$$\text{s.t.} \quad \sum_i k_{ni} w_i = q_n \qquad \forall \quad n \in N$$

$$x_{ias} - r_i^{as} y_{as} \leq 0 \qquad \forall \quad i \in I, \quad a \in A, \quad s \in S$$

$$\sum_a \sum_s x_{ias} \leq 1 \qquad \forall \quad i \in I$$

$$\sum_s \sum_a c_s y_{as} \leq B$$

$$w, x, y \in \{0,\ 1\}$$

The SPNIP-LB formulation is defined by the same constraint sets as the binary SPNIP-M. In terms of modeling, SPNIP-LB differs in its derivation of non-detection

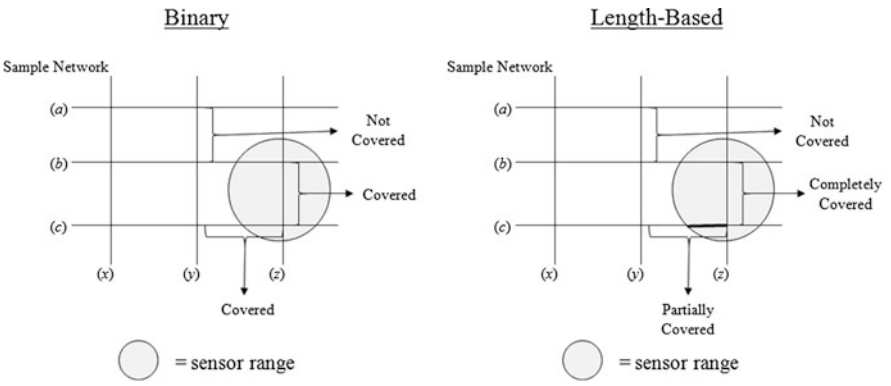


Fig. 5 Differentiating between the binary SPNIP-M and length-based SPNIP-LB formulations (Yates and Sanjeevi 2012)

probability and in composition of the objective function. In SPNIP-LB, calculation of non-detection probability is performed through a more complex function which more realistically models the connection between detection probability and time spent inside a sensor's range (Przemieniecki 2000). Figure 5 illustrates the concept of partial coverage and how it's able to be captured through formulation of the SPNIP-LB, adding yet another dimension of complexity for interdiction modelers.

Obtaining Solutions Through Bender's Decomposition

The interdiction formulations previously discussed share one major property that allows for a separation-based solution approach; no single set of constraints contains both attacker and defender variables. This means that the formulations may be divided into sub-problems for the attacker and defender respectively. Once these sub-problems are composed, they may be solved iteratively and linked together in a way that the solutions obtained in one sub-problem are used to feed the other cyclically. This approach is known as Bender's Decomposition (Bard 1998).

Implementing Bender's Decomposition for the aforementioned interdiction models, we derive an attacker sub-problem which maximizes path non-detection and is constrained by the conservation of flow constraints and fixed defender decision variables $\bar{\bar{x}}_{ist}$ (this results in the formation of a node-arc incidence matrix where the LP relaxation will return integral solutions (Nemhauser and Wolsey 1999). We derive the defender sub-problem to minimize path non-detection subject to the remaining constraints with the defender variables x and y and with fixed attacker decision variables $\bar{\bar{w}}_i$. Both sub-problems are now provided, along with an illustration of the decomposition approach in Fig. 6.

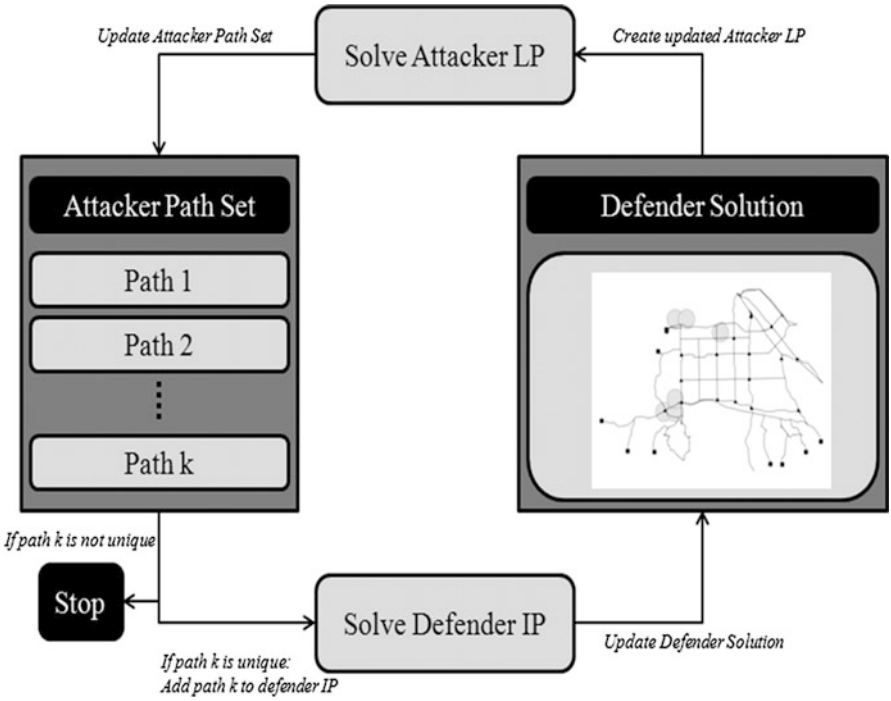


Fig. 6 Illustration of the Bender's decomposition method (Yates and Casas 2010)

[**Attacker Sub-Problem**]

$$z^{att} = \max \sum_{i,s,t} \log(u_{ist}) w_i \; \bar{\bar{x}}_{ist}$$
$$s.t. \; \sum_i k_{ni} w_i \leq q_n \quad \forall \; n$$
$$w_i \in R^+$$

[**Defender Sub-Problem**]

$$z^{def} = \min \; V$$
$$s.t. \; x_{ist} - \tfrac{1}{t} \sum_a r_i^{as} y_{as} \leq 0 \quad \forall \; i, s, t$$
$$\sum_{i,s,t} x_{ist} = 1 \quad \forall \; i$$
$$\sum c_s y_{as} \leq B$$
$$\sum \log(u_{ist}) \, \bar{\bar{w}}_i \, x_{ist} \leq V$$
$$y_{as}, x_{ist} \in Z^+$$

As in Fig. 6, the technique iteratively solves the attacker and defender sub-problems, passing solution information sequentially. Within each iteration, the defender allocates its resources optimally considering only those attacker paths found through previous iteration. This optimal allocation is then used to update all arc non-detection probabilities and used to obtain the attacker's path of maximum non-detection given the current resource allocation. If the path obtained by the attacker has already been considered in the constraint set (i.e. it has already been found in a previous iteration), an optimal defense resource allocation has been found

and the method stops. If the obtained path is not currently in the attacker path set, it is added and the defender sub-problem is solved again with the additional path considered. Bender's Decomposition yields provably optimal solutions and, in the worst-case, will iterate once for every unique, complete, simple path of the network (resulting in a worst-case performance of complete enumeration). In practice, Bender's Decomposition requires significantly less iterations to find the optimal resource allocation strategy.

Examining Network Interdiction Solutions

To assess interdiction solutions, we develop an experimental design that is used to examine two sub-networks of the Los Angeles County region. Multiple factors make up the experimental design, including road network and formulation type. Each identified factor has at least two test levels. Table 6 provides information pertaining to the initial experimental design and Fig. 7 illustrates the two test networks.

M-AST (Additive, Single Sensor Type)
M-NAST (Non-additive, Single Sensor Type)
M-NAMT (Non-additive, Multiple Sensor Types)

In Table 6, the parameter settings that define each of the six formulation levels are provided. The SPNIP-Length Based model has three distinct levels varying by sensor power (LB-1, LB-2 and LB-3). For SPNIP-M, three variations result. In M-NAST (Non-additive, Single Sensor Type), only a single sensor type is considered and no overlapping sensor coverage is allowed (i.e. $t = 1$). In M-AST (Additive, Single Sensor Type), only a single sensor type is considered but overlapping coverage is allowed until a given threshold, beyond which additional sensor coverage will not reduce arc non-detection. In M-NAMT (Non-additive, Multiple Sensor Types), multiple sensor types with various costs and sensor power parameters are considered, however no overlapping coverage is allowed (i.e. $t = 1$).

In Fig. 7, a uniform grid structure was used to establish the atom locations. The grid's scale is consistent for both Lancaster and Northridge, with these networks being chosen for experimental study due to their diversity in scale, complexity and density. Table 7 provides specific information on the network and atom sets for Lancaster and Northridge. Figure 8 illustrates the specific origin and critical infrastructure target locations for Lancaster and Northridge and Fig. 9 shows how arc influence is determined for SPNIP-M and SPNIP-LB on each network. U.S. Census Bureau classification (CFCC) was used to determine appropriate targets as follows (U.S. Census Bureau 2008). Green = {all regional airports}, Blue = {all regional airports and hospitals}, Yellow = {all regional airports, hospitals and police/fire stations}, Orange = {all regional airports, hospitals, police/fire stations and landmarks}, Red = {all regional airports, hospitals, police/fire stations, landmarks and schools/universities}. Note that any given target level includes all targets identified at preceding levels (i.e. all blue targets are included in the yellow target set). Origins were chosen randomly from the set of external/boundary nodes for each network.

Table 6 Experimental design factors, levels and problem parameters to evaluate SPNIP-M performance

Experimental design factors and levels					Parameter levels				
Level	Network	Destination	Budget	Formulation	Formulation	Sensor power, $\eta(0)$	Sensor power, $\eta(1)$	Sensor cost	Initial arc non-detection, $u(i01)$
1	Lancaster	Red	\$800	M-AST	M-AST	1	0.5	\$200	Uniform [0.3,0.7]
2	Northridge	Orange	\$1,200	M-NAST	M-NAST	1	0.5	\$200	Uniform [0.3,0.7]
3		Yellow	\$1,600	LB-1	M-NAMT	1	Varies	Varies	Uniform [0.3,0.7]
4		Blue		LB-2	LB-1	0	2	\$200	Uniform [0.3,0.7]
5		Green		LB-3	LB-2	0	3	\$200	Uniform [0.3,0.7]
					LB-3	0	4	\$200	Uniform [0.3,0.7]

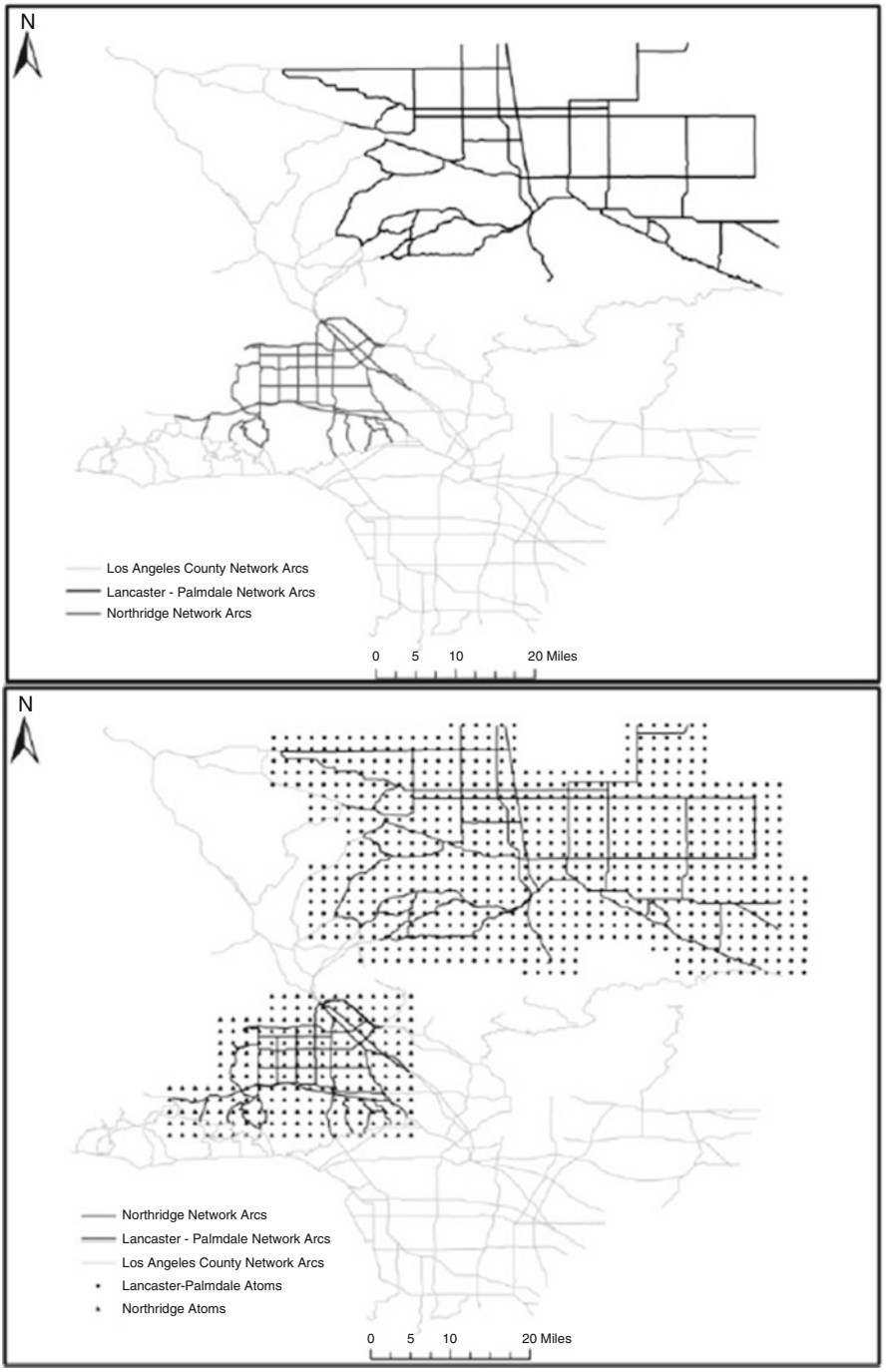


Fig. 7 Lancaster and Northridge test-case networks and their position within the Los Angeles County road network (Yates and Casas 2010)

Table 7 General regional and network data for Lancaster and Northridge

	Regional data			Road network data[a]				
	Area (sq. miles)	# Atoms	Atom density	Total	Density	Min	Max	Mean
Lancaster	1,295.61	743	0.57	465.61	0.36	0.12	14.51	4.47
Northridge	361.18	220	0.61	246.64	0.68	0.01	8.16	1.32

[a]Total, min, max and mean are measured in miles

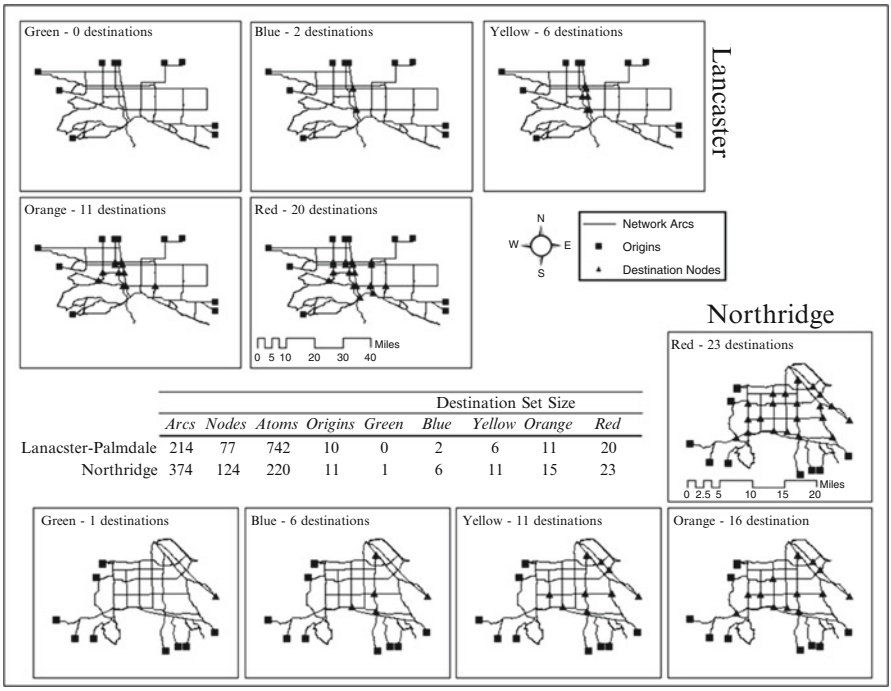


Fig. 8 Lancaster-Palmdale (top) and Northridge (bottom) network entry and CIKR target points at each identified threat level (Yates and Sanjeevi 2012)

Recalling that Benders Decomposition was used to solve for optimal SPNIP values, the resulting experimental design defined 30 individual problem instances for each of the six formulation levels (a total of 180 individual problem instances). Table 8 gives the aggregated results for five of the six formulation levels (M-NAMT is excluded as it is the only formulation which includes multiple sensor types). Tables 9 and 10 provide the individual run results for all SPNIP-M instances while Fig. 10 illustrates a typical SPNIP-M solution for the Northridge network.

In examining these solutions, we notice that, as expected, network and formulation choice directly impact objective value and computation time (significant differences are present across all formulation levels). Specifically, we begin to notice that SPNIP-M and SPNIP-LB performance is not monotonic. From Table 8, the results of the experimental design show that SPNIP-M is more efficient computationally in solving instances on the Lancaster network while SPNIP-LB is more efficient with respect to Northridge.

Table 8 Statistical summary of experimental results (Yates and Sanjeevi 2012)

	Lancaster				Northridge				
	Average	Std dev.	Min	Max	Average	Std dev.	Min	Max	
SPNIP-M NAST	0.052	0.023	0.013	0.087	2.76E−02	0.024	5.30E−06	5.59E−02	Objective
SPNIP-M AST	0.027	0.024	0.003	0.087	1.38E−02	0.019	5.19E−09	5.59E−02	
SPNIP-LB 1	0.004	0.009	1.17E−06	0.023	8.76E−04	0.002	5.36E−24	8.08E−03	
SPNIP-LB 2	0.004	0.009	2.21E−08	0.023	2.67E−04	0.001	7.10E−34	2.72E−03	
SPNIP-LB 3	0.004	0.009	2.80E−10	0.023	9.03E−05	2.43E−04	9.41E−44	9.18E−04	
SPNIP-M NAST	11	4.3485	6	18	17.2	11.3528	8	39	Iterations
SPNIP-M AST	15.25	3.7689	9	22	20.73	9.55784	5	35	
SPNIP-LB 1	14	4.4313	7	22	10	4.86973	2	19	
SPNIP-LB 2	14.67	6.3865	7	29	9.8	4.76895	2	17	
SPNIP-LB 3	14.75	5.7228	6	27	10.6	5.65433	2	21	
SPNIP-M NAST	N/A				N/A				Comp time
SPNIP-M AST	4.58	6.43	0.67	22.09	2,063.64	3,829.42	8.73	13,012.98	
SPNIP-LB 1	28.85	23.59	4.30	81.58	19.66	13.85	0.41	42.34	
SPNIP-LB 2	43.82	56.17	4.14	193.80	18.18	14.36	0.42	49.95	
SPNIP-LB 3	51.26	60.22	7.22	206.15	18.20	13.88	0.41	51.55	

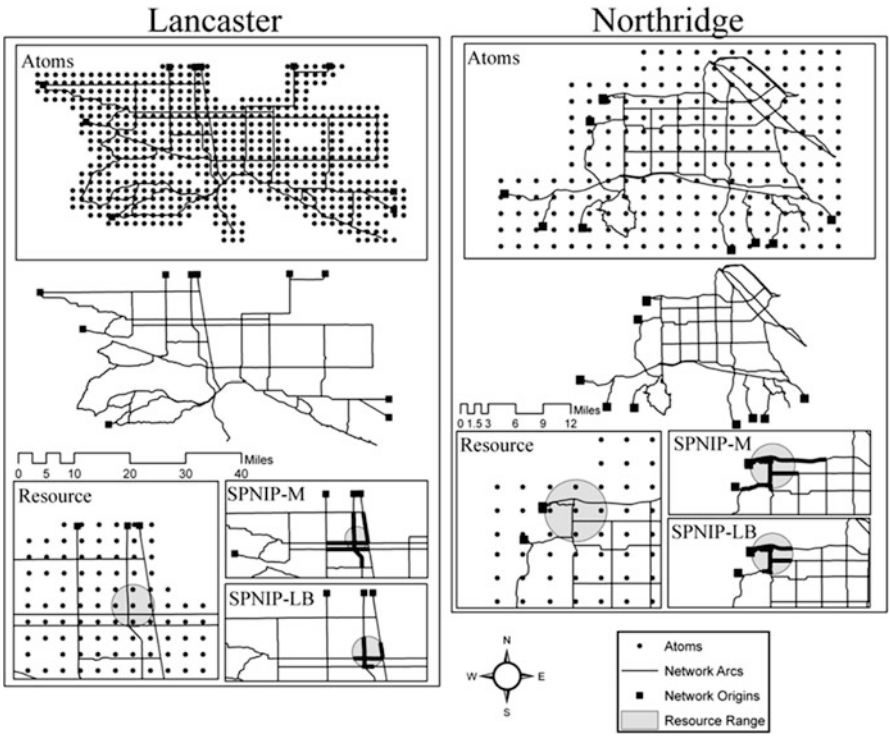


Fig. 9 Illustration of the network atom sets and a sample resource allocation (Yates and Sanjeevi 2012)

Extending SPNIP-LB

In SPNIP-LB, modification to the sensor model creates a simple function that enables the modeling of dynamic sensors. In contrast with the sensors used to this point, dynamic sensors are allocated to arcs within the network, repeatedly looping (i.e. covering) these arcs in similar fashion to a local law enforcement vehicle on patrol. The available dynamic sensor paths are finite and pre-determined and are represented by the set P. The collection of all sensor locations, which includes the set of atoms A for immobile or static sensors, is written as $C = A \cup P$.

Call t_i the amount of time spent traversing arc i of path p and t_p the total path traversal time for path p. A uniform distribution determines the probability that the sensor is present on arc i. The definition in (5), we can calculate the probability of non-detection for arc i under the influence of mobile sensor c as in (6)

$$P(T_{ip}) = \frac{t_i}{t_p} \tag{5}$$

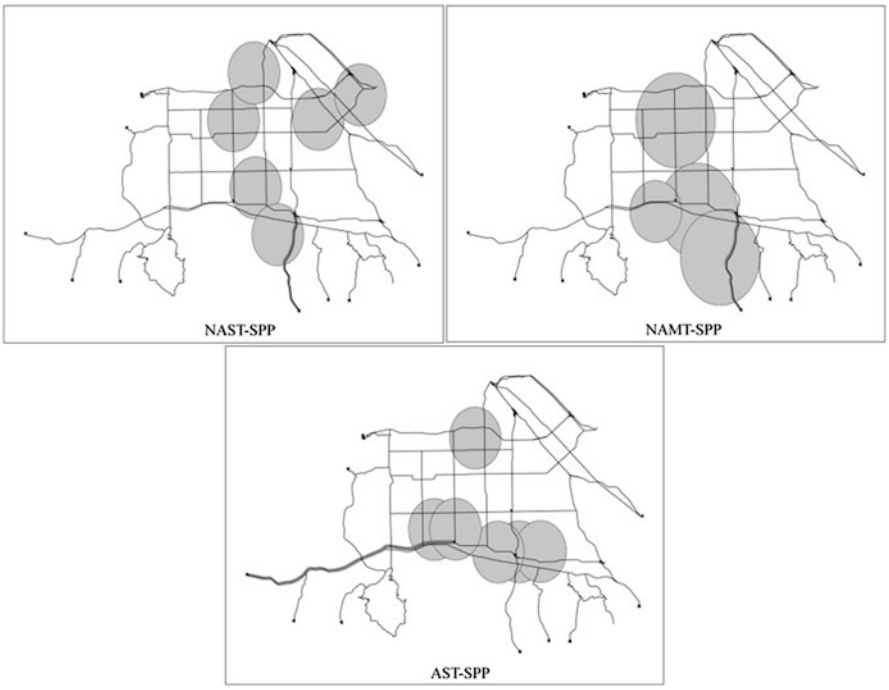


Fig. 10 Illustration of representative sensor placement results (Northridge shown)

$P(ND_{cip}|T_{ip})$ is the probability of non-detection for arc i covered by sensor located on path $p \in C$ (note that if path p contains the single arc i, then $t_i = t_p$ and $P(N\,D_{cip},\ T_{ip})$ equals the static sensor detection rate).

$$P(N\,D_{cip}|T_{ip}) = e^{-\eta l_{ic}}$$

$$\begin{aligned} P(N\,D_{cip}, T_{ip}) &= P(N\,D_{cip}|T_{ip})P(T_{ip}) \\ &= e^{-\eta l_{ic}}\,\frac{t_i}{t_p} \\ &= u_{ic} \quad \forall\ c \in P \end{aligned} \tag{6}$$

$P(ND_{cip}|T_{ip})$ is the probability of non-detection for arc i covered by sensor located on path $p \in C$ (note that if path p contains the single arc i, then $t_i = t_p$ and $P(N\,D_{cip},\ T_{ip})$ equals the static sensor detection rate. Also, dynamic sensors are assumed to have an influence range of 1.5 miles such that a 0.5 mile arc would have $l_{ic} = 0.5$ while a 2.4 mile arc would have $l_{ic} = 1.5$). Figure 11 illustrates an adaptation of the Northridge network to include six dynamic/mobilesensor paths in

Table 9 Objective function values at all threat levels for the Lancaster and Northridge test networks (objective function represents the min-max probability of non-detection for the given case)

		Green			Blue			Yellow			Orange			Red		
Lancaster	NAST	N/A			0.021	0.013	0.013	0.056	0.056	0.056	0.063	0.063	0.063	0.086	0.063	0.063
	NAMT				0.021	0.013	0.013	0.056	0.056	0.056	0.063	0.063	0.063	0.086	0.063	0.063
	AST				0.015	0.007	0.003	0.028	0.015	0.007	0.057	0.023	0.015	0.086	0.043	0.025
Northridge	NAST	6.80E−06	5.3E−06	5.3E−06	0.013	0.013	0.013	0.013	0.013	0.013	0.056	0.056	0.056	0.056	0.056	0.056
	NAMT	6.80E−06	5.3E−06	5.3E−06	0.013	0.013	0.013	0.013	0.013	0.013	0.056	0.056	0.056	0.056	0.056	0.056
	AST	4.30E−07	4.15E−08	5.19E−09	0.003	0.001	0.000	0.000	0.012	0.003	0.056	0.019	0.013	0.056	0.028	0.014
	Budget	$800	$1,200	$1,600	$800	$1,200	$1,600	$800	$1,200	$1,600	$800	$1,200	$1,600	$800	$1,200	$1,600

Table 10 Number of paths used at all threat levels for the Lancaster and Northridge test networks (number of paths represents the number of Bender's Decomposition iterations necessary to obtain an optimal solution)

		Green			Blue			Yellow			Orange			Red		
Lancaster	NAST	N/A			12	18	17	12	10	7	6	6	6	9	15	14
	NAMT				16	16	19	10	9	10	7	10	8	12	16	14
	AST				11	15	22	13	15	18	11	15	18	9	18	18
Northridge	NAST	38	38	39	9	9	9	18	16	16	12	8	8	13	13	12
	NAMT	36	34	32	8	9	10	24	22	18	9	9	9	12	12	11
	AST	21	30	33	16	19	32	15	32	35	5	12	18	8	15	20
	Budget	$800	$1,200	$1,600	$800	$1,200	$1,600	$800	$1,200	$1,600	$800	$1,200	$1,600	$800	$1,200	$1,600

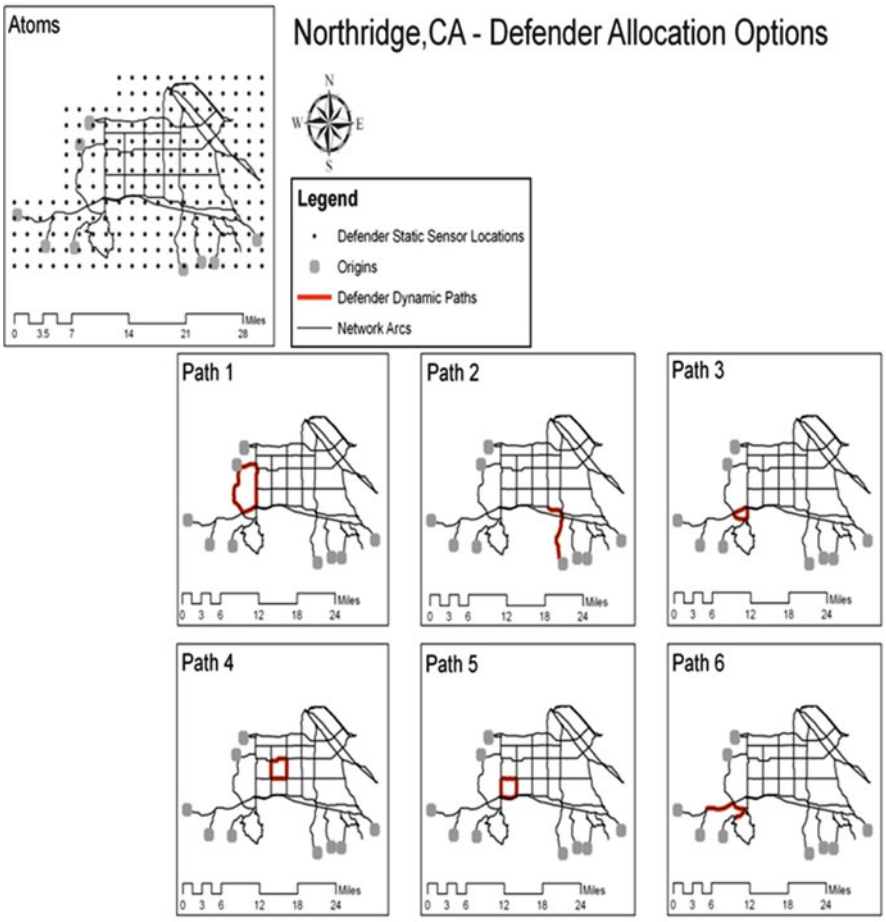


Fig. 11 Static and dynamic sensor locations for Northridge, CA (Yates and Sanjeevi unpublished)

addition to the static atom locations for the SPNIP-LB formulation. Figure 12 and Table 11 provide solution results after running several instances of SPNIP-LB with dynamic sensors.

