IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 15, 2022, accepted April 12, 2022, date of publication April 19, 2022, date of current version April 27, 2022.

*Digital Object Identifier 10.1109/ACCESS.2022.3168736*

# A Novel Smart Ambulance System–Algorithm Design, Modeling, and Performance Analysis

**MOHAMMAD A. R. ABDEEN[1], MOHAMED HOSSAM AHMED[2], (Senior Member, IEEE), HAFEZ SELIEM[3], TAREK RAHIL SHELTAMI[4], TURKI M. ALGHAMDI[1], AND MUSTAFA EL-NAINAY[5], (Senior Member, IEEE)**

[1]Faculty of Computer Science and Information Systems, Islamic University of Madinah, Medina 42351, Saudi Arabia
[2]School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada
[3]Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11517, Egypt
[4]Department of Computer Engineering, Interdisciplinary Research Center of Smart Mobility and Logistics, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
[5]Department of Computer Engineering, AlAlamein International University, New Alamein City 51718, Egypt

Corresponding author: Mohamed Hossam Ahmed (mahme3@uottawa.ca)

This work was supported in part by the Deputyship for Research and Innovation, Ministry of Education, Saudi Arabia, under Project 15/20, and in part by the King Fahd University of Petroleum and Minerals (KFUPM), IRC SML Research under Grant INML 2104.

**ABSTRACT** Smart ambulance is a novel system where modern communication, computation, and sensing technologies are employed to revolutionize ambulance and emergency systems. We propose a smart system that aims to minimize the ambulance response time, travel time from patient's location to the hospital, and the waiting time at the hospital. We utilize the road traffic conditions and hospital loading information (collected in real-time basis) to make optimal decisions (which hospital responds to the patient's request and which ambulance it sends, which route the ambulance takes to reach the patient, which hospital the ambulance heads to after picking up the patient, and which route it should take to the selected hospital). The first two decisions are used to minimize the response time while the last two decisions are employed to minimize the door-to-needle time. We analyze the performance of the proposed algorithm; both analytically and by simulation for verification. The results showed very good consistency between simulation results and analytical results, which confirms the correctness and accuracy of the analysis. In addition, we compare the performance of our proposed smart algorithm with a previous algorithm that is reported in the literature and that minimizes the drop-off delay. The results confirmed the superiority of our smart algorithm under considered operating conditions and scenarios.

**INDEX TERMS** Smart health, smart ambulance, emergency service, performance analysis.

## I. INTRODUCTION

There is a growing demand for smart systems nowadays where provided services can be improved using modern technologies of sensing, communication, high computing performance, signal processing and multimedia. Such technologies can be utilized to improve ambulance and emergency services. In [1], authors analyzed data for emergency medical service in urban and rural areas in the United States. Measured data showed that the average response time was 7 Minutes for urban areas and 14 Minutes for rural areas. Authors in [2] showed that longer response time is associated with worse outcomes for trauma patients. In [3], it was indicated that in rural cases modest delay in response time can be life threatening. In [4], the authors proposed an ubiquitous emergency medical service system. The proposed system aims to minimize the response time (call-to-site and site-to-door) and to improve the pre-hospital treatment (including patient's site before ambulance arrival as well as the ambulance vehicle. In [5], it was shown that reducing the response time by one minute improves the survival rate of patients with sudden cardiac arrest by 24%. Authors in [6] used several techniques to reduce the response time by 50%. These techniques include controlling traffic lights (through central traffic system). In [7], it was shown that dynamic rerouting can be used to relocate nearest ambulance vehicle to the site. In [8], the authors analyzed the transportation delay in emergency medical service of suspected

The associate editor coordinating the review of this manuscript and approving it for publication was Young Jin Chun.

ST-elevation-myocardial infarction in the VIENNA-STEMI network. They used records of 4751 patients to measure the call-to-site and site-to-door times.

Smart ambulance has been proposed recently to improve the performance of ambulance. Smart ambulance uses technologies such as Internet-of-things, real-time data communication and video streaming, connected vehicles, road traffic monitoring, big data, biomedical sensing, and body area networks to improve the emergency service, minimize response time, and provide medical support with the least possible delay [9]–[11]. However, smart ambulance needs high speed data transmission to support high-quality and real-time video, data and audio communication between ambulance and hospitals. Authors in [12] showed that existing communication networks (e.g., Long-Term Evolution (LTE) wireless networks) are unable to meet the requirements of smart ambulance and proposed a 5G-based wireless network to facilitate the smart ambulance. Authors in [13] designed mobile apps to locate the nearest ambulance to the patient. When the ambulance reaches the patient's location, the mobile app shows nearby hospitals, their locations from which the nearest hospital is selected. Finally, the app at the ambulance finds the shortest route to the hospital.

A fuzzy logic -based system was proposed in [14] to manage the traffic system. The objective of the proposed system is to minimize the ambulance travel time. The fuzzy controller receives traffic conditions (in terms of the average vehicle speed and occupancy level) and produces a congestion value parameter. The congestion level and other parameters (emergency priority level received from the ambulance and traffic conditions) are applied to the traffic management system which controls the traffic lights, reserved lane, maximum speed, vehicle rerouting, and lane clearance. A system was proposed in [15] to alert the drivers (using mobile app-generated messages) about the emergency route (route selected by a coming ambulance). By doing so, the drivers make the road clear for the ambulance early enough before it arrives, which can minimize the response time.

In [16], authors proposed an algorithm to minimize the ambulance drop-off time. They classified hospitals as urban (inside the city) or community (outside the city). When a patient requests ambulance, the closest hospital deploys an ambulance to the patient. If there is no congestion at urban hospitals, the ambulance takes the patient back to the closest hospital. Otherwise, the ambulance takes the patient to the closest community hospital instead. Markov-decision process is used to model the system and determine the optimal solution. Results showed that the proposed algorithm reduced the drop-off time. However, their work has some drawbacks. First, authors combined all waiting patients in urban hospitals in one virtual queue. Second, they assumed that the congestion status is the same for all urban hospital, i.e. either all urban hospitals are congested or all of them are not. Furthermore, they assumed that community hospitals are always uncongested. Moreover, this work does not take the traffic and road conditions into consideration in the system model. These limitations make the work less practical in general situations and restricts its outcomes to small cities with uniform traffic conditions and uniform distribution of city inhabitants and uniform distribution of resources at the urban hospitals.

Authors in [17] proposed a queuing model to analyze the performance of ambulance system. They analyzed the impact of patient routing (from patient's location to one of the hospitals) by the dispatcher in the performance particularly the drop-off delay. The authors extended their work in [18] by optimizing the routing probability to minimize the drop-off delay. Despite the effectiveness of the proposed algorithm in minimizing the drop-off delay, it has some limitations. First, it is a static algorithm where the routing probabilities is calculated using the long-term statistical parameters (e.g., average arrival rate, average treatment (service) time, number of servers (treatment teams) at each hospital, etc.). Hence, the algorithm lacks the adaptation mechanism in case some of those parameters change. Second, due to the static nature of the algorithm, it does not take into account the loading at each hospital (which is dynamic by its nature). Third, the algorithm does not take into consideration the patients' locations and the traffic and road conditions. Fourth, the authors assume that ambulance requests are lost if they arrive while no ambulance is available. This assumption is not realistic since the patient's request should be kept in a virtual queue to be served by one of the next available ambulances.

In [19], we proposed a novel algorithm that uses smart health technology in ambulance service to minimize the response time (time from receiving the request to the arrival of the ambulance at the patient's location) and to minimize the door-to-needle time which is the sum of the delivery time (from the patient's location to the hospital) and the waiting time (at the hospital). Therefore, we minimize the time from request until the patient starts to receive treatment at the emergency department in the hospital. In this paper we extend the work proposed in [19] as follows:

- We enhance the accuracy and practicality of the system model by adding the ambulance drop-off delay at hospitals, the non-preemption priority to emergency patients, and the ambulance centres.
- We improve the proposed smart algorithm by adding one more feature which is the treatment time reduction due to the use of smart ambulance.
- We develop a queuing model that represents all phases of ambulance system including the ambulance travel from hospitals/ambulance centres to patients' location, ambulance travel from patients' location to hospitals, and waiting at hospitals unlike the queuing model in [17], [18] that does not include the ambulance travel and provides only the optimal routing decision that minimized the waiting time at hospitals. Analytical results are obtained using our queuing model and are compared with the simulation results for verification.
- We compare our results with existing work in the literature (namely work in [18]) and show that our algorithm

outperforms that in [18] as discussed in details in the Results section.
- We analyze the proposed algorithm under realistic time-varying loading conditions.
- We apply the proposed algorithm to Madinah city in Saudi Arabia to analyze the performance results of the proposed algorithm in practical settings.

The remainder of this paper are organized as follows. In the next section, we explain the system model and the proposed algorithm. In Section III, we present the queuing model which we developed to determine the performance of the proposed algorithm analytically. Results are given and discussed in Section IV. Finally, conclusions and future work are given in Section V.

## II. SYSTEM MODEL AND PROPOSED ALGORITHM

We consider a virtual city[1] with a circular shape of a radius $R_{max}$ as shown in Fig. 1. Modern cities and especially in the ''new world'' usually exhibit a regular shape such as square or rectangular (Manhattan model). This is mainly due to the fact that lands were available in vast sizes and there were no natural barriers that limit city planners to use other shapes [20]. Ancient cities (such as Madinah and Mecca in Saudi Arabia), on the other hand, were designed and built in areas that are normally populated and that are built around a specific centre of attraction (city centre and thereby the name) [21]. The assumed city shape and hospital locations are chosen to resemble the centre of Madinah city.

The proposed smart ambulance provides interconnectivity between hospitals, ambulance centres, ambulance vehicles, central dispatcher, and patients. The high-speed connectivity between the hospitals/ambulance centres and ambulance vehicles can be provided using 5G wireless networks. This interconnectivity makes information (including ambulance vehicle location, patients' locations, load at hospitals, congestion and road conditions, etc.) available in real-time basis at the central dispatcher as well as the ambulance vehicles. Thus, this information is used to minimize the time between requesting the ambulance service and the start of treatment at the emergency departments at the hospitals. It is assumed that the considered virtual city has $S$ hospitals and $N$-$S$ ambulance centres. Hence, we have a total of $N$ locations of ambulance vehicles. However, it should be noted that hospitals have emergency departments and ambulance vehicles while the ambulance centres have ambulance vehicles only. The virtual city has main roads (shown as thick solid lines) and inner roads (shown as thin dashed lines) with maximum speed limits of 70 km/h and 40 km/h, respectively. Furthermore, we assume ambulance vehicles move at the maximum speed limit of the roads. The city inhabitants are uniformly distributed across the city area. The patients' arrival is modelled using Poisson distribution with an average arrival rate $\lambda_a$ patients/hour and $\lambda_w$ patients/hour for ambulance patients and walk-in patients, respectively.

[1]We consider a regular city shape to facilitate the mathematical analysis. However, we consider a practical city model (Madinah City in Saudi Arabia) afterwards in the Results Section.

