

## Research Article

# Cooperative Hypercube Queuing Model for Emergency Service Systems

Han Liu <sup>1,2</sup>, Hao Yin <sup>3</sup>, Yang Zhou <sup>4</sup>, and Meng Li <sup>1</sup>

<sup>1</sup>Department of Civil Engineering, Tsinghua University, Beijing 100084, China

<sup>2</sup>The Research Institute of Emergency Science, The Central Research Institute, China Coal Technology and Engineering Group (CCTEG), Beijing 100013, China

<sup>3</sup>SF Technology Co., Ltd., Shenzhen 518000, China

<sup>4</sup>College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China

Correspondence should be addressed to Meng Li; [mengli@tsinghua.edu.cn](mailto:mengli@tsinghua.edu.cn)

Received 11 November 2020; Revised 9 December 2020; Accepted 18 December 2020; Published 8 January 2021

Academic Editor: Yong Wang

Copyright © 2021 Han Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

As a useful descriptive tool for emergency service effectiveness, the hypercube queuing model has been applied in systems of many countries, such as the SAMU system in Brazil. However, the traditional hypercube queuing model and its extended forms assume that the service provider performs independent services, lacking a compelling description of the situation where emergency vehicles perform cooperative services (e.g., NEPPHE in China). To this end, we assume that vehicles in the same fleet simultaneously start and end services and propose a cooperative hypercube queuing (CHQ) model that can describe the state of emergency systems which apply multivehicle dispatches. In order to verify the accuracy of the model, we apply Arena simulation software in Wuhan case. The results show that the CHQ model can illustrate cooperative performance effectively. Sensitivity analyses under more general parameters are conducted to reveal insights into the model application.

## 1. Introduction

The emergency system's rapid response plays a crucial role in the mitigation and control of disasters. Especially for catastrophic events, response delay will bring about irreparable losses to people's lives and property. However, a lot of these incidents require several types of vehicles to respond together to complete the rescue mission [1], which casts a big challenge if the required vehicles have a long way to dispatch. In urban fire rescue, major hazardous chemical fires require the cooperation of water tank fire trucks, foam fire trucks, and biochemical fire trucks; for skyscraper fires, lifting jet trucks are also indispensable vehicles among various fire trucks. For example, the Beirut explosion in 2020 and the fire in the China Central Television Building in 2009 require various fire vehicles in cooperative fleets [2, 3]. The applicable scenarios of cooperative response are not limited to the field of fire rescue but also are widely seen in medical assistance. For example, Covid-2019 has brought about huge losses to people

worldwide [4, 5]. In high-risk areas, frequent emergency medical needs are usually accompanied by social incidents of quarrels and provocations. Therefore, police cars need to clear the way for ambulances, and sometimes they need to be escorted by military armed vehicles to ensure the smooth completion of the mission [6]. As the cooperation-needed disaster scales and diversities continue to develop, capturing a complex cooperative system's emergency performance poses a significant challenge to practitioners and researchers. In particular, according to the system vehicle's configuration with professional functions or purposes, evaluating cooperative service effectiveness is essential for discovering and designing better configuration plans.

In recent decades, deterministic and probabilistic models on the emergency facility location have been studied. The location set covering problem (LSCP) model by Toregas et al. [7] and the maximum coverage location problem (MCLP) introduced by Church and ReVelle [8] are two classic models that regard the emergency system as static and deterministic.

The former model assumes that resources are unlimited and strategically determine the minimum number of facilities for entire demand coverage. The latter assumes that resources are limited and aims to make their full use for maximal demand coverage. Based on these two classic models, researchers have proposed extended models in line with the emergency field's reality. Marianov and ReVelle extended LSCP model to a probabilistic version by adding an availability/busy constraint [9]. Daskin (1983) combined vehicle availability with the MCLP model and proposed the maximum expected coverage location model [10]. Erkut et al. incorporated the survival function (i.e., a monotonically decreasing function of survival rate with response time) into the coverage model and constructed the maximum survival location problem (MSLP) [11]. Gendreau et al. considered the characteristics of different coverage radii of various facilities and proposed the dual-standard coverage model (DSM) [12]. The optimization result of facility locations ensures that at least a certain percentage of demand points are located within the smaller standard radius while locating all demand points within the larger standard radius. Wang et al. proposed a logistics location-routing problem under a cooperation mechanism for the benefit and risk-sharing [13]. Later, a collaborative multiple center vehicle routing (CMCVR) problem based on a state-space-time network was presented by designing a suitable profit distribution mechanism [14, 15]. He et al. proposed a two-stage stochastic programming model for the last mile delivery system [16]. Motivated by China's fire emergency services mechanism, Wang et al. proposed a strategic location model for coordinated rescue facilities by considering the impact of the vehicle number and distance on the progressive coverage level [17]. Yufeng Zhou et al. proposed a hybrid genetic algorithm for facility location and relief material transportation problem on the background of the 5.12 Wenchuan Earthquake [18].

For decades, scholars have also conducted numerous academic researches on spatial queuing theory. The hypercube queuing model, developed by Larson [19], is a useful descriptive tool for emergency system performance analysis. It considers geographical and temporal complexities in a server-to-customer system under consideration and illustrates server state transactions under Markovian equilibriums. Based on Larson's work, Jarvis [20] proposed an approximation model by determining servers' busy probabilities in their specific location. Chelst and Barlach [21] extended a multiple unit dispatches model for hypercube queuing services. Their model avoided the exponential growth of problem complexity and thus did not identify each server's actual state for each type of call. Some extensions of the hypercube model have been applied in SAMU emergency system in Brazil [22–25]. For example, Iannoni and Morabito [25] proposed a hypercube queuing model under multiple dispatches and partial backup policies. However, these models did not allow simultaneous transition of cooperative servers from busy states to available states, which may not provide sufficient insight into the hypercube model application in cooperation scenarios.

To the best of our knowledge, hypercube queuing equilibrium under emergency server cooperation has not

been carefully studied. The impact of the supply/demand relationship under the cooperative preference design has generally been ignored in the hypercube queuing literature. Although the classic hypercube queuing model has been widely used in emergency systems in many countries, it may result in inaccurate operational estimations when applied in severe public health incidents such as the Covid-2019 virus outbreak, where necessary cooperative policies are frequently adopted. In light of this gap, we extend the hypercube model under cooperative service equilibrium. Our model relaxes the traditional assumptions that servers dispatched to the same call are independent with different mean service time by assuming that servers in the same fleet become free simultaneously when they finish a cooperative service. The state probabilities in some traditional models are eliminated, which helps reduce the computational complexity of the model. Moreover, our model adds more necessary states for differentiating cooperative server states among their possible combinations, thus providing more accurate dispatching details. The cooperative hypercube model's validation is conducted under different supply/demand scenarios to provide practical insight.

