Les Cahiers du GERAD ISSN: 0711–2440

Ambulance Location and Relocation Models

L. Brotcorne, G. Laporte
F. Semet

G–2000–53

October 2000
Revised: November 2001

Ambulance Location and Relocation Models

by

Luce Brotcorne [1]
Gilbert Laporte [2,3]
and
Frédéric Semet [1,4]

October, 2000
Revised: November, 2001

Les Cahiers du GERAD
G–2000–53

---

[1] LAMIH-ROI, Université de Valenciennes, Le Mont-Houy, 59304 Valenciennes Cedex 9, France.

[2] GERAD and Canada Research Chair in Distribution Management, École des Hautes Études Commerciales, 3000 chemin de la Côte-Sainte-Catherine, Montréal, Canada H3T 2A7.

[3] Centre de recherche sur les transports, Université de Montréal, Case postale 6128, Succursale "Centre-ville", Montréal, Canada H3C 3J7.

[4] GRIS, Département d'administration de la santé, Université de Montréal, Case postale 6128, succursale "Centre-Ville", Montréal, Canada H3C 3J7.

Abstract

This article traces the evolution of ambulance location and relocation models proposed over the past thirty years. The models are classified in two main categories. *Deterministic* models are used at the planning stage and ignore stochastic considerations regarding the availability of ambulances. *Probabilistic* models reflect the fact that ambulances operate as servers in a queueing system and cannot always answer a call. In addition, *dynamic* models have been developed to repeatedly relocate ambulances throughout the day.

Key words: Ambulances, emergency medical services, location, relocation, redeployment.

Résumé

Cet article passe en revue l'évolution des modèles de positionnement et de repositionnement d'ambulances proposés au cours des trente dernières années. Les modèles *dynamiques* sont utilisés au stade de la planification et font abstraction de la disponibilité des ambulances. Les modèles *probabilistes* reflètent le fait que les ambulances fonctionnent comme des serveurs dans un système de files d'attente et ne peuvent pas toujours répondre à un appel. De plus, des modèles *dynamiques* ont récemment été développés pour relocaliser de façon répété les ambulances pendant la journée.

Mots clefs: Ambulances, systèmes de véhicules d'urgence, localisation, relocalisation, redéploiement.

1 Introduction

This review article traces the evolution of ambulance location and relocation models proposed over the past thirty years. This period was marked by an unprecedented growth not only in computer technology, but also in modeling and algorithmic sophistication, in the performance of mathematical programming solvers, and in the widespread adoption of computer software at several levels of decision making. The literature on ambulance positioning systems truly reflects this evolution. The first models proposed were unsophisticated integer linear programming formulations, but over time more realistic features were gradually introduced, and solution techniques also evolved.

Most of the early models dealt with the *static* and *deterministic* location problem. These were meant to be used at the planning stage and they ignored stochastic considerations. Several *probabilistic* models were then developed to reflect the fact that ambulances operate as servers in a queueing system and are sometimes unavailable to answer a call. *Dynamic* models are more recent. They address the problem of repeatedly relocating ambulances in the same day to provide better coverage. In recent years, the development of powerful local search algorithms, particularly tabu search (Glover and Laguna, 1997), coupled with the growth of parallel computing (Crainic and Toulouse, 1998) have given rise to a new stream of research that deals effectively with the dynamic nature of the problem. With the newest models and algorithms, large scale problems can be solved rapidly and dynamically in real time, with a high level of accuracy.

There exists a rich literature on emergency vehicles siting models. The survey by Marianov and ReVelle (1995) provides an overview of the most important models published until that date. Our review is less general since it focuses on ambulance services, but it unavoidably covers some of the same material, albeit with a different emphasis. The Marianov-ReVelle survey ends with an indirect reference to dynamic relocation models integrated within geographic information systems: "rarely have ambulances been positioned at free standing stations" ..., "Lastly we have a warming competition ... The technique of GIS" (p.223). Our survey precisely addresses this issue by devoting a section to dynamic relocation models which have just started to emerge. We also report on actual implementations of ambulance location and relocation models. Finally, we provide a synthetic overview, in table form, of all the models we discuss.

The article is structured as follows. In Section 2, we briefly describe the functioning of emergency medical services. Two early models developed for the static case are described in Section 3. An important shortcoming of these models is that they may no longer guarantee adequate coverage as soon as ambulances dispatched to a call become unavailable. Two types of models have been developed to handle the need to provide extra coverage: deterministic models and probabilistic models. These are presented in Sections 4 and 5, respectively. We have chosen to concentrate on the most important models, leaving aside several minor variants already listed in the articles of ReVelle (1989), Swersey (1994) and Marianov and ReVelle (1995). In Section 6, we provide an account of some of the emerging research in the area of dynamic ambulance repositioning. A summary and conclusions follow in Section 7.

2 How emergency medical services operate

The chain of events leading to the intervention of an ambulance to the scene of an incident includes the following four steps: 1) incident detection and reporting, 2) call screening, 3) vehicle dispatching and 4) actual intervention by paramedics. Decisions made by emergency services managers are concerned with the second and third steps. The main function of the screening process is to determine the severity of the incident and its degree of urgency (e.g., on a one-to-four scale), and to make a decision on the type and number of ambulances to dispatch. Since time is vital in emergency situations, it is critical that vehicles be at all time located so as to ensure an adequate coverage and a quick response time (Wisborg et al., 1994). The United States Emergency Medical Services Act (see Ball and Lin, 1993) sets some standards: in urban areas, 95% of requests should be served within 10 minutes; in rural areas, they should be served within 30 minutes. This is where ambulance location and relocation models and algorithms come into play. Advanced information technologies are now often used to assist the ambulance management process. These include road network surveillance (Heddebaut, 1997; Cohen, 1999), vehicle positioning systems (Bouveyron and Didier, 1993; American TriTech, 1996), geographical information systems (Bernhardsen, 1999), and artificial intelligence based call screening systems (Clawson and Dernocoeur, 1991). Ideally, these systems should be fully integrated and interconnected within an ambulance relocation module.

