Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

DYNAMIC OPTIMIZATION OF DRONE DISPATCH FOR SUBSTANCE OVERDOSE RESCUE

Xiaoquan Gao

School of Industrial Engineering
Purdue University
315 Grant St 2
West Lafayette, IN 47907, USA

Nan Kong

Weldon School of Biomedical Engineering
Purdue University
206 S Martin Jischke Dr
West Lafayette, IN 47907, USA

Paul M. Griffin

Harold and Inge Marcus Department of Industrial and Manufacturing Engineering
Pennsylvania State University
310 Leonhard Building
University Park, PA 16802, USA

ABSTRACT

Opioid overdose rescue is very time-sensitive. Hence, drone-delivered naloxone has the potential to be a transformative innovation due to its easily deployable and flexible nature. We formulate a Markov Decision Process (MDP) model to dispatch the appropriate drone after an overdose request arrives and to relocate the drone to its next waiting location after having completed its current task. Since the underlying optimization problem is subject to the curse of dimensionality, we solve it using ad-hoc state aggregation and evaluate it through a simulation with higher granularity. Our simulation-based comparative study is based on emergency medical service data from the state of Indiana. We compare the optimal policy resulting from the scaled-down MDP model with a myopic policy as the baseline. We consider the impact of drone type and service area type on outcomes, which offers insights into the performance of the MDP suboptimal policy under various settings.

1 INTRODUCTION

Over the past three decades, drug overdose deaths in the US have tripled (Warner et al. 2011). More than 72,000 Americans died from a drug overdose with opioids making up the majority of these deaths. This is more than the number of lives lost in car accidents or gun-related homicides (Scholl et al. 2019). Naloxone is a medication commonly used for respiratory failure from opioid overdose. It is included as part of an emergency overdose response kit distributed to opioid drug users and emergency responders. The efficacy of naloxone depends on how quickly it can be delivered. Death from overdose typically occurs within 1 to 3 hours. It is therefore important to reduce the time it takes to administer naloxone by getting it into the hands of those best positioned to respond rapidly (Kim et al. 2009). However, due to insufficient emergency medical services (EMS) resources and traffic congestion, many opioid overdosed people miss the best time to receive naloxone. Unmanned aerial vehicles (UAVs) can be easily deployed to augment the current logistic system and potentially cost-effective for not staffing additional EMS personnel and independent of road conditions. To reduce the response time by a minute traditionally requires staffing more ambulances, which can cost $200,000. However, a $10,000 drone can reduce the response time by two minutes. Furthermore, NARCAN Nasal Spray is now available, which allows for naloxone to be sprayed into the nose. If an overdosed person is unconscious, a bystander can come to help. Therefore, it is practical and efficient for UAVs to use the spray on the overdosed person. Introducing UAVs to EMS makes timely treatment practical, and consequently, increases the chance of survival.

Once a 9-1-1 call for opioid overdose is received, the administrator needs to decide which UAV to dispatch. The most common and natural dispatch strategy is to send the UAV closest to the overdosed person, since the objective, in general, is to minimize the response time, thus maximizing the chance of survival. Unfortunately, this strategy tends to be myopic (Carter et al. 1972). Consider the case where two UAVs are in areas A and B, and area A has a higher call volume than area B. In this case, the mean response time will decrease if the UAV stationed in area B is allowed to respond to some of the calls from area A. Further complicating the EMS operations is that a UAV can be relocated after completing its EMS mission, i.e., choosing its site to wait for the next request. This is different from ambulance, which we are not able to relocate frequently for the sake of its medical staff. In summary, UAVs bring greater flexibility and uncertainty in EMS delivery, and thus present bigger challenges on making logistic operation decisions.

In this paper, we optimize sequential dispatch and relocation decisions jointly for UAVs in an EMS delivery logistics context. Our contribution is threefold.

1. We propose a novel stochastic dynamic model for sequential dispatch and relocation joint optimization problem, which incorporates practical relevance of UAVs (e.g., energy consumption).
2. We design an action disaggregation and adjustment theme based on a sub-optimal policy (or strategy) from solving a scaled-down version of the MDP problem which is induced by state aggregation.
3. We identify the settings with respect to the UAV type and service area type, under which a computationally attainable MDP suboptimal policy outperforms a commonly used myopic policy.

The rest of this paper is structured as follows. Section 2 provides a literature review on related work. In Section 3, we describe the problem in detail and present our MDP model, followed by a solution based on state aggregation. We use a discrete-event simulation to assess the performance of our MDP-based policy under various settings in Section 4. We draw conclusions and outline future work in Section 5.

2 LITERATURE REVIEW

This paper contributes to the literature in three areas: (1) EMS supply chain and logistics, (2) dynamic EMS vehicle dispatch/assignment, and (3) state aggregation based MDP numerical solution. We briefly describe each of these below.

According to the scope and time scale that such decisions are made, the EMS planning problems can be divided into three different categories (Reuter-Oppermann et al. 2017): (1) strategic level where decisions are made for several years, e.g., facility location problems; (2) tactical level where decisions often hold for several months, e.g., staff scheduling; and (3) operational level where decisions are made on a daily basis or even in real time, e.g., patient transport scheduling. Our work falls into the third category. For each of these three levels, different planning problems arise, including demand and workload foresting , call-handling strategy design, and ambulance planning (Zarkeshzadeh et al. 2016).