The remainder of this paper is organized as follows. Section 2 introduces the cooperative rescue mechanism by taking an emergency system in Wuhan as an example. Section 3 develops a novel spatial queuing model under cooperative hypercube equilibrium and conducts the model validation through a simulation method. In Section 4, we conduct parameter sensitivity analysis to explore the model application under various scenarios. Finally, Section 5 provides concluding remarks and future research directions.

## 2. Cooperative Dispatching Preferences

The emergency service system is typically modelled as a zero-capacity waiting line system [25]. The continuous-time Markov chain in the hypercube queuing model can be described as server-status vertices. In the basic hypercube queuing model, the total number of possible states is  $2^N$ , since each server must be the only one of two states (idle (0) or busy (1)). For example, a four-server system generates 16 possible states according to the hypercube vertices; the state {1010} denotes that the first server and the third server are busy and the other two servers are idle. Based on the aggregation of similar events in adjacent locations, the service area can be divided into subareas named atoms in the literature. Each atom is modelled as a single point (marked as  $j_1, j_2, \dots$ ) in the atom's central area to indicate where the emergency call occurs. Emergency calls at each geographical atom generate in a Poisson process with mean arrival rates  $\lambda_j$ , independent from other atoms. The service time of server follows a negative exponential distribution. Table 1 shows the dispatching preference lists for a noncooperative system (e.g., SAMU system in Brazil), which can be described by the classic hypercube model. Taking Type 2 calls (calls at Atom 2) as an example, while the SAMU system also applies the double dispatch policy, there are three independent candidates in its preference list. When it comes to a Type 2 call that requires double dispatch, the first and second preferred

TABLE 1: Noncooperative dispatching preference lists.

Atom	Dispatch policy	Type of call	First	Second	Third	Fourth
1	A	1	V1	V3	—	—
2	A or B	2	V2	V1	V3	—
3	A or B or C	3	V4	V2	V3	V1

servers (V2 and V1) are simultaneously dispatched; if any one of the two servers is busy, the third one (V3) would be dispatched as a backup one; moreover, if any two servers of the three candidates are both busy, the only idle one would be dispatched as a single dispatch. If the three candidates are all unavailable, the call is lost. One can observe that SAMU's policy may not be suitable for major disasters, which should be simultaneously coped with by multiple types of servers. See [25] for more details.

However, some severe demand has a chance to be saved if and only if vehicles of all essential types are available. Taking China's emergency system as an example, a high-level hazard demand (e.g., a virus outbreak, nuclear power plants, or explosive energy facility) requires specific combinations of professional emergency vehicles (e.g., advanced ambulance, police vehicles, and fire-fighting turntable ladders) to form a cooperative rescue fleet. The essential responder types in cooperation for serving catastrophic demand are confirmed by the National Emergency Plan for Public Health Emergencies (NEPPHE) [1] of China. It describes incidents' physical or chemical properties, including virus-affected patients and other rescue events, and specifies necessary responders' cooperative service standards. If any type of required vehicle stipulated in the NEPPHE is not available, the specific demand's cooperative mission cannot go through.

Motivated by the cooperative mechanism (e.g., NEPPHE) for coping with major disasters, we extend the classic hypercube queuing model to the cooperative hypercube queuing (CHQ) model. We assume that cooperative servers in the same fleet operated by NEPPHE become free simultaneously when they finish a cooperative service. Call types are classified into three levels according to the incident risks: low-risk calls, medium-risk calls, and high-risk calls. The descriptions of dispatching preference policies are given in Table 2:

- (i) A low-risk call requires a single dispatch of an ambulance. For example, to provide basic emergency care, one ambulance is highly competent at this work.
- (ii) A medium-risk call requires the dispatched fleet, including servers of two types (an ambulance and a police car together). When a medium-risk call arrives, the fleet composed of the nearest ambulance and a police car will serve as the first preferred fleet; if either one is busy, then a further available server of the absent type would be selected to formulate the second preference fleet; otherwise, the call is lost to other systems.

TABLE 2: Cooperative dispatching preference lists for NEPPHE, China.

Atom	Dispatch policy	Type of call	First vehicle	Second vehicle
1	Type A	1	V1	V3
2	Types A & B	2	V1 & V2	V3 & V2
3	Types A & B & C	3	V3&V2&V4	V1&V2&V4

- (iii) A high-risk call requires cooperative servers of three types (an ambulance, a police car, and a fire engine). Calls of this type respond to the highest-level severity. For example, in a terrorist attack or a radioactive material accident, the combinatory fleet of three-type vehicles is necessary for a response. When a high-risk call arrives, the fleet composed of the nearest ambulance, a police car, and a fire engine would be dispatched as the first preferred fleet. If any one of the three nearest vehicles is busy, the further available vehicle, which is the same type as the busy one, would be combined with the other available vehicles for the coordinated response. If there is no suitable vehicle to substitute the busy one, the call is lost.

In this way, the traditional model assumption that servers dispatched to the same call are independent (with different mean service time) is relaxed. The suitable cooperative fleet is assigned to the corresponding call before identifying essential server status.

To capture the inherent mechanism of CHQ model, we apply the model in an emergency system in Wuhan (China) according to the released data from the Health Commission of Hubei Province [26]. Figure 1 shows a real network layout of three atoms and four vehicles (i.e., two ambulances: V1 and V3 with service rates  $\mu_1 = 0.86$  and  $\mu_3 = 0.73$ , respectively; one police car: V2 with service rate  $\mu_2 = 1.16$ ; and one fire engine: V4 with service rate  $\mu_4 = 1.37$ ) hosted in three emergency stations that share the same locations of the atoms. The ambulances V1 and V3 are hosted at emergency stations 1 and 3, respectively. The police car V2 is at emergency station 2, and the fire engine V4 is at emergency station 3. According to the free travel time from each station to each atom (AMAP data), the preference list is as follows:

Type 1 calls with arrival rate  $\lambda_1 = 0.13$  require the single-Type A dispatch, and there are two Type-A candidate servers in NEPPHE dispatching preference list (the nearest server V1 as the first preference and the further V3 as the second preference). When a Type 1 call arrives, V1 takes the response; if V1 is already in the busy state, V3 will be dispatched as a backup server, although the further travel distance may cause delay; if both V1 and V3 are busy, the call is lost.

