OR Spectrum
https://doi.org/10.1007/s00291-024-00744-4

ORIGINAL ARTICLE

Online algorithms for ambulance routing in disaster response with time-varying victim conditions

Davood Shiri[1] · **Vahid Akbari**[2] · **F. Sibel Salman**[3]

Received: 10 December 2022 / Accepted: 4 January 2024
© The Author(s) 2024

Abstract

We present a novel online optimization approach to tackle the ambulance routing problem on a road network, specifically designed to handle uncertainties in travel times, triage levels, required treatment times of victims, and potential changes in victim conditions in post-disaster scenarios. We assume that this information can be learned incrementally online while the ambulances get to the scene. We analyze this problem using the competitive ratio criterion and demonstrate that, when faced with a worst-case instance of this problem, neither deterministic nor randomized online solutions can attain a finite competitive ratio. Subsequently, we present a variety of innovative online heuristics to address this problem which can operate with very low computational running times. We assess the effectiveness of our online solutions by comparing them with each other and with offline solutions derived from complete information. Our analysis involves examining instances from existing literature as well as newly generated large-sized instances. One of our algorithms demonstrates superior performance when compared to the others, achieving experimental competitive ratios that closely approach the optimal ratio of one.

Keywords Disaster relief · Ambulance routing · Online optimization · Competitive ratio

✉ Davood Shiri
d.shiri@sheffield.ac.uk

Vahid Akbari
vahid.akbari@nottingham.ac.uk

F. Sibel Salman
ssalman@ku.edu.tr

[1] Sheffield University Management School, University of Sheffield, Conduit Road, Sheffield S10 1FL, United Kingdom

[2] Nottingham University Business School, University of Nottingham, Jubilee Campus, Nottingham NG8 1BB, United Kingdom

[3] College of Engineering, Koç University, Sariyer, Istanbul 34450, Turkey

Published online: 07 February 2024

1 Introduction

Disasters have profound consequences on the well-being of people worldwide. As per data from the emergency events database (EM-DAT) (CRED 2013), there has been a notable rise in both the frequency and impact of natural and man-made disasters on a global scale. According to the findings reported by CRED (2020), from 2000 to 2019, a total of 7,348 disaster events were documented. These incidents tragically resulted in the loss of 1.23 million lives and had a profound impact on the lives of 4.2 billion individuals, while also causing over 2.97 trillion dollars in economic losses globally. It is widely recognized that the effectiveness of relief operations during the response phase plays a crucial role in mitigating the severity of these disasters. Unfortunately, a considerable portion of the victims may perish due to lack of timely medical aid in the aftermath of a disaster (Farahani et al. 2020; Lu et al. 2019). Therefore, it is essential to prioritize victims to provide them with the needed treatments and to efficiently utilize the medical resources in post-disaster situations and mass casualty incidents. The prioritization of the victims is done by *triage*, which refers to the dynamic process of categorizing the victims (into different triage levels) with respect to the degree of severity of their injury and how likely their conditions could get worse due to untimely treatment (Oksuz and Satoglu 2020).

In this study, among the disaster response operations, we concentrate on the transportation of the casualties by ambulances. Namely, we address the online routing and scheduling of ambulances. The main mission of the paramedics following a disaster is to provide initial medical care to individuals with minor injuries on-site and to transport critically injured patients to hospitals using ambulance services (Shiri et al. 2023; Aringhieri et al. 2022; Rabbani et al. 2022; Talarico et al. 2015). However, efficient planning of these tasks is tremendously complicated in post-disaster scenarios. The ambulances must be dispatched under extremely challenging circumstances such as damaged communication and transportation infrastructure, limited supplies and resources, as well as uncertainty about the condition of the victims who are in need of medical assistance (Akbari and Shiri 2022; Najafi et al. 2014). Therefore, having an online decision support tool that produces effective solutions which can operate under uncertain information and can be executed in a short time for the optimization of routing and scheduling of ambulances plays a vital role in disaster response operations (Shiri et al. 2023, 2020; Bélanger et al. 2019).

Over the last few years, several articles have analyzed the *Ambulance Routing Problem* (ARP) by considering an offline setting for problem inputs, where all the information about the victims and the roads are available before making decisions, e.g., Rabbani et al. (2022); Tlili et al. (2018); Talarico et al. (2015). Some other articles studied the ARP under a stochastic setting, where the triage level of each victim is associated with a known probability distribution, e.g., Yoon and Albert (2020); Schilde et al. (2011). However, in the emergency disaster response phase, due to the uncertainty caused by the disaster, it is not realistic to assume predetermined or probabilistic problem inputs as each disaster

situation is unprecedented, and historical data are not pertinent. Recently, Shiri et al. (2023) studied an online version of the ARP where the triage levels and the treatment times of certain victims are not initially known to the ambulances. They analyzed the problem under the assumption that the unknown information associated with a victim is revealed as soon as one of the ambulances examines the victim. In this study, we build upon the research conducted in Shiri et al. (2023) by adding two realistic features to the online ARP in mass casualty incidents: (1) time-varying victim conditions and (2) real-time road conditions. Notably, the integration of these dynamic elements represents a novel and distinct methodological approach, significantly diminishing the efficacy of conventional solution techniques and making mathematical models impractical due to extremely incomplete information. Hence, necessitating the exploration of alternative adaptive online heuristics to effectively cope with these challenges.

We emphasize that in the online ARP, decisions concerning prioritization of the treatment of the victims as well as the selection of the available hospitals to deliver the critically injured victims must be made online with respect to the information available up to that point of time. In fact, the online variation of the ARP involves several discrete decision-making time instants. At each of these decision-making points, a new piece of information is revealed and decisions should be updated accordingly. Here, we introduce the online ambulance routing and scheduling problem with time-varying victim conditions and real-time road information and refer to it as the *Online Ambulance Routing Problem* (OARP).

We first analyze the OARP through a worst-case competitive analysis lens and demonstrate that there exists no online algorithm with a finite competitive ratio for this problem. In this way, we show the theoretical value of having access to complete information prior to solving the problem. In order to handle real-life instances, we propose various online policies to tackle the OARP. We test the performance of our policies on instances adopted from Talarico et al. (2015) and larger size instances which we have introduced in our study. To assess the effectiveness of our proposed solutions, we compare them against each other and the offline solutions derived from complete information. Notably, our algorithms exhibit remarkably fast execution times, often taking less than 1 s, even when dealing with the largest tested instances consisting of 100 nodes. Consequently, they are exceptionally well-suited for practical use in real-life emergency and post-disaster scenarios. These algorithms excel in managing problem uncertainties and swiftly generating and updating solutions based on new information.

The structure of this article is outlined as follows. We commence with a review of relevant literature in Sect. 2. In Sect. 3, we offer comprehensive explanations of both the offline and online problems. Section 4 is dedicated to a theoretical worst-case competitive analysis of the OARP. In Sect. 5, we introduce our online heuristic policies tailored for solving the OARP. Detailed results from our computational experiments are presented in Sect. 6. Finally, the article is concluded with a discussion of closing remarks and avenues for future research in Sect. 7.

2 Literature review

In this section, we commence by examining prior research in offline optimization, specifically focusing on studies related to ambulance routing and scheduling that operate with complete information. Subsequently, we delve into a review of relevant literature in the fields of stochastic and online optimization.

2.1 Offline optimization literature

Within the realm of articles dedicated to studying offline vehicle routing problems, our focus centers on papers that tackle the routing and scheduling of ambulances. More specifically, our attention is drawn to articles that engage with the ambulance routing problem while integrating considerations such as hospital capacities and the categorization of victim groups. The ARP takes center stage in the study by Talarico et al. (2015). Their research addressed the task of identifying optimal ambulance routes for serving a predetermined set of red and green code victims, operating under the premise of complete input information. The objective function in their problem is the minimization of a weighted linear combination of the latest completion times for both red and green victims. Talarico et al. (2015) introduces two exact mathematical formulations tailored to tackle the ARP. Furthermore, the authors proposed an effective large neighborhood search procedure designed to tackle larger instances of the problem. Noteworthy is their introduction of a comprehensive set of 1296 ARP instances, which serves as the foundation for testing the validity of their mathematical models and heuristic approach.

Tlili et al. (2017) formulated the ARP as an open vehicle routing problem, seeking to minimize the cumulative travel distance. They introduced a cluster-first route-second procedure based on a combination of the particle swarm optimization approach and the *petal algorithm*. We note that the petal algorithm is a technique that generates numerous routes (petals) for the vehicle routing problem and employs a set partitioning approach to identify the optimal combination of these routes. The problem was further investigated by Tlili et al. (2018) in the same vein, presenting an efficient heuristic algorithm and evaluating its performance through small case studies.

Distinctly, Salman and Gül (2014) sidestepped triage considerations and proposed a mixed-integer programming formulation which concurrently optimizes capacity allocation and victim transportation decisions over multiple periods, accounting for varying casualty numbers at different disaster-affected locations. Their model minimizes a weighted linear combination of the total cost of setting up new facilities as well as the total travel and waiting time of victims. Zidi et al. (2019) investigated the ARP by considering only one type of victims where the goal is to minimize the total travel distance. Their solution methodology involves a combination of the cluster-first route-second technique, tabu search, and simulated annealing. Talebi et al. (2022) employed the multi-objective bees algorithm to address a version of the ARP, where two distinct groups of victims are considered, namely red and green. The objective in their work is to simultaneously minimize both the total travel

distance and the latest service completion time. In a recent research endeavor, Aringhieri et al. (2022) studied the ARP with a distinctive emphasis on ensuring fairness and equity in the delivery of services to victims. To tackle this challenge, they conceived a novel problem variant, drawing inspiration from the team orienteering problem. Subsequently, they introduced an innovative hybrid solution approach, which functions based on a neighborhood search procedure.

Other facets of the ARP have garnered attention in recent literature. Tikani and Setak (2019) investigated a variation of the ARP with three different triage levels for victims and different types of ambulances. They provided a mathematical formulation to optimize routes with the goal of minimizing the latest service completion time for the victims. Additionally, they introduced two meta-heuristic procedures based on genetic algorithm and tabu search to tackle larger problem instances. In a more contemporary perspective, Rabbani et al. (2022) extended the scope of the offline problem by accounting for dynamic changes in victim conditions over time and incorporating three triage levels. They introduced two competing objective functions: one targeting the minimization of service completion times and another aimed at reducing the count of victims experiencing worsened conditions due to delays. Treating these objectives as inherently conflicting, the study emphasizes the pursuit of Pareto optimal solutions. Their methodology encompasses mixed-integer programming formulations, as well as innovative techniques such as the multi-objective particle swarm optimization as well as the genetic algorithm.

In a recent study, Yazdani and Haghani (2023) considered the ARP in response to pandemic outbreaks. Their problem does not consider triage and time-varying victim conditions. They proposed various heuristic algorithms which are based on local search and genetic algorithm to address their problem. An indirectly associated research domain with the ARP is disaster relief routing, encompassing an extensive array of articles. For an in-depth exploration of this field of study, we direct the reader to review papers of Wadi Khalid et al. (2021), Aringhieri et al. (2017), and De la Torre et al. (2012).

The articles covered in this section deviate from our study as they function with complete input information. In our paper, we approach the ARP by integrating the uncertainty regarding the victim conditions and road travel information, thereby capturing the impact of lack of accurate information in the disaster response phase in our problem and analysis.

2.2 Stochastic and online optimization literature

A body of literature delves into dynamic vehicle routing problems, but these cases typically fall beyond the confines of ambulance routing or disaster management, e.g., see the surveys of Soeffker et al. (2022); Rios et al. (2021); Pillac et al. (2013); Ghiani et al. (2003). In a remotely related vein, certain studies explore post-disaster relief operations with a dynamic or real-time perspective, e.g., see the surveys of Tippong et al. (2022) and Farahani et al. (2020). This stream of research, however, diverges from our specific problem by not incorporating elements such as victim groups, ambulance and hospital capacities, as well as the routing of ambulances and

the transportation of victims to medical facilities. Given our focus on ambulance routing and scheduling in the presence of uncertainties, this literature review section primarily centers on the routing of ambulances while accounting for varying uncertainty models.