Dynamic assignment models have been applied extensively to EMS dispatch problems focusing on ambulance operations. One stream of these studies aims to develop decision support tools for ambulance dispatch and relocation decisions through threshold policies. For example, Andersson and Värbrand (2007) proposed a "preparedness" measure to evaluate the ability of an EMS system serving potential patients at present and in the future. However, to deploy a preparedness threshold-based policy requires the pre-tuning of a set of parameters, which is highly dependent on the situation of each distinct area. Aringhieri et al. (2017) and Belanger et al. (2019) provided a comprehensive reviews of optimization models in location, relocation and dispatching of emergency medical vehicles. To deal with the complexity of the EMS system, Lanzarone et al. (2018) and Belanger et al. (2020) proposed a recursive simulation-optimization framework which encompasses a mathematical formulation and a discrete-event simulation model that produces both empirical estimations of the ambulance availability and system performances.

In addition to the simulation-optimization approaches, many researchers applied Markov decision processes (MDPs) for system modeling and optimization. For example, Grannan (2014) formulated an infinite-horizon, average-cost MDP model to solve a medical asset dispatching and casualty transport problem. His MDP model contains only 12 states and 16 actions. Our MDP model is, instead, more similar to those presented in McLay and Mayorga (2013) and Jagtenberg et al. (2017). These models are discrete-time MDP models, which include the status of EMS request times and vehicle locations in the state space, and make dispatch and relocation decisions. To avoid the curse of dimensionality, the authors on both papers assumed that the service time is exponential so that they could employ the standard value iteration algorithm. To make the model more realistic, we relax the above exponential assumption and incorporate the remaining service time in our state space to capture energy consumption. Note that medical UAVs for their agility and low cost need to be battery powered which can last many hours (or multiple EMS missions) with the current battery technology. The consideration of energy consumption drastically increases the size of the state space, which makes the real-world MDP problem we deal with intractable with standard solution methods such as value iteration and policy iteration.

To address the computational challenges, we apply a straightforward but effective version of state aggregation to reduce the problem size so it is solvable in a reasonable amount of time through a standard implementation of the value iteration method. In addition, choosing state aggregation helps to identify managerial insights in our studied context. Much of the existing literature on state aggregation falls into two categories: (1) assuming a fixed aggregation and evaluating various bounds on the performance loss , and (2) designing algorithms on the basis of state aggregation to solve large-scale MDPs (Mendelssohn 1982) . However, the question of finding an appropriate method for state aggregation to minimize the performance loss that remains open. By assuming the availability of the value function or a reasonably good estimate on it, Jia (2010) shed insights on how to find simple aggregation policies with good performance. Nevertheless, assumptions on the necessary derivation required in the above literature, are not realistic for our problem. We instead choose to aggregate the states based on simplistic geographic partitioning and compare it with a myopic policy via simulation.

3 METHODOLOGY

We solve the sequential UAV dispatch and relocation joint decision optimization problem with an infinite-horizon discrete-time MDP model. In each state s (Section 3.1), we check the status of UAVs and emergency calls and then choose an action from the allowable actions: $\mathcal{A}_s \subseteq \mathcal{A}$ (Section 3.2). This process evolves over time according to the transition probabilities (Section 3.4), which depend on the current state and the chosen action. Our goals in this problem are (1) maximizing the discounted total service rate and (2) minimizing the unsatisfied requests (Section 3.3). We describe our solution approach in Section 3.5.

Table 1: Notation.

I	set of demand nodes, $\lvert I\rvert = M$
J	set of UAV charging stations, $\lvert J\rvert = N$
$K = I \cup J$	set of candidate relocation destinations of UAVs
$(d_{ij})_{(M+N)\times(M+N)}$	distance matrix between demand nodes and charging stations
L	total number of UAVs
R	an upper bound on service time
Z	maximum level of UAV's battery power
λ	expected number of requests that arrive per unit of time (Poisson parameter)
P_i	conditional probability that a request arrives at demand node i, given that a request arrives, $i = 1, \cdots, N$, such that $\sum_{i=1}^{N} P_i = 1$

3.1 States

We define the *states* of the system at time t as $\mathbf{s}(t) := (s_1(t), \cdots, s_L(t))$. More specifically, we define

$$\mathbf{s}(t) := \{D_1(t), s_1(t), E_1(t), \cdots, D_L(t), s_L(t), E_L(t), \varepsilon_1(t), \cdots, \varepsilon_N(t)\}.$$

The state space comprises two sections. The first section indicates the status of each UAV, which further involves three features: $D_l(t) = k$, $l = 1, \cdots, L$, $k \in K$, if at time t, UAV l's destination is k; $E_l(t) = e$, $e \in [0,1]$, if at time t, the battery power level of UAV l is e; and

$$s_l(t) = \begin{cases} r & \text{if UAV } l\text{'s remaining service time is } r \text{ at time } t, r \in \{1, \cdots, R\}, \\ 0 & \text{otherwise (i.e., UAV } l \text{ is idle).} \end{cases}$$

The other part implies the status of the requests: $\boldsymbol{\varepsilon}(t) = \{\varepsilon_1(t), \cdots, \varepsilon_N(t)\}$, where $\varepsilon_i(t) = \varepsilon$, representing the number of requests that occur at node i in decision epoch $(t-1, t]$, $\varepsilon \geq 0$, $i = 1, \cdots, N$. In our model, we assume that at most one request occurs at each demand node within a decision epoch, i.e., $\varepsilon \in \{0,1\}$, with the number of occurrence following a Poisson process. Hence the smaller the decision epoch is, the more accurate our model is. However, there is a tradeoff because smaller decision epochs lead to heavier computational burden. Throughout this paper, we consider the length of each decision epoch to be five minutes, which we believe balances the solution accuracy and computation time.

