UAV-enabled Human Internet of Things

Kelly Rael, Georgios Fragkos, Jim Plusquellic, Eirini Eleni Tsiropoulou
Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA
kelrael@unm.edu, gfragkos@unm.edu, jplusq@unm.edu, eirini@unm.edu

***Abstract*—In this paper, an Unmanned Aerial Vehicles (UAVs) - enabled human Internet of Things (IoT) architecture is introduced to enable the rescue operations in public safety systems (PSSs). Initially, the first responders select in an autonomous manner the disaster area that they will support by considering the dynamic socio-physical changes of the surrounding environment and following a set of gradient ascent reinforcement learning algorithms. Then, the victims create coalitions among each other and the first responders at each disaster area based on the expected-maximization approach. Finally, the first responders select the UAVs that communicate with the Emergency Control Center (ECC), to which they will report the collected data from the disaster areas by adopting a set of log-linear reinforcement learning algorithms. The overall distributed UAV-enabled human Internet of Things architecture is evaluated via detailed numerical results that highlight its key operational features and the performance benefits of the proposed framework.**

***Index Terms*—Public Safety Systems, Unmanned Aerial Vehicles, Human Internet of Things, Reinforcement Learning**

I. Introduction

Public Safety Systems (PSSs) support the emergency response and disaster relief services during and after emergency events, including catastrophic events, which can be natural or man-made. PSSs consist of special service personnel, i.e., first responders, such as emergency medical personnel, police officers, firefighters, law enforcement experts, military personnel, local, national and non-government organizations and others. Recently, Unmanned Aerial Vehicles (UAVs) have been used to assist the rescue operations. Due to the UAVs unique characteristics, they are able to establish resilient, reliable, scalable, and robust communication even during catastrophic events, where the infrastructure-based communication is damaged or non-existing [1]. The UAVs' salient attributes are their fast, flexible, and effortless deployment, mobility, maneuverability, line-of-sight (LoS) communication links due to their ability to hover above a disaster area, adaptive altitude, adjustable usage, and low-cost [2]. The cooperation of the UAVs with the first responders is of vital importance for a successful rescue operation. Some of the research challenges that have attracted the interest of the industrial and research community are the UAVs path planning, the UAVs' positioning and allocation to disaster areas, the spectrum sharing among the UAVs and the ground base stations, the first responders' autonomous and distributed data and computation tasks offloading to the UAVs, and others [3].

The research of Mr. Fragkos and Dr. Tsiropoulou was conducted as part of the NSF CRII-1849739.

A. Related Work & Motivation

UAVs have been extensively used to support the communication in disaster areas, where part of the ground communication infrastructure is damaged [4], [5]. In [6], the problem of spectrum sharing among the UAVs and the licensed terrestrial networks is studied. A distributed multi-agent reinforcement learning (RL) algorithm is proposed in order to enable the UAVs to decide if they will act as sensing nodes (i.e., collecting information from the first responders and the victims) or relaying nodes (i.e., retransmitting the already collected data to the rescue operation center) based on the spectrum availability. This work has been extended in [7], where the UAVs are autonomously allocated in disaster areas based on the priority of the catastrophic event by following a distributed RL algorithm. The ultimate goal of the UAVs is to maximize their achievable throughput and prolong their battery lifetime. In [8], a game-theoretic approach is introduced to create clusters of victims in a PSS and a reinforcement learning mechanism is proposed to determine the clusterheads that communicate directly with the UAVs.

In [9], a UAVs' data collection from the vehicles scenario is studied in a disaster area by introducing a blockchain-based collaborative aerial-ground network architecture. The authors introduce a credit-based consensus algorithm to monitor the data transactions from the vehicles in an energy efficient and secure manner. In [10], the impact of victims' behavior on their transmission power investment to communicate with the UAVs and ground base stations is studied under the principles of Prospect Theory. In [11], the problem of UAVs deployment in the disaster areas towards satisfying the victims' communication needs and establishing connections with optimal throughput is studied via introducing a matching game algorithm among the UAVs and the groups of victims.

The UAVs have also been considered in wireless powered communication networks in disaster areas, where the victims devices harvest energy from the UAVs' radio frequency signals [12]. In [13], the authors jointly optimize the UAVs' position and the UAVs', victims', and first responders' transmission power to improve the reliability of data collection by the UAVs. The concepts of information quality and criticality, as well as value of information collected by the first responders in a disaster area are introduced in [14] to quantify each first responder's contribution in the rescue operation. In [15], the PSS is supported by mobile and static UAVs hovering above the disaster area and the victims' data offloading and the corresponding transmission power to the multiple UAVs are determined following the principles of Prospect Theory and the theory of the Tragedy of the Commons.

However, despite the significant advances achieved by these efforts, the joint problem of first responders' allocation in the different disaster areas, the autonomous organization of the victims in rescue groups, and the communication-driven association of the first responders with the UAVs has been neglected or partially studied in the existing literature. In this paper we aim to address this research gap by introducing a three layers approach, i.e., (i) the distributed association of the first responders to different disaster areas based on different gradient ascent reinforcement learning algorithms; (ii) the victims' coalition formation mechanism to form rescue groups following the expectation-maximization approach; and (iii) the autonomous association of the first responders with the UAVs to enable the robust data flow to the Emergency Control Center (ECC) based on a set of log-linear reinforcement learning algorithms.

B. Contributions & Outline

The key technical contributions of this research work are summarized as follows.