Analyzing the Spatial Properties of SPNIP-M and SPNIP-LB Solutions

We use ArcGIS 10 to perform common spatial analysis techniques such as clustering and autocorrelation on the SPNIP-M and SPNIP-LB solutions. We note that much exists in the analysis of global and local clustering measures (local indicators of

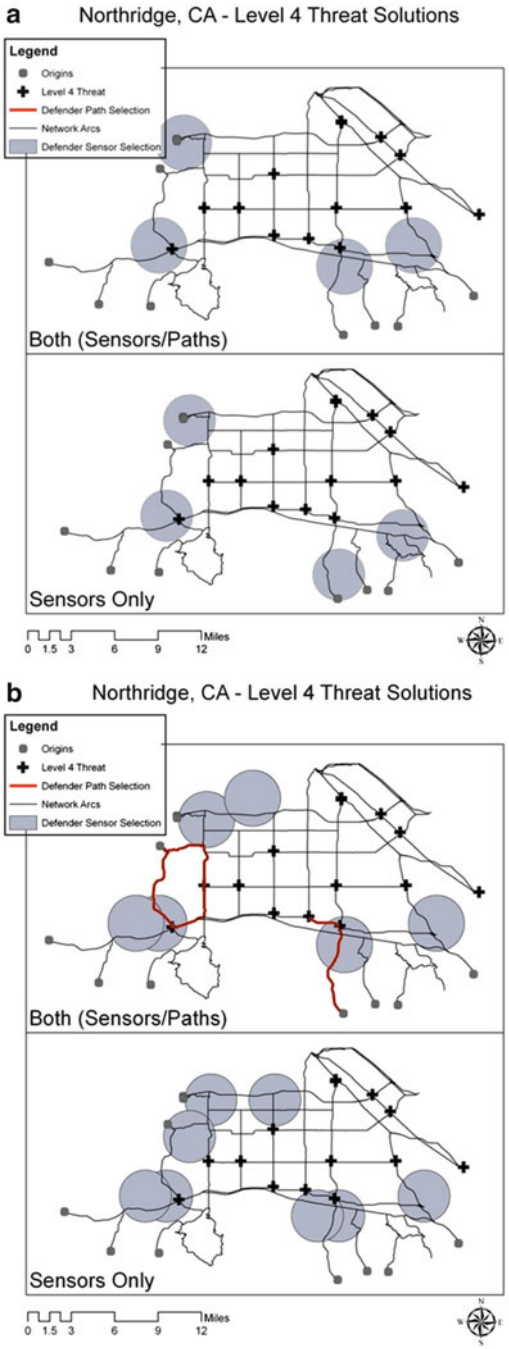


Fig. 12 Sample SPNIP-LB solution: "Level 4 Threat" with $\eta = 2$; (**a**) $B = \$800$ (**b**) $B = \$1{,}600$ (Yates and Sanjeevi unpublished)

Table 11 Atom and path allocations for level 4 threat (Yates and Sanjeevi unpublished)

Allocation type	Budget	Atom usage								Path usage	
Sensors only	$800	52	131	159	167						
	$1,600	20	62	131	134	154	172	186	201		
Both	$800	52	117	131	172						
	$1,600	20	35	131	154	172	201			1	2

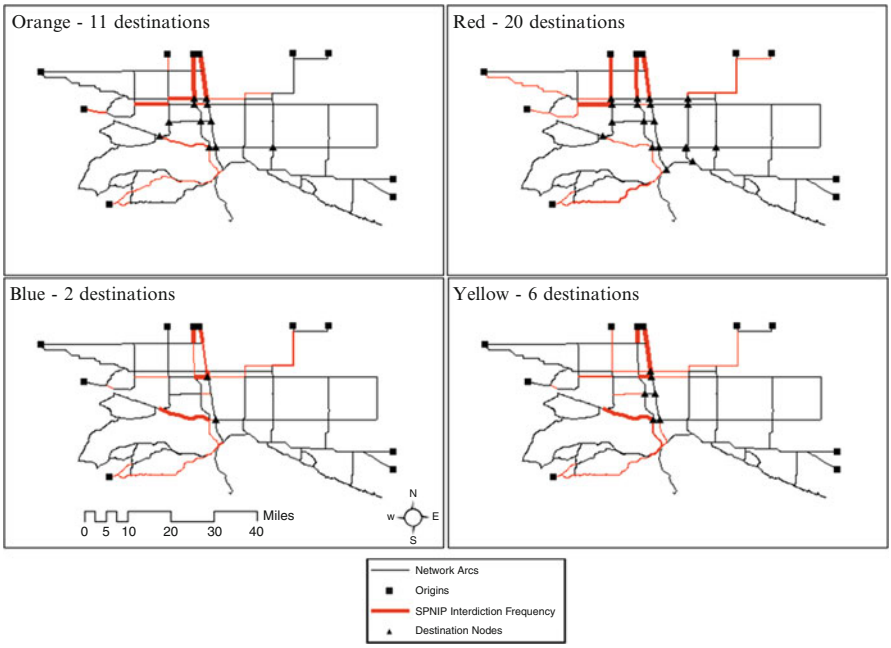


Fig. 13 Map of aggregate Lancaster SPNIP results at each destination level (Yates and Sanjeevi 2012)

spatial autocorrelation, LISA) on a network (Anselin 1995; Yamada and Thill 2010). With SPNIP-M and SPNIP-LB, there is an arc influence decision variable (x_{ist} and x_{ias} respectively) in addition to a variable indicating a sensor's location at an atom (y_{as} in both models). While atom locations have a regular spatial pattern (to this point, all atom locations have been grid-structured), their independence from the road network itself enables the application LISA measures.

Figures 13 and 14 illustrate the aggregate results of solving a traditional arc-based SPNIP model such as Israeli and Wood (2002) and the aggregate atom solutions across all previously defined factor levels for SPNIP-M and SPNIP-LB in Lancaster and Northridge respectively.

Kernel density and map algebra techniques were implemented in an effort to better understand the observed similarity between the aggregated models of Figs. 13 and 14 (see ESRI 2009) for discussion on kernel density and map algebra). At a high

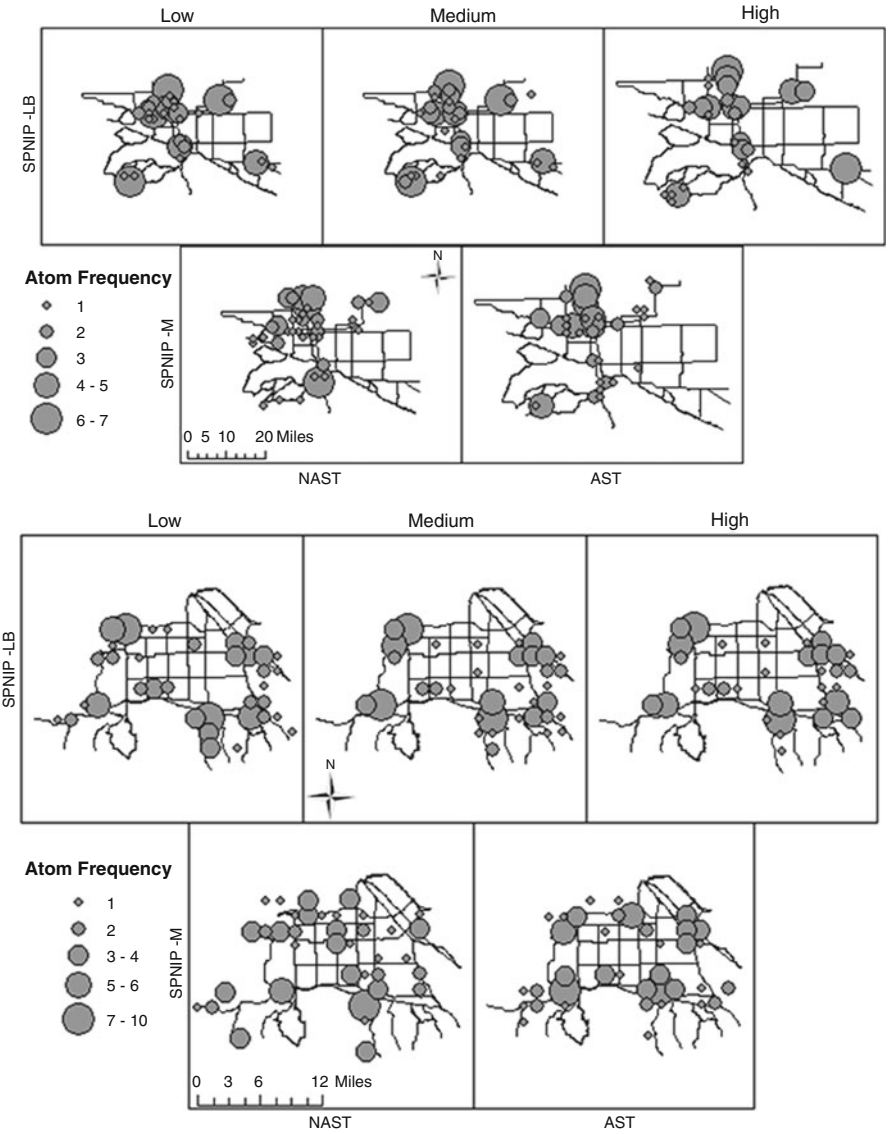


Fig. 14 Map of atom frequency in Lancaster-Palmdale (left) and Northridge (right) (Yates and Sanjeevi 2012)

level, kernel density is used to create a continuous-space image from the discrete atom frequencies that are obtained when SPNIP-M and SPNIP-LB solutions are aggregated. The continuous kernel density image is akin to a pixilated image and is different from the individual discrete point-based atom locations (all pixels are given a value through implementation of the kernel density and atom frequencies

are "smoothed" through the continuous space occupied by the atoms). Map algebra techniques are used to perform calculations on one or multiple continuous kernel density files. We use map algebra techniques here to define similarity metrics for the obtained kernel densities. Figure 15 illustrates the obtained kernel densities from the atom frequencies of Fig. 14.

We define two measures for comparing aggregated solutions. The first, Less-Than-Or-Equal-To (LessEQ), returns the percentage of individual pixels for kernel density input A that are less than or equal to the individual pixels for kernel density input B. The second, Equal-To (EQ), only returns the percentage of pixels for input A and input B that are identical. Both LessEQ and EQ are used to evaluate the similarity in aggregate coverage between SPNIP-M and SPNIP-LB. Table 12 provides the obtained similarity results using the kernel densities from Fig. 15 and the two raster similarity measures.

The motivation for this analysis stems from discussion in the previous section where the computational performance of SPNIP-M was more efficient in Lancaster than in Northridge. If it can be shown that results from the two formulations are similar, then they may be used interchangeably. This would provide flexibility for the modeler to choose or continue to use formulations exhibiting efficient performance. Additionally, such information on similarity can be useful in the development of approximation techniques to reformulate and solve interdiction problems (as will be discussed in the next section). In the case of Table 12, LB-3 exhibits strong similarity with M-NAST and M-AST, especially in Lancaster where average equality is between 63 and 70 %. Similarly, LessEQ demonstrates that both M-NAST and M-AST are capable of meeting or exceeding LB coverage in 85 % of the experimental runs.

Developing Interdiction Approximations and Heuristics

The previous section highlighted the computational performance and spatial characteristics of certain shortest path network interdiction problem variants. Though the aforementioned represents a small subsection of interdiction formulations, there were inherent trade-offs in computational performance across different formulations. A knowledgeable modeler or public policy maker could use these trade-offs to more effectively obtain information on the region and critical infrastructure being examined. While such gains would be beneficial to the modeler, the spatial similarities in solution characteristics support the assertion that there are inherent solution properties that may be replicable or decipherable either through alternative, approximate formulations or new solution techniques. This section is devoted to examining how the shortest path network interdiction problem can be re-modeled and re-evaluated for the purposes of developing faster, stronger approximation and solution techniques. We will begin by discussing a knapsack approximation to the SPNIP-M problem of the previous section. After the approximation is introduced, we will

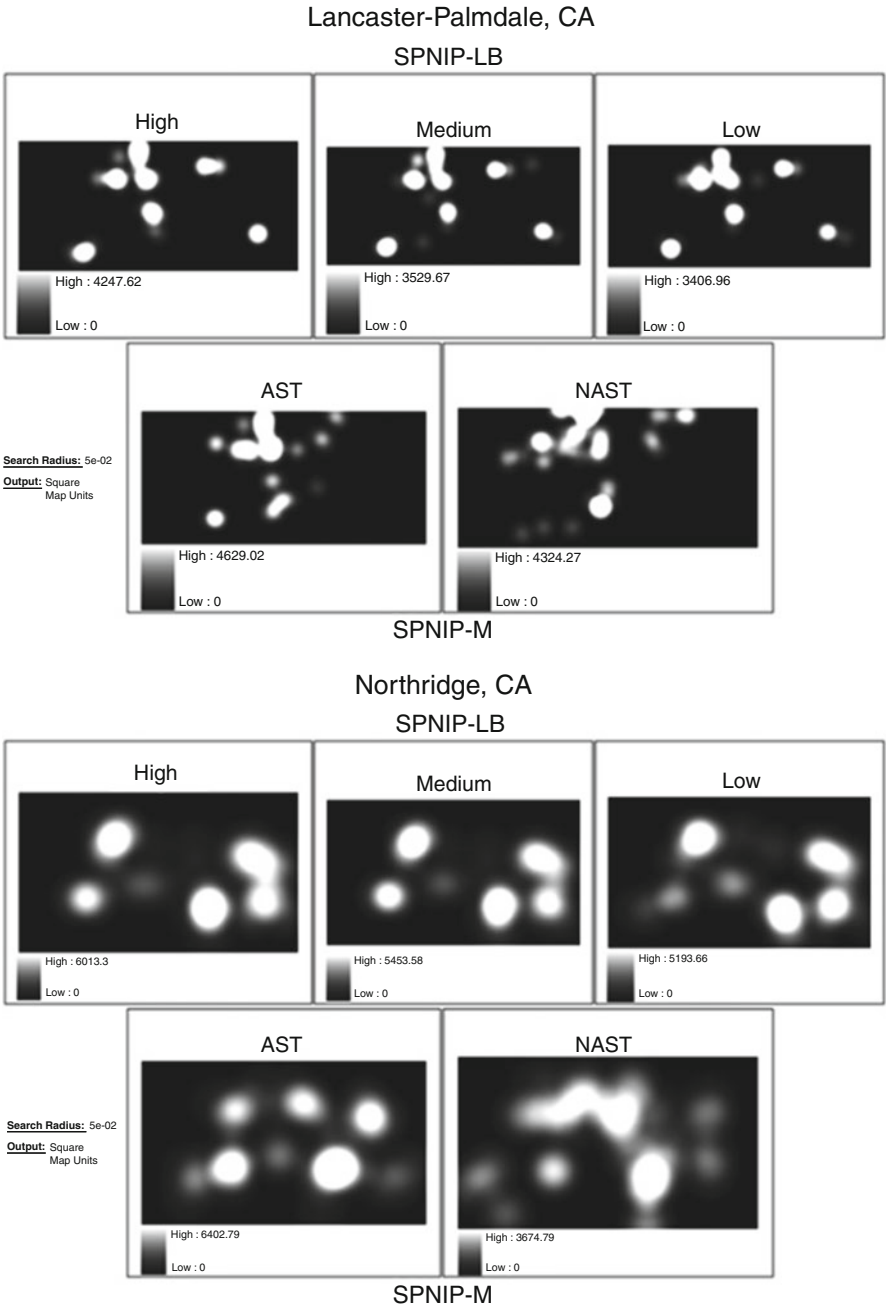


Fig. 15 Kernel density obtained from aggregate atom frequencies (Yates and Sanjeevi 2012)

Table 12 Spatial similarity of formulation results

First entry	Second entry	Raster	Lancaster		Northridge	
			Mean	Std dev.	Mean	Std dev.
LB-3	M-NAST	LessEQ	0.858	0.349	0.648	0.478
LB-3	M-AST	LessEQ	0.887	0.317	0.683	0.465
LB-1	M-NAST	LessEQ	0.838	0.369	0.630	0.483
LB-1	M-AST	LessEQ	0.859	0.348	0.601	0.490
LB-2	M-NAST	LessEQ	0.826	0.379	0.636	0.481
LB-2	M-AST	LessEQ	0.836	0.370	0.665	0.472
LB-3	M-NAST	EQ	0.665	0.472	0.168	0.374
LB-3	M-AST	EQ	0.705	0.456	0.224	0.417
LB-1	M-NAST	EQ	0.640	0.480	0.156	0.362
LB-1	M-AST	EQ	0.685	0.465	0.194	0.396
LB-2	M-NAST	EQ	0.629	0.483	0.151	0.364
LB-2	M-AST	EQ	0.658	0.474	0.212	0.409

examine ways to increase the performance of Bender's Decomposition and conclude with provocation of a new, heuristic approach to solving interdiction problems.

A Knapsack Approximation

Within an interdiction problem, the primary concern of the defender is to identify arcs and/or nodes that are most critical to the protection of critical infrastructure and protect, reinforce, or otherwise deter an attacker from using those arcs. The Bender's Decomposition approach to solving interdiction models is essentially an iterative approach that uses an attacker-based sub-problem to build a set of likely attacker paths through the network. These attacker paths contain what can be considered the critical arcs. Severing or preventing the attacker from using these critical arcs by allocating regional resources is the primary defender concern.

Identifying critical, vulnerable, or salient network arcs is not a new problem. Many approaches use a cut-set mentality to identify arcs whose absence will cut off or extremely inhibit flow between origins and destinations. A classic application of such an approach is given in Matisziw et al. (2007). Many interdiction models, especially those in early hazardous materials literature, are predicated on the maximum flow-minimum cut paradigm, using a minimum cut as the primary method to identify critical network arcs. The following list contains such max flow-min cut incorporated models: Burch et al. (2003), Corley and Chang (1974), Cunningham (1985), Ghare et al. (1971), Phillips (1993), Ratliff et al. (1975), Wollmer (1964), and Wood (2003).

The approximation technique discussed here is essentially a constrained knapsack optimization problem that identifies the attacker's critical network arcs, again relying on the max flow-min cut property as its foundation. The major premise of the approximation is to consider minimum capacity cuts in the network as a means to identify likely attacker critical arcs. In doing so, the following lemma guides the approximation (the proof for this lemma is by contradiction and may be found in Yates and Lakshmanan (2011). The knapsack approximation formulation [KNAP] follows the lemma.

Lemma *For any given network $G(N, \Lambda)$ having probability of non-detection as its flow metric, the maximum flow in any path is an upper bound on the total path non-detection probability.*

$$
\begin{aligned}
& \text{[KNAP]} \\
z^* = \max & \sum_b \varphi_b v_b \\
s.t. & \sum_b c_1 v_b \le B \qquad\qquad \varphi_b = \sum_{j \in R^{b1}} \left(\alpha\, m_j + \tfrac{1-\alpha}{p_j} \right) \\
& \sum_{b \in \phi^j} v_b \le \tau \quad \forall\ j \in \Lambda \\
& v \in \{0,\ 1)
\end{aligned}
$$

z^* is the weighted objective utility for KNAP and contains two components. The first, m_j, is the aggregated maximum flow for arc j considering all origin and critical infrastructure pairs. The second, $\frac{1}{p_j}$ is the inverse of the node count between arc j and its closest origin. α Determines the emphasis placed on the objective with the goal of KNAP to locate sensor resources at atoms of the network such that z^* is maximized. In addition to the standard knapsack budgetary constraint, KNAP also constrains the amount of tolerable sensor overlap within a sensor allocation scheme in the same way that DSPNI used the subscript index t to control the degree to which sensor overlap was counted when calculating z.

When examining solutions to the knapsack approximation, it is import to note that only defender allocation schemes are determined under this method (i.e. no attacker path information is provided). As a method for defender's to gain useful information on situational awareness, the knapsack approximation provides a fast and reliable approximation to defender sensor location. Figures 16 and 17 and in Table 13 provide computational results of the knapsack and SPNIP-M formulation solutions on the Lancaster and Northridge networks and using the same experimental design discussed in the previous section. In Figs. 16 and 17, z is the optimal objective value for SPNIP-M and z' is the objective value obtained when the KNAP sensor solution is evaluated for the SPNIP-M objective. Also in the figures, KNAP parameters were set at $B = \$3{,}600$, $\alpha = 0.02$ and $\tau = 3$.

Comparing the computational results of the knapsack approximation illuminates a few important trends. First, the approximation reliably captures the form of the SPNIP-M objective function through a simple approximation based on a well-

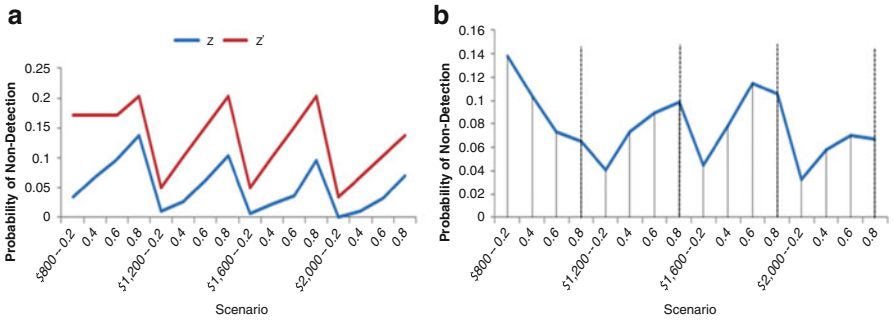


Fig. 16 Lancaster-Palmdale, random initial non-detection probability (Yates and Lakshmanan 2011)

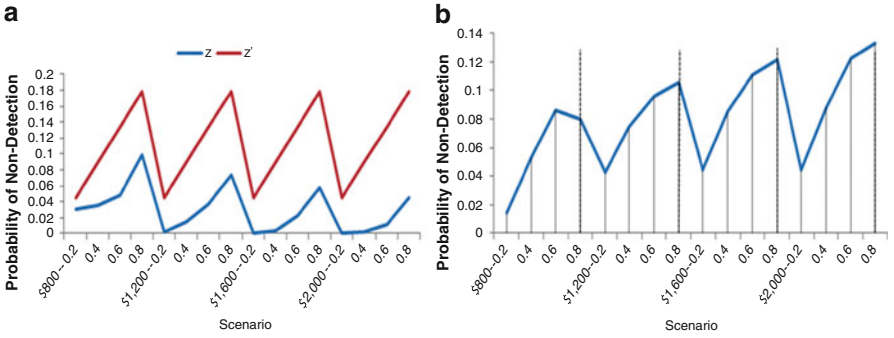


Fig. 17 Northridge, random initial non-detection probability (Yates and Lakshmanan 2011)

known and understood optimization principle, in this case maximum flow-minimum cut. Second, the approximation demonstrates insensitivity to network choice. Third, the approximation demonstrates computational insensitivity to changes in SPNIP-M problem parameters. Computationally, the knapsack approximation performs well for the cases examined and appears to be a well suited alternative to model defender sensor location in SPNIP-M. We now introduce and discuss the knapsack's ability to spatially approximate SPNIP-M solutions. To do this, the same spatial analysis techniques (kernel density and map algebra) were applied as in the previous section. Figure 18 illustrates the obtained kernel densities while Table 14 gives the LessEQ and EQ values.

The approximation's spatial performance is promising, though not as strong as its computational capabilities. With roughly 75 % similarity to the SPNIP-M solution, the approximation performs well in Lancaster and appears to increase its performance as overlap (τ) increases. While the approximation does not perform as well in Northridge, only a small number of possible parameter combinations are provided in Table 14 and it is highly probable that the knapsack approximation could be strengthened through a more comprehensive pareto analysis.