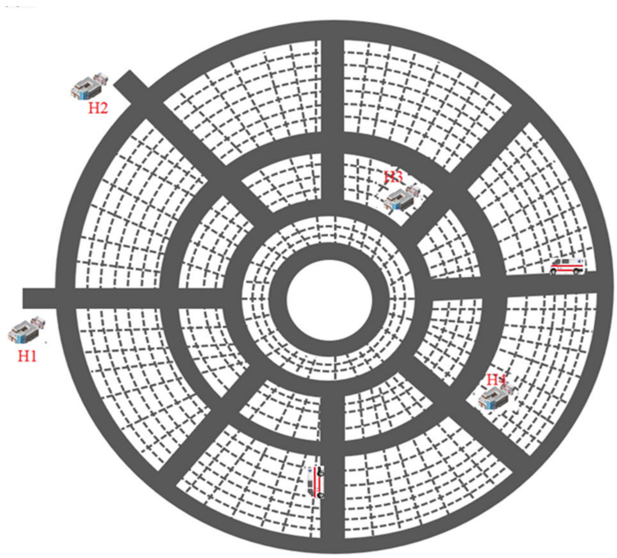



**FIGURE 1.** Virtual city map.

Each hospital/ambulance centre has $K_i$ ambulance vehicles, where $i$ is the index of the hospital/ambulance centre. When a patient calls the emergency service, the central dispatcher uses the road traffic information, the patient's location and the locations of all ambulance vehicles to assign the request to the hospital/ambulance centre with the minimum expected response time ($T_R$), where the response time is the sum of the travel time from the hospital/ambulance centre to the patient's location ($T_{H2P}$) and the waiting time until an ambulance vehicle becomes available if none of the $K_i$ ambulance vehicles is available. Then, the hospital/ambulance centre sends an ambulance vehicle to the patient's location. After the ambulance reaches the patient's location, the patient is delivered to one of the $S$ hospitals. The ambulance selects the hospital that minimizes patient door-to-needle time which is the sum of the travel time from the patient's location to the hospital ($T_{P2H}$) and the waiting time at the hospital ($T_{WT}$). The optimal routes to be followed by the ambulance (in its way to the patient and back to the hospital) are also determined. We assume that each hospital has a maximum capacity of $M_i$ patients; which means that $M_i$ patients at most can be treated simultaneously by the emergency department at the $i^{th}$ hospital. When a patient arrives to the emergency service, he/she is admitted and served if there is enough capacity. Otherwise, the ambulance patient is kept in the ambulance vehicle due to the need of life-support equipment, while the walk-in patient is kept in the hospital waiting room. When resources become free at the emergency departments, waiting patients are admitted in first-in first-out (FIFO) basis with a non-preemption priority to the ambulance patients. We also assume that the service time (i.e, treatment time) is exponentially distributed with an average duration $T_T$. When an ambulance patient is admitted and transferred from the ambulance to be treated, the ambulance vehicle returns to its initial location at one of the $S$ hospitals or $N$-$S$ ambulance centres. In this work we analyze the performance of the proposed algorithm using the following performance metrics:

1) Average ambulance response time ($T_R^{avg}$) : The average time duration between receiving the call for an ambulance service and the arrival of the ambulance vehicle at the patient's location.
2) Average ambulance patients door-to-needle time ($T_{P2H\_a}^{avg} + T_{WT\_a}^{avg}$): The sum of the average time the ambulance takes to transport the patient to one of the hospitals and the average drop-off time (time between the ambulance patient's arrival at the hospital and the beginning of the treatment).
3) Average walk-in patients door-to-needle time ($T_{P2H\_w}^{avg} + T_{WT\_w}^{avg}$): The sum of the average time the walk-in patient takes to go from the patient's location to one of the hospitals and the average waiting time (time between the patient's arrival at the hospital and the beginning of the treatment).

We have selected these three performance metrics because the first metric indicates how fast the ambulance arrives at the patients' location while the second and third ones indicate how fast the patients receive treatment after departing form their locations. Furthermore, these three metrics are the functions that our proposed smart algorithm aims to minimizes.

Algorithms 1, 2, and 3 (given below) illustrate the 3 components of the proposed smart system (ambulance dispatching, selecting a hospital for an ambulance patient, and selecting a hospital for a walk-in patient, respectively). As shown in Algorithm 1, when a patient's request is received, the central dispatcher estimates the response time (which is the travel time from the ambulance vehicle's location to the patient's location) based on the patient's location, available ambulance vehicles' locations, and road traffic conditions. The ambulance vehicle with the minimum response time is dispatched to the patient's location. In Algorithm 2, the travel time from the patient's location to each hospital is estimated based on the patient's location, hospitals' locations, and road traffic conditions. Algorithm 2 also estimates the ambulance patient's waiting time (also called drop-off time) at each hospital by $n_{wap}(i)\, T_T \big/ M_i$, where $n_{wap}(i)$ is the number of waiting ambulance patients at the $i^{\text{th}}$ hospital. Finally, the estimated door-to-needle time is calculated as the sum of the estimated travel time and the estimated drop-off time. The hospital with the minimum estimated door-to-needle time is selected and the ambulance vehicle drives the patient from her/his location to the selected hospital using the route with the minimum travel time. Finally, Algorithm 3 estimates the door-to-needle time of a walk-in patient in a similar way to Algorithm 2 except that the waiting time at the hospital of walk-in patients is estimated by $n_{twp}(i)\, T_T \big/ M_i$, where $n_{twp}(i)$ is the total number of waiting patients (including ambulance patients and walk-in patients) at the $i^{\text{th}}$ hospital. This difference between the estimated waiting time of ambulance patients and walk-in patients is due to the non-preemption priority of the ambulance patients.

Table 1 lists all mathematical symbols used in this paper. We compare the proposed algorithm with traditional ambulance which does not take into consideration the real-time information about the hospitals and the roads.

## III. ANALYTICAL MODEL

In order to analyze the performance of the proposed system analytically we developed the queueing system shown in Fig. 2. The block at the top models the central dispatcher that receives the patients' requests. The central dispatcher responds to these requests by assigning the patient to the $i^{\text{th}}$ hospital/ambulance centre with the minimum response time ($T_R$). This step is modelled using the first bank of parallel queues (below the central dispatcher). Each one of these $N$ queues has multiple servers, namely, queue $i$ has $K_i$ where $i = 1, 2, 3, \ldots, N$, and $K_i$ is the number of ambulance vehicles at the $i^{\text{th}}$ hospital/ambulance centre. Since the central dispatcher assigns the patient to the hospital/ambulance centre with the minimum expected response time (sum of the expected queuing delay and average service (travel) time to the requesting patients' location), this bank of parallel $N$ queues can be modelled as parallel queues with multiple servers per queue, where the shortest delay first (SDF) routing policy is employed to assign requests to queues.

The travel time from the patient's location to the hospitals is represented by the $S$ parallel servers after the decision-making block. The last part of the queuing model is the bank of $S$ parallel queues with multiple servers per queue ($M_i$ servers, where $i = 1, 2, 3, \ldots, S$) representing the emergency departments at the hospitals. It is worth noting that arrivals to the queues consist of ambulance patients and walk-in patients. The routing policy of these $S$ parallel queues can also be considered as the SDF routing policy. It was shown in [22] and [23] that the performance of parallel queues (with multiple servers per queue) using SDF routing policy is almost identical to that of a single queue with the same total number of parallel servers and FIFO policy. Hence, we can analyze the two banks of parallel queues shown in Fig. 2 using an equivalent single queue with multiple servers and FIFO

$$T_R^{avg} = T_{H2P}^{avg} + \frac{\left(T_{H2P}^{avg} + T_{P2H\_a}^{avg} + T_{WT\_a}^{avg}\right) C\left(\sum_{i=1}^{N} k_i, \lambda_a \left(T_{H2P}^{avg} + T_{P2H\_a}^{avg} + T_{WT\_a}^{avg}\right)\right)}{\sum_{i=1}^{N} k_i - \lambda_a \left(T_{H2P}^{avg} + T_{P2H\_a}^{avg} + T_{WT\_a}^{avg}\right)}, \tag{1}$$

**Algorithm 1** Ambulance Dispatching

**input**:
  **Queue** PatientRequest ▷ FIFO queue for the received calls
  **List** AvailableAmbulances ▷ List of available ambulances ready for order

```
while true do
  if PatientRequest.isEmpty or
    AvailableAmbulance.isEmpty
  then continue ▷ We do not have any request, or we do
      not have available ambulance
  else
      Patient ← PatientRequest.dequeue
      SelectedAmbulance ←AvailableAmbulances [0]
      MinResponseTime ← Integer.Max
      forall Ambulance ▷ AvailableAmbulances do
          ResponseTime ← calculateTravelTime
          (Ambulance, Patient) ▷ get the shortest
          travel time path between an ambulance and
          patient
          if MinResponseTime > ResponseTime then
              SelectedAmbulance ← Ambulance
              MinResponseTime ← ResponseTime
          end
      end
  end
    AvailableAmbulances.remove(SelectedAmbulance)
    send (AvailableAmbulance, Patient)
end
```

**Algorithm 2** Selecting a Hospital for the Ambulance-Patient

**input**:
  List Hospitals ▷ List of Hospitals
  AmbulancePatient ▷ The ambulance patient after the ambulance arrived to his/her location

```
MinDoor2NeedleTime ← Integer.Max
SelectedHospital ← Hospitals[0]
forall Hospital ∈ Hospitals do
    ExpectedTravelTime ← calculateTravelTime
    (Hospital, AmbulancePatient) ▷ get the
    shortest travel time path between the current
    location of the patient and the hospital
    ExpectedDropOffTime ←
    Hospital.calculateDropOff() get the expected
    drop-off time for the hospital based on its
    current status
    Expected Door2NeedleTime ←
    ExpectedTravelTime +
    ExpectedDropOffTime
    if Min Door2NeedleTime > Expected
      Door2NeedleTime
    then
        SelectedHospital ← Hospital
        Min Door2NeedleTime ← Expected
        Door2NeedleTime
    end
end
send(AmbulancePatient, SelectedHospital)
```

policy. Thus, the performance metrics can be approximated as follows:

### A. AVERAGE RESPONSE TIME ($T_R^{avg}$)

Eq. (1), as shown at the bottom of the previous page, where $T_{H2P}^{avg}$ is the average travel time from the selected hospital/ambulance centre to the patient, $T_{P2H\_a}^{avg}$ is the average travel time from the patient's location to the selected hospital, $T_{WT\_a}^{avg}$ is the average ambulance drop-off time at the hospital and $C$ (., .) is the Erlang-C formula [24].

### B. AVERAGE AMBULANCE PATIENT DOOR-TO-NEEDLE TIME ($T_{P2H\_a}^{avg} + T_{WT\_a}^{avg}$)

$$T_{P2H\_a}^{avg} + \frac{1}{1-\sigma}\left(\frac{c!\,(1-\rho)\,c}{T_T}\sum_{n=0}^{c-1}\frac{(c\rho)^{n-c}}{n!}+\frac{c}{T_T}\right)^{-1}, \tag{2}$$

where $\sigma$ is the server utilization ratio by the ambulance patients which is given by $\lambda_a T_T/c$, $\rho$ is the server utilization ratio by either ambulance patients or walk-in patients which is given by $(\lambda_a + \lambda_w)\,T_T/c$, and $c$ is the total number of servers in the single queue FIFO system which is given by $c = \sum_{i=1}^{s} M_i$ [25].

