Knowledge-Based Systems 70 (2014) 268–280

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Dynamic coordination of ambulances for emergency medical assistance services

Holger Billhardt [a,*], Marin Lujak [a], Vicente Sánchez-Brunete [b], Alberto Fernández [a], Sascha Ossowski [a]

[a] CETINIA, Universidad Rey Juan Carlos, Calle Tulipán s/n, 28933 Móstoles, Madrid, Spain
[b] Medical Emergency Service SUMMA 112, Comunidad de Madrid, C./ Antracita 2 Bis. 2, 28045 Madrid, Spain

ARTICLE INFO

Article history:
Received 2 October 2013
Received in revised form 24 June 2014
Accepted 14 July 2014
Available online 23 July 2014

Keywords:
Emergency vehicle management
Multiagent systems
Coordination
Dynamic optimization
Dynamic fleet management
Dispatch systems

ABSTRACT

The main objective of emergency medical assistance (EMA) services is to attend patients with sudden diseases at any possible location within an area of influence. This usually consists in providing "in situ" assistance and, if necessary, the transport of the patient to a medical center. The potential of such systems to reduce mortality is directly related to the travel times of ambulances to emergency patients. An efficient coordination of the ambulance fleet of an EMA service is crucial for reducing the average travel times. In this paper we propose mechanisms that dynamically improve the allocation of ambulances to patients as well as the redeployment of available ambulances in the region under consideration. We test these mechanisms in different experiments using historical data from the EMA service of the Autonomous Region of Madrid in Spain: *SUMMA112*. The results empirically confirm that our proposal reduces the average response times of EMA services significantly.

1. Introduction

The domain of medical assistance in general, and in emergency situations in particular, includes many tasks that require flexible on-demand negotiation, initiation, coordination, information exchange and supervision among different involved entities (e.g., ambulances, emergency centers, hospitals, patients, physicians, etc.). Among such tasks the coordination of the available resources to provide assistance to emergency patients as fast as possible is of crucial importance for obtaining an efficient service. The main goal here is to improve one of the key performance indicators: the response time (time between a patient call and the moment an ambulance arrives and the patient can receive medical assistance).

There is a general understanding that shorter response times are an essential starting point to improve care and reduce mortality [1,2]. This holds especially for severe injuries. In order to assure the quality of emergency services many countries and regions specify response time limits for EMA service provider organizations either by law or through contractual norms. In Europe and in the Unites States such limits usually lie between 8 and 15 min. In the UK, for instance, a national standard defines that at least 75 per cent of Category A (immediately life-threatening) calls should be responded within 8 min.

Even though response time standards are often fixed by norms, they can be rather considered as targets that EMA service providers continuously try to reach.

One way to reduce response time consists in reducing the part that depends on the logistic aspects of an EMA service: the travel or arrival times of ambulances to emergency patients. There are two main problems EMA managers are faced with in the logistic part of a service: allocation and redeployment of ambulances. The allocation problem consists in determining an ambulance that should be sent to assist a given patient. Redeployment consists in relocating available ambulances in the region of influence in a way that new patients can be assisted in the shortest time possible.

In this article we present a novel coordination model for ambulance fleets that combines a mechanism for dynamic redeployment of available ambulances with a dynamic allocation of ambulances to patients. Regarding ambulance redeployment, we propose a method, based on centroidal Voronoi tessellations, that tends to optimize the allocation of idle ambulances in each moment with respect to the probability distribution of possible emergency cases. Regarding ambulance allocation, we propose to use a dynamic auction-based assignment of patients to ambulances that tends to optimize the sum of the expected arrival times in each particular moment. EMA services are highly dynamic; e.g., new emergency patients will have to be attended and previous missions will finish.

* Corresponding author. Tel.: +34 916647458.
E-mail addresses: holger.billhardt@urjc.es (H. Billhardt), marin.lujak@urjc.es (M. Lujak), vicente.sanchezbrunete@salud.madrid.org (V. Sánchez-Brunete), alberto.fernandez@urjc.es (A. Fernández), sascha.ossowski@urjc.es (S. Ossowski).

http://dx.doi.org/10.1016/j.knosys.2014.07.006
0950-7051/

We present an event-driven system that dynamically executes ambulance assignment and redeployment and, thus, continuously tends to optimize the situation of the ambulance fleet with regard to the changes that occur in an EMA service.

The outline of the rest of the paper is as follows. Section 2 presents related work and relates our approach to others. In Section 3 we provide a brief description of the operation of EMA services. Then we present our ambulance allocation and redeployment mechanisms, and we expose an event-driven architecture for employing both mechanisms dynamically in real time. In Section 4 we present an experimental evaluation of our proposal and compare it with the operation strategies currently used by SUMMA112, the EMA service provider organization in the Autonomous Region of Madrid in Spain.[1] The experiments have been carried out in a simulated environment and using real patient data from Madrid. Finally, Section 5 gives some conclusions and points out some aspects of our current and future research.