From a medical and economic point of view, it seems advisable for an urban emergency medical services system to operate several vehicle types serially (Stout, Pepe and Mosesso, 2000). This has been shown to increase the global system performance in trauma cases and cardiac incidents (Haas et al., 1995). Emergency medical services typically work with two types of providers having different capabilities: basic life support (BLS) units and advanced life support (ALS) units, both of which are often dispatched to the same incident, but within different time standards (Mandell, 1998). In several North-American cities, BLS is assured by firemen trained as paramedics. They are based at local fire stations and are often the first to arrive on scene. ALS is covered by ambulances. Most calls can be served by only one ambulance, but on occasions two or more are required.

There are important differences between the operations of emergency medical services and those of fire companies or police departments. First, ambulances are not always based in a building, but often at a very rudimentary location such as a parking lot. More importantly, they are periodically relocated to insure a good coverage at all times. Ambulances do not normally patrol streets between calls, but once they are dispatched to the scene of an incident, they may be diverted to a more important call. Police cars, on the contrary, regularly perform patrol duties since their presence on city streets acts as a crime deterrent. For further readings on fire and police operations, the interested reader is referred to Larson (1972), Walker, Chaiken and Ignall (1979), Swersey (1994), and Adams (1997).

3 Two early models for the static ambulance location problem

Ambulance location models are defined on graphs. The set of demand points is denoted by V and the set of potential ambulance location sites is denoted by W. The shortest travel time t_{ij} from vertex i to vertex j of the graph is known. As is common in location theory, assigning demands to a discrete set can be achieved through an aggregation process which unavoidably results in a loss of accuracy. Various techniques have been proposed to measure and control the error bound (Erkut and Bozkaya, 1999; Francis, Lowe and Tamir, 2000). A demand point $i \in V$ is said to be *covered* by site $j \in W$ if and only if $t_{ij} \leq r$, where r is a preset coverage standard. Let $W_i = \{j \in W : t_{ij} \leq r\}$ be the set of location sites covering demand point i.

In the *Location Set Covering Model* (LSCM) introduced by Toregas, ReVelle, Swain and Bergman (1971), the aim is to minimize the number of ambulances needed to cover all demand points. It uses binary variables x_j equal to 1 if and only if an ambulance is located at vertex j:

$$\text{(LSCM)} \qquad \text{Minimize} \qquad \sum_{j \in W} x_j \tag{1}$$

$$\text{subject to} \qquad \sum_{j \in W_i} x_j \geq 1 \qquad (i \in V) \tag{2}$$

$$x_j \in \{0,1\} \qquad (j \in W). \tag{3}$$

This model ignores several aspects of real-life problems, the most important probably being that once an ambulance is dispatched, some demand points are no longer covered. Some of the more sophisticated models described in Section 4 adequately address this shortcoming. The model also assumes that up to $|W|$ ambulances are available, which is not always the case in practice. It does, however, provide a lower bound on the number of ambulances required to ensure full coverage.

An alternative approach proposed to counter some of the shortcomings of the LSCM is to maximize population coverage subject to limited ambulance availability. In the *Maximal Covering Location Problem* (MCLP) originally proposed by Church and ReVelle (1974), d_i denotes the demand of vertex i, and p is the number of available ambulances. The binary variable y_i is equal to 1 if and only if vertex i is covered by at least one ambulance. The model is then:

$$\text{(MCLP)} \qquad \text{Maximize} \qquad \sum_{i \in V} d_i y_i \tag{4}$$

$$\text{subject to} \qquad \sum_{j \in W_i} x_j \geq y_i \qquad (i \in V) \tag{5}$$

$$\sum_{j \in W} x_j = p \tag{6}$$

$$x_j \in \{0,1\} \qquad (j \in W) \tag{7}$$

$$y_i \in \{0,1\} \qquad (i \in V). \tag{8}$$

Each of the two models LCSM and MCLP makes sense in its own right. The first can be used as a planning tool to help determine the right number of vehicles to cover all demand, while the second attempts to make the best possible use of available limited resources. Several extensions of both models have been proposed in the ambulance location literature. A sensible approach is to repeatedly solve MCLP with increasing values of p until all demand is covered. A tradeoff between cost and coverage can then be made.

Eaton *et al.* (1985) have used MCLP to plan the reorganization of the emergency medical service in Austin, Texas. The proposed plan has saved the city \$3.4 million in construction costs, and \$1.2 million annually in operating costs, in 1984. In addition, average response time has been reduced despite an increase in calls for service.

4 Deterministic static models with extra coverage

Neither LSCM nor MCLP recognizes the fact that on occasions vehicles of several types may be dispatched to the scene of an incident. Also, even if only one vehicle type is used, solving MCLP alone may not provide a sufficiently robust location plan. We present in this section a number of deterministic models developed to deal with the issue of multiple coverage. Probabilistic models will be presented in Section 5.

One of the first models developed to handle several vehicle types is the *Tandem Equipment Allocation Model*, or TEAM (Schilling *et al.*, 1979). It applies naturally to fire companies that operate with two types of equipment (pumpers and rescue ladders), but it is also relevant in an ambulance location context where BLS and ALS units are used. Denote by p^A and p^B the number of vehicles of types A and B available, let r^A and r^B be the coverage standards for each vehicle type, and define $W_i^A = \{j \in W : t_{ij} \leq r^A\}$, $W_i^B = \{j \in W : t_{ij} \leq r^B\}$. Let $x_j^A(x_j^B)$ be a binary variable equal to 1 if and only if a vehicle of type $A(B)$ is located at vertex i, and let y_i be a binary variable equal to 1 if and only if vertex $i \in V$ is covered by two types of vehicle. The TEAM model can be written as follows:

$$\text{(TEAM)} \qquad \text{Maximize} \qquad \sum_{i \in V} d_i y_i \tag{9}$$

$$\text{subject to} \qquad \sum_{j \in W_i^A} x_j^A \geq y_i \qquad (i \in V) \tag{10}$$

$$\sum_{j \in W_i^B} x_j^B \geq y_i \qquad (i \in V) \tag{11}$$

$$\sum_{j \in W} x_j^A = p^A \tag{12}$$

$$\sum_{j \in W} x_j^B = p^B \tag{13}$$

$$x_j^A \leq x_j^B \qquad (j \in W) \tag{14}$$

$$x_j^A, x_j^B \in \{0,1\} \qquad (j \in W) \tag{15}$$

$$y_i \in \{0,1\} \qquad (i \in V). \tag{16}$$

This model is a direct extension of MCLP except for constraints (14) which impose a hierarchy between the two vehicle types. This constraint can of course be removed if circumstances warrant it. In the *Facility-Location, Equipment-Emplacement Technique*, or *FLEET* model (Schilling *et al.* 1979), constraints (14) are relaxed, but only p location sites may be used. A more elaborate model for fire protection siting, and belonging to the same family, was later developed by Marianov and ReVelle (1992). It can be used to locate capacitated fire stations with two types of equipment, subject to constraints ensuring that each demand point is adequately covered by the right number of pumper and rescue ladders.