Type 2 calls with arrival rate  $\lambda_2 = 0.28$  require the double-type (e.g., one Type A server and one Type B server) dispatch, and there are two candidate fleets (the combinatory fleet of V1 and V2 is prior to that of V3

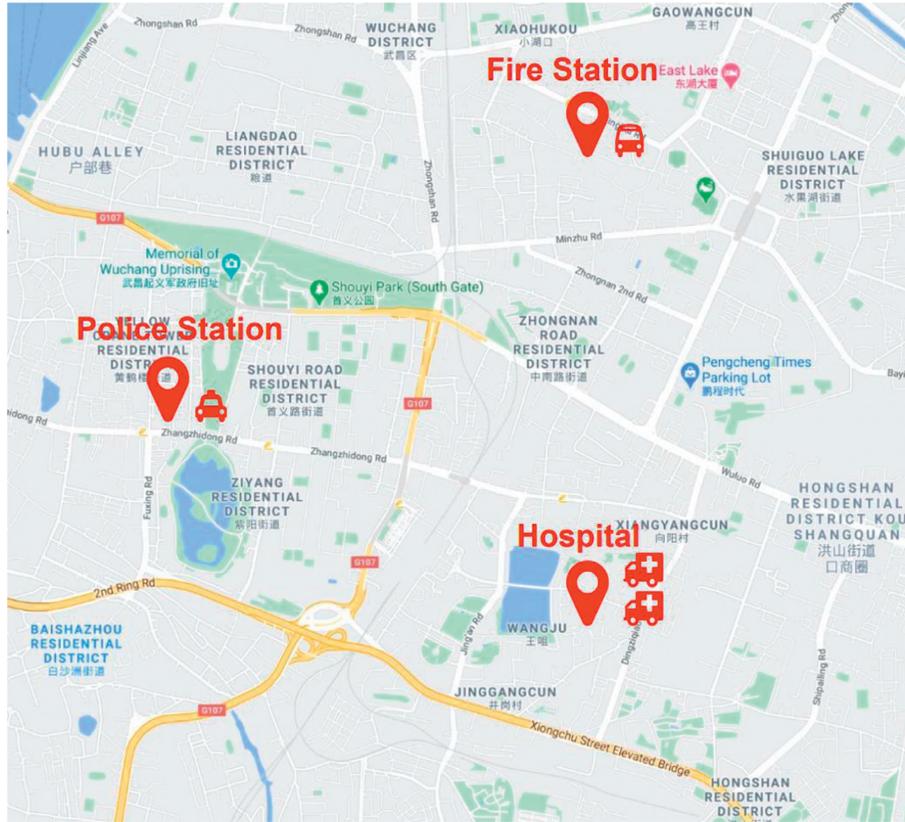


FIGURE 1: A real network with three atoms and four servers in urban areas of Wuhan.

and V2 because the nearer server in a candidate fleet is preferred) in the NEPPHE list.

Type 3 calls with arrival rate  $\lambda_3 = 0.09$  require the triple-type (e.g., one Type A server, one Type B server, and one Type C server) dispatch. There are also two candidate fleets, and each fleet contains three essential types of servers in the NEPPHE list. For a Type 3 call, the first fleet of V2 & V3 & V4 is dispatched; if V3 in this fleet is busy, the further Type 1 vehicle (i.e., V1) will be dispatched together with V2 & V4, which may also cause a delay in the cooperative emergency service; if the two candidate fleets are not busy, the call is lost.

Compared to SAMU, NEPPHE attaches great importance to cooperative characteristics in emergency services because essential server absence probably causes task failure. Since emergency response systems in many regions have applied the cooperative dispatch policy, the hypercube queuing model needs further study to accurately describe system performance. A new extension of the classic hypercube model is then proposed in the next section.

### 3. CHQ Model and Validation

Using the spatial queuing theory and Markovian analysis application, we formulate the CHQ model by a set of flow equilibrium equations. We use  $B$  to represent each possible state of the system (e.g.,  $B = \{1100\}$ ).  $P_B$  is the equilibrium probability of state  $B$ . Compared with the basic hypercube queuing model, the total number of possible states in the CHQ model may not be  $2^N$  because the server combination changes the original candidate unit set. On the one hand, some states in partial dispatch policy never exist in the CHQ model and thus should be eliminated. For example, the state  $\{0100\}$  is not a candidate unit for dispatch in the CHQ model, since V2 cannot be dispatched alone according to the cooperative preference list (see Table 2). The states such as  $\{0001\}$ ,  $\{1010\}$ ,  $\{1001\}$ ,  $\{0101\}$ ,  $\{0011\}$ , and  $\{1011\}$  are also the states that should be eliminated in the CHQ model.

On the other hand, new states will be added in the CHQ model to identify in which type of calls the server is busy. For example,  $\{1110\}^2$  (the first fleet: V1; the second fleet: V2 & V3) and  $\{1110\}^1$  (the first fleet: V3; the second fleet: V2 & V1) are substituted for  $\{1110\}$ . Besides,  $\{1111\}^1$  (the first

fleet: V1; the second fleet: V2 & V3 & V4) and  $\{1111\}^2$  (the first fleet: V3; the second fleet: V2 & V1 & V4) are substituted for  $\{1111\}$ . Possible hypercube and queue states for the

example of the CHQ model are shown in Figure 2. The CHQ model with three atoms and four servers is as follows:

$$(\lambda_1 + \lambda_2 + \lambda_3) \cdot P\{0000\} = \mu_1 \cdot P\{1000\} + (\mu_1 + \mu_2) \cdot P\{1100\} + (\mu_2 + \mu_3 + \mu_4) \cdot P\{0111\} + \mu_3 \cdot P\{0010\} + (\mu_2 + \mu_3) \cdot P\{0110\} + (\mu_1 + \mu_2 + \mu_4) \cdot P\{1101\}, \quad (1)$$

$$(\lambda_2 + \lambda_3 + \mu_1) \cdot P\{1000\} = \lambda_1 \cdot P\{0000\} + (\mu_2 + \mu_3) \cdot P\{1110\}^2 + (\mu_2 + \mu_3 + \mu_4) \cdot P\{1111\}^1, \quad (2)$$

$$(\lambda_2 + \lambda_3 + \mu_3) \cdot P\{0010\} = (\mu_1 + \mu_2) \cdot P\{1110\}^1 + (\mu_1 + \mu_2 + \mu_4) \cdot P\{1111\}^2, \quad (3)$$

$$[(\mu_1 + \mu_2) + \lambda_1] \cdot P\{1100\} = \mu_3 \cdot P\{1110\}^1 + \lambda_2 \cdot P\{0000\}, \quad (4)$$

$$[\lambda_1 + (\mu_2 + \mu_3)] \cdot P\{0110\} = \mu_1 \cdot P\{1110\}^2, \quad (5)$$

$$[(\mu_1 + \mu_2 + \mu_4) + \lambda_1] \cdot P\{1101\} = \mu_3 \cdot P\{1111\}^2, \quad (6)$$

$$[(\mu_2 + \mu_3 + \mu_4) + \lambda_1] \cdot P\{0111\} = \lambda_3 \cdot P\{0000\} + \mu_1 \cdot P\{1111\}^1, \quad (7)$$

