European Journal of Operational Research 198 (2009) 452–463

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing and Logistics

An optimization based approach for deployment of roadway incident response vehicles with reliability constraints

Raktim Pal [a,*], Indranil Bose [b,1]

[a] College of Business, James Madison University, Showker Hall, MSC 0202, Harrisonburg, VA 22807, USA
[b] School of Business, The University of Hong Kong, Pokfulam Road, Hong Kong

ARTICLE INFO

Article history:
Received 9 August 2007
Accepted 9 September 2008
Available online 19 September 2008

Keywords:
Transportation
Facilities planning and design
Optimization
Incident management
Emergency response

ABSTRACT

In recent years transportation agencies have introduced patrol based response programs to remove roadway incidents rapidly. With the evolution of technology incident detection and notification from remote traffic operation centers is possible and patrols to detect incidents are not necessary. Instead, the response units can be placed at various depots in urban areas and dispatched to incident sites upon notification. In this paper, we propose a reliability based mixed integer programming model to find best locations of incidence response depots and assign response vehicles to these depots so that incidents can be cleared efficiently at a minimum cost. The approach is unique as it considers fixed and variable costs of vehicles and depots, occurrences of major and minor incidents, and reliability of response service in the same model. Numerical results are generated for an example problem and sensitivity analysis is conducted to explore the relationships between parameters of the problem.

1. Introduction

Roadway congestion is a common phenomenon causing delays and frustrations for millions of road users. Incidents are a major cause of congestion (FHWA, 2004). Incidents can be defined as non-recurrent events causing reduction in roadway capacity or abnormal increase in demand. There are many types of incidents such as crashes, disabled vehicles, spilled loads, debris on road. In the US urban area congestion caused 3.7 billion hours of delay and 2.3 billion gallons of wasted fuel, costing travelers more than \$63 billion in 2003 (Schrank and Lomax, 2005). Every year more than 42,000 people die in road crashes in the US (NHTSA, 2005).

In response to the growing adverse impacts of incidents, several metropolitan areas have developed incident management programs (FHWA, 2001). Several patrol programs have been introduced where incident response vehicles patrol highway corridors or road networks to clear incidents as soon as possible and keep the traffic moving. With electronic surveillance of roadways, incident detection is possible from a remote traffic operation center and random patrols are not necessary. Instead, incident response vehicles can be assigned to depots located in different parts of an urban area. Upon detection of an incident at the traffic operation center, the response vehicle can be dispatched to the incident site from an appropriate depot. Such arrangements not only save the cost of patrol, but also reduce the response time (Larson and Odoni, 1981). For urban areas where Advanced Traffic Management Systems (ATMS) is deployed, depot based incident response system can be quite useful. ATMS employs a variety of sensory and communication equipments to monitor traffic, coordinate signal timings, and manage traffic flow (Hellinga et al., 2005). A number of traffic monitoring technologies including magnetic loop detectors, infrared sensors, microwave radars, ultrasonic detectors, and video cameras are available to be deployed in ATMS (Berka and Lall, 1998). For example, the Colorado Department of Transportation has deployed cameras to monitor and manage the State of Colorado's roadways (Mesenbrink, 2002). Also, the incident detection techniques that process information collected by the traffic monitoring technologies to identify the location and nature of incidents are getting more reliable (Karim and Adeli, 2003). Advanced Crash Notification (ACN) systems are capable of notifying the crash locations of vehicles and are being installed as an additional option on a limited number of high value luxury vehicles. Akella et al. (2003) discussed the effectiveness of the ACN system in reducing incident response time. With the development of automated incident detection and notification systems depot based incident response programs appear to be more promising than ever before. The knowledge of optimal location of depots and optimal number of vehicles at each depot is necessary for effective operation of such programs. In the following sections of this paper, we present an optimization model that determines locations of such depots and assigns vehicles to the depots for responding to incidents efficiently.

* Corresponding author. Tel.: +1 540 568 7094; fax: +1 540 568 3273.
E-mail addresses: palrx@jmu.edu (R. Pal), bose@business.hku.hk (I. Bose).
[1] Tel.: + 852 2241 5845; fax: +852 2858 5614.

0377-2217/\$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2008.09.010

2. Literature review

Emergency response has been a popular area of study in the operations research community. Toregas et al. (1971) formulated a set covering problem to determine the locations of service stations from where the emergency vehicles responded to emergency calls and the number of service stations was minimized. Church and ReVelle (1974) developed a model for the maximum coverage location problem (MCLP) that found the locations of a pre-specified number of facilities so as to maximize the demand that is covered by at least one facility. The model assumed that response vehicle (or server) located at a facility would always be available to serve demand from zones it has been assigned to and would never be busy. However, this might not be the case in reality. The hypercube model introduced by Larson (1974) proved to be a useful tool in planning emergency services in urban areas. The performance of the system was analyzed by using the model of spatially distributed queues with multiple servers. The model represented each server individually, determined the probability that each server would become busy, and incorporated a complex dispatching policy. It was necessary to find all possible state probabilities by solving a system of linear equations, which were used to compute system performance measures such as servers' travel time, dispatch frequency, and workload. Although the hypercube model was useful in analyzing various scenarios, it did not provide a direct solution to location problems.