Table 13 Time to solution (CPU seconds)

Sensor power	0.2			0.4			0.6			0.8			KNAP		
Budget	$800	$1,200	$1,600	$800	$1,200	$1,600	$800	$1,200	$1,600	$800	$1,200	$1,600	$800	$1,200	$1,600
Lancaster	13.14	35	53.424	11.344	14.642	33.952	9.611	16.188	21.702	6.61	7.938	11.45	2.935	2.904	2.889
Northridge	32.626	72.606	221.767	31.983	59.982	75.433	21.311	27.268	40.827	14.875	22.672	22.579	8.743	8.743	8.712

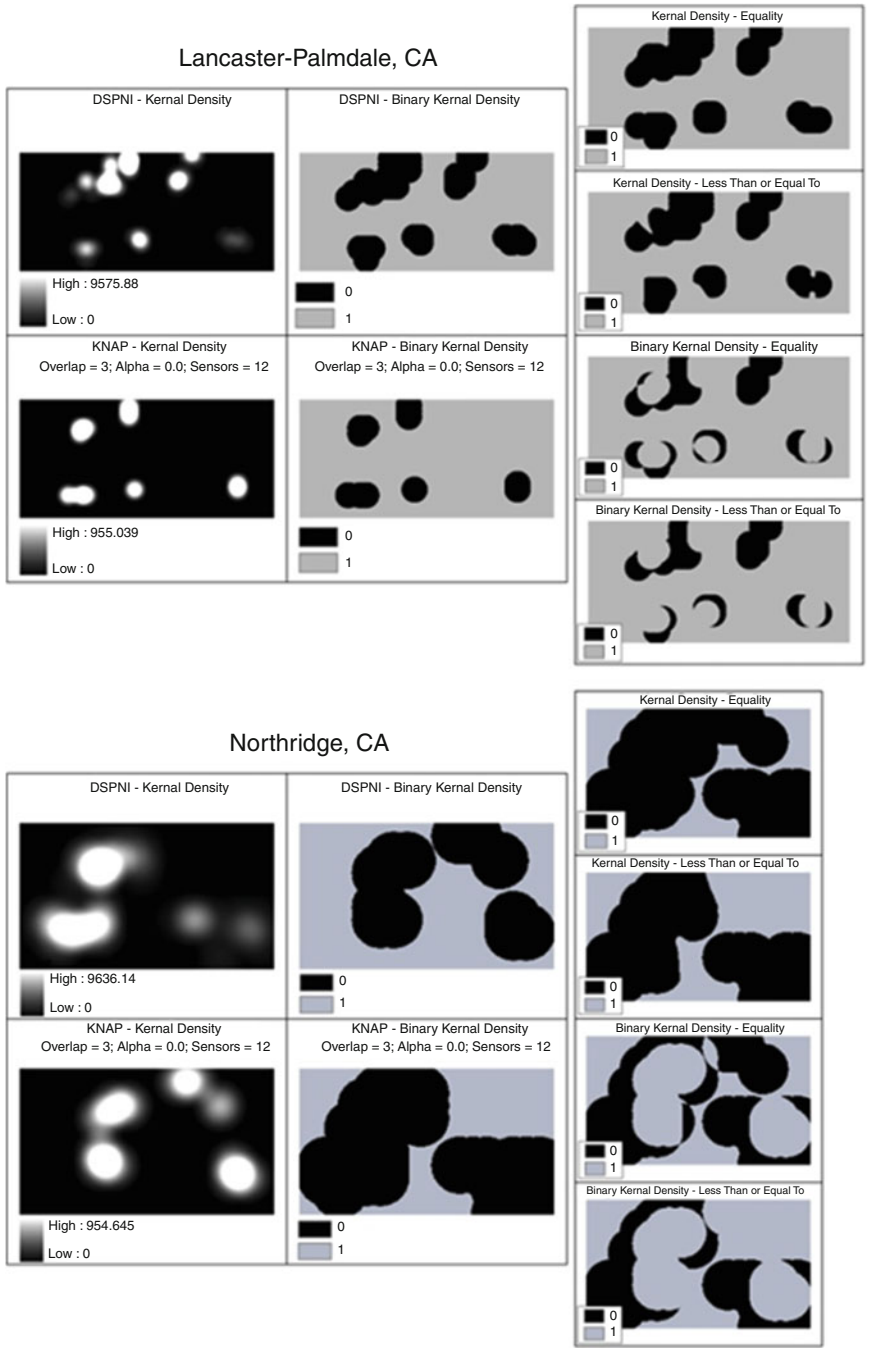


Fig. 18 Kernel and binary kernel density and comparison for the Lancaster-Palmdale and Northridge case study regions (Yates and Lakshmanan 2011)

Table 14 EQ and LessEQ similarity values for Lancaster and Northridge under varying KNAP parameters

	Overlap	Type	Alpha = 0.0		Alpha = 0.50		Alpha = 1.0	
			Mean	Std dev.	Mean	Std dev.	Mean	Std dev.
Lancaster	1	EQ	0.684	0.465	0.667	0.471	0.664	0.472
		LessEQ	0.762	0.426	0.759	0.428	0.761	0.426
	2	EQ	0.740	0.439	0.710	0.454	0.710	0.454
		LessEQ	0.758	0.428	0.760	0.427	0.761	0.426
	3	EQ	0.738	0.440	0.701	0.458	0.701	0.458
		LessEQ	0.764	0.425	0.759	0.428	0.760	0.427
	4	EQ	0.742	0.437	0.724	0.447	0.703	0.457
		LessEQ	0.764	0.425	0.761	0.426	0.762	0.426
Northridge	1	EQ	0.182	0.386	0.187	0.390	0.187	0.390
		LessEQ	0.429	0.495	0.434	0.496	0.429	0.495
	2	EQ	0.208	0.406	0.227	0.419	0.227	0.419
		LessEQ	0.399	0.490	0.408	0.492	0.408	0.492
	3	EQ	0.248	0.432	0.217	0.412	0.217	0.412
		LessEQ	0.381	0.486	0.393	0.489	0.393	0.489
	4	EQ	0.248	0.432	0.221	0.415	0.221	0.415
		LessEQ	0.384	0.486	0.399	0.490	0.399	0.490

The knapsack approximation demonstrates how an entirely new formulation can be developed to take advantage of the unique structure of the network interdiction problem. With a relatively simplistic formulation capable of being solved efficiently, close approximations to the SPNIP-M formulation's optimal solutions were obtained with significantly less computational effort. The next section will discuss how a similarly simple idea can be used to eliminate costly Bender's Decomposition iterations in solving SPNIP-M.

Approximating with k-Shortest Paths

In lieu of redefining a new optimization problem for network interdiction or reformulating an existing one, this section will discuss the benefits of modifying the interdiction solution approach. For this discussion, the same basic experimental design introduced in section "Formulating Interdiction Models" will be used and applied to the Northridge network. Recalling that SPNIP-M is solved by implementing Bender's Decomposition, the iterative attacker-defender sub-problem format is revised in an effort to decrease overall computation time. Additionally, this work nicely complements the knapsack approximation approach by enabling modelers a quick and easy approach to obtaining quality attacker paths (recall that the knapsack approximation only provided defender resource allocation solutions. Figure 19 illustrates the original Bender's Decomposition structure and the suggested revision to be examined here.

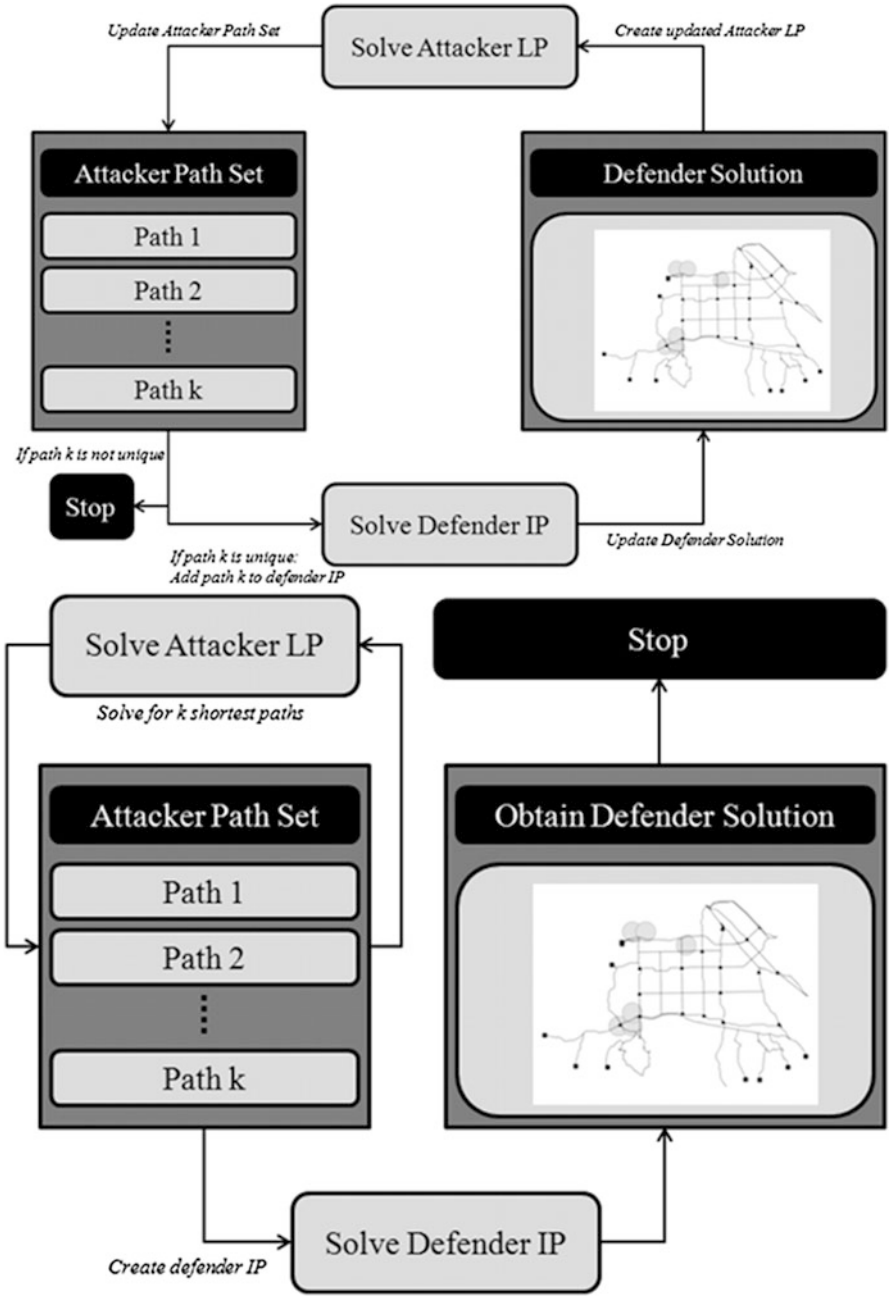


Fig. 19 Illustration of the original and modified Bender's decomposition (Yates et al. in submission)

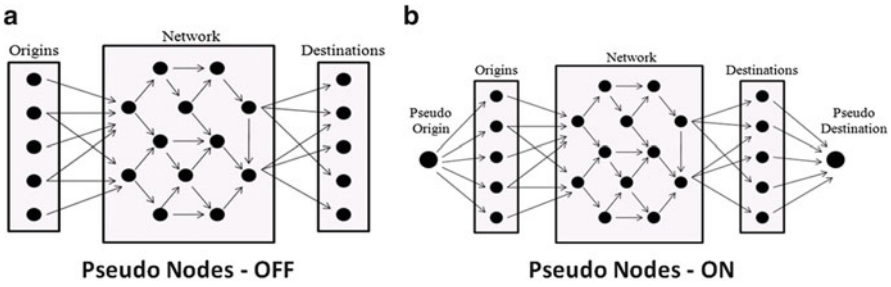


Fig. 20 Illustration of the exclusion and inclusion of pseudo nodes to solve for the *k*-shortest paths (Yates et al. in submission)

The traditional Bender's Decomposition approach decomposes the SPNIP-M into attacker and defender sub-problems which are then iteratively solved until an equilibrium point has been reached, signifying an optimal solution. In this approach, the defender sub-problem is a complex mixed integer problem that accounts for a majority of the computational effort. In the revised method, it is surmised that gains in computational performance can be achieved by solving the defender problem only once instead of iteratively and repeatedly. In Fig. 19, this method is illustrated on the right-hand-side, where the attacker sub-problem is looped, essentially solving a *k*-Shortest path problem to identify the *k* paths of high non-detection given the initial, random arc non-detection metrics (see Yen 1971) for discussion on the *k* shortest path problem). Each of the individual *k* paths then becomes a constraint in the defender sub-problem of SPNIP-M and the defender sub-problem is solved once to determine an acceptable defense allocation. At its core, this approach essentially asks "what is the appropriate *k* value to capture the optimal attacker path in SPNIP-M?"

To assert whether there is an acceptable *k* value, a modified experimental design approach was developed. First, a global origin and target set were identified within Northridge. Four levels were used to dictate the origin set size [2,4,6,8] and target set size [3,9,15,21], within which five random origin and destination sets were generated. When calculating the *k* shortest paths, pseudo nodes were either included or not included, resulting in two additional experimental levels (Fig. 20 illustrates these two levels). Lastly, the value of *k* was set to [1,2,3,5] when pseudo nodes were off and [5,10,15,20] when pseudo nodes were on. In the Northridge network, initial arc non-detection values were assigned randomly using a *Uniform[0.3, 0.7]* distribution.

To analyze and compare the solutions obtained using the *k*-Shortest approach with the SPNIP-M optimal, *Gap V%*, *Gap T%*, *% Under* and *% Over* are used. *Gap V%* is the percentage difference in the *k*-Shortest solution from the SPNIP-M optimal while *Gap T%* measures the difference in computational time (in CPU seconds). *% Under* and *% Over* are spatial metrics to assess coverage similarity as a function of length-of arc covered and are aggregated over all network arcs

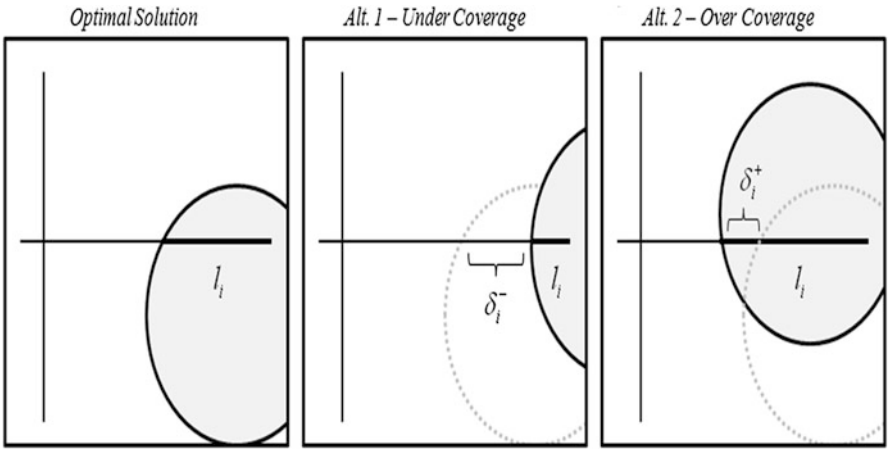


Fig. 21 Illustration of under coverage and over coverage in the calculation of % Under and % Over (Yates et al. in submission)

for a given experimental run. In this way, the *k*-Shortest path solution may either exactly replicate SPNIP-M arc coverage, under-cover and arc, or over-cover an arc. Figure 21 illustrates the later two cases. Table 15 provides information on the comparative performance of the *k*-Shortest path approach.

Table 15 provides the obtained averages from the five random experimental instances at each origin-target level. From the table, there are observable benefits in adapting traditional Bender's Decomposition with a *k*-Shortest path approach. First, the *k*-Shortest approach identified the optimal SPNIP-M solution in at least one of the five random replications at each origin-target level (as indicated in Table 15 by a minimum *Gap V%* of "–"). In all but two levels, computational effort was also saved. Spatially, the ability of the *k*-Shortest approach to identify optimal SPNIP-M solutions is acceptable (∼5 % under coverage with pseudo nodes off and 15 % with pseudo nodes on). Over coverage is relatively consistent in cases, with an average of 5 %.

Figure 22 visualizes two select, comparative cases between the optimal SPNIP-M solution and the obtained solution based on the *k*-Shortest path approach. In the figure, the left-hand case demonstrates an instance of strong computational and spatial coverage. In this case, the optimal SPNIP-M solution was obtained through the *k*-Shortest approach with a 23 % time savings . The right-hand case, however, illustrates a scenario where *k*-Shortest fails to perform. With 33 % under coverage and a 45 % optimality gap, the SPNIP-M optimal solution was not well approximated. The latter case signifies that the choice of k, in this case $k = 5$, was not high enough to adequately capture any attacker path trends.

The simplicity of this *k*-Shortest approach can be a huge advantage in decision-making and public policy, where modelers would desire the ability to test large numbers of scenarios and network compositions repeatedly. Though the Northridge

Table 15 Summary statistics of computational performance for the k-shortest approximation

	Gap *V*%					Gap *T*%				% Under				% Over			
	Level	Avg.	Std dev.	Min	Max	Avg.	Std dev.	Min	Max	Avg.	Std dev.	Min	Max	Avg.	Std dev.	Min	Max
Off	1	1.788	3.530	–	9.569	65.108	42.575	−90.960	93.764	0.087	0.097	–	0.308	0.023	0.030	–	0.131
	2	1.229	3.270	–	14.183	8.597	64.795	−124.812	79.724	0.039	0.040	–	0.127	0.020	0.025	–	0.089
	3	–	–	–	0.000	−208.038	272.568	−960.824	47.464	0.055	0.059	–	0.204	0.048	0.045	–	0.145
	4	–	–	–	0.000	−443.798	461.330	−1,391.150	7.908	0.082	0.088	–	0.245	0.023	0.021	–	0.058
On	1	8.742	12.537	–	31.360	38.907	53.681	−136.874	92.306	0.179	0.221	–	0.573	0.040	0.042	–	0.131
	2	26.488	27.271	–	68.591	65.623	22.462	21.882	94.724	0.221	0.214	–	0.651	0.033	0.024	–	0.081
	3	23.017	31.230	–	97.515	43.124	42.471	−65.230	93.829	0.098	0.095	–	0.327	0.053	0.066	–	0.253
	4	26.034	33.593	–	143.604	51.555	33.229	−9.090	94.241	0.140	0.130	–	0.522	0.047	0.038	–	0.111

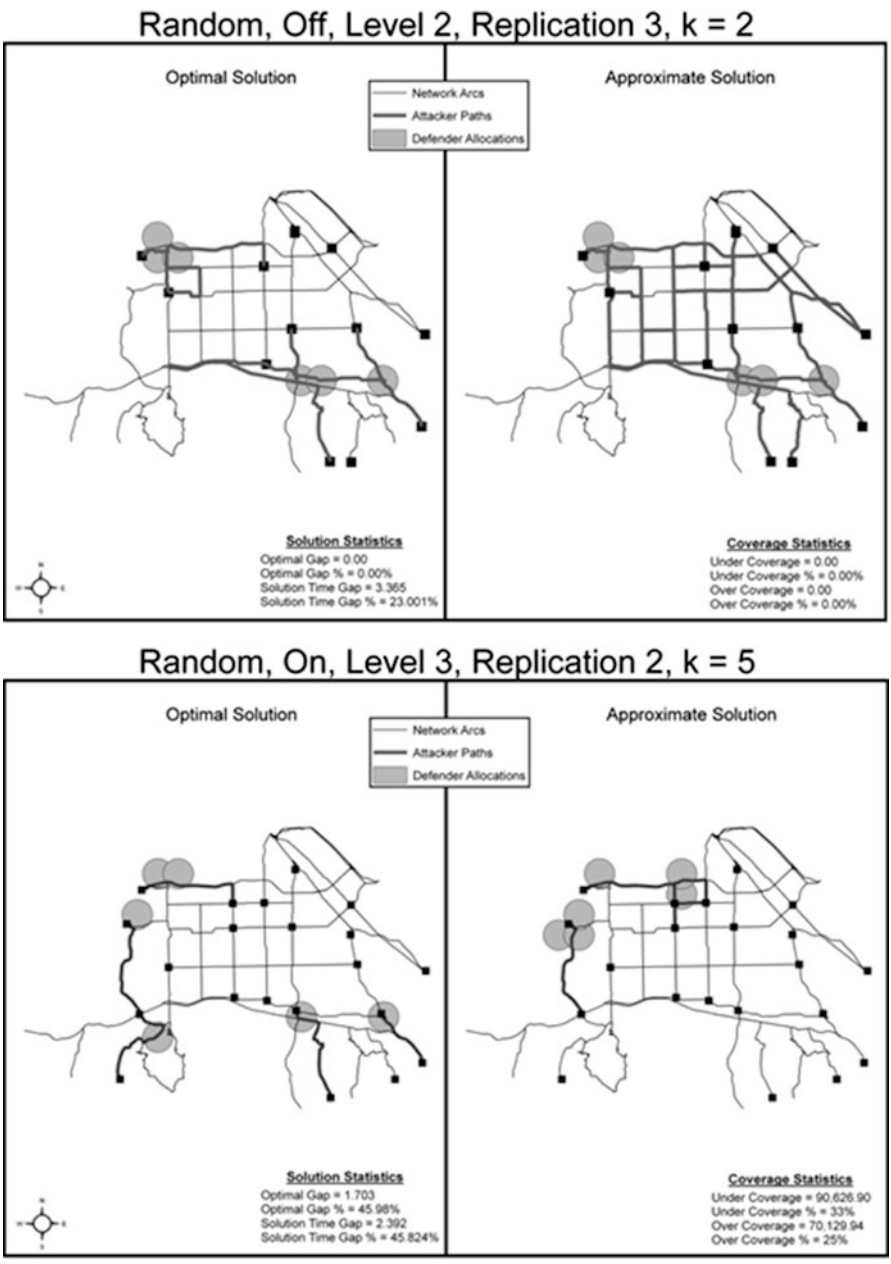


Fig. 22 Example of poor spatial performance by the approximation (Yates et al. in submission)

case is decidedly small, the SPNIP-M formulation grows exponentially with the size of a network. In Northridge (384 directed arcs), SPNIP-M was formulated with 3,437 constraints while a network of 1,000 directed arcs and 500 nodes produces 8,869 SPNIP-M constraints. Increasing network size by a factor of 10, a 100,000 direct arc network with 1,000 nodes has nearly 100 times the number of SPNIP-M constraints at 801,369. Using the entire Los Angeles County road network (U.S. Census Bureau 2008) easily leads to 2,000,000+ constraints. While the complexity of these problems may lead to elongated solution times and the necessity to include more advanced and strategic computational techniques, known *k*-shortest path algorithms enable this simple solution approach to handle problems of realistic scale without significant alteration. Stated previously, such a tool would be invaluable to emergency planners, responders and public policy makers alike when analyzing network performance/vulnerability/accessibility as they could test a plethora of event scenarios.

Additionally, the *k*-shortest approximation approach preserved much of the spatial integrity of SPNIP-M solutions (i.e. small under and over coverage measures), implying that this simple approximation approach could be used as a capable technique for other modified shortest path network interdiction models. As an example, the under and over coverage values indicate that quality SPNIP-LB solutions, where arc length was used directly in the determination of non-detection probability (Przemieniecki 2000) could be well approximated by this approach.

Identifying Basic Network Trends

The knapsack and *k*-Shortest path approximations previously discussed illustrate how simple concepts and applications in the optimization of network interdiction can be used to develop alternative techniques in obtaining interdiction solutions. Additionally, the similarity between solutions of the SPNIP-M and SPNIP-LB problems in "Formulating interdiction models" combined with the approximation accuracy of the knapsack and *k*-Shortest approaches implies that there are certain problem properties and parameters (e.g. network complexity/structure and atom set composition) that have a high level of impact in determining defender and attacker interdiction solutions. Specifically pertaining to the SPNIP-M and SPNIP-LB, four problem properties were identified and tested to determine their significance in influencing the corresponding interdiction solutions. An experimental design similar to those already used was developed and tested within three real-world network: New York City, Boston and Houston. These networks were chosen because of their diverse network structure and were tested in combination with three distinct atom sets (Casas et al. 2012). The atom sets used maintain uniform, grid-like spacing between atoms but change density to increase the number of potential sensor locations.

To test whether a given problem property was influential or not, a Negative Binomial Regression (NBR) was used. The NBR was chosen instead of other

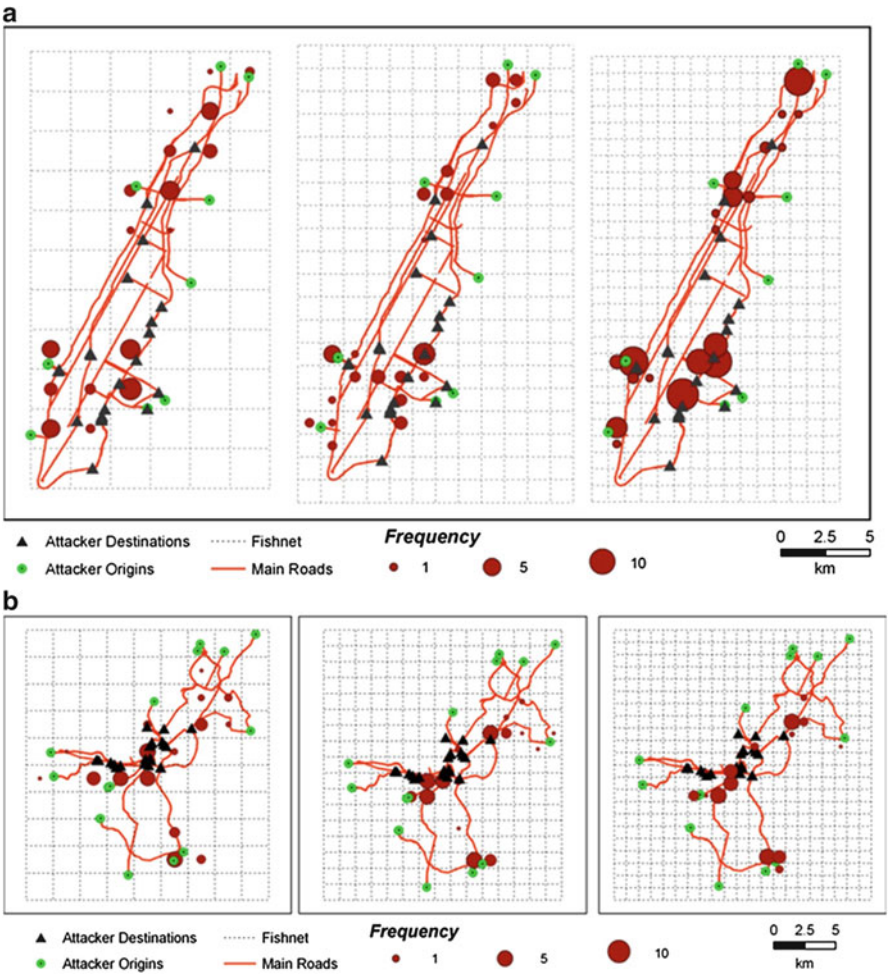


Fig. 23 Cumulative atom usage frequency from SPNIP-M experimental design for NYC (**a**) and Boston (**b**) (Casas et al. 2012)

regressions such as OLS or Poisson Regression due to the fact that most available atoms in SPNIP-M and SPNIP-LB are not used to locate sensors in the final solution. This creates sparse solution sets that NBR is better adept to evaluate. For information on NBR, please see Hilbe (2007). When implementing NBR to assess correlation, the frequency of atom use in aggregated SPNIP-M optimal solutions across the experimental design is used as the measuring variable.

Figure 23 illustrates the SPNIP-M optimal solutions for NYC and Boston while Fig. 24 illustrates the solutions for Houston. In the figures, frequency of use represents the cumulative number of sensors located at the particular atom across

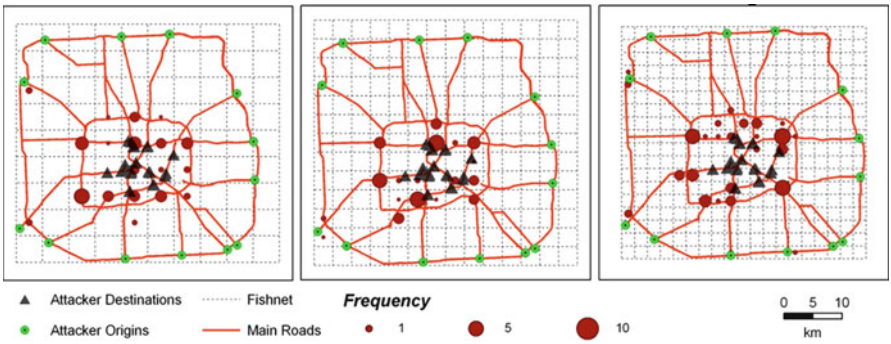


Fig. 24 Cumulative atom usage frequency from SPNIP-M experimental design for Houston (Casas et al. in submission)

all experimental runs. Atom density increases from left-to-right in each figure, with the three atom levels being low, med, and high.