- A UAV-enabled human Internet of Things (IoT) public safety system is introduced consisting of UAVs, first responders, and victims interacting among each other (Section II). The first responders act as autonomous decision makers, making stable decisions regarding the disaster areas that they will assist by considering the dynamic socio-physical changes of the surrounding disaster environment. Their decisions are supported by a set of different gradient ascent reinforcement learning algorithms that enable them to adapt to the dynamically changing needs of the disaster areas in a real time manner (Section III).
- At each disaster area, the victims create coalitions among each other, i.e., rescue groups, by being associated with a first responder that acts as a coalition-head. A K-means algorithm is introduced to build the victims coalitions based on the expected-maximization approach (Section IV).
- Given the first responders allocation to the disaster areas and the rescue groups coalition formation, each first responder selects to report the collected information from the disaster field to a UAV in a distributed and autonomous manner by considering the UAVs' physical and communication characteristics. The UAVs selection by the first responders is performed by adopting a set of log-linear reinforcement learning algorithms (Section V).
- The UAV-enabled human Internet of Things architecture is introduced by combining the three individual mechanisms into an operational system to support the rescue operations in PSSs (Section VI).
- A set of detailed numerical results is presented to evaluate the performance of the proposed framework, while a comparative study demonstrates its superiority compared to alternative approaches (Section VII). Finally, Section VIII concludes the paper.

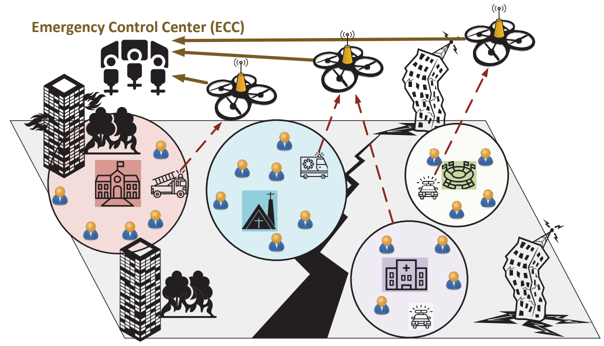

Fig. 1: UAV-enabled Human Internet of Things Topology

II. System Model

We consider a disaster-struck environment consisting of $|A|$ disaster areas and their set is denoted as $A = \{1, ..., a, ..., |A|\}$. A set of first responders $F = \{1, ..., f, ..., |F|\}$, such as police officers, firefighters, and medical personnel, is available to support the rescue operation of the Emergency Control Center (ECC). At each disaster area a, a set of victims $V_a = \{1, ..., v, ..., |V_a|\}$ requests help from the first responders. A set of UAVs $U = \{1, ..., u, ..., |U|\}$ enables the communication of the first responders with the ECC given that the infrastructure-based communication is limited or non-existing due to the disaster, e.g., earthquake, wildfire, terrorists attack [16].

Each critical disaster area is characterized by a minimum number of first responders N_a needed to cover the victims' rescue needs and contributes to a successful rescue operation. The distance of a first responder f from a critical disaster area a is denoted as $d_{f,a}$[m]. Moreover, each first responder f, based on its capabilities and specialty (e.g., police officer, firefighter, medical personnel) wants to offer its services to the candidate critical areas that a specific catastrophic event has occurred. For example, a police officer has a higher interest to offer its services to a disaster area that a shooting has occurred compared to an area that a wildfire has occurred. The first responders' service interest for a disaster area is denoted as $i_{f,a}$, $i_{f,a} \in [0, 1]$. Furthermore, the first responders tend to collaborate among each other in order to conclude to a successful rescue operation. Thus, a first responder has a higher interest to go to a critical area that other first responders with complementary specialties, expertise, and capabilities reside in order to complement the rescue process. The interest of a first responder to go to a critical disaster area a that another first responder f' with complementary specialty already offers its services is denoted as $CS_{f,f'}, CS_{f,f'} \in [0,1]$, $f, f' \in F_a$ where F_a is the set of first responders that have decided to support disaster area a. Moreover, each critical disaster area a is characterized by an importance factor $I_a, I_a \in [0,1]$ depending on the criticality of the catastrophic event that occurred in it. For example, a shooting event at a school resulting in multiple victims has a greater importance factor compared to a car accident. The overall considered UAV-enabled human Internet of Things topology in a public safety environment is presented in Fig.1.

III. Autonomous First Responders Allocation

In this section, the problem of enabling the first responders to select a disaster area to offer their services in a distributed and autonomous manner is studied. During a catastrophic event, the first responders have limited available information to make their decisions, and little or even no collaboration exists among the various public safety agencies (e.g., police department, fire department) given their rescue operation protocols [17]. Moreover, the disaster environment changes dynamically and requires multiple sequential decisions by the first responders, which makes the problem of allocating the first responders to the critical disaster areas even more complex. Also, the problem becomes even more complicated given the partial available information to the first responders, who may not even be aware of the presence of other first responders, thus making the public safety environment seem non-stationary.

Considering the aforementioned challenges that the public safety environment imposes to the first responders' decisions who offer their services to the victims of a disaster area, the principles of multi-agent reinforcement learning have been adopted in this research work to address the examined problem. The first responders are characterized by strategic interactions among them. In other words, the first responders act as autonomous entities, having individual goals and independent decision making capabilities, while at the same time they experience the impact of each others' decisions. Thus, we adopt a set of *gradient ascent reinforcement learning approaches*, where the first responders learn their environment by performing gradient updates of their perceived reward.

