OPAL – Optimized Ambulance Logistics

Tobias Andersson* Sverker Petersson† Peter Värband*

*Linköping University
ITN/Campus Norrköping
SE-601 74 Norrköping, Sweden
{toban,petva}@itn.liu.se

†SOS Alarm AB
Box 5776
SE-114 87 Stockholm, Sweden
sverker.petersson@sosalarm.se

1 Introduction

SOS Alarm is the company in Sweden responsible for receiving all calls to the national emergency number, 112, and also for controlling all ambulance movements. The operations are run from a SOS central of which there is one in each county (administrative district) in Sweden. One of the services offered by SOS Alarm is called *ambulance logistics*, and the customers are the county councils.

An ambulance can be requested by the public or by the health care, and the request is received by a SOS operator who prioritizes it according to three degrees:

- Prio 1: Urgent, life threatening symptoms.
- Prio 2: Urgent, not life threatening symptoms.
- Prio 3: Non-urgent calls.

A Prio 3 call is often a transport to or from the hospital and the patient's home, or between hospitals. These requests are often made by the health care but can also be made by the public at a special ambulance order number.

Orders received more than a day before they are to be executed can be used to create a transportation schedule. Today, the planning mostly consists of checking that it is possible to perform the pre-ordered transports and still have enough ambulances left to cover for incoming urgent calls, i.e. without compromising the *preparedness* in the county.

Calls that are received the same day as they are to be executed may be urgent or non-urgent, but in both cases they are prioritized by a SOS operator. An ambulance controller (that may or may not be the same person as the operator) then has to decide which ambulance to allocate to the incoming call. This decision depends on a number of different parameters; e.g. the location of the ambulances, their equipment and also the preparedness in the area. Preparedness refers to the ability to get one or more ambulances to a potential call site, i.e. where the patient is located, within a reasonable time. At a certain location, the preparedness depends on how many ambulances that can reach the location within a certain time frame, as well as their travel time. Of course, the preparedness also depends on the expected need for ambulances in the area; e.g. in a densely populated area it may be necessary to have more than one ambulance close by in order to get an adequate preparedness.

For a Prio 1 call, the closest ambulance (or ambulances if more than one is needed) is always assigned. This assignment may however cause a drop in the preparedness, probably for the zones close to the ambulance that was assigned to the new call. This might make it necessary for the controller to relocate another ambulance to the affected areas, in order to cover for the ambulance that is busy. For Prio 2 and 3 calls, the controller may choose to assign another ambulance than the closest one in order to preserve the preparedness, if this can be done without compromising the safety for the patient.

Keeping an adequate preparedness is one of the most complex tasks for the ambulance controller. It requires knowledge of where call sites are likely to appear and of how fast the ambulances can travel through different parts of the county, as well as knowledge of where the ambulances currently are located and if they are available. Today, most ambulances in Sweden have GPS receivers and transmit their position and status to the SOS central. However, to know where ambulances might be needed in the future, and how fast they can get there requires experience.

2 Preparedness

In ambulance logistics, preparedness has been used as a qualitative measure for a long time, but two people does not always mean the same thing when using the word. Also, two ambulance controllers may have different opinions on what can be considered good or bad preparedness, depending on their experience, their risk aversion and their personality. For example, one controller may think that less than twenty available ambulances in the county means that the preparedness is low, while another controller might think that twenty ambulances is more than enough for an adequate preparedness. Thus, there is a need for a clear definition of what preparedness actually is and a quantitative measure than can be used to calculate it.

Definition 1 (Preparedness): *In ambulance logistics, preparedness refers to the ability of being able to, within a reasonable time, offer qualified ambulance health care to the inhabitants in a specific geographical area.*

Definition 1 states that the population in a county in Sweden should be able to expect to get an ambulance within a *reasonable time* when calling 112. This reasonable time depends on how urgent the call is, and what standards and goals the county council has set. Today some counties use waiting time goals to measure the quality of the ambulance health care. E.g. the county of Stockholm, which is the largest county in Sweden in terms of number of calls and ambulances, has different waiting time goals, depending on the priority of the call. For Prio 1 calls, the goals say that 75% of all calls should be served within 10 minutes, 95% within 15 minutes and 99% within 20 minutes. For Prio 2 and 3 calls, the response times are allowed to be longer. It is SOS Alarm's responsibility to ensure that there are enough available ambulances, positioned in such a way that these goals will be fulfilled.

The waiting time goals are useful to evaluate the ambulance logistics service, but as they are reactive measures, a certain time has to pass before it is possible to draw any conclusions. They are therefore not directly useful for an ambulance controller. A measure of the preparedness should have the quality that it can help an ambulance controller in the operative work. It should be able to distinguish between different parts of the county; i.e. one part might suffer from a low preparedness while another part has more ambulances than probably will be needed. For this to be possible it is necessary to divide the area under consideration into different zones, all with a specific preparedness. How many zones a county should be divided into is a choice between complexity (more zones increase the complexity) and quality (more zones will give a better model of the real situation). The final choice is however often directly dependent on the available data. It is for example necessary to have travel times for the ambulances in the county in order to predict how long time it will take for an ambulance to get from its initial position to a call site; preferably travel times that changes dynamically with time. Also, some way of measuring the need for ambulances in the different zones is necessary.

