Article:
Xu, Y.-H., Liu, X., Zhou, W. et al. (2021) Generative adversarial LSTM networks learning for resource allocation in UAV-served M2M communications. IEEE Wireless Communications Letters, 10 (7). pp. 1601-1605. ISSN: 2162-2337

https://doi.org/10.1109/lwc.2021.3075467

IEEE WIRELESS COMMUNICATIONS LETTERS

Generative Adversarial LSTM Networks Learning for Resource Allocation in UAV-served M2M Communications

Journal:	*IEEE Wireless Communications Letters*
Manuscript ID	WCL2021-0065
Manuscript Type:	Original Article
Date Submitted by the Author:	12-Jan-2021
Complete List of Authors:	Xu, Yi-Han; Nanjing Forestry University; University of New South Wales Liu, Xin; Nanjing Forestry University Zhou, Wen; Nanjing Forestry University Yu, Gang; The University of Sheffield
Key Words:	Unmanned aerial vehicles, M2M communications, Resource allocation, Long short-term memory, Generative adversarial networks

Generative Adversarial LSTM Networks Learning for Resource Allocation in UAV-served M2M Communications

Yi-Han Xu[1,2], Xin Liu[1], Wen Zhou[1] and Gang Yu[3]

***Abstract*—This letter investigates the resource allocation problem for multiple Unmanned Aerial Vehicles (UAVs)-served Machine-to-Machine (M2M) communications. Our goal is to maximize the sum-rate of UAVs-served M2M communications by jointly considering the transmission power, transmission mode, frequency spectrum, relay selection and the trajectory of UAVs. In order to model the uncertainty of stochastic environments, we formulate the resource allocation problem to be a Markov game, which is the generalization of Markov Decision Process (MDP) for the case of multiple agents. However, owning to the UAVs mobility poses the difficulty of perceiving the environment, we propose a Long Short-Term Memory (LSTM) with Generative Adversarial Networks (GANs) framework to better track and forecast the UAVs mobility and improving the network reward. Numerical results demonstrate that the proposed framework outperforms the conventional LSTM and Deep Q-Network (DQN) algorithms.**

***Index Terms*—Unmanned aerial vehicles, M2M communications, Resource allocation, Long short-term memory, Generative adversarial networks**

I. Introduction

MACHINE-to-Machine (M2M) communication emerges as a facilitator for Internet of Things (IoTs), has currently attracted extensive interests from both industry and academia. Owning to its inherent nature of massive and pervasiveness of Machine-Type Devices (MTDs) connectivity, most existing M2M communications rely on cellular infrastructure since the Base Stations (BSs) are capable of providing centralized, Quality of Service (QoS) guaranteed and secured services. However, in contrast to Human-to-Human (H2H) communications, parts of the MTDs in M2M communications are environmental-oriented which are typically deployed in remote areas. In such situation, the utilization of conventional cellular infrastructure is impracticable.

Fortunately, Unmanned Aerial Vehicles (UAVs) have been recently proposed to serve as aerial BSs for providing cost-effective and on-demand wireless coverage services in future wireless communications, attributed to its flexible deployment [1]. In addition, the channel provided by UAV-served communications are probably Line-of-Sight (LoS), which is also beneficial to the performance of wireless communications [2]. Consequently, in this letter, we intend to investigate multiple UAVs served as aerial BSs to facilitate the M2M communications in the area, where the conventional cellular infrastructure is unavailable. The goal of this work is to maximize the sum-rate of UAVs-served M2M communications from the perspective of resource allocation. More specifically, in order to model the uncertainty of stochastic environments, we formulate the resource allocation problem to be a Markov game by jointly considering the transmission power, transmission mode, frequency spectrum, relay selection and the trajectory of UAVs. In the game, each UAV acts as a learning agent and enables to effectively learn from the environment to make the allocation decision. However, owning to the fact that each UAV has different mobility pattern and the conventional Reinforcement Learning (RL) algorithms have shed little light on the possible influence of UAVs mobility on the perceived demand of resource [3]. Therefore, we propose a Long Short-Term Memory (LSTM) [4] with Generative Adversarial Networks (GANs) [5] framework, ca-called Generative Adversarial LSTM Networks to better track and forecast UAVs' mobility and thus improving the network reward. Numerical results demonstrate that the proposed framework is superior to the conventional LSTM and Deep Q-Network (DQN) algorithms.