$$[(\mu_1 + \mu_2) + \mu_3] \cdot P\{1110\}^1 = \lambda_2 \cdot P\{0010\} + \lambda_1 \cdot P\{1100\}, \quad (8)$$

$$[\mu_1 + (\mu_2 + \mu_3)] \cdot P\{1110\}^2 = \lambda_2 \cdot P\{1000\} + \lambda_1 \cdot P\{0110\}, \quad (9)$$

$$[\mu_1 + (\mu_2 + \mu_3 + \mu_4)] \cdot P\{1111\}^1 = \lambda_3 \cdot P\{1000\} + \lambda_1 \cdot P\{0111\}, \quad (10)$$

$$[\mu_3 + (\mu_1 + \mu_2 + \mu_4)] \cdot P\{1111\}^2 = \lambda_3 \cdot P\{0010\} + \lambda_1 \cdot P\{1101\}, \quad (11)$$

$$P\{0000\} + P\{1000\} + \dots + P\{1111\}^2 = 1. \quad (12)$$

Equation (12) ensures that state probabilities' summation should be equal to 1, called equilibrium probability normalization.

To illustrate the proposed hypercube queuing mechanism, we take equation (2), which is the state transaction of the state  $\{1000\}$ , as an example; that is  $(\lambda_2 + \lambda_3 + \mu_1) \cdot P\{1000\} = \lambda_1 \cdot P\{0000\} + (\mu_2 + \mu_3) \cdot P\{1110\}^2 + (\mu_2 + \mu_3 + \mu_4) \cdot P\{1111\}^1$ .

On the left side of this equation (outflow rate of state  $\{1000\}$ ), all types of calls generated can be serviced in the system. On the right side, the probabilities of all the possible states that transition into  $\{1000\}$  are:

$\{0000\} \rightarrow \{1000\}$ : V1 goes into service when a Type 1 call arrives at atom 1

$\{1110\}^2 \rightarrow \{1000\}$ : the fleet of V2 & V3 finishes a Type 2 call service, and the two servers become available simultaneously

$\{1111\}^1 \rightarrow \{1000\}$ : the fleet of V2 & V3 & V4 finishes a Type 3 call service, and the three servers become available simultaneously

Note that the state  $\{1010\}$  cannot be the past state of  $\{1000\}$  because it is impossible to release another Type A server V3 when V1 is serving the Type 1 call in the case. There are similar equilibrium probability equations for other states. Performance measures such as system lost probability, server workload, and dispatch frequency probability can be evaluated based on state probabilities calculated in the proposed CHQ model.

The hypercube queuing model often needs to be verified to show its effect in describing the entire service system. Therefore, we used Arena simulation software to calculate the system service performance under different parameter values. The results can be compared with those of the mathematical model to find the matched service system mode of our proposed model.

In the Arena simulation software, we enter these parameters and build logic flow modules to implement the cooperative dispatch policy. We set the warm-up time as 730 days (about two years) and conducted a replication length of 3650 days (about ten years) so that the system is in a stationary state, and the half-width of each variable is correlated. The simulation module design is shown in Figure 3.

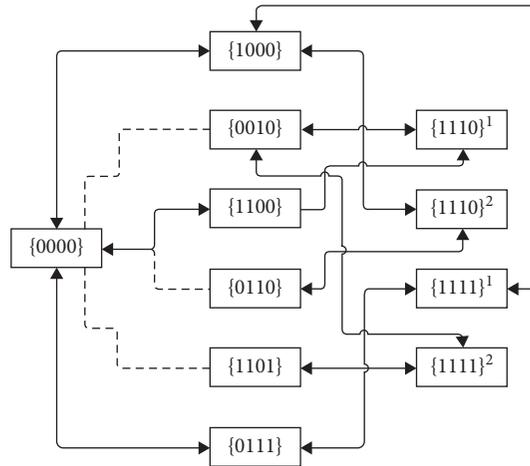


FIGURE 2: Possible hypercube and queue states for the example of the CHQ model.