Specifically, we adopt the theory of Learning Automata (LA), where an action probability vector $\mathbf{P_f}$ characterizes the decisions of each first responder. The first responders explore their available potential actions, i.e., visit a disaster area, and based on the reward that they receive by this action, they update the corresponding action probabilities. The first responders make their final decision to visit a disaster area, if $P_{f,a} \geq P_{thres}, \forall f \in F, \forall a \in A$, where P_{thres} is a threshold value of the action probability. The set of the first responders' actions is the number of critical disaster areas $A = \{1, ..., a, ..., |A|\}$, where they can offer their services to contribute in the rescue operation. Thus, each first responder's action probability vector is denoted as $\mathbf{P_f} = [P_{f,1}, ..., P_{f,a}, ..., P_{f,|A|}]$. The most commonly applied learning update rule is called Linear Reward-Penalty (LRP) and each first responder's action probabilities are updated as follows:

$$P_{f,a}^{(ite+1)} = P_{f,a}^{(ite)} + \lambda_1 \hat{r}_{f,a}^{(ite)} (1 - P_{f,a}^{(ite)}) - \lambda_2 (1 - \hat{r}_{f,a}^{(ite)}) P_{f,a}^{(ite)}, \quad if \quad a^{(ite+1)} = a^{(ite)} \tag{1a}$$

$$P_{f,a}^{(ite+1)} = P_{f,a}^{(ite)} - \lambda_1 \hat{r}_{f,a}^{(ite)} P_{f,a}^{(ite)} + \lambda_2 (1 - \hat{r}_{f,a}^{(ite)}) \left(\frac{1}{|A|-1} - P_{f,a}^{(ite)}\right), \quad if \quad a^{(ite+1)} \neq a^{(ite)} \tag{1b}$$

where Eq. 1a represents the probability that the first responder f selects the same critical disaster area in the next iteration of the LRP algorithm, and Eq. 1b captures the probability to choose a different disaster area. The iteration of the LRP algorithm is denoted as *ite*. The parameter λ_1 and λ_2, $\lambda_1, \lambda_2 \in [0,1]$ are learning rate constant parameters, representing the reward and the penalty that a first responder will experience by exploring an action. The parameter $\hat{r}_{f,a}^{(ite)}$ represents the normalized reward that a first responder experiences by selecting an action, i.e., a disaster area $a^{(ite)}$. The normalized reward is defined as follows.

$$\hat{r}_{f,a}^{(ite)} = \frac{r_{f,a}^{(ite)}}{\sum\limits_{\forall a \in A} r_{f,a}^{(ite)}} \tag{2}$$

where $r_{f,a}^{(ite)}$ is the actual reward that the first responder experiences by potentially offering its services to the disaster area a, and is defined as follows.

$$r_{f,a}^{(ite)} = \frac{\frac{N_a}{\sum\limits_{\forall a \in A} N_a} \cdot \frac{|V_a|}{\sum\limits_{\forall a \in A} |V_a|} \cdot i_{f,a}^{(ite)} \cdot I_a \cdot \sum\limits_{f'=1, f' \neq f}^{|F_a|} CS_{f,f'}^{(ite)}}{\frac{d_{f,a}}{\sum\limits_{\forall f \in F} \sum\limits_{\forall a \in A} d_{f,a}} \cdot c_{f,a}} \tag{3}$$

where $c_{f,a} \in (0,1]$ is the personal cost, e.g., commuting cost to reach the disaster area a, that the first responder f experiences. The physical notion of the first responders personalized reward function $r_{f,a}^{(ite)}$ is that a first responder prefers to offer its services to a disaster area that: (i) needs an increased number of first responders N_a in order to deal successfully with the rescue operation; (ii) a large number of victims reside in this area; (iii) has high service interest based on the catastrophic event; (iv) the importance and criticality of the catastrophic event in the disaster area is high; (v) can perform the rescue operation by collaborating with other first responders with complementary specialties; (vi) has low personal cost to reach that area; and (vii) is in close proximity to the first responder.

The LRP action probabilities update rule, as presented in Eq.1a, 1b, allows the first responders to thoroughly explore the available disaster areas and finally, select the one that they can constructively and successfully offer their services. For the LRP algorithm, we consider $\lambda_1 = \lambda_2$. In the special case that $\lambda_1 >> \lambda_2$, the learning algorithm allows the first responders to explore less other potential disaster areas to offer their services and it is called Linear Reward - ϵ Penalty (LR-ϵP). Furthermore, in the case that $\lambda_2 = 0$, then the first responders probabilistically select the disaster area that provides them the best reward with very limited exploration of alternative actions and the algorithm is called Linear Reward-Inaction (LRI). Given that the LRP algorithm allows the first responders to thoroughly explore their available choices compared to the LR-ϵP and the LRI algorithms, it is expected to need increased computational time to converge to a stable decision. However, the first responders will experience higher rewards by their well-thought choices following the LRP learning approach. A detailed comparative analysis of the three presented decision-making algorithms, i.e., LRP, LR-ϵP, and LRI, is provided in Section VII-A, highlighting their benefits and drawbacks.

IV. Victims Coalition Formation

In this section, given that the first responders have selected to which disaster area a they will offer their services, we propose a coalition formation mechanism among the first responders and the victims in each disaster area a, $\forall a \in A$. The coalitions among the victims and the first responders are constructed by adopting the K-means algorithm [18]. The K-means algorithm gets as input the number of coalitions that should be created and follows the expectation-maximization approach to build the coalitions. In the examined problem in this research work, the victims want to reach help from the first responders $|F_a|$ that have arrived in the disaster area. Thus, the number of coalitions that will be created is $|F_a|$, and each first responder acts as coalition-head at the corresponding created coalition.