In Schilde et al. (2011), a dynamic stochastic ARP is explored, omitting triage considerations. The study focuses on transferring patients between their locations and hospitals, in both directions. In their problem, certain requests are predetermined, while others exhibit dynamic characteristics or are linked to probability distributions. They formulated the problem as a dynamic stochastic dial-a-ride problem involving expected return transports and introduced four distinct adaptations of a meta-heuristic algorithm. In another recent relevant research, Ritzinger et al. (2022) provide a comprehensive comparative analysis of dynamic anticipatory algorithms for the dial-a-ride problem in stochastic environments.

In Oksuz and Satoglu (2020), they proposed a two-stage stochastic programming framework for planning victim transportation, which considers triage levels and strategically placing temporary medical centers. They demonstrated the practicality of their model through a case study focused on the possibility of an earthquake occurring in Istanbul, Turkey. In a separate investigation, Yoon and Albert (2020) explored the dynamic ARP involving two ambulance types while facing uncertainty regarding victim triage levels. In their problem, whenever a new victim request reveals, the appropriate ambulance type for dispatch to the location of the victim must be determined. Their analysis incorporates prior probabilistic knowledge related to uncertain information, and they presented a Markov decision process model for determining the dispatched vehicle types.

Lee et al. (2022) conducted an analysis on ambulance routing and relocation, taking probabilistic demand into account. Their primary goal is to minimize the total time required to transport patients to medical facilities. The authors analyzed the problem under the assumption that victim requests follow a Poisson distribution and provided a hybrid solution methodology involving a Lagrangian dual decomposition procedure and a branch-and-bound procedure. More recently, Khoshgehbari et al. (2023) addressed the challenge of handling uncertainties by introducing an ambulance location-routing problem. This study formulates a two-stage integer stochastic programming model that incorporates the treatment golden time as a critical determinant of service quality. Moreover, it accommodates a heterogeneous fleet of ambulances and various victim groups. The authors presented a novel heuristic approach to solve this problem effectively.

It is worth noting that the methodologies proposed in the previously mentioned articles differ from our approach in this study. Our research focuses on addressing the uncertainty for the ARP, specifically through competitive analysis and online optimization. In contrast to the above-mentioned prior approaches, we do not rely on any predefined probabilistic assumptions regarding the incomplete information. Recent literature has extensively explored vehicle routing problems that involve online uncertainty, where a post-disaster road network is considered. In these studies, locations to be serviced (Shiri et al. 2020) and/or the road conditions (Akbari and Shiri 2022; Akbari et al. 2021; Shiri and Salman 2020) are not known at the beginning and are explored dynamically over time. However, none of these online

optimization problems address the online routing and scheduling of ambulances as they do not involve transportation of victims to hospitals as well as any uncertainty about the triage levels of victims.

The most related paper to our work is the recent study by Shiri et al. (2023), which explores a reduced version of the OARP under simplifying assumptions of known travel times and constant victim conditions. They proved various finite lower bounds on the competitive ratio of online solutions. Furthermore, they proposed three different online algorithms, namely a formulation-based algorithm which relies on solving a mathematical model, a machine learning-based clustering approach, as well as a novel heuristic algorithm to solve their problem. They empirically confirmed the good performance of their algorithms on the instances from (Talarico et al. 2015). Building upon this foundation, our paper extends their findings by incorporating two crucial aspects of the ARP in mass casualty situations: the dynamic nature of victim conditions and real-time road conditions. This broader perspective acknowledges the potential changes in victim severity and medical requirements over time, crucial in immediate post-disaster scenarios. Moreover, we integrate real-time road conditions, equipping each ambulance with access to real-time travel times of only those routes whose starting node is the current location of the ambulance. This introduces a level of uncertainty, as precise travel times of any other routes are concealed from the ambulance in real time preventing the use of traditional routing models.

3 Problem descriptions

We start by formally describing the offline version of the problem (ARP). In the ARP, all the input parameters are known beforehand. Following that, we delve into the online version (OARP), providing a detailed description along with the online input parameters and how to obtain their exact values. For convenient reference, we present the notations used in both the ARP and the OARP in Table 1.

3.1 The offline problem

The offline version of the ARP is defined assuming that ambulances have complete information about the status of the victims and road conditions before they start their routes. With regard to the existing standard literature, three triage levels for the victims are considered in our study; **red code**: a critically injured victim who should be transferred to a hospital by one of the ambulances (Talarico et al. 2015), **green code**: a slightly injured victim who should be helped directly on-site by the medical team of an ambulance (Talarico et al. 2015), and **black code**: a passed away victim (Rabbani et al. 2022). Hence, the set of all victims is denoted by $V = R \cup G \cup B$, where R, G, and B are sets of red, green, and black code victims, respectively. Furthermore, the possibility of changes in the health conditions of the victims over time has been recently addressed in the offline problem (Rabbani et al. 2022). To this end, each green victim $g \in G$ is associated with two parameters T^g_{gr} and $T^g_{gb} \geq T^g_{gr}$ which are time thresholds after which the triage level of g turns to red and black, respectively.

Table 1 Table of notation for the ARP and the OARP

Notation	Description	ARP Inputs	OARP Inputs
$N = H \cup R \cup G \cup B$	Set of all nodes	✓	✓
$V = R \cup G \cup B$	Set of all victims	✓	✓
R	Set of red code victims	✓	
G	Set of green code victims	✓	
B	Set of black code victims	✓	
H	Set of hospitals	✓	✓
A such that ($\|A\| \leq \|V\|$)	Set of ambulances	✓	✓
$E = \{V \times V\} \cup \{V \times H\} \cup \{H \times V\}$	Set of directed links	✓	✓
S_v	Treatment time of victim $v \in V \setminus B$	✓	
T^g_{gr}	Time threshold by which triage level of a green victim $g \in G$ turns to red	✓	
T^g_{gb}	Time threshold by which triage level of a green victim $g \in G$ turns to black	✓	
T^r_{rb}	Time threshold by which triage level of a red victim $r \in R$ turns to black	✓	
C_h	Capacity of hospital $h \in H$	✓	✓
$d_a \subset H$	Initial location of ambulance $a \in A$	✓	✓
A_h	Number of ambulances initially located in hospital $h \in H$	✓	✓
t_{ij}	Travel time of link $(i,j) \in E$	✓	
S_h	service time to drop off a red code victim at hospital $h \in H$	✓	✓
W_R	Weight of urgency of treatment of red code victims	✓	✓
W_G	Weight of urgency of treatment of green code victims	✓	✓

We note that the notation T^g_{gb} denotes the cumulative time from time zero until a green victim g meets the condition for turning black. Similarly, each red victim $r \in R$ is associated with a parameter T^r_{rb} which is a time threshold after which the triage level of r turns to black. We remark that in the offline problem, the location of a black code victim is not visited by any of the ambulances since it is assumed that a black code victim must be serviced by non-medical teams due to limited medical resources in mass casualty incidents. In the online problem, however, because the conditions of the victims are not known beforehand, all of the victim locations should be visited by the ambulances.

In the ARP, which involves several hospitals and ambulances, each ambulance can transport at most one red victim before promptly delivering them to a hospital. As for green victims, they can receive on-site care, allowing an ambulance to proceed directly to another victim's location after providing assistance. This flexibility enables ambulances to efficiently assist multiple victims during their routes before returning to a hospital. We also note that if for a green victim g, if they are not visited before T^g_{gr}, once they are visited, they should be transferred to a hospital as their triage changes to red. To provide clarity, it is important to mention that a route is referred to as a *tour* within this context. A tour begins at a hospital, visits at least one victim, and concludes at a hospital, which can be different from the initial hospital. Accordingly, for an ARP solution, it is possible for an ambulance to carry out more than one tour. Let H and C_h denote the set of hospitals and the capacity of hospital $h \in H$, respectively. We represent the ambulance set by A and the starting location of an ambulance $a \in A$ by $d_a \in H$. Accordingly, we show the set of all nodes by $N = G \cup R \cup H = V \cup H$ and the set of directed links by $E = \{V \times H\} \cup \{H \times V\} \cup \{V \times V\}$. We represent the travel time of an arc $(i, j) \in E$ by t_{ij}. Furthermore, each hospital $h \in H$ has a service time S_h to drop off a victim at hospital h and each victim in $v \in V$ has a treatment time S_v. We also let W_G and W_R represent the urgency of treatment of green and red victims, respectively.

For a solution of the ARP, we let n_{rb} represent the number of red victims whose condition changes to black, n_{gr} be the number of green victims whose condition changes to red, and n_{gb} denote the number of victims whose condition changes from green to black. For a green coded victim to turn to black, they must first turn to a red code victim and then if first aid is not provided in the required time, they will perish and turn to black code victims. We also let e_G and e_R be decision variables representing the latest service completion times of green and red victims, respectively. As mentioned, the ARP with the setting discussed above was defined first in Rabbani et al. (2022), where they defined two objective functions. Their first objective minimizes a weighted linear combination of the latest service completion times of green and red victims (i.e., e_G and e_R), and their second objective minimizes a weighted linear combination of the number of victims whose condition gets worse (i.e., n_{rb} and n_{gr}). In their study, they regarded these objectives as two contradicting objective functions and hence the focus of their solution method is on finding Pareto optimal solutions for this problem. Departing from their study, we believe that the most significant objective in such scenarios is to minimize the number of passed away victims rather than minimizing service completion times (Farahani et al. 2020; Jat and Rafique 2020). In fact, we opine that these

Table 2 Decision variables used in the mathematical model of ARP

Notation	Type	Description
x_{ij}	$\{0, 1\}$	Equals 1 when an ambulance traverses link (i, j) in the direction from i to j
y_v	$\{0, 1\}$	Equals 1 when victim v turns to black
z_g	$\{0, 1\}$	Equals 1 when green victim g turns to red
u_{vh}	$\{0, 1\}$	Equals 1 when victim $v \in V$ is brought to hospital $h \in H$
b_v	≥ 0	Visiting time of victim v
e_R	≥ 0	Service completion time for the last red victim
e_G	≥ 0	Service completion time for the last green victim

objectives must be considered hierarchically. To achieve this, we consider three objective functions which are ordered hierarchically as follows.

- **Objective (1):** our first objective is to minimize the number of victims whose condition changes to black, i.e., minimizing $n_{rb} + n_{gb}$.
- **Objective (2):** our second objective is to minimize the number of green victims whose condition changes to red, i.e., minimizing n_{gr}.
- **Objective (3):** our third objective is to minimize the weighted linear combination of latest completion times of green and red victims, i.e., minimizing $W_G \cdot e_G + W_R \cdot e_R$.

Table 2 gives the decision variables used in the mathematical formulation we devised to tackle the offline problem.