3.2 Decisions

We next define the actions as $\mathbf{u}(t) = \mathbf{u}(\mathbf{s}(t)) := (u_1(t), \cdots, u_L(t))$, where

$$u_l(t) = \begin{cases} k & \text{if UAV } l \text{ is dispatched or relocated to location } k,\ k \in K; \\ 0 & \text{otherwise,} \end{cases}$$

for $l = 1, \cdots, L$. Note that for some states, not all actions are necessarily permissible. Denote the set of feasible actions in state s as $\mathcal{A}_s \in \mathcal{A}$. Feasible actions should meet the following requirements:

1. If $s_l(t) > 0$, $u_l(t) = 0$, i.e., dispatch and relocation decisions can only be made on idle UAVs;
2. If $\min_{j \in J} 2e(d_{D_l(t),j}) > e_l(t)$, $u_l(t) \in J$, i.e., the UAV must go to a charging station for recharge if its battery power is lower than twice of the battery needed to fly to the nearest charging station.

3.3 Rewards

We consider a reward dependent upon the service time, which is a combination of travel time and on-scene time (see Figure 1). The on-scene time is assumed to be constant and the travel time is assumed to only dependend on the distance. In EMS practice, a typical goal is to minimize the fraction of late arrivals with respect to some cut-off time. However, this type of objective penalizes solutions that lead to many missions with service times just above the cut-off. Hence we choose survival rate as the performance measure, which is a function of the service time. Dietze et al. (2019) estimate the Kaplan-Meier survival curve for opioid overdose. To approximate the survival rate, we define the reward as a step function of the service time:

$$R\big(\mathbf{s}(t), \mathbf{u}(t)\big) = \textstyle\sum_{l=1}^{L} \mathbf{1}_{\{0 < u_l(t) \leq N\}} f\big(d_{u_l(t), D_l(t)}\big)$$

where

$$f\big(d_{u_l(t), D_l(t)}\big) = \sum_{h=1}^{H} \alpha_h \mathbf{1}_{\{d_{u_l(t), D_l(t)} < p_h\}}.$$

Parameters $\alpha_h, p_h, h = 1, 2, ..., H$, and H are estimated from the Kaplan-Meier curve and $\mathbf{1}_{\{\cdot\}}$ is the indicator function.

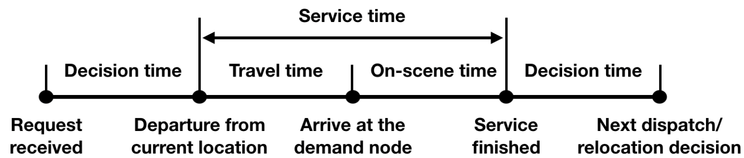


Figure 1: An illustration of UAV-based EMS process timeline.

Note that if a request arrives when there is no idle UAV, it is deemed as an unsatisfied request, and no UAV will be dispatched. Since UAVs are not the only way to deliver naloxone and the opioid overdose rescue is time-sensitive, we consider it unreasonable to hold the request to the next decision epoch. This kind of request can be referred to ambulances or other first responders. Given the difficulty of modeling the number of unsatisfied requests with a functional form, we do not incorporate it as part of the reward in the MDP model. Instead, we evaluate this outcome through a high-fidelity simulation (Section 4).

3.4 Transition Probabilities

Two events drive the state transitions: (1) the decision maker makes a dispatch or relocation decision, and (2) UAV's remaining service time and its battery power decrease with time. Accordingly, we can write out the system dynamics, i.e., $(\mathbf{s}(t),\mathbf{u}(t)) \longmapsto \mathbf{s}(t+1)$. For each UAV $l = 1,\cdots,L$,

$$D_l(t+1) = \begin{cases} u_l(t+1) & \text{if } u_l(t+1) \neq 0; \\ D_l(t) & \text{otherwise.} \end{cases} \tag{1}$$

If $s_l(t) = 0$,

$$s_l(t+1) = \begin{cases} \tau_{u_l(t),D_l(t)} & \text{if } u_l(t+1) \neq 0; \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

If $s_l(t) > 0$,

$$s_l(t+1) = [s_l(t) - 1]^+. \tag{3}$$

$$e_l(t+1) = \begin{cases} e_l(t) & \text{if } u_l(t+1) = 0; \\ e_l(t) - e(d_{u_l(t),D_l(t)}) & \text{If } u_l(t+1) \in \{1,2,...,N\}; \\ e_{max} & \text{If } u_l(t+1) \in \{N+1,...,N+M\}. \end{cases} \tag{4}$$

Equation (1) shows the dynamics of the states: unless a new dispatch or relocation decision is made, the destination of the UAV will remain the same. Equations (2) and (3) indicate the transitions of remaining service time. Whenever a dispatch or relocation decision is made, the remaining service time is set as a function of the distance between the UAV's current location and its destination. We assume constant payload and flight speed. Equation (4) shows the energy consumption of the UAV for one flight $e(d_{u_l(t),D_l(t)})$, which depends on the flight distance.

We now present the stochastic dynamic optimization problem as:

$$\max_{\pi} \quad \mathbb{E}[\sum_{t=1}^{\infty} \alpha^t R(\mathbf{s}(t),\mathbf{u}(t))]$$

where α is the discounting factor and $\pi : \mathcal{S} \to \mathcal{A}_s$ denotes the policy to be optimized.