Each victim $v \in V_a$ ideally wants to reach out for help from a first responder in the closest proximity during a catastrophic scenario. We consider the coordinates of each victim $(x_v, y_v), \forall v \in V_a$, and each first responder $(x_f, y_f), \forall f \in F_a$. Thus the objective function of the K-means coalition formation algorithm is defined as follows.

$$I = \sum_{v=1}^{|V_a|} \sum_{f=1}^{|F_a|} ||(x_v, y_v) - (x_f, y_f)||^2 \quad (4)$$

where $||\cdot||$ denotes the Euclidean distance.

The K-means algorithm aims to minimize the objective function in Eq.4 by assigning the victims to their closest first responder. Thus, it concludes in creating homogeneous coalitions among the victims and the first responders in terms of grouping together victims that are in the neighborhood (i.e., close proximity) of each first responder. The outcome of the K-means algorithm is the set $V_{f,a}$ of the victims in the disaster area a that are supported by the first responder f. Indicative numerical results are presented in Section VII-B, showing the outcome of creating coalitions among the victims and the first responders based on the K-means algorithm.

V. UAVs – First Responders Association

In this section, a distributed solution is proposed to enable the first responders to select a UAV in an autonomous manner in order to report the collected information from the disaster area. The UAVs hover in various positions above the overall disaster environment and act as relay nodes to retransmit the first responders' collected information to the ECC, which oversees and performs the rescue operation planning. The first responders, given their limited available information regarding the situation in the surrounding environment, should be able to make autonomous communication decisions with the UAVs in a fast manner by exploiting the UAVs' physical and communications characteristics.

Specifically, each UAV is characterized by its normalized flying time, $FT_u, FT_u \in [0, 1]$, and its normalized flying cost $C_u, C_u \in [0, 1]$ capturing its fuels and cost. Each UAV $u \in U$ can collect (i.e., process and retransmit) an amount of data B_u [bits] from the disaster field. Also, the distance of a UAV u from a first responder f is denoted as $d_{u,f}[m]$. By considering the aforementioned physical and communication characteristics of the UAVs, the utility that a first responder f experiences by selecting a UAV u is defined as follows.

$$U_{f,u}^{[\tau+1]} = \frac{FT_u^{[\tau]} \cdot B_u}{d_{u,f}^{[\tau]} \cdot C_u} \quad (5)$$

where τ denotes the time instance of examining the public safety system. The physical notion of Eq.5 is that a first responder prefers to select a UAV that: (i) has a long flying time, thus allowing the first responder to establish a stable communication link; (ii) is in the first responder's close proximity, thus the latter spends low levels of power to communicate with it; (iii) can handle a large amount of data; and (iv) has a low flying cost, thus contributing to a cost-efficient rescue operation. Based on Eq.5, the normalized utility that each first responder experiences by selecting to communicate with a UAV is given as follows.

$$\hat{U}_{f,u}^{[\tau+1]} = \frac{U_{f,u}^{[\tau+1]}}{\sum\limits_{\forall u \in U} U_{f,u}^{[\tau+1]}} \quad (6)$$

Towards enabling the first responders to select a UAV to report their data and maximize their perceived utility, the log-linear reinforcement learning algorithms have been adopted. The log-linear reinforcement learning algorithms enable the first responders, who may have competing interests in terms of selecting UAVs, to converge to a stable decision with high probability and no information exchange among each other. The main characteristic of this set of learning algorithms is that they allow the decision-makers, i.e., first responders, to "make mistakes", i.e., select suboptimal actions, in order to thoroughly explore their actions space, i.e., UAVs choices [19].

Two log-linear reinforcement learning algorithms are studied in this research work, namely Binary Log-Linear Learning (BLLL) and Max Log-Linear Learning (MLLL) algorithms. Each first responder's action space is the set of UAVs $U = \{1, ..., u, ..., |U|\}$ and initially selects randomly a UAV with equal probability, i.e., $P_{f,u}^{(\tau=0)} = \frac{1}{|U|}$. At each iteration τ of the BLLL and MLLL algorithms, one first responder is randomly selected and performs exploration and learning. Specifically, the first responder f explores an alternative choice of UAV $\tilde{u}_f^{(\tau)}$ as its new strategy with equal probability $\frac{1}{|U|}$ and receives its corresponding utility $\tilde{U}^{(\tau)}_{f,\tilde{u}_f^{(\tau)}}$ *(exploration phase)*. Then, the first responder updates its strategy based on the probabilistic rules of Eq.7a, 7b regarding the BLLL algorithm and the Eq.8a, 8b regarding the MLLL algorithm *(learning phase)* [14].

$$P_{f,u}^{(\tau+1)}[u_f^{(\tau+1)} = \tilde{u}_f^{(\tau)}] = \frac{e^{\tilde{U}^{(\tau)}_{f,\tilde{u}_f^{(\tau)}} \cdot \beta}}{e^{\tilde{U}^{(\tau)}_{f,\tilde{u}_f^{(\tau)}} \cdot \beta} + e^{U^{(\tau)}_{f,u_f^{(\tau)}} \cdot \beta}} \tag{7a}$$

$$P_{f,u}^{(\tau+1)}[u_f^{(\tau+1)} = u_f^{(\tau)}] = \frac{e^{U^{(\tau)}_{f,u_f^{(\tau)}} \cdot \beta}}{e^{\tilde{U}^{(\tau)}_{f,\tilde{u}_f^{(\tau)}} \cdot \beta} + e^{U^{(\tau)}_{f,u_f^{(\tau)}} \cdot \beta}} \tag{7b}$$

$$P_{f,u}^{(\tau+1)}[u_f^{(\tau+1)} = \tilde{u}_f^{(\tau)}] = \frac{e^{\tilde{U}^{(\tau)}_{f,\tilde{u}_f^{(\tau)}} \cdot \beta}}{\max\{e^{\tilde{U}^{(\tau)}_{f,\tilde{u}_f^{(\tau)}} \cdot \beta}, e^{U^{(\tau)}_{f,u_f^{(\tau)}} \cdot \beta}\}} \tag{8a}$$

$$P_{f,u}^{(\tau+1)}[u_f^{(\tau+1)} = u_f^{(\tau)}] = \frac{e^{U^{(\tau)}_{f,u_f^{(\tau)}} \cdot \beta}}{\max\{e^{\tilde{U}^{(\tau)}_{f,\tilde{u}_f^{(\tau)}} \cdot \beta}, e^{U^{(\tau)}_{f,u_f^{(\tau)}} \cdot \beta}\}} \tag{8b}$$