In the following, we present the offline model that solves our modified ARP:

$$\min \quad \sum_{v \in V} y_v \tag{1}$$

$$\min \quad \sum_{g \in G} z_g \tag{2}$$

$$\min \quad W_R \cdot e_R + W_G \cdot e_G \tag{3}$$

$$s.t.$$

$$\sum_{j \in V \cup H} x_{hj} \leq \left|A_h\right| \quad \forall h \in H \tag{4}$$

$$\sum_{j \in V \cup H} x_{ji} = \sum_{j \in V \cup H} x_{ij} \quad \forall i \in V \tag{5}$$

$$\sum_{j \in V \cup H} x_{ji} = 1 - y_i \quad \forall i \in V \tag{6}$$

$$\sum_{h \in H} u_{rh} = 1 - y_r \quad \forall r \in R \tag{7}$$

$$\sum_{h \in H} u_{gh} = z_g - y_g \quad \forall g \in G \tag{8}$$

$$\sum_{i \in V} u_{ih} \leq C_h \quad \forall h \in H \tag{9}$$

$$b_v + S_v + t_{vj} \leq b_j + M \cdot \left(1 - x_{vj}\right) \quad \forall v \in G \cup H; \forall j \in V \tag{10}$$

$$b_v + S_v + t_{vh} + S_h + t_{hj} \leq b_j + M \cdot \left(2 - x_{vj} - u_{vh}\right) \quad \forall v \in V; j \in V; h \in H \tag{11}$$

$$b_r - M \cdot y_r \leq T^r_{rb} \quad \forall r \in R \tag{12}$$

$$b_g - M \cdot y_g \leq T^g_{gb} \quad \forall g \in G \tag{13}$$

$$b_g - M \cdot z_g \leq T^g_{gr} \quad \forall g \in G \tag{14}$$

$$e_G \geq b_g + S_g - M \cdot u_{gh} \quad \forall g \in G; h \in H \tag{15}$$

$$e_R \geq b_r + S_r + u_{rh} \cdot \left(t_{rh} + S_h\right) \quad \forall r \in R; h \in H \tag{16}$$

$$e_R \geq b_g + S_g + u_{gh} \cdot \left(t_{gh} + S_h\right) - M \cdot (1 - z_g) \quad \forall g \in G; h \in H \tag{17}$$

$$x_{ij} \in \{0, 1\} \quad \forall (i, j) \in E \tag{18}$$

$$y_v \in \{0, 1\} \quad \forall v \in V \tag{19}$$

$$z_g \in \{0, 1\} \quad \forall g \in G \tag{20}$$

$$b_v \geq 0 \quad \forall v \in V \cup H \tag{21}$$

$$u_{vh} \in \{0, 1\} \quad \forall v \in V; h \in H \tag{22}$$

Here, (1), (2) and (3) define the hierarchical objective functions in which the first objective is to minimize the number of victims that turn to black, the second objective is to minimize the number of victims that turn from green to red and the third objective is to minimize the weighted time at which the last green and red victims are visited. We note that the constraints in the formulation are adopted from Talarico et al. (2015) and Rabbani et al. (2022). Constraints in (4) control the number of ambulances originating from each hospital. Constraint sets (5) and (6) guarantee the balance of flow and the connectivity of ambulance routes. They also ensure that all victims, except those coded as black, are visited by an ambulance. Constraints (7) and (8) ensure that victims in red code are allocated to a hospital. By (7), red victims that are not turned to black are allocated to a hospital and by (8), green victims that turn to red but not black are assigned to a hospital. Constraints (9) impose capacity limits on the hospitals. Constraints (10) and (11) compute the time progression for green and red victims. Constraints (12), (13) and (14) are to set the relations between variables and the threshold values. Constraints (15), (16) and (17) determine the final time at which a green and a red victim are served. In constraint (15), we consider only those green victims whose triage does not turn to red. In (16), we consider the red victims and in constraint (17) those green victims whose triage level turned to red. The remaining constraints are designed to define the domains of the variables. We remark that Talarico et al. (2015) and Letchford et al. (2006) observed that two-indexed formulations are more efficient than three-indexed versions for similar vehicle routing problems. This is why we opted to use the two-indexed formulation in our study.

3.2 The online problem

In the OARP, the triage level (i.e., sets R, G, B), the required treatment time of each victim (i.e., S_v for $v \in V$), and the time threshold(s) based on which the triage level of each victim deteriorates (i.e., $T^g_{gr}, T^r_{gb}, T^r_{rb}$ for $g \in G$ and $r \in R$) are not known in advance. For victim $v \in V$, the incomplete information about the triage level, treatment time, and thresholds is revealed only when victim v is examined by the medical staff of one of the ambulances. In this case, the revealed information is shared among all the ambulances and the victim may be serviced straight away by the same ambulance or later by another ambulance. By accounting for time-varying victim conditions, we recognize that the severity and medical needs of victims can change over time, especially in the immediate aftermath of a mass casualty event.

Furthermore, the travel time of the links are not known a priori in the OARP and can only be sensed in real time due to the dynamic nature of post-disaster scenarios. With respect to the standard literature of online routing problems with arc uncertainty (Akbari and Shiri 2021; Zhang et al. 2019; Büttner and Krumke 2016), we assume that the travel time of an arc $(i, j) \in E$ can be learned online by an ambulance only when the ambulance is at node i. Through the incorporation of real-time road conditions, each ambulance is equipped with access solely to the real-time travel times of the routes connecting its present location to other nodes within the

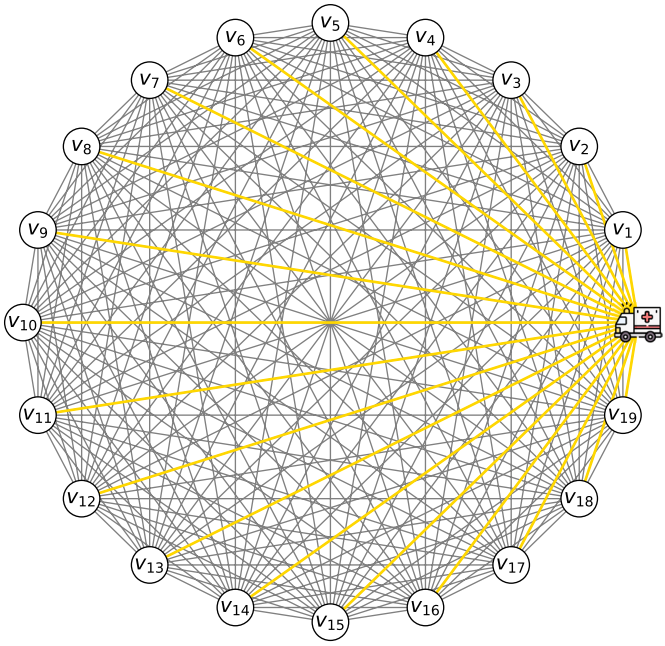


Fig. 1 In this complete graph with 20 nodes, the ambulance, positioned at the only hospital, has access to real-time travel times for $\frac{20}{2\times\binom{20}{2}-20} \approx 5\%$ of the routes. Arc directions have been omitted for the sake of clarity in the illustration

road network. This means that the real-time travel times for $2\binom{|V_0|}{2} - |V_0|$ routes between locations in V_0 remain hidden from each ambulance. In other words, to obtain precise real-time traffic information for a route connecting two locations, an ambulance must be positioned and prepared for dispatch at the starting location of the route. Consequently, a significant portion of the travel times in the road network is concealed from each ambulance in real time. This incomplete information hinders the use of state-of-the-art mathematical models, which are effective for solving vehicle routing problems, in the design of online algorithms for addressing the OARP. Figure 1 provides a schematic representation of real-time road information in our problem context. To maintain the realism of real-time road information in our model, we ensure that online algorithms are oblivious to travel times for already traversed arcs, as these times may be updated with real-time data. As a result, ambulances cannot gradually collect travel time information of the complete network.

We suppose that the time required for assessing the triage level of the victims is negligible compared to the time spent on travel and treatment. All other parameters of the OARP are known and identical to those of the ARP. The differences in the problem inputs of the ARP and the OARP are highlighted in the third and

fourth columns of Table 1, respectively. As the OARP generalizes the ARP by incorporating online uncertainty, the objective function of the OARP is the same as the objective function of the corresponding ARP that we defined in Sect. 3.1.

Remark 3.1 In the OARP, although certain parameters are not predetermined, the set of victims (i.e., set V) and their locations are known to the ambulances. These data can be gathered through technologies such as drones and satellites or via reports furnished by rescue teams and individuals. Similarly, information about blocked roads can be obtained using the same technologies (Daud et al. 2022; Akbari et al. 2021; Dukkanci et al. 2023), allowing the exclusion of blocked roads from the road network. Consequently, the underlying road network reduces to a complete graph containing the nodes in set $N = H \cup G \cup R \cup B$.

4 Theoretical competitive analysis

In academic literature, the formal investigation of online optimization problems is commonly known as *competitive analysis* (Ma et al. 2021; Akbari et al. 2021; Ma and Simchi-Levi 2020; Zhang et al. 2019; Jaillet and Wagner 2008). This approach involves comparing the performance of an *online solution*, which is derived with incomplete information, to that of the *offline solution*, which represents the optimal outcome with complete information. This comparison is made from a worst-case point of view, and it is quantified using a metric called the *competitive ratio*.

Online algorithms are typically categorized as either deterministic or randomized. In the case of a deterministic online algorithm, its output remains consistent when applied to the same online problem instance multiple times. Conversely, a randomized online algorithm may produce varying solutions for the same problem instance on different runs, leading to the consideration of expected objective function values in the case of randomized algorithms.

The concept of competitive ratio, initially proposed for single-objective optimization problems in Sleator and Tarjan (1985), has become a widely accepted performance criterion for the theoretical analysis of online solutions in the Operations Research community, as evident in works such as (Ma and Simchi-Levi 2020; Ma et al. 2021; Zhang et al. 2019; Akbari and Shiri 2021; Jaillet and Wagner 2008). This notion was later extended to a multi-objective setting in Tiedemann et al. (2015). For an online multi-objective minimization problem with z different objective functions, let Δ denote the global set of all problem instances and $\Theta = (\theta_1, \theta_2, ..., \theta_z)$ be a vector representing z objective functions. Formally, an online algorithm (ALG) is called Θ-competitive if the supremum of the ratio of the (expected) i^{th} objective function value of the online solution to the i^{th} objective function value of the offline optimal solution (OPT) for any instance $\delta \in \Delta$ is less than or equal to θ_i for $i \in \{1, 2, ..., z\}$, i.e., $1 \leq \sup_{\delta \in \Delta} \frac{\mathbb{E}(ALG(\delta))}{OPT(\delta)} \leq \theta_i$ for $i = 1, 2, ..., z$.

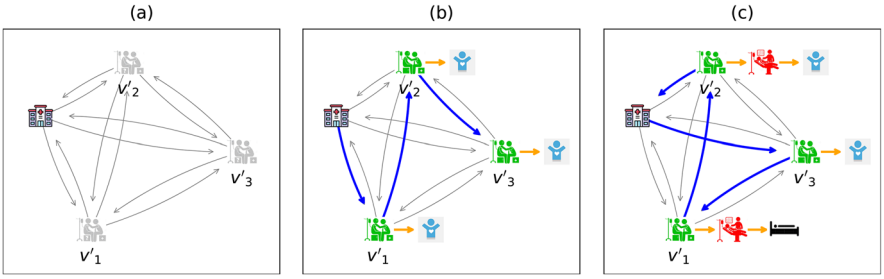


Fig. 2 A scenario corresponding to the proof of Lemma 4.1 on an instance with three victims. Part (**a**) illustrates that the triage of victims is unknown to online algorithms. Part (**b**) demonstrates the route of the offline optimal solution where the triage of all three victims remains green before they receive treatment. Part (**c**) depicts the route of an online algorithm in which the triage of one victim remains green before receiving treatment, the triage of one victim first turns to red and then to black, and the triage of one victim turns to red before receiving treatment

In the next lemma, we demonstrate that any online algorithm is arbitrarily powerless against the offline optimal solutions for our problem.

Lemma 4.1 *No online algorithm* (*ALG*) *achieves a finite competitive ratio for the OARP in the worst-case, i.e., ALG is* $\Theta -$ *competitive where* $\Theta = (\theta_1, \theta_2, \theta_3)$ *denotes the vector of competitive ratio such that none of the* θ_i *values is bounded for* $i = 1, 2, 3$.