3 Calculating the preparedness

Suppose that the geographical area is divided into a set of zones, N. A weight c_j is assigned to each zone *j*. This weight mirrors the probability that an ambulance will be needed in the zone and can for example be calculated as

$$c_j = k * \frac{\text{number of calls in zone } j}{\text{total number of calls}}$$

where *k* is a suitable scaling factor and calls are collected for a sufficiently long time period. It is also possible to base the weight on the population, advance knowledge of special events and other information that may affect the need for ambulances in the zone. The weights may also be time dependent, e.g. c_{jt} = the weight for zone *j* in time period *t*, as the need for ambulances often varies with time.

It can be reasoned that the preparedness in a zone mainly depends on three things:

- The number of ambulances that can reach the zone (within a certain time).
- The time it takes for the ambulances to reach the zone (i.e. the travel time).
- The expected need for ambulances in the zone (i.e. c_j).

It is possible to construct a measure in a number of different ways, while still taking all these aspects into consideration. Depending on the construction, the different measures will of course have different qualities. This makes it important to carefully consider what the measure can and will be used for. The measure then has to be tested to see if it possesses the desired qualities.

The preparedness in zone *j*, p_j =

$$\frac{1}{c_j}\sum_{i\in N_j}\sum_{k\in A}\frac{x_{ik}}{t_{ij}} \qquad (1)$$

A = the set of ambulances.
N_j = the set of zones from where an ambulance can reach zone *j* within *T* time units.
x_{ik} = 1 if ambulance *k* is located in zone *i*.
t_{ij} = the travel time between zone *i* and *j*.
c_j = the weight for zone *j*.

$$\rho y_j + \frac{1}{c_j}\sum_{i\in N_j}\sum_{k\in A}\frac{x_{ik}}{t_{ij}} \qquad (2)$$

A, N_j, x_{ik}, t_{ij} and c_j as for (1).
$y_j = 1$ if there is an ambulance that can reach zone *j* in T^1 minutes.
ρ = a contribution factor.

$$\frac{1}{c_j}\sum_{l\in L_j}\frac{\gamma^l}{t_j^l} \qquad (3)$$

c_j as for (1).
L_j = the ordered set of ambulances that contribute to the preparedness in zone *j*.
γ^l = the contribution factor for ambulance *l* (*l* = 1 is the closest, 2 the second closest etc.).
t_j^l = the travel time to zone *j* for ambulance *l*.

Above, three different ways of calculating the preparedness in a zone is proposed. The three measures differ in complexity, i.e. how computationally expensive they are to calculate in real-time. All three depends on where the ambulances in the set A are located, their status, and how these factors change. In (1), it is easy to update the preparedness when the status or location of an ambulance changes, since all ambulances are independent of each other. However, this have some negative effects on the qualities of the measure. For example, two ambulances 40 minutes away from zone j, will give the same contribution to p_j as one ambulance that is 20 minutes away. If c_j is not large, it is easy to see that one close ambulance is better that two that are far away. This problem is somewhat corrected in (2), where a term has been added to the measure, giving an extra contribution, ρ, if there is an ambulance close to the zone. An ambulance is defined as close to the zone if it is within T^1 minutes of travel time, a time that can be set e.g. in accordance with the waiting time goals. A problem with (2) is that the influence of this first term will decrease as the number of ambulances in A increases, making the contribution from the second term, the back-up coverage, stronger. None of the disadvantages of (1) and (2) mentioned above can be found in (3), a measure with the major disadvantage that the contribution from an ambulance to the preparedness in a zone is dependent on where the other ambulances are located. This makes (3) more computationally expensive than (2) and (1).

4 Decision support for ambulance logistics

Deciding where ambulance stations should be located, and how many ambulances that are needed, is often referred to as *ambulance location*, and is a fairly well studied problem. A recent review of ambulance location and relocation models can be found in Brotcorne et al (2003). Another recent review of the area is Goldberg (2004), which also extends to related areas beyond ambulance location and relocation, including data gathering and simulation. One of the main problems with ambulance location, in contrast to many other location problems, is that a static location of ambulance stations does not at all guarantee that a patient will be able to get an ambulance quickly. Since it is not possible to predict exactly where and when calls will appear, the ambulances that are located near a new call might be busy. Different ways have been used to deal with this fact, among them to use probabilistic models with a busy period, which basically is a measure of the probability that an ambulance will be busy when a new call arrives. Another way of handling the fact that ambulances may be busy is to try to cover the demand more than once, which often is used in deterministic models. This is similar to using the measure of preparedness when locating the ambulances, which is done here.
Two basic ambulance location models are formulated:

MAPAL:
$$\max \min_{j \in N} \quad p_j(\boldsymbol{x})$$
$$\text{s.t.} \quad \boldsymbol{x} \in X$$

MINAL:
$$\min \quad |\boldsymbol{x}|$$
$$\text{s.t.} \quad p_j(\boldsymbol{x}) \geq P \quad \forall j \in N$$
$$\boldsymbol{x} \in X$$

The variables $\boldsymbol{x}$ represent the locations of a given set of ambulances, which the preparedness depends on, making $p_j(\boldsymbol{x})$ a function of $\boldsymbol{x}$. Depending on how p_j is calculated, the number of variables and constraints in X will vary.