In any of the above models, coverage may become inadequate when vehicles become busy. A strategy employed in the case of a single vehicle type is to modify MCLP in order to provide better multiple coverage, without increasing the total number of vehicles beyond p. As suggested by Daskin and Stern (1981) and by Hogan and ReVelle (1986), a second objective can be incorporated within MCLP to better distinguish between multiple optima of (4). In the first case the authors use a hierarchical objective to maximize the number of demand points covered more than once. In the second case, the total demand covered twice is maximized. Hogan and ReVelle (1986) also present two models *Backup Coverage Formulations*, called BACOP1 and BACOP2, incorporating binary variables y_i equal to 1 if and only if demand point $i \in V$ is covered once by an ambulance within a coverage standard r, and binary variables u_i equal to 1 if and only if i is covered twice within r. The two models are:

$$\text{(BACOP1)} \qquad \text{Maximize} \qquad \sum_{i \in V} d_i u_i \tag{17}$$

$$\text{subject to} \qquad \sum_{j \in W_i} x_j - u_i \geq 1 \qquad (i \in V) \tag{18}$$

$$\sum_{j \in W} x_j = p \tag{19}$$

$$0 \leq u_i \leq 1 \qquad (i \in V) \tag{20}$$

$$x_j \geq 0 \qquad (i \in V) \tag{21}$$

and

$$\text{(BACOP2)} \qquad \text{Maximize} \qquad \theta \sum_{i \in V} d_i y_i + (1-\theta) \sum_{i \in V} d_i u_i \tag{22}$$

$$\text{subject to} \qquad \sum_{j \in W_i} x_j - y_i - u_i \geq 0 \qquad (i \in V) \tag{23}$$

$$u_i - y_i \leq 0 \qquad (i \in V) \tag{24}$$

$$\sum_{j \in W} x_j = p \tag{25}$$

$$0 \leq u_i \leq 1 \qquad (i \in V) \tag{26}$$

$$0 \leq y_i \leq 1 \qquad (i \in V) \tag{27}$$

$$x_j \geq 0, \qquad (i \in W) \tag{28}$$

where θ is a weight chosen in $[0, 1]$.

In the model proposed by Gendreau, Laporte and Semet (1997), two coverage standards are used: r_1 and r_2, with $r_1 < r_2$. All demand must be covered by an ambulance located within r_2 time units, and a proportion α of the demand must lie within r_1 time units of an ambulance, which may possibly coincide with the ambulance that covers that demand within r_2 units. The United States Emergency Medical Services Act of 1973 sets a value of 10 minutes for r_1, but no value for r_2, and $\alpha = 0.95$. The *Double Standard Model* (DSM) of Gendreau, Laporte and Semet seeks to maximize the demand covered twice within a time standard of r_1, using p ambulances, at most p_j ambulances at site j, and subject to the double covering constraints.

Let $W_i^1 = \{j \in W : t_{ij} \leq r_1\}$ and $W_i^2 = \{j \in W : t_{ij} \leq r_2\}$. The integer variable y_j denotes the number of ambulances located at $j \in W$ and the binary variable x_i^k is equal to 1 if and only if the demand at vertex $i \in V$ is covered k times ($k = 1$ or 2) within r_1 time units. The formulation is then:

$$\text{(DSM)} \qquad \text{Maximize} \qquad \sum_{i \in V} d_i x_i^2 \tag{29}$$

$$\text{subject to} \qquad \sum_{j \in W_i^2} y_j \geq 1 \qquad (i \in V) \tag{30}$$

$$\sum_{j \in V} d_i x_i^1 \geq \alpha \sum_{i \in V} d_i \tag{31}$$

$$\sum_{j \in W_i^1} y_j \geq x_i^1 + x_i^2 \qquad (i \in V) \tag{32}$$

$$x_i^2 \leq x_i^1 \qquad (i \in V) \tag{33}$$

$$\sum_{j \in W} y_j = p \tag{34}$$

$$y_j \leq p_j \qquad (j \in W) \tag{35}$$

$$x_i^1, x_i^2 \in \{0, 1\} \qquad (i \in V) \tag{36}$$

$$y_j \quad \text{integer} \qquad (j \in W). \tag{37}$$

Here, the objective function computes the demand covered twice within r_1 time units, constraints (30) and (31) express the double coverage requirements. The left-hand side of (32) represents the number of ambulances covering vertex i within r_1 units, while the right-hand side is 1 if i is covered within r_1 units, and 2 if it is covered at least twice within r_1 units. The combinations of constraints (31) and (32) ensures that a proportion α of the demand is covered and the coverage standard must be r_1. Constraints (33) state that vertex i cannot be covered at least twice if it is not covered at least once. In constraints (35), p_j can be set equal to 2 since an optimal solution using this value always exists.

