ARTICLE IN PRESS
JID: EOR [m5G;January 11, 2016;21:41]

European Journal of Operational Research 000 (2016) 1–13

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Innovative Applications O.R.

The effect of ambulance relocations on the performance of ambulance service providers

T. C. van Barneveld[a,b,*], S. Bhulai[b,a], R. D. van der Mei[a,b]

[a] Centrum Wiskunde en Informatica, Science Park 123, 1098XG Amsterdam, The Netherlands
[b] Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081HV Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 20 February 2015
Accepted 14 December 2015
Available online xxx

Keywords:
OR in health services
Multiple criteria analysis
Dynamic smbulance management
Dynamic relocation
Response times

ABSTRACT

Dynamic Ambulance Management (DAM) is generally believed to provide means to enhance the response-time performance of emergency medical service providers. The implementation of DAM algorithms leads to additional movements of ambulance vehicles compared to the reactive paradigm, where ambulances depart from the base station when an incident is reported. In practice, proactive relocations are only acceptable when the number of additional movements is limited. Motivated by this trade-off, we study the effect of the number of relocations on the response-time performance. We formulate the relocations from one configuration to a target configuration by the Linear Bottleneck Assignment Problem, so as to provide the quickest way to transition to the target configuration. Moreover, the performance is measured by a general penalty function, assigning to each possible response time a certain penalty. We extensively validate the effectiveness of relocations for a wide variety of realistic scenarios, including a day and night scenario in a critically and realistically loaded system. The results consistently show that already a small number of relocations lead to near-optimal performance, which is important for the implementation of DAM algorithms in practice.

1. Introduction

In emergency situations, the location of ambulances has a huge impact on the response time to an incident, i.e., the total time between an incoming emergency call and the moment that an ambulance arrives at the emergency scene. The evaluation of ambulance services providers, judged by the authorities, heavily relies on their performance regarding these response times. For instance, in The Netherlands, the response time of an ambulance may not exceed 15 minutes in 95 percent of the highest priority emergency cases. To realize short response times, it is crucial to plan ambulance services well. This encompasses a variety of planning problems at the strategic, tactical, and operational level. At the strategic level, the locations of the ambulance base stations are determined. Then, at the tactical level, the number of ambulances and thus crews per base station is specified. At the operational level, real-time dispatching of ambulances to incidents and real-time relocation of ambulances is considered.

In this paper, we focus on the last part of the operational level: the relocation of ambulances. Ambulance vehicles are relocated in real-time, using dynamic and proactive relocation strategies, in order to achieve shorter response times to incidents. These relocation decisions are typically made when an event happens, e.g., when an ambulance is dispatched or when an ambulance is newly free after the service of a patient. However, whether relocations are allowed, and if so, to which locations, depend on regulatory rules. For instance, in Vienna, Austria, moving around ambulances unoccupied by a patient is not allowed, cf. Schmid (2012) as opposed to Edmonton, Canada, cf. Alanis, Ingolfsson, and Kolfal (2013). Moreover, the number of locations at which an ambulance is allowed to park up differs per country. This number can exceed the number of ambulances, as in Montreal, Canada, cf. Gendreau, Laporte, and Semet (2006). Many of these waiting sites are just street corners or different hot spots. In contrast, in The Netherlands, ambulances always must return to a base station, cf. Jagtenberg, Bhulai, and van der Mei (2015). This is a building with several facilities where the ambulance crew can spend its shift when idle. Another difference between countries is the average hospital transfer time. In North America this time can be very large, cf. Carter et al. (2015), as opposed to The Netherlands where the transfer time is usually short.

We consider the Dutch setting in this paper: short transfer times and the dispatcher is allowed to relocate ambulances unoccupied from base station to another one, but the number of

* Corresponding author at: Centrum Wiskunde en Informatica, Science Park 123, 1098XG Amsterdam, The Netherlands. Tel.: +31 205934010.
E-mail address: t.c.van.barneveld@cwi.nl (T.C. van Barneveld).

http://dx.doi.org/10.1016/j.ejor.2015.12.022
0377-2217/

Please cite this article as: T.C. van Barneveld et al., The effect of ambulance relocations on the performance of ambulance service providers, European Journal of Operational Research (2016), http://dx.doi.org/10.1016/j.ejor.2015.12.022

locations on which an ambulance can idle, i.e., base stations, is rather small.

Ambulance relocations are not popular among ambulance crews, especially when the crew is idle at a base station and it is relocated to a different one. Instead, they prefer to spend their shift at a base station and not on the road. To keep the personnel motivated, the number of relocations they have to perform is not allowed to increase excessively. If the ambulance crews spend too much time on the road, the ambulance service provider probably will be condemned by an Occupational Safety and Health organization. Furthermore, costs for the ambulance service provider are associated with each relocation. Some ambulance service providers namely have the policy, especially at night, that the salary of the ambulance crew partly depends on their busy time in which their relocation time is included. Therefore, decision makers must make a consideration between the number of ambulance relocations and the effect of these relocations on the performance of the ambulance service provider.