It is noted that Eq. 7a and Eq. 8a capture the probability that a first responder f explores the communication with an alternative UAV $\tilde{u}_f^{(\tau)}$ in the iteration $\tau + 1$, while the Eq. 7b and Eq. 8b represent the probability of selecting the same UAV in the next iteration $\tau + 1$, regarding the BLLL and MLLL algorithms, respectively. Moreover, the learning parameter $\beta, \beta \in \mathbb{R}^+$ captures the allowance of the first responder to explore alternative actions. Specifically, for large values of β, the first responder is allowed to explore more thoroughly its communications choices with the UAVs and converge to a more beneficial choice, however, by spending more time to converge to its best decision. The BLLL and MLLL algorithms converge when the system's social welfare remains approximately the same for a small number of consecutive iterations of the algorithms. Detailed numerical results regarding the drawbacks and benefits of the BLLL and MLLL algorithms are presented in Section VII-C.

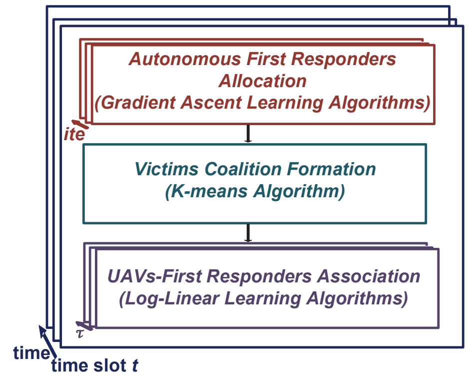

Fig. 2: UAV-enabled Human Internet of Things Architecture

VI. UAV-enabled Human Internet of Things Architecture

In this section, the overall architecture of the proposed UAV-enabled human Internet of Things (IoT) is introduced by combining the three individual mechanisms presented in Sections III - V into an operational system that can support the rescue operations during and after disaster scenarios. The overall UAV-enabled human IoT architecture is presented in Fig. 2 and consists of three layers that are periodically executed over the time.

Initially, the first responders in the time t select which disaster area to visit and offer their services following one of the three introduced gradient ascent learning algorithms. Then, the victims at each disaster area create coalitions among them and the first responders that have arrived, by implementing the K-means algorithm. Finally, the first responders select in an autonomous manner to which UAV they will report their collected data from the disaster area by exploiting the UAVs physical and communication characteristics and following one of the proposed log-linear learning algorithms.

VII. Numerical Results

In this section, we provide a set of detailed numerical results to illustrate the performance of the proposed UAV-enabled human Internet of Things architecture to support the public safety systems in terms of the following aspects: autonomous first responders allocation based on the gradient ascent learning algorithms (Section VII-A), victims coalition formation based on the K-means algorithm (Section VII-B), and the first responders' association with the UAVs based on the log-linear learning algorithms (Section VII-C). Additional comparative results of the various examined learning algorithms are presented to study their drawbacks and benefits in terms of applying them in the rescue operation planning in public safety scenarios.