### C. AVERAGE WALK-IN PATIENT DOOR-TO-NEEDLE TIME ($T_{P2H\_w}^{avg} + T_{WT\_w}^{avg}$)

$$T_{P2H\_w}^{avg} + \frac{1}{(1-\sigma)(1-\rho)} \times \left(\frac{c!\,(1-\rho)\,c}{T_T}\sum_{n=0}^{c-1}\frac{(c\rho)^{n-c}}{n!}+\frac{c}{T_T}\right)^{-1}. \tag{3}$$

In order to calculate the average travel times of the ambulance ($T_{H2P}^{avg}$, $T_{P2H\_a}^{avg}$ and $T_{P2H\_w}^{avg}$), we need to determine the areas that belong to each hospital by finding the closest hospital to point in the city circular shape. As shown in Fig. 3 (for $N = S = 4$), we draw the green borderlines that separate the city area based on the closest hospitals. This is done by connecting each pair of hospitals using the blue lines. Then, drawing a perpendicular green line that pass by the middle point of the blue line indicates the borderline between these two hospitals. The area that belongs to each hospital is defined by the green borderlines and the part of the circumference of the outer area with the maximum radius ($R_{max}$).

Hence, $T_{H2P}^{avg}$ can be determined by averaging $T_{H2P}$ over the city area. Hence, $T_{H2P}^{avg}$ can be expressed as

$$T_{H2P}^{avg} = \sum_{i=1}^{N}\left(\iint_{A_i}\sum_{l}\frac{d_{H2P}(i,r,\theta,l)}{v_a(i,r,\theta,l)}f_{R,\Theta}(r,\theta)\,rdrd\theta\right), \tag{4}$$

**TABLE 1.** List of symbols and their definitions.

| Symbol | Definition |
|---|---|
| $R_{max}$ | Virtual cell radius |
| $S$ | Number of Hospitals |
| $N$ | Number of Hospitals + number of ambulance centres |
| $\lambda_a$ | Arrival rate of ambulance patients |
| $\lambda_w$ | Arrival rate of walk-in patients |
| $K_i$ | Number of ambulance vehicles at the $i^{th}$ hospital |
| $i$ | The hospital/ambulance index |
| $T_R$ | Response time (time between ambulance request and ambulance arrival at the patient's location) |
| $T_{H2P}$ | Travel time from the hospital/ambulance centre to the patient's location |
| $T_{P2H}$ | Travel time from the patient's location to the hospital |
| $T_{WT}$ | Waiting time at the hospital |
| $M_i$ | Capacity of the emergency department at the $i^{th}$ hospital |
| $T_T$ | Average treatment time at the emergency department |
| $T_R^{avg}$ | Average response time (i.e., average value of $T_R$) |
| $T_{P2H\_a}^{avg}$ | Average delivery (travel) time of the ambulance patient to the hospital (i.e., average value of $T_{P2H}$ of the ambulance patients) |
| $T_{WT\_a}^{avg}$ | Average ambulance drop-off (waiting) time at the hospital (i.e., average value of $T_{WT}$ of the ambulance patient) |
| $T_{P2H\_w}^{avg}$ | Average delivery time of the walk-in patient to the hospital (i.e., average value of $T_{P2H}$ of the walk-in patients) |
| $T_{WT\_w}^{avg}$ | Average walk-in patient waiting time at the hospital (i.e., average value of $T_{WT}$ of the walk-in patient) |
| $n_{wap}(i)$ | The number of waiting ambulance patients at the $i^{th}$ hospital |
| $n_{twp}(i)$ | The total number of waiting patients at the $i^{th}$ hospital (including ambulance patients and walk-in patients) |
| $C(.\,,.)$ | Erlang-C formula |
| $d_{H2H}(i,j)$ | The length of the $l^{th}$ road segment of the route with the minimum travel time between the $i^{th}$ hospital/ambulance centre and the patient's location |
| $\sigma$ | Server utilization by ambulance patients = $\frac{\lambda_a T_T}{c}$ |
| $\rho$ | Server utilization = $\frac{(\lambda_a+\lambda_w)T_T}{c}$ |
| $c$ | Total number of servers = $\sum_{i=1}^{s} M_i$ |
| $(R, \Theta)$ | Random variable pair representing the polar coordinates of the patient's location |
| $(r, \theta)$ | Particular value of the random variable pair $(R, \Theta)$ |
| $d_{H2P}(i,r,\theta,l)$ | The length of the $l^{th}$ road segment of the route with the minimum travel time between the $i^{th}$ hospital/ambulance centre and the patient's location |
| $v_a(i,r,\theta,l)$ | The ambulance speed in the $l^{th}$ road segment of the route with the minimum travel time between the $i^{th}$ hospital/ambulance centre and the patient's location |
| $l$ | The index of the road segment of the route |
| $A_i$ | The area that belongs to the $i^{th}$ hospital/ambulance centre |
| $f_{R,\Theta}(r,\theta)$ | The probability density function (pdf) of the polar coordinates $(R, \Theta)$ of the patients' location |
| $d_{P2H}(i,r,\theta,l)$ | The length of the $l^{th}$ road segment of the route with the minimum travel time between the patient's location and the $i^{th}$ hospital/ambulance centre |
| $v_w(i,r,\theta,l)$ | The walk-in speed in the $l^{th}$ road segment of the route with the minimum travel time between the patient's location and the $i^{th}$ hospital/ambulance centre |
| $\tau_l$ | Average delay at the intersection between the $l^{th}$ road segment and the next road segment |

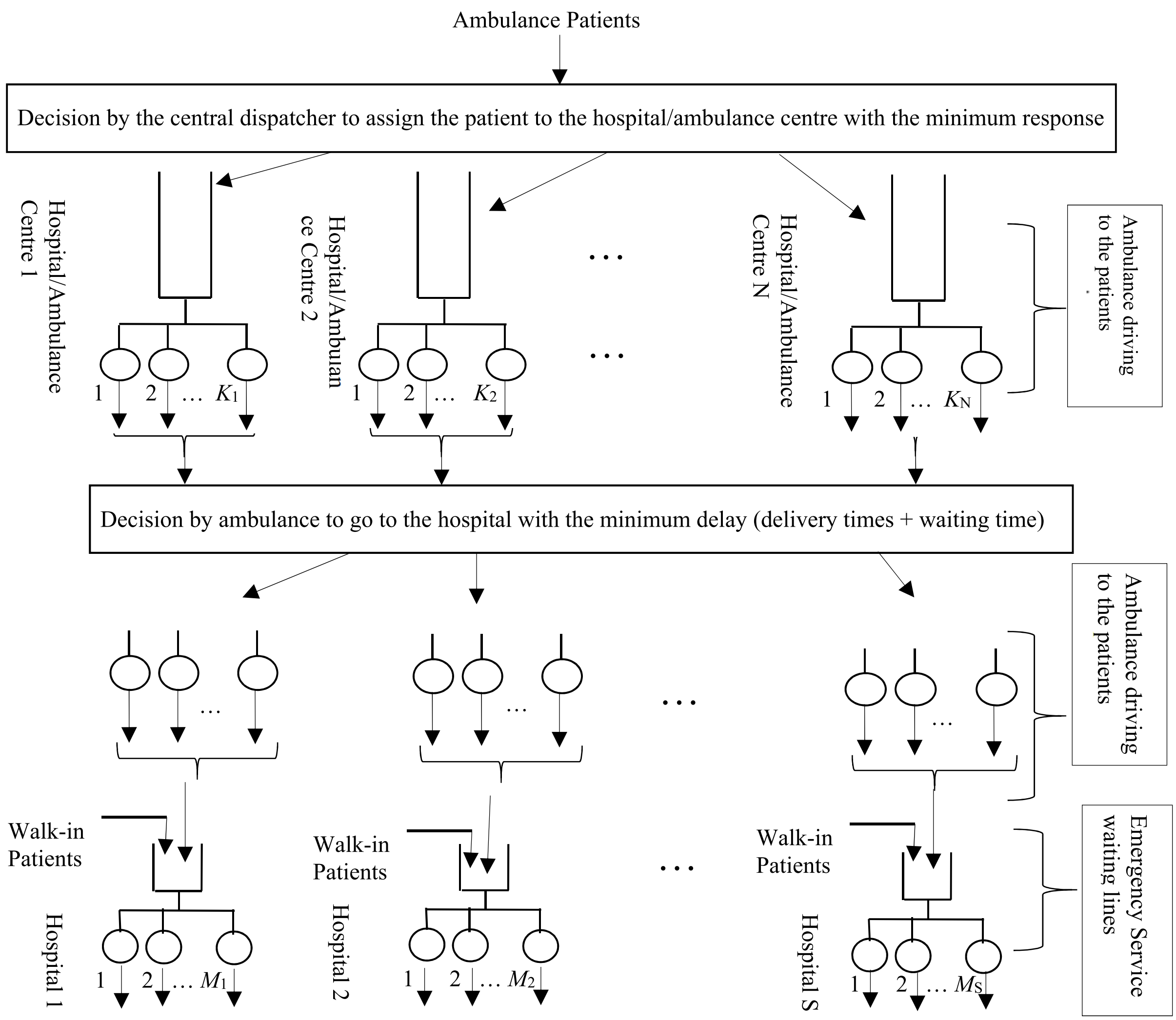



**FIGURE 2.** Queuing model of the smart ambulance system.

where $A_i$ is the area that belongs to the $i^{th}$ hospital/ambulance centre, $d_{H2P}(i, r, \theta, l)$ is the length of the $l^{th}$ road segment of the route with the minimum travel time between the $i^{th}$ hospital/ambulance centre and the patient's location expressed in the polar coordinates $(r, \theta)$, $v_a(i, r, \theta, l)$is the corresponding ambulance speed in the $l^{th}$ road segment, and $f_{R,\Theta}(r, \theta)$ is the probability density function (pdf) of the polar coordinates $(R, \Theta)$ of the patients' location. Assuming uniform distribution of the patient's locations and independence between $R$ and $\Theta$, it can be shown that

$$f_{R,\Theta}(r,\theta) = \begin{cases} \dfrac{2r}{R_{max}}\dfrac{1}{2\pi} = \dfrac{r}{R_{max}\pi}, \\ \quad 0 \le r \le R_{max} \quad 0 \le \theta \le 2\pi \\ 0 \quad otherwise \end{cases} \tag{5}$$

Similarly, $T^{avg}_{P2H\_a}$ and $T^{avg}_{P2H\_w}$ can be approximated by

$$T^{avg}_{P2H\_a} = \sum_{i=1}^{S}\left(\iint_{A_i}\sum_{l}\frac{d_{P2H}(i,r,\theta,l)}{v_a(i,r,\theta,l)} f_{R,\Theta}(r,\theta)\, r\, dr\, d\theta\right), \tag{6}$$

and

$$T^{avg}_{P2H\_w} = \sum_{i=1}^{S}\left(\iint_{A_i}\sum_{l}\left[\frac{d_{P2H}(i,r,\theta,l)}{v_w(i,r,\theta,l)} + \tau_l\right] \times f_{R,\Theta}(r,\theta)\, r\, dr\, d\theta\right). \tag{7}$$

where $d_{P2H}(i, r, \theta, l)$ is the length of the $l^{th}$ road segment of the route with the minimum travel time between the patient's location and the $i^{th}$ hospital and $v_w(i, r, \theta, l)$ is the corresponding walk-in speed in the $l^{th}$ road segment, $\tau_l$ is the average delay at the intersection between the $l^{th}$ road segment and the next road segment.