5 Probabilistic static models with extra coverage

One of the first probabilistic models for ambulance location is the *Maximum Expected Covering Location Problem* formulation (MEXCLP) due to Daskin (1983). In this model,

it is assumed that each ambulance has the same probability q, called the *busy fraction*, of being unavailable to answer a call, and all ambulances are independent. The busy fraction can be estimated by dividing the total estimated duration of calls for all demand points by the total number of available ambulances. Thus, if vertex $i \in V$ is covered by k ambulances, the corresponding expected covered demand is $E_k = d_i(1 - q^k)$, and the marginal contribution of the k^{th} ambulance to this expected value is $E_k - E_{k-1} = d_i(1 - q)q^{k-1}$. In MEXCLP, up to p ambulances may be located in total, but more than one vehicle may be located at the same vertex. Let y_{ik} be a binary variable equal to 1 if and only if vertex $i \in V$ is covered by at least k ambulances. The model can be written as follows:

$$\text{(MEXCLP)} \qquad \text{Maximize} \qquad \sum_{i \in V} \sum_{k=1}^{p} d_i(1-q)q^{k-1} y_{ik} \tag{38}$$

$$\text{subject to} \qquad \sum_{j \in W_i} x_j \geq \sum_{k=1}^{p} y_{ik} \qquad (i \in V) \tag{39}$$

$$\sum_{j \in W}^{p} x_j \leq p \tag{40}$$

$$x_j \quad \text{integer} \qquad (j \in W) \tag{41}$$

$$y_{ik} \in \{0, 1\} \qquad (i \in V,\ k = 1, \ldots, p). \tag{42}$$

The validity of this model stems from the fact that the objective function is concave in k. Therefore, if $y_{ik} = 1$, then $y_{ih} = 1$ for $h \leq k$. Since the objective is to be maximized, both (39) and (40) will be satisfied as equalities. It follows that the two sides of (39) will be equal to the number of ambulances covering vertex $i \in V$.

An application of MEXCLP to the city of Bangkok is described in Fujiwara, Makjamroen and Gupta (1987). The authors have solved MEXCLP heuristically on an instance with $|V| = 59$, $|W| = 46$ and $10 \leq p \leq 30$. One conclusion of their study is that by reducing the number of ambulances from 21, as in the current situation, to 15, a similar expected covering and average response time could be obtained. An extension of MEXCLP, called TIMEXCLP, was also developed by Repede and Bernardo (1994) and applied to the Louisville, Kentucky, data. In TIMEXCLP, variations in travel speed throughout the day are explicitly considered. The authors have combined this model with a simulation module to provide an assessment of the proposed solutions. The main result was an increase of the proportion of calls covered in ten minutes or less from 84% to 95%. In addition, the response time went down by 36%. Finally, Goldberg *et al.* (1990b) have developed yet another variant of MEXCLP in which stochastic travel times are considered. The objective was to maximize the expected number of calls covered within eight minutes. The authors classify the potential location sites in order of preference. They compute the probability of reaching a demand point within this time standard, based on the following three probabilities: 1) the probability that an ambulance at the k^{th} preferred site for a demand point be able to reach this point within eight minutes; 2) the probability that this ambulance is available; 3) the probability that the ambulances located at the $k - 1$ less preferred site are not available. On data from the city of Tucson, Arizona, they showed

that a better location plan could yield a one percent increase in the expected number of calls covered within eight minutes and that the worst covering ratio of a zone in time could be increased from 24% to 53.1%.

Two other probabilistic models were proposed by ReVelle and Hogan (1989) to maximize the demand covered with a given probability α. These authors formulate the *Maximum Availability Location Problem* (MALP I and MALP II) as a chance constrained stochastic program (Charnes and Cooper, 1959). In MALP I, the busy fraction q is assumed to be the same for all potential location sites. The minimum number of ambulances required to serve each demand point i with a reliability level of α is determined by the constraints

$$1 - q^{\sum_{j \in W_i} x_j} \geq \alpha \qquad (i \in V) \tag{43}$$

which can be linearized as

$$\sum_{j \in W_i} x_j \geq \lceil \log(1-\alpha)/\log q \rceil \qquad (i \in V). \tag{44}$$

To formulate MALP I, define binary variables y_{jk} as in MEXCLP. The model can be written as:

(MALP I) Maximize
$$\sum_{i \in V} d_i y_{ib} \tag{45}$$

subject to
$$\sum_{k=1}^{b} y_{jk} \leq \sum_{j \in W_i} x_j \qquad (i \in V) \tag{46}$$
$$y_{ik} \leq y_{i,k-1} \qquad (i \in V,\ k = 2, \ldots, b) \tag{47}$$
$$\sum_{j \in W} x_j = p \tag{48}$$
$$x_j \in \{0,1\} \qquad (j \in W) \tag{49}$$
$$y_{ik} \in \{0,1\} \qquad (i \in V,\ k = 1, \ldots, p). \tag{50}$$

Here, constraints (47) are required since the concavity property observed in MEXCLP no longer holds.

In MALP II, the assumption that the busy fraction is identical for all sites is relaxed. Instead, ReVelle and Hogan compute an estimate of the busy fraction q_i associated with each $i \in V$, as the ratio of the total duration of all calls associated with i to the total availability of all ambulances in W_i. This value is a lower bound since some ambulances in W_i may be dispatched to calls unrelated to i, but a valid albeit conservative model can be constructed along the lines of MALP I. In MALP II, instead of b, a value b_i is computed for each $i \in V$. ReVelle and Hogan (1989) rightly point out the difficulty of working with a busy fraction q_j specific to each $j \in W$ since these values are an *output* of the model and cannot be known *a priori*. However, given an ambulance location plan and demand levels, probabilities can be estimated using analytical tools such as the hypercube model

(Larson, 1974, 1975; Burwell, McKnew and Jarvis, 1992), an iterative optimization algorithm (Jarvis, 1975; Fitzsimmons and Srikar, 1982), or simulation (Davis, 1981; Goldberg, 1990a).

Several articles are devoted to the estimation of the busy fraction associated with the whole system or with a specific vertex $j \in W$. Thus, Batta, Dolan and Krishnamurty (1989) have developed the *Adjusted MEXCLP Model* (AMEXCLP) in which each term of the objective function (38) is multiplied by a correction factor that accounts for the fact that ambulances do not operate independently, but may be viewed as servers in a queueing system to which the hypercube model (Larson, 1974) can be applied to compute busy fractions. Whereas Batta *et al.* assume the same busy fraction for the entire system, Marianov and ReVelle (1994) propose the *Queueing Probabilistic Location Set Covering Problem* (QPLSCP) in which busy fractions are site specific. These authors compute the minimum number b_i of ambulances necessary to cover a demand point $i \in V$ in such a way that the probability of all of them being simultaneously busy does not exceed a given threshold. This number is then used as an input in MALP II.