3.5 Solution Methods

Standard discounted infinite-horizon MDP problems can be solved by value iteration, policy iteration, or linear programming. To solve a problem, all the above algorithms need to first parameterize the associated transition matrix ($\mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$) and reward matrix ($\mathcal{S} \times \mathcal{A} \to \mathbb{R}$). In our MDP model, the transition matrix has dimension of $|\mathcal{S}|^2 \times |\mathcal{A}| = (((N+M+1)(R+1)(Z+1))^L \cdot 2^N)^2(N+M+1)^L = 2^{2N}(N+M+1)^{3L}(R+1)^{2L}(Z+1)^{2L}$. Even for a toy instance with 4 demand nodes and 2 UAVs with minimally distinguishable remaining service time and battery power, the value iteration algorithm will terminate prematurely on a standard personal computer for exhausting its computing resource. Furthermore, for a real instance with 64 demand nodes and 2 UAVs, the dimension of the policy lookup table is about 3.1×10^{27}. Finally, even if we could solve such a problem in a reasonable amount of time, it would have been impractical to store the resultant policy lookup table and query it for every service logistic operation.

To address the above challenges, We first apply an ad-hoc state aggregation technique to deal with the inherent curse of dimensionality. We then *disaggregate* each action in the resultant policy lookup table to each original state. We term the MDP model induced by state aggregation the *scaled-down MDP model*. Recall for our MDP model the state variable includes: (1) destination of each UAV, i.e., $D_l(t) \in I, l = 1,...,L$, (2) number of requests from each demand node, i.e. $\varepsilon_i(t), i \in I$, (3) remaining service time of each UAV, i.e., $s_l(t) \in \{0,1,...,R\}, l = 1,...,L$, and (4) remaining battery power of each UAV, i.e., $E_l(t) \in \{0,1,...,Z\}, l = 1,...,L$. The first two components are pertaining to the set of demand nodes. We thus aggregate the demand nodes in the scaled-down MDP model based on some notion of geographical similarity, since adjacent demand nodes imply comparable flight time, and in general, request intensity as well. Thus, we aggregate the first two components of the state set. In addition, the levels of the remaining service time build on the total number of demand nodes, through which the flight time between different nodes can be distinguished. While keeping the idle state $s_l(t) = 0$ unchanged, we cluster positive remaining service times uniformly. Given the computational burden and the importance of each component to the model, we ignore the component of energy consumption in the scaled-down MDP model but recover it in the ensuing simulation-based comparative study. With state aggregation, we can efficiently construct a lookup table to represent the MDP suboptimal policy from state aggregation for real instances.

4 PERFORMANCE EVALUATION

In this section, we present a simulation-based comparative study, which is based on real EMS data from the state of Indiana. We use the practical strategy that always dispatches the closest idle UAV as the baseline. We construct a number of real instances (i.e., with 2 UAVs and 64 demand nodes) with respect to the types of UAV and service area, and offer insights into the performance of the MDP suboptimal policy under these settings.

4.1 Simulation-based Evaluation Procedure

The dimension of any policy lookup table is the same as that of the state set, so it is impractical to store and use a lookup table policy with respect to the original state set. Instead, we evaluate the optimal policy obtained from solving the scaled-down MDP model with a self-developed discrete-event stochastic simulation. To evaluate the policy derived from the scaled-down MDP model, we propose a scheme to disaggregate actions mapped from the aggregated states for the scaled-down MDP. Further, we consider the geographic distribution of the requests and energy consumption of the UAVs as additional features in the simulation. We present the simulation-based evaluation procedure as follows.

1. Generate a stream of requests as a Poisson process with parameter λ and request distribution P_i, $i = 1,...,N$. Divide the requests into time intervals of equal length.
2. Aggregate states $s \in \mathcal{S}_k$ into s'_k of the scale-down MDP. Specifically, aggregate the remaining service time uniformly as $0 \to 0$, $\{1,2,...,\frac{R}{R'}\} \to 1$,..., $\{\frac{(R'-1)R}{R'}+1,...,R\} \to R'$. The demand nodes $i \in I_{i'}$ are aggregated into area i' according to the partition $I = \bigcup_{i'} I_{i'}$. Note that if there are more than one requests in areas $i \in I_{i'}$, we still have $\varepsilon_{i'} = 1$ in the aggregated model.

3. Get the optimal action for the aggregated state from the input policy lookup table resulted from solving the scaled-down MDP.
4. Disaggregate the action based on the following theme. If $u'_l = 0$, $u_l = 0$. If $u'_l = i'$ and $\sum_{i \in I_{i'}} \varepsilon_i = 0$, then u'_l is a relocation decision with $u_l = \arg\max_{i \in I_{i'}} P_i$. If $u'_l = i'$ and $\sum_{i \in I_{i'}} \varepsilon_i = 1$, then u'_l is a dispatch decision with $u_l = \sum_{i \in I_{i'}} \varepsilon_i \cdot i$. If $u'_l = i'$ and $\sum_{i \in I_{i'}} \varepsilon_i > 1$, then u'_l is a dispatch decision with $u_l = \arg\min_{i \in I_{i'}} d_{D_l(t),i} - M\varepsilon_i$, where M is a large enough constant.
5. Add recharge actions. If $\min_{j \in J} 2e(d_{D_l(t),j}) > e_l(t)$, $u_l(t) \in J$, then change u_l to $\arg\min_{j \in J} d_{D_l(t),j}$.
6. Simulate the system dynamics and calculate the rewards following the model described in Section 3.3 and 3.4. Furthermore, we label the request if it is unsatisfied and tally all the unsatisfied requests as the additional outcome.

In Step 4 of the above procedure, we explain how we translate actions resulted from the scaled-down MDP into relocation operations with the following two points. First, with any resultant relocation action, we dispatch the closest UAV to the request in the simulation and assign it to the demand area with the highest request frequency since the relocation is intended to better prepare for future requests (Andersson and Värbrand 2007). In addition, with aggregation, the policy lookup table cannot deal with multiple requests that arrive within the current time interval. Thus, in the simulation, we select the UAV closest to a request and have it to fly to the area of that request. To address the concern on energy consumption, we, as described in Step 5, assign a UAV to the closest charging station when the UAV's battery power is lower than twice of the battery power needed to fly to the nearest charging station. To make a fair comparison, we specify the myopic baseline policy as follows. With the baseline policy, we always dispatch the idle UAV closest to the request and take the same operation as before to recharge the UAVs.