## IV. RESULTS

In this section we discuss the results of the proposed algorithm using simulation and the analytical model derived in the previous section. We simulate the system model describe in

**Algorithm 3** Selecting a Hospital for the Walk-in Patient

**input**:
List Hospitals ▷ List of Hospitals
WalkinPatient ▷ The walk-in patient who is looking for a hospital

Min Door2NeedleTime ← Integer.Max
SelectedHospital ← Hospitals[0]
**forall** Hospital ∈ Hospitals **do**
ExpectedTravelTime ← calculateTravelTime (Hospital,WalkinP atient)
▷get the shortest travel time path between the current location of the patientand the hospital
ExpectedWaitingTime ← Hospital.calculateWaitingTime()
▷ get the expected waiting time for walk-patient at the hospital based on its status
ExpectedDoor2NeedleTime ← ExpectedDoor2T ravelTime+ ExpectedWaitingTime
**if** MinNeedleTime > ExpectedNeedleTime **then**
SelectedHospital ← Hospital
MinNeedleTime ← ExpectedNeedleTime
**end**
**end**
send(WalkinPatient, SelectedHospital)

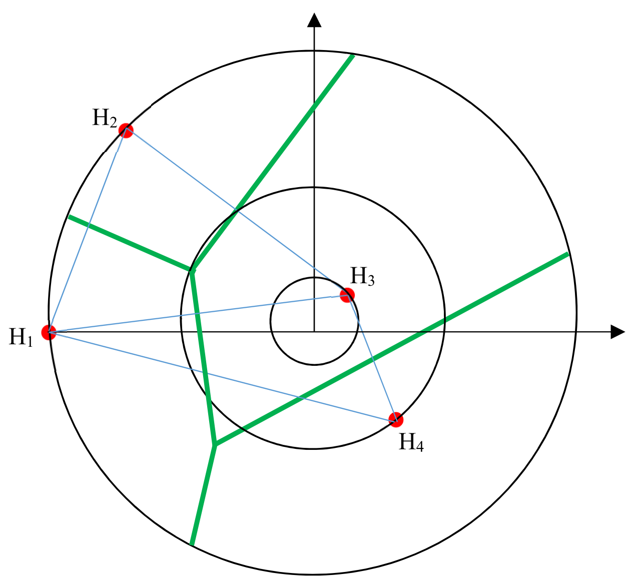



**FIGURE 3.** Borderlines of areas served with each hospital.

Section II using an event-driven dynamic simulation (implemented in Java) to simulate the patients' arrivals, ambulance vehicle selection, ambulance vehicle mobility, hospital selection, route selection, and patients' service (treatment) start and end. We compare the simulation results with the analytical results for verification. The analytical results are obtained by numerical substitution and numerical solution of the integrations in the equations developed in the analytical model. We also compare the performance of the proposed algorithm with that proposed in [18]. Finally, we evaluate the performance of the proposed algorithm using a real model and data of Madinah city in Saudi Arabia using AnyLogic® simulation package.

### A. CAPACITY PROPORTIONAL TO THE ARRIVAL RATE

As shown in Fig. 3 the areas $A_1$, $A_2$, $A_3$, and $A_4$ associated with the four hospital $H_1$, $H_2$, $H_3$, and $H_4$, respectively are not equal. Numerical results showed that $A_1$, $A_2$, $A_3$, and $A_4$ are approximately given by 0.2, 0.2, 0.3, and 0.3 of the total area. Since it is assumed that the city inhabitants are uniformly distributed, the patient arrival rates to the four hospital $H_1$, $H_2$, $H_3$, and $H_4$, can be expressed as 0.2, 0.2, 0.3, and 0.3 of the total patient arrival rate, respectively. In order to balance the load at the four hospitals, the capacity is assumed to be proportional to the arrival rate. Specifically, the capacity of $H_1$, $H_2$, $H_3$, and $H_4$ are set to $M_1 = 4$, $M_2 = 4$, $M_3 = 6$, and $M_4 = 6$, respectively. The numerical values of the remaining system parameters are listed in Table 2. Fig. 4 depicts the performance metrics versus the patient arrival rate. It is evident that analytical results are in a very well agreement with the simulation results, which verifies the correctness and accuracy of the results. We also can observe that the proposed system and the traditional system have the same performance. This is because the hospital capacity is proportional to the arrival rate which balances the load over the four hospitals.

Also, traffic conditions are assumed to be homogenous over the city. Therefore, there is no room for improvement since the traditional ambulance can be considered optimal in this case. In addition, it can be seen that ambulance patients have shorter delay to get treatment (door-to-needle time) compared to walk-in patients due to the higher priority of the former.

### B. CAPACITY NOT PROPORTIONAL TO THE ARRIVAL RATE

Unlike the previous subsection, here we assume that the capacities of the four hospitals are not proportional to the arrival rates. Specifically, the capacity of $H_1$, $H_2$, $H_3$, and $H_4$ are set to $M_1 = 3$, $M_2 = 3$, $M_3 = 7$, and $M_4 = 7$, respectively, while the traffic arrival rates and other parameters are kept as in subsection (a). Fig. 5 shows the performance metrics versus the patient arrival rate. It is evident that our proposed smart system improves the performance and significantly reduces the delay especially at high arrival rate. This is because our smart algorithm almost kept the performance metric unchanged (very close to the optimal values in the previous subsection) while the performance of traditional system is negatively affected by the unbalanced distribution of the capacity. The performance of our smart algorithm was almost unchanged because it directs the patients to the least loaded hospital that is equivalent to parallel queues with the

**TABLE 2.** System parameters.

| *Parameter* | *Numerical Value* |
|---|---|
| $\lambda_a = \lambda_w$ | 5, 10, 15, 20, and 22.5 Patients/hour |
| $S$ | 4 |
| $N$ | 4 |
| $T_T$ | 24 minutes |
| $K_i$ | 5 |
| $v_a = v_w$ | 40 km/h for inner roads & 70 km/h for main roads |
| $\tau_l$ | 9 sec. for stop signs and 30 sec. for traffic lights |

shortest queue first route policy, which is an optimal policy and equivalent to a single queue with multiple servers using a FIFO policy.

As mentioned in the Introduction section, authors in [18] proposed an algorithm to minimize the drop-off time (waiting time of ambulance patients). However, they didn't consider the response time and the delivery time. Therefore, we compare the performance of our algorithm and the proposed algorithm in [18] in terms of the average drop-off time and the waiting time waiting time of walk-in patients.

Fig. 6 shows the average waiting time of the ambulance patients and walk-in patients for the algorithm proposed in [18] (referred to as "Almehdawe") and our proposed smart algorithm. In both cases (ambulance patients and walk-in patients) our algorithm has shorter waiting time compared with Almehdawe's algorithm. As a matter of fact, for the ambulance patient drop-off time our algorithm is slightly better than Almehdawe's algorithm and this is because the latter is designed with the objective of minimizing the drop-off time. The slight improvement of our algorithm (especially at high arrival rate) is due to the fact that our algorithm uses real-time data and directs patients to the hospital that are least crowded, while Almehdawe's algorithm uses optimal routing probability, but it directs patients to hospitals arbitrarily. For instance, if Almehdawe's algorithm uses optimal routing probabilities of 20%, 20%, 30%, and 30% for the four hospitals $H_1$, $H_2$, $H_3$, and $H_4$, respectively, the dispatcher will direct 20% of the patients to $H_1$ regardless of the instantaneous loading at the hospital at this particular time instant.

In addition, the improvement in the waiting time of walk-in patients is significantly higher. For example, our algorithm reduces the average waiting-time from 26 minutes to 12.78 minutes. This is because our algorithm allows walk-in patients to access the instantaneous loading of hospitals in terms of the expected average waiting time (e.g., using a mobile app) while Almehdawe's algorithm does not direct walk-in patients to hospitals and they are left to choose hospitals arbitrarily (e.g., the closest hospital).

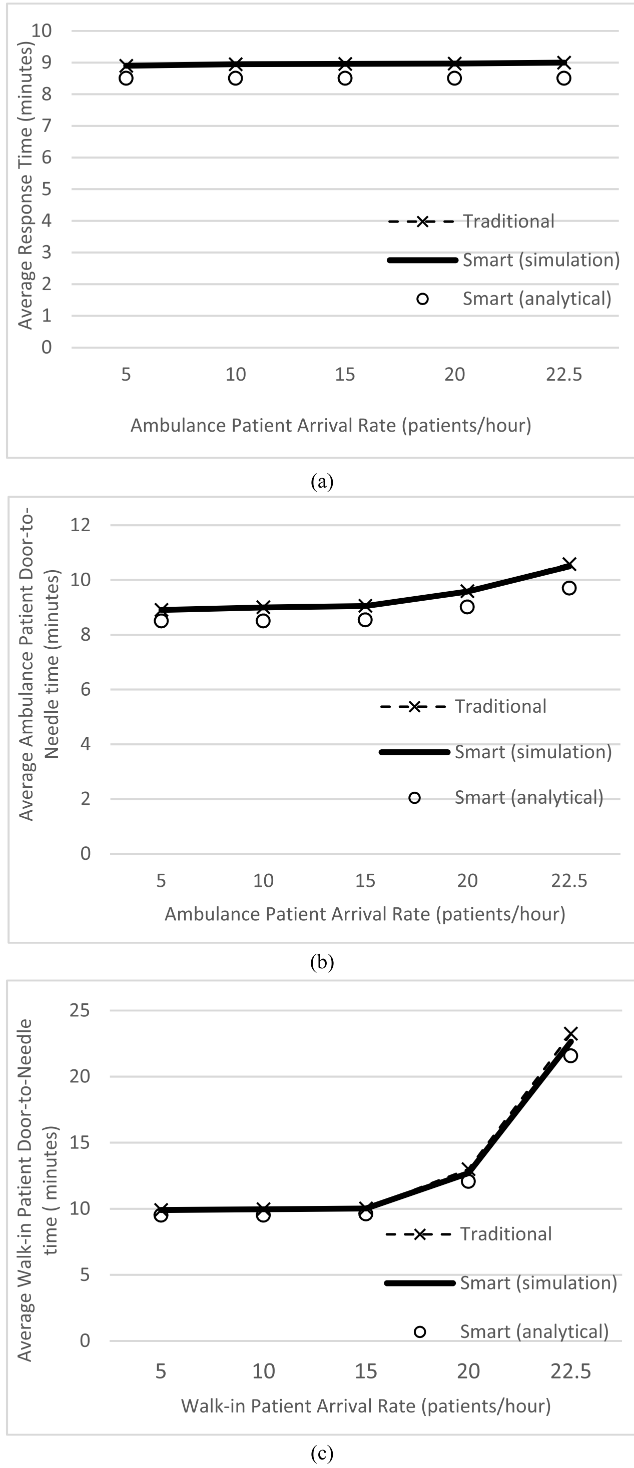
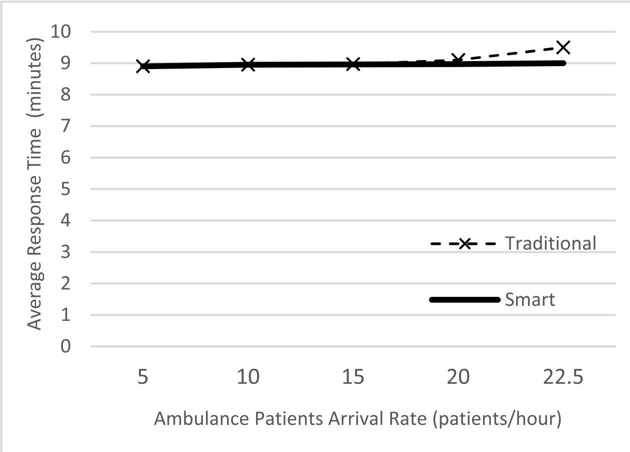