II. System Model

In this letter, we consider a time-slotted multi-UAVs-served M2M communications scenario. We deploy M UAVs denoted as U_m $(m \in \{1, 2, \dots, M\})$ and N M2M pairs denoted as D_n $(n \in \{1, 2, \dots, N\})$. The ground MTDs in M2M pairs are randomly distributed in a square area of $R \times R$. Each M2M pair has a transmitter (*MTD_Tx*) and a receiver (*MTD_Rx*). In this model, W_{total} bandwidth is available and which is equally divided into F orthogonal sub-channels as $W_{each} =$

This work was supported by National Natural Science Foundation of China under Grant of 61601275.

Yi-Han Xu is with the College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037 China and School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney 2052, Australia (e-mail: xuyihan@njfu.edu.cn).

Xin Liu is with the College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037 China (e-mail: 572860097@qq.com).

Wen Zhou is with the College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037 China (e-mail: wenzhou@ustc.edu).

Gang Yu is with the Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield S10 2TN, UK (e-mail: 1781895340@qq.com).

$\{\frac{W_{total}}{F}, \frac{2 \cdot W_{total}}{F}, \dots, \frac{(F-1) \cdot W_{total}}{F}, W_{total}\}$. Without loss of generality, we assume that $F = M$ and each sub-channel is pre-assigned to the UAV. Meanwhile, M2M links are allowed to reuse the pre-occupied UAV sub-channel to enhance the spectrum efficiency. Thus, we define a binary indicator $\alpha_{m,n} \in \{0,1\}, \forall m \in (1,2,\dots,M)$ and $\forall n \in (1,2,\dots,N)$ to denote if the *n-th* M2M link currently reuses the *m-th* UAV spectrum. It is reasonable to note that each M2M link can reuse no more than one spectrum of UAV link at a time. Therefore, one constraint can be derived as Eq. (1).

$$\sum_{m=1}^{M} \alpha_{m,n} \leq 1, (\forall n \in N) \quad (1)$$

In addition, we assume that the Channel State Information (CSI) between *MTD_Tx* and *MTD_Rx* and that between MTDs and UAV are known by UAV [6]. Therefore, in every time slot, the M2M pairs can be categorized into two transmission modes: 1) M2M mode and 2) M-U-M mode. If the channel gain between *MTD_Tx* and *MTD_Rx* is better than that between MTDs and UAV, the M2M pair communicates in M2M mode. Otherwise, it works in M-U-M mode. Hence, we define another indicator $\beta_n \in \{0,1\}, \forall n \in (1,2,\dots,N)$ to denote which transmission mode is utilized by *n*-th M2M pair. We set $\beta_n = 1$ indicates the M-U-M mode is currently used. Otherwise, M2M pair communicates in M2M mode. Notably, each M2M pair can only operate in one transmission mode and each UAV can only serve as a relay for one M2M pair in a time slot. Therefore, we can get another constraint as Eq. (2).

$$\sum_{n=1}^{N} \beta_n \leq M, (\forall n \in N) \quad (2)$$

Moreover, it is worth noting that which UAV is served as a relay for a certain M2M pair in M-U-M mode is critical to the performance of network. Therefore, we define another parameter $\delta_{m,n} \in \{0,1\}, \forall m \in (1,2,\dots,M)$ and $\forall n \in (1,2,\dots,N)$ as an indicator of the *m*-th UAV is served as the relay for the *n*-th M2M pair.

$$\sum_{n=1}^{N} \delta_{m,n} \leq 1, (\forall n \in N) \quad (3)$$

In M2M mode, we can obtain the instantaneous transmission rate of *n*-th M2M pair in *t*-th time slot based on Shannon's theorem.

$$TR_n^{M2M}(t) = W_{each} \cdot log_2(1 + SINR_n^{M2M}(t)) \quad (4)$$

Where, $SINR_n^{M2M}(t)$ can be given by Eq. (5).

$$SINR_n^{M2M}(t) = \frac{p_n^{M2M} \cdot g_n^{Tx\_Rx}}{I_n^{M2M}(t) + n_0} \quad (5)$$

in which,

$$I_n^{M2M}(t) = \sum_{\substack{n,l,j \in N \\ M2M_j \neq M2M_n \\ M2M_j \neq M2M_l}} \frac{\delta_{m,j} \cdot (p_j^{MTD\_Tx-UAV} \cdot g_m^{MTD\_Tx-UAV_j} + p_m^{UAV-MTD\_Rx} \cdot g_m^{UAV-MTD\_Rx_j})}{2} + \sum_{\substack{n,i \in N \\ M2M_i \neq M2M_n}} \alpha_{m,i} \cdot p_i^{M2M} \cdot g_i^{Tx\_Rx} \quad (6)$$

The expression of $I_n^{M2M}(t)$ includes two items, the first item indicates the interference from M-U-M link and the second item is the interference from other M2M links which recently uses the same pre-assigned spectrum of *m*-th UAV.