Table 16 provides the statistical results from the NBR for all three networks under all three atom density sets. Again, we assume here that only one type of sensor is being allocated within the region. The four variables examined included *Coverage* (the total network distance covered by a sensor located at a given atom), *Count* (the number of arcs covered by locating a sensor at an atom), *Min Dist to Origin* (the distance from any given atom to its closest origin) and *Min Dist to Target* (the distance from any given atom to its closest target).

Highlighting the main points of interest (detailed discussion can be found in Casas et al. (in submission)), *Min Dist to Origin* plays an important role in determining whether an atom is used frequently in the various SPNIP-M experimental design cases. The closer an atom is to an identified origin, the higher its usage in SPNIP-M solutions. The NBR analysis also shows that *Min Dist to Target* is the least influential of the four variables examined. Simply stated, a defender in SPNIP-M creates a more influential detection system when they focus on early detection rather than target-specific sensor allocation. If the defender focuses on covering critical infrastructure independently, there is a "one-to-one" effect while a defender focusing on the coverage of origins experiences a "one-to-many" impact (many destinations are potentially covered or protected by the allocation of a single sensor). Additionally, the coverage capability of a sensor (*Count* in Table 16) is a quintessential factor in determine sensor location. The frequency of atom use is positively related to the number of arcs a sensor placed at that atom covers.

This last relationship is used to discuss the final interdiction-related problem of this chapter. Knowing that *Count* acts an indicator of atom usage, three strategic atom allocation schemes will be examined for New York City, Boston and Houston. These schemes will be developed using network topology and connectivity to create the atom set that guides sensor location. Using these intelligently created atom sets, the SPNIP-M optimal solutions will be examined and compared to the standard grid-like atom sets used in the previous analysis. The goal of this analysis is to determine

Table 16 Regression results for NYC, Boston and Houston at each atom density level

	New York City			Boston			Houston			
	Estimate	Std error	Pr (>|z|)	Estimate	Std error	Pr (>|z|)	Estimate	Std error	Pr (>|z|)	
(Intercept)	3.006	1.086	0.00564	3.11E+00	1.21E+00	0.009902	−0.4703	1.652	0.7759	Low density
Coverage	1.331e−0.6	6.88E−07	0.05299	2.37E−06	1.42E−06	0.094174	6.83E−07	1.03E−07	3.51E−11	
Count	0.009112	0.4175	0.98259	−2.11E−02	2.89E−01	0.941637	−0.1295	0.06834	0.0581	
Min dist to origin	−0.000233	8.50E−05	0.00606	−1.00E−03	2.75E−04	0.000271	−1E−06	3.12E−05	0.9744	
Min dist to target	−0.000217	9.48E−05	0.02214	−2.78E−04	1.75E−04	0.113408	−4.5E−05	2.79E−05	0.1036	
(Intercept)	2.86	1.12E+00	0.0103	2.32E+00	1.24E+00	0.06213	1.438	1.868	0.44116	Med density
Coverage	1.462e−0.5	2.17E−06	1.63E−12	2.68E−05	5.81E−06	3.99E−06	2.6E−06	3.67E−07	1.3E−12	
Count	−1.426	5.24E−01	0.006499	−1.44E+00	4.98E−01	0.00394	−0.2472	0.1274	0.0524	
Min dist to origin	−3.000412	1.08E−04	0.000133	−801E−04	3.33E−04	0.0163	−3.8E−05	3.56E−05	0.28488	
Min dist to target	−0.252	1.01E−04	0.013031	−5.77E−04	1.77E−04	0.00112	−0.00011	3.74E−05	0.00221	
(Intercept)	−1.052	1.05E+00	0.31685	−1.19E+00	8.63E−01	0.1679	−0.7876	1.518	0.6038	High density
Coverage	0.0000377	6.44E−06	4.85E−09	1.32E−04	1.84E−05	7.58E−13	5.92E−06	6.06E−07	<2e−16	
Count	2.897	6.64E−01	1.28E−05	−3.94E+00	1.01E+00	9.14E−05	−0.336	0.153	0.0281	
Min dist to origin	−0.000276	1.05E−04	0.00893	−8.07E−04	3.24E−04	0.0128	−1.2E−05	2.86E−05	0.683	
Min dist to target	−8.29E−05	1.01E−04	0.41022				−7E−05	2.74E−05	0.0104	

the effect that atom location sets have on optimal SPNIP-M sensor locations and to evaluate computational and spatial trade-offs in determining whether the additional fidelity obtained from intelligent atom locations provides significantly improved or diverse defender sensor schemes.

Intelligently Locating Atoms

The problem of locating atoms intelligently is motivated by the observation that not all networks are created equal and that not all uniform, grid-based atom allocations are capable of providing an adequate set of sensor location points for network interdiction. The hypothesis of intelligent atom design is that atom allocations that are derived based on individual network properties will provide higher fidelity, more accurate solutions to network interdiction problems. Given that network structure is relatively unique, devising intelligent atom sets will give more sensor location options in areas that are denser, or which have a higher concentration of arcs with low initial detection probabilities.

Observing that analyzers have different interests in network features, we develop three methods to add atoms intelligently. For the network interdiction problem, methods are based on the initial arc non-detection values, the density of arcs in a pre-defined space and the number of arcs in a pre-defined space. The algorithmic approach to creating these atom structures is executed using ESRI ArcGIS 10 and is now provided. Figure 25 illustrates algorithmic implementation for the Boston network.

Intelligent atom algorithm

1. Determine the geographic area the network occupies (x, y or latitude, longitude coordinates).
2. Build an initial grid-based structure (user decided initial grid size).
3. For each grid, calculate the {average arc non-detection, arc density, number of arcs}.
4. If the grid value exceeds the decision threshold, further decompose grid into quarters.
5. Stopping criteria.
   (a) If the iteration number is pre-determined and has been reached, STOP.
   (b) If the iteration number is not pre-determined and no existing grids require decomposition, STOP.
   (c) If at least one grid was decomposed in Step 4, return to Step 3.

Figure 25d illustrates the final atom set for Boston using the number of arcs within a grid as the decomposition measure. Locating atoms intelligently, the atom set is clearly more contoured to the individual structure of the Boston network. To examine whether the increased complexity of an intelligent atom set is useful from a modeling standpoint, an experimental design similar to those previously discussed is invoked. The individual solutions from the experimental design runs are aggregated

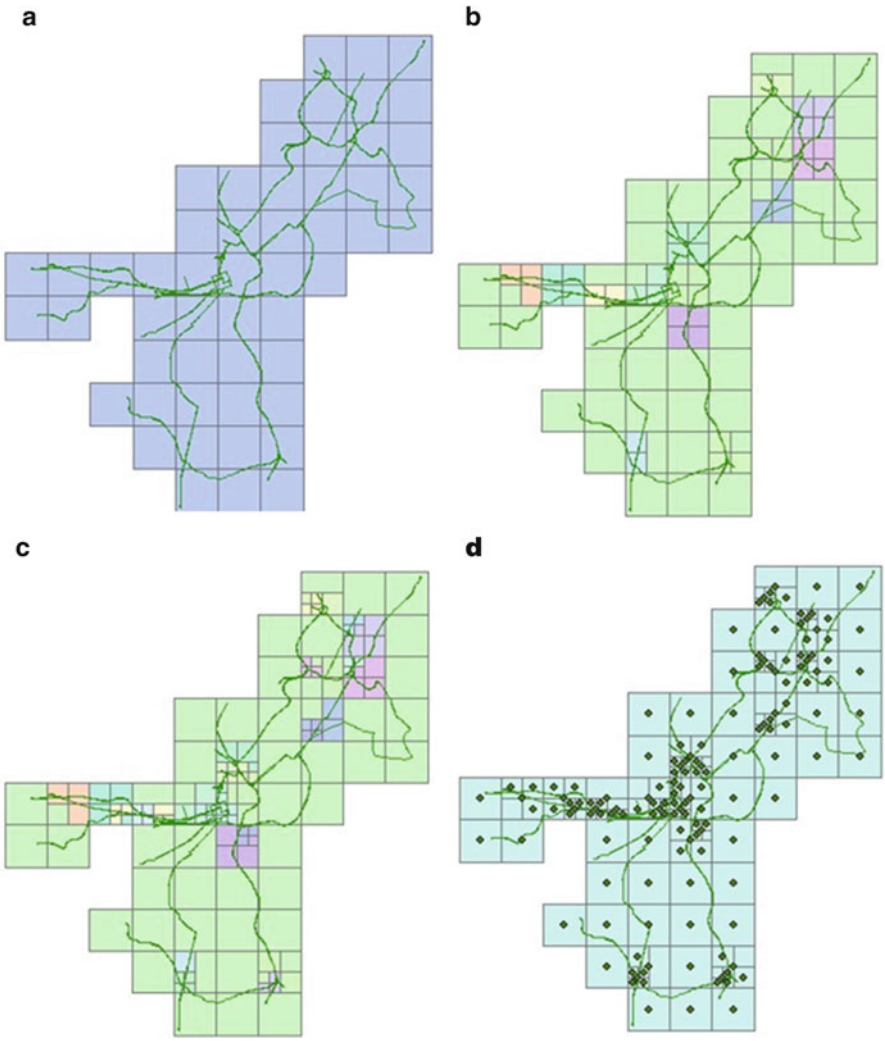


Fig. 25 Using the arc number to intelligently locate atoms

and used to obtain average computational information for road networks in New York City (NYC), Washington DC (DC), Boston, MA and Houston, TX. In total six atom allocations were considered and compared. They are: low resolution grid-based (low), med resolution grid-based (med), high resolution grid-based (high), arc length intelligent (length), arc number intelligent (number) and arc non-detection intelligent (non-det). For each network, 15 randomly generated origin-target pairs were identified with each origin and target being pulled from a pre-defined origin and destination set (origin and target size for each of the 15 pairs was also randomly generated using a uniform distribution between one and the corresponding

Table 17 Better caption

		Grid allocations			Intelligent allocations		
		Low	Med	High	Length	Number	Vulnerability
NYC	Objective	0.23	0.21	0.23	0.21	0.22	0.22
	No. paths	9.79	10.98	8.17	11.32	11.45	10.92
	Sol. time	28.70	76.40	39.05	748.00	717.00	753.00
DC	Objective	0.16	0.18	0.22	0.06	0.06	0.06
	No. paths	30.48	21.51	14.12	11.10	10.06	11.26
	Sol. time	269.70	268.30	373.60	877.00	572.00	2, 188.00
Boston	Objective	0.18	0.19	0.18	0.13	0.14	0.13
	No. paths	12.56	10.60	10.73	21.97	22.37	25.68
	Sol. time	27.99	25.58	27.31	216.60	205.90	326.00
Houston	Objective	0.11	0.24	0.20	0.14	0.12	0.11
	No. paths	18.35	19.11	17.12	14.06	15.84	13.02
	Sol. time	584.70	281.40	646.00	402.00	1, 431.00	103.30

origin/target set size). Four budget levels enabled the allocation of 4, 6, 8 and 10 sensors within the region. The computational results from this experimental design are given in Table 17, with each experimental design case being solved to optimality using Bender's Decomposition.

From Table 17, it can be shown that the computational comparison between the standard grid-based atom sets and intelligent atom sets is inconclusive at best. In certain networks like NYC and Boston, there is little different in the optimal objective values of grid and intelligent atom sets but a great gain in solution time. Though the actual number of Bender's Decomposition iterations does not increase dramatically, the intelligent atom allocations create a substantial rise in defender sub-problem solution times. With intelligent atom location, many individual atoms will have similar coverage schemes, especially in SPNIP-M where coverage is binary and the actual length of coverage is not counted. This similarity extends the amount of branching required to solve the defender sub-problem, increasing its solution time. For the DC and Houston networks, Table 17 shows that there is more significant dissimilarity between objective function values, though the intelligent atom allocations actually reduce the number of Bender's Decomposition iterations. In all, it appears that there may be some usefulness to an intelligent atom design, though in those network cases examined to date, the increased computational effort does not appear to be worth the limited fidelity gained over a grid-based approach.

Conclusions

In this chapter, we began by discussing the hazardous materials transportation problem as it is addressed in optimization. We saw how traditional hazardous

materials modeling in the 1970s and 1980s transitioned towards and motivated development of the network interdiction problem in the mid-1990s and early 2000s. Focusing on the network interdiction problem, discussion was provided on the standard shortest path network interdiction formulation and two initial variations (one modified and one length-based shortest path network interdiction problem). In the modified problem (SPNIP-M), we saw how sensor location could be separated from the network, with sensors instead located at geographic points called atoms. Examples were provided on the computational and spatial performance of SPNIP-M and SPNIP-LB, with various measures of spatial similarity introduced to compare and contrast the solutions of these models.

In the later part of this chapter, we used the obtained knowledge from SPNIP-M and SPNIP-LB to develop approximations with the goal of reducing computation time while maintaining a similar spatial distribution of sensors. We saw how the max flow-min cut theorem could be used to motivate development of a constrained knapsack approximation. This approximation was able to maintain consistent computational solution times across each of the two networks examined while reasonably replicating spatial sensor locations. In addition to the knapsack model, we pursued a spatially-based regression study which used a detailed experimental design to statistically identify two spatial properties which were shown to be pertinent factors in allocating defender resources. Lastly, used one of these properties (the number of arcs within a sensor's range) to motivate development of an algorithm to strategically determine potential sensor locations for a given network. Computational results from this intelligent atom location show that the more simplistic grid-based solutions to provide a strong base-line for sensor allocation strategies with increased fidelity in defender sensor solutions from the intelligent atom allocations coming at a steep computational price.

Acknowledgements Special acknowledgement is given to Christopher Hill and Nannan Chen, who contributed in the development and analysis of select sections in this work while students at Texas A&M University.

References

Anselin L (1995) Local indicators of spatial association – LISA. Geogr Anal 27:93–115

Bard J (1998) Practical bilevel optimization; algorithms and applications. Kluwer Academic Publishers, Boston

Batta R, Chiu S (1998) Optimal obnoxious paths on a network: transportation of hazardous materials. Oper Res 36(1)

Brown G et al (2006) Defending critical infrastructure. Interfaces 36(6):530–544

Burch C et al (2003) A decomposition-based approximation for network inhibition. Network interdiction and stochastic programming. Kluwer, Boston, pp 51–69

Casas et al. (Louisiana Tech University, 2012, unpublished)

U.S. Census Bureau (2008) TIGER/line and TIGER-related products. http://www.census.gov/geo/www/tiger/. Accessed 1 Aug 2008

Church R, Scaparra M (2007) Protecting critical assets: the r-interdiction median problem with fortification. Geogr Anal 39:129–146
Church R, Scaparra M, Middleton S (2004) Identifying critical infrastructure: the median and covering facility interdiction problems. Ann Assoc Am Geogr 94(3):491–502
Corley HW, Chang H (1974) Finding the n most vital nodes in a flow network. Manage Sci 21:362–364
Cova T, Conger S (2003) Transportation hazards. In: Kutz M (ed) Handbook of transportation engineering. McGraw Hill, New York, pp 17.1–17.24
Cunningham WH (1985) Optimal attack and reinforcement of a network. J Assoc Comput Mach 32:549–561
Department of Homeland Security (2009) Critical infrastructure and key resources. http://www.dhs.gov/files/programs/gc_1189168948944.shtm. Accessed 7 Jul 2009
Erkut E, Verter V (1998) Modeling of transport risk for hazardous materials. Oper Res 46(5)
ESRI (2009) ArcGIS: a complete integrated system. http://www.esri.com/software/arcgis/index.html. Accessed 10 Feb 2010
Federal Emergency Management Agency (2006) Software program for estimating potential losses from disasters. http://www.fema.gov/plan/prevent/hazus. Accessed 25 May 2006
Field M (2004) Highway security and terrorism. Rev Policy Res 21(3):71–91
Ghare P, Montgomery D, Turner WC (1971) Optimal interdiction policy for a flow network. Nav Res Logist Q 18:37–45
Gopalan R et al (1990) Modeling equity of risk in the transportation of hazardous materials. Oper Res 38(6)
Grubesic T, Murray A (2006) Vital nodes, interconnected infrastructures, and the geographies of network survivability. Ann Assoc Am Geograph 96(1):64–83
Hilbe JM (2007) Negative binomial regerssion. Cambridge University, Cambridge
Huang B, Cheu RL, Liew YS (2004) GIS and genetic algorithms for HazMat route planning with security considerations. Int J Geogr Inf Sci 18(8):769–787
Israeli E, Wood K (2002) Shortest-path network interdiction. Networks 40(2):97–111
Jenelius E, Petersen T, Mattsson LG (2006) Importance and exposure in road network vulnerability analysis. Transp Res A 40:537–560
Jin H, Batta R (1996) On the analysis of two new models for transporting hazardous materials. Oper Res 44(5)
Jin H, Batta R, Karwan M (1996) On the analysis of two new models for transporting hazardous materials. Oper Res 44(5)
Kara B, Verter V (2004) Designing a road network for hazardous materials transportation. Transp Sci 38(2):188–196
Laefer D, Pradhan A (2006) Evacuation route selection based on tree-based hazards using light detection and ranging and GIS. J Transp Eng 132(4):312–320
Lim C, Smith JC (2007) Algorithms for discrete and continuous multicommodity flow network interdiction problems. IIE Trans 39:15–26
List G et al (1991) Modeling and analysis for hazardous materials transportation: risk analysis, routing/scheduling and facility location. Transp Sci 25(2)
Luedtke J, White C (2002) HazMat transportation and security: survey and directions for future research. Department of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA
Matisziw T, Murray A, Grubesic T (2007) Bounding network interdiction vulnerability through cut-set identification. Adv Spat Sci 61:243–255
Matisziw T, Murray A (2009) Modeling s-t path availability to support disaster vulnerability assessment of network infrastructure. Comput Oper Res 36:16–26
MCEER (2006) REDARS 2 methodology and software for seismic risk analysis of highway systems. http://mceer.buffalo.edu. Accessed 25 May 2006
Morton D, Pan F, Saeger K (2007) Models for nuclear smuggling interdiction. IIE Trans 39:3–14
Murray A, Matisziw T, Grubesic T (2007) Critical network infrastructure analysis: interdiction and system flow. J Geogr Syst 9:103–117

National Highway Institute (1996) Highway routing of hazardous materials, guidelines for applying criteria. The Institute, USA, p 123
Nemhauser G, Wolsey L (1999) Integer and combinatorial optimization. Wiley-Interscience, New York
Phillips (1993) The network inhibition problem. Proceedings of the 25th annual ACM symposium on the theory of computing (STOC), May 1993, pp 776–785
Przemieniecki JS (2000) Mathematical methods in defense analyses. AIAA Education Series, Reston, Virginia
Qiao J et al (2007) Allocating security resources to a water supply network. IIE Trans 39:95–109
Ratliff HD, Sicilia G, Lubore S (1975) Finding the n most vital links in flow networks. Manage Sci 21:631–639
Scaparra M, Church R (2008) A bilevel mixed-integer program for critical infrastructure protection planning. Comput Oper Res 35:1905–1923
Southworth F (2008) Multi-criteria sensor placement for emergency response. Appl Spatial Anal 1:37–58
Texas Transportation Institute (2011) The 2011 urban mobility report. Texas A&M University, College Station
Transportation Research Board (2002) Deterrence, protection, and preparation: the new transportation security imperative – special report 270
Transportation Security Administration (2012) Highway and motor carrier division. http://www.tsa.gov/what_we_do/tsnm/highway/index.shtm. Accessed 10 Jan 2012
US Department of Transportation (2006) Biennial report on hazardous materials transportation. Washington, DC. Report (obtained online). Accessed 12 Aug 2012
US Department of Transportation (2012) PHMSA: pipeline and hazardous materials safety administration. http://www.phmsa.dot.gov/. Accessed 10 Jan 2012
Wein L, Atkinson M (2007) The last line of defense: designing radiation detection-interdiction systems to protect cities from a nuclear terrorist attack. IEEE Trans Nucl Sci 54:654–669
Wollmer R (1964) Removing arcs from a network. Oper Res 12:934–940
Wood K (1993) Deterministic network interdiction. Math Comput Model 17(2):1–18
Wood RK (2003) Deterministic network interdiction. Math Comput Model 17(2):1–18
Yamada I, Thill JC (2010) Local indicators of network-constrained clusters in spatial patterns represented by a link attribute. Ann Assoc Am Geogr 100(2):269–285
Yates J, Sanjeevi S (2013) A length-based multiple resource type approach to shortest path network interdiction. *Int J Crit Infrastruct Protect* (Spring)
Yates J, Wang X, Chen N (2013) Assessing k-Shortest approximation accuracy for shortest path network interdiction. Optimiz Eng (Summer)
Yates J, Casas I (2010) Role of spatial data in the protection of critical infrastructure and homeland defense. Appl Spat Anal Policy 5:1–18
Yates J, Lakshmanan K (2011) A constrained binary knapsack approximation for shortest path network interdiction. Comput Ind Eng 61:981–992
Yates J, Sanjeevi S (2012) Assessing the impact of vulnerability modeling in the protection of critical infrastructure. J Geogr Syst 14:415–435
Yen J (1971) Finding the k shortest loopless paths in a network. Manage Sci 17(11):712–716

Optimal Emergency Resources Deployment Under a Terrorist Threat: The Hazmat Case and Beyond

Rodrigo A. Garrido

Introduction

Most logistics systems are designed to operate under standard conditions, i.e., when the transportation and communication networks are fully operative, the suppliers are able to deliver what they are asked for and the demand patterns fluctuate within (somewhat) known boundaries, as well as availability of human resources and vehicles to distribute products and services from production sites to consumption points. Even under this scenario, the logistics strategy and operations are rather complex tasks both for the size of the instances to be solved and for the type of models to be solved when trying to optimize sensible variables such as inventory levels or sequences of vehicles' stops under time constraints. However, this already complex situation becomes much more complicated when there is uncertainty in the systems components. That is the case with the logistics of emergencies right after a natural or human-made disaster. Typically, there will be victims that need prompt attention in several dimensions: health care, food, water, safety, and childcare among others. However, the means to deliver the assistance may have been severely damaged by the disaster and hence the standard supply plans of the various industries involved in the provision of these goods, services and human resources do not hold in the disastrous scenario. Additionally, improvised logistics plans to reach the affected area have proven to fail due to the lack of information and preparedness (see for example Holguin-Veras et al. 2007). Consequently, there is the need for designing strategic logistics plans to cope with the consequences of catastrophic events in advance of their occurrence. These plans should be capable of dealing with shortages in networks capacity, lack of transportation services and possible stock outs of the main urgent supplies. This chapter deals with the logistics of emergency

R.A. Garrido (✉)
Dean Facultad de Ingeniería, Universidad Diego Portales, Chile
e-mail: rodrigo.garrido@udp.cl

R. Batta and C. Kwon (eds.), *Handbook of OR/MS Models in Hazardous Materials Transportation*, International Series in Operations Research & Management Science 193,
DOI 10.1007/978-1-4614-6794-6_8,

systems in the context of human-made catastrophies, which we will call terrorist attacks. The attacks may take several forms, one of those is the use of vehicles transporting a load that might be potentially used as a weapon. Examples of these are all the goods categorized as hazardous materials (hazmat) or byproducts categorized as hazardous waste (hazwaste). These type of loads can be easily transported to a strategic destination to cause damage to the people or property. A special case in these categories is the fuel, which can always pose a threat by just being within the fuel tank of a vehicle (e.g. the airplanes that crashed into the WTC towers on September 11, 2001 were both almost full of fuel, which increased the proportions of the disaster).

The problem of hazmat (or hazwaste) vehicles as potential weapons has attracted considerable attention after the events of September 11, 2001. See for example Nune (2007), Murray-Tuite (2008), Murray-Tuite and Fei (2010), Dadkar et al. (2010), Szyliowicz (2012), and Lepofsky (2012). However, once a hazmat vehicle is identified as a terrorist threat, the problem of what to do with it remains as a topic not thoroughly studied yet. In fact, regardless of the outcome of the attack attempt, the authorities should send specialized teams not only to deter or contain the attack but also to assist possible victims in the case that the attack does take place. Thus, there is an interaction between attackers, defenders and other actors of the system that deserves to be studied in detail.

The problem of interaction between authorities trying to control terrorist's attacks and terrorists trying to reach their goals has been studied from a variety of perspectives and disciplines. It is convenient to tackle this problem quantitatively from the strategic aspect of behavior and equilibrium between attackers and defenders. This approach has been followed by some authors in search for the impact of several actions of both parties in the final outcome (see for example Sandler and Siqueira 2009; Siqueira and Sandler 2008, 2010; Siqueira et al. 2009; Clauset et al. 2007; Arce and Sandler 2003). Usually, the conflicts between terrorists and authorities exhibit asymmetric information. The authority is uncertain (to some extent) about the level of resources available to the terrorists to plan their attacks, but also the terrorists are unsure about the exact level of resources dedicated to block their actions; in addition, neither of these players know with certainty the location of events to deploy counterattacking resources or the extent of resources allocated for protection in different locations. Some studies have modeled this interaction as a signaling game where the magnitude of terrorist attacks signals (i.e. becomes a proxy) the terrorist's level of resources (see Overgaard 1993). The asymmetric nature of this type of information gives an incentive to the terrorists to believe that they have a larger amount of resources than they really do, to affect the retaliation decisions.

From the perspective of the terrorists decisions, Murray-Tuite et al. (2007) and Garrido (2010) studied the problem of a malicious entity that hijacks a hazmat vehicle from its normal path and immediately transports the substances to a target. They presented a methodology for determining the malicious entity's set of pareto optimal paths from the hijacking point to a given target, based on the competing objectives of minimizing distance and maximizing consequence, conditioned on the

probability of being intercepted by law enforcement prior to reach the target. With this knowledge, the authority may tell apart those paths that are most likely used by regular drivers from those that are likely to be used by a terrorist. However, the information about probabilities of an attack and/or capture of the perpetrators is not perfect and is based on perceptual parameters on both sides; therefore the set of paths computed by each party is likely to be different. Nevertheless, the identification of a vehicle following a path that belongs to a potential terrorists pareto solution, acts as a red flag to the authority to double-check the status of that vehicle to identify whether or not it has become an actual threat. In most cases, this information provide a basis more accurate than other methodologies to estimate attack probabilities for general cases of terrorists attacks. The fact that hazmats are transported in special vehicles and the authority is informed about their movements gives plenty of information to be used to derive a probability of attack, which does not exist in many other situations.

Once a hazmat threat is positively identified, the authority must act quickly to avoid a significant impact or else to assist the victims in the shortest possible time. Depending on the type of hazmat and the magnitude of the attack, there are different types of teams to be sent to the affected location. For instance, if the threat is originated from a vehicle containing a toxic load, the authorities will probably send not only law enforcement teams but also paramedics and hazmat crews; whereas a vehicle with an explosive load may need a different assembly of human resources to be sent.

The (broad) concept studied in this chapter departs from the assumption that a hazmat threat has been positively identified and consequently, the authority must allocate resources in a given area in such a way that any point with a potential threat can be reached in the minimum time interval within a given budget. Throughout the chapter, different cases and scenarios will be analysed and their corresponding models will be presented.

The Context

Terrorism risk shares some characteristics with other types of risk in terms of correlations in time or space, periodicity, level of damage, etc. In particular they have commonalities with complex engineering systems, such as reliability engineering, control systems, among others. Many risk analyses in those fields rely on decision and/or probability trees to deal with risk. These trees show all the possible events branching down possible courses of action to represent all possible outcomes of each event, along with the outcome's probabilities associated to each branch. However, terrorist attacks bear two particular features that make them different from the rest: unlike natural disasters they feature human intelligence and unlike industrial disasters they feature human intent (Major 2002). Therefore, the analysis of terrorism risk should rely on complex operations research methods

and robust theories about the behavior of each one of the actors involved. In the case of hazmats transportation, both situations may occur, i.e., a spill may be due to a traffic accident (which can be categorized as a natural disaster) but also a terrorist attack can be started using a hazmat vehicle.