2. Related work

There have been many proposals for the problem of coordinating ambulance fleets for EMA services. Brotcorne et al. provides a good review of ambulance allocation and redeployment strategies from the early 1970s through 2003 [3]. More recent reviews have been done by Li at all. [4] and Aboueljinane et al. [5]. Whereas the former concentrates on covering models and optimization techniques for facility location, the latter analyzes the use of simulation models in emergency medical service operations.

To the best of our knowledge, most of the work has been dedicated to the redeployment or coverage problem, e.g., the optimal location of ambulances in an region such that all points can be reached within a predefined time standard. Early approaches concentrate on a static distribution of ambulances. The Location Set Covering Problem, proposed by Toregas et al. [6] tries to find the minimum number of emergency facilities and their locations to cover all demand assuming that the demand occurs at a finite set of points. In the Maximal Covering Location Problem proposed by Church and Re Velle [7], the aim is to locate a fixed number of facilities in order to maximize the population covered within some service distance. Such static methods do not take into account the relation between mobility and coverage of ambulances. In particular, demand points will be uncovered if one or more ambulances are called for service. To overcome this problem, researchers have proposed to maximize the coverage of demand points by more than one ambulance or by using double standards for coverage (e.g., [8,9]). Another trend has been to establish probabilistic models, that explicitly model the availability or the travel and assistance times of ambulances (e.g., [10,11]).

More recent research on the covering problem has concentrated on the dynamic location of ambulances, where methods are proposed to redeploy ambulances during the operation of a service in order to take into account the intrinsic dynamism of EMA services. In [12], Gendreau et al. extend their Double Standard Model to reflect the dynamic nature of the problem. They propose to use tabu search heuristics and solve the model through a (non-exhaustive) pre-computation of redeployment scenarios. In [13], the same authors propose the Maximal Expected Coverage Relocation Problem and present a strategy for dynamically relocating idle ambulances that are located in low demand areas. Rajagopalan et al. [14] developed another dynamic model for redeploying ambulances to predictable demand fluctuations in time and space. The objective of the model is to determine the minimum number of ambulances and their locations for each time cluster in which significant changes in demand patterns might occur while meeting coverage requirement with a predetermined reliability. Whereas the previous methods require solving integer programs, in [15], Maxwell et al. propose to use an approximate dynamic programming approach for ambulance redeployment. To deal with the high-dimensional state space in the dynamic program, they construct approximations to the value function that are formulated in terms of the percentage of calls that are reached within a time standard. Naoum-Sawaya and Elhedhli [16] present a two-stage stochastic optimization model that minimizes ambulance relocations while maintaining acceptable service level. While many approaches are based on centralized optimization, the solution approach of Ibri et al. in [17] is decentralized. The authors propose a multi-agent system that integrates a dynamic ambulance dispatching and redeployment method. However, there are a couple of drawbacks to this method. To limit deviations of vehicles, they allow assigning a vehicle to another call only if this latter has higher priority than the first one, thus, not leaving space for a real dynamic optimization of (current) travel times. Secondly, the vehicles are represented (and grouped) by the station agents and the redeployment is performed among a fixed number of stations, which in the case of an insufficient number and/or position of stations, may result in insufficient coverage.

Most proposals on dynamic redeployment of ambulances (like the ones mentioned before) only consider the possibility to relocate ambulances among different, predefined sites (stations). This requirement is relaxed in the work proposed by Andersson and Varbrand [18]. These authors propose a decision support tool that recommends the redeployment of a fixed number of ambulances to areas with less preparedness (a criteria for coverage) and ambulances can be relocated to any place in the region.

Regarding dispatching strategies for ambulances (the patient allocation problem), there has been less research that treats this problem explicitly. Most works use the "nearest available ambulance" rule for assigning ambulances to patients in a first-came first-served manner. Some works analyze priority dispatching strategies. For instance in [19], Baranda et al. analyze dispatching strategies that take into account the severity level of patients and evaluate the survival probability of patients for different strategies. Also in [18], for calls with the highest priority the vehicle with the shortest travel time is assigned, whereas for less severe patients a vehicle is dispatched that reaches the patient in a given time limit but harms less some coverage criterion. López et al. [20] propose a multiagent system where ambulances are also assigned based on the severity of the patients. Besides the distance, the system also takes into account a trust value, that reflects the belief that an ambulance can fulfil its obligations in time. In this sense, expert drivers will have higher trust values than novice drivers. A more complex approach is presented by Haghani et al. [21], where the system dynamically optimizes the total travel time (ambulances to patients, ambulances to base stations and ambulances to hospitals).

Our redeployment approach differs from others in the sense that we do not try to maximize the zones in a region that are covered with respect to some time limits. Instead, we use an approach based on geometric optimization [22] that tends to optimize in each moment the positions of all ambulances that are still available such that the expected arrival time to potential new emergency patients is minimized. Using centroidal Voronoi tessellations, that are scalable with the number of agents in the network [23], we compute optimal ambulance positions dynamically in real time. The latter takes into account the probability distribution of emergency cases in the region (at different times of the day), based on historical data. Furthermore, in our approach, ambulances can be redeployed to any point in the region and all idle ambulances (and not only a limited number) are dynamically redeployed

[1] www.madrid.org/cs/Satellite?pagename = SUMMA112/Page/S112_home.