In practice, we offline compute the optimal MDP policy and set up the aggregation and disaggregation strategies, as stated above. At each decision epoch, we collect requests' locations and UAVs' statuses and aggregate them for the state space of the scale-down MDP. Then we compute the optimal action for the aggregated state from the input policy lookup-table, resulting from solving the scale-down MDP. Afterward, we disaggregate the action based on the given strategy and add recharging actions.

4.2 Experimental Setup

We consider three different types of UAVs (Table 2), namely *consumer drones*, *top-of-range drones*, and *ideal drones*. The consumer drones are what can typically be found in the market, which does not allow relocation without a recharge after one mission. We consider ideal drones to be those that have infinite cruise range. In other words, they can continuously carry out missions without a recharge. We consider top-of-range drones to be those between consumer drones and ideal drones in terms of their cruise range. We assess their respective service system performances based on the optimal strategy from solving the scaled-down MDP (i.e., without energy consumption consideration).

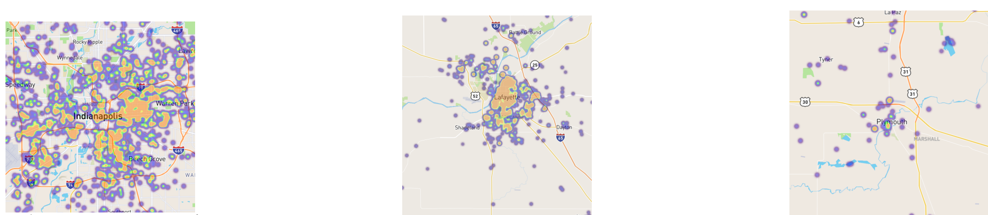
Table 2: Features of different types of UAVs.

Type of UAVs	Engineering features	Eligibility for dispatch and relocation
consumer drone	able to fly for up to one hour with an average speed of $70km/h$	need to fly back to a charging station after each mission for recharge and supply replenishment
top-of-range drone	able to fly for up to four hours with an average speed of $70km/h$	can continuously execute three to four missions (dependent on the flight distance) without needing recharge or resupply
ideal drone	have an infinite cruise range and with an average speed of $70km/h$	can execute missions in a nonstop fashion for a long period of time

For our case studies, we extracted real-world data from the naloxone administration heatmap in the state of Indiana. This heatmap provides information of suspected opioid overdose incidence locations from

2014 to 2019. Their encounters (demands) were initially requested by patients themselves and bystanders. The geographic distribution of these request incidences (measured by the centrality of the distribution) is consistent with the rural-urban area delineation (Figure 2). We thus considered three types of catchment areas in terms of their geographic area delineation, namely *urban*, *semiurban*, and *rural*. The three catchment areas selected are Marion county, Tippecanoe county, and Plymouth county, respectively. We thus simulated geographic-specific demand arrivals based on the heatmaps of the three counties above.

The pattern of the demand differed not only by its geographic distribution but also its overall arrival (Table 3). For each of the three catchment area types, we simulated different overall demand arrival rates for the likely time-varying nature of the arrival process. When we solved for the MDP-based optimal policy, we set the request arrival rate to be 1 (i.e., $\lambda = 1$) without knowing the real-world overall arrival rate. Then through preliminary experiments, we identified the possible range of λ. We found that when $\lambda = 0.5$, both strategies can satisfy over 80% requests and when $\lambda = 4$, both strategies can satisfy only 20% requests. We argue such a service quality guarantee is reasonable with the 2-UAV capacity configuration. Further, in our ensuing simulation-based comparative studies, we specified four overall arrival rates to be $\lambda = 0.5, 1, 2, 4$, which represent low, medium-low, medium-high, and high demand scenarios, respectively. These four rates are representative for capturing the time-varying demand intensity over a day based on historical events that occurred in the state of Indiana.



(a) Urban county: Marion (b) Semiurban county: Tippecanoe (c) Rural county: Plymouth

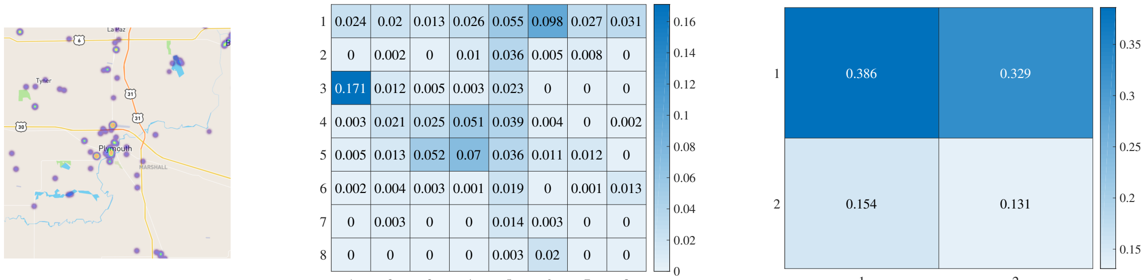
Figure 2: Distributions of request incidences.

Table 3: Features of different incidences distribution.

Country type	Heatmap used	Overall arrival rate	Distribution centrality
urban county	Marion	high	yes
semiurban county	Tippecanoe	medium	yes
rural county	Plymouth	low	no

We partitioned each county into an 8×8 grid and extracted the geographic distribution of the request arrivals through analyzing the RGB features of the corresponding heatmap. With the state aggregation approach, we constructed the scaled-down MDP model based on an orthogonal 2×2 aggregate grid. Taking the rural county (Plymouth county) as an example, we show the geographic distribution of the arrivals as an input to our simulation and scale-down MDP models in Figure 3.