Ball and Lin (1993) have developed an extension of LSCM, called Rel-P, that incorporates a linear constraint on the number of vehicles required to achieve a given reliability level. The model contains binary variables x_{jk} equal to 1 if and only if k ambulances are located at vertex $j \in W$, and constants c_{jk} equal to the cost of locating k vehicles at site j. An upper bound p_j is imposed on the number of ambulances located at site j. Their model is as follows:

$$\text{(Rel-P)} \qquad \text{Minimize} \qquad \sum_{j \in J} \sum_{1 \le k \le p_j} c_{jk} x_{jk} \tag{51}$$

$$\text{subject to} \qquad \sum_{1 \le k \le p_j} x_{jk} \le 1 \qquad (j \in W) \tag{52}$$

$$\sum_{j \in W_i} \sum_{1 \le k \le p_j} a_{jk} x_{jk} \ge b_i \quad (i \in V) \tag{53}$$

$$x_{jk} \in \{0,1\} \qquad (j \in W,\ 1 \le k \le p_j). \tag{54}$$

In constraints (53), the constants a_{jk} and b_i are computed to ensure that given the number of ambulances covering demand point i, the probability of being unable to answer a call does not exceed a certain value. The computation of the a_{jk} and b_i coefficients are in fact carried out by using an upper bound on that probability.

Finally, Mandell (1998) describes a two-tiered system in which ALS and BLS units are to be located. The system is *inclusive* in the sense that ALS units can provide BLS services. The probability that a call originating at vertex $i \in V$ is adequately served depends on the number h of ALS units within travel time r^A of i, the number k of ALS units within r^B of i, and the number ℓ of BLS units within r^B of i, where $r^A \ge r^B$. Using a queueing model, Mandell computes the associated probability $\theta_{ihk\ell}$. The problem is to locate p^A ALS units and p^B BLS units in W. Let $x_j^A(x_j^B)$ be the number of ALS (BLS) units located at site $j \in W$. Also define binary variables $y_{ihk\ell}$ equal to 1 if and only if h ALS units are located within r^A of i, k ALS units are located within r^B of i, and ℓ BLS units are located within r^B of i. Letting $W_i^A = \{j \in W : t_{ij} \le r^A\}$ and $W_i^B = \{j \in W : t_{ij} \le r^B\}$,

the range of $y_{ihk\ell}$ goes from 0 to $h_i = \min\{p^A, |W_i^A|\}$ for h, from 0 to $k_i = \min\{h, p^A, |W_i^B|\}$ for k, and from 0 to $\ell_i = \min\{p^B, |W_i^B|\}$ for ℓ. The *Two-Tiered Model* (TTM) proposed by Mandell is to maximize the expected covered demand:

$$\text{(TTM)} \qquad \text{Maximize} \sum_{i\in V}\sum_{h=1}^{h_i}\sum_{k=0}^{k_i}\sum_{\ell=0}^{\ell_i} d_i\,\theta_{ihk\ell}\,y_{ihk\ell} \tag{55}$$

$$\text{subject to} \quad \sum_{h=1}^{h_i} h \sum_{k=0}^{k_i}\sum_{\ell=0}^{\ell_i} y_{ihk\ell} \le \sum_{j\in W_A^i} x_j^A \quad (i\in V) \tag{56}$$

$$\sum_{k=1}^{k_i} k \sum_{h=k}^{h_i}\sum_{\ell=\ell_0}^{\ell_i} y_{ihk\ell} \le \sum_{j\in W_B^i} x_j^B \quad (i\in V) \tag{57}$$

$$\sum_{\ell=1}^{\ell_i} \ell \sum_{h=1}^{h_i}\sum_{k=0}^{k_i} y_{ihk\ell} \le \sum_{j\in W_B^i} x_j^B \quad (i\in V) \tag{58}$$

$$\sum_{h=1}^{h_i}\sum_{k=0}^{k_i}\sum_{\ell=\ell_0}^{\ell_i} y_{ihk\ell} \le 1 \quad (i\in V) \tag{59}$$

$$\sum_{j\in W} x_j^A \le p^A \tag{60}$$

$$\sum_{j\in W} x_j^B \le p^B \tag{61}$$

$$y_{ihk\ell} \in \{0,1\} \quad (i\in V,\ 0\le h\le h_i,\ 0\le k\le k_i,\ 0\le \ell\le \ell_i) \tag{62}$$

$$x_j^A, x_j^B \in \{0,1\} \quad (j\in W), \tag{63}$$

where $\ell_0 = 1$ if $k = 0$ and $\ell_0 = 0$ if $k > 0$. Constraints (56)–(58) ensure that the values taken by the coverage variables $y_{ihk\ell}$ are consistent with the number of located ALS and BLS units. Constraints (59) mean that at most one combination h, k and ℓ units of different types is used for any demand point i.

6 A dynamic model

When siting emergency vehicles, relocation decisions must periodically be made in order not to leave areas unprotected. This was recognized by Kolesar and Walker (1994) who designed a relocation system for fire companies. The ambulance relocation problem is more difficult to tackle since it has to be solved more frequently at very short notice. More powerful solution methodologies are called for in this case. With the development of faster heuristics and advanced computer technologies, it is now possible to quickly solve the ambulance location problem in real-time. What this means is that a new ambulance redeployment strategy can be recomputed at any time t, using the available information.

As far as we are aware, only one such model exists in the area of ambulance relocation. It was developed by Gendreau, Laporte and Semet (2001) and makes use of the Double Standard Model put forward by the same authors in 1997 (see Section 4). In addition to the standard coverage and site capacity constraints, the model takes into account a number of practical considerations inherent to the dynamic nature of the problem: 1) vehicles moved in successive redeployments cannot always be the same; 2) repeated round trips between the same two location sites must be avoided; 3) long trips between the initial and final location sites must be avoided.