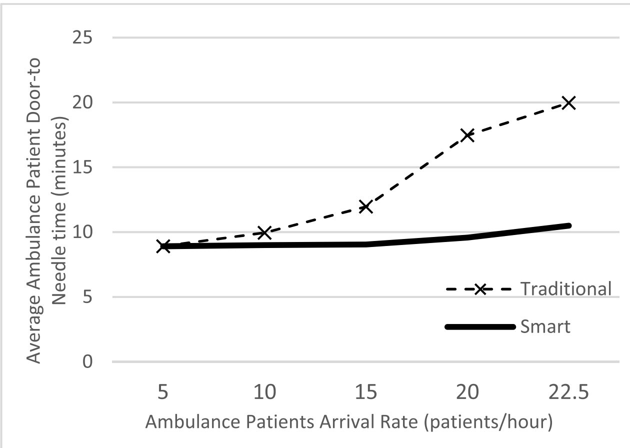

**FIGURE 4.** (a) Average ambulance response time ($T_R^{avg}$) for proportional capacity. (b). Average ambulance patient door-to-needle time for proportional capacity. (c). Average walk-in patient door-to-needle time for proportional capacity.

## C. COMPARISON WITH PREVIOUS WORK

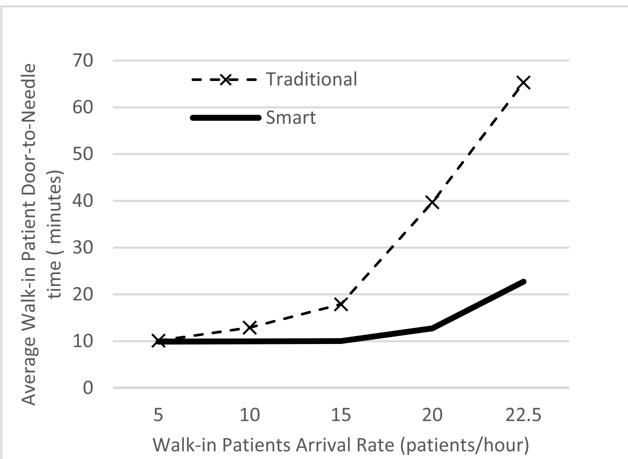
In practice patients' arrival to hospitals changes with time (during the day) and even from a day to day (e.g., weekdays




(a)




(b)




(c)

**FIGURE 5. (a). Average ambulance response time ($T_R^{avg}$) for non-proportional arrival rate. (b). Average ambulance patient door-to-needle time for non-proportional capacity. (c). Average walk-in patient door-to-needle time for non-proportional capacity.**

versus weekends). Fig. 7(a) shows a record of the average ambulance patient arrival rate over a week in a Calgary, Alberta, Canada hospital [26], where the first five days are the weekdays and the last two days are the weekend days. We use this record of the average patient arrival rate to analyze the proposed algorithm and compare its performance with Almedhawe's algorithms under a realistic time-varying average arrival rate. In order to deal with the variation in the arrival rate versus time, we assume that the capacities at the hospitals are adjusted over time so that higher capacity is offered at periods with a high average patient arrival rate and lower capacity is employed at periods with a low average patient arrival rate.

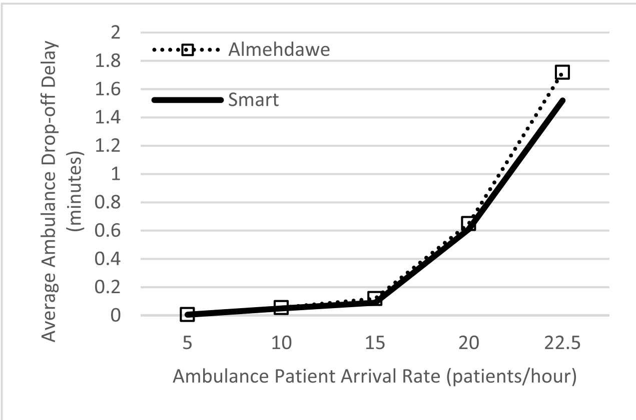
The adaptation of the total capacity (capacity of the four hospitals given by $\sum_{i=1}^{4} M_i$) is depicted in Fig. 7(a). We also assume that the total capacity is distributed over the four hospitals using the same ratio of the arrival rates at the four hospitals. For instance, if the total capacity is set to 10 and the arrival rates of the four hospitals $H_1$, $H_2$, $H_3$, and $H_4$ are 0.2, 0.2, 0.3, and 0.3 of the total arrival rate, respectively, then the capacity of the 4 hospitals are adjusted to $M_1 = 2$, $M_2 = 2$, $M_3 = 3$, and $M_4 = 3$, respectively. Although Almehdawe's algorithm is static where the routing probabilities are once calculated using the long-term average statistical parameters of the emergency systems in the hospitals (e.g., average arrival rate, service rate, etc.) we allow this algorithm to adjust the routing probability every time the capacity is changed at the hospitals (which is assumed to be done 3 times a day). This modification allows Almehdawe's algorithm to be more efficient since it updates the optimal routing probability when there is a significant change in the average patient arrival rate and the capacity.

As shown in Figs. 7(b) and 7(c) our algorithm outperforms the other two algorithms (traditional ambulance and Almehdawe's algorithm). Our proposed smart algorithm achieves smaller waiting time at hospitals for both ambulance patients and walk-in patients. For instance, at 8:00 am of the second day, the average drop-off (ambulance waiting) time is 5.7, 3.78, and 2.8 minutes for traditional ambulance, Almehdawe's algorithm, and our proposed smart algorithm, respectively. At the same time instant, the average walk-in patient waiting time is 42.56, 31.4, and 11.3 minutes for traditional ambulance, Almehdawe's algorithm, and our proposed smart algorithm, respectively.

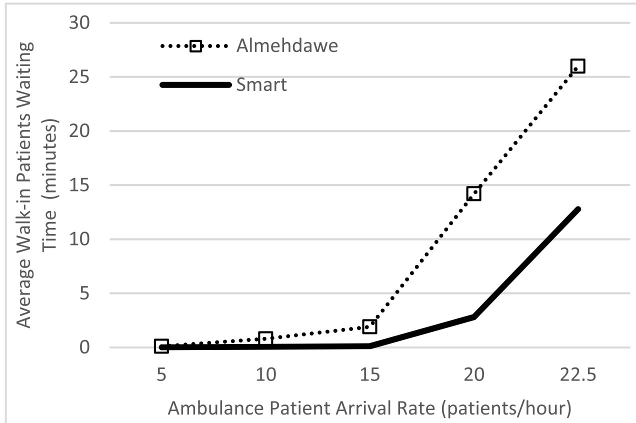
### D. NON-HOMOGENOUS TRAFFIC CONDITIONS

In the previous results we assumed that the traffic conditions are homogenous and uniform across the city. However, in reality, traffic congestion happens due to car accidents, construction work, variation in road capacities, etc. In this section, we show the results of our algorithm and the traditional ambulance taking traffic congestion into consideration. We don't take traffic congestion due to rush hours since it usually affects most of the city areas. We rather take into consideration congestion due to car accidents and short-term construction work. We use the parameters given in Table 3 to model the congestion across the city.

Fig. 8 shows the response time, door-to-needle of ambulance patients, and door-to-needle of walk-in patients versus time under non-homogenous traffic conditions. As depicted




(a)




(b)

**FIGURE 6. (a). Average ambulance drop-off time ($T_{WT\_a}^{avg}$) for non-proportional arrival rate. (b). average waiting time ($T_{WT\_w}^{avg}$) for non-proportional arrival rate.**

**TABLE 3. Spatial-temporal parameters of the road congestion model.**

| *Parameter* | *Description/Numerical Value* |
|---|---|
| Spatial Distribution | Uniform |
| Temporal Distribution | Poisson with a rate of Λ |
| Accident Rate (Λ) | 2 accidents/day (main roads) and 1 accident/day (internal roads) |
| Road Speed Distribution | Uniform with a range 10-20 km/h (main roads) and 5-10 km/h (internal roads) |
| Congestion Duration Distribution | Exponential with an average duration of 4 hours (main roads) and 2 hours (internal roads) |
| $\lambda_a = \lambda_w$ | 20 Patients/hour |

in the figures, the three performance metrics experience fluctuations due to spatial and temporal variations in traffic conditions. The peaks correspond to congestions periods due to accidents and short-term construction. However, it is apparent that our proposed smart algorithm is able to minimize the negative impact of the congestion compared with the traditional ambulance. For instance, at 3:00 o'clock in the fourth day, our algorithm reduces the response time from 15.61 minutes to 8.92 minutes, the door-to-needle time of ambulance patients from 14.95 minutes to 9.64, and door-to-needle time of walk-in patients from 21.6 minutes to 12.38 minutes. This is due to the fact that our smart algorithm collects road conditions in real-time and selects the routes and the destined hospitals based on the collected information. For instance, hospital $H_1$ can be the closest to a patient who requested an ambulance service. However, due to an accident or short-term construction work in the roads between $H_1$ and the patient's location, the travel time from $H_2$ to the patient is shorter than that from $H_1$. In this case, our algorithm will select $H_2$ to send an ambulance to the patient and deliver him/her back to the hospital while the traditional algorithm would select $H_1$ instead, which increases the response time and travel time.

### E. REDUCED TREATMENT TIME

It was shown in [27] that the use of smart ambulance not only reduces the response and door-to-needle time but also shortens the treatment time. This is due to the fact that smart ambulance allows paramedics in the ambulance to exchange biomedical data, images, and videos (recorded or in real-time basis) with the emergency departments and specialists. The exchanged data, images and videos, as well as the potential intervention by paramedics supervised by specialists, can help in reaching initial diagnoses for the case and treatment preparation at the hospital, which reduces the treatment time. As depicted in Fig. 9, the average ambulance drop-off time ($T_{WT\_a}^{avg}$) and the average walk-in waiting time ($T_{WT\_w}^{avg}$) are considerably shortened due to the reduction in the treatment ($T_T$), especially at high arrival rate ($\lambda_a = \lambda_w$). For instant, at $\lambda_a = 22.5$ patients/hour, $T_{WT\_a}^{avg}$ and $T_{WT\_w}^{avg}$ are reduced by 60% and 80%, respectively for 10% reduction in $T_T$ while $T_{WT\_a}^{avg}$ and $T_{WT\_w}^{avg}$ are reduced by 85% and 95%, respectively for 20% reduction in $T_T$. This significant reduction in the ambulance drop-off and walk-in waiting time is very important for patients with critical conditions and life-threatening conditions. Furthermore, the reduction of the ambulance drop-off time speeds up the return of ambulance to its original hospital and the return to the stand-by condition and ready to pick-up a new patient. As, a result the reduction in $T_T$ does not only reduce the ambulance drop-off time and walk-in waiting-time but also the response time since ambulance availability at hospitals will increase, especially at high arrival rates.