Proof **Design of a worst-case instance.** We consider an instance of the problem with a single ambulance a_1 located at the only hospital h_1 and an initial scenario where all the victims ($v_i \in V$) are green. For $i \in \{1, 2, ..., |V|\}$, we fix the travel time from h_1 to v_i at $t_{(h_1,v_i)} = \epsilon$ and the travel time from v_i to h_1 at $t_{(v_i,h_1)} = 1$. Also, we set the travel time between v_i and v_j to $1 + \epsilon$ in both directions for $i, j = 1, 2, ..., |V|$, i.e., $t_{v_i v_j}$ is not known by the ambulance unless it arrives at node v_i or v_j. We suppose that the treatment times of all victims and the service times to drop the red victims to hospital h_1 are ϵ, i.e., $S_{v_1} = S_{v_2} = ... = S_{v_{|V|}} = S_{h_1} = \epsilon$. Note that the thresholds are such that if the treatment is not started before a time threshold for a victim, the triage level of that victim will be deteriorated the moment after the threshold is triggered. Given that the thresholds are online parameters, they are unknown to the ambulances from the beginning. Without loss of generality, we re-label the victims with respect to their thresholds by $v'_1, v'_2, ..., v'_{|V|}$, i.e., these labels are unknown to the ambulances initially. Figure 2a illustrates the representation of the problem inputs for the online algorithms, with only three victims in the road network. For v'_1, we assume that the thresholds are $T_{gr}^{v'_1} = T_{gb}^{v'_1} = \epsilon$. For v'_2, we assume that the thresholds are $T_{gr}^{v'_2} = 1 + 3\epsilon$ and $T_{gb}^{v'_2} = M$ where M is a large enough positive value. For v'_i and

$i = 3, 4, ..., |V|$, we assume that the thresholds are $T_{gr}^{v_i'} = T_{gb}^{v_i'} = i - 1 + (2i - 1) \times \epsilon$. Finally, we let $W_G = \epsilon$ and $W_R = 1$.[1]

Analysis of the offline optimum. In the offline optimum, the order of treatment of the victims is $v_1' \rightarrow v_2' \rightarrow ... \rightarrow v_{|V|}'$. Therefore, all the victims remain green code after the termination in the offline optimum, i.e., $n_{gr} = n_{rb} = 0$ in the offline optimum. Also, the weighted linear combination of the latest completion times of the green and red victims is $W_G \times (|V| - 1 + (2|V| - 1) \times \epsilon) = \epsilon \times (|V| - 1 + (2|V| - 1) \times \epsilon)$ in the offline optimum. Hence, the values of objective functions 1, 2, and 3 would be 0, 0, and $\epsilon \times (|V| - 1 + (2|V| - 1) \times \epsilon)$, respectively. Figure 2b depicts the offline optimal route where there are only three victims in the road network.

Analysis of an arbitrary online solution. We remark that an online algorithm has no prior information about the thresholds of the victims. We let v_1' and v_2' be uniformly distributed among the victims, i.e., it is equally likely that any of the $|V|$ victims be v_1' or v_2'. Observe that if v_1' is not visited by time ϵ, it will first turn from green to red and then it will turn from red to black since $T_{gr}^{v_1'} = T_{gb}^{v_1'} = \epsilon$. Also, if v_2' is not visited by time $1 + 3\epsilon$, it will turn from green to red since $T_{gr}^{v_2'} = 1 + 3\epsilon$ and $T_{gb}^{v_2'} = M$. In this case, n_{gr} and n_{gb} would be at least 2 and 1, respectively. Also, the latest service completion time of v_2' would be a positive value which is greater than $1 + 3\epsilon$, i.e., the weighted latest service completion time of green and red victims would be at least $W_R \times (1 + 3\epsilon)$. Figure 2c illustrates such a scenario where there are only three victims in the road network. Given the travel times between locations, the probability that v_1' and v_2' are not visited by times ϵ and $1 + 3\epsilon$ is at least $\frac{|V|-2}{|V|} \times \frac{|V|-3}{|V|-1}$. Hence, the expected objective function values of an arbitrary online algorithm against the discussed uniform probability distribution would be at least $\frac{|V|-2}{|V|} \times \frac{|V|-3}{|V|-1} \times 1$ for objective function 1, $\frac{|V|-2}{|V|} \times \frac{|V|-3}{|V|-1} \times 2$ for objective function 2, and $W_R \times \frac{|V|-2}{|V|} \times \frac{|V|-3}{|V|-1} \times 1 + 3\epsilon = \frac{|V|-2}{|V|} \times \frac{|V|-3}{|V|-1} \times 1 + 3\epsilon$ for objective function 3. Given the offline optimal objective function values, the lemma is derived by Yao's principle (Shiri et al. 2023; Andrew Chi-Chih 1977) when $|V|$ is sufficiently large and ϵ is small enough. □

Remark 4.2 While finite lower bounds on the competitive ratio of online algorithms have been recently proven for the reduced version of the OARP, which omits the time-varying victim conditions (Shiri et al. 2023), our analysis highlights the challenges of devising an efficient online algorithm for the OARP when considering time-varying victim conditions. Specifically, we show that online algorithms cannot achieve finite competitive ratios in such scenarios.

[1] We have deliberately chosen this specific setting for the worst-case instance since it imposes the infinite competitive ratio to the online algorithm in all the three objective functions 1, 2, and 3.

Lemma 4.1 is a strong theoretical statement on the worst-case performance of online algorithms, confirming that none of them achieves a finite competitive ratio for the OARP. However, it fails to discriminate between the average performance of different online algorithms on real-world instances. To address this concern and to effectively measure the computational performance of online algorithms on instances with real-world characteristics, the notion of *experimental competitive ratio* has been recently applied in the literature (Zhang et al. 2019; Akbari et al. 2021; Yao et al. 2022). For an online algorithm and a given set of problem instances, the experimental competitive ratio is the ratio of the average cost of the online algorithm over the average cost of the offline optimal solution on that given set of problem instances.

5 Online algorithms

In an online algorithm designed for tackling the OARP, there are three decisions that must be made for each ambulance dynamically (in an online manner) with respect to the revealed information about the condition of the victims. These decisions are: (1) the next victim node to be visited by the ambulance, (2) the next hospital a newly found red victim must be delivered to by the ambulance, and (3) if a newly found green victim by an ambulance must be serviced immediately or must be bypassed without providing service to prioritize servicing other potential victims with urgent conditions.

To understand and compare the effectiveness of different heuristic rules on the performance of online algorithms for the OARP, we first propose a general procedure, called the *Decision-Based* (DB) procedure, which we apply for constructing various online algorithms. The DB procedure is mainly based on a combination of heuristic rules which address the three aforementioned types of decisions, i.e., the procedure applies a heuristic rule for each of these decisions and constructs a solution.

For each of the aforementioned three decisions, we analyze the effect of two heuristic rules on the performance of the algorithm, i.e., we design and develop $2 \times 2 \times 2 = 8$ different versions of the DB algorithm (i.e., $DB_1, DB_2, ..., DB_8$) that follow the combinations of the heuristic rules. In the following, we first present the general structure of the DB procedure and then discuss the details of our heuristic rules. For that, we need to define some notations first.

- We call the heuristic rule, based on which the next victim to be visited by an ambulance is determined, the *NV* heuristic rule.
- We call the heuristic rule, based on which the next hospital for delivering a newly found red victim is determined, the *NH* heuristic rule.
- We call the heuristic rule, based on which an ambulance decides whether to service a newly found green victim immediately or bypass the victim for a later treatment, the *GS* heuristic rule.

5.1 The DB procedure

The inputs of the DB procedure are listed in the last column of Table 1. The DB procedure employs the unassigned victims list, denoted by set $V^{'}$, and initializes it by equating it to the complete list of all victims. Furthermore, this procedure establishes and initializes an empty list to monitor the green victims who have been observed but not yet served, denoted by set OUG. Initially, each ambulance is allocated to a victim based on an NV rule, and the list $V^{'}$ is subsequently updated. Afterward, the ambulances are sent toward their designated victims. Upon reaching their designated victims and ascertaining the victims' conditions, the ambulances communicate information regarding the triage level and necessary treatment time to other ambulances. Then, three cases may occur:

1. The victim's triage level is categorized as green. In this case, the ambulance decides whether to service or bypass the victim based on a GS rule. If the victim must be serviced, the ambulance provides the service. Otherwise, the green victim is added to the set OUG. Then, the next location to be visited by the ambulance is determined based on an NV rule and the list $V^{'}$ is updated.
2. The victim's triage level is categorized as red. Consequently, the ambulance administers the necessary treatment to the victim. Subsequently, the ambulance follows an NH rule to determine its destination hospital. This information is then relayed to the other ambulances, and the capacity of the chosen hospital is decremented by one.
3. The victim's triage level is categorized as black. In this case, the ambulance must bypass the victim without providing service and shares this information with non-medical teams, i.e., the black victim will be serviced by non-medical teams due to limited medical staff and resources. The next location to be visited by the ambulance is determined based on an NV rule and the list $V^{'}$ is updated.

After transporting a victim to a hospital, an ambulance is assigned to a new victim location using an NV rule and the list $V^{'}$ is updated. The pseudo-code of the main steps of the DB procedure is presented in Appendix A.

5.2 Details of the proposed heuristic rules

In this paper, we investigate six heuristic rules. Namely, two heuristics rules NV_1 and NV_2 for assigning a victim node to an available ambulance, two heuristic rules NH_1 and NH_2 for assigning a hospital to an ambulance carrying a red victim, and two heuristic rules GS_1 and GS_2 to decide whether to service or bypass a green victim. Before we explain the rules, we remark some important aspects of the online problem definition in the following. As discussed in Sect. 3.2, the arc travel times are unknown initially and can only be revealed at the starting node of an arc with respect to the real-time information. In addition, we do not have any prior information about the triage levels, treatment times, and thresholds. Given these, we propose the heuristic rules embedded in different versions of our DB procedure as follows.

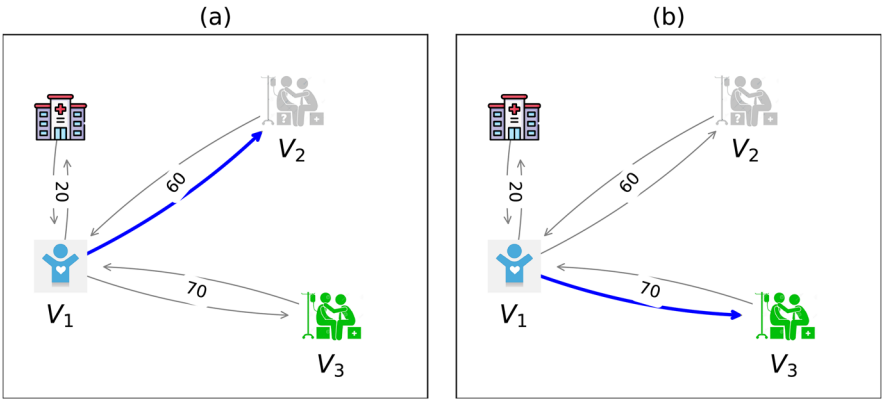


Fig. 3 An illustration of the NV rules where $v_3 \in OUG$ and the ambulance is located at v_1 who has recently received treatment. Part (**a**) represents the route chosen by NV_1. Part (**b**) demonstrates the route chosen by NV_2

5.2.1 The NV rules

NV_1. The ambulance follows a greedy policy and chooses the closest victim from the set of unassigned victims (V').