The time horizon for the simulation was one day, which was divided evenly into 288 5-minute decision epochs. This specification balances the requirement of setting up the MDP as an infinite-horizon problem and the requirement of making realistic UAV logistic operations. Taking the semiurban county (Tippecanoe county) under the medium overall arrival rate as an example, we show in Figure 4 that the simulation duration is long enough to stabilize the average reward over time for both strategies and all types of drones, which leads to meaningful comparison. Finally, we simulated each setting (i.e., one service area type, one UAV type, and one overall arrival rate level) for 1000 replications to ensure statistical confidence in the comparisons.



(a) Request distribution heatmap. (b) Request distribution for the simulation. (c) Request distribution for the scale-down MDP.

Figure 3: Request distribution extraction and the state aggregation.

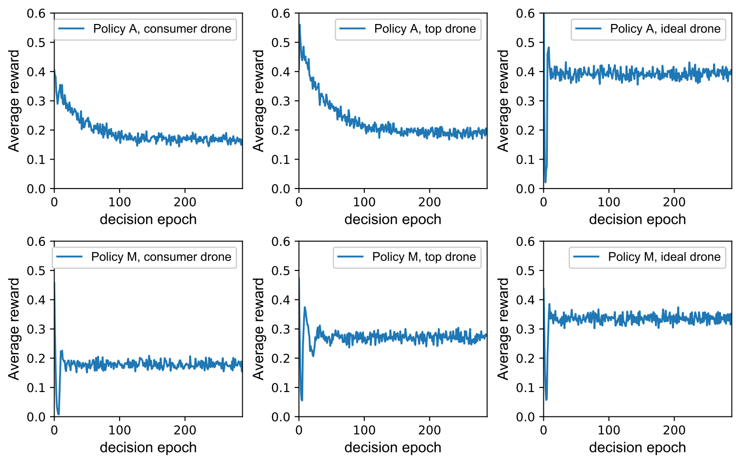


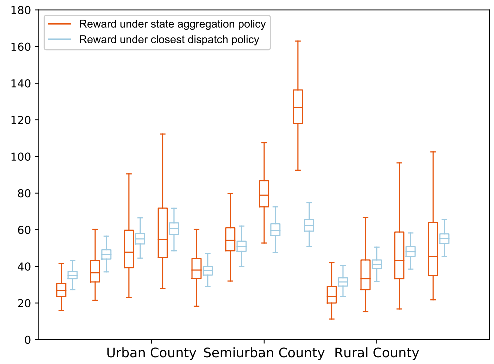
Figure 4: The trend of average reward (i.e., $\frac{\sum_{\tau=1}^{t} R\left(\mathbf{s}(\tau),\mathbf{u}(\tau)\right)}{t}$ vs. t). Policy A refers to the MDP suboptimal policy with state aggregation. Policy M refers to the nearest dispatch based myopic policy.

4.3 Simulation Results and Analysis

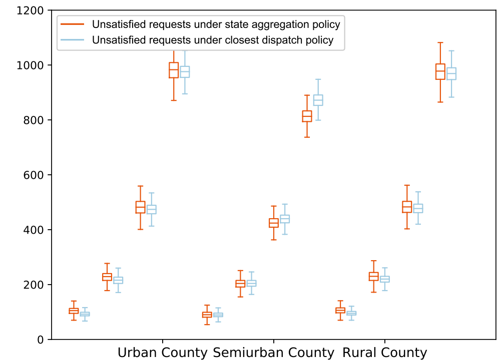
We compared the two strategies for UAV logistic operations in terms of two performance measures: (1) rewards, i.e., the survival rate based on the Kaplan-Meier curve approximation; (2) costs, i.e., the number of unsatisfied requests. Figure 5 shows the performance comparisons.

First, the figure suggests the consistency of strategy superiority between the rewards and costs. In other words, whenever a strategy achieves higher rewards in Figures 5(a), 5(c) and 5(e), the policy incurs lower costs in Figures 5(b), 5(d) and 5(f). This alleviates the concern that the action disaggregation step in the simulation could distort the tradeoff between the rewards and the costs.

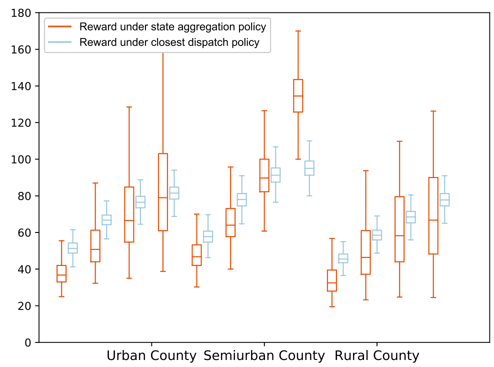
Our results suggest that Policy A (or state aggregation policy; MDP suboptimal policy) performs better than Policy M (or nearest dispatch policy; myopic policy) in high-demand scenarios. It is the nature of an MDP model that its solution takes future opportunity rewards/costs into consideration. When the demand is high, there are more requests emerging continuously, and thus Policy A shows its advantage in better preparing for future requests. Note that this advantage is not offset by the inferiority brought in by the state aggregation. On the other hand, our results suggest that Policy M outperforms Policy A in low-demand cases. When the demand is low, the decision process is close to the combination of multiple segments of one-shot decisions in that there is a long inter-arrival time between two consecutive requests. Thus, it



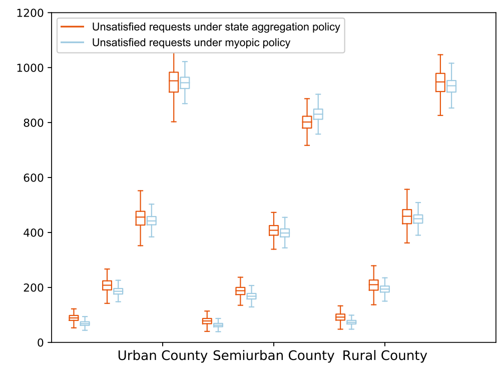
(a) Consumer drone: rewards.