### F. RESULTS OF MADINAH

Fig. 10 shows the Madinah central-area map with eight hospitals (four large hospitals and four small hospitals) and six ambulance centres. Each hospital has five

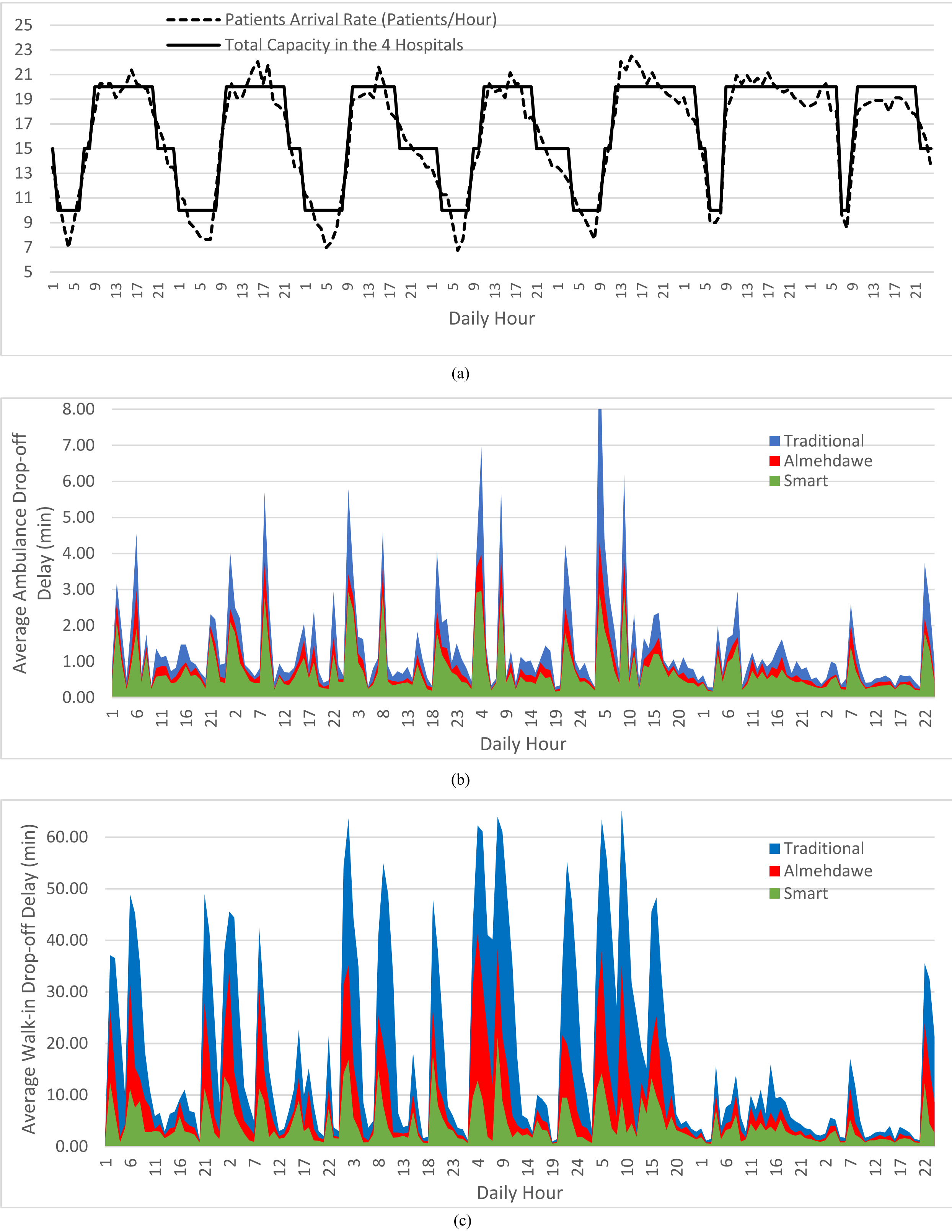
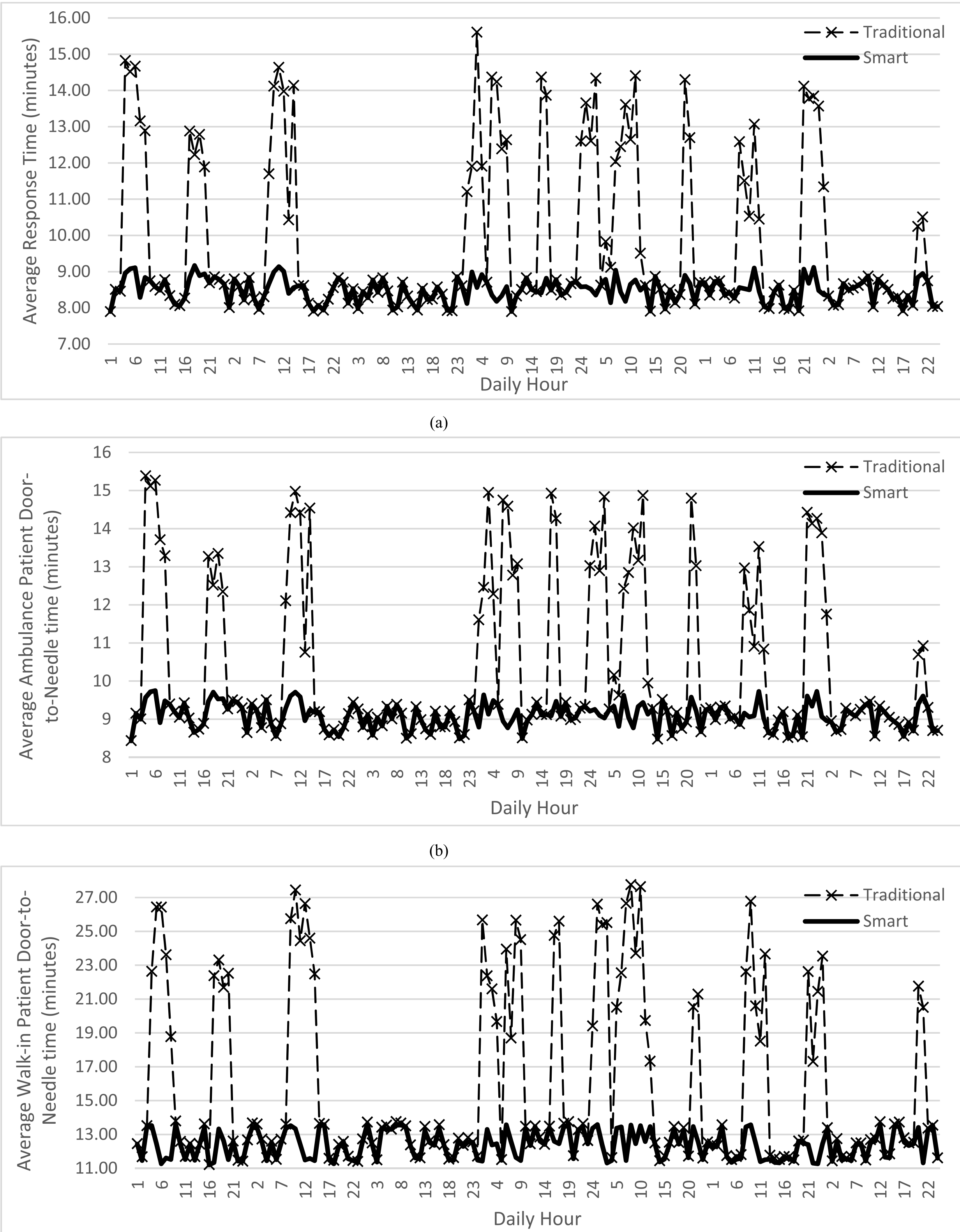



**FIGURE 7.** **(a). Variation of ambulance arrival rate ($\lambda_a = \lambda_w$) and total capacity ($\sum_{i=1}^{4} M_i$) versus time. (b). Average ambulance drop-off time ($T_{WT\_a}^{avg}$) for dynamic arrival rate. (c). Average walk-in waiting time ($T_{WT\_w}^{avg}$) for dynamic arrival rate.**

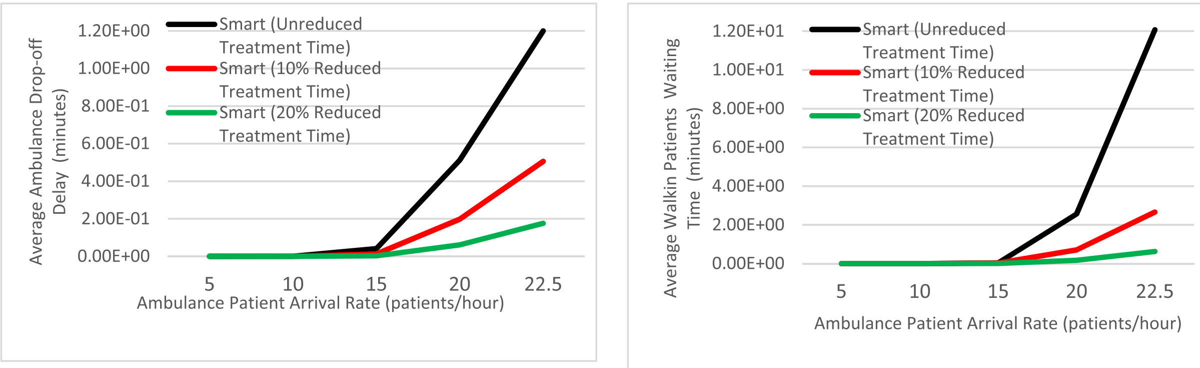



(a)

(b)

(c)

**FIGURE 8.** (a). Average ambulance response time ($T_R^{avg}$) with road congestion. (b). Average ambulance patient door-to-needle time with road congestion. (c). Average walk-in patient door-to-needle time with road congestion.