In M-U-M mode, we consider a 3-Demensional (3D) Cartesian coordinate, in which $C_{UAV,m}(t) = [x_{UAV}, y_{UAV}, z_{UAV}]^T \in \mathbb{R}^{3\times1}$ and $C_{MTD,n}(t) = [x_{MTD}, y_{MTD}, z_{MTD}]^T \in \mathbb{R}^{3\times1}$ denote the coordinate of *m*-th UAV and *n*-th M2M pair in *t*-th time slot, respectively. We assume that all MTDs are located on the ground. For explicitly, we use $C_{MTD\_Tx,n}(t) = [x_{MTD\_Tx}, y_{MTD\_Tx}, z_{MTD\_Tx}]^T$ and $C_{MTD\_Rx,n}(t) = [x_{MTD\_Rx}, y_{MTD\_Rx}, z_{MTD\_Rx}]^T$ to denote the coordinates of the *n*-th *MTD_Tx* and *MTD_Rx* in *t*-th time slot, respectively. Therefore, we can obtain the distance between *m*-th UAV and *n*-th *MTD* in *t*-th time slot as Eq. (7).

$$D_{UAV,m-MTD,n}(t) = \sqrt{\| C_{UAV,m}(t) - C_{MTD,n}(t) \|^2} \quad (7)$$

Meanwhile, we suppose that all the UAVs will fly back to the base after complete the mission, so the Eq. (8) should be satisfied. Similarly, the trajectories of each UAV also constrained to its mobility speed and the minimum distance with other UAVs. Therefore, Eq. (9) and Eq. (10) should be satisfied.

$$C_{UAV,m}(1) = C_{UAV,m}(T) \quad (8)$$

$$\| C_{UAV,m}(t+1) - C_{UAV,m}(t) \| \leq V_{max} \quad (9)$$

$$\| C_{UAV,m \in M}(t) - C_{UAV,\forall j \neq m, \forall j \in M}(t) \| \geq D_{max} \quad (10)$$

Despite the flying UAVs are capable of providing a LoS link with ground MTDs, it also depends on the practical environment. Therefore, the randomness appearance of the LoS and Non-LoS should be considered. One commonly used expression can be given as Eq. (11) [7, 8].

$$Prb_{UAV,m-MTD,n}^{LoS}(\mathrm{t}) = \frac{1}{1 + a \cdot exp[-b \cdot (\theta_{UAV,m-MTD,n}(t) - a)]} \quad (11)$$

Where, *a* and *b* are the constants that depend on the carrier frequency and environment. $\theta_{UAV,m-MTD,n}(t)$ is the elevation angle, which is expressed as Eq. (12).

$$\theta_{UAV,m-MTD,n}(t) = \frac{180}{\pi \cdot \sin(z_{UAV}/D_{UAV,m-MTD,n}(t))} \quad (12)$$

Herein, we can obtain the average path loss between *m*-th UAV and *n*-th MTD in *t*-th time slot as expressed by Eq. (13).

$$PL_{m,n}(t) = Prb_{UAV,m-MTD,n}^{LoS}(t) \cdot PL_{UAV,m-MTD,n}^{LoS}(t) + (1 - Prb_{UAV,m-MTD,n}^{LoS}(t)) \cdot PL_{UAV,m-MTD,n}^{NLoS}(t) \quad (13)$$

where,

$$PL_{UAV,m-MTD,n}^{LoS}(t) = PL_{UAV,m-MTD,n}^{FS}(t) + \eta^{LoS} \quad (14)$$

$$PL_{UAV,m-MTD,n}^{NLoS}(t) = PL_{UAV,m-MTD,n}^{FS}(t) + \eta^{NLoS} \quad (15)$$

Here, $PL_{UAV,m-MTD,n}^{FS}(t)$ is the free space path loss between *m*-th UAV and *n*-th MTD in *t*-th time slot. η^{LoS} and η^{NLoS} are the mean additional losses for LoS and NLoS, respectively.