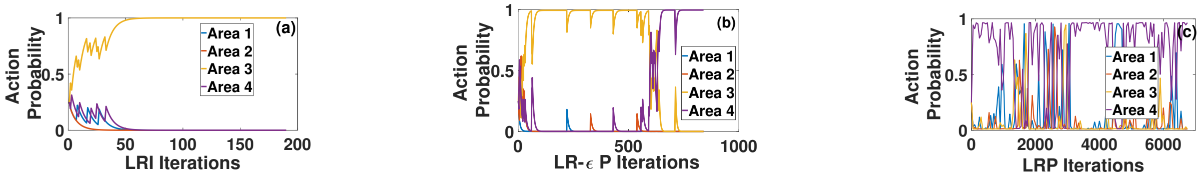

Fig. 3: Gradient Ascent Learning Algorithms – Convergence

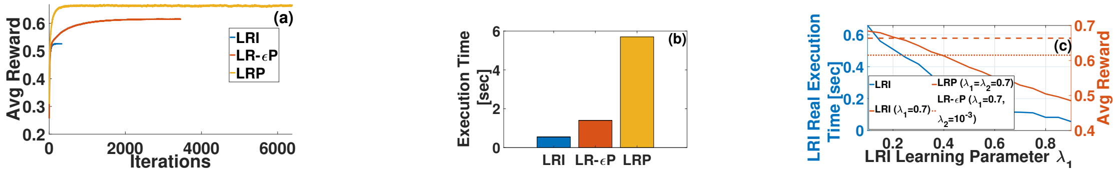

Fig. 4: Gradient Ascent Learning Algorithms – Comparative Evaluation

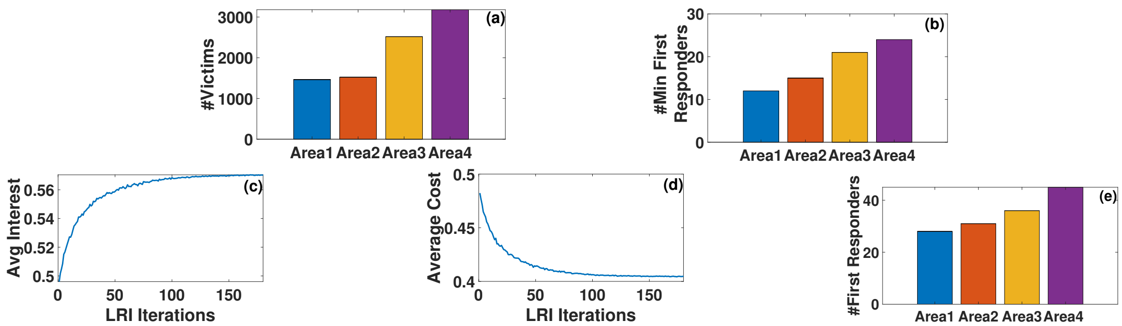

Fig. 5: Gradient Ascent Learning Algorithms – Pure Framework Evaluation

We considered a disaster-struck environment consisting of $|A| = 4$ disaster areas and each area $a \in A$ has $V_a \in [800, 3500]$ victims, who need at least $N_a \in [5, 35]$ first respondres. Moreover, we assume that there are $|F| = 140$ first responders available, where each first responder's distance $d_{f,a}$ from each disaster area is randomly and uniformly distributed in the interval $[80m, 250m]$. Furthermore, regarding the coordinates of both the victims and first responders we consider that $x_u, y_u, x_f, y_f \in [10m, 800m]$. The available UAVs in our scenario are $|U| = 7$, with the following characteristics: $B_u \in [50, 100] MBytes$ and $d_{u,f} \in [100m, 300m]$. Finally, for all the gradient ascent learning algorithms, we consider $\lambda_1 = 0.7$, whereas for the LRI approach $\lambda_2 = 0$, for the LR-ϵP approach $\lambda_2 = 0.001$ and for the LRP approach $\lambda_2 = 0.7$. The proposed framework's evaluation was conducted via modeling and simulation and executed in a MacBook Pro Laptop, 2.5GHz Intel Core i7, with 16GB LPDDR3 available RAM.

A. Autonomous First Responders Allocation based on Gradient Ascent Learning

In this subsection, we initially study and analyze the convergence of the three alternative gradient ascent reinforcement learning algorithms, i.e., LRI, LR-ϵP, and LRP as they were introduced in Section III, to the first responders' autonomous and stable decision to visit a disaster area and offer their services to the victims. Fig. 3a - 3c present the convergence of one indicative first responder's action probabilities to its stable decision to offer its services to a specific critical disaster area. Also, Fig. 4a shows the average reward achieved by all the first responders in the examined public safety system as a function of the iterations for the three examined gradient ascent learning algorithms. Fig. 4b illustrates the convergence time of the three gradient ascent learning algorithms in order all the first responders in the examined setup to converge to their stable solutions. The results reveal that the LRI algorithm allows the first responders to make a fast decision (Fig. 3a), i.e., 180 iterations cooresponding on average to 0.55 seconds (Fig. 4b) compared to the LR-ϵP algorithm (Fig. 3b) that needs approximately 800 iterations, i.e., 1.4 seconds on average (Fig. 4b), and the LRP algorithm (Fig. 3c), where almost 7,000 iterations are needed to converge, i.e., 5.7 seconds on average (Fig. 4b). This behavior of the three gradient ascent learning algorithms stems from the fact that the LRI algorithm increases the probability of a selected action that resulted in a good reward value (Eq. 3), while decreases the probability of selecting any other disaster area. Thus, the LRI algorithm provides limited freedom to the first responders to explore their strategy space and concludes to low achieved rewards for the first responders (Fig. 4a). On the other hand, the LR-ϵP algorithm allows the first responders to slightly explore alternative actions other than that one which provides them the highest reward, by setting $\lambda_1 >> \lambda_2$ (Eq. 1a, 1b). Thus, the first responders explore slightly more their alternative choices compared to the LRI algorithm and need more time to converge to their stable decision (Fig. 4b), however, they achieve higher reward (Fig. 4a). The greatest exploration freedom

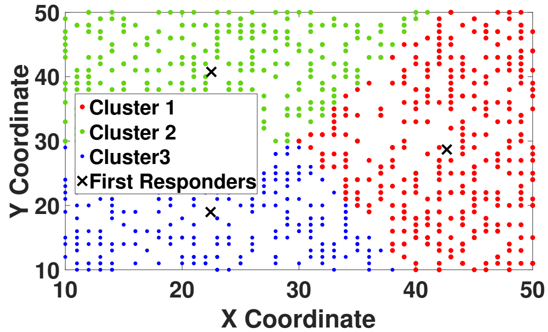

Fig. 6: K-Means Coalitions Formation Mechanism

is provided by the LRP algorithm to the first responders by setting $\lambda_1 = \lambda_2$. Thus, the LRP algorithm presents the slowest convergence to the first responders' stable decisions (Fig. 4b), while they achieve the highest reward (Fig. 4a) by thoroughly learning their surrounding disaster areas.

Moreover, Fig. 4c presents the execution time and the corresponding first responders' average reward as a function of the learning parameter λ_1 considering the LRI algorithm. The results reveal that as the learning parameter λ_1 increases, the first responders converge faster to their stable decisions, however, they achieve lower reward, as they perform very limited exploration of their available choices. Also, for an indicative value of the learning parameter $\lambda_1 = 0.7$, we observe that the LR-ϵP and the LRP algorithms contribute to 16.2% and 25.3% improvement of the first responders' average normalized reward compared to the LRI algorithm, respectively.