The fact that terrorist attacks have a significant amplification effect on the public perception of risk makes the problem harder to tackle for the corresponding authorities. In fact, regardless of the low probabilities associated to these events, the perception and reaction of the regular citizen seems to adhere to a phenomenon called *probability neglect* (Sunstein 2003) in which people often weigh the risks of terrorism significantly higher than those they confront in ordinary life. The practical consequence of this phenomenon is that the authorities often respond to probability neglect with disproportioned actions and regulations that may be unjustified or in some cases even counterproductive. Therefore, there is a need for the authorities to take actions with the aim of reducing both actual and unjustified public fear, assuring that the benefits of the response can outweigh the corresponding costs.

There are a great number of terrorism-related scenarios for which an authority could aim its efforts. One such scenario is an attack with catastrophic consequences (i.e., an event whose severity is significantly larger than that of a regular criminal attack) on one of many possible targets. This scenario is sufficiently general to cover a wide range of possibilities and yet allows the modeling of optimal decisions within tractable limits. This scenario is the one studied in this chapter, assuming that the authorities budget is entirely available to fight terrorist attacks (other authors consider scenarios where the budget is shared between terrorism and natural disasters; see for example Zhuang and Bier 2007). The basic assumptions are the following:

1. There are a defender and an unknown number of attackers.
2. Both actors are rational and seek to optimize certain objective functions.
3. Neither have infinite resources.
4. Defender's budget is entirely available to fight terrorism.
5. Defender's budget is greater than that of the attacker's.
6. Both the target, timing and severity of an attack are unknown to the defender.

Note that the assumptions are general enough to cover not only terrorist attacks coming from a hazmat situation but also more general types of attack.

Problem Description

Some Characteristics of a Terrorist Attack

Terrorists attacks are always traumatic events that obtain considerable media attention, regardless of their actual consequences. These attacks may take several different forms but their ultimate goal is to cause significant damage to either people

or property in order to generate a sense of terror among a certain target. Perhaps the most widely known example of a vehicle transporting a hazmat involved a terrorist attack was the case of the airplanes deliberately crashed into the towers of the World Trade Center in New York City, in September 11, 2001, killing all the passengers aboard (including the hijackers). We may consider these airplanes as hazmat vehicle since they were almost full of fuel (although the damage was not attributable only to the fuel). These attacks had the intention not only to kill people but also spread the sense of terror among the US population (the significance of the numbers and the towers themselves served that purpose). Thus, there exists a variety of types of attacks with different degrees of potential damage and different characteristics that makes them hard to forecast and prevent. Although the general perception might be that it is impossible to predict these events, in practice there are some situations in which the forecasting of (at least) some parameters of an attack can be reasonably done with the aid of operations research techniques. In addition to the literature previously cited above, regarding the probability of a hazmat vehicle as a potential weapon, for more general cases there are some articles that study the computation of these probabilities. In fact, Clauset et al. (2007) show that there are significant temporal correlation in frequencies of severe attacks. More so, they proved, as a robust feature of terrorists attacks, that the frequency of these attacks exhibits the phenomenon of *scale invariance*, i.e., the frequency scales as the inverse power of the severity. After the occurrence of a terrorist attack, various negative effects may impact the population and/or environment in the surrounding area and hence the authority must decide what action to take in order to minimize those negative impacts. However, not only the time and place of these events is unknown but also the severity and consequently the type of assistance that will be needed after the event. Therefore, a working hypothesis in this chapter is that it is possible to estimate probabilities of the occurrence of a terrorist attack and its intensity through exogenous econometric models based on observations of past events; in some cases like hazmat transportation, these probabilities might be easier or more accurate to estimate than other cases but the departing point will be the existence of some measure of these probabilities.

From the tactical perspective, the authority should aim to allocate sufficient human and material resources to satisfy the demand for them after an event, while at the same time keeping the system's cost within the available budget. That means that the deployment of personnel and supplies at certain locations should guarantee the satisfaction of demand to a certain confidence level, i.e., to an acceptable degree of failure. For instance, a desired objective could be to find the levels of resource allocation that assures that the teams will reach any affected location within a specified time in at least 95% of the cases, and this level must be met with the available budget.

The uncertainty in the time, place and intensity of a terrorist attack makes the demand function to change quickly from zero to a high level of resource necessity (in time and space). This feature is typical in the demand for emergency services but

in this case it may be more abrupt due to the severity and possible media coverage that a terrorist attack attracts when compared to other emergencies situations. For example, in a certain town during a normal period, the demand for high energy protein biscuits or water purification systems may be zero, until a tsunami hits it and then the demand for these items might rise to a very high level. If you plotted that on a graph, the figure would be an horizontal line with no change for a long period of time until abruptly a sudden jump shows up and later on the line would keep the same original low level indefinitely. The media amplification effect can be gauged when, for example, a news is transmitted on TV warning the danger of getting close to a zone with potential harmful radiation; this news will probably deter any attempt to reach the zone with trucks with humanitarian aid, regardless of the actual level of radiation.

System's Components

Considering the characteristics of the occurrence of a terrorist attack, the following components are defined to play a key role in the system.

Emergency Teams

These are human resources (e.g. firefighters, paramedics, hazmat experts, etc.) with base in different locations spread out within an area significantly larger than any affected zone. If an event occurs, a whole team may be assembled with personnel based at different locations, therefore they must be transported from one or more emergency stations in the numbers that the emergency calls for (following a minimum cost flow program). Note that the crew costs may vary between locations as well as within the same location across time periods.

Emergency Vehicles

The transportation of emergency teams from their location to the affected points can be done by special vehicles not necessarily located at the same site that human resources are based; more so, under budget constraints there is the possibility that there are more crew bases than specialized vehicles. Each time a team is assembled, an assignment problem arises when choosing what vehicle picks them up and transport them to an affected site. Given the different nature of emergency crews, the vehicles are not necessarily homogeneous (e.g. firetrucks, ambulances, etc.).

Demand Points

They correspond to zones affected by an attack or under the risk of it. These zones are represented by nodes at which the demand for an emergency team is needed. Each node is assumed to have a known probability of being attacked at any given period during the planning horizon. Note that the assessment of these probabilities is extremely difficult not only due to the intrinsic stochasticity of the problem itself but also because it is a rather new topic within the research community and no single methodology has been adopted nor accepted to produce acceptable figures for all the cases (sensitivity analysis emerges as a natural way to cope with this uncertainty). A good description of many efforts along this line of research can be found in Ezell et al. (2010).

Simplifying the Problem

Both the time horizon and the geographic scope are defined a priori by the decision maker. The geographic area is subdivided into smaller zones, which can be represented by nodes in a network. The time horizon is divided into discrete periods. The need for assistance is a derived demand generated by the terrorist attack/threat which can be forecasted on a spatio-temporal grid. For instance, a hazmat truck that deviates from its original expected path alerts the authorities of a potential problem; the works by Nune (2007) and Murray-Tuite et al. (2007) elaborate on this aspect. For other general cases not related to hazmat, a real life example that deserves attention is that of RAND Corporation (the Santa Monica think tank) contributed with the company Risk Management Solutions (RMS) from San Francisco to develop a mathematical model of potential terrorist actions in the US territory (Willis et al. 2005). Consequently, it is assumed that in one of the regions of the specified geographic partition, at a certain time period a terrorist attack/threat may occur with an estimated probability. This probability is exogenous to the model and it is assumed to be predictable. Thus, once the event strikes a populated area, there is an immediate generation of demand for a specific type of emergency team (e.g. firefighters, paramedics). Then, the crew size and composition will change randomly according to type of event, its severity and scope of the emergency.

Notation

The following notation will be used throughout the chapter:

Parameters and Sets

p_i^t	Probability that a threat is identified at location i during period t.
$\rho_{i,j}^{t_a,t_b}$	Correlation between $p_i^{t_a}$ and $p_j^{t_b}$, for locations i and j, and periods t_a and t_b.
N	Set of locations within the area of interest (set of network's nodes).
Π	Set of types of emergency teams (e.g. law enforcement and hazmat teams)
T	Number of periods within the planning horizon.
Ψ	Set of all the periods.
Φ	Set of vehicles' classes.
θ	Value of time, i.e., the monetary assessment of a time unit.
D_{ci}^t	Demand for emergency resources of type c, in location i, during period t.
w_{cv}	Compatibility matrix between crew and vehicle's class. Typical element (c, v)takes value 1if crew of type c, can be transported on the vehicle class $v$and 0otherwise.
V_{vi}^t	Number of vehicles of class v, available at location i, at the beginning of period t.
L_{cj}	Capacity to accommodate crew members of type c at location j.
tt_{cjk}^t	Travel time of crew type c, to the location k, from its base in location j during period t.
ca_{cj}^t	Cost of allocating a member of a crew of type c, in a base located in j, during period t.
cd_{cjk}^t	Cost of dispatching a member of a crew of type c, in a base located in j, to the location $k$during period t.
B_D	Budget available to the defender
B_A	Budget available to the attacker

Decision Variables

The following are the modeling variables:

x_{cjk}^t	Flow of crew members of type c, from base located at j sent to location k, during period t.
y_{vij}^t	Flow of vehicles of class v, relocated from location i to location j, during period t.
AV_{cj}^t	Availability of crew members of type c, in base located in j, at the beginning of period t.
A_i	Aggregated resources allocated by the attacker to location i.
D_i	Aggregated resources allocated by the defender to location i.
μ	Lagrange multiplier.

Different Cases of Terrorist Attacks

Case 1: Allocating Resources on the Basis of an Expected Event

One of the difficulties in developing an investment plan to control terrorist attacks is the uncertainty, both in time and space, regarding the occurrences of events. In fact, the authority must allocate its scarce resources to minimize the potential damage to people and property, before an event takes place. Thus, the following problem definition tackles the issue of the different times at which decisions must be made for the pre-positioning of resources and the posterior assignment of those resources, once the event actually takes place. There are two stages in this problem. First, the authority decides where to locate and allocate human and material resources. Second, once a terrorist attack takes place, the same authority must decide what type of resources and from what location will be sent to the affected area (e.g. paramedics and firefighters). Clearly both decisions are sequential and interdependent.

Problem Statement

There are a set of n possible targets $S = \{1, 2, \ldots, n\}$, spread over a network connecting each zone. The defender has limited information about the attack, which makes it difficult to compute the defenders' objective function. In fact, the defender will allocate human and material resources following some forecast information about the place and type of attack and then when the attack materializes the defender will assign resources from various sources to assist the affected points. It is clear then that both events do not happen simultaneously and hence their translation to costs is not direct.

The problem can be modeled assuming a first stage in which the cost of allocated human and material resources can be represented by the following expression:

$$\sum_{t \in \Psi} \sum_{c \in \Pi} \sum_{j \in N} ca_{cj}^{t} AV_{cj}^{t} \tag{1}$$

When the attack does take place, the problem consists of finding a way to send assistance as quickly as possible from a set of points previously defined in the first stage. This problem corresponds to a well-known two-stage stochastic program, extensively studied in the mathematical literature (see for example Bienstock and Shapiro 1998; Ahmed and Shapiro 2002; Shapiro and Homem-de-Mello 1998).

The standard use of these two-stage stochastic programs is to minimize an objective function compound by two different terms. One is the actual cost of a decision made in the first stage and the other term represents the expected value of the costs that would be derived from the realization of the random action corresponding to the second stage.

In this case the objective function would take the following form:

$$Min \sum_{t\in\Psi}\sum_{c\in\Pi}\sum_{j\in N} \lambda ca_{cj}^{t} AV_{cj}^{t} + (1-\lambda)\,\theta E\left[\Gamma\left(x, D\right)\right] \tag{2}$$

where λ is a non-negative weight parameter. Note that the two parts of this expression do not have the same dimension and hence the parameter θ corrects for this effect. The expected value $E\left[\Gamma\left(x, D\right)\right]$ corresponds to the solution of the following problem:

$$E\left[\Gamma\left(x, D\right)\right] = Min \sum_{t\in\Psi}\sum_{c\in\Pi}\sum_{j\in N}\sum_{k\in\Phi} tt_{cjk}^{t} x_{cjk}^{t} \tag{3}$$

$$\sum_{j\in N} x_{cjk}^{t} = D_{ck}^{t} \quad \forall k \in \Phi,\ c \in \Pi,\ t \in \Psi \tag{4}$$

$$\sum_{v\in\Omega}\sum_{j\in N}\sum_{k\in N} w_{cv} x_{cjk}^{t} \leq \sum_{v\in\Omega}\left(u_{v}\sum_{i\in N}\sum_{j\in N} y_{vij}^{t}\right), \quad \forall t \in \Psi,\ c \in \Pi \tag{5}$$

$$\sum_{k\in N} x_{cjk}^{t} \leq AV_{cj}^{t}, \quad \forall j \in N,\ c \in \Pi,\ t \in \Psi \tag{6}$$

$$\sum_{j\in N} y_{vij}^{t} \leq V_{vi}^{t}, \quad \forall i \in N,\ t \in \Psi,\ v \in \Omega \tag{7}$$

Therefore, the complete expression of the two-stage model is as follows:

$$Min \sum_{t\in\Psi}\sum_{c\in\Pi}\sum_{j\in N} \lambda ca_{cj}^{t} AV_{cj}^{t} + (1-\lambda)\,E\left[\Gamma\left(x, D\right)\right] \tag{8}$$

s.t.

$$AV_{cj}^{t} \geq 0, \quad AV_{cj}^{t} \leq L_{cj}, \quad \forall j \in N,\ c \in \Omega,\ t \in \Psi \tag{9}$$

Expression (3) corresponds to the total travel time needed to reach the affected locations with the emergency teams. Expression (4) is a set of all the demand satisfaction equations. The constraint set (5) restricts the available transportation capacity for each period and type of crew. The constraint set (6) controls the availability of each crew type at each location and period. The constraint set (7) ensures that the flow of vehicles does not exceed the available fleet size.

Case 2: At Least α% of the Demand Must Be Satisfied

In this case, the shortage of budget justifies a partial demand satisfaction, i.e. less than a 100% of the locations will be served. Here, again, no information is available about the value of any possible target and the probability of a successful attack is exogenous to the model.

Mathematical Formulation of Case 2

The demand for emergency resources of type c at location i, during period t is defined as follows:

$$D_{ci}^{t} = \begin{cases} d_{ci}^{t} & \text{with probability } p_i^t \\ 0 & \text{with probability } 1 - p_i^t \end{cases} \tag{10}$$

$$d_{ci}^{t} \sim F_{ci}^{t}(x) \tag{11}$$

$F_{ci}^{t}(x)$ is a probability distribution function over a real non-negative domain.

The Optimization Problem

The solution of this problem is aimed at assisting the decision maker in tactical aspects of the emergency logistics after a hazmat vehicle has been reported as hijacked or the detection of other terrorist threat . Thus, the model will answer the following questions:

1. What type of emergency team and crew size should be set at each location within the relevant network?
2. How to transport the emergency teams from their location to the threat points in minimum time intervals?

The following mixed integer programming model can be designed:

$$\min \sum_{t\in\Psi}\sum_{c\in\Pi}\sum_{j\in N}\sum_{k\in N} t t_{cjk}^{t} x_{cjk}^{t} \tag{12}$$

$$P\left\{\sum_{j\in\Phi} x_{cjk}^{t} \geq D_{ck}^{t} \quad \forall\, k \in \Phi,\ c \in \Pi,\ t \in \Psi\right\} \geq \alpha \tag{13}$$

$$\sum_{v\in\Phi}\sum_{j\in N}\sum_{k\in N} w_{cv} x_{cjk}^{t} \leq \sum_{v\in\Phi}\left(u_v \sum_{i\in N}\sum_{j\in N} y_{vij}^{t}\right) \quad t \in \Psi,\ c \in \Pi \tag{14}$$

$$\sum_{k\in N} x_{cjk}^{t} \leq AV_{cj}^{t} \quad \forall\ j \in N,\ c \in \Pi,\ t \in \Psi \tag{15}$$

$$\sum_{t\in\Psi}\sum_{c\in\Pi}\sum_{j\in N} c a_{cj}^{t} AV_{cj}^{t} + \sum_{t\in\Psi}\sum_{c\in\Pi}\sum_{j\in N}\sum_{k\in N} c d_{cjk}^{t} x_{cjk}^{t} \leq B_D \tag{16}$$

$$AV_{cj}^{t} \leq L_{cj} \quad \forall\ j \in N,\ c \in \Pi,\ t \in \Psi \tag{17}$$

$$x_{cjk}^{t}, AV_{cj}^{t} \in \mathbb{R}^{+},\ j \in N,\ k \in N,\ c \in \Pi,\ t \in \Psi \tag{18}$$

The objective function (12) represents the total travel time for the teams to get to the threat points. The set of constraints (13) ensures that the demand will be satisfied at least to an $\alpha\%$. The constraint set (14) restricts the available transportation capacity for each period and type of crew. The constraint set (15) controls the availability of each crew type at each zone and period. The constraint set (16) ensures that the cost of assignment and dispatch for all crew types, zones and periods does not exceed the total defender's budget. The constraint set (17) ensures that the capacity to accommodate crew members of each type does not exceed the assigned availability to each zone, type and period. Finally, the constraint sets (18) impose non negativity.

A Solution Approach

The optimization problem is rather difficult (it is NP-Complete) and hence is unlikely that a standard mathematical programming approach could find an exact solution within reasonable boundaries of time and computational resources. In fact, with no information about a joint probability function to characterize the set of constraints (13) it is impossible to attempt an exact solution search. Consequently, in this section, an implementation of a heuristic scheme called *Sample Average Approximation* is presented (as proposed by Pagnoncelli et al. 2009). The vector of stochastic demands are generated through Monte Carlo simulation, considering all the spatio-temporal correlations. For each generated instance, the equivalent optimization problem is solved, considering a given level of service to satisfy the demand for emergency items.

Let z_{kp}^{th} be binary variables that measure the number of times that a demand constraint is not satisfied. Thus, the following modified optimization model is defined for the generated samples.

$$\min \sum_{h=1}^{H} \sum_{t\in\Psi} \sum_{p\in\Pi} \sum_{k\in N} \sum_{j\in N} \sum_{i\in N} cv_{ijkp}^{t} x_{ijkp}^{th} + \sum_{h=1}^{H} \sum_{c\in\Omega} \sum_{j\in N} \sum_{i\in N} cl_{ijc}^{t} y_{ijc}^{th} + \sum_{h=1}^{H} \sum_{t\in\Psi} \sum_{p\in\Pi} \sum_{j\in N} ci_{jp}^{t} I_{jp}^{th} \quad (19)$$

$$\sum_{j\in N} \sum_{i\in N} x_{ijkp}^{th} + z_{kp}^{th} D_{kp}^{th} \geq D_{kp}^{th} \quad \forall\, k \in N,\ p \in \Pi,\ t \in \Psi,\ h = 1, \ldots, H \quad (20)$$

$$\sum_{p\in\Pi} \sum_{k\in N} \sum_{j\in N} w_{pc} x_{ijkp}^{th} \leq u_c \sum_{j\in N} y_{ijc}^{th} \quad \forall\, i \in N,\ t \in \Psi,\ c \in \Omega,\ h = 1, \ldots, H \quad (21)$$

$$\sum_{j \in N} y_{ijc}^{th} \leq V_{ic}^{t} \quad \forall\, i \in N,\ t \in \Psi,\ c \in \Omega,\ h = 1, \ldots, H \tag{22}$$

$$\sum_{k \in N} \sum_{i \in N} x_{ijkp}^{th} \leq I_{jp}^{th} \quad \forall\, j \in N,\ p \in \Pi,\ t \in \Psi,\ h = 1, \ldots, H \tag{23}$$

$$I_{jp}^{tn} \leq L_{jp} \quad \forall\, j \in N,\ p \in \Pi,\ t \in \Psi,\ h = 1, \ldots, H \tag{24}$$

$$\sum_{h=1}^{H} \sum_{t \in \Psi} \sum_{p \in \Pi} \sum_{k \in N} z_{kp}^{th} \leq N(1 - \gamma) \tag{25}$$

$$x_{ijkp}^{th}, I_{jp}^{th} \in \mathbb{R}^{+} \quad \forall\, i \in N,\ j \in N,\ k \in N,\ p \in \Pi,\ t \in \Psi,\ h = 1, \ldots, H \tag{26}$$

$$y_{ijc}^{th} \in \mathbb{Z}^{+} \quad \forall\, i \in N,\ j \in N, c \in \Omega,\ t \in \Psi,\ h = 1, \ldots, H \tag{27}$$

$$z_{kp}^{th} \in \{0, 1\} \quad \forall\, k \in N,\ p \in \Pi,\ t \in \Psi,\ h = 1, \ldots, H \tag{28}$$

where, the upper index h is the sample number, H is the sample size and γ is the desired level of service to solve the approximated problem. Constraint (25), allows that the number of times that the demand is not satisfied (constraint (20)), does not exceeds $1 - \gamma$.

Lower Bound

To obtain a lower bound a sample average approximation scheme is applied (see Pagnoncelli et al. 2009). The first step is to find two integer numbers M and H such that:

$$\theta_H := \sum_{i=1}^{\lfloor \gamma H \rfloor} \binom{H}{i} \alpha^i \, (1 - \alpha)^{H-i} \tag{29}$$

and L being an integer number such that:

$$\sum_{i=1}^{L-1} \binom{M}{i} \theta_H^i \, (1 - \theta_H)^{M-i} \leq 1 - \beta \tag{30}$$

Then, a set of M independent samples must be generated $D_{kp}^{t1m}, \ldots, D_{kp}^{tHm}$, $m = 1, \ldots, M$ each one of size H.

For each generated sample, the above modified optimization problem must be solved.

The optimal solution for each sample, called $\hat{\theta}_H^m, m = 1, \ldots M$, must be arranged in non decreasing order, $\hat{\theta}_H^{(1)}, \ldots, \hat{\theta}_H^{(M)}$, where $\hat{\theta}_H^{(i)}$ is the ith smallest value.

Finally, the value $\hat{\theta}_H^{(L)}$ will be a lower bound for the optimal solution of the original problem, with a significance level of at least β.

Upper Bound

To obtain an upper bound, the method put forward by Luedtke and Ahmed (2008) will be applied. One of the findings of that article is the size H of a sample to guarantee that the solution of the modified optimization problem be in fact a feasible solution for the original problem, with a significance level of β. The latter is obtained as follows:

$$H \geq \frac{2}{(\alpha-\gamma)^2} \log\left(\frac{1}{1-\beta}\right) + \frac{2m}{(\alpha-\gamma)^2} \log\left\lceil \frac{2DL}{\alpha-\gamma} \right\rceil \tag{31}$$

This result gives a theoretical guide for the search of the sample size H. However, as explained in Pagnoncelli et al. (2009) and Luedtke (2008), the problem size (with the obtained value of H) could be prohibitively large.

An alternative to this method is to solve the modified optimization problem with a smaller value of H and then check (a posteriori) the fulfillment of the stochastic constraint.

This a posteriori checking can be done by using a sample of size N', and then for the samples $D_{kp}^{t1}, \ldots, D_{kp}^{tH'}$ counting the number of times that this expression holds: $\sum_{j\in N} \sum_{i\in N} x_{ijkp}^{th} \geq D_{kp}^{th}$.

The upper bound will be the objective function value corresponding to the solution with the highest value within all the feasible solutions, once the a posteriori checking was performed (see Luedtke 2008).

Case 3: The Probability of Attack Is a Function of the Defender's Resources Allocated to Each Zone: A Stackelberg Game Approach

In this case, each target i has a value v_i and a probability of being (successfully) attacked of p_i. The defender has a total budget of B_D which must be optimally allocated among all the possible targets. The attacker's total budget is unknown to the defender, but the assumption is that whatever resources the attacker has must be optimally allocated at each possible target. Under these circumstances the lesser the resources assigned by the defender to a zone the higher the probability of an attack at that zone. Thus, the defender acts first by investing in defense at each target (some targets may have zero resources allocated) which makes it the *leader* while the attacker waits until the allocation is known (or forecasted) and then he acts as a *follower*.

The implicit assumption in this case is that the probability of being attacked at zone i is inversely proportional to its allocation of emergency resources. In other words, the more resources are deployed to protect a zone the less vulnerable it becomes:

$$\frac{\partial p(Alloc_i)}{\partial Alloc_i} \leq 0 \tag{32}$$

where $Alloc_i = \sum_{c \in \Pi} AV_{ci}$

Accordingly, the demand expression (10) needs to be modified as follows:

$$D_{ci} = \begin{cases} d_{ci} & \text{with probability} \quad p_i(Alloc_i) \\ 0 & \text{with probability} \quad 1 - p_i(Alloc_i) \end{cases} \tag{33}$$

Clearly, the probability is a function of decision variables AV_{ci} whose values are not known in advance, i.e., the values of p_i depend on AV_{ci}, which depends on p_i as well. Therefore, the problem becomes an equilibrium one. Both players have opposite objectives: minimizing and maximizing the expected loss (EL), which is the following discrete expectation:

$$EL = \sum_i v_i \, p_i(Alloc_i) \tag{34}$$

Consequently, as the probability depends only on the vulnerability of the target (i.e. the defender's resources allocated on it) the leader must allocate resources in an effort to minimize EL and then the follower will decide his actions depending on that result.

Property 1. The optimal Stackelberg resource allocation for the defender (leader) is to assign resources to each target in such a way that the value of each target weighted by the marginal effect of the investment remains constant for all available targets.

Proof. The optimization problem that is consistent with the defender's objective is the following:

$$\underset{\{Alloc_i\}}{Min} \sum_{i \in N} v_i \, p_i(Alloc_i) \tag{35}$$

s.t.

$$\sum_{i \in N} Alloc_i \leq B_D \tag{36}$$

The model (35) and (36) can be conveniently solved through the Lagrange multipliers approach:

$$\underset{\{Alloc_i\}}{Min} \sum_{i \in N} v_i \, p(Alloc_i) + \mu(B_D - \sum_{i \in N} Alloc_i) \tag{37}$$

Minimizing expression (37) with respect to $Alloc_i$ yields the following rule:

$$v_i \frac{\partial p(Alloc_i)}{\partial Alloc_i} = \mu \quad \forall i \in N \tag{38}$$

which completes the proof. □

Property 1 gives a rule to assign the defense budget optimally when all the players play Stackelberg and the probability of a successful attack depends only on the amount of resources assigned to each target. But, what would happen if that probability also depends on the resources that the attacker assigns to each target?

Case 4: The Probability of Attack Is a Function of All the Resources Allocated to Each Zone: A Finite-Dimension Variational Inequality Approach

Problem Statement and Modeling

There are a set of n possible targets $N = \{1, 2, \ldots n\}$, each one with a fixed value v_i. There are two players: a defender (e.g. the authority) and an attacker (e.g. any malicious entity which plans to attack an unknown subset of these targets). Each player has a fixed available budget (B_D and B_A respectively) to accomplish their own task. The vector (D_i,A_i) represents a resources' allocation by each player, i.e. the resources used by each player to defend/attack the value v_i of target i. The destruction of any target is a stochastic process, governed by a probability function $p_i\,(D_i, A_i)$, which represents the probability of a successful attack on target i, as a function of the resources' allocation. The vectors $A = (A_1, A_2, \ldots, A_n)$ and $D = (D_1, D_2, \ldots, D_n)$ are respectively called attack and defense *patterns*.

p_i is a smooth known function that satisfies the following conditions:

$$\frac{\partial p_i}{\partial D_i} < 0, \quad \frac{\partial p_i}{\partial A_i} > 0 \tag{39}$$

Thus, the defender aims to find the vector of resources that minimizes the expected outcome of an attack, whereas the attacker tries to find the optimal allocation of his resources to maximize the payoff of his attack.