As an alternative, one could also consider the relocation times. Especially if the above-mentioned payment structure is used, it can be cheaper to minimize these relocation times. However, in this paper, we treat the crew's perspective as our major critical success factor, instead of the financial aspect related to relocations. The number of relocations is a good measure for the crew's perspective, since crews in general prefer to perform one long relocation rather than several short ones. Of course, there is also a trade-off between number of relocations and relocation times. We will pay attention to this trade-off as well.

The relationship between performance and the number of ambulance relocations is complex. The consequences of moving an ambulance to a different base station are not known a priori, due to uncertainty that plays an important role in the process. It is usually not the case that 'more' is 'better', i.e., the more relocations are made, the better the performance of the ambulance service provider. But even if this was the case, there is still a trade-off: would one carry out extra ambulance relocations for only a small gain in performance? Opinions of different ambulance providers differ on this question and it is hard to set a standard concerning the execution of relocations. Therefore, useful insights about the relationship between performance and the number of ambulance relocations are desirable.

1.1. Related work

As stated before, the planning of ambulance services falls apart in three levels. Comprehensive studies of ambulance location and relocation models are done by Brotcorne, Laporte, and Semet (2003) and Li, Zhao, Zhu, and Wyatt (2011). In these papers several deterministic, probabilistic, and dynamic models and their solution procedures are reviewed. Another study on ambulance facility location problems is performed by Owen and Daskin (1998). The operational level falls apart in dispatching and relocation of ambulances. A dispatching algorithm based on the preparedness concept explained by Andersson and Värbrand (2007), is proposed by Lee (2011). Another dispatch method, based on the maximal covering location problem developed by Church and ReVelle (1974), is presented by Lim, Mamat, and Bräunl (2011) and it is shown by simulation that response times to urgent calls can be reduced.

A common way to solve the dynamic ambulance relocation problem is the offline approach: redeployment decisions are precomputed for different states of the system. For instance, compliance tables are computed, which prescribe desired locations for idle ambulances by Gendreau et al. (2006). With this purpose, the Maximal Expected Coverage Relocation Problem (MECRP) is proposed and solved, by formulating this problem as an integer linear program. It is stated by Maleki, Majlesinasab, and Sepehri (2014), that computing compliance tables is just the first part of the computation of relocation decisions. The second part involves the actual assignment of ambulances to base stations. Therefore, the Generalized Ambulance Assignment Problem (GAAP) and Generalized Ambulance Bottleneck Assignment Problem (GABAP) are proposed. Compliance tables are the subject of Alanis et al. (2013) as well: a two-dimensional Markov chain is proposed and analyzed to obtain optimal compliance tables. A two-stage stochastic optimization model for the ambulance redeployment problem that minimizes the number of relocations while maintaining an acceptable service level is presented by Naoum-Sawaya and Elhedhli (2013).

In addition to the offline approach, a large part of the ambulance literature focuses on the online computation of relocation decisions. Whenever an event occurs, e.g., an ambulance becomes available again, the dispatcher has the opportunity to control the system. Based on the information of the state of the system, one computes a relocation decision. Such a relocation decision needs to be obtained in a very short time, and thus is the main focus of this literature on heuristics. For instance, a heuristic called the Dynamic Maximum Expected Coverage Location Problem (DMEXCLP) is proposed by Jagtenberg et al. (2015). This problem, based on the MEXCLP presented by Daskin (1983), computes a new location for an ambulance that just finished service of a patient. Moreover, a parallel tabu search heuristic is used for the real-time redeployment of ambulances by Gendreau, Laporte, and Semet (2001). Andersson and Värbrand (2007) use the notion of preparedness. This preparedness is a measure for the ability to serve potential patients now and in the future. Moreover, a dynamic relocation model named DYNAROC and a heuristic to solve this model is presented. In addition, some papers use approximate dynamic programming for determining relocation strategies, for instance, Maxwell, Restrepo, Henderson, and Topaloglu (2010); Maxwell, Henderson, and Topaloglu (2013) and Schmid (2012). Relocation decisions are made at the time of call arrivals and when an ambulance becomes available again by Maxwell (2011). In this work, it is shown that making relocation decisions at such times is equivalent to the usage of a nested compliance table policy. At last, a comprehensive study on both online and offline redeployment is executed by Zhang (2012).

1.2. Our contribution

In this paper, we study the relationship between number of ambulance relocations and the performance of the ambulance service provider. Theretofore, we present an ambulance redeployment model, in which we are able to incorporate different performance criteria. We use a heuristic method that computes an action concerning the relocation of ambulances in such a way that the expected performance is maximized. This computation is done at decision moments: the time of occurrence of a new incident or the time of the idle report of an ambulance. We use a heuristic policy instead of the optimal one because computation of the optimal policy is very complex, if not impossible. Besides, even if it was possible to compute, the optimal policy is probably a complex one: it is not easy to understand and to execute by the dispatcher. Instead, we use a heuristic method that is not too far-fetched, while it is highly likely that this heuristic policy contains the same characteristics as the optimal one.

This paper differs from the mainstream literature in two respects

1. Most of the papers in the literature, e.g., Jagtenberg et al. (2015), assume that the computed action is always carried out. However, it may be the case that the expected gain in performance by taking this action is very small and that this benefit does not outweigh the disadvantages regarding the