Status: -1 idle; -2 busy; -4 failure

Call type: 1-am1; 2-am2; 3-am3

Service type: 1-first\_choice; 2-second\_choice; -1-failure

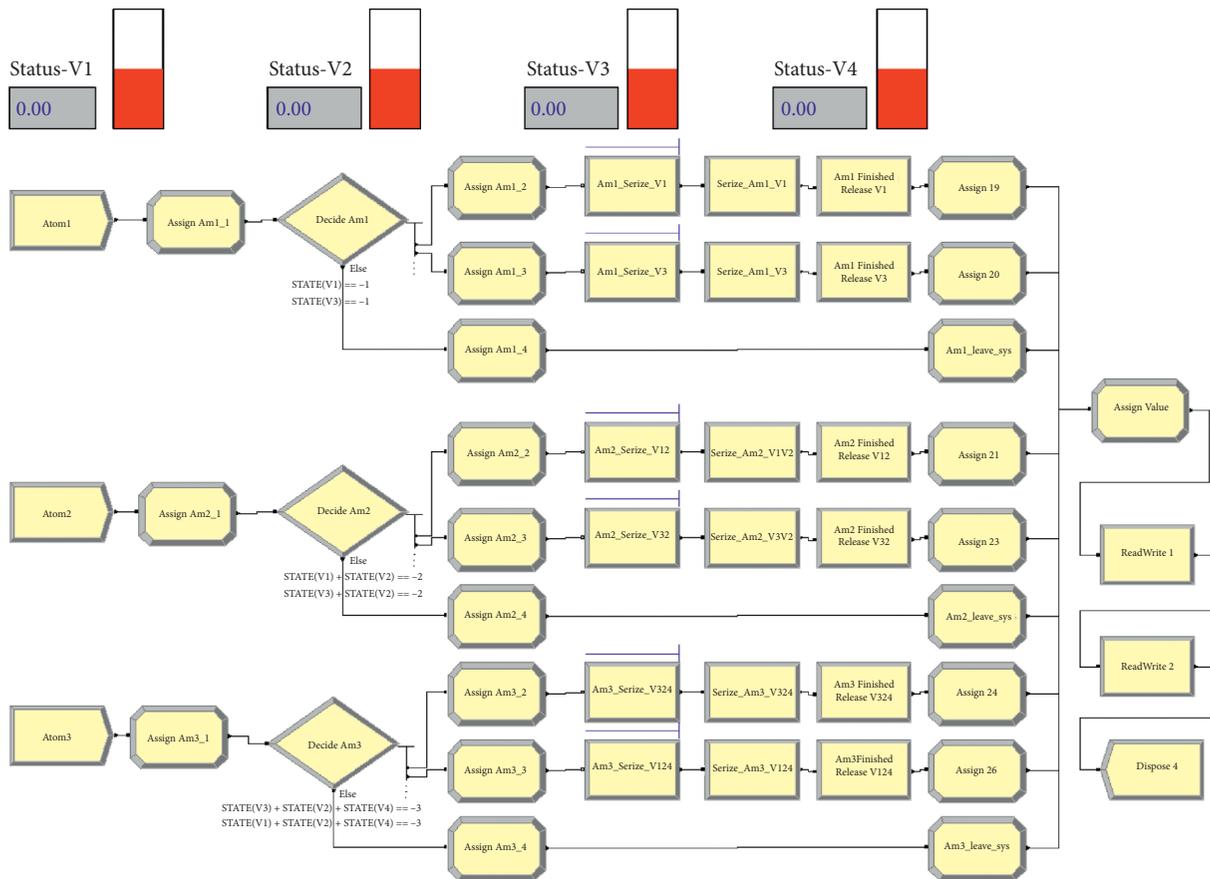


FIGURE 3: The simulation module design of the cooperative response process in the Arena simulation software.

Table 3 shows the system performance results from the CHQ model and the Arena simulation software in Wuhan case study. The call lost probability results indicate that nearly 90% of calls can be served under the cooperative policy in the four-vehicle system. We can observe relatively

small differences in lost probability performance of Type 1, Type 2, and Type 3 (e.g., 0.46%, 2.72%, and 0.42%). The workload statistics show that V1 has the highest probability of being busy. Because V1 is the first response candidate for more frequent calls (e.g., Type 1 and Type 2), since Type 3

TABLE 3: System performance results from the CHQ model and the Arena simulation software in Wuhan case.

System performance	CHQ model	Arena simulation	Difference (%)
Call lost probability			
Type 1	0.005	0.000385	0.46
Type 2	0.08	0.052797	2.72
Type 3	0.026	0.030193	0.42
Workload			
$V_1$	0.221	0.232246	1.12
$V_2$	0.143	0.148918	0.59
$V_3$	0.056	0.07391	1.79
$V_4$	0.024	0.023217	0.08
Dispatch frequency			
$V_1$ to $j_1$	0.203	0.212726	0.97
$V_3$ to $j_1$	0.052	0.047627	0.44
$V_1$ to $j_2$	0.422	0.436617	1.46
$V_2$ to $j_2$	0.48	0.505857	2.59
$V_3$ to $j_2$	0.057	0.06924	1.22
$V_1$ to $j_3$	0.002	0.004015	0.20
$V_2$ to $j_3$	0.154	0.150415	0.36
$V_3$ to $j_3$	0.152	0.1464	0.56
$V_4$ to $j_3$	0.154	0.150415	0.36

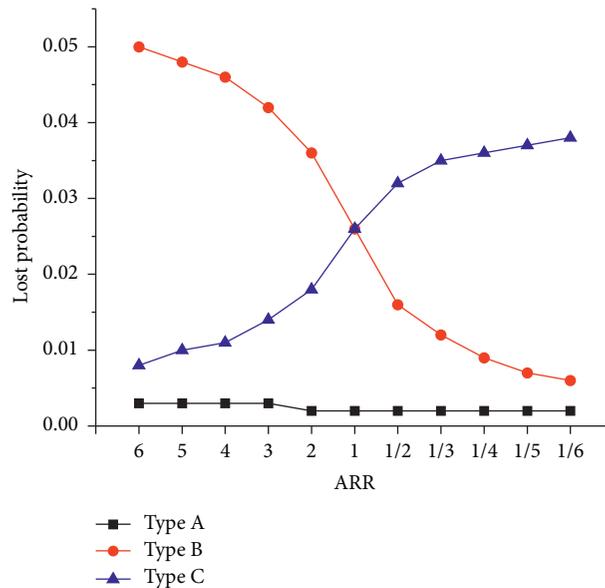


FIGURE 4: Lost probability performance under various ARR.

demand is not as frequent as the other two types of demand,  $V_4$  has the highest idle rate. The workload differences of servers from the two methods are 1.12%, 0.59%, 1.79%, and 0.08%, showing effectiveness in describing system servers' status.

We can also calculate the dispatch frequency statistics under the cooperative policy, such as single-dispatch frequency of  $V_1$  or  $V_3$  to Type 1 calls; double-dispatch frequency of  $V_1$  &  $V_2$  (or  $V_3$  &  $V_2$ ) to Type 2 calls; triple-dispatch frequency of  $V_3$  &  $V_2$  &  $V_4$  (or  $V_1$  &  $V_2$  &  $V_4$ ) to Type 3 calls. According to the model results, the dispatch frequency of first choice  $V_1$  (i.e., 0.203) is almost four times that of the second choice  $V_3$  (0.056) in serving Type 1 calls. Regarding Type 2 calls, the dispatch frequency measures of  $V_1$ ,  $V_2$ , and  $V_3$  are

0.422, 0.48, and 0.057, respectively.  $V_2$ 's dispatch frequency is relatively large because it connects with either  $V_1$  or  $V_3$ . For the triple dispatch to Type 3 calls, the frequency measures of  $V_2$ ,  $V_3$ , and  $V_4$  are nearly the same (i.e., 0.154, 0.152, and 0.154). This measure for  $V_1$  is the least because of its non-priority in the preference list and the less frequency of Type 3 calls. The dispatching frequency differences between the CHQ model and the Arena simulation software are from 0.20% to 2.59%.

In the validation, the simulation results are close to those calculated by the CHQ model. Most of the difference values are under 3%. This shows that the model we proposed can accurately describe the server's status and system performances under cooperation conditions. As a result, satisfied

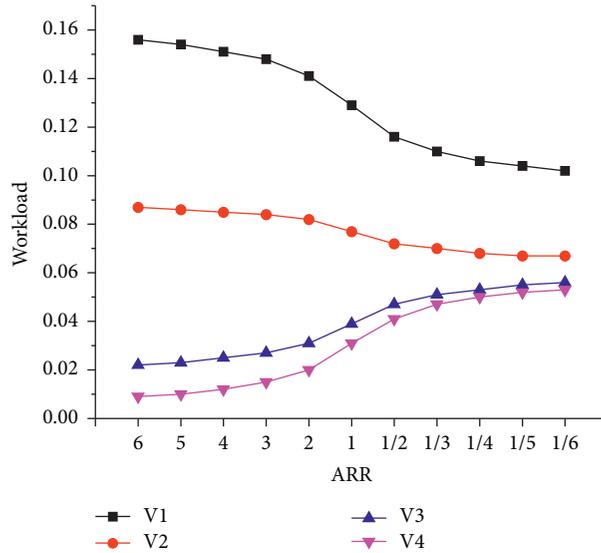


FIGURE 5: Workload performance under various ARR.