The defender faces an expected loss at target i given by the following expression:

$$EL_i\,(D_i, A_i) = p_i\,(D_i, A_i) \times v_i, \quad \forall i \in N \tag{40}$$

Under these conditions, there are two scenarios of interest:

1. The defender tries to minimize the *maximum expected loss*. The latter means that, among all the possible targets, and their probability of a successful attack, the objective function is the worst-case scenario:

$$Min\left\{\underset{i}{Max}\left[EL_i\left(D_i, A_i\right) = p_i\left(D_i, A_i\right) \times v_i\right], \quad \forall i \in N\right\} \tag{41}$$

2. The defender tries to minimize the *total expected loss*. The latter means that, among all the possible targets, and their probability of a successful attack, the objective function is the expected value of all the possible outcomes:

$$Min\left\{TEL\left(D, A\right) = \sum_{\forall\, i \in N} p_i\left(D_i, A_i\right) \times v_i\right\} \tag{42}$$

The scenario 1 (expression (41) is analyzed in Major (2002) who establishes that under these conditions the defender should use all the resources available to this end, i.e. there is no reason to invest less than B_D. In this case, the defender will allocate her budget in a subset of targets. In fact, the minimax criterion in expression (41) implies that there will be two groups of targets: defended and undefended. Indeed, regardless of the attackers' preferences and actions, the defender must lower the highest EL among all the targets, in order to avoid leaving a target with a more attractive payoff than the others (whenever possible). Thus, as the value of EL_i decreases with the increase in the defensive resources (D_i) allocated to target i, the defender will try to decrease the value of the highest EL_i (which will change as the defender allocates her budget in different targets).

The analysis presented in Major (2002) can be further extended. Indeed, following the same logic the assignment strategy would start by investing in the target with the highest EL and would continue until a point where increasing the investment in one target would need to decrease it in another one, which would increase the EL of the latter. At this point a Nash equilibrium will be reached. In fact, at this point there will be a set of targets N' whose EL value will be exactly EL^*, after investing optimally in each one of them, and another set N'' containing the rest of the targets whose EL values are lower than EL^* and consequently no budget should be allocated to them. In fact, any target whose EL is lower than EL^* should be left undefended because a successful attack on them would yield an expected loss that is lower than that of a successful attack on any target on N'. At this point no target can benefit from a change in the defense investment without a loss in the EL of another target. Hence, the optimal defense investment in target i, D_i^*, represents a Nash equilibrium.

The latter can be mathematically expressed as follows:

$$EL_i\left(D_i, A_i\right)\left\{\begin{array}{lll} = EL^* & \text{if} & D_i^* > 0 \\ \leq EL^* & \text{if} & D_i^* = 0 \end{array}\right\} \quad \forall i \in N. \tag{43}$$

Under a resource's deployment according to (43), the attacker faces a set of targets with expected value EL^* as candidates for an attack, and another set of targets with less expected value. Thus, a rational attacker would aim to the targets with higher EL, expecting a payoff of EL^*in each one of them.

Expression (43) can be expressed as a variational inequality (see Nagurney 1999 for a general definition of this concept). To the knowledge of the author, this is the first time that this case has been presented as a variational inequality. First, the following property is needed.

Property 2. The equilibrium pattern of defense resources' allocation expressed by (43) is equivalent to the following formulation:

$$\left(EL^* - EL_i\left(D_i, A_i\right)\right) \cdot \left(D_i - D_i^*\right) \geq 0, \quad \forall D_i \geq 0 \tag{44}$$

Proof. If $D_i^* > 0 \Rightarrow EL_i\left(D_i, A_i\right) = EL^*$ and (44) holds. If $D_i^* = 0 \Rightarrow EL_i\left(D_i, A_i\right) \leq EL^*$and (44) also holds. Now we need to prove that any non-negative D^*that satisfies (44) also satisfies the equilibrium condition (43). Let's consider an arbitrary defense investment vector $D = (D_1, D_2, \ldots, D_n) \neq D^*$. There are two cases:

(a) For any$D_i^* > 0$ we have either $D_i^* - D_i < 0$ or $D_i^* - D_i > 0$, therefore Property 2 will hold only if $[EL^* - EL_i\left(D_i, A_i\right)] = 0$
(b) For $D_i^* = 0$ it follows that $D_i - D_i^* \geq 0$ and hence Property 2 yields $EL^* \geq EL_i\left(D_i, A_i\right), \quad \forall i \in N$

with the cases a and b the demonstration is complete. □

Property 3. For any attack pattern AV, the equilibrium defense pattern expressed by (43), can be expressed as the following variational inequality:

$$EL\left(D^*, A\right) \cdot \left(D^* - D\right) \geq 0, \quad \forall D \subset \Omega \tag{45}$$

where Ω denotes a feasible set defined by:

$$\Omega = \left\{ D_i \in \mathbb{R} / \exists \quad D_i \geq 0 \wedge \sum_{\forall i \in N} D_i = B_D \right\}.$$

Proof. Note that Property 2 holds for each target in N, hence the following summation is valid:

$$\sum_{\forall i \in N} \left(EL^* - EL_i\left(D_i, A_i\right)\right) \cdot \left(D_i - D_i^*\right) \geq 0$$

This may be rewritten as follows:

$$\sum_{\forall i \in S} EL^* \cdot \left(D_i - D_i^*\right) - EL_i\left(D_i, A_i\right) \cdot \left(D_i - D_i^*\right) = EL^* \sum_{\forall i \in N} \left(D_i - D_i^*\right) - \sum_{\forall i \in N} EL_i\left(D_i, A_i\right) \cdot \left(D_i - D_i^*\right) \geq 0 \quad (46)$$

But considering that the defender is better off spending her whole budget (Major 2002), i.e.:

$$\sum_{\forall i \in N} D_i = B_D, \quad \forall D_i \geq 0 \quad (47)$$

Then the term $EL^* \sum_{\forall i \in N} \left(D_i - D_i^*\right)$ in expression (46) cancels out and the expression may be rewritten as follows:

$$-\sum_{\forall i \in N} EL_i\left(D_i, A_i\right) \cdot \left(D_i - D_i^*\right) \geq 0, \quad \forall D_i \in \Omega \quad (48)$$

Expression (48) written in vector form yields expression (45), which completes the proof. □

On the Existence of Equilibrium for Scenario 1

The set of defense allocations Ω is compact, since all the D_i are non negative and lie on the hyperplane $\sum_{\forall i \in N} D_i = B_D$ therefore by the theory of variational inequalities we only need $EL_i(\cdot)$ to be continuous to guarantee the existence of equilibrium D^*.

A well known result in variational inequalities (see for example Ortega and Rheinboldt 2000) is that in the special case where the Jacobian matrix is symmetric, i.e. $\frac{\partial EL_i}{\partial D_j} = \frac{\partial EL_j}{\partial D_i}, \quad \forall i \neq j$, then the variational inequality problem (45) has an equivalent optimization problem. In fact, if $EL(\cdot)$ in (45) is the gradient of a differentiable function $f(\cdot)$, then (45) is equivalent to the following minimization problem:

$$Min\ f(x) \quad \text{s.t.} \quad x \in \Omega \quad (49)$$

Applying (49) to the function EL we obtain:

$$\underset{D \in \Omega}{Min} \left\{ -\sum_{\forall i \in N} \int_0^{D_i} EL_i(x, A) dx \right\} \quad (50)$$

s.t.

$$\sum_{\forall i \in N} D_i = B_D$$

A Functional Form for EL

Different approaches have been developed to give a functional form to the probability of a successful attack. For instance, Harris and Hall (1992) presents the case of Probabilistic Safety Analysis through event trees. Major (2002) put forward an approach based on a two-events tree in which the probability of a successful attack is the product of the probability of escape detection and the probability of success given that the attacker has escaped. Major's success probability function relies on two different ground bases: search theory (see Frost 1999) and dose–response modeling (see Hoel 1985). The expression for Major's probability is the following:

$$p_i\left(D_i, A_i, v_i\right) = \exp\left(-\frac{D_i \times A_i}{\sqrt{v_i}}\right) \times \left(\frac{A_i^2}{A_i^2 + v_i}\right) \tag{51}$$

Later Powers and Shen (2009) extended Major's model to an event tree with four components:

p_1 Probability of at least one attack
p_2^i Probability of target i being attacked given that at least one attack has occurred
p_3^i Probability that attack to target i goes undetected given that target i has been attacked and given that at least one attack has occurred
p_4^i Probability that attack to target i is successful given that attack to target i goes undetected, given that target$i$ has been attacked and given that at least one attack has occurred.

Thus, the expression for the probability of a successful attack on target i is given by the following expression:

$$p_i\left(D_i, A_i, v_i\right) = \pi_i \exp\left(-\frac{A_i D_i}{v_i}\right) \frac{A_i}{A_i + D_i} \tag{52}$$

where $\pi_i = p_1 p_2^i$

For Major's model (expression (51)), the corresponding optimization model would be the following:

$$\underset{\{D_i\}}{Max} \sum_{\forall i} \int_0^{D_i} EL_i(x)dx = \sum_{\forall i} \int_0^{D_i} \exp\left(-\frac{x \times A_i}{\sqrt{v_i}}\right) \times \left(\frac{A_i^2}{A_i^2 + v_i}\right) \times v_i dx =$$

$$\underset{\{D_i\}}{Max} \sum_{\forall i} \left(\frac{A_i}{A_i^2 + v_i}\right) \times v_i^{3/2} \left(1 - \exp\left(-\frac{D_i \times A_i}{\sqrt{v_i}}\right)\right)$$

s.t.

$$\sum_{\forall i \in N} D_i = B_D \tag{53}$$

For scenario 2 (i.e., expression (42)) the objective is conceptually different from the latter. The aim is to minimize not the highest EL but the total EL. This scenario can also be modeled as a variational inequality. In fact, if we perform the same analysis done for scenario 1 but instead of the function EL in expression (43) we consider its first derivative, i.e. the marginal expected loss

$$\nabla EL_i = \frac{\partial EL_i}{\partial D_i}, \quad \forall i \in N,$$

(note that this marginal vale is negative by definition) and then we replace this marginal value in the corresponding optimization model (expression (50)) would be the following:

$$\underset{D \in \Omega}{Min} \left\{ -\sum_{\forall i \in N} \int_0^{D_i} \nabla EL_i(x, A)dx = -\sum_{\forall i \in N} \int_0^{D_i} \frac{\partial EL_i}{\partial x}(x, A)dx = \sum_{\forall i \in N} EL_i \right\} \quad (54)$$

s.t.

$$\sum_{\forall i \in N} D_i = B_D$$

which is exactly the minimization sought in scenario 2 (expression (42)). Therefore, the above discussion shows the following property:

Property 4. For any attack pattern A, the equilibrium defense pattern represented by (42), can be expressed as the following variational inequality:

$$\nabla EL\left(D^*, A\right) \cdot \left(D^* - D\right) \geq 0, \quad \forall D \subset \Omega \quad (55)$$

where Ω denotes a feasible set defined by:

$$\Omega = \left\{ D_i \in \mathbb{R} / \exists \quad D_i \geq 0 \wedge \sum_{\forall i \in N} D_i = B_D \right\}.$$

In other words, properties 3 and 4 show that the Nash equilibrium in scenario 1 is reached through the expected loss function whereas for the scenario 2 the equilibrium is reached through the marginal loss functions.

Final Remarks

This chapter had the aim to present different cases and scenarios of defense planning when an unknown terrorist attack is expected to happen with a given probability. The planning for the defender consisted of allocating resources (human and material) to different targets or space locations trying to optimize certain criteria. The different cases analyzed in this chapter had a corresponding optimization model; its solution

may guide the decision makers and authorities to make rational and optimal investment decisions minimizing the risk associated to a terrorist attack.

Due to the novelty of the models proposed in this chapter, it is not possible to present here one of the most desired outcomes of any modelling experience: validation against real data or other published material. Unfortunately, to the knowledge of this author, there was not a single result reported in the public literature that can be used for comparison purposes.

An interesting challenge emerges now: running these models with actual data and try to interpret the results against the outcomes of actual attempts to attack a given set of targets.

Acknowledgements The author wants to thank the Chilean National Fund for Development of the Science and Technology FONDECYT through Grant 1120046 as well as Vicerectoria Academica at Universidad Diego Portales.

References

Ahmed S, Shapiro A (2002) The sample average approximation method for stochastic programs with integer recourse. Optimization online. www.optimization-online.org. Accessed Apr 2013

Arce D, Sandler T (2003) Terrorism and game theory. In Simulation and gaming, 34:319 (accessed online april 2013 at sag.sagepub.com/content/34/319

Bienstock D, Shapiro A (1988) Optimizing resource acquisition decisions by stochastic programming. Manage Sci 34:215–229

Clauset A, Young M, Gledistch KS (2007) On the frequency of severe terrorist attacks. J Conflict Resolut 51(1):58–88

Dadkar Y, Nozick L, Jones D (2010) Routing of hazardous material shipments under the threat of terrorist attack. In: Bell M, Hosseinloo S, Kanturska U (eds) Security and environmental sustainability of multimodal transport. Springer with NATO Public Diplomacy Division, NATO Science for Peace and Security Series C, pp 89–110

Ezell BC, Bennett SP, von Winterfeldt D, Sokolowski J, Collins AJ (2010) Probabilistic risk analysis and terrorism risk. Risk Anal 30(4):5175–5189

Frost JR (1999) Principles of search theory, Part I: detection. Response 17(2):1–7

Garrido RA (2010) Terrorists and hazmat: a methodology to identify potential routes. In: Bell M, Hosseinloo S, Kanturska U (eds) Security and environmental sustainability of multimodal transport. Springer with NATO Public Diplomacy Division, NATO Science for Peace and Security Series C, pp 149–166

Harris J, Hall S (1992) Comparison of event tree, fault tree and Markov methods for probabilistic safety assessment and application to accident mitigation. Major hazards onshore and offshore, vol 15, p 59

Hoel D (1985) Mathematical dose-response models and their application to risk estimation. In: Vouk VB, Butler GC, Hoel DG, Peakall DB (eds) Methods for estimating risk of chemical injury: human and nonhuman biota and ecosystems. SCOPE, John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, pp 347–360

Holguin-Veras J, Prez N, Ukkusuri S, Wachtendorf T, Brown B (2007) Emergency logistics issues affecting the response to Katrina: a synthesis and preliminary suggestions for improvement. Transport Res Rec J Transport Res Board 2022(1):76–82

Lepofsky M (2012) Security of hazmat transports by road. In: Reniers GLL, Zamparini L (eds) Security aspects of uni and multimodal hazmat transportation systems. Wiley-VCH; 1 edition (April 16, 2012) ISBN-10: 3527329900 pp 29–48

Luedtke J, Ahmed S (2008) A sample approximation approach for optimization with probabilistic constraints. SIAM J Optim 19(2):674–699

Major J (2002) Advanced techniques for modeling terrorism risk. J Risk Finance Fall 15–24

Murray-Tuite PM (2008) Transportation network risk profile for an origin-destination pair: security measures, terrorism, and target and attack method substitution. Transport Res Rec 2041:19–28

Murray-Tuite PM, Fei X (2010) A methodology for assessing transportation network terrorism risk with attacker and defender interactions. Comput Aided Civ Infrastruct Eng 25:396–410

Murray-Tuite PM, Garrido RA, Nune R (2007) Path prediction methodology for hazardous materials transported by malicious entities. Conference proceedings of the 11th world conference on transport research, University Berkeley, USA

Nune R (2007) Path prediction and path diversion identifying methodologies for hazardous materials transported by malicious entities. M.Sc. thesis, Virginia Polytechnic Institute and State University, Falls Church, Virginia

Nagurney A (1999) Network economics: a variational inequality approach. Kluwer Academic Publishers, Boston

Ortega JM, Rheinboldt WC (2000) Iterative solution of nonlinear equations in several variables. Academic, New York

Overgaard PB (1993) The scale of terrorist attacks as a signal of resources. Economics Working Papers 1993-20, School of Economics and Management, University of Aarhus

Pagnoncelli BK, Ahmed S, Shapiro A (2009) Sample average approximation method for chance-constrained programming: theory and applications. J Optim Theory Appl 142(2):399–416

Powers M, Shen Z (2009) Colonel Blotto in the war on terror: implications for event frequency. J Homeland Security Emerg Manage 6(1):1–16

Siqueira K, Sandler T (2008) Defensive counterterrorism measures and domestic politics. Defense Peace Econ 19:405–413

Sandler T, Siqueira K (2009) Games and terrorism: recent developments. Simulat Gaming 40: 164–192

Siqueira K, Sandler T (2010) Terrorist networks, support, and delegation. Public Choice 142: 237–253

Siqueira K, Sandler T, Cauley J (2009) Common agency and state-owned enterprise reform. China Econ Rev 20:208–217

Shapiro A, Homem-de-Mello (1998) A simulation-based approach to two-stage stochastic programming with recourse. Math Program 81:301–325

Sunstein CR (2003) Terrorism and probability neglect. J Risk Uncertainty 26(2/3):121–136

Szyliowicz JS (2012) Safeguarding hazmat shipments in the US: policies and challenges. In Reniers GLL, Zamparini L (eds) Security aspects of uni and multimodal hazmat transportation systems. Wiley-VCH; 1 edition (April 16, 2012) ISBN-10: 3527329900, pp 237–261

Willis H, Morral A, Kelly T, Medby J (2005) Estimating terrorism risk. Rand Corp. Center for Terrorism Risk Management Policy Monograph Series

Zhuang J, Bier VM (2007) Balancing terrorism and natural disasters—defensive strategy with endogenous attacker effort. Oper Res 55(5):976–991

The Role of OR in Emergency Evacuation from Hazmat Incidents

Brian Wolshon and Pamela Murray-Tuite

Introduction

Evacuations occur more frequently in the United Sates (US) than is often realized. A recent US Nuclear Regulatory Commission (NRC) report showed that, on average, an evacuation involving 1,000 or more people occurs nearly once every 2 weeks somewhere in the country (Dotson and Jones 2005). The NRC study statistics, compiled by the Sandia National Laboratories over the 10 year period between 1993 and 2003, also suggest that the vast majority of incidents requiring an evacuation are small, localized events. Of the 230 evacuation events documented in the study, 171 or nearly three quarters of them involved 1,000–5,000 people. While wildfires and floods were the hazards for which an evacuation was most commonly required, hazardous material incidents and transportation-related events also accounted for more than a quarter of the documented evacuations over this period.

As the movement of hazardous materials continues to increase across all modes of transport, it is expected that the frequency of major incidents requiring evacuations will also likely rise. And while many of these will continue to be of the smaller-scale nature that are currently most common, some are also likely to encompass large areas and involve many thousands, if not tens of thousands, of

B. Wolshon (✉)
Department of Civil and Environmental Engineering, Gulf Coast Research Center for Evacuation and Transportation Resiliency, Louisiana State University, Baton Rouge, LA 70803, USA
e-mail: brian@rsip.lsu.edu

P. Murray-Tuite
Department of Civil and Environmental Engineering, Virginia Tech, 7054 Haycock Rd, Falls Church, VA 22043, USA
e-mail: murraytu@vt.edu

R. Batta and C. Kwon (eds.), *Handbook of OR/MS Models in Hazardous Materials Transportation*, International Series in Operations Research & Management Science 193,
DOI 10.1007/978-1-4614-6794-6_9,

people. To mitigate the potential for injury and loss-of-life, it is necessary to plan for and implement appropriate protective actions during such events. Chief among these is the use of evacuations.

Specialized techniques such as those developed in the field of operations research (OR) can play an important role in the planning and operation of evacuations. During emergencies, time is limited and initial response resources are nearly never adequate to address all the immediate needs of the event. OR techniques have been useful in simulation models which are used to assess hazard and response scenarios. They can be applied to meet these challenges and support decision-making in situations in which there are many competing needs and various response alternatives that each have varying levels of costs and benefits.

Currently, the level of complexity in evacuation planning and response for transportation-related hazardous material incidents is limited in practice. These types of emergencies are typically directed at the local level by emergency response personnel with considerable practical training and experience who must make very fast decisions and react equally as quickly to any number of changing conditions. Because of the fluidity of these events and the somewhat random locations of their occurrence, there has also been little opportunity for the application of complex quantitative and analytical methods for guiding decisions and responses. However, the growing availability, portability, and speed of handheld devices and computer technology, remote sensing systems, and storage of demographic and geographic information now makes data available virtually anywhere and accessible at a moment's notice. When this is also combined with the types of sophisticated computational and optimization algorithms developed within the field of operations research, such information can be used to develop and implement evacuation orders with considerably higher speed, precision, and effectiveness than has been possible in the past.

This chapter discusses the ways in which operations research knowledge and tools can be applied for improving evacuation planning, management, and operation for transportation related hazardous material incidents. The chapter is divided into three primary sections that generally discuss characteristics of hazmat events and evacuations; the existing state of the practice for analyzing evacuation transportation processes, including a summary of current dedicated- and general-purpose traffic simulation systems that have been applied by analysts; and the ways in which OR techniques are or can be applied to improve evacuation conditions. Then a forward-looking discussion examines the emerging frontiers of OR knowledge and research that is expected to one day permit ever greater complex to be dealt with. Within this discussion, the chapter will also highlight how evacuations for hazmat incidents differ from those of other types of hazards; the application of OR in emergency management including, preparedness and response; and the ways knowledge of today can be harnessed to provide the highest levels of protection during any size hazmat emergency.

Although the information presented in this chapter focuses on emergency evacuation for hazmat incidents, it should also be noted that the information and techniques that are included here can also be applied to many other types of

hazard events that require evacuations. Since most hazard-response scenarios can be characterized in terms of key spatial and temporal variables, a common approach is to scale these key parameters to be appropriate to a specific threat scenario. Among the key spatial variables are the geographic extent of the hazard; the distribution of the population within the protective action zone; the layout and capacity of the evacuation road network; and the location of the safe shelter destinations, among many others. Key temporal variables include the amount of advanced warning time to the onset of hazardous conditions; the rate at which evacuees receive and heed the evacuation order; the time and day on which the incident occurs; and so on. Within this context, the various parameters used in the OR models can be appropriately modified to fit the specific scenario.

Evacuation from Hazmat Incidents

The fundamental purpose of an evacuation is to move people away from hazardous conditions. The need for and specifics of them is based on a set of time and space parameters that describe the extent and movement of the hazard as well as the area being threatened. The duration and size of these conditions dictates the requirements of the evacuation including, how fast it needs to be carried out, how many evacuees need to leave, how far they need to travel, the urgency at which they must flee, as well as what other types of traffic control and emergency proactive actions can be taken to expedite the evacuation process. While some of these are based on the characteristics of the hazard, including its level of danger, speed of movement, and the extent of harmful effect, others are based on the configuration and capacity of the transportation network and the availability of transportation assets and resources. Another, and perhaps most critical set of conditions, is the response and decision-making behavior of the evacuees.

Although this chapter focuses on evacuations associated with hazmat incidents, the events for which evacuations may be ordered include many types of naturally-occurring and man-made hazards. Hazmat-related evacuations are not necessarily unique among all other hazards. They, like an evacuation from any other hazard, are governed by the same temporal and spatial constraints of a threat condition. With that in mind, it is still possible to broadly generalize about the characteristics of hazmat incidents, the types of threat conditions they create, and the evacuation responses that they generate.

Hazardous material incidents, though widely varying, generally fall into the "man-made" incident category. Although they can include large-scale disasters, most hazmat incidents are also considerably smaller than natural disasters like hurricanes, wildfires, and floods which routinely threaten areas of hundreds, if not thousands of square miles. Depending on the population characteristics near the incident site, smaller impact areas also mean smaller numbers of evacuees, although this is not always the case. Not only can hazmat incidents be large and complex, they can also accompany natural disasters.

Recent experience has led to the creation of a new group of incidents called "natech disasters." Natech events are those in which a natural disaster results in a cascading series of one or more related technological disasters within the built environment in impacted areas (Steinberg et al. 2008). A recent example of such an event occurred in 2011 at the Fukushima Daiichi Nuclear Power Plant in Japan when an earthquake and resulting tsunami disabled the reactor cooling systems leading to nuclear radiation leaks from the plant. Other, similar types of conditions occurred in Turkey when the 1999 Kocaeli earthquake caused large scale petrochemical refinery fires and spills and, yet another was observed after Hurricane Katrina when flooding caused oil containment breeches near residential neighborhoods of Chalmette in suburban New Orleans.

Hazmat incidents that involve evacuation can also feature numerous other complicating characteristics that increase the difficulty of ordering and carrying out evacuations. Unlike a hurricane or tsunami or other fixed-site incident where the location or general approach direction of the hazard is known, hazmat incidents can occur virtually anywhere, at any time, and can involve any one of many hazardous materials. For example, a train derailment or over turned tractor trailer can occur along any stretch of railroad or highway. First, the threat it poses to the public must be evaluated on short notice to determine if an evacuation or other form of protective action is warranted. In many cases, a shelter-in-place order (the opposite of an evacuation) may be determined to be most appropriate if travel to a safe location cannot be achieved. Next, the area under threat has to be determined. For airborne contaminants, plume dynamics and dissipation characteristics vary for different substances. Similarly, the area under threat will change with ambient wind patterns. Then, an appropriate protective action must be communicated to the threatened population. With little-to-no-advanced warning time this can be difficult, particularly late at night and in less densely populated areas without reverse 911 systems, civil defense sirens, and so on. Even post-event reentry can be made problematic by potential residual effects of hazardous materials.

Evacuation Transportation Processes

The last 15 years has seen tremendous advances in the way that evacuations are planned, implemented, managed, and analyzed. Techniques, such as contraflow, staged and phased evacuations, cross-state regional coordination, special needs and transit assisted evacuation as well as planning and analysis techniques like regional multimodal traffic simulation, have come about due, in large part, to a series of high profile failures, a greater list of hazards for which evacuations are now required to serve as a protective action, changing population demographics, and conditions which have resulted in natural hazards that occur both with higher frequency and increasing levels of severity.

The following sections discuss the key spatial and temporal considerations that are commonly considered in evacuation planning and analysis as well as how

various hazard conditions impact the planning process. This section also highlights the existing state of the practice for analyzing evacuation transportation processes, including a summary of current dedicated- and general-purpose traffic simulation systems that have been applied by analysts. Each of these discussions also includes examples of how these various parameters and systems have been applied, adapted, and modified for use in evacuation.

From an OR modeling perspective, framing the components of the evacuation process in terms of time and space is helpful from two perspectives. First, it is helpful to disaggregate the often enormously complicated and interrelated processes of the evacuation into separate, smaller components that are easier to observe and record. Then, it permits each of these key components to be represented quantitatively, as equations or as a distribution of continuous data.

Temporal Parameters

Among the temporal evacuation variables, one of the most critical is the amount of advanced notice available prior to the onset of hazardous conditions. In evacuation planning there are two primary types of prior-warning conditions; those that give advanced notice and those that do not. "No notice" events include a variety of natural (e.g., earthquakes) and unintentional (e.g., levee/damn break, train derailment, chemical explosion, etc.) as well as intentional (e.g., terrorist attack, biotoxin release, etc.) manmade hazards. "With-notice" events vary more widely and have been categorized as "short notice" events (those hazards that may give up to an hour's notice prior to the onset of hazardous conditions) such as fires, floods, some nuclear/biological/chemical releases and "long notice" events including hazards such as wildfires and hurricanes that give several hours or even days of advanced notice. The majority of evacuations from hazardous material emergencies, and the focus of this chapter, typically give only short- to no-advanced notice.