(a) (b)

**FIGURE 9.** **(a). Impact of treatment time reduction on the average ambulance drop-off time ($T^{avg}_{WT\_a}$). (b) Impact of treatment time reduction on the average walk-in waiting time ($T^{avg}_{WT\_w}$).**

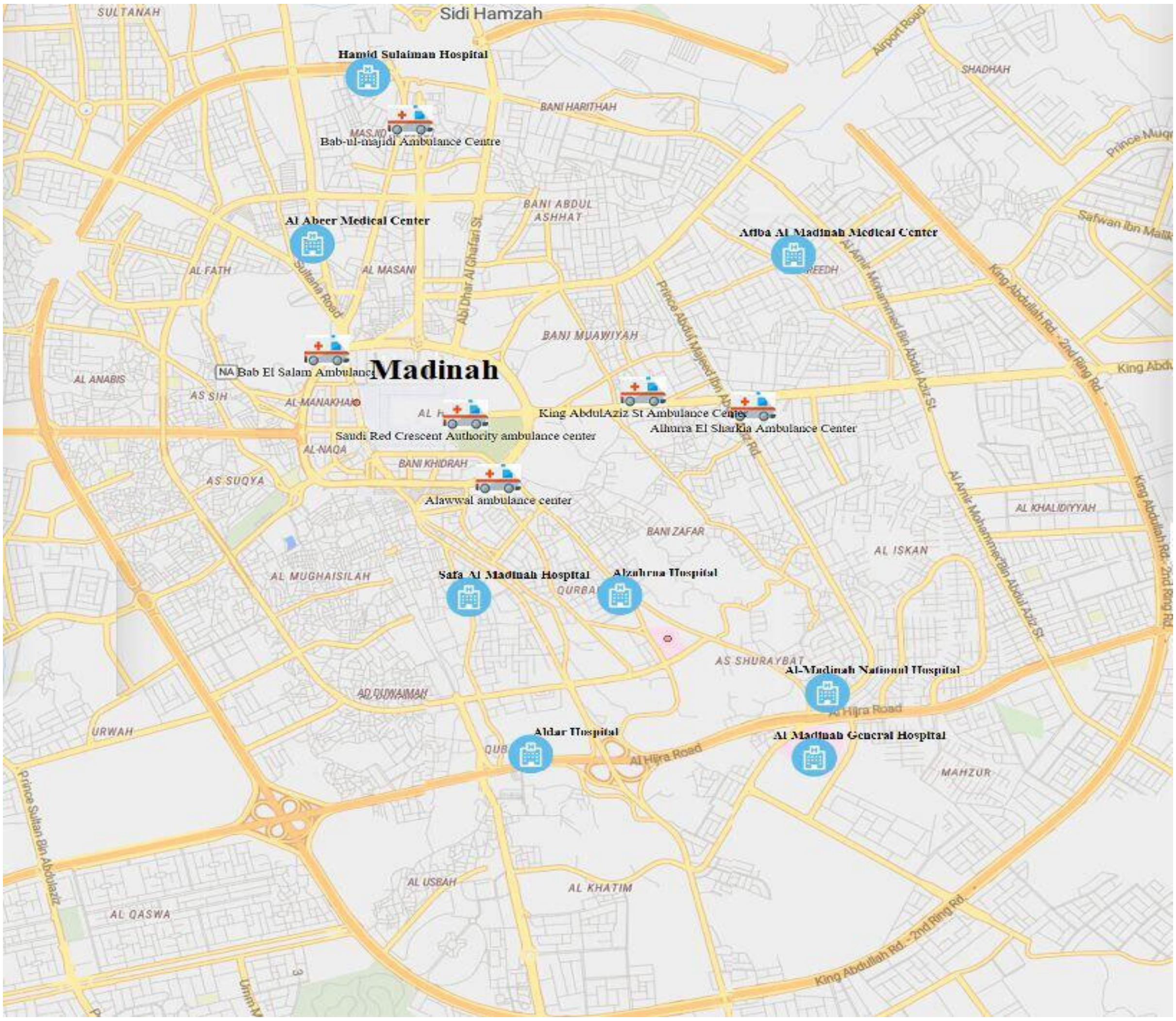
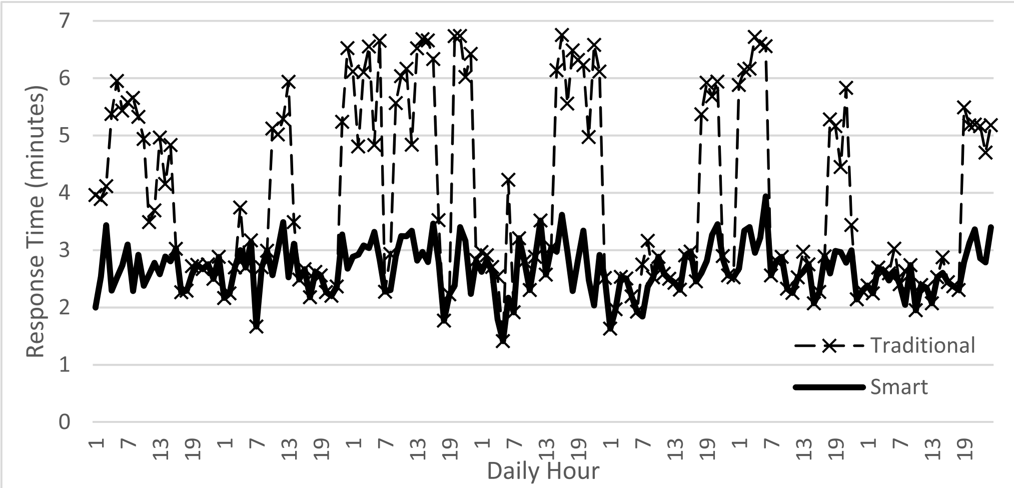
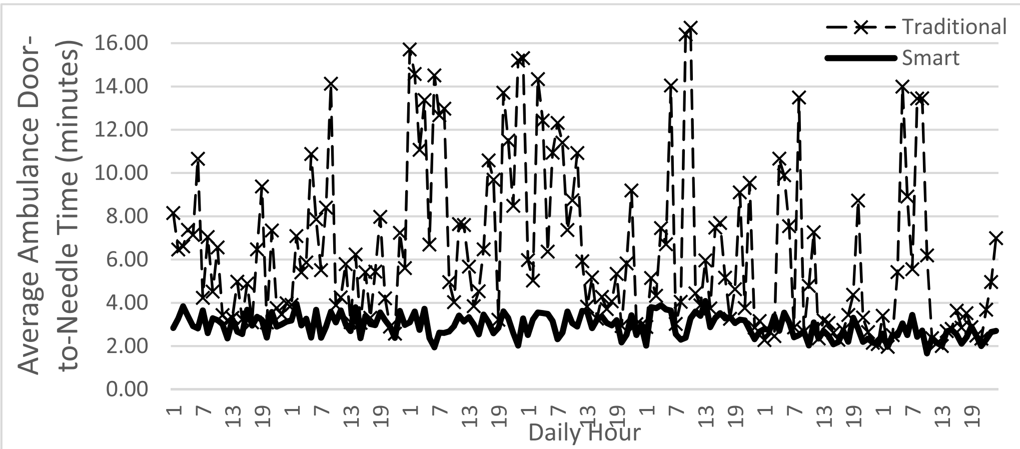



**FIGURE 10.** **Madinah (central-area) map with highlighted hospitals and ambulance centres used in Anylogic simulation.**

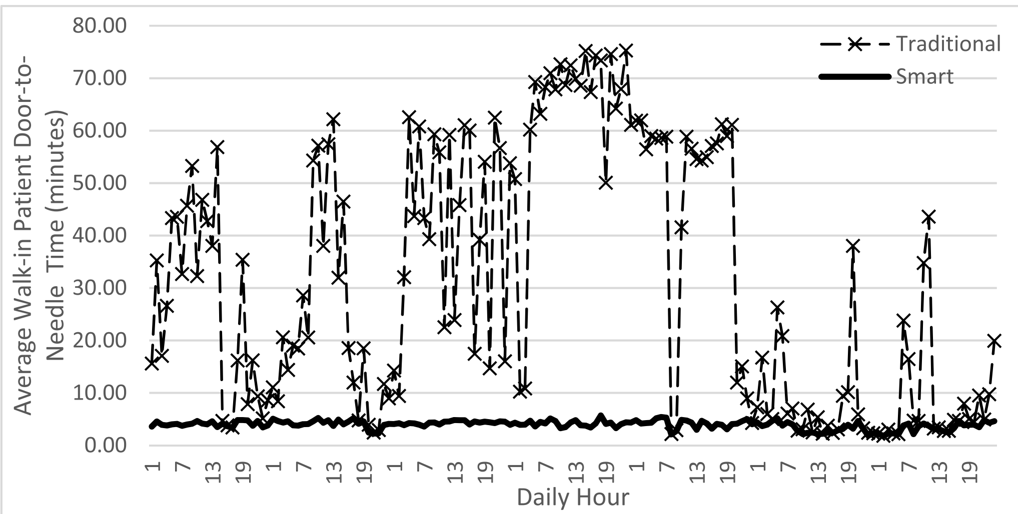
ambulance vehicles while each ambulance centre has three ambulance vehicles. The eight hospitals are classified as four large ones (Al-Madinah General, Alzahraa, Al-Abeer, and Attba Al-Madinah Medical Centre) and four small ones (Safa Al-Madinah, Aldar, Madinah National, and Hamid Suliman). In this section, we include the time-varying




(a)




(b)




(c)

**FIGURE 11. (a). Average ambulance response time ($T_R^{avg}$) in madinah central area. (b). Average ambulance patient door-to-needle time in madinah central area. (c). Average walk-in patient door-to-needle time in Madinah central area.**

arrival rate and capacity similar to that in Fig. 7(a). However, the total capacity is twice that shown in Fig. 7(a) since the capacity of large hospitals are 6, 4, and 3 at high arrival rate, medium arrival rate, and low arrival rate, respectively, while small hospitals have 4, 3, and 2 at high arrival rate, medium arrival rate, and low arrival rate, respectively. In addition, we take into consideration the traffic congestion due to accidents and short-term construction work as discussed in Subsection (d). We consider three speeds (110 km/h for circular road King Abdullah Rd, 80 km/h for other main roads, and 50 km/h for the inner roads. Fig. 11 shows the performance metrics versus time over one week. As in the virtual city case, our proposed smart algorithm significantly improves the three performance metrics. It is evident that our smart algorithm reduces the average response time by more than 50%, the average ambulance patient door-to-needle time by more than 75%, and the average walk-in patient door-to-needle time by more than 95%.

## V. CONCLUSION

In this paper we proposed a novel smart ambulance system which is shown to significantly improve the ambulance performance metrics. The proposed system utilizes the real-time information about the hospital loading and road traffic conditions.

We analyzed the performance of the proposed system analytically and by simulation. The results showed good agreement between the analytical results and simulation results. We also compared the performance of the proposed smart systems with that of traditional ambulance (without real-time traffic information utilization) and the results confirmed the superiority of our proposed smart algorithm. In addition, we compared the performance of our algorithm with that of one of the existing algorithms in the literature which minimizes the ambulance drop-off time (Almehdawe's algorithm [18]) and we showed that our algorithm outperforms Almehdawe's algorithm. Furthermore, we took into consideration the treatment time reduction due to the exchange of data, image and video data between ambulance and emergency department before patient arrival. The performance superiority of our proposed smart algorithm is verified in two environments (circular virtual city and the central-area in Madinah, Saudi Arabia). From the obtained results, it is apparent that smart ambulance is a very promising technology that can revolutionize the emergency service.

We plan to extend this work to take into consideration the different urgency levels of patients and the availability of some specialists in specific hospitals. More consideration of reducing the treatment time will be taken into account by adopting smart diagnostic model. Such a model will be responsible to provide emergency departments with initial checkup of patients such as heartbeat rate, blood pressure measure, glucose level, and body temperature.

## ACKNOWLEDGMENT

The authors would like to thank the King Fahad University of Petroleum and Minerals for their help with this research.

## REFERENCES

[1] H. K. Mell, S. N. Mumma, B. Hiestand, B. G. Carr, T. Holland, and J. Stopyra, "Emergency medical services response times in rural, suburban, and urban areas," *JAMA Surg.*, vol. 152, no. 10, pp. 983–984, Oct. 2017.