NV_2. This rule is sensitive about the observed information in the sense that before making a decision, the ambulance checks if there is any victim in the set of observed and unserviced green victims (OUG) who must be serviced immediately due to approaching their thresholds for triage level changes. To quantify this for a green victim $g \in OUG$, we use the following parameters in this rule: (1) T^g_{gr}, (2) current time CLK, (3) average revealed travel time between the current node of the ambulance and other locations t_{AVG}, and (4) average revealed service time of locations (i.e., victims and hospitals). Using t_{AVG} and S_{AVG} we construct an approximate measure called $AVG = 3 \times t_{AVG} + 2 \times S_{AVG}$ for the required time for servicing another (red) victim and arriving back at g, i.e., AVG is just an approximate measure of time for quantifying the emergency of visiting g. In the NV_2 rule, if $T^g_{gr} - CLK \leq AVG$, the treatment of g is regarded as *urgent*. If such a victim $g \in OUG$ with $T^g_{gr} - CLK \leq AVG$ exists, the ambulance chooses a victim whose $T^g_{gr} - CLK$ has the least value. Otherwise, by considering the revealed travel times, the ambulance chooses the closest victim node in the set of victims who belongs to $V' \setminus OUG$. This feature ensures that an ambulance will not be assigned to already observed green victims such that servicing them is not urgent. In case $V' \setminus OUG = \emptyset$, the ambulance chooses the closest victim in OUG. We note that due to ethical considerations, we utilize T^g_{gr} rather than T^g_{gb} as a criteria for decision making in order to avoid deterioration in the condition of an already observed green victim.

An example for NV **rules.** To distinguish between NV_1 and NV_2, consider a scenario where the set of remaining victims is $\{v_2, v_3\}$, $V' = \{v_2\}$ and $OUG = \{v_3\}$ such that the observed information about v_3 are $S_{v_3} = 10$, $T^{v_3}_{gr} = 200$, and $T^{v_3}_{gb} = 400$. In this

scenario, we focus on one of the ambulances that has completed servicing a green victim at v_1 with observed average real-time travel time between v_1 and other nodes being $t_{AVG} = 50$, and observed average service time of nodes being $S_{AVG} = 20$ at time $CLK = 100$. Suppose that the travel time between v_1 and v_2 is $t_{v_1v_2} = 60$ and the travel time between v_1 and v_3 is $t_{v_1v_3} = 70$. In this case, NV_1 chooses v_2 as the next victim to be visited since it is closer in terms of travel time. However, NV_2 chooses v_3, because $v_3 \in OUG$ and $T_{gr}^{v_3} - CLK = 200 - 100 = 100$ is less than $AVG = 3 \times t_{AVG} + 2 \times S_{AVG} = 190$. This example has been depicted in Fig. 3.

5.2.2 The NH rules

NH_1. The ambulance employs a greedy policy, selecting a hospital with available capacity that minimizes the sum of its travel time from the victim's location and the service time required to drop off the red victim at that hospital.

NH_2. The ambulance selects a hospital using a *hospital utility score* (HUS) as follows. To calculate the utility of each hospital ($h \in H$) with remaining capacity of more than one, considering an ambulance at the location of a red victim r, we take into account (1) the travel time from the victim's location to hospital h (i.e., t_{rh} is known because the ambulance is at r), and (2) the service time required to deliver the red victim to that hospital (S_h), (3) the fraction of the remaining capacity of that hospital ($C_h^{'} > 0$) over the maximum remaining capacity of hospitals which yet have available capacities (C_{max}), which can be denoted as $0 \leq \Pi_h = \frac{C_h^{'}}{C_{max}} \leq 1$. Using these parameters, we compute the value of HUS as $HUS_h = \frac{1}{1+\Pi_h} \times (t_{rh} + S_h)$. In this equation, given the range for Π_h, we have $\frac{1}{2} \leq \frac{1}{1+\Pi_h} \leq 1$. Here, the term $\frac{1}{1+\Pi_h}$ is a tuning coefficient that incorporates the reverse effect of the residual capacities of the hospitals in the calculations of HUS. In here, based on our computational experiments, we have deliberately set the lower bound value for $\frac{1}{1+\Pi_h}$ equal to $\frac{1}{2}$ to tune and control the effect of the residual capacities in calculations of HUS. After calculating the HUS values, the ambulance opts for the hospital with the lowest HUS score. It is important to note that when developing the HUS metric, we take into consideration both time-related aspects (travel and service times) and the capacity status of available hospitals. This approach allows us to provide swift service to red victims while reserving hospitals with lower remaining capacities for potential red victims who may need assistance later and could be closer to these hospitals.

An example for *NH* rules. Consider a scenario where the ambulance has serviced a red victim at v_r and must decide to transport the victim to one of the two remaining hospitals h_1 and h_2 such that the travel time between v_r and h_1 is $t_{v_rh_1} = 100$, the travel time between r and h_2 is $t_{v_rh_2} = 150$, and the hospital service time for both hospitals is $S_{h_1} = S_{h_2} = 50$. Suppose that the residual capacity of h_1 is $C_{h_1}^{'} = 1$ and the residual capacity of h_2 is $C_{h_2}^{'} = 4$. In this case, NH_1 chooses h_1 since the service completion

time for v_r can be done earlier at this hospital. However, NH_2 chooses h_2 since $HUS_{h_2} = \frac{1}{1+1}(150+50) = 100$ is less than $HUS_{h_1} = \frac{1}{1+\frac{1}{4}}(100+50) = 120$.

5.2.3 The GS rules

GS_1. The ambulance follows a greedy policy with the aim of providing service for every newly visited victim as soon as possible. In this scenario, the ambulance provides immediate service to any green victim upon reaching them.

GS_2. The ambulance decides whether to service or bypass the victim according to a *green utility score* (GUS) as follows. It is important to note that when an ambulance bypasses a green victim, that victim will be included in the group of observed and unserviced victims (OUG). Later on, either the same ambulance or a different one may provide treatment to these victims (as indicated by NV_1 and NV_2). For an ambulance positioned at the location of a newly discovered green victim $g \in G$, the GUS is calculated in real time, taking into account: (1) the time threshold by which the triage level of the green victim turns to red (T^g_{gr}), (2) the time threshold by which the triage level of the green victim turns to black (T^g_{gb}), (3) the average revealed travel time between g and other nodes t_{AVG}, (4) the average service time of already observed nodes (S_{AVG}), (5) the revealed treatment time of the green victim S_g, (6) the weight of the green victims W_G, (7) the weight of the red victims W_R, (8) the proportion of known red victims to all victims with known triage denoted by $0 \leq \Pi_R \leq 1$, where $\Pi_R = \dfrac{|R^1|}{|R^1 \cup G^1|}$ at time 0, and (9) the current time CLK. Given the values above, the ambulance first computes

$$\alpha = \frac{T^g_{gb} - CLK}{3 \times t_{AVG} + 2 \times S_{AVG}} = \frac{T^g_{gb} - CLK}{AVG} \tag{23}$$

and

$$\beta = \frac{T^g_{gr} - CLK}{3 \times t_{AVG} + 2 \times S_{AVG}} = \frac{T^g_{gr} - CLK}{AVG}. \tag{24}$$

In these equations, α and β are measures which correspond to the first and second objective functions of the OARP. These measures compare the remaining time until triggering thresholds for the green victim (i.e., $T^g_{gb} - CLK$) with the average time for servicing a potential red victim and returning back to the node of the bypassed green victim (i.e., $AVG = 3 \times t_{AVG} + 2 \times S_{AVG}$). Here, we clarify that this rule does not mean that in our algorithms a bypassed green victim must be necessarily serviced by the same ambulance.

In addition to α and β, the ambulance computes

$$\xi = \frac{1}{1+\Pi_R} \times \frac{W_G}{W_R} \times \frac{t_{AVG}}{S_g}, \tag{25}$$

such that ξ is a measure for the third objective function, where $\frac{t_{AVG}}{S_g}$ enforces higher GUS values for green victims with relatively small treatment times, i.e., if the required time for treatment of the green victim (g) is relatively higher than the average observed travel time between g and other nodes (t_{AVG}), then the GUS would be smaller for g. Also, the term $0 \leq \frac{W_G}{W_R} \leq 1$ in the equation above is a tuning coefficient that takes into account the weights of red and green victims for GUS calculation such that the GUS value for the green victims would decrease where $\frac{W_G}{W_R}$ decreases, i.e., the probability of providing immediate service to green victims during their initial encounters diminishes as the weight of red victims substantially outweighs that of green victims. Additionally, there is another tuning coefficient denoted as $\frac{1}{2} \leq \frac{1}{1+\Pi_R} \leq 1$, which dynamically considers the current proportion of known red victims as a decision factor for the algorithm. If Π_R is high, indicating a significant number of known red victims, the GUS decreases. Consequently, the likelihood of bypassing a green victim increases. In simpler terms, when ambulances observe a considerable number of red victims during their operations, the algorithm is more likely to bypass green victims. Conversely, if few or no red victims have been observed, the algorithm does not permit bypassing green victims. We remark that, based on our computational experiments, we have deliberately set the lower bound value for this coefficient equal to $\frac{1}{2}$ to tune and control the effect of this coefficient.

Once all α, β, and ξ have been computed, the GS_2 rule sorts the objective functions of the OARP hierarchically in terms of their importance as follows. If $\alpha \leq \epsilon_1$, the value of the GUS is set to 1. Otherwise, β determines the value for GUS as follows. If $\beta \leq \epsilon_2$, the value of GUS is set to 1. Otherwise, the value of the GUS is set to ξ. Once GUS is calculated, the ambulance bypasses the victim if $GUS < 1$, and services the victim otherwise. We note that in our computational experiments, we have set the values of ϵ_1 and ϵ_2 equal to 5 and 3, respectively. By considering these values, we ensure that a green victim is serviced if their thresholds are relatively close to the current time.

An example for *GS* rules. To differentiate between GS_1 and GS_2, consider a newly visited green victim g where the time threshold by which the triage of g turns to red is $T^g_{gr} = 250$, the time threshold by which triage of g turns to black is $T^g_{gb} = 500$, the average observed travel time between g and other locations is $t_{AVG} = 10$, the average service time of already observed nodes is $S_{AVG} = 10$, the revealed treatment time of the green victim is $S_g = 10$, the weight of green victims is $W_G = 1$, the weight of red victims is $W_R = 5$, the proportion of known red victims to all victims with known triage is $\Pi_R = 1$, and the current time is $CLK = 50$. In this

Table 3 Different versions of the DB algorithm

Algorithm	Heuristic Rules
DB_1	NV_1, NH_1, GS_1
DB_2	NV_1, NH_2, GS_1
DB_3	NV_1, NH_1, GS_2
DB_4	NV_2, NH_1, GS_1
DB_5	NV_2, NH_1, GS_2
DB_6	NV_2, NH_2, GS_1
DB_7	NV_1, NH_2, GS_2
DB_8	NV_2, NH_2, GS_2

Table 4 Characteristics of the data sets

	Number of instances	$\|V\|$	$\|\bar{R}\|$	$\|\bar{G}\|$	$\|\bar{A}\|$	$\|\bar{H}\|$	Average overall hospital capacity
First data set	20	10	5.8	4.2	2.9	2.5	10.0
	20	25	13.9	11.1	8.9	2.3	25.9
	20	50	23.4	26.6	13.1	2.5	41.4
Second data set	55	100	50.6	49.4	29.1	5.4	90.0

case, GS_1 services g immediately. However, GS_2 calculates $\alpha = \frac{T^g_{gb} - CLK}{3\times t_{AVG} + 2\times S_{AVG}} = \frac{500-50}{3\times 10+2\times 10} = \frac{450}{50} = 9$, $\beta = \frac{T^g_{gr} - CLK}{3\times t_{AVG} + 2\times S_{AVG}} = \frac{250-50}{3\times 10+2\times 10} = \frac{200}{50} = 4$, and $\xi = \frac{1}{1+\Pi_R} \times \frac{W_G}{W_R} \times \frac{t_{AVG}}{S_g} = \frac{1}{1+1} \times \frac{1}{5} \times \frac{10}{10} = 0.1$. In this case, since $\alpha > 5$ and $\beta > 3$, the value of GUS is determined by $GUS = \xi = 0.1$. Since $GUS < 1$, the GS_2 rule bypasses the victim.