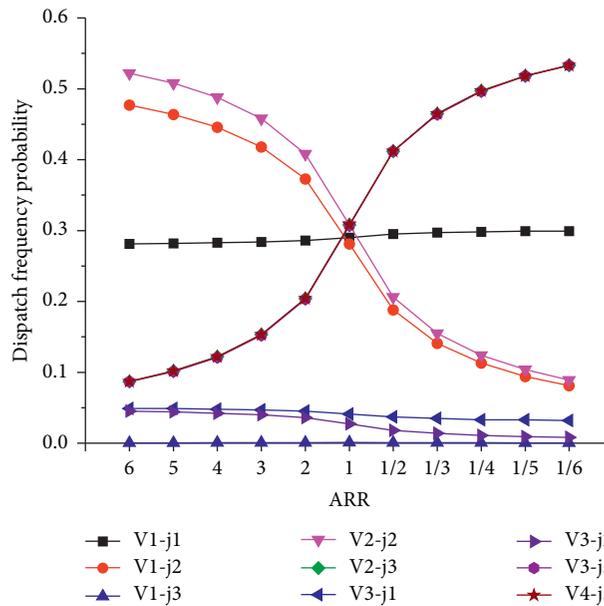


FIGURE 6: Dispatch frequency performance under various ARR.

configurations for engineering guidelines and policy enhancements can be explored through CHQ model application.

#### 4. Sensitivity Analysis

In this section, we conduct additional analysis with homogeneous and heterogeneous demand arrival rates and service rates. The extended hypercube model’s potential application can be explored in more circumstances.

*4.1. Arrival Rate Value of Different Type Calls.* We fix the value of  $\lambda_1$ , the arrival rate of Type 1 demands, as 0.1 and assume that  $\lambda_2$  is symmetric to  $\lambda_3$  with respect to  $\lambda_1$ . The ratio  $\lambda_2/\lambda_3$  is defined as the arrival rate ratio (ARR) of demands that require cooperation responses (i.e.,  $\lambda_1 = 0.1$ ,  $\lambda_2 = 2(1 + (1/ARR))^{-1}$ , and  $\lambda_3 = 2(1 + ARR)^{-1}$ ). The service rate of each vehicle is assumed to be 1. Figures 4–6 show the system performances (e.g., lost probability, server workload, and dispatch frequency) with different ARR (i.e., 1/6, 1/5, 1/4, 1/3, 1/2, 1, 2, 3, 4, 5, and 6). As ARR decreases, the lost probability of Type 1 decreases slightly, while the lost

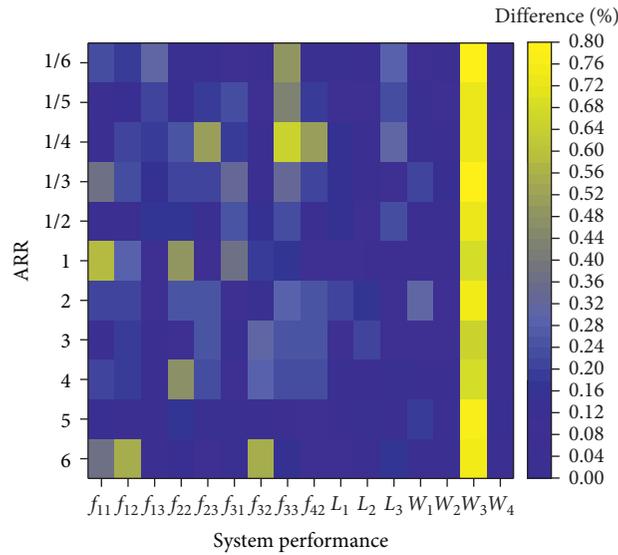


FIGURE 7: System performance difference under various ARR.

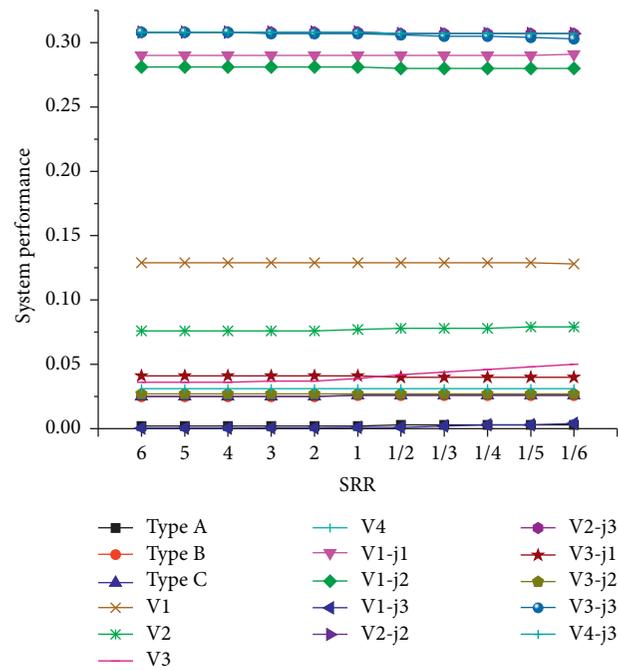


FIGURE 8: System performance under various SRR.

probabilities of Types 2 and 3 show the opposite trend. The centralization degree of vehicle workload probability has gradually strengthened. This is because a smaller ARR parameter leads to more frequent cooperation, promoting vehicles' full utilization.

Regarding dispatch frequency, the probabilities of V1 and V2 to j2 decrease, while the probabilities of V2, V3, and V4 to j3 increase. The dispatch probability of other cooperative vehicles has changed slightly. This shows that the dispatch probabilities of prior vehicles in cooperation are

very sensitive to ARR. Figure 7 illustrates the differences between the CHQ model and the Arena simulation software. The workload difference of V3 remains below 1%, although it is slightly larger than other system performance measures. The proposed model can accurately describe the system performance under a variety of ARR parameters.

4.2. Service Rate Value of Different Servers. The arrival rate of each type of call is assumed to be 0.1. We fix the values of  $\mu_1$

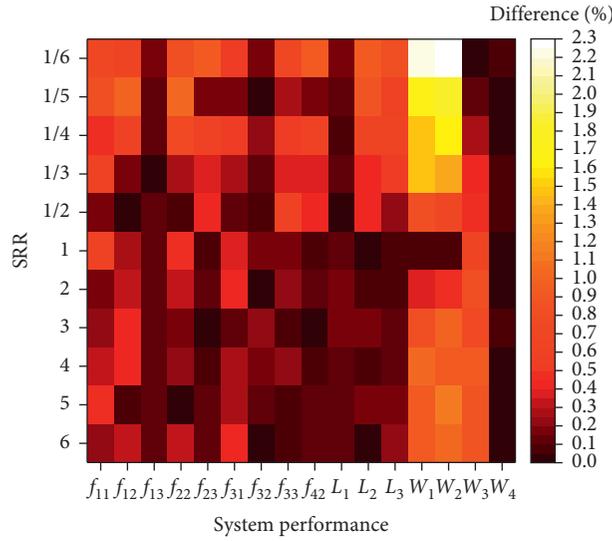


FIGURE 9: System performance difference under various SRR.