In the following analysis, we adopt the LRI algorithm, i.e., worst case scenario given that it allows the first responders to perform the most limited exploration of their actions and converges fast to a stable decision, to study the performance of the proposed autonomous first responders allocation to the disaster areas. Fig. 5a, 5b present the number of victims $|V_a|$ and the minimum number of first responders that are needed per disaster area, respectively. Fig. 5c, 5d illustrate the first responders' average service interest $i_{f,a}$ and personal cost $c_{f,a}$, respectively, as a function of the LRI algorithms iterations. Also, Fig. 5e shows the number of first responders that selected to offer their services to each disaster area after the LRI algorithm has converged. The results reveal that even with the simplest selected gradient ascent learning algorithm (i.e., LRI algorithm – worst case scenario), the first responders successfully learn during the algorithm's iterations to select the disaster areas that have higher interest to serve (Fig. 5c) and lower personal cost to reach them (Fig. 5d). Also, the proposed approach achieves to successfully allocate more first responders (Fig. 5e) to the disaster areas that are in greater need of support, i.e., they have more victims (Fig. 5a) and they need a greater minimum number of first responders (Fig. 5b).

B. Victims Coalition Formation Evaluation

In this section, we present the operation of the victims' coalition formation mechanism in an indicative sub-area of a disaster area. Fig. 6 presents the topology of the sub-area, where three first responders are offering their services to the victims, who are presented with the multicolored dots. Each created coalition is presented with a different color. Indeed, the results reveal that the victims tend to create coalitions among each other and the first responder in the closest proximity to them. Thus, homogeneous coalitions are created, where their members are close to each other and can help each other during the rescue operation process.

C. UAVs – First Responders Association based on Log-Linear Learning

In this section, we evaluate the proposed distributed decision framework based on the introduced log-linear algorithms, i.e., BLLL and MLLL algorithms, that enable the first responders to select a UAV in order to report their collected data from the disaster field to the ECC. Fig. 7a, 7b present the first responders' average utility (Eq.5) as a function of the iterations regarding the BLLL and MLLL algorithms, respectively, for different values of the learning parameter β. The results reveal that for greater values of the learning parameter β, both the BLLL and MLLL algorithms allow the first responders to thoroughly explore their communication choices with the UAVs by spending more iterations in order to converge to their stable decisions that conclude to greater average utility. Also, it is confirmed that the MLLL algorithm converges faster to a stable decision for all the first responders compared to the BLLL algorithm given the form of its probability updating rule, i.e., Eq. 8a, 8b. A detailed Monte Carlo simulation analysis is illustrated in Fig. 7c, 7d for 10,000 executions of the BLLL and MLLL algorithms. Specifically, Fig. 7c and 7d present the first responders' average utility after the algorithms' convergence and the corresponding real execution time, respectively. The results confirm that the MLLL algorithm achieves better utility for the first responders in a shorter execution time, thus, it outperforms compared to the BLLL algorithm.

Fig. 8 presents the operation of the first responders' autonomous selection of the UAVs to report their information to the ECC based on the UAV's physical and communication characteristics by adopting the MLLL algorithm. Fig. 8a-8d present the UAVs normalized flying cost C_u, amount of data that they can collect B_u, normalized flying time FT_u, and averaged distance from the first responders in the examined public safety system, respectively. Also, Fig. 8e and Fig. 8f show the first responders' average exploring probability (Eq. 8a, 8b) and the number of first responders that selected each UAV after the MLLL algorithm has converged. The results reveal that more first responders (Fig. 8f) select with higher average probability (Fig. 8e) the UAVs that are characterized by low flying cost (Fig.

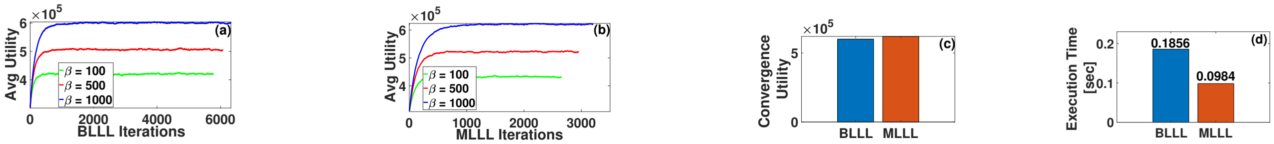

Fig. 7: Log-Linear Learning Algorithms - Comparative Evaluation

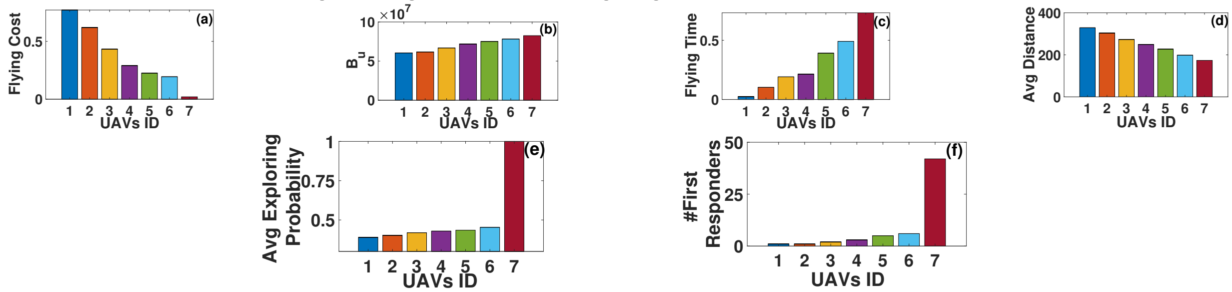

Fig. 8: Log-Linear Learning Alghorithms - Pure Framework Evaluation

8a), high flying time (Fig. 8c), small average distance from them (Fig. 8d), and they are robust in terms of collecting a large amount of data (Fig. 8b).