The ambulance relocation problem is solved at each instant t at which a call is registered. The dynamic aspect of the redeployment model is captured by time dependent constants $M_{j\ell}^t$ equal to the cost of repositioning, at time t, ambulance ℓ from its current site to site $j \in W$. This includes the case where site j coincides with the current location of the ambulance, i.e., $M_{j\ell}^t = 0$. The constant $M_{j\ell}^t$ captures some of the history of ambulance ℓ. If it has been moved frequently prior to time t, then $M_{j\ell}^t$ will be larger. If moving ambulance ℓ to site j violates any of the above constraints, then the move is simply disallowed. Binary variables $y_{j\ell}$ are equal to 1 if and only if ambulance ℓ is moved to site j. The *Dynamic Double Standard Model at Time t* (DDSM^t) can now be described:

$$\text{(DDSM}^t\text{)} \qquad \text{Maximize} \sum_{i \in V} d_i x_i^2 - \sum_{j \in W} \sum_{\ell=1}^{p} M_{j\ell}^t y_{j\ell} \tag{64}$$

$$\text{subject to} \qquad \sum_{j \in W_i^2} \sum_{\ell=1}^{p} y_{j\ell} \geq 1 \qquad (i \in V) \tag{65}$$

$$\sum_{i \in V} d_i x_i^1 \geq \alpha \sum_{i \in V} d_i \tag{66}$$

$$\sum_{j \in W_i^1} \sum_{\ell=1}^{p} y_{i\ell} \geq x_i^1 + x_i^2 \qquad (i \in V) \tag{67}$$

$$x_i^2 \leq x_i^1 \qquad (i \in V) \tag{68}$$

$$\sum_{j \in W} y_{j\ell} = 1 \qquad (\ell = 1, \ldots, p) \tag{69}$$

$$\sum_{\ell=1}^{p} y_{\ell j} \leq p_j \qquad (j \in W) \tag{70}$$

$$x_i^1, x_i^2 \in \{0, 1\} \qquad (i \in V) \tag{71}$$

$$y_{j\ell} \in \{0, 1\} \qquad (j \in W,\ \ell = 1, \ldots, p). \tag{72}$$

Apart from the variables $y_{j\ell}$, all variables, parameters and constraints of this model can be interpreted as in the static case. The objective function is the demand covered twice within r_1 time units, minus the sum of penalties associated with vehicle moves at time t.

To solve DDSM^t, Gendreau, Laporte and Semet have developed a fast tabu search heuristic implemented on parallel processors. This algorithm runs non-stop and continuously computes the best possible redeployment plans associated with the current positions of ambulances, in response to each potential anticipated ambulance request. In other words, the algorithm computes the best relocation plan for all h, where h represents the next ambulance assigned to answer a call. In effect, a table is built in which each line h contains a solution to DDSM^t. When h becomes known, the associated redeployment strategy can readily be applied. If the elapsed time between two successive requests is sufficiently long, all possible redeployments plan can be computed in time for the next request. Otherwise, a suitable redeployment solution may not be available when needed and no redeployment then takes place. Several secondary features relative to the relative urgency of calls, rerouting of ambulances on their way to a call, etc. are incorporated into the system.

The system was run on a network on eight parallel Sun Ultra–1/140 workstations and tested on six problems generated using real data sets from the Island of Montreal. In these test problems, between 120 and 140 calls were randomly generated from 5 a.m. to 12 noon, according to a Poisson distribution. There are 2521 aggregated demand points on the Island of Montreal, and between 40 to 51 available ambulances, depending on the time of day. The covering standards used were $r_1 = 7$ minutes and $r_2 = 15$ minutes. Speeds varying between 35 and 50 km/h were applied. Results obtained on the six test problems show that the parallel tabu search algorithm was capable of precomputing a redeployment strategy in 95% of all cases. It failed when two calls arrived within less than 32 seconds of each other. Out of all calls, 38% required at least one ambulance relocation and 99.5% of all relocations involved at most five ambulances, with an average of 2.08. Comparisons with exact solutions obtained by CPLEX on a sample of scenarios show a maximal departure of 2% from optimality.

7 Summary and conclusions

There has been an important evolution in the development of ambulance location and relocation models over the past thirty years. The first models were very basic and did not take into account the fact that some coverage is lost when an ambulance is dispatched to a call. Nevertheless, these early models served as a sound basis for the development of all subsequent models. The question of ambulance non-availability was addressed in two main ways. Deterministic models yield solutions in which demand points are overcovered, but the actual availability of ambulances is not considered. Probabilistic models work with the busy fraction of vehicles, which can be estimated in a number of ways, including sophisticated queueing calculations. Dynamic models have just started to emerge. They can be used to periodically update ambulance positions throughout the day. Tests have shown that such models can work in practice provided that fast heuristics exist and sufficient computing power are available. We summarize in Tables 1 and 2 the main available deterministic and probabilistic models for ambulance location and relocation.

As noted by Marianov and ReVelle (1995) the development of ambulance location and relocation models will likely parallel the growth of information technologies. We anticipate that much of the evolution will take place in the field of dynamic models which

Table 1: Summary of deterministic static and dynamic models

Reference	**Model**	**Objective**	**Coverage constraints**	**Constraints on location sites**	**Ambulances**
ReVelle, Swain and Bergman (1971)	Location Set Covering Model (LSCM).	Minimize the number of ambulances.	Cover each demand point at least once.	At most one ambulance per site.	One type. Number unlimited.
Church and ReVelle (1974)	Maximal Covering Location Problem (MCLP).	Maximize the demand covered.	None.	At most one ambulance per site.	One type. Number given.
Schilling *et al.* (1979)	Tandem Equipment Allocation Model (TEAM).	Maximize the demand covered.	None.	At most one ambulance of each type per site. Type A can only be located if B is located.	Two types. Numbers given.
Schilling *et al.* (1979)	Facility-Location, Equipment-Emplacement Technique (FLEET).	Maximize the demand covered.	None.	At most one ambulance per site. Only p sites can be used.	Two types. Numbers given.
Daskin and Stern (1981)	Modified MCLP.	Maximize the demand covered, then the number of demand points covered more than once.	Cover each demand point at least once.	At most one ambulance per site.	One type. Number given.
Hogan and ReVelle (1986)	Modified MCLP. (BACOP1 and BACOP2).	Maximize the demand covered twice, or a combination of the demand covered once or twice.	Cover each demand point at least once.	At most one ambulance per site.	One type. Number given.
Gendreau, Laporte and Semet (1997)	Double Standard Model (DSM).	Maximize the demand covered at least twice within r_1.	All demand covered within r_2. Proportion α of all demand covered within r_1.	Upper bound on the number of ambulances per site.	One type. Number given.
Gendreau, Laporte and Semet (2001)	Dynamic Double Standard Model at Time t (DDSM^t)	Dynamically maximize the demand covered at least twice within r_1, minus a redeployment penalty term.	All demand covered within r_2. Proportion α of all demand covered within r_1.	Upper bound on the number of ambulances per site.	One type. Number given.