Advanced warning time is also important because it dictates the amount of notice that response agencies have to implement control and management measures like contraflow, road closures, and emergency signal timing plans as well as the time to activate assisted evacuation plans like evacuation bus services, medical special needs evacuations for the elderly, infirm, and disabled. It can also limit or extend the amount of time that evacuees have for pre-evacuation mobilization activities, such as picking up children from school, coordinating with mobility-limited friends and relatives, closing homes and businesses, gathering materials and supplies, and so on. Although it may seem that there is little time for these types of time consuming activities during short notice hazmat incidents, there are numerous cases where they could still be used. Nuclear power plant emergencies, in particular, are commonly planned under a set of assumptions that includes several hours will be available between the time when a reactor emergency occurs to the time at which a containment breech may begin to pose a risk to persons within the plume exposure area. This time permits, among other activities, police control

check points and contraflow lane reversals to be implemented and school and other assisted evacuation plans to be initiated.

Advanced warning time also affects the ability of officials to issue evacuation orders because they must be communicated through various formal (e.g., media releases, reverse 911 calls, etc.) and informal means, including social networks (e.g., friends, family, co-workers, neighborhoods, etc.). This lead time is particularly important for non-resident transient populations who may be within the evacuation zone for work, shopping, and other recreational activities. Research and development work, particularly related to hurricane evacuations, in studying behavioral responses under the various advanced notice conditions is available in the literature (Lindell and Perry 1992, 2012; Wilmot and Mei 2004; Fu and Wilmot 2004; Gudishala and Wilmot 2012). Much of this information has also been quantified and adapted to create evacuee departure times and response distributions which can be used in OR models of evacuation. With the additional calibration or adjustment, it is expected that even hurricane evacuation behavioral responses could be adapted for use in analyzing short- to no-notice events such as those commonly associated with hazmat incidents.

Once the evacuation is underway, there are numerous other temporal parameters which can be used to evaluate the performance of evacuation processes. From a transportation analysis perspective, these can include evacuation travel time and delay as well as the time needed to implement and/or remove evacuation traffic management and control measures like contraflow, road and bridge closures, and police control points. From an emergency management perspective, key temporal parameters may include onset time and duration of hazard conditions, time to issue evacuation orders, evacuee mobilization time, and so on. Again these processes have been quantified in prior work for use in the planning and evaluation of evacuation alternatives (Chen et al. 2007; Wolshon et al. 2005, 2006).

Spatial Parameters

Like the temporal parameters described previously, evacuations also encompass a range of spatial parameters which also influence the manner in which evacuations are planned and carried out. They can also be used to evaluate the effectiveness of evacuation plans and identify areas of need and improvement.

Among the most essential spatial parameters that dictate the size of the evacuation protective action zone is the spatial extent of the hazard. Obviously, the larger the area of threatening conditions, the larger the area that must be evacuated. However, spatial distributions of the resident and transient populations within the threatened population are what actually influence the amount of people and vehicles that would be in the evacuation. For example, several past hurricanes in Texas have threatened thousands of square miles but ultimately made landfall in sparsely populated areas that did not require major evacuations. While the 9/11 terrorist attacks in New York affected several city blocks, they required the evacuation

of several million people. Although most hazmat incidents would be expected to be considerably smaller in scale than a hurricane, there are many hazmat hazard scenarios that could affect hundreds of square miles.

Two other hazard parameters that influence the urgency, extent, and direction of the evacuation are the approach direction and movement of threatening conditions. An illustrative example of the effect of hazard movement on an evacuation lies in wildfires in urban-wildland interface areas. Because of the highly variable development and movement of wildfires, which are themselves a function of weather and fuel conditions, it is not possible to develop specific detailed evacuation plans. In Southern California, for example, emergency preparedness and response agencies find it more effective to work from a general evacuation framework, rather than a plan, to permit greater flexibility to respond to rapidly changing fire conditions. This includes designating the geographic extent of threat region, amount of available advanced warning time, available routes, and even shelter destinations. With the exception of nuclear power plant facilities which are typically planned to assume a fixed site emergency, the majority of hazmat incidents would be expected to create evacuation conditions that are more similar to wildfires where the locations of the protective action zone and urgency at which the evacuation needs to be conducted is not known in advance and can change rapidly based on wind strength and direction.

Other important spatial parameters that affect evacuation processes are the location and required travel distance to safe shelters, arrangement and access to transportation networks, and the location and frequency of downstream bottlenecks. For hazards like nuclear power plant emergencies and hurricanes, shelters are planned well in advance of the emergency and, as such, evacuation travel is planned to reach them. However, for wildfires, safe shelter destinations and the routes recommended to reach them may change from event to event or even several times within a single event. Similarly, the available road network including intersections, merges, terminal points, and capacity restrictions (e.g., bridges, tunnels, etc.) influence the direction of movement away from the threat.

OR models can be effective for planning and analyzing the various spatial components and characteristics of evacuations. Although most of the recent past research and development work has been focused on evacuations related to hurricanes and terrorist-related incidents, these applications could be adapted to hazmat incidents. OR models could be used to minimize travel and clearance time and/or optimize lane utilization, network capacity, loading, traffic, etc.

Evacuation Simulation and Analysis Techniques

Most modeling in the traffic planning and operations field is undertaken to evaluate routine traffic conditions such as the impacts of local and regional development, capital roadway improvements, and traffic control alternatives. It is also commonly accomplished using one of an assortment of commercially available and other

non-proprietary simulation packages. Although these software systems are useful and effective at modeling the activities and performance of systems and agents that range from regional sized road networks down to individual vehicles and pedestrians, they typically do not include sophisticated algorithms or strategies for optimizing cost and/or performance.

A considerable amount of recent OR work in the area of evacuation has been focused on improving traffic modeling and simulation in these areas. Although there are many existing traffic simulation systems, research is always ongoing to develop descriptions and models to more accurately and realistically capture various travel behaviors and characteristics under both routine and emergency conditions. One area where this has been particularly true has been in the area of travel demand modeling where considerable effort has gone into integrating dynamic algorithms that can reassign traffic to alternate travel paths based on the evolving occurrence of incident-related and/or routine recurring traffic congestion. Numerous studies have examined how network models can be modified and adapted to model the key aspects of dynamic traffic assignment to the network as threat and traffic volume condition change continuously throughout an evacuation process (Chiu et al. 2005; Brown et al. 2009; Lin et al. 2009; Pel et al. 2012).

The following sections present and discuss the three traditional categories of traffic simulation models with a focus on how some of them have been adapted to or applied for the specific problem of evacuation. Several recent reviews of traffic simulation software applications for evacuation planning and analysis have been conducted. Among the most useful was completed by Hardy et al. (2009) as part of a United States Department of Transportation (USDOT) technical study. The USDOT's *Evacuation Management Operations Transportation Modeling Inventory* included discussion of 30 systems that have been applied for evacuation-related traffic analyses. The report also comments on the relative strengths and weaknesses that should be considered when selecting any of them. The discussion that follows summarizes many of the key findings from this work as well as recent projects and studies conducted by various transportation and emergency management agencies and firms across the country. Typical performance measures produced at various modeling scales and the validation and calibration processes needed to assure its utility for evacuation applications are discussed as well as the ability of these systems to incorporate multimodal system aspects, pedestrians, and visualization capabilities is also included.

Modeling Scale

Simulation models for operational- and planning-level analyses generally fit into one of three model categories; macro, meso, or micro. These categories are distinguished based on the level of abstraction of relevant components of the transportation system, including the vehicles, roadway components, control devices, etc. Like any system that is modeled, the level of abstraction has a direct impact on

the quantity and level of detail requirements of the input data as well as the amount time and labor effort required for coding and entry. Correspondingly, it also dictates the level of detail at which output data is produced.

Macro-level

Macroscopic models are typically used for developing and assessing "Big Picture" views of traffic conditions. They are most effective for analyzing regional level networks such as those that may encompass a city or metropolitan area. In most macro models, specific network characteristics and individual vehicles are aggregated and represented as "average" conditions over road segments. Within such a representation, individual vehicles are grouped into platoons with average speeds and spacing. Similarly, segments of roads are represented generically without changing posted speeds, numbers of lanes, minor street crossings and the like. As such, these models do not capture the effects of vehicle-to-vehicle interactions or the effects that they can have on the overall flow conditions of the traffic. Because of this, they are not useful for identifying localized conditions such as bottlenecks caused by stopped, slowing, merging, and diverging traffic along a link. Macroscopic models are also usually temporally aggregated over incremental time periods of 15 min to an hour, so changing flow conditions such as that which occurs during queue formation and dissipation is not captured within a simulation. However, when this level of detail is not required, these models can be quite effective. In Florida, a statewide model was developed to look at traffic patterns and clearance times during regional mass movements of traffic during hurricane evacuations (Florida Department of Community Affairs 2010; Lindell and Perry 2012). In the Florida Keys where there is effectively a single route of egress and road cross sections are fairly consistent over a long stretch of US-1, a spreadsheet-based macro model has been used for more than a decade to estimate clearance time and the improvements that could be gained from various capacity enhancement improvement projects.

Micro-level

Micro-level or "agent-based" traffic models represent the variety of systems on the opposite end of the spectrum from the macro-models. In microscopic models, traffic flow is represented in terms of individual vehicles moving in the system. As such, specific movements such as accelerating, decelerating, and lane changing can be tracked separately for individual vehicle or pedestrian "agents." Similarly, behavioral characteristics can be assigned to individual agents. To reflect a representative mix of driver types and vehicle performance characteristics, these attributes are assigned to individual entities based on statistical distributions that can be modified to reflect a locally observed population of drivers and vehicles. Because of the level of detail required to code a microscopic traffic model, the scope of these models are typically limited to networks consisting of several dozen intersections or corridors

extending over 20 or less miles. Although micro-scale models have been used to represent networks ranging from corridors and networks much greater than this, the amount of time and labor required to create these larger models does not typically justify the cost.

Micro-scale models typically use time- or discrete event-based updating of the model state and its agents. In time-based models, activities are incrementally updated at fixed intervals, commonly 1 s. At each interval, vehicle speeds, locations, etc. are changed as a function of intra-vehicle spacing, acceleration, deceleration, gap availability and so on. Some models use increments as short as one tenth of a second and string together successive movements to create smooth movements for graphical visualizations. In an event-based system, updates are made based on the occurrence of events that result in a change of system state such as would occur when a vehicle is released from queue at a toll booth or traffic signal.

Meso-level

Mesoscopic models have aspects of both micro and macro systems. One type of meso-scale modeling system that is growing in use is the cell transmission model (CTM). In a CTM, road segments and other system features are disaggregated from macroscopic model levels, but not quite down to the agent level of micro models. In the TRANSIMS (TRansportation ANalysis SIMulation System), a type of CTM and one of several meso-scale models used for regional scale evacuation traffic analysis, individual vehicles are not explicitly modeled, rather road segments are broken up into 7.5 m cells (about the length of a single passenger vehicle) that are occupied or vacant during each second of the simulation. Vehicle speed is represented by the occupation of several successive cells and lane change by the occupation change between adjacent cells on a second-by-second basis. Traffic control and various other features are also modeled at similar aggregated/disaggregated levels. Within this context, meso models permit the simulation of considerably larger areas over longer durations than micro-models with more precise results than macro models.

Mixed Modeling Approaches

One other technique for applying simulation models for evacuation traffic analysis is the mixed-model approach. In the mixed-model approach, the process begins with a high-level macro analysis. Based on the results of the big-picture perspective, micromodels are developed to examine specific locations within the network. If areas of significant traffic congestion are apparent, then these segments can be micro-analyzed for closer inspection. These sub-areas can vary from as large as a corridor down to as specific as an intersection or interchange. At this level it becomes possible to assess specific geometric design and traffic control features that may be contributing to flow turbulence. A mixed modeling approach is currently being used to analyze emergency and evacuation scenarios in the City of

Philadelphia. Another benefit from this development will be ability to apply these models for routine peak-hour and other planned event conditions.

While the mixed-modeling approach saves the effort of coding significant areas of the network in detail that have little general effects on the overall operation of the network, it may still require significant time and labor investments since multiple sets of models may still be required. Moving from macro to micro may also incur risks of not accounting for the multitude of circumstances and conditions such as incidents, network loading, and the like which can also contribute to network congestion.

OR Applications to Traffic Modeling

OR techniques can add additional value to traffic simulation modeling systems in a number of different ways and both from within and outside of these programs. Routines for optimizing and balancing traffic volumes to alternative routes and assigning traffic signal timings to signalized intersections have been incorporated into traffic simulation models for many years. Dynamic traffic assignment (DTA) routines form the basis of systems that variably assign traffic to routes based on time-varying network conditions and travel demand conditions. This arrangement much more realistically represents real-world traffic processes.

External OR applications are for pre- and post-processing of traffic scenarios. For example, OR techniques can be used to devise scenarios to allocate transit buses within a network to best utilize available capacity, personnel, and vehicles, then use a simulation system to view and compute specific performance measures of these plans under varying levels of network volume. Similarly, operational performance measures like intersection arrival volume from a simulation model can be used to optimize and coordinate signal systems in the field. OR can also be used for the calibration and validation of simulation models to most accurately represent actual traffic conditions.

OR Applications for Evacuation Management Strategies

The desire to improve evacuation conditions or develop lower bounds on evacuation times has led to the use of OR models for a variety of evacuation management strategies. A selection of these strategies is discussed below; and where the formulations are reasonably short or can be understood without reiterating the majority of the related article, sample formulations are presented. Most of these strategies interact with traffic simulation or route optimization models. (Note that contraflow strategies, while popular for hurricane evacuations, are not likely to be appropriate for no-notice evacuations since they can require significant set up times, e.g., 3–5 h (Zimmerman et al. 2007).)

As demonstrated below, bi-level techniques are often used with an optimization problem as the upper level and the determination of travel times at the lower level (optimization or simulation). The general approach uses the upper level to make a change to the transportation network or demand and the lower level determines travel times which are fed into the upper level's objective function. The solution techniques can be viewed in three categories, depending on the complexity of the formulation: (1) strict optimization (using optimal algorithms for single layer problems); (2) meta-heuristics; and (3) iterative optimization and simulation. Combinations of techniques are also possible. In the third option, and at times the second, the modification to the network or demand selected in the upper level problem may require additional code or external modification to inputs for the lower level; thus creating a potential solution that is evaluated in the lower level and iterations between the two levels continuing.

Ramp Management Optimization

The ramp management strategies are intended to improve evacuation times by smoothing traffic flow, typically on freeways. These strategies can involve metering (e.g., Daganzo and So 2011; Edara et al. 2010; McGhee and Grimes 2006; So and Daganzo 2010) or closure (e.g., Fonseca et al. 2009; Ghanipoor Machiani et al. 2012, 2013).

As a simplified example of the bi-level model for ramp closure, Ghanipoor Machiani et al.'s (2012) model is presented below and is solved using a combination of techniques (2) and (3). In their study, the lower level involved traffic simulation and the overall model is solved using the heuristic approach simulated annealing.

Upper level problem:

$$\min z = \sum_{k \in K, s \in S} \sum_{j=1}^{E^{ks}} T_j^{ks} \tag{1}$$

Subject to:

$$x_i^f + x_{i+1}^f \leq 1 \quad \forall i \in \{R^f - l\} (l = \textit{last element of } R^f),\ f = 1, 2, \ldots, F \tag{2}$$

$$\sum_{f=1}^{F} \sum_{i=1}^{|R^f|} x_i^f \leq B \tag{3}$$

$$x_i^f \in \{0, 1\} \qquad \forall i \in R^f,\ f = 1, 2, \ldots F \tag{4}$$

Lower level:
traffic assignment that results in $T_j^{ks} \qquad \forall k \in K,\ s \in S.$

where K is the set of impacted zones located in a sub-network G' of the overall network G''', S is the set of all safe zones in $G''' - G'$, T_j^{ks} is evacuee j's travel time from origin k to destination s, E^{ks} is the number of evacuees traveling from origin k to destination s, x_i^f is the decision variable (1 if ramp i on freeway f is closed, 0 otherwise), B is the upper bound on the number of ramps that can be closed, R^f is the set of ramps on freeway f in an area larger than G' but smaller than G''', and F is the set of freeways considered for ramp closures.

The upper level's objective function (1) minimizes mandatory evacuees' total travel time from their origins to their destinations. Constraint (2) prohibits closing two consecutive ramps; this constraint is intended to limit stress on the evacuees who are seeking quick routes out of danger. Constraint (3) limits the maximum number of ramps that can be closed. Finally, constraint (4) indicates that the decision variables are binary integers. This formulation was used in a no-notice evacuation study in Virginia. The results suggested that ramp closures could improve evacuees' total travel time by as much as 8% under certain conditions. While this may seem like a small percentage, it equated to over 20,500 h (Ghanipoor Machiani et al. 2013).

Unlike ramp closure, ramp metering permits continuous, rather than integer, variables. Daganzo and So's (2011) approach for determining optimal freeway on-ramp control is based on an "innermost first out" control strategy, which limits the number of evacuees released by a ramp onto the freeway to the residual capacity of a downstream bottleneck. This strategy prioritizes upstream freeway flow over ramp flows (Daganzo and So 2011) and is determined using technique (1)—optimal algorithms.

Crossing Elimination Optimization

Crossing elimination is intended to decrease evacuation time by minimizing or reducing vehicle conflicts at intersections. Figure 1 illustrates one possible restructuring of an intersection. This strategy prohibits selected movements at selected intersections and smoothes flow by removing some flow interruptions. It requires careful implementation since there is a tradeoff between faster speeds and longer distances (Cova and Johnson 2003).

Cova and Johnson developed lane-based routing strategies with a linear program that treated crossing elimination as a minimum-cost flow problem, as shown below. The objective of the minimum-cost flow optimization problem is to minimize the cost of transporting supply from source nodes to meet demand (destinations) given a capacitated network. The network simplex technique (general technique (1)) can be used to solve this problem (Cova and Johnson 2003).

$$\min Z_2 = \sum_i \sum_j d_{ij} x_{ij} \tag{5}$$

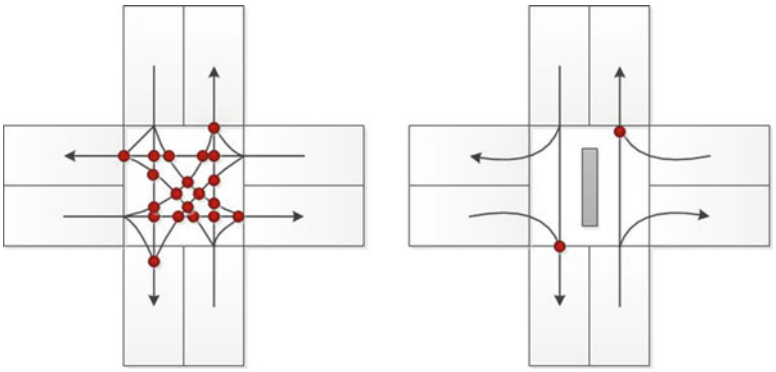


Fig. 1 Sample intersection reconfiguration with reduced conflict points

Subject to:

$$\sum_j x_{ij} - \sum_j x_{ji} = b_i \quad \text{for all } i \tag{6}$$

$$x_{ij} \leq u_{ij} y_{ij} \quad \text{for all } i \rightarrow j \text{ that cross } k \rightarrow l \tag{7}$$

$$x_{kl} \leq u_{kl}(1 - y_{ij}) \quad \text{for all } k \rightarrow l \text{ crossed by } i \rightarrow j \tag{8}$$

$$\sum_j y_{ji} \leq z_i + 1 \quad \text{for all } i \text{ with a potential merge} \tag{9}$$

$$\sum_i z_i \leq M \tag{10}$$

$$0 \leq x_{ij} \leq u_{ij} \quad \text{for all i} \rightarrow \text{j} \tag{11}$$

where i is the index of network nodes, $i \rightarrow j$ indicates a directed arc from node i to node j, b_i is the net flow at node i, d_{ij} is the distance along $i \rightarrow j$, u_{ij} indicates the capacity for $i \rightarrow j$, M is the upper bound on the number of merges, x_{ij} indicates vehicle flow on $i \rightarrow j$, y_{ij} is $\begin{cases} 1 \text{ if the flow on arc } ij \text{ is positive} \\ 0 \text{ otherwise} \end{cases}$, z_i is the number of traffic streams that merge at node i (formulation and notation quoted from Cova and Johnson (2003, p. 584)).

Cova and Johnson's (2003) evacuation routing problem is an integer extension of the minimum-cost flow problem, which minimizes total travel distance for the vehicles in the network, subject to several constraints. First, flow must be conserved, as in the min-cost flow problem. Second, intersection crossing-conflicts must be prevented. Third, merges must be identified and three-way merges at a single lane can be prevented by setting bounds on certain variables. The total number of merges may be bounded for the network. Finally, the flow variables are bounded between

0 and an upper bound, specific to individual links. Their model routes vehicles to their closest evacuation zone exits and minimizes the number of merging and crossing conflicts at intersections. Cova and Johnson applied their model to a section of downtown Salt Lake City, Utah, (20 intersections) where evacuations occurred due to armed assailants, tornados, and bomb threats. Their research showed that overall travel time could be reduced under some scenarios, but these improvements often came at the price of longer and in some cases more indirect routes (Cova and Johnson 2003).

Several years later, other researchers developed additional approaches to crossing elimination (Luo and Liu 2012) and combined crossing elimination with lane reversal (e.g., Xie et al. 2010, 2011; Xie and Turnquist 2009, 2011). These approaches are largely bi-level optimization models. Xie and Turnquist's (2009) upper level minimizes total network evacuation time and the lower level problem is a stochastic traffic assignment problem. It was assumed that evacuees choose their routes and destinations simultaneously. Unique constraints included reserving a sufficient number of lanes for emergency vehicles. They applied their model to the area near a nuclear power plant in Minnesota with an estimated 42,000 evacuees and an evacuation network of approximately 44 nodes and found that full link reversal heading away from the hazard guaranteed no traffic crossing at the intersections. However, access for emergency vehicles requires some lanes operating in the opposite direction and a trade-off must be made between emergency vehicles and evacuation traffic (Xie and Turnquist 2009).

Xie's further development in this area continued with bi-level problems but used combined Lagrangian relaxation and Tabu search (Xie et al. 2010; Xie and Turnquist 2011) solution methods. Relaxation based heuristics were also used by Bretschneider and Kimms (Bretschneider and Kimms 2011). Similar to the 2009 and 2010 work, Xie et al. (2011) also converted the intersections into subnetworks where each movement is represented by a link. Their solution technique was simplex based.

Signal Retiming Optimization

Another strategy to modify intersection control operations is to retime the traffic signals (Chen et al. 2007; Liu et al. 2008). Liu et al. (2007) developed a model reference adaptive control framework that adjusts the timings based on desired traffic states from system optimal objectives and current prevailing conditions. Ren et al. (2012) presented a bi-objective, bi-level program to determine traffic signal timings and flows on evacuation routes with the added complication of uncertain background traffic. Similar to the other bi-level problems, the lower level represents traffic assignment. Their specific approach is a logit based routing method. At the upper level, the objectives are to minimize the total travel time for evacuees and the performance index for the entire network flow. The performance index is a weighted combination of delay and background traffic impact degree (BTID). The BTID is a measure of the spillback due to the background traffic (non-evacuees). The background traffic is determined from the lower-level model. The problem is

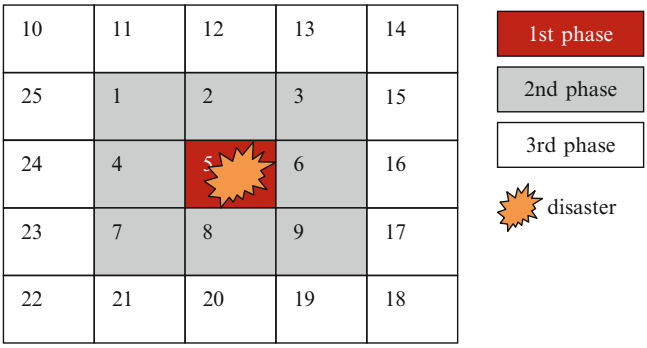


Fig. 2 Notional staged evacuation plan

solved with the non-dominated sorting genetic algorithm II (general technique (2)). They tested the formulation for the Sioux Falls network and the Jianye district of Nanjing city, which holds the Nanjing Olympic Sports Center and found that minimal evacuation time occurs with higher spillback probabilities while plans with lower spillback probabilities yield higher evacuation times (Ren et al. 2012).

Staged/Phased Evacuation

The basic idea behind staged or phased evacuation is to stagger evacuee departures (as in Fig. 2) to reduce (Chen and Zhan 2004; Mitchell and Radwan 2006) or minimize overall evacuation time (Bish et al. 2006; Dixit and Radwan 2009), minimize the time for those in the highest risk areas to reach safety, minimize risk (Bish et al. 2006), or some combination of these.

Sbayti and Mahmassani (2006) used a bi-level formulation to determine time-dependent route assignments at the upper level and the route travel times at the lower level. The upper level is solved with the method of successive averages while the lower level is solved with traffic assignment-simulation software (general technique (3)). Their model produces departure time, route, and destination for each vehicle. They then used the output to generate a time-dependent staging plan for each origin. They stated that their model incorporates an inherent objective function that minimizes evacuee total trip time, maximizes the number of evacuees reaching safety in each time period, and minimizes network clearance time. They tested their model on the Fort Worth, Texas network and reduced total evacuation trip time by 31–71% and network clearance time by 20–61%, depending on the number of evacuees, which were under 6,000 out of 47,300 travelers in the three examined scenarios (Sbayti and Mahmassani 2006).

Similar to Sbayti and Mahmassani (2006), Bish et al.'s (2006) model determines the evacuee's destinations and departure times and Chiu et al.'s (2006) model determines evacuees' destinations, routes, and stages. Chiu et al.'s work uses a linear

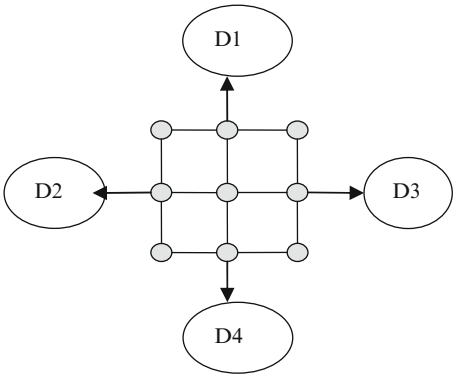


Fig. 3 Notional representation of destination choice

programming version of the cell transmission model to account for traffic and the well-known interior-point solution method (general technique (1)). Their objective function minimizes travel time of all evacuees in the network during the evacuation operation time horizon (system optimal), but they suggest that maximizing evacuation flow or minimizing exposure would also be suitable objectives. Outcomes of these staging models can then be incorporated into simulation tools, such as Liu et al.'s (2008) work.

Destination Assignment

From an optimization perspective, the destination choice should minimize the time for evacuees to reach safety. As mentioned in the discussion of staged evacuation, several researchers (e.g., Chen and Zhan 2004; Sbayti and Mahmassani 2006; Chiu et al. 2006) combined destination choice optimization with staging and routing (in some cases). Others (e.g., Chen 2005) have modeled destination choice using planning methods, such as the intervening opportunities model and gravity model (Wilmot et al. 2006).

Han et al. (Han et al. 2006) combined the optimization of destination and route, taking departure time as given. They treat the network as having one destination (e.g., connected to D1, D2, D3, and D4 in Fig. 3 with dummy links). The true destination (e.g., D1–4) is found by tracing the flow on the dummy links. Their optimization formulation is a traffic assignment problem with the objective function to minimize user optimal or system optimal travel time. They used traffic simulation, which involves iterations to approximate user equilibrium or system optimum traffic assignment, to solve their problem, which was single level, rather than the bilevel problems frequently discussed above. They tested their 1-destination framework in the context of a county-wide special event evacuation and found up to a 60% reduction in total evacuation time (Han et al. 2006).