By incorporating different combinations of these rules into the DB procedure described in Sect. 5.1, we construct eight different online algorithms, i.e., DB_1, DB_2 ,..., DB_8. These algorithms are formally presented in Table 3.

6 Experimental competitive analysis

In this section, we evaluate the performance of all eight versions of the DB algorithm. The experiments were implemented in Python 3.9 and conducted on a computer equipped with an Intel Core i5 processor, 32 GB of RAM, and a 64-bit Windows 10 operating system. First, we provide details about the data sets employed in our computational experiments, followed by the presentation of the experimental results.

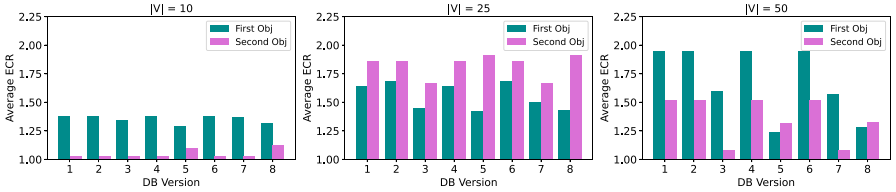


Fig. 4 Results of the first and second objectives on the first data set

6.1 Data set descriptions

In order to computationally analyze our algorithms, we tested them on two data sets that are both based on the instances introduced in Talarico et al. (2015). In this study, instances with 10, 25 and 50 victims with varying hospital capacity and number of ambulances were generated. Given that in their instances, only red victims had to be transferred to hospitals, and in some of the instances, the cumulative hospital capacities are only large enough to accommodate red victims. In our problem, however, if a green victim g is served at a time between T^g_{gr} and T^g_{gb}, they have to be transferred to a hospital. Thus, for the first data set, we selected instances that have higher capacities. In particular, we selected 20 instances from each of the cases with 10, 25 and 50 victims. In order to analyze the performance of our algorithms on larger instances, we have also tested them on instances with 100 victims that are generated based on Talarico et al. (2015) instances with 50 victims. For that, we selected ten instances from the set of instances with 50 victims and generated 55 cases with 100 victims. Each of these cases is generated from accumulation of two instances with 50 victims. Table 4 gives the characteristics of the data sets used in our computational experiments. In this table, $|V|$ gives the total number of victims, $|\bar{R}|$ and $|\bar{G}|$ give the average number of red and green victims in those instances and $|\bar{A}|$ and $|\bar{H}|$ gives the average number of ambulances and the average number of hospitals, respectively.

In the context of online optimization, uncertainty is introduced to the online algorithm by concealing a portion of the input related to uncertain parameters, as highlighted in previous works (Zhang et al. 2019; Akbari and Shiri 2021; Jaillet and Wagner 2008; Ausiello et al. 2001). To incorporate this uncertainty into our model, we have chosen to withhold information regarding triage levels, treatment times, travel times, and thresholds from the online algorithms. This unknown information is partially observable to the online algorithms in real time. For our instances, which are derived from Talarico et al. (2015), travel times are generated as Euclidean distances between nodes. It is worth noting that this method for generating travel times aligns with the standard benchmark instances in the literature, rather than being a decision made independently. In relation to the thresholds, we draw upon existing literature, which suggests that prompt treatment within a few hours in mass casualty incidents can significantly reduce casualties (Farahani et al. 2020) Additionally,

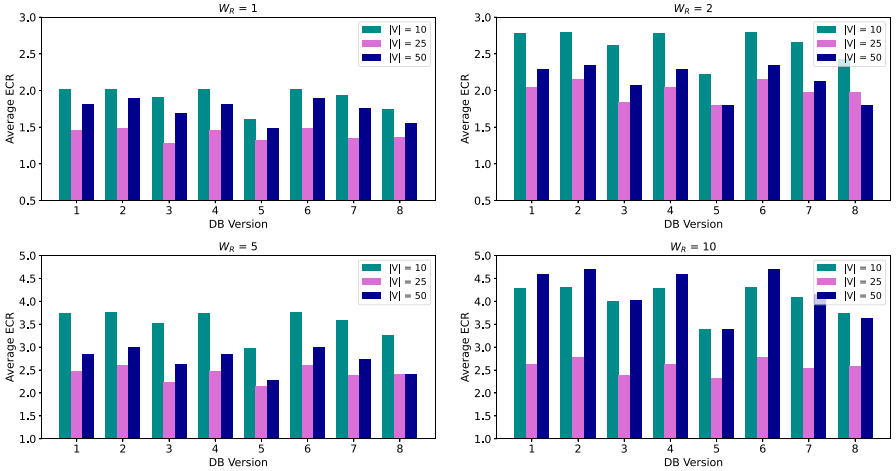


Fig. 5 Results of the third objective function on the first data set

studies on earthquakes in China highlighted in Brown et al. (2017) have indicated that unless aid is provided within 2 to 6 h, fewer than half of the victims are likely to survive. Furthermore, Liang et al. (2001) has pointed out that timely medical attention within the first 6 h following the Armenia earthquake could have prevented many deaths. In accordance with the findings in this literature, we selected threshold values of less than 6 h. For the sake of general applicability, we opted for different threshold values for transitioning from red to black for initially red victims and for the time taken to transition from red to black for victims initially classified as green. Specifically, in Sect. 6.2, we set the thresholds to specific values: $T^g_{gr} = 180$ minutes, $T^g_{gb} = 360$ minutes, and $T^r_{rb} = 240$ minutes. In Sect. 6.3, to conduct sensitivity analyses on the performance of our algorithms, we relaxed these thresholds by assigning sufficiently large values.

6.2 Experimental analysis of the algorithms

We first provide the results of testing the eight variations of the DB algorithm on the instances from the first data set and then present the results of this algorithm on the larger instances from the second data set.

Figures 4 and 5 give the results of the DB algorithm on the first data set. Figure 4 presents the results of the obtained experimental competitive ratio values over the first and second objectives, and Fig. 5 presents the results of the third objective function with varying W_R values. Given that changing the value of W_R does not impact the first and second objective functions, we separated the results corresponding to the objectives to these two figures.

Figure 4 presents the average experimental competitive ratio (ECR) values over the 20 selected instances with 10, 25 and 50 victims. The first objective minimizes

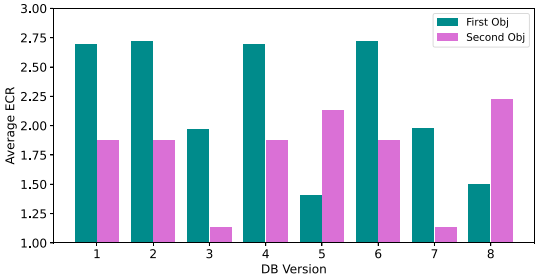


Fig. 6 Results of the first and second objectives on the second data set

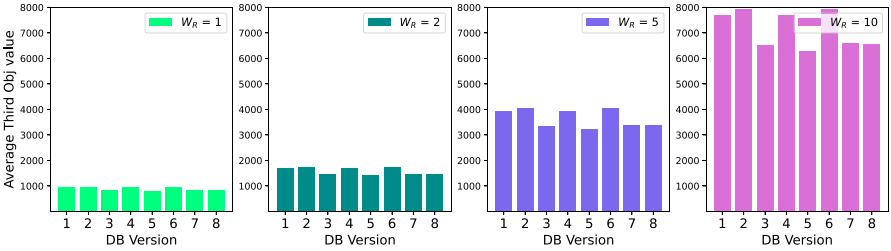


Fig. 7 Results of the third objective function on the second data set

the number of victims that turn to black code and the second objective minimizes the number of victims that turn from green to red code. For a green victim that first turns to red code and then turns to black code, their triage level change will be counted once in the second objective and once in the first objective. As can be observed in these results, in general, all the variations of the DB algorithm are able to find solutions for which the average ECR remains under 2. For instance, the worst average ECR of the first objective over all the instances belongs to variations 2 and 6 of the DB algorithm and they are both at 1.72. The same variations of the DB algorithm also have the worst average ECR over the second objective, which is at 1.42. On the other hand, variation 5 is able to find the best average ECR over the first objective, which is at 1.27 and variation 7 is able to find the best average ECR over the second objective with a value of 1.13. It is important to note that since the objectives are hierarchical, the first objective has higher priority and variation 5 of the DB algorithm outperforms the rest of the variations in this sense. Looking at the average ECR of the first and second objectives, again, variation 5 outperforms the rest with an average of 1.29.

Similar to the experimental studies conducted in Talarico et al. (2015), we kept the value of $W_G = 1$ in all the cases and only changed W_R to investigate its impact on the algorithms and the solutions. Since the first and second objectives do not depend on the value of W_R, changing it does not impact the solutions obtained for the first and second objectives. However, this can impact the third objective function which minimizes the weighted time at which the last green and red victims are serviced.

Figure 5 gives the results of the DB algorithm over the third objective function when W_R is set to 1, 2, 5 and 10. When $W_R = 1$, for the cases with 10 and 50 victims, variation 5 of the DB algorithm has the best performance with average ECRs of 1.61 and 1.49, respectively. For the case with 25 victims, variation 3 has the best performance with an average ECR of 1.29 which is followed by variation 5 with average ECR of 1.32.

For rest of the cases when W_R increases from 2 to 5 and then to 10, the performance of the algorithms are negatively impacted and solutions tend to produce higher ECR values. An important observation in these cases is that variation 5 of the DB algorithm consistently outperforms the rest of the algorithms over all the instances with 10, 25 and 50 victims and $W_R = 2$, 5 and 10. This confirms that over these instances, the combination of strategies NV_2, NH_1 and GS_2 that form the fifth version of the DB algorithm outperforms the rest of the variations. This confirms the importance of considering learning-based methods to select the next victim's location and bypassing some green victims while justifying the use of a simple greedy based strategy for selection of hospitals to deliver the victims.

The results of testing the eight variations of the DB algorithm on the second data set is presented in Figs. 6 and 7. Similar to the case with the first data set, we present the results for the first and second objectives and the third objective separately. Figure 6 gives the average ECR values of the first and second objectives for the eight variations of the DB algorithm. As can be observed, even with these larger instances, some of the variations of the DB algorithm are able to find reasonably good solutions within a very short time. For example, over these 55 instances with 100 victims, DB_5 has an average ECR of 1.41 over the first objective function and DB_3 and DB_7 have an average ECR of 1.13 over the second objective function. For DB_5, while the first objective is lower on average, the second objective is a bit higher compared to some of the other variations. This is because this variation aims to prevent death by minimizing the first objective while sacrificing from the second objective that corresponds to triage level change from green to red. Given the utmost importance of preventing death, we can conclude that DB_5 still outperforms the rest of the variations of the DB algorithm.

Figure 7 presents the average objective function values obtained from testing the variations of the DB algorithm on the instances from the second data set. We used the mathematical model presented in Sect. 3.1 to solve the offline problem for the instances from the first and second data sets. We set a time limit of 3 h (10800 s) to solve these problem instances. However, our model was only able to find the solutions of the first and second objectives, and in most of the cases, the problem remained unsolved for the third objective function and only loose upper bounds were found. We note that since the focus of our study is not providing an efficient solution method to solve the offline problem, we relied on the solutions obtained from our mathematical model to verify and test the performance of our online algorithms.

