Simulation Modelling Practice and Theory 60 (2016) 170–184

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

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

A novel hybrid method for improving ambulance dispatching response time through a simulation study

Mehdi Zarkeshzadeh, Hadi Zare *, Zainabolhoda Heshmati, Mehdi Teimouri

Department of Network Science and Technology, Faculty of New Sciences and Technologies, University of Tehran, Iran

ARTICLE INFO

Article history:
Received 23 June 2015
Received in revised form 30 October 2015
Accepted 31 October 2015
Available online 14 November 2015

Keywords:
Emergency medical services
Ambulance dispatching
Response time
Centrality measures
Simulation

ABSTRACT

Response time is the most important factor in evaluating the performance of various Emergency Medical Services (EMS). In this paper, a novel hybrid method has been proposed to improve response time for ambulance dispatching. The proposed approach uses a linear hybrid metric based on network centrality measures, nearest neighbor method and first-in first-out (FIFO) policy. Other important parameters in ambulance dispatching such as the operating environment, rate of incoming emergency calls, available resources, hospitalization probability of the patients as well as distances and locations of units are all part of information used in this proposed approach. In line with the traditional metrics used in previous works, we have adopted a linear combined metric which is adjusted according to environment parameters. Results of extensive simulation experiments show reductions in response time by as much as 42% as compared to previous methods.

1. Introduction

In recent years, the number of natural disasters has increased sharply and the management of aid and relief in these situations is of particular importance. During the first six months of 2013, 127 disaster events have occurred with associated fatalities of 8250, influencing over 16 million people, and estimated economic damages to 45,129 million dollars in total [1]. One of the major causes of death and disability in the world is due lack of attention and high response times in emergencies [2]. Each year, half of the 270,000 people who suffer a heart attack in Britain, lose their life before they reach the hospital. In the first 10 min of cardiac arrest, reduction of response time by every minute, increases the odds of survival by between 7% and 10% [3]. In Wales and England the economy of reducing the response time by one minute is estimated about 54 million Euros [4].

There is a vast amount of researches on simulation-driven decision making and expert systems for disaster and critical situations based on a variety of techniques and applications such as simulation-based modelling for hospital emergency service design [5], incident information management system [6], simulation of the reliability of healthcare systems [7], Viable System Model for increasing information processing [8], simulation study on the performance of an EMS department [9] and simulation-based decision support system for prediction strain situations in an emergency department [10].

In ambulance dispatching, response time is the most popular indicator for the assessment of the quality of emergency medical services [11]. The time interval between a patient's call to the emergency services and the arrival of the ambulance

* Corresponding author.
E-mail addresses: mehdi.zarkesh@ut.ac.ir (M. Zarkeshzadeh), h.zare@ut.ac.ir (H. Zare), zheshmati@ut.ac.ir (Z. Heshmati), mehditeimouri@ut.ac.ir (M. Teimouri).

http://dx.doi.org/10.1016/j.simpat.2015.10.004
1569-190X/

is called response time, [12–17]. This time is also sometimes called the waiting time [12]. It relates directly to people's health, and it is therefore used as a performance index in the related literature [14].

Some of the important factors affecting the response time are number of ambulance fleets, their location, traffic conditions and possible delays in communicating with the emergency control center [16]. Many of the previous literature have focused on finding the optimal location for the ambulances and their relocations [18,19]. Ambulance dispatching and relocation decisions are mainly based on changes in temporal and geographical patterns of requests in order to cover more number of calls in a certain time [13]. Ambulance dispatching in emergency medical services provides the appropriate decisions for allocation of ambulances to calls in order to reduce the response time [20].

In EMS, ambulance dispatching decisions are usually made under two general situations; first is when the number of ambulances is greater than the number of calls and hence the system load is low [21]. This situation is also known as call-initiated [22], because when a call is established, an ambulance is available and ready which is dispatched to serve the call. In the second and most vital situation, the number of calls are greater than the number of ambulances available and hence the system load is high. In this case, calls are listed in a waiting queue for ambulances to be released and dispatched to serve the call. This situation is also called ambulance-initiated, where the incoming calls await a free ambulance [22]. This situation is more common in natural or man-made hazards or disasters where they lead to increased traffic, more emergency calls, injuries and deaths [12]. This paper focuses on situations that are ambulance-initiated where loading of system is high.

There are several methods for ambulance dispatching. The most common method is the nearest neighbor (NN) where the closest (geographically) available unit is dispatched to the call [23,24]. Although dispatching the closest ambulance to the call reduces the response time of the current call, it may cause an increase in the response time of the next incoming calls. In certain conditions, dispatching the ambulance to other calls based on alternative policies may lead to better performance with respect to the NN policy [13].

First-in first-out (FIFO) and prioritizing the emergency calls, are among the simplest methods of ambulance dispatching policies. The former dispatches ambulances based on the order of incoming calls and the latter is based on the priority of the incoming emergency call. A hybrid method would be to sort the waiting queue based on FIFO and the priority of the emergency call and then dispatch ambulances based on the NN policy [25].

According to statistics, a significant percentage of incidents are treated and served on site and do not require transportation to the hospital. The percentage of such incidents in United Kingdom has been reported as 34% nationwide with a maximum of 54% in a particular area [26]. Therefore an ambulance may have the chance to serve multiple calls before having to transfer a patient to the hospital.

Because of the greedy behavior of the NN method and the lack of attention to the long-term performance, various algorithms are presented for dispatching, [16,23,25,27,28]. In [16,25] some rules are presented for ambulance dispatching. In [16], if ambulance A is dispatched to call C, and then ambulance B becomes free which is closer to call C, in this case ambulance B is dispatched to call C. In [25], if an ambulance is responsible for service in area P and is busy when a call is received in area P, another ambulance should be dispatched from an area with a lower probability of calls.

Other ambulance dispatching approaches have been proposed based on the priority of the incoming emergency calls and the preparedness criteria of the area of responsibility [12]. The calls are divided into three categories based on their urgency and are given priorities. The nearest ambulances are dispatched to first priority calls. However, dispatching decisions for the second and third (lower) priority calls are based on the preparedness of the area to serve other potential patients with ambulances available. Later on a composite algorithm was introduced to improve the proposed preparedness approach [13].

Another method for ambulance dispatching has been introduced based on the centrality of calls using network analysis approaches [28]. Using this approach, in disaster situations, an ambulance is dispatched to a call which is situated in a dense area of calls in order to reduce the average response time in the long run.

While the previous works have concentrated on a variety of situations, they suffer from other properties such as lack of attention to the area of the call or different call priorities. Hence, in this paper, we propose to overcome such problems by introducing a novel hybrid method for more efficient ambulance dispatch and lower response times in critical conditions.

In line with the traditional metrics used in previous works, we have adopted a linear combined metric which is adjusted according to environment parameters. In our combined approach, linear combination of network centrality measures, the nearest neighbor metric and FIFO method are considered based on adjusted weights according to environmental setting. The optimal weights for environmental settings are derived through extensive simulation studies.

The overall structure of this paper is as follows. Section 2 describes the centrality based ambulance dispatching method. The initial proposed linear hybrid method based on Centrality and NN are introduced in Section 3. The novel weighted hybrid method based on Centrality, NN and FIFO is then presented in Section 4. Furthermore, the extensive experiments and the corresponding results based on proposed methods are illustrated in Sub section 4.1. Finally, Section 5 concludes the results and discusses possible future works.

2. Ambulance dispatching based on centrality

Centrality of a node in a network indicates the importance of that node in the operational efficiency of the network. When incoming emergency calls are prioritized based on their centrality, ambulances are dispatched to a call which is most central