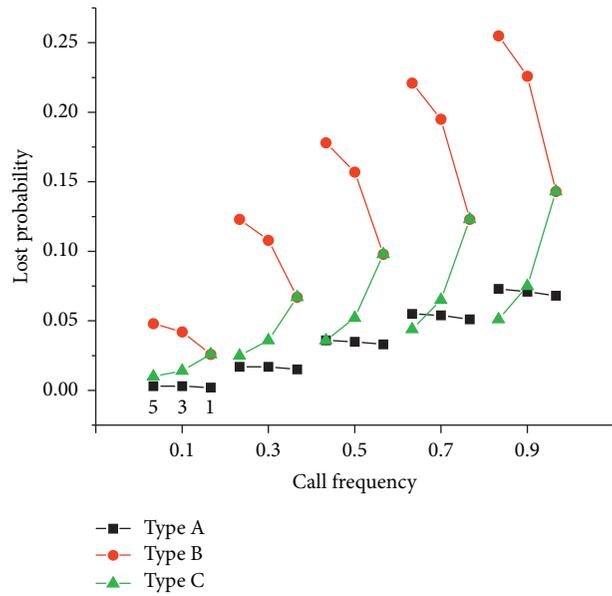


FIGURE 10: Lost probability performance under various call frequencies.

and  $\mu_3$  as 1 and assume that  $\mu_2$  is symmetric to  $\mu_3$  with respect to  $\mu_1$  (or  $\mu_3$ ). The ratio  $\mu_2/\mu_3$  is defined as the service rate ratio (SRR) of necessary vehicles for cooperation responses (i.e.,  $\mu_1 = \mu_3 = 1$ ,  $\mu_2 = 2(1 + (1/SRR))^{-1}$ , and  $\mu_3 = 2(1 + SRR)^{-1}$ ). System performances under different SRR (i.e., 1/6, 1/5, 1/4, 1/3, 1/2, 1, 2, 3, 4, 5, and 6) scenarios are calculated. Figure 8 shows that each system performance measures slight changes, and its sensitivity to SRR is relatively weak. Figure 9 illustrates the differences between the CHQ model and the Arena simulation software. When the SRR value is more considerable, the CHQ model can describe the system performance more accurately.

**4.3. Relative Value of Demand/Service Rate.** We set the value of  $\lambda_1$  as 0.1, 0.3, 0.5, 0.7, and 0.9, respectively. For each value of  $\lambda_1$ , the ARR with the values of 1, 3, and 5 are designed. Lost probability and workload performance under various call frequencies are shown in Figures 10 and 11, respectively. Figure 12 also illustrates the differences between the CHQ model and the Arena simulation software. Note that when the frequency of call increases (e.g.,  $\lambda_1$  is greater than 0.3), the CHQ model's accuracy gradually decreases. Especially when it equals 0.9, the workload difference of V3 can reach more than 10%. This shows that the CHQ model is suitable for cooperative emergency systems with low system

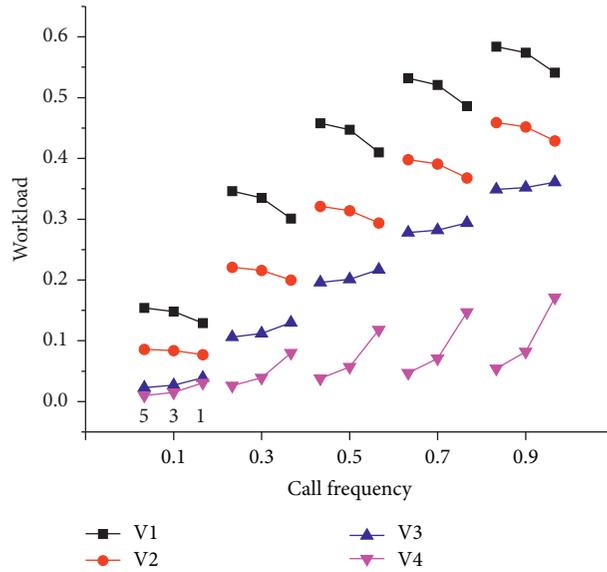


FIGURE 11: Workload performance under various call frequencies.

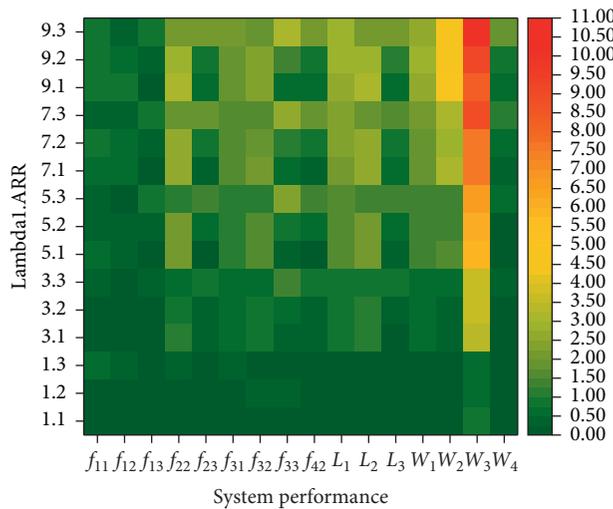


FIGURE 12: System performance difference under various call frequencies.

requirements or strong service capabilities. When the cooperative system’s saturation is large, the accuracy of CHQ’s description for cooperative vehicle workload may decrease.

### 5. Conclusions

Over the past few decades, the hypercube queuing model has been widely used in engineering practice because it can effectively describe the single-dispatch state of emergency response systems. However, for complex disasters that require cooperative service, the classic hypercube queuing model and its extensions may not provide accurate cooperation performance. To this end, we propose the cooperative hypercube queuing model by assuming that certain types of calls require cooperative service; cooperative servers can become free simultaneously from busy states. We test

the model in an empirical emergency system in which two ambulances, one police car, and one fire engine are combined as response candidates. The validation results by the Arena simulation software show that the CHQ model can accurately describe system performance in the case study. To further verify the potential of model application in general scenarios, we test the model’s accuracy under different ARR, SRR, and call frequency values. We find that ARR values have a significant impact on the dispatch frequency of priority vehicles, and its smaller value can promote the balanced use of vehicles. Compared with ARR, the value of SRR has a small effect on system performance. However, larger values of SRR help to improve the accuracy of the model description. The sensitivity analysis on the call frequency shows that the abrupt increase of call frequency may have a negative impact on the accuracy of the CHQ model.