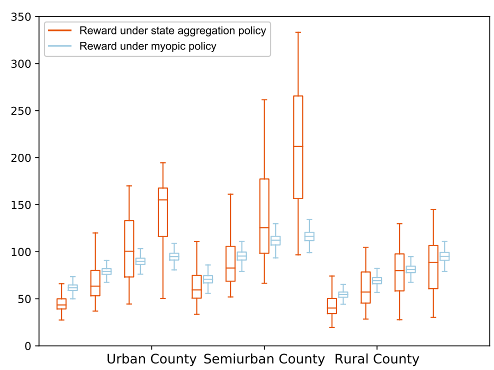
(b) Consumer drone: costs.



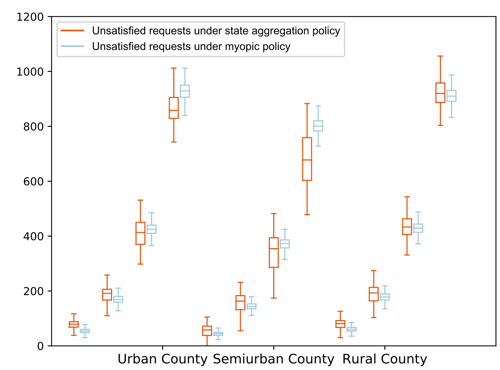
(c) Top-of-range drone: rewards.



(d) Top-of-range drone: costs.



(e) Ideal drone: rewards.



(f) Ideal drone: costs.

Figure 5: Performance comparison. For each UAV type, 12 settings are simulated for the overall arrival rate $\lambda = 0.5, 1, 2, 4$ and for the catchment area type being urban, semiurban, and rural (from left to right) and the results are presented in the two subfigures in each row.

makes more sense to capture the present reward and avoid the present cost as this myopic view is likely still better than considering an MDP solution with the stage aggregation.

Regarding the geographic distribution of the requests, our results suggest Policy A performs the best for a semiurban county and the worst in a rural county, regardless of the overall arrival rate and drone type. Table 3 and Figure 3 suggest that the request is centralized in a semiurban county while more scattered in a rural county. Intuitively, it is less beneficial to make relocation decisions with consideration of future demand when the demand is more scattered, as relocation decisions would likely bring little benefit for the location randomness of the next request. The limited benefit increment would then be offset by the suboptimality of the MDP solution with the state aggregation. On the other hand, in a semiurban county, Policy A tends to relocate the UAV in anticipation of next requests emerging at locations closer to the center of the service area with high probability. As a result, the relocation operation would likely create a large enough benefit. The cases with an urban service area are likely between the above two types as the request distribution is centralized but the area of high arrival rate is relatively large. So the performance superiority between Policy A and Policy M is highly dependent on the overall arrival rate.

For consumer drones and ideal drones, Policy A outperforms Policy M mostly in cases with a semiurban county and with high overall arrival rates, as concluded above. However, for top-of-range drones, Policy A becomes less likely to outperform Policy M under the same circumstances. This is likely because the superiority between the two policies is now related to the state of battery usage. Note that due to the computation burden, we ignore the energy level in the state space when solving the MDP. Thus, the MDP model with the state aggregation becomes a system model of even lower fidelity relative to the simulation model in cases with top-of-range drones than in cases with other two types of drones.

To summarize, Policy A gains relative superiority when the request distribution is centralized and the overall request intensity is high due to the tendency of dispatching and relocating UAVs for the sake of future rewards/costs with the MDP-based policy. On the other hand, Policy M performs well in service areas with low overall intensity and uniformly distributed requests since in these cases, the decision process resembles a series of multiple one-shot decisions as the locations of future requests are hard to predict.

5 CONCLUSIONS AND FUTURE WORK

We studied a joint optimization problem of dynamic dispatch relocation decisions in the context of deploying UAVs for emergency medical services. We formulated a discounted, infinite-horizon total-cost MDP model. This model relaxes the common but unrealistic assumption of exponential service time in the EMS dispatch decision optimization literature by incorporating the remaining service time into the state space. In addition, our model incorporates specific and significant features of drones such as energy consumption and cruise speed. To deal with the curse of dimensionality from a realistic model, we applied a state aggregation technique to obtain a policy lookup table while ensuring the balance between solution quality and tractability.

Extensive full-fledged simulation-based comparative studies with real-world data have shown that the state aggregation idea can lead to high-quality solutions within a reasonable amount of time and outperform a myopic policy currently used in practice. In the comparative studies, we selected three counties of Indiana representing three distinct types of service areas. We also considered three different types of drones, which leads to their ability of being operated without recharging. We simulated four levels of the overall request intensity, referring to the peak hours and off-hours in a day. The simulation results demonstrate that the MDP based policy gains superiority when the request distribution is more centralized and the overall request intensity is higher.