Using Fig. 7, we can compare the performance of the variants of the DB algorithm. With $W_R = 1$, on average, DB_5 found the best solutions and DB_2 and DB_6 had the worst performance. DB_5 had the best average performance over the instances

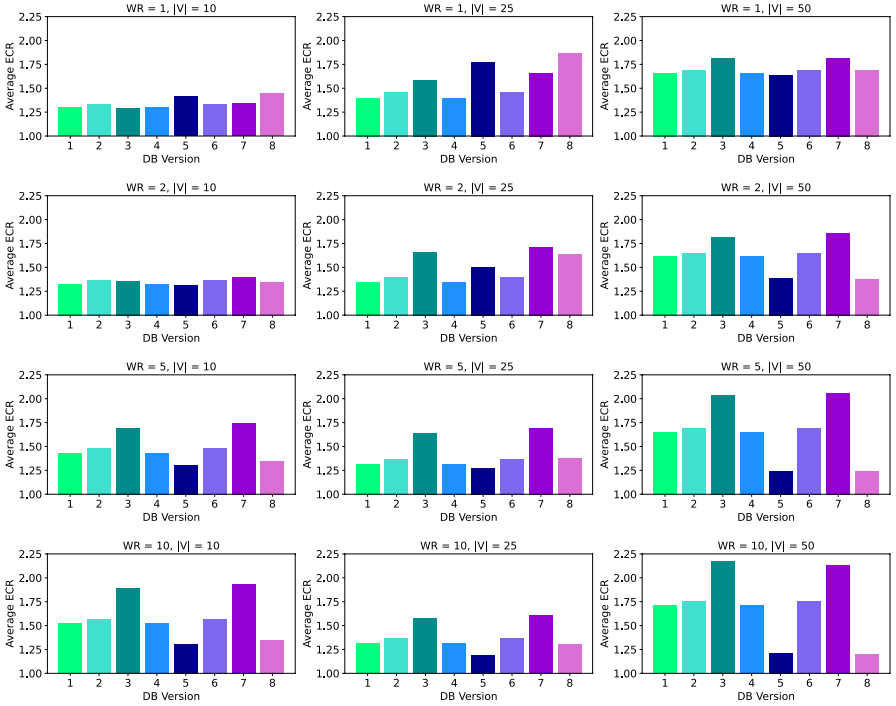


Fig. 8 Sensitivity analysis of threshold values on the first data set

with $W_R = 2$, 5 and 10 as well. Similar to the cases with the first data set, we again observe that DB_5, for which NV_2, NH_1 and GS_2 strategies are used, outperform the rest of the combinations. Meanwhile, DB_2 which uses NV_1, NH_2 and GS_1 consistently shows the worst performance. DB_2 and DB_5 are based on opposite strategies in the three main decisions pointed out in Sect. 5 and that is probably the reason why one of them is the best and one of them is the worst variation of the DB algorithm.

6.3 Sensitivity analysis on the threshold values

We analyze the performance of our algorithm on the same set of instances but when considering large threshold values such that the condition of none of the victims gets worse. In other words, we investigate the case when all T^g_{gr}, T^g_{gb} and T^r_{rb} values are large. When this is considered, the optimal objective function values of the first and second objectives are equal to 0 i.e., $\sum_{v \in V} y_v = 0$ and $\sum_{g \in G} z_g = 0$. In this version, only the third objective will be activated and the problem reduces to the version introduced in Talarico et al. (2015). In the following, we first present the results on the first data set and then analyze the performance of our algorithms on larger instances given as the second data set.

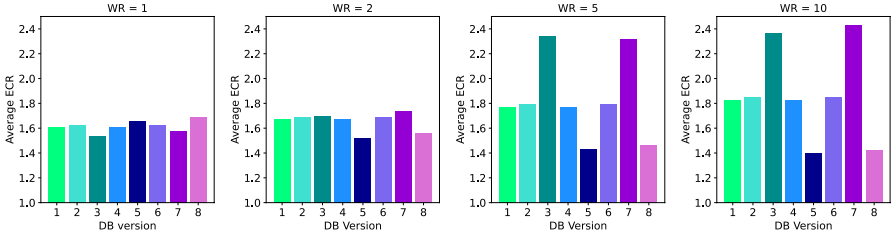


Fig. 9 Sensitivity analysis of threshold values on the second data set

As stated above, for this special case of the problem, we only represent the results from the third objective function as the optimal solutions of the first and second objectives are 0 in all the cases. Similarly, the solutions obtained from variations of the DB algorithm for the first and second data sets are also equal to 0 as none of the victims will have a change in the triage level. Figure 8 represents the results of the variations of the DB algorithm on the instances of the first data set for 10, 25 and 50 nodes and varying W_R values in the range 1, 2, 5 and 10. As can be observed in this figure, the average performance of the DB algorithm is acceptable and its best average performance over all variations in the worst case remains under 1.63 (DB_5 when $W_R = 1$ and $|V| = 50$).

For the case when $W_R = 1$, DB_1 and DB_4 consistently obtain the best average ECR values over the instances with 10, 25 and 50 victims. DB_1 corresponds to NV_1, NH_1 and GS_1 and DB_4 corresponds to NV_2, NH_1 and GS_1. This shows that when there are no threshold values after which the victim's condition gets worse, when the weight of the green and red victims are the same (i.e., $W_R = W_G$), greedy decisions can perform very well. When W_R increases to 2, when the average ECR over the three cases with 10, 25 and 50 victims is considered, DB_5 outperforms the rest of the DB variations. However, the average performance of DB_5 is not the best for the instances with 25 victims and in those instances, DB_1 and DB_4 show the best average performance.

For rest of the cases when $W_R = 5$ and $W_R = 10$, DB_5 consistently outperforms the rest of the variations. Over these cases, while DB_5 is based on NV_2, NH_1 and GS_2 decisions, DB_8 in which all the decisions are made using comparative learning-based methods (NV_2, NH_2, GS_2), is consistently the second best variation of the DB algorithm. While it might be expected that the performance of DB_5 deteriorates when more victims are considered, we see that this is in fact not correct and the number of victims does not seem to have a negative impact on the performance of DB_5. On the other hand, the performances of DB_3 and DB_7 are negatively impacted by increasing the number of victims or increasing the value of W_R. Given that DB_3 is based on the NV_1, NH_1 and GS_2 rules and DB_7 is based on NV_1, NH_2 and GS_2 rules, we see the importance of considering NV_2 in designing online algorithms as both DB_5 and DB_8 are based on NV_2.

Finally, the results of testing the DB algorithm on second data set instances when sufficiently large T^g_{gr}, T^g_{gb} and T^r_{rb} values are considered is presented in Fig. 9. Overall, it can be observed that even with these larger instances with 100 victims, we can obtain reliable and good quality solutions using the DB algorithm. In this figure, it is observed

that when no difference is put between red and green victims ($W_R = 1$), the performance of DB_3 is the best among the variants of the DB algorithm with an ECR of 1.53. For the rest of the cases, however, DB_5 consistently outperforms the rest of the variations and DB_8 is the second best version of the DB algorithm. When we compare Figs. 8 and 9, we can see that when the number of victims increases, the performance of the DB_5 algorithm remains at a very good level. Especially when W_R increases and more importance is given to the red victims, the performances of DB_5 and DB_8 are considerably better than rest of the variations. This testifies the importance of a decision-making mechanism that incorporates known information. Similar to the cases in Fig. 8, here we again observe that DB_3 and DB_7 are negatively impacted when W_R increases.

6.4 Analysis of the running times of the algorithms

All the variants of the DB algorithm were able to find their solutions in **less than 1 s** in all of the tested instances including the largest instances. As a result, we did not report the run time of our online algorithms. This highlights that our algorithms are appropriate for the emergency post-disaster scenarios where decisions should be made quickly. We note that a very high computational running time (e.g., more than 3 h) is required to produce optimal solutions for the offline optimization problem (ARP) in the instances with 50 or 100 victims. Therefore, utilizing the offline optimization approach is not suitable in emergency post-disaster scenarios due to the uncertainty in problem inputs and time limitations.

7 Conclusions and future research directions

In this study, we explored an online optimization problem, referred to as the OARP, with the aim of determining efficient routes and schedules for ambulances in post-disaster scenarios, where the goal is to provide timely assistance to a large number of casualties in limited time. In particular, we considered victims with time-varying conditions and different triage levels, i.e., green, red, black. The unknown parameters in our problem that are revealed over time are: travel times between locations, triage levels and treatment times of the victims, and time thresholds by which the condition of each victim worsens. We analyzed this problem by considering three hierarchical objective functions such that the first objective is to minimize the number of victims who pass away, the second objective is to minimize the number of victims whose condition (i.e., triage level) goes from green to red, and the third objective is to minimize the total weighted service completion time of the red and green victims.

We first studied this problem from a theoretical worst-case competitive ratio perspective and proved that a finite lower bound on the competitive ratio of online solutions does not exist for this problem. This confirms how challenging the problem is under incomplete information and a worst-case scenario. Furthermore, to address

real-life instances of the problem, we proposed eight different online heuristics all of which are capable to provide solutions in very low running times (i.e., less than 1 s). Our algorithms are compared against each other and the offline solutions on instances from the literature as well as newly generated large instances. One of our algorithms (i.e., DB 5) appears to outperform the other competitors in the majority of the tested scenarios. This algorithm intelligently incorporates piece-by-piece observed information into its decision-making process by employing the following rules:

- NV_2 rule which prioritizes victims whose conditions are known to be deteriorating imminently.
- NH_1 rule which focuses on delivering victims to the closest hospitals quickly and servicing remaining victims promptly to reduce the number of victims whose conditions worsen. Note that this rule does not have considerations of the real-time residual capacities of the hospitals.
- GS_2 rule which takes into account the thresholds of observed green victims. It allows the algorithm to make informed decisions about whether to bypass green victims if their condition is expected to remain stable for an extended period. By doing so, the algorithm ensures that it can allocate resources to service potentially red victims in urgent situations while minimizing unnecessary stops for green victims whose conditions are unlikely to deteriorate in the short term.

It is remarkable that DB5, which follows a greedy rule for hospital selection, outperforms the DB8 algorithm, which takes into account hospital residual capacities. A possible explanation for this phenomenon is the presence of thresholds. Specifically, delivering victims to nearby hospitals allows ambulances to serve more victims before their conditions deteriorate to the point of triggering thresholds. This approach is particularly beneficial in serving more green victims before their conditions change to red, as well as more red victims before they turn black. On average, the DB5 algorithm finds quality solutions against the offline solutions as well. Hence, it can be considered to be applied in the aftermath of disasters in real-world mass casualty incidents.

A future research direction is collecting real-life data to further investigate the performance of the algorithms on real-life scenarios. From a theoretical competitive analysis point of view, specifying real-world limitations for the worst-case instance to achieve finite competitive ratios would be a possible way for deepening the theoretical analysis.

Appendix A. Pseudo-code of the DB algorithm

The pseudo-code of the main steps of the DB procedure is presented in Algorithm 1.

