Computers & Industrial Engineering 78 (2014) 271–284

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

Two new models for redeployment of ambulances

Mohammad Maleki [a,b,*], Nahidsadat Majlesinasab [c], Mohammad Mehdi Sepehri [a,b]

[a] Department of Industrial Engineering, Tarbiat Modares University, Tehran, Iran
[b] Health Systems Optimization, Engineering and Informatics Laboratory, Tehran, Iran
[c] Department of Industrial Engineering, Isfahan University of Technology, Isfahan, Iran

ARTICLE INFO

Article history:
Available online 5 June 2014

Keywords:
Ambulance
Location
Redeployment
Generalized Ambulance Assignment Problem
Generalized Ambulance Bottleneck Assignment Problem

ABSTRACT

Nowadays, in all countries, public resources for healthcare are inadequate to meet the demands for the services. Therefore, policy makers should provide the most effective healthcare services to citizens within the limited available resources. During the past decades, lots of research using operations research techniques, tools, and theories has been applied to a wide range of problems in healthcare. Ambulance service planning is a branch of healthcare. Ambulance location and redeployment problems are considered two important issues in ambulance service planning. In these models, an attempt is made to maximize coverage in the districts and properly service the patients in emergency situations.

In this paper, at first we use an existing model, named MECRP[1], to locate ambulances in four districts of Isfahan, Iran. Then, we formulate a generalized assignment model with the aim of minimizing the total time traveled by ambulances. We also formulate a generalized bottleneck assignment model. The goal of this model is to minimize the maximum travel time. The proposed models specify the movement of the ambulances, using the out of MECRP determines the relocation of the ambulances in just one run for all possible combinations. In fact, each of these models specifies the movement of the ambulances from hospitals to stations or from stations to other stations based on its aim.

In addition, to shed light on the merits of the proposed models, computational results on experimental data from Isfahan EMS agency are provided. The results, using these models, show travel times can be significantly reduced. By the end of this paper, the corresponding conclusions are expressed.

1. Introduction

The goal of EMS[2] is to reduce mortality, disability, and suffering. EMS decision makers deal with the difficult task of locating ambulances, to quickly and optimally service emergency medical calls. Furthermore, they are often under pressure from contractual obligations or managerial goals to meet the standard levels of performance criteria. This problem becomes more complicated when the number of emergency calls, increases, operational costs increase or traffic conditions get worse. Ambulance location and relocation models can improve the levels of the performance criteria and alleviate such intricacies (Maxwell, Restrepo, Henderson, & Topaloglu, 2010).

These models will provide the maximum level of preparedness for different zones. They provide good coverage in the zones where dispatching of ambulances has led to a low-level of preparedness (Andersson & Värbrand, 2007). The demand for ambulances fluctuates throughout the week, depending on the day of the week, and even time of the day. Therefore, EMS operators can improve system performance by dynamic relocation of ambulances in response to fluctuating demand patterns (Rajagopalan, Saydam, & Xiao, 2008). Ambulance location and redeployment models are classified in three main categories as follows: deterministic static models, probabilistic static models, and dynamic models.

1.1. Deterministic static models

Early proposed ambulance location models were linear integer formulations (Brotcorne, Laporte, & Semet, 2003). Since these models did not consider the probability that a particular ambulance might be busy at a given time, they were deterministic (Rajagopalan & Saydam, 2009). Furthermore, because ambulances are not allowed to relocate among stations, these models were static (Brotcorne et al., 2003). Toregas, Swain, ReVelle, and Bergman (1971) published the first coverage model known as the set covering location problem (SCLP). This model looks for the least number of ambulances needed to cover all demand points within a given

* Corresponding author at: Department of Industrial Engineering, Tarbiat Modares University, Tehran, Iran. Tel.: +98 913 229 3775.
E-mail address: Maleki65@Gmail.com (M. Maleki).

[1] Maximal Expected Coverage Relocation Problem.
[2] Emergency Medical Services.

http://dx.doi.org/10.1016/j.cie.2014.05.019
0360-8352/

response-time. Response-time is the period between an emergency call is recorded and the time the first ambulance arrives at the scene in a life-threatening case. This model ignores several aspects of real-life problems, the most important being, that once an ambulance is dispatched, some demand points are no longer covered. It also assumes that up to W ambulances are available, which is not always the case in practice (Brotcorne et al., 2003). Since SCLP treats all demand points equally, the solution may require more servers than needed or underestimates the number of ambulances needed for those locations with relatively heavy demand (Saydam & Aytug, 2003). To avoid such outcomes, Church and ReVelle developed the maximal covering location problem (MCLP) (Church & ReVelle, 1974). The goal of their model was to maximize the population covered with limited resources.

In each of the above models, coverage may become inadequate when vehicles become busy. Hence, to compensate for this shortcoming, Daskin and Stern (1981) extended the set covering model to include an objective of maximizing the number of zones covered by more than one vehicle. In fact, with high levels of multiple coverage, a partially congested system may more often be capable of responding to demand within the distance standard even when the most desirable unit is busy. Eaton & Morgan, 1986 applied a multi objective formulation to determine the optimal locations of emergency medical services in Santo Domingo, Dominican Republic. The model minimized the number of facilities and maximized multiple demand coverage. Hogan and ReVelle (1986) extended the maximal covering model and the set covering model by adding a second objective to maximize the number calls covered by 2 or more vehicles. Gendreau, Laporte, and Semet (1997) developed a model known as the double standard model (DSM). This model maximizes the demand covered by at least two vehicles, implicitly taking into account the fact that vehicles might become unavailable, while ensuring certain extra requirements concerning coverage are met.