Cooperative emergency systems with low demand or high service rates may be more suitable for applying the CHQ model. An interesting topic for future research would be the CHQ model calibration for an emergency system with frequent calls and limited-service abilities. Another exciting line of research would be to develop heuristic algorithms for large system applications of the CHQ model.

### Data Availability

The data were obtained from AMap (<http://www.amap.com>).

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

The authors thank Professor Yanfeng Ouyang from the University of Illinois at Urbana-Champaign for his valuable support in this research. This work was supported by the National Key Research and Development Project of China (2019YFF0301300) and the National Natural Science Foundation of China (51828802).

### References

- [1] Central People's Government of the People's Republic of China, "National public health emergency response plan," 2006, [http://www.gov.cn/yjgl/2006-02/26/content\\_211654.htm](http://www.gov.cn/yjgl/2006-02/26/content_211654.htm).
- [2] J. Simons and N. Faour, "Mail online: exclusive: female firefighter dubbed "Bride of Beirut" is buried in emotional funeral as her heroic colleagues who broke into doomed warehouse as blaze took hold are named," 2020, <https://www.dailymail.co.uk/news/article-8599961/Hero-firefighters-trying-open-door-doomed-Beirut-warehouse-tragic-photo-named.html>.
- [3] A. Jacobs, "Fire ravages renowned building in Beijing," 2009, <https://www.nytimes.com/2009/02/10/world/asia/10beijing.html>.
- [4] European Centre for Disease Prevention and Control, "Increase in COVID-19 cases in many EU countries," 2020, <https://www.ecdc.europa.eu/en/covid-19-pandemic>.
- [5] Disease Control and Prevention (CDC), "CDC COVID data tracker," 2020, [https://covid.cdc.gov/covid-data-tracker/#cases\\_casesper100klast7days](https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days).
- [6] U.S. Department Of Health and Human Services, *COVID-19 and HIPAA: Disclosures to Law Enforcement, Paramedics, Other First Responders and Public Health Authorities*, U.S. Department Of Health and Human Services, Washington, DC, USA, 2020, <https://www.hhs.gov/sites/default/files/covid-19-hipaa-and-first-responders-508.pdf>.
- [7] C. Toregas, R. Swain, C. ReVelle, and L. Bergman, "The location of emergency service facilities," *Operations Research*, vol. 19, no. 6, pp. 1363–1373, 1971.
- [8] R. Church and C. ReVelle, "The maximal covering location problem," *Papers of the Regional Science Association*, Springer, vol. 32, no. 1, pp. 101–118, Berlin, Germany, 1974.
- [9] V. Marianov and C. Revelle, "The queuing probabilistic location set covering problem and some extensions," *Socio-Economic Planning Sciences*, vol. 28, no. 3, pp. 167–178, 1994.
- [10] M. S. Daskin, "A maximum expected covering location model: formulation, properties and heuristic solution," *Transportation Science*, vol. 17, no. 1, pp. 48–70, 1983.
- [11] E. Erkut, A. Ingolfsson, and G. Erdoğan, "Ambulance location for maximum survival," *Naval Research Logistics (NRL)*, vol. 55, no. 1, pp. 42–58, 2008.
- [12] M. Gendreau, G. Laporte, and F. Semet, "The maximal expected coverage relocation problem for emergency vehicles," *Journal of the Operational Research Society*, vol. 57, no. 1, pp. 22–28, 2006.
- [13] Y. Wang, S. Peng, X. Zhou, M. Mahmoudi, and L. Zhen, "Green logistics location-routing problem with eco-packages," *Transportation Research Part E: Logistics and Transportation Review*, vol. 143, p. 102118, 2020.
- [14] Y. Wang, Y. Yuan, X. Guan et al., "Collaborative two-echelon multicenter vehicle routing optimization based on state-space-time network representation," *Journal of Cleaner Production*, vol. 258, p. 120590, 2020.
- [15] Y. Wang, X. Ma, Z. Li, Y. Liu, M. Xu, and Y. Wang, "Profit distribution in collaborative multiple centers vehicle routing problem," *Journal of Cleaner Production*, vol. 144, pp. 203–219, 2017.
- [16] Y. He, M. Qi, F. Zhou, and J. Su, "An effective metaheuristic for the last mile delivery with roaming delivery locations and stochastic travel times," *Computers & Industrial Engineering*, vol. 145, p. 106513, 2020.
- [17] J. Wang, H. Liu, S. An, and N. Cui, "A new partial coverage locating model for cooperative fire services," *Information Sciences*, vol. 373, pp. 527–538, 2016.
- [18] Y. Zhou, B. Zheng, B. Zheng, J. Su, and Y. Li, "The joint location-transportation model based on grey bi-level programming for early post-earthquake relief," *Journal of Industrial & Management Optimization*, vol. 13, no. 5, 2017.
- [19] R. C. Larson, "A hypercube queuing model for facility location and redistricting in urban emergency services," *Computers & Operations Research*, vol. 1, no. 1, pp. 67–95, 1974.
- [20] J. P. Jarvis, "Approximating the equilibrium behavior of multi-server loss systems," *Management Science*, vol. 31, no. 2, pp. 235–239, 1985.
- [21] K. R. Chelst and Z. Barlach, "Multiple unit dispatches in emergency services: models to estimate system performance," *Management Science*, vol. 27, no. 12, pp. 1390–1409, 1981.
- [22] L. F. Rodrigues, R. Morabito, F. Y. Chiyoshi, A. P. Iannoni, and C. Saydam, "Towards hypercube queuing models for dispatch policies with priority in queue and partial backup," *Computers & Operations Research*, vol. 84, pp. 92–105, 2017.
- [23] S. Ansari, S. Yoon, and L. A. Albert, "An approximate hypercube model for public service systems with co-located servers and multiple response," *Transportation Research Part E: Logistics and Transportation Review*, vol. 103, pp. 143–157, 2017.
- [24] F. Y. Chiyoshi and R. Morabito, "A Tabu search algorithm for solving the extended maximal availability location problem," *International Transactions in Operational Research*, vol. 18, no. 6, pp. 663–678, 2011.
- [25] A. P. Iannoni and R. Morabito, "A multiple dispatch and partial backup hypercube queuing model to analyze emergency medical systems on highways," *Transportation Research Part E: Logistics and Transportation Review*, vol. 43, no. 6, pp. 755–771, 2007.
- [26] Health Commission of Hubei Province, "Information released of COVID-19 in Hubei Province," 2020, <http://wjw.hubei.gov.cn/bmdt/ztlz/fkxxgzbdgrfyyq/xxfb>.