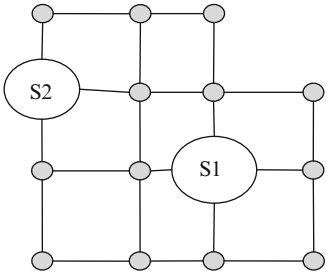


Fig. 4 Shelters within the evacuation network

Combined Departure Time, Destination, and Route Optimization

As discussed in other sections, several elements of demand (e.g. demand scheduling, destination choice, and route) are often combined (e.g., Chiu et al. 2006). Abdelgawad et al. (2010) also considered these three aspects for auto evacuees in their multimodal model. As objective functions for auto evacuees, they considered minimizing vehicle travel time, waiting time, and waiting time and vehicle travel time. For the transit dependents, they used a variant of the vehicle routing problem that allows multiple depots, time constraints, and multiple pickup and delivery locations. In the transit context, Abdelgawad et al. (2010) optimize scheduling and routing to minimize in-vehicle travel time, minimize waiting time, and optimize fleet size, considering vehicle cost. Their solution techniques combined genetic algorithms and dynamic traffic assignment simulation (a combination of techniques (2) and (3)). They combine the three transit objectives into a single weighted objective function. They tested their model for a hypothetical event in downtown Toronto, Ontario, Canada and found accounting for waiting time at the origins and travel times yields a compromise solution between excessive waiting time and gridlock resulting from simultaneous evacuation (Abdelgawad et al. 2010).

Shelter Location and Assignment Optimization

Determining the optimal shelter location or assignment is closely related to destination choice, except the type of facility implied may be different within the immediate network (see Fig. 4) and the travel distance to the shelter may be shorter than the distance to an evacuee's preferred final destination of a friend or relative's home. Similar to the destination choice work, optimization of shelter choice has been considered in conjunction with routing or both routing and departure time (e.g. Afshar and Haghani 2008).

Afshar and Haghani (2008) sought to minimize total travel time and network clearance time by spreading traffic over time and squeezing the demand to speed the departure of the last evacuee from danger. Their approaches are heuristic based. Traffic is assigned according to the system optimum approach, which provides

evacuees with departure times and shelter locations, as well as routes. The objective function is the minimization of evacuees' total travel times.

Ng et al. (2010) also used the system optimal approach for the upper level of their bi-level optimization model. The upper level assigns evacuees to shelters while evacuees choose their routes to those shelters according to user equilibrium routing assumptions. Their model is quoted below (Ng et al. 2010) p. 549.

$$minimize \sum_a v_a t_a(v_a) \tag{12}$$

Subject to

$$\sum_s q_{rs} = O_r, \quad \forall r \in V \tag{13}$$

$$\underline{q}_{rs} \leq q_{rs} \leq \bar{q}_{rs}, \quad \forall r, s \in V \tag{14}$$

$$minimize \sum_a \int_0^{v_a} t_a(x)dx \tag{15}$$

Subject to

$$\sum_k f_{rs}^k = q_{rs}, \quad \forall r, s \in V \tag{16}$$

$$v_a = \sum_{k,r,s} \delta_{ars}^k f_{rs}^k, \quad \forall a \in A \tag{17}$$

$$f_{rs}^k \geq 0, \quad \forall r, s \in V, \forall k \tag{18}$$

where O_r is the estimated number of evacuees from origin node r, $r \in V$, $\underline{q}_{rs}$ is the nonnegative lower bound on q_{rs}, $\bar{q}_{rs}$ is the nonnegative upper bound on q_{rs}, f_{rs}^k is the flow on path k between origin r and shelter s, v_a is the flow on link a, $a \in A$, δ_{ars}^k is an indicator variable taking the value 1 if link a is on path k between origin r and shelter s, 0 otherwise, and q_{rs} is the number of evacuees from origin node r assigned to shelter s.

Equation (12) is the objective function of the upper level and minimizes total evacuation time. Equation (13) forces all evacuees to be assigned to some shelter. Equation (14) ensures capacity of the shelter is not exceeded and that the number of evacuees assigned there meets the planner-specified lower bound. Equations (15)–(18) represent the lower level problem which is the well-known deterministic user equilibrium problem that assigns evacuees to their shortest paths while meeting demand (16), obtaining link flow from path flow (17), and requiring non-negative path flow (18). The model is solved using the meta-heuristic technique simulated annealing. Testing on the Sioux Falls network revealed that shorter evacuation times can be achieved if evacuees are spread among the available shelters (Ng et al. 2010).

Li et al. (2011) also used a bilevel formulation to optimize the locations of shelters for a range of (hurricane) events. Routing was a key element of their model, for which they adopted a stochastic user equilibrium approach. Their optimization formulation consists of two levels, SUE at the lower level and the selection of which shelters to open at the upper level. The upper level is treated as a two-stage stochastic programming problem; the first stage locates shelters prior to observing a scenario and the second determines which shelters to open in a given situation and how to assign evacuees to them. The solution technique is heuristic based, involving iterations between the two levels.

Transit Operations Optimization

Transporting evacuees by transportation modes other than personal vehicles reduces the number of vehicles, and related congestion, and is also instrumental for evacuees without vehicles, either due to commuting mode choice or the lack of such a vehicle. To plan evacuation transit service, optimization and combinations of optimization and traffic simulation tools are frequently employed. From the optimization perspective, the decision variables often determine routes and schedules (e.g., Abdelgawad and Abdulhai 2010), stop locations (e.g., Chen and Chou 2009), and routes (e.g., Bish 2011). Zhang and He (2008) also developed an optimization approach for transit operations. Their model is a location-routing problem and has deterministic and stochastic versions. They considered the objective functions to minimize total evacuee travel time, minimize the longest bus running time, balance the loads and tasks for each bus, minimize total transport distance, and minimize the number of buses, but ultimately selected minimizing total evacuee travel time. Their constraints include restricting each bus stop to be served by only one vehicle, respecting shelter and vehicle capacities, route continuity, routing from and to one shelter, prohibiting direct connections between shelters, and only using open shelters. The solution technique involves the use of hybrid genetic heuristics. They found that deterministic plans applied in stochastic conditions could lead to exceeding the capacity of some of the transit vehicles, implying that some evacuees would be left behind (Zhang and He 2008) or presumably that some additional trips would have to be made that were not part of the original plan.

Summary and Concluding Comments

As the list of conditions for which evacuations could be used as a protective action grows, the need to better plan for and implement evacuations also grows. Over the past decade, many new planning and operational practices have come into use. Unfortunately, much of this new knowledge has come at the price of the difficult and sometimes tragic lessons learned from recent failures. However, in addition to revealing the limitations of current practice, the past two decades of evacuations

for floods, tsunamis, fires, chemical spills, nuclear power plant emergencies, and hurricanes has also permitted the collection and analysis of response data in quantities and levels of detail never before available. From these observations and datasets, it has been possible to develop new techniques to model and forecast the conditions that commonly accompany such events. One of the areas where some of the most significant advances have been in modeling and forecasting has been made is in the field of operations research.

This chapter provides an overview of the key components of evacuation processes and how these conditions can be represented quantitatively and adapted for a range of conditions, most importantly here, for evacuations associated with short- to no- advance notice hazmat events. The first group of these OR techniques is described for the representation of traffic flow conditions. Recent applications for traffic analysis have been used for, among other uses, to optimize traffic signal timing during emergencies; balance traffic between available evacuation routes, and examine how the closure of freeway ramps and the elimination of intersection turning movement can actually increase rates of flow and reduce overall clearance time. The OR models presented in this chapter have also been developed for high-level regional planning, including examining the effects of staged evacuation, how evacuees select the location of shelter destinations and how these resources can best be optimized. The last set of examples included in this chapter looked at evacuations for individuals who may chose or be required to utilize transit services and how to best serve their needs.

Although these techniques represent significant advancements over prior practice just a few years earlier, much improvement remains. First and foremost, many of these models are quite complex and generally reside beyond the ability of first responders to quickly and effectively apply them during most, if any, but a few emergency situations. To work their way into wide use, the OR techniques must be simple to use even for lower-trained individuals in transportation response organizations.

Next, in order to be useful and effective the information these models generate must be reliable, actionable, and helpful to evacuees and emergency responders and decision-makers. As with most transportation-related management techniques, all guidance must be timely, accurate, and useful to maintain credibility with the public. As such, the results from these systems must be predictive, rather than purely reactive, in nature so that issues and needs cannot merely be reacted to and but forecast well in advanced of adverse conditions. This allows plans to be modified and adapted to reduce the adverse impacts of not only present conditions but conditions that are anticipated to exist 1 or 2 or 8 or 24 h later.

Among the chief difficulties in achieving any of these goals is likely to be collecting, storing, collating, and accessing the potentially overwhelming amount of input data required to make them work. Questions remain of where such data would come from, where it would be stored, how would it be accessed, then applied quickly and simply to support operational decision-making. Finally, to be truly effective, it is also suggested that these models be unified into systems that can be used together transparently and in sequence rather than separately with each component focusing on specific aspects of the problem.

While none of these issues is able to be addressed in a single overview book chapter, it is anticipated that readers will be able to use the information presented here to move in these directions. To this end, interested readers are encouraged to review the list of references provided as well as to tap into the wide range of related works by these and other authors to seek ways of advancing the field by building on the existing body of work.

References

Abdelgawad H, Abdulhai B (2010) Managing large-scale multimodal emergency evacuations. J Transport Saf Secur 2(2):122–151

Abdelgawad H, Abdulhai B, Wahba M (2010) Multiobjective optimization for multimodal evacuation. Transport Res Rec 2196:21–33

Afshar AM, Haghani A (2008) Heuristic framework for optimizing hurricane evacuation operations. Transport Res Rec 2089:9–17

Bish DR (2011) Planning for a bus-based evacuation. OR Spectrum 33(3):629–654

Bish DR, Sherali HD, Hobeika AG (2006) Optimal evacuation strategies with staging and routing. Working Paper. Virginia Polytechnic Institute and State University, Blacksburg, VA

Bretschneider S, Kimms A (2011) A basic mathematical model for evacuation problems in urban areas. (Compendex).

Brown C, White W, van Slyke C, Benson JD (2009) Development of a strategic hurricane evacuation–dynamic traffic assignment model for the Houston, Texas, Region. Transportation Research record no. 2137, pp 46–53

Chen B (2005) Modeling destination choice in hurricane evacuation with an intervening opportunity model, civil engineering. Louisiana State University, Baton Rouge

Chen C-C, Chou C-S (2009) Modeling and performance assessment of a transit-based evacuation plan within a contraflow simulation environment. Transport Res Rec 2091:40–50

Chen X, Zhan FB (2004) Agent-based modeling and simulation of urban evacuation: relative effectiveness of simultaneous and staged evacuation strategies, 83rd annual meeting of the Transportation Research Board. Transportation Research Board, Washington, DC

Chen M, Chen L, Miller-Hooks E (2007) Traffic signal timing for urban evacuation. J Urban Plann Dev 133(1):30–42

Chiu Y-C, Korada P, Mirchandani PB (2005) Dynamic traffic management for evacuation, 84th annual meeting of the Transportation Research Board. Transportation Research Board, Washington, DC

Chiu Y-C, Villalobos J, Gautam B, Zheng H (2006) Modeling and solving optimal evacuation destination-route-flow-staging problem for no-notice extreme events, 85th annual meeting of the Transportation Research Board. Transportation Research Board, Washington, DC

Cova TJ, Johnson JP (2003) A network flow model for lane-based evacuation routing. Transport Res A Pol 37(7):579–604

Daganzo CF, So SK (2011) Managing evacuation networks. Transport Res B Meth 45:1424–1432

Dixit VV, Radwan E (2009) Hurricane evacuation: origin, route, and destination. J Transport Saf Secur 1(1):74–84

Dotson LJ, Jones J (2005) Identification and analysis of factors affecting affecting emergency evacuations - main report. U.S. Nuclear Regulatory Commission report no. NUREG/CR-6864, Vol. 1, SAND2004-5901, Washington, DC, p 48

Edara P, Sharma S, McGhee C (2010) Development of a large-scale traffic simulation model for hurricane evacuation—methodology and lessons learned. Nat Hazards Rev 11(4):127–139

Florida Department of Community Affairs (2010) Statewide regional evacuation study program, volumes 1–8, Tallahassee, FL http://www.sfrpc.com/sresp.htm

Fonseca DJ, Moynihan GP, Johnston J, Jennings J (2009) A simulation tool for hurricane evacuation planning. Model Simulat Eng 2009:10p (Article ID 729570)
Fu H, Wilmot CG (2004) A sequential logit dynamic travel demand model for hurricane evacuation. Transport Res Rec 1882:19–26
Ghanipoor Machiani S, Murray-Tuite P, Jahangiri A, Yin W, Liu S, Park B, Chiu Y-C (2012) Modeling ramp closure as a strategy to reduce evacuee travel time, second international conference on evacuation modeling and management, Chicago, IL
Ghanipoor Machiani S, Murray-Tuite P, Jahangiri A, Liu S, Park B, Chiu Y-C, Wolshon B (2013) No-notice evacuation management: ramp closures under varying budgets and demand scenarios, 92nd annual meeting of the Transportation Research Board, Washington, DC
Gudishala R, Wilmot CG (2012) A comparison of time-dependent sequential logit and nested logit for modeling hurricane evacuation demand. Transportation Research record no. 3959
Han LD, Chin S-M, Hwang H-L, Yuan F (2006) Proposed framework for simultaneous optimization of evacuation traffic destination and route assignment, 85th annual meeting of the Transportation Research Board. Transportation Research Board, Washington, DC
Hardy M, Wunderlich K, Bunch J (2009) Evacuation management operations (EMO) modeling assessment: transportation modeling inventory. USDOT Research and Innovative Technology Administration project no. 04050002-01, Washington, DC 2007, pp 48, http://www.its.dot.gov/its_publicsafety/emo/emo.pdf
Li A, Xu N, Nozick L, Davidson RA (2011) BiLevel optimization for integrated shelter location analysis and transportation planning for hurricane events. J Infrastruct Syst 17:184–192
Lin D-Y, Eluru N, Waller ST, Bhat CR (2009) Evacuation planning using the integrated system of activity-based modeling and dynamic traffic assignment. Transportation Research record no. 2132, pp 69–77
Lindell MK, Perry RW (1992) Behavioral foundations of community emergency planning. Hemisphere Publishing Company, Washington, DC
Lindell MK, Perry RW (2012) The protective action decision model: theoretical modifications and additional evidence. Risk Anal 32(4):616–632
Liu HX, Ban JX, Ma W, Mirchandani PB (2007) Model reference adaptive control framework for real-time traffic management under emergency evacuation. J Urban Plann Dev 133(1):43–50
Liu Y, Chang G-L, Liu Y, Lai X (2008) Corridor-based emergency evacuation system for Washington, DC: system development and case study. Transport Res Rec 2041:58–67
Luo Z, Liu Y (2012) Optimal location planning of signalized and uninterrupted-flow intersections in urban network during emergency evacuation, 91st annual meeting of the Transportation Research Board, Washington, DC
McGhee CC, Grimes MC (2006) An operational analysis of the hampton roads hurricane evacuation traffic control plan, Virginia Department of Transportation, Virginia Transportation Research Council, Richmond
Mitchell S, Radwan E (2006) Heuristic priority ranking of emergency evacuation staging to reduce clearance time. Transport Res Rec 1964:219–228
Ng M, Park J, Waller ST (2010) A hybrid bilevel model for the optimal shelter assignment in emergency evacuations. Comput Aided Civ Infrastruct Eng 25:547–556
Pel AJ, Bliemer MCJ, Hoogendoorn SP (2012) A review on travel behaviour modelling in dynamic traffic simulation models for evacuations. Transportation 39(1):97–123
Ren G, Huang Z, Cheng Y, Zhao X, Zhang Y (2012) An integrated model for evacuation routing and traffic signal optimization with background demand uncertainty. J Adv Transport 47:4–27
Sbayti H, Mahmassani HS (2006) Optimal scheduling of evacuation operations, 85th annual meeting of the Transportation Research Board. Transportation Research Board, Washington, DC
So SK, Daganzo CF (2010) Managing evacuation routes. Transport Res B Meth 44(4):514–520
Steinberg LJ, Sengu HN, Cruz AM (2008) Natech risk and management: an assessment of the state of the art. Nat Hazards 46(2):143–152
Wilmot CG, Mei B (2004) Comparison of alternative trip generation models for hurricane evacuation. Nat Hazards Rev 5(4):170–178

Wilmot CG, Modali N, Chen B (2006) Modeling hurricane evacuation traffic: testing the gravity and intervening opportunity models as models of destination choice in hurricane evacuation, p 60

Wolshon B, Urbina E, Wilmot C, Levitan M (2005) National review of hurricane evacuation plans and policies, part I: planning and preparedness. Nat Hazards Rev 6(3):129–142

Wolshon B, Catarella-Michel A, Lambert L (2006) Louisiana highway evacuation plan for hurricane Katrina: proactive management of regional evacuations, J Transport Eng 132(1):1–10

Xie C, Turnquist MA (2009) Integrated evacuation network optimization and emergency vehicle assignment. Transport Res Rec 2091:79–90

Xie C, Turnquist MA (2011) Lane-based evacuation network optimization: an integrated Lagrangian relaxation and tabu search approach. Transport Res C Emer 19(1):40–63

Xie C, Lin D-Y, Travis Waller S (2010) A dynamic evacuation network optimization problem with lane reversal and crossing elimination strategies. Transport Res E Log 46(3):295–316

Xie C, Waller ST, Kockelman KM (2011) Intersection origin–destination flow optimization problem for evacuation network design. Transport Res Rec 2234:105–115

Zhang L, He S (2008) optimum transit operations during the emergency evacuations, p 54

Zimmerman C, Brodesky R, Karp J (2007) Using highways for no-notice evacuations: routes to effective evacuation planning primer series. Federal Highway Administration, Office of Operations, Washington, DC, p 116

Index

A
Accident consequence, 2, 4, 5, 129, 130, 138, 142, 144–146
Accident probability, 2, 4, 51, 52, 114, 119, 130, 132, 133, 145, 150
Accident probability (or accident rate), 14, 16, 18, 21, 42, 106
Active set method, 94, 95
Agent-based traffic model, 277
Air dispersion, 3, 25
Air pollution dispersion, 4, 109–113, 122, 124
Analytic network process (ANP), 163
Anti-cent-dian, 5, 160, 173–175, 177
Anticenter, 159, 161
Anticenter-maxian, 160
Anti-Condorcet, 159
Antimedian, 159–161
Atmospheric condition, 26, 106, 110
Atom, 202–205, 208, 211, 213, 217, 219–222, 224, 234–241
Attacker, 44, 202–204, 206–208, 223, 224, 228, 230, 231, 234, 246, 248, 252, 258, 260–262, 264

B
Bender's decomposition, 200, 202, 206–208, 211, 216, 223, 228–231, 240
Bilevel optimization, 60, 63, 65, 84, 89
Blocking, 2, 12, 13, 32, 33, 42
Branch and bound, 94, 160

C
Carrier, 2–4, 50, 51, 60–73, 75, 76, 78, 80–84, 86–92, 95, 96, 128, 130, 188, 194, 198
Cell transmission model (CTM), 278, 285
Center problem, 160, 168–171, 176
Coherent risk measure, 137, 139
Conditional probability, 2, 5, 14–16, 18–20, 24, 25, 38, 106, 131
Conditional value-at-risk (CVaR) (or expected shortfall), 2, 4, 5, 127–152
Confidence level, 5, 23, 132, 133, 135–138, 141–151, 249
Continuous location, 157
Crossing elimination, 7, 281–283

D
Danger circle, 17, 27, 38, 194
Defender, 6, 21, 200, 202–208, 223–225, 228, 230, 234, 236, 238, 240, 246, 248, 252, 253, 256, 258–261, 263, 265
Dense gas model, 4, 113
Derailment, 2, 9, 13–16, 18–24, 32, 33, 38, 272, 273
Derailment rate (or derailment probability), 14, 15, 18, 19, 21, 22, 32, 33
Destination assignment, 7, 285–286
Detection, 200–206, 214, 236, 238, 255, 264
Discrete location, 157, 158
Dispersion of pollutants, 104, 105, 108–113
Dissimilar path, 3, 54–60
Disutility (DU), 5, 106, 131, 132, 162
Dual toll pricing, 93
Dynamic programming, 53, 54
Dynamic traffic assignment (DTA), 276, 279, 286

R. Batta and C. Kwon (eds.), *Handbook of OR/MS Models in Hazardous Materials Transportation*, International Series in Operations Research & Management Science 193, DOI 10.1007/978-1-4614-6794-6, © Springer Science+Business Media New York 2013

E
Emergency evacuation, 2, 7, 269–290
Emergency resource deployment, 2, 6, 245–266
Emergency response, 16, 17, 25, 59, 161, 196, 270
Emergency team, 250–252, 254, 255
Emergency vehicle, 250, 283
Equity, 3, 4, 51, 54–60, 62, 64, 73, 74, 87, 89, 95, 160–162, 195, 197
Evacuation distance, 17, 29, 30, 38
Evacuation simulation, 275–279
Expected consequence (or traditional risk), 2, 3, 10, 14–20, 25, 33, 35, 36, 38–41, 45, 63, 103, 131, 132, 141, 149, 151, 195
Exposure band, 2, 17

F
Facility location, 2, 5, 155–183
Follower, 60–63, 65–69, 74–76, 78–81, 85, 89, 198, 203, 258, 259

G
Game theory, 44
Gaussian plume model, 3, 4, 25, 109–113, 115, 116, 121, 122, 124
Geographic information system (GIS), 6, 106, 112, 123, 160, 183, 270
Global routing (or global route planning), 1–3, 49–96, 130
Governmental authority, 3, 50, 51, 60, 62, 63, 74, 95, 130

H
Hazardous materials (hazmat), 1, 9, 50, 127, 188, 246
Hazardous-network design problem (HDP), 198
Hazmat transportation network design (HTND), 3, 51, 54, 55, 60–81
Heuristic, 6, 39, 49, 53, 54, 58, 59, 61–63, 70, 71, 79–81, 106, 119, 188, 202, 203, 221–240, 256, 280, 283, 286, 288
Hijack, 196, 246

I
Incident probability, 2, 5, 10, 14, 18–25, 32, 43, 45, 52, 131, 189, 194, 196, 197
Intrinsic road characteristics, 107
Itinerary, 2, 3, 10–12, 34, 36–38

K
Karush-Kuhn-Tucker (KKT) optimality conditions, 67, 76, 91
Knapsack approximation, 221, 223–228, 241
k-shortest path, 55, 56, 58, 228–234

L
Lagrangean dispersion model, 113
λ-anti-cent-dian, 6, 174–179, 182
λ-cent-dian, 160
Leader, 60–62, 65–71, 74–76, 78–81, 85, 86, 89, 198, 258, 259
Length-based shortest-path network interdiction problem (SPNIP-LB), 203, 205–206, 208, 211–222, 234, 235, 241
Local routing (or local route planning), 1–4, 50–54, 130
Location Analysis, 156–159
Location theory, 156–158
Long-notice event, 273

M
Macro-level model (macroscopic model), 7, 277, 278
Man-made disaster, 7
M-AST, 208, 209, 212, 221, 223
Mathematical Program with Equilibrium Constraints (MPEC), 4, 60, 64, 87, 89, 90, 93, 94
Maxian, 5, 159, 161, 168, 171–174, 177, 179, 182, 183
Maximin, 159, 160, 162, 163, 168, 173, 174, 176
Maximum risk, 5, 57, 131, 132, 136, 141, 145
Maxisum, 159, 161, 171, 173, 174, 176
Maxmax, 159, 169, 176
Mean-variance, 5, 131, 132
Median problem, 176, 182
Meso-level model (mesoscopic model), 7, 278
Micro-level model (microscopic model), 7, 277
Minimax, 56, 57, 159, 162, 169, 176, 261
Minisum, 159, 163, 176
M-NAMT, 208, 209, 211
M-NAST, 208, 209, 212, 221, 223
Modified shortest-path network interdiction problem (SPNIP-M), 202–206, 208, 209, 211, 212, 217–222, 224, 225, 228, 230, 231, 234–236, 238, 240, 241
Multi-commodity flow network interdiction problem (MFNIP), 198, 199

Multicriteria (or multiple criteria), 5, 6, 155, 156, 158, 162–164, 166–168, 175–179, 182, 183, 197
Multicriteria networks, 6, 156, 163, 176–182

N
Nash equilibrium, 89–91, 261, 265
Nash game, 64, 87, 88
Natech disaster, 7, 272
Natural disaster, 7, 197, 247, 248, 271, 272
Network design, 2, 3, 51, 54, 55, 60–83, 87, 96, 163
Network interdiction, 2, 6, 187–241
Network location, 156–159, 161, 163, 165, 166, 171
Network segmentation, 195
No-notice event, 273, 274
Noxious facility, 5, 155, 176

O
Obnoxious facility, 5, 159, 160, 162, 163
Ordered median problem (OMP), 158

P
Pareto optimal, 34, 168, 246
Perceived risk, 5, 131, 132, 162
Point of derailment (POD), 2, 18–23
Population exposure, 2, 3, 5, 10, 14, 16, 17, 19, 21, 25–31, 34, 35, 38, 39, 41, 43, 45, 52, 54, 82, 86, 87, 94, 106, 131, 196, 198, 201
Probability neglect, 248
Protective distance, 109

Q
Quantile-based risk measure, 137

R
Railroad, 2, 4, 9–45, 112, 272
Ramp management, 7, 280–281
Risk assessment, 1, 2, 4, 9–45, 50, 112, 113, 129, 130, 199
Risk-averse, 128, 132, 136, 145, 151
Risk management, 10, 13, 31–42, 45, 128, 251
Routing, 2–5, 10, 13, 34–43, 45, 49–54, 57–59, 61, 64, 69, 70, 91, 95, 106, 112, 113, 127–152, 159, 160, 162, 188, 197–199, 281–283, 285–288

S
Scale invariance, 249
Scheduling, 34, 42–43, 49, 52–54, 57, 158, 188, 286
Security, 3, 42–44, 53, 55, 62, 159, 187–189, 197, 199
Sensor, 6, 202–206, 208, 209, 211, 213, 214, 217, 219, 224–226, 234–236, 238, 240, 241
Service leg, 11, 34, 38, 39
Service network, 2, 11, 12
Shelter location, 7, 286–288
Shortest-path network interdiction problem (SPNIP), 6, 200, 202–205, 208, 211, 219, 221, 241
Shortest-path problem, 5, 53, 55, 58, 66, 104, 114, 115, 119, 131, 136, 141, 194, 198, 230
Short-notice event, 273
Signal retiming, 283–284
Spatial parameter, 7, 274–275
Stackelberg game, 6, 258–260
Staged/phased evacuation, 7, 272, 284–285

T
Technological disaster, 7, 272
Temporal parameter, 7, 273–274
Terrorist decision, 63, 246
Terrorist threat, 2, 6, 245–266
Threshold distance approach, 4, 108–109, 118
Time dependent shortest path problem, 104, 114, 115, 119
Toll setting, 3, 4, 51, 54, 55, 63–64, 81–96
Transit operations, 7, 288
Transportation phase, 128
Travel time, 4, 43, 52–55, 58, 59, 94, 103, 106, 113, 115–118, 164, 166, 198, 252, 254, 256, 274, 280, 281, 283–288

U
Uncenter, 5, 169–171, 173, 174, 177, 179, 182, 183
Undesirable center problem, 168–171, 176
Undesirable facility, 2, 5, 6, 155, 156, 158–163, 168–182

V
Value-at-risk (VaR), 2–5, 127–152
Variational inequality, 6, 260–265
Vehicle routing problem (VRP), 49, 50, 54, 188, 286

W
Weather condition, 4, 19, 104, 106–108, 112–115, 124, 194
Weather systems, 3, 4, 103–124
Wind direction, 27, 29, 104, 106, 108, 113, 122, 123
Wind speed, 3, 26, 27, 104, 111, 113, 122, 123
With-notice event, 273

Z
Zone segmentation, 195