Table 2: Summary of probabilistic models

Reference	**Model**	**Objective**	**Coverage constraints**	**Constraints on location sites**	**Ambulances**	**Busy period**
Daskin (1983)	Maximum Expected Covering Location Problem (MEXCLP).	Maximize the expected demand covered.	None.	None.	One type. Upper bound given (always reached).	Same for each ambulance. Given value.
ReVelle and Hogan (1989)	Maximum Availability Location Problem (MALP I).	Maximize the total demand covered with a probability α.	None.	None.	One type. Number given.	Same for all potential location site.
ReVelle and Hogan (1989)	Maximum Availability Location Problem (MALP II).	Maximize the total demand covered with a probability at least α.	None.	None.	One type. Number given.	Varies according to each demand point.
Batta, Dolan and Krishnamurty (1989)	Adjusted MEXCLP (AMEXCLP).	Maximize the expected demand covered.	None.	None.	One type. Number given.	Varies according to each demand point. Ambulances not independent.
Goldberg *et al.* (1990b)	Adjusted MEXCLP.	Maximize the expected demand covered within eight minutes.	None.	At most one ambulance per site.	One type. Number given. Two types of calls.	Same for each ambulance. Given value.
Ball and Lin (1993)	Modified LSCM (Rel-P).	Minimize the sum of ambulance fixed costs.	Proportion α of all demand covered within r_1.	At most p_j ambulances at site j.	One type. Number unlimited.	Upper bound computed on busy period.
Repede and Bernardo (1994)	Time dependent MEXCLP (TIMEXCLP).	Maximize the expected demand covered.	None.	None.	One type. Number given. Varying speeds.	Same for each ambulance. Given value.
Marianov and ReVelle (1994)	Queueing Probabilistic Location Set Covering Problem (QPLSCP).	Maximize the total demand covered with a probability at least α.	None.	None.	One type. Lower bound computed for each demand point.	Varies according to demand points.
Mandell (1998)	Two-Tiered Model (TTM).	Maximize the expected total demand.	None.	Bounds on each type per site.	Two types. Inclusive system. Numbers given.	Computed using a queueing model.

are not only dependent on sophisticated system technologies, but also on the availability of fast and accurate search heuristics. In addition, we expect that advanced algorithms in the area of stochastic programming with recourse (see, e.g., Birge and Louveaux, 1997) and new techniques for dynamic shortest path computations (see, e.g., Pallottino and Scutella, 1998) will soon become standard ingredients of dynamic models. In the first case, the idea is to incorporate into the models the expected cost incurred when no suitable ambulance is available to answer a call, as opposed to just imposing a probabilistic constraint in the model. In the second case, variations in travel times during the day should be reflected in shortest path computations used in repeated applications of the dynamic model.

Acknowledgments

This research was partially funded by the Canada Research Chair in Distribution Management, by the Fonds de recherche en santé du Québec (FRSQ) under grant 980861, and by the Canadian Natural Sciences and Engineering Research Council (NSERC) under grants OGP0039682 and OGP0184123. This support is gratefully acknowledged. Thanks are due to two referees for their valuable comments.

References

Adams, T.F., 1997. *Police Operations*, Prentice-Hall, Englewood Cliffs, NJ.

American TriTech, 1996. *VisiCAD for EMS*. San Diego, CA.

Ball, M.O., Lin, L.F., 1993. A Reliability Model Applied to Emergency Service Vehicle Location. *Operations Research* **41**, 18–36.

Batta, R., Dolan, J.M., Krishnamurty, N.N., 1989. The Maximal Expected Covering Location Problem: Revisited. *Transportation Science* **23**, 277–287.

Bernhardsen, T., 1999. *Geographic Information Systems: An Introduction*. Wiley, Chichester.

Birge, J.R., Louveaux, F., 1997. *Introduction to Stochastic Programming*, Springer, New York.

Bouveyron, C., Didier, M., 1993. *Guide Économique et Méthodologique des SIG*. Hermès, Paris.

Burwell, T.H., McKnew, M.A., Jarvis, J.P., 1992. An Application of a Spatially Distributed Queueing Model to an Ambulance System. *Socio-Economic Planning Sciences* **26**, 289–300.

Charnes, A., Cooper, W.W., 1959. Chance Constrained Programming. *Management Science* **5**, 73–79.

Church, R.L., ReVelle, C.S., 1974. The Maximal Covering Location Problem. *Papers of the Regional Science Association* **32**, 101–118.

Clawson, J.J., 1981. Dispatch Priority Training: Strengthening the Weak Link. *Journal of Emergency Medical Services* **6**, 32–36.

Cohen, S., 1999. In-Response: détection, alerte rapide et traitements des incidents routiers. *Second Carrefour Predit*, 418–422.

Crainic, T.G., Toulouse, M., 1998. *Parallel Metaheuristics.* In: Crainic, T.G., Laporte, G. (Eds), *Fleet Management and Logistics*, Kluwer, Boston, pp. 205–251.

Daskin, M.S., 1983. A Maximum Expected Location Model: Formulation, Properties and Heuristic Solution. *Transportation Science* **7**, 48–70.

Daskin, M.S., Stern, E.H., 1981. A Hierarchical Objective Set Covering Model for Emergency Medical Service Vehicle Deployment. *Transportation Science* **15**, 137–152.