Although the state aggregation based strictly on the grid partition brings favorable results, we believe there exist flexible aggregation methods with anticipated better performance. However, it is hard to obtain a good estimate of the value function as shown in many previous studies utilizing state aggregation; otherwise, we can simply extract the policy from the approximate value function. In the future, we will adaptively aggregate the state space to avoid request losses from the service area aggregation. For example, we can construct a scaled-down MDP where each decision epoch corresponds to the locations and intensity of emergent requests. We also plan to try alternative approximate dynamic programming (ADP) approaches, such as using a neural network to approximate the value function. Furthermore, flexible state aggregation and other ADP methods are able to provide higher-quality solutions to our problem under the assumption that the request distribution is temporarily steady. However, real-world requests present time-varying nature. One way to address it in the future is to combine the request prediction with the MDP solution. Specifically, we periodically predict the distribution of future requests based on the historical data and solve the MDP model based on the predicted request distribution. The other way is reinforcement learning where we keep updating our policy through real-time policy evaluation.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation under Grant No.1761022.

REFERENCES

Andersson, T., and P. Värbrand. 2007. "Decision Support Tools for Ambulance Dispatch and Relocation". *Journal of the Operational Research Society* 58(2):195–201.

Aringhieri, R., M. E. Bruni, S. Khodaparasti, and J. T. van Essen. 2017. "Emergency Medical Services and Beyond: Addressing New Challenges through A Wide Literature Review". *Computers & Operations Research* 78:349–368.

Bélanger, V., E. Lanzarone, V. Nicoletta, A. Ruiz, and P. Soriano. 2020. "A Recursive Simulation-Optimization Framework for the Ambulance Location and Dispatching Problem". *European Journal of Operational Research* 286(2):713–725.

Bélanger, V., A. Ruiz, and P. Soriano. 2019. "Recent Optimization Models and Trends in Location, Relocation, and Dispatching of Emergency Medical Vehicles". *European Journal of Operational Research* 272(1):1–23.

Carter, G. M., J. M. Chaiken, and E. Ignall. 1972. "Response Areas for Two Emergency Units". *Operations Research* 20(3):571–594.

Dietze, P., M. Jauncey, A. Salmon, M. Mohebbi, J. Latimer, I. van Beek, C. McGrath, and D. Kerr. 2019. "Effect of Intranasal vs Intramuscular Naloxone on Opioid Overdose: A Randomized Clinical Trial". *JAMA Network Open* 2(11):e1914977–e1914977.

Grannan, B. 2014. *Dispatch, Delivery, and Location Logistics for the Aeromedical Evacuation of Time-Sensitive Military Casualties under Uncertainty*. Ph.D. thesis, Department of Systems Modeling and Analysis, Virginia Commonwealth University, Richmond, Virginia. https://scholarscompass.vcu.edu/etd/3536/.

Jagtenberg, C. J., S. Bhulai, and R. D. van der Mei. 2017. "Dynamic Ambulance Dispatching: Is the Closest-Idle Policy Always Optimal?". *Health Care Management Science* 20(4):517–531.

Jia, Q.-S. 2010. "On State Aggregation to Approximate Complex Value Functions in Large-Scale Markov Decision Processes". *IEEE Transactions on Automatic Control* 56(2):333–344.

Kim, D., K. S. Irwin, and K. Khoshnood. 2009. "Expanded Access to Naloxone: Options for Critical Response to the Epidemic of Opioid Overdose Mortality". *American Journal of Public Health* 99(3):402–407.

Lanzarone, E., E. Galluccio, V. Bélanger, V. Nicoletta, and A. Ruiz. 2018. "A Recursive Optimization-Simulation Approach for the Ambulance Location and Dispatching Problem". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, and A. Skoogh, 2530–2541. Gothenburg, Sweden: Institute of Electrical and Electronics Engineers, Inc.

McLay, L. A., and M. E. Mayorga. 2013. "A Model for Optimally Dispatching Ambulances to Emergency Calls with Classification Errors in Patient Priorities". *IIE Transactions* 45(1):1–24.

Mendelssohn, R. 1982. "An Iterative Aggregation Procedure for Markov Decision Processes". *Operations Research* 30(1):62–73.

Reuter-Oppermann, M., P. L. van den Berg, and J. L. Vile. 2017. "Logistics for Emergency Medical Service Systems". *Health Systems* 6(3):187–208.

Scholl, L., P. Seth, M. Kariisa, N. Wilson, and G. Baldwin. 2019. "Drug and Opioid-Involved Overdose Deaths—United States, 2013–2017". *Morbidity and Mortality Weekly Report* 67(5152):1419–1427.

Warner, M., L. H. Chen, D. M. Makuc, R. N. Anderson, and A. M. Miniño. 2011. "Drug Poisoning Deaths in the United States, 1980-2008". *NCHS Data Brief* 81(81):1–8.

Zarkeshzadeh, M., H. Zare, Z. Heshmati, and M. Teimouri. 2016. "A Novel Hybrid Method for Improving Ambulance Dispatching Response Time through A Simulation Study". *Simulation Modelling Practice and Theory* 60:170–184.

AUTHOR BIOGRAPHIES

XIAOQUAN GAO is a current Ph.D. student in the School of Industrial Engineering at Purdue University. She received her B.S. in Theoretical and Applied Mechanics from Peking University in 2019. Her research interest is predictive analytics, stochastic modeling, and decision optimization under uncertainty in healthcare systems. Her e-mail address is gao568@purdue.edu.

NAN KONG received his PhD in industrial engineering from the University of Pittsburgh in 2006. His research focuses on applying operations engineering techniques (especially stochastic optimization) to improving healthcare delivery outcomes. He is the corresponding author and his e-mail address is nkong@purdue.edu.

PAUL M. GRIFFIN received his PhD in industrial engineering from Texas AM University. His research focuses on health systems engineering, health analytics, and cost-effectiveness analysis. His email address is pmg14@psu.edu.