[2] R. P. Gonzalez, G. R. Cummings, H. A. Phelan, M. S. Mulekar, and C. B. Rodning, "Does increased emergency medical services prehospital time affect patient mortality in rural motor vehicle crashes? A statewide analysis," *Amer. J. Surg.*, vol. 197, no. 1, pp. 30–34, Jan. 2009.

[3] A. Ashour, P. Cameron, S. Bernard, M. Fitzgerald, K. Smith, and T. Walker, "Could bystander first-aid prevent trauma deaths at the scene of injury?" *Emergency Med. Australasia*, vol. 19, no. 2, pp. 163–168, Apr. 2007.

[4] T.-H. Tan, M. Gochoo, Y.-F. Chen, J.-J. Hu, J. Chiang, C.-S. Chang, M.-H. Lee, Y.-N. Hsu, and J.-C. Hsu, "Ubiquitous emergency medical service system based on wireless biosensors, traffic information, and wireless communication technologies: Development and evaluation," *Sensors*, vol. 17, no. 12, p. 202, Jan. 2017.

[5] C. O'Keeffe, J. Nicholl, J. Turner, and S. Goodacre, "Role of ambulance response times in the survival of patients with out-of-hospital cardiac arrest," *Emergency Med. J.*, vol. 28, no. 8, pp. 703–706, Aug. 2011.

[6] K. Tokuda and S. Ohmori, "Experiments of SAFER (speedy ambulance first-aid, emergency, rescue operations supporting) system," in *Proc. 14th Int. Symp. Wireless Pers. Multimedia Commun.*, Brest, France, Oct. 2011, pp. 1–5.

[7] C. S. Lim, R. Mamat, and T. Braunl, "Impact of ambulance dispatch policies on performance of emergency medical services," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 624–632, Jun. 2011.

[8] B. Jäger, P. M. Haller, E. Piackova, A. Kaff, G. Christ, W. Schreiber, F. Weidinger, T. Stefenelli, G. Delle-Karth, G. Maurer, and K. Huber, "Predictors of transportation delay in patients with suspected ST-elevation-myocardial infarction in the VIENNA-STEMI network," *Clin. Res. Cardiol.*, vol. 109, no. 3, pp. 393–399, Mar. 2020.

[9] A. Dumka and A. Sah, "Smart ambulance system using concept of big data and Internet of Things," in *Healthcare Data Analytics and Management*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 155–176, doi: 10.1016/B978-0-12-815368-0.00006-3.

[10] X. James. (Apr. 2021). *Smart Ambulances: The Future of Emergency Healthcare*. The Healthcare Guys. [Online]. Available: https://www.healthcareguys.com/2021/04/01/smart-ambulances-the-future-of-emergency-healthcare/

[11] M. Al-khafajiy, H. Kolivand, T. Baker, D. Tully, and A. Waraich, "Smart hospital emergency system: Via mobile-based requesting services," *Multimedia Tools Appl.*, vol. 78, no. 14, pp. 20087–20111, Jul. 2019.

[12] Y. Zhai, X. Xu, B. Chen, H. Lu, Y. Wang, S. Li, X. Shi, W. Wang, L. Shang, and J. Zhao, "5G-network-enabled smart ambulance: Architecture, application, and evaluation," *IEEE Netw.*, vol. 35, no. 1, pp. 190–196, Jan. 2021, doi: 10.1109/MNET.011.2000014.

[13] T. Akca, O. K. Sahingoz, E. Kocyigit, and M. Tozal, "Intelligent ambulance management system in smart cities," in *Proc. Int. Conf. Electr. Eng. (ICEE)*, Sep. 2020, pp. 1–7.

[14] S. Djahel, N. Smith, S. Wang, and J. Murphy, "Reducing emergency services response time in smart cities: An advanced adaptive and fuzzy approach," in *Proc. IEEE 1st Int. Smart Cities Conf. (ISC)*, Oct. 2015, pp. 1–8, doi: 10.1109/ISC2.2015.7366151.

[15] A. Karkar, "Smart ambulance system for highlighting emergency-routes," in *Proc. 3rd World Conf. Smart Trends Syst. Secur. Sustainablity (WorldS)*, Jul. 2019, pp. 255–259.

[16] M. Li, A. J. E. Carter, J. Goldstein, T. Hawco, J. Jensen, and P. Vanberkel, "Determining ambulance destinations when facing offload delays using a Markov decision process," *Omega*, vol. 101, pp. 1–14, Mar. 2020.

[17] E. Almehdawe, B. Jewkes, and Q.-M. He, "A Markovian queueing model for ambulance offload delays," *Eur. J. Oper. Res.*, vol. 226, no. 3, pp. 602–614, May 2013.

[18] E. Almehdawe, B. Jewkes, and Q.-M. He, "Analysis and optimization of an ambulance offload delay and allocation problem," *Omega*, vol. 65, pp. 148–158, Dec. 2016.

[19] M. Abdeen, M. H. Ahmed, H. Seliem, M. El Nainay, and T. R. Sheltami, "Improving the performance of ambulance emergency service using smart health systems," in *Proc. Workshop Artif. Intell. Internet Things Digit. Health Conjunct. (IEEE/ACM CHASE)*, Washington DC, USA, Dec. 2021, pp. 205–209.

[20] M. Batty and P. Longley, "The shape of cities: Geometry, morphology, complexity and form," in *Fractal Cities: A Geometry of Form and Function*, San Diego, CA, USA: Academic, 1994, ch. 1.

[21] V. Meena. *City Forms*. Accessed: Sep. 10, 2021. [Online]. Available: https://www.slideshare.net/vjspa/city-forms?next_slideshow=1

[22] D. J. Houck, "Comparison of policies for routing customers to parallel queueing systems," *Oper. Res.*, vol. 35, no. 2, pp. 306–310, Apr. 1987.

[23] Y. Kogan, Y. Levy, and R. A. Milito, ''Call routing to distributed queues: Is FIFO really better than MED?'' *Telecommun. Syst.*, vol. 7, nos. 1–3, pp. 299–312, Jun. 1997.

[24] A. L. Garcia, *Probability, Statistics and Random Processes for Electrical Engineering*, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2008.

[25] D Gross, J. Shortle, J. Thompson, and C. Harris, *Fundamentals of Queueing Theory*. Hoboken, NJ, USA: Wiley, 2008.

[26] N. Channouf, P. L'Ecuyer, A. Ingolfsson, and A. N. Avramidis, ''The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta,'' *Health Care Manage. Sci.*, vol. 10, no. 1, pp. 25–45, 2007.

[27] N. Fearn, ''How connected ambulances could revolutionize patient treatment and reduce hospital numbers,'' Forbes, Jersey City, NJ, USA, Nov. 2019. [Online]. Available: https://www.forbes.com/sites/nicholasfearn/2019/11/18/how-connected-ambulances-could-revolutionise-patient-treatment-and-reduce-hospital-numbers/#56455712498d

**MOHAMMAD A. R. ABDEEN** received the M.S. degree in electrical and computer engineering from the University of Victoria, Canada, and the Ph.D. degree in electrical and computer engineering from the University of Ottawa, Canada. He assumed numerous academic positions in international institutions, including the University of Ottawa, Carleton University, and The British University in Egypt. He was also an Industrial Consultant of Nortel, Alcatel, and ViStar Telecommunications. He is currently an Associate Professor of computer science with the Islamic University of Madinah, Saudi Arabia, and an Adjunct Professor with the School of Electrical Engineering and Computer Science, University of Ottawa. He is leading numerous research projects in the area of smart city and autonomous vehicles. His research interests include machine learning, distributed systems, natural language processing, and data mining. He is a member of the Research Council, Islamic University of Madinah. He is the Editor-in-Chief of the *Journal of Engineering, Science and Computing*.

**MOHAMED HOSSAM AHMED** (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Carleton University, Ottawa, Canada, in 2001. From 2001 to 2003, he was a Senior Research Associate with Carleton University. In 2003, he joined the Faculty of Engineering and Applied Science, Memorial University of Newfoundland, Canada, where he worked as a Full Professor, until December 2019, and has been an Adjunct Professor, since January 2020. He is currently a LTA Professor with the University of Ottawa. He is a Registered Professional Engineer (P.Eng.) in the Province of Newfoundland, Canada. He has published more than 180 papers in international journals and conferences. His research is sponsored by NSERC, CFI, QNRF, Bell/Aliant, and other governmental and industrial agencies. His research interests include radio resource management in wireless networks, multi-hop relaying, cooperative communication, vehicular *ad hoc* networks, cognitive radio networks, and wireless sensor networks. He served as an Editor for IEEE COMMUNICATION SURVEYS AND TUTORIALS, from 2007 to 2018, and a Guest Editor for the Special Issue on Fairness of Radio Resource Allocation of *EURASIP JWCN*, in 2009.

**HAFEZ SELIEM** received the B.Sc. and M.Sc. degrees from Ain Shams University, Cairo, Egypt, in 2009 and 2014, respectively, and the Ph.D. degree in computer engineering from the Memorial University of Newfoundland, Canada, in 2019. Currently, he is a Lecturer with Ain Shams University, Cairo. His research interests include vehicular *ad hoc* networks, mobile *ad hoc* networks, the IoT, big data, and software design.

**TAREK RAHIL SHELTAMI** received the Ph.D. degree in electrical and computer engineering from the Electrical and Computer Engineering Department, Queen's University, Canada. He has been an Adjunct Professor with the Jodrey School of Computer Science, Acadia University, Canada, since 2010. He is currently a Professor with the Computer Engineering Department and a member of the Interdisciplinary Research Center of Smart Mobility and Logistics, King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia. Before joining the KFUPM, he was a Research Associate Professor with the School of Information Technology and Engineering, University of Ottawa, Canada. He worked at GamaEng Inc., as a Consultant on wireless networks (2002–2004). Also, he worked in several joined projects with Nortel Network Corporation. He is currently collaborating with the Transportation Research Institute (IMOB), Hasselt University, Belgium. He authored/coauthored more than 200 research articles. His research interests include *ad hoc* networks, WSNs, the IoT, digitizalition, the IoD, computer network security, fog computing, quality, and software defined networks.

**TURKI M. ALGHAMDI** received the B.Sc. degree in computer science, in 2005, the M.Sc. degree in software engineering, in 2008, and the Ph.D. degree in software engineering from De Montfort University, U.K., in 2012. He is currently an Associate Professor with the Faculty of Computer and Information Systems, Islamic University of Madinah, Saudi Arabia. His research interests include big data analysis, the IoT, wireless sensor networks, and IS policy management.

**MUSTAFA EL-NAINAY** (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in computer engineering from Alexandria University, in 2001 and 2005, respectively, and the Ph.D. degree in computer engineering from Virginia Tech, in 2009. He is currently on leave and affiliated with the Faculty of Computer Science and Engineering, AlAlamein International University. He is also a Professor of computer engineering with the Computer and Systems Engineering Department, Alexandria University, Egypt. His focus is on the use of artificial intelligence to solve problems in different domains, including communications and networking, indoor localization, and software engineering. His research interests include wireless and mobile networks, cognitive radio and cognitive networks, and software testing automation and optimization. He has served as a reviewer, a TPC member, and the TPC chair/the track chair for various international journals and conferences.