MAPAL maximizes the lowest preparedness in any of the zones in N given a certain fleet of ambulances. MINAL on the other hand minimizes the number of ambulances needed, for each zone to have a preparedness of at least P. Since the measure of preparedness is used, both of these models in some sense take into account the dynamics of ambulance operations, i.e. they try to locate more ambulances where the need for ambulances is large, while they at the same time ensure a good spread of the ambulances throughout the county. Both models can easily be extended to take into account e.g. constraints ensuring a maximum number of ambulance stations, a maximum number of ambulances at each station, coverage requirements and other aspects.

No matter how well it is possible to solve an ambulance location model, it is still very hard to evaluate the solution. Basically, two options are available; try the solution in a real life situation or simulate a real life situation. Either way, operational rules for how to control the ambulances are needed. For example, if ambulances always return to their station after a call, one solution might be good. Another solution might however have been better if ambulances are used more dynamically and are allowed to switch stations or stay for extended times at hospitals.

In order to create a simulated ambulance control scenario, it is also necessary to simulate how the ambulance controller works operationally, i.e. to create rules for how to allocate ambulances to calls among other things. If these rules are based on optimization theory, they ought to be good enough to provide decision support in the tactical and operational planning. In Henderson et al (1999), various strategic ambulance issues are tackled with a simulation approach, which was later developed to a commercial software package called Siren.

A non-urgent ambulance transportation that is requested at least the day before it will take place, can be referred to as a *planned transport*. Planned transports are usually requested as transportation from an origin at a certain time to a destination. This means that the goal for the ambulance controller is not to get the ambulance to the origin as fast as possible, but to get it there as close to the requested time as possible. The ambulance controller checks if it is possible to make the transportation with preserved preparedness, and if it is possible to find other planned transports that it can be combined with in order to make a more efficient routing.

With a sufficient number of planned transports, a pick up and delivery problem, with multiple depots and preparedness constraints (PDP-P), can be solved to find efficient routings for the ambulances. These routings, or transportation schedules, might of course be disrupted by the operational decisions, in which case some sort of schedule recovery scheme is needed. The PDP-P has similarities to routing problems where a robust schedule is sought for, but to the authors' knowledge, there has been no attempt to solve routing problems with preparedness constraints.

Decision support for operational ambulance control may include suggestions on ambulance allocations and on dynamic relocations. The ambulance allocation part includes deciding which ambulance to allocate to a new call. For Prio 1 calls this is fairly easy if it is possible to tell which ambulance that is closest to the call site. However, this decision might lead to a reassignment for an ambulance, e.g. if it is on its way to a Prio 3 call but is the closest ambulance for a new Prio 1 call. Thus, the allocation of ambulances to old calls, which have had their ambulances reassigned, is also included in problem area of ambulance allocation. If the old call was part of an ambulance schedule from the tactical planning, this should also be considered. For Prio 2 and 3 calls, a decision support tool should ideally regard the consequences of the allocation before making the decision.

Dynamic relocation is made to preserve a high preparedness. It may be necessary after a number of unfortunate Prio 1 calls, where the preparedness could not be taken into account when the allocations were made. It may however also be advantageous to relocate ambulances depending on how the need for ambulances varies during the day and during the season. There has been little previous work on operations research based decision support tools for operational ambulance control. However, in Gendreau (2001) a tabu search heuristic for the ambulance relocation problem is presented.

5 Conclusions and future work

Here we have presented a number of problems within the area of ambulance logistics where optimization based decision support tools might make a contribution. We have also suggested a few ways of calculating the preparedness in an area, or a zone, something that is useful when dealing with the location, planning and control of emergency vehicles.

This project is still in its initial phase, so only models for ambulance location is presented. Models and solution schemes for the other parts of the planning process are to be developed next. There are many interesting problems within the area of ambulance logistics that are relatively unexplored by the operations research society. This project, OPAL, tries to challenge some of them, and it is the authors' hope that we also might inspire others to do the same.

References

L. Brotcorne, G. Laporte, F. Semet, "Ambulance location and relocation models", *European Journal of Operational Research* 147, 451-463 (2003).

M. Gendreau, G. Laporte, F. Semet, "A dynamic model and parallel tabu search heuristic for real-time ambulance relocation", *Parallel Computing* 27, 1641-1653 (2001).

J. Goldberg, "Operations research models for the deployment of emergency service vehicles", *EMS Management Journal* 1, 20-39, 2004.

S. Henderson, A. Mason, "Estimating ambulance requirements in Auckland, New Zealand", Proceedings, Winter Simulation Conference 99, Phoenix, Arizona, December 1999.