Darvis, S., 1981. Analysis of the Deployment of Emergency Medical Services. *Omega* **9**, 655–657.

Eaton, D.J., Daskin, M.S., Simmons, D., Bulloch, B., Jansma, G., 1985. Determining Emergency Medical Deployment in Austin, Texas. *Interfaces* **15**(1), 96–108.

Erkut, E., Bozkaya, B., 1999. Analysis of Aggregation Errors for the p-Median Problem. *Computers & Operations Research* **26**, 1075–1096.

Fitzsimmons, J.A., Srikar, B., 1982. Emergency Ambulance Location Using the Contiguous Zone Search Routine. *Journal of Operations Management* **2**, 225–237.

Francis, R.L., Lowe, T.J., Tamir, A., 2000. Aggregation Error Bounds for a Class of Location Models. *Operations Research* **48**, 294–307.

Fujiwara, O., Makjamroen, T., Gupta, K.K., 1987. Ambulance Deployment Analysis: A Case Study of Bangkok. *European Journal of Operational Research* **31**, 9–18.

Gendreau, M., Laporte, G., Semet, F., 1997. Solving an Ambulance Location Model by Tabu Search. *Location Science* **5**, 75–88.

Gendreau, M., Laporte, G., Semet, F., 2001. A Dynamic Model and Parallel Tabu Search Heuristic for Real-Time Ambulance Relocation. *Parallel Computing*, **27**, 1641–1653.

Glover, F., Laguna, M., 1997. *Tabu Search.* Kluwer, Boston.

Goldberg, J., Dietrich, R., Chen, J.M., Mitwasi, M.G., 1990a. A Simulation Model for Evaluating a Set of Emergency Vehicle Base Locations: Development, Validation and Usage. *Socio-Economic Planning Sciences* **24**, 124–141.

Goldberg, J., Dietrich, R., Chen, J.M., Mitwasi, M.G., 1990b. Validating and Applying a Model for Locating Emergency Medical Services in Tucson, AZ. *European Journal of Operational Research* **49**, 308–324.

Haas, N.P., Hoffman, R.F., Mauch, C., Von Fournier, C., Sudkamp, N.P., 1995. The Management of Polytraumatized Patients in Germany. *Clinical Orthopaedics & Related Research* **318**, 23–35.

Heddebaut, M., 1997. Télébornes d'appel d'urgence et alerte des passagers. *Trans Flash*, 221.

Hogan, K., ReVelle, C.S., 1986. Concepts and Applications of Backup Coverage. *Management Science* **34**, 1434–1444.

Jarvis, J.P., 1975. Optimization in Stochastic Service with Distinguishable Servers. TR–19–75, Innovative Resource Planning Project in Urban Public Safety Systems. Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA.

Kolesar, P., Walker, W.E., 1974. An Algorithm for the Dynamic Relocation of Fire Companies. *Operations Research*, **22**, 249–274.

Larson, R.C., 1972. *Urban Police Patrol Analysis*, The MIT Press, Cambridge, MA.

Larson, R.C., 1974. A Hypercube Queueing Model for Facility Location and Redistricting in Urban Emergency Services. *Computers & Operations Research* **1**, 67–75.

Larson, R.C., 1975. Approximating the Performance of Urban Emergency Service Systems. *Operations Research* **23**, 845–868.

Mandell, M.B., 1998. Covering Models for Two-Tiered Emergency Medical Services Systems. *Location Science* **6**, 355–368.

Marianov, V., ReVelle, C.S., 1992. The Capacitated Standard Response Fire Protection Siting Problem: Deterministic and Probabilistic Models. *Annals of Operations Research* **40**, 303–322.

Marianov, V., ReVelle, C.S., 1994. The Queueing Probabilistic Location Set Covering Problem and Some Extensions. *Socio-Economic Planning Sciences* **28**, 167–178.

Marianov, V., ReVelle, C.S., 1995. Siting Emergency Services. In: Z. Drezner, (Ed.), *Facility Location*, Springer, New York, pp. 199–223.

Pallottino, S., Scutella, M.G., 1998. Shortest Path Algorithms in Transportation Models: Classical and Innovative Aspects. In: P. Marcotte and S. Nguyen (Eds), *Equilibrium and Advanced Transportation Modelling*, Kluwer, Boston, pp. 245–281.

Repede, J.F., Bernardo, J.J., 1994. Developing and Validating a Decision Support System for Locating Emergency Medical Vehicles in Louisville, Kentucky. *European Journal of Operational Research* **75**, 567–581.

ReVelle, C.S., 1989. Review, Extension and Prediction in Emergency Services Siting Models. *European Journal of Operational Research* **40**, 58–69.

ReVelle, C.S., Hogan, K., 1989. The Maximum Availability Location Problem. *Transportation Science* **23**, 192–200.

Schilling, D.A., Elzinga, D.J., Cohon, J., Church, R.L., ReVelle, C.S., 1979. The TEAM/FLEET Models for Simultaneous Facility and Equipment Siting. *Transportation Science* **13**, 163–175.

Stout, J., Pepe, P.E., Mosesso, V.N., 2000. All-Advanced Life Support vs Tired-Response Ambulance Systems. *Prehospital Emergency Care* **4**, 1–6.

Swersey A.J., 1994. The Deployment of Police, Fire, and Emergency Medical Units. In: S.M. Pollock, M.H. Rothkopf and A. Barnett (Eds), *Operations Research and the Public Sector*, Handbooks in Operations Research and Management Science, Vol. 6, North-Holland, Amsterdam, pp. 151–200.

Toregas, C.R., Swain, R., ReVelle, C.S., Bergman, L., 1971. The Location of Emergency Service Facilities. *Operations Research* **19**, 1363–1373.

Walker, W.E., Chaiken, J.M., Ignall, E.J., (Eds), 1979. *Fire Department Deployment Analysis*, North-Holland, New York.

Wisborg, T., Guttmorsen, A.B., Sorensen, M., Flaatten, H.K., 1994. The Potential of Anaesthesiologist-Manned Ambulance Service in Rural/Urban District. *Acta Anaesthesiologica Scandinavica* **38**, 657–661.