Algorithm 1 DB Procedure

DB Procedure

1: **Input:** an OARP instance, NV rule, NH rule, GS rule
2: **Initiation:**
a: $V^{'} = V$ ▷ set of unassigned victims
b: $OUG = \emptyset$ ▷ OUG : set of Observed and Untreated Green victims
3: **Main body:**
4: **for** $a_\lambda \in A$: ▷ A is set of ambulances
5: choose $v_* \in V^{'}$ as the next location of a_λ based on NV rule and set $V^{'} = V^{'} \setminus v_*$
6: **dispatch** the ambulances in set A to their next allocated locations
7: **if** an ambulance $a_\lambda \in A$ reaches to its next location (v_λ) **then:**
8: **if** v_λ is a green victim **then:**
9: observe S_{v_λ} and the travel times of adjacent arcs to v_λ
10: decide whether to service or bypass based on GS rule
11: **if** the service must be provided, **then:**
12: service v_λ ▷ service the green victim
13: set $OUG = OUG \setminus v_\lambda$ ▷ if the green victim belongs to OUG, remove it from OUG
14: **else:**
15: add v_λ to OUG ▷ share information and add the green victim to OUG if $v_\lambda \notin OUG$
16: **if** $V^{'} \neq \emptyset$, **then:**
17: choose $v_* \in V^{'}$ as the next location of a_λ based on NV rule and set $V^{'} = V^{'} \setminus v_*$
18: **else**: ▷ if $V^{'} = \emptyset$
19: set $A = A \setminus a_\lambda$ ▷ there is no unserviced victim to be allocated to ambulance a_λ
20: **elif** v_λ is a red victim **then:**
21: observe S_{v_λ} and the travel times of adjacent arcs to v_λ
22: service v_λ ▷ service the red victim
23: determine $h_* \in H$ based on NH rule
24: set next location $a_\lambda = h_*$
25: set $C_{h_*} = C_{h_*} - 1$ ▷ share information and reduce the capacity of hospital h_* by one
26: **elif** v_λ is a black victim **then:**
27: observe the travel times of adjacent arcs to v_λ
28: choose $v_* \in V^{'}$ as the next location of a_λ based on NV rule and set $V^{'} = V^{'} \setminus v_*$
29: bypass the black victim without providing service
30: **elif** if v_λ is a hospital, **then:**
31: **if** $V^{'} \neq \emptyset$, **then:**
32: choose $v_* \in V^{'}$ as the next location of a_λ based on NV rule and set $V^{'} = V^{'} \setminus v_*$
33: **else**: ▷ if $V^{'} = \emptyset$
34: set $A = A \setminus a_\lambda$ ▷ there is no untreated victim to be allocated to ambulance a_λ
35: **if** $A \neq \emptyset$ **then:** ▷ if there is at least one more victim to be served
36: Go to **dispatch**, i.e., line 6
37: **else:**
38: **stop** ▷ termination point

Data deposition information In our study, we selected a subset of the data sets presented in Talarico et al. (2015) to undertake the computational experiments. A detailed description of the used data sets is presented in Sect. 6.1 and the relevant files can be openly accessed through this link: https://figshare.com/ndownloader/files/38509442.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Akbari V, Shiri D (2021) Weighted online minimum latency problem with edge uncertainty. Eur J Oper Res 295(1):51–65

Akbari V, Shiri D (2022) An online optimization approach for post-disaster relief distribution with online blocked edges. Comput Op Res 137:105533

Akbari V, Shiri D, Sibel Salman F (2021) An online optimization approach to post-disaster road restoration. Trans Res Part B: Methodol 150:1–25

Anuar, Wadi Khalid, Lee, Lai Soon, Pickl, Stefan, & Seow, Hsin-Vonn. 2021. Vehicle Routing Optimisation in Humanitarian Operations: A Survey on Modelling and Optimisation Approaches. Appl Sci **11**(2)

Aringhieri R, Bruni ME, Khodaparasti S, van Essen JT (2017) Emergency medical services and beyond: Addressing new challenges through a wide literature review. Comput Op Res 78:349–368

Aringhieri R, Bigharaz S, Duma D, Guastalla A (2022) Fairness in ambulance routing for post disaster management. CEJOR 30:189–211

Ausiello G, Feuerstein E, Leonardi S, Stougie L, Talamo M (2001) Algorithms for the on-line travelling salesman. Algorithmica 29(4):560–581

Brown LC, David B, Smith LC, Micah J, Chibi LC, Tazi M, Hassani N, Lotfi CB (2017) Minimizing postdisaster fatalities. Fed Pract 34(2):10

Bélanger V, Ruiz A, Soriano P (2019) Recent optimization models and trends in location, relocation, and dispatching of emergency medical vehicles. Eur J Oper Res 272(1):1–23

Büttner S, Krumke SO (2016) The Canadian Tour Operator Problem on paths: tight bounds and resource augmentation. J Comb Optim 32:842–854

CRED (2013) People affected by conflict: Humanitarian needs in numbers. Tech. rept, Centre for Research on the Epidemiology of Disasters, Brussels

CRED, & UNDRR. 2020. Human cost of disasters: An overview of the last twenty years 2000-2019. Tech. rept. Centre for Research on the Epidemiology of Disasters and UN Office for Disaster Risk Reduction

Daud SM, Mohd S, Yusof MY, Mohd P, Heo CC, Khoo LS, Singh MK, Chainchel M, Shah M, Nawawi H (2022) Applications of drone in disaster management: A scoping review. Sci Justice 62(1):30–42

De la Torre, Luis E., Dolinskaya, Irina S., & Smilowitz, Karen R. 2012. Disaster relief routing: Integrating research and practice. Socio-Economic Planning Sciences, **46**(1), 88–97. Special Issue: Disaster Planning and Logistics: Part 1

Dukkanci O, Koberstein A, Kara BY (2023) Drones for relief logistics under uncertainty after an earthquake. Eur J Oper Res 310(1):117–132

Farahani RZ, Lotfi MM, Baghaian A, Ruiz R, Rezapour S (2020) Mass casualty management in disaster scene: A systematic review of OR & MS research in humanitarian operations. Eur J Oper Res 287(3):787–819

Ghiani G, Guerriero F, Laporte G, Musmanno R (2003) Real-time vehicle routing: Solution concepts, algorithms and parallel computing strategies. Eur J Oper Res 151(1):1–11

Jaillet P, Wagner MR (2008) Generalized Online Routing: New Competitive Ratios, Resource Augmentation, and Asymptotic Analyses. Oper Res 56:745–757

Jat MN, Rafique RA (2020) Mass-Casualty distribution for emergency healthcare: a simulation analysis. Int J Disaster Risk Sci 11(3):364–377

Khoshgehbari, Farnaz, Al-e, S Mohammad J Mirzapour, et al. 2023. Ambulance Location Routing Problem Considering all Sources of Uncertainty: Progressive Estimating Algorithm. Comput Op Res, 106400

Lee Y-C, Chen Y-S, Chen AY (2022) Lagrangian dual decomposition for the ambulance relocation and routing considering stochastic demand with the truncated Poisson. Trans Res Part B: Methodol 157:1–23
Letchford Adam N, Salazar-González, & Juan-José. (2006) Projection results for vehicle routing. Math Progr 105(2):251–274
Liang N-J, Shih Y-T, Shih F-Y, Wu H-M, Wang H-J, Shi S-F, Liu M-Y, Wang BB (2001) Disaster epidemiology and medical response in the Chi-Chi earthquake in Taiwan. Ann Emerg Med 38(5):549–555
Lu, Lili, & Wang, Shuaian. 2019. Literature Review of Analytical Models on Emergency Vehicle Service: Location, Dispatching, Routing and Preemption Control. Pages 3031–3036 of: 2019 IEEE Intelligent Transportation Systems Conference (ITSC)
Ma W, Simchi-Levi D (2020) Algorithms for online matching, assortment, and pricing with tight weight-dependent competitive ratios. Oper Res 68(6):1787–1803
Ma W, Simchi-Levi D, Teo C-P (2021) On policies for single-leg revenue management with limited demand information. Oper Res 69(1):207–226
Najafi M, Eshghi K, de Leeuw S (2014) A dynamic dispatching and routing model to plan/ re-plan logistics activities in response to an earthquake. OR Spectrum 36(2):323–356
Rios O, Humberto B, Xavier EC, Miyazawa FK, Amorim P, Curcio E, Santos MJ (2021) Recent dynamic vehicle routing problems: A survey. Comput Indus Eng 160:107604
Oksuz MK, Satoglu SI (2020) A two-stage stochastic model for location planning of temporary medical centers for disaster response. Int J Disaster Risk Reduct 44:101426
Pillac V, Gendreau M, Guéret C, Medaglia AL (2013) A review of dynamic vehicle routing problems. Eur J Oper Res 225(1):1–11
Rabbani M, Oladzad-Abbasabady N, Akbarian-Saravi N (2022) Ambulance routing in disaster response considering variable patient condition: NSGA-II and MOPSO algorithms. J Indus Manag Optim 18(2):1035–1062
Ritzinger U, Puchinger J, Rudloff C, Hartl RF (2022) Comparison of anticipatory algorithms for a dial-a-ride problem. Eur J Oper Res 301(2):591–608
Salman FS, Gül S (2014) Deployment of field hospitals in mass casualty incidents. Comput Indus Eng 74:37–51
Schilde M, Doerner KF, Hartl RF (2011) Metaheuristics for the dynamic stochastic dial-a-ride problem with expected return transports. Comput Oper Res 38(12):1719–1730
Shiri D, Salman FS (2020) Online Optimization of First-responder Routes in Disaster Response Logistics. IBM J Res Dev 64:1–9
Shiri D, Akbari V, Salman FS (2020) Online routing and scheduling of search-and-rescue teams. OR Spectrum 42(3):755–784
Shiri, Davood, Akbari, Vahid, & Tozan, Hakan. 2023. Online optimisation for ambulance routing in disaster response with partial or no information on victim conditions. Comput Oper Res, 106314
Sleator D, Tarjan R (1985) Amortized efficiency of list update and paging rules. Commun ACM 28:202–208
Soeffker N, Ulmer MW, Mattfeld DC (2022) Stochastic dynamic vehicle routing in the light of prescriptive analytics: A review. Eur J Oper Res 298(3):801–820
Talarico Luca, Meisel Frank, Sörensen Kenneth (2015) Ambulance routing for disaster response with patient groups. Comput Opera Res 56:120–133
Talebi E, Shaabani M, Rabbani M (2022) Bi-objective model for ambulance routing for disaster response by considering priority of patients. Int J Supply Oper Manag 9(1):80–94
Tiedemann, Morten, Ide, Jonas, & Schöbel, Anita. 2015. Competitive analysis for multi-objective online algorithms. Pages 210–221 of: International Workshop on Algorithms and Computation. Springer
Tikani H, Setak M (2019) Ambulance routing in disaster response scenario considering different types of ambulances and semi soft time windows. J Indus Syst Eng 12(1):95–128
Tippong D, Petrovic S, Akbari V (2022) A review of applications of operational research in healthcare coordination in disaster management. Eur J Oper Res 301(1):1–17
Tlili, Takwa, Harzi, Marwa, & Krichen, Saoussen. 2017. Swarm-based approach for solving the ambulance routing problem. Procedia Computer Science, **112**, 350–357. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 21st International Conference, KES-20176-8 September 2017, Marseille, France
Tlili T, Abidi S, Krichen S (2018) A mathematical model for efficient emergency transportation in a disaster situation. Am J Emerg Med 36(9):1585–1590
Yao, Andrew Chi-Chih. (1977) Probabilistic computations: Towards a unified measure of complexity. Proceedings of the 18th Annual IEEE Symposium on the Foundations of Computer Science, 222–227

Yao C, Chen S, Yang Z (2022) Online distributed routing problem of electric vehicles. IEEE Trans Intell Transp Syst 23(9):16330–16341

Yazdani M, Haghani M (2023) Optimisation-based integrated decision model for ambulance routing in response to pandemic outbreaks. Progr Disaster Sci 18:100288

Yoon S, Albert LA (2020) A dynamic ambulance routing model with multiple response. Trans Res Part E: Logist Trans Rev 133:101807

Zhang H, Tong W, Lin G, Xu Y (2019) Online minimum latency problem with edge uncertainty. Eur J Oper Res 273:418–429

Zidi, Issam, Al-Omani, Mohammad, & Aldhafeeri, Karim. 2019. A New Approach Based On the Hybridization of Simulated Annealing Algorithm and Tabu Search to Solve the Static Ambulance Routing Problem. Procedia Computer Science, **159**, 1216–1228. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 23rd International Conference KES2019

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.