VIII. Conclusions

In this paper, a distributed UAV-enabled human Internet of Things architecture is introduced, to support the rescue operations during and after catastrophic events in public safety systems. The first responders select the disaster area that they will provide their services in a distributed manner by adopting various gradient ascent learning algorithms and sensing the dynamic changes of the disaster environment. Then, the victims at each disaster area create coalitions among each other and the first responders that have arrived based on the K-means algorithm. Also, the first responders choose the UAVs to which they will report the collected information from the disaster field in an autonomous manner following the log-linear learning approach. A set of detailed numerical and comparative results are presented. Part of our current and future work examines the rescue operations in public safety systems by adopting concepts from the labor economics and contract theory to create stable relationships among the first responders and the victims.

References

[1] C. Wu, B. Ju, Y. Wu, X. Lin, N. Xiong, G. Xu, H. Li, and X. Liang, "Uav autonomous target search based on deep reinforcement learning in complex disaster scene," *IEEE Access*, vol. 7, pp. 117 227–117 245, 2019.

[2] P. Vamvakas, E. E. Tsiropoulou, and S. Papavassiliou, "Exploiting prospect theory and risk-awareness to protect uav-assisted network operation," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, p. 286, 2019.

[3] M. Erdelj and E. Natalizio, "Drones, smartphones and sensors to face natural disasters," in *Proc. of the 4th ACM Workshop on Micro Aerial Vehicle Net., Systems, and Appl.*, 2018, pp. 75–86.

[4] B. Wang, Y. Sun, N. Zhao, and G. Gui, "Learn to coloring: Fast response to perturbation in uav-assisted disaster relief networks," *IEEE Transactions on Vehicular Technology*, 2020.

[5] G. Fragkos, P. A. Apostolopoulos, and E. E. Tsiropoulou, "Escape: Evacuation strategy through clustering and autonomous operation in public safety systems," *Future Internet*, vol. 11, no. 1, p. 20, 2019.

[6] A. Shamsoshoara, M. Khaledi, F. Afghah, A. Razi, and J. Ashdown, "Distributed cooperative spectrum sharing in uav networks using multi-agent reinforcement learning," in *16th IEEE Annual Consumer Comm. & Net. Conf.* IEEE, 2019, pp. 1–6.

[7] A. Shamsoshoara, F. Afghah, A. Razi, S. Mousavi, J. Ashdown, and K. Turk, "An autonomous spectrum management scheme for unmanned aerial vehicle networks in disaster relief operations," *preprint arXiv:1911.11343*, 2019.

[8] D. Sikeridis, E. EleniTsiropoulou, M. Devetsikiotis, and S. Papavassiliou, "Self-adaptive energy efficient operation in uav-assisted public safety networks," in *IEEE SPAWC*, 2018, pp. 1–5.

[9] Z. Su, Y. Wang, Q. Xu, and N. Zhang, "Lvbs: Lightweight vehicular blockchain for secure data sharing in disaster rescue," *IEEE Trans. on Dependable and Secure Comput.*, pp. 1–1, 2020.

[10] P. Vamvakas, E. E. Tsiropoulou, and S. Papavassiliou, "On the prospect of uav-assisted communications paradigm in public safety networks," in *IEEE INFOCOM WKP*, 2019, pp. 762–767.

[11] K. Ali, H. X. Nguyen, Q.-T. Vien, P. Shah, and M. Raza, "Deployment of drone based small cells for public safety communication system," *IEEE Systems Journal*, 2019.

[12] D. Sikeridis, E. E. Tsiropoulou, M. Devetsikiotis, and S. Papavassiliou, "Wireless powered public safety iot: A uav-assisted adaptive-learning approach towards energy efficiency," *Journal of Network and Computer Appl.*, vol. 123, pp. 69–79, 2018.

[13] Z. Huang, C. Chen, and M. Pan, "Multi-objective uav path planning for emergency information collection and transmission," *IEEE Internet of Things Journal*, pp. 1–1, 2020.

[14] G. Fragkos, E. E. Tsiropoulou, and S. Papavassiliou, "Disaster management and information transmission decision-making in public safety systems," in *IEEE GLOBECOM*, 2019, pp. 1–6.

[15] P. Vamvakas, E. E. Tsiropoulou, and S. Papavassiliou, "Risk-aware resource management in public safety networks," *Sensors*, vol. 19, no. 18, p. 3853, 2019.

[16] E. Tsiropoulou, K. Koukas, and S. Papavassiliou, "A socio-physical and mobility-aware coalition formation mechanism in public safety networks," *EAI Endorsed Trans. Future Internet*, vol. 4, p. 154176, 2018.

[17] L. E. Miller, "Wireless technologies and the safecom sor for public safety communications," *NIST*, pp. 1–68, 2005.

[18] M. Filippone, F. Camastra, F. Masulli, and S. Rovetta, "A survey of kernel and spectral methods for clustering," *Pattern recognition*, vol. 41, no. 1, pp. 176–190, 2008.

[19] J. R. Marden and J. S. Shamma, "Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation," *Games and Economic Behav.*, vol. 75, no. 2, pp. 788–808, 2012.