The traditional definition used in MCLP model is that a demand node is assumed to be covered completely if it is within the target distance, otherwise it will not be covered. Since the optimal solution to a MCLP is likely sensitive to the choice of the target distance, this definition may lead to erroneous results. Karasakal and Karasakal (2004) allowed the coverage to change from "covered" to "not-covered" within a distance range instead of a single target distance and called this intermediate coverage level *partial coverage*. Then, they formulated the MCLP in the presence of partial coverage. Doerner & Hartl, 2008 modified the objective function of the DSM, and implemented the *Tabu Search* in Gendreau, Laporte, and Semet (2001) to obtain near optimal solutions. Drezner, Drezner, and Goldstein (2010) proposed a stochastic gradual coverage model in which the short and long distance standards are random variables. Berman, Drezner, and Krass (2010) formulated the Cooperative Location Set Covering Problem (CLSCP) and the Cooperative Maximum Covering Location Problem (CMCLP). The goal of these models is to replace the "individual coverage" assumption with a mechanism where all facilities contribute to the coverage of each demand point.

1.2. Probabilistic static models

Deterministic models did not account for the probability of a particular ambulance being busy at a given time. As a result, they either underestimated the number of ambulances needed, or overestimated the actual coverage provided (Rajagopalan & Saydam, 2009). Hence, to compensate for this shortcoming, probabilistic models were developed. Probabilistic location models recognize any given ambulance may be busy when it is called. Such uncertainty can be modeled within the mathematical programming formulation or using a queuing framework (Rajagopalan et al., 2008).

One of the first probabilistic models for ambulance location is the Maximum Expected Coverage Location Problem (MEXCLP) due to Daskin (1983). He removed the implicit assumption of deterministic models, all units are available at all times, by assuming each ambulance has the same probability q, called the busy fraction, of being unavailable to answer a call. He also assumed all ambulances are independent. Given a predetermined number of response units, MEXCLP maximizes the expected coverage subject to a response–time standard. Saydam and McKnew (1985) used a separable programming approach to reformulate the MEXCLP into a nonlinear form.

ReVelle and Hogan (1989b) extended the notion of MEXCLP by introducing the probabilistic location set covering problem (PLSCP) model to utilize a region-specific busy fraction instead of a system-wide busy fraction as in the MEXCLP. PLSCP includes a set of constraints on the reliability of a server being available. Since PLSCP will usually lead to a potentially large number of servers being assigned or required, ReVelle and Hogan (1989a) developed a chance constrained stochastic model called MALP (Maximum Availability Location Problem). This model distributes a fixed number of servers to maximize the population covered within a response–time standard and with a predetermined reliability. Since the MALP attempts to make the best possible use of available limited resources, this model is more applicable than the PLSCP to real world problems. There are two versions of the MALP. The MALP-I assumed that the facilities had the same busy fraction q. However, in the MALP-II, the busy fraction q_i associated with demand point i was computed as the ratio of the total duration of all calls generated from demand point i to the total availability of all facilities in W_i.

The common assumptions to the MEXCLP and its new forms, such as the same busy fraction and ambulance independence, make models easier to build and solve. However, MEXCLP models lack an accurate estimation of the expected coverage. Some researchers resorted to hypercube queuing models to obtain better estimation of the expected coverage. Batta, Dolan, and Krishnamurthy (1989) developed an approximate way to relax the server independence assumption in MEXCLP in using the hypercube correction factor developed by Larson (1981). This correction factor, applied to the MEXCLP objective function, led to an adjusted model, which the authors called AMEXCLP. Ball and Lin (1993) formulated a new version of PLSCP called Rel-P. This model incorporates a linear probabilistic constraint on the number of vehicles required to achieve a given reliability level. Later Marianov and ReVelle (1994) extended PLSCP using the assumption of region-specific busy fraction in MALPII to formulate queuing probabilistic location set covering problem (Q-PLSCP). Marianov and ReVelle (1996) proposed the queuing maximal availability location problem (Q-MALP). The main difference between MALP and Q-MALP is the methodology for the calculation of the smallest integer which satisfies the required reliability.

Beraldi, Bruni, and Conforti (2004) developed a stochastic programming model with probabilistic constraints aimed to solve both the location and the dimensioning problems, i.e. where service sites must be located and how many emergency vehicles must be assigned to each site, in order to achieve a reliable level of service and minimize the overall costs. In fact, they incorporated probabilistic constraints in their model to ensure all requests are served with a prescribed high probability. Galvão, Chiyoshi, & Morabito, 2005 presented a unified view of MEXCLP and MALP, named EMALP, by dropping their simplifying assumptions of servers operating independently and with the same busy probability. Alsalloum and Rand (2006) extended MCLP and developed Goal Programming models. They first determined locations of facilities to maximize expected demand coverage and then adjusted the capacity of each station.