Daskin (1983) used a model for maximum expected covering location problem (MEXCLP) to analyze the location of public service facilities. The objective was to maximize the expected coverage of demand by the facilities. The availability of servers was treated as a random variable and it was assumed that each server had equal probability of becoming busy and servers operated independently from each other. Batta et al. (1989) suggested that an approximate approach for relaxing the server independence assumption of MEXCLP would be to use the hypercube correction factor developed by Larson (1975). Saydam and Aytug (2003) combined MEXCLP with the hypercube approximation algorithm developed by Jarvis (1985) and proposed a genetic algorithm based approach to solve MEXCLP with a higher level of accuracy. Hogan and ReVelle (1986) incorporated the concept of backup coverage in their model. ReVelle and Hogan (1989) extended the MEXCLP by incorporating the concept of probabilistic location set covering problem (PLSCP) that included constraints for reliability of servers. ReVelle and Hogan's maximum availability location problem (MALP) model found locations of servers so that maximum demand was covered with stated reliability. It allowed server's busy fractions to be different at different parts of the service area. Like the MEXCLP model of Daskin (1983), it also assumed that servers worked independent of each other. This assumption was relaxed by Marianov and ReVelle (1996). They modeled reliability using some results from M/G/s-loss queuing system. The model was called queuing maximum availability location problem (Q-MALP).

Batta and Mannur (1990) proposed a framework for locating response units where multiple units usually responded to an emergency event. They examined both set covering and maximum covering problems. Melachrinoudis (1994) used a cost based approach for locating facilities using integer programming. Akella et al. (2005) proposed a mixed integer programming based model that maximized coverage of emergency calls over cellular network and found optimal locations for base stations that were used to process calls to or from the mobile units in the network. Yi and Özdamar (2007) developed an integrated location-distribution model using mixed integer multi-commodity network flow for coordinating relief and evacuation operations during disaster.

Pirkul and Schilling (1988) formulated a model to select emergency service facilities as well as backup services that would minimize both fixed and variable cost of facilities under work load capacity constraints. Several other site specific applications appeared in Plane and Hendrick (1977), Schilling et al. (1979), and Eaton et al. (1985). A framework using a combination of optimization and simulation techniques that attempted to keep incident induced delay within some acceptable limit was proposed by Zografos et al. (1993). The issue of reliability in designing an emergency response system was modeled by a number of researchers as detailed in Daskin (1983) and ReVelle and Hogan (1989). However, not much information exists on how to determine the optimal location of depots as well as the optimal assignment of vehicles to these depots simultaneously. This problem was addressed in Ball and Lin (1993). Their work is an extension of the PLSCP model and they modeled reliability of service by keeping the probability of uncovered demand within an upper bound specified by the system designer. They assumed that all vehicles are of the same type and did not explicitly include the initial and operating costs of vehicles in their formulation. This is a restrictive assumption as the initial and operating costs of vehicles usually constitute a major portion of the total cost of an incident response system. In fact, it is reported that these costs constituted more than seventy percent of the total annual equivalent cost of the Hoosier Helper freeway patrol service program in Indiana (Latoski et al., 1999). Alsalloum and Rand (2006) extended the maximal covering location problem that was originally proposed by Church and ReVelle (1974). They considered the probability of covering a demand within a target time and identified optimal locations that would maximize the expected coverage. The second objective was to find minimum number of ambulances that would provide target service level. A goal programming framework was used to combine these two objectives. Although this work was a good extension of existing models, it had few limitations. There was an underlying assumption that demand was covered by the nearest open station, which might not be realistic when all ambulances in the nearest station were busy and a backup unit was dispatched from another station. Such backup coverage was not considered in the model. Also, the model needed to calculate the probability of reaching a demand node from a station within a target time. These values for all demand node/station pair combinations were not easily available and estimation of these values was quite tedious.

Our proposed model addresses some of the limitations in the available literature. Unlike previous studies, all types of major costs involved in operating an incident response system are included in the model described in this paper. This includes fixed and variable costs of both response vehicles and depots. To the best of our knowledge this has not been addressed before in related research. Also, a distinction between major and minor incidents is made to accommodate prioritized service provided by an incident response vehicle while explicitly considering all types of cost. In addition, the reliability of the response service is incorporated in the model.

The remainder of the article is organized as follows: the details of the model including notations, decision variables, model parameters, and formulation are presented in Section 3. Section 4 describes a numerical example. Section 5 discusses the results and provides insights on the results. The conclusions and future research directions are provided in Section 6.

3. Model formulation

The problem is addressed by developing a model so that each zone in the response area is serviced by at least one response vehicle. Incidents are classified into two categories, namely minor and major incidents. Most of the incidents encountered on a roadway