After above discussion, we can get the instantaneous transmission rate of *n*-th M2M link in *t*-th time slot in M-U-M mode. However, different from M2M mode, M-U-M link consists of two parts: *MTD_Tx*-UAV and UAV-*MTD_Rx* links.

$$TR_{m,n}^{MTD\_Tx-UAV}(t) = W_{each} \cdot log_2\left(1 + SINR_{m,n}^{MTD\_Tx-UAV}(t)\right) \quad (16)$$

$$TR_{m,n}^{UAV-MTD\_Rx}(t) = W_{each} \cdot log_2\left(1 + SINR_{m,n}^{UAV-MTD\_Rx}(t)\right) \quad (17)$$

where,

$$SINR_{m,n}^{MTD\_Tx-UAV}(t) = \frac{p_n^{MTD\_Tx-UAV} \cdot g_n^{MTD\_Tx-UAV} \cdot PL_{m,n}^{-1}(t)}{I_{m,n}^{MTD\_Tx-UAV}(t) + n_0} \quad (18)$$

$$SINR_{m,n}^{UAV-MTD\_Rx}(t) = \frac{p_m^{UAV-MTD\_Rx} \cdot g_n^{UAV-MTD\_Rx} \cdot PL_{m,n}^{-1}(t)}{I_{m,n}^{UAV-MTD\_Rx}(t) + n_0} \quad (19)$$

in which,

$$I_n^{MTD\_Tx-UAV}(t) = I_n^{UAV-MTD\_Rx}(t) = \sum_{\substack{n,i\in N \\ M2M_i \neq M2M_n}} \frac{\alpha_{m,i} \cdot p_i^{M2M} \cdot g_i^{Tx\_Rx}}{2} \quad (20)$$

It is worth noting that in M-U-M mode, the first half time slot is used for transmission of *MTD_Tx*-UAV link and the second half time slot is used by UVA for forwarding the data to the *MTD_Rx*. Hence, the effective transmission rate of the M-U-M mode is given by Eq. (21).

$$TR_{m,n}^{M-U-M}(t) = \frac{1}{2} \cdot \min\left(TR_{m,n}^{MTD\_Tx-UAV}(t), TR_{m,n}^{UAV-MTD\_Rx}(t)\right) \quad (21)$$

Therefore, we can get the sun-rate of the UAV-served M2M Communication systems as Eq. (22).

$$SR = \sum_t \sum_{m=1}^{M} \sum_{n=1}^{N} (1-\beta_n) \cdot TR_n^{M2M}(t) + \beta_n \cdot TR_{m,n}^{M-U-M}(t) \quad (22)$$

To this end, the resource allocation problem can be formulated as:

$$\max_{\substack{\{\alpha_{m,n},\ \beta_n,\ \delta_{m,n}\} \\ \{p_n^{M2M},\ p_n^{MTD\_Tx-UAV}, p_m^{UAV-MTD\_Rx}\}}} SR \quad (23)$$

s.t.

$$\sum_{m=1}^{M} \alpha_{m,n} \leq 1, (\forall n \in N)$$

$$\sum_{n=1}^{N} \beta_n \leq M, (\forall n \in N)$$

$$\sum_{n=1}^{N} \delta_{m,n} \leq 1, (\forall n \in N)$$

$$\alpha_{m,n}, \beta_n, \delta_{m,n} \in \{0,1\}\ \forall m \in (1,2,\dots,M), \forall n \in (1,2,\dots,N)$$

$$C_{UAV,m}(1) = C_{UAV,m}(T)$$

$$\| C_{UAV,m}(t+1) - C_{UAV,m}(t) \| \leq V_{max}$$

$$\| C_{UAV,m\in M}(t) - C_{UAV,\forall j\neq m,\forall j\in M}(t) \| \geq D_{max}$$

$$0 \leq p_n^{M2M} \leq p_{max}^{MTD}$$

$$0 \leq p_n^{MTD\_Tx-UAV} \leq p_{max}^{MTD}, 0 \leq p_m^{UAV-MTD\_Rx} \leq p_{max}^{UAV}$$

From Eq. (23), we found that the problem is a mixed integer nonlinear programming problem and owning to the reward of each UAV is only dependent on the current state and action that is satisfied by the properties of Markov chain [9]. Therefore, we formulate the problem as a Markov game, which is defined by a 4-elements tuple (UAV, S, $\{A^m\}_{m\in M}$, $\{r^m\}_{m\in M}$) that is described as below:

1) $UAV \triangleq \{1,\dots,m,\dots,M\}$ is a set of agents, which are the UAVs in the work.

2) $S \triangleq \{S^1,\dots,S^m,\dots,S^M\}$ is the global state spaces for the all UAVs and the S^m denotes the state space of the *m*-th UAV. In this work, the state of each UAV $S_m^k = (g_k^{Tx\_Rx}, g_k^{MTD\_Tx-UAV}, g_k^{UAV-MTD\_Rx}, I_k^{M2M}, I_k^{MTD\_Tx-UAV}, I_k^{UAV-MTD\_Rx})$, where, k is the number of time steps.

3) $\{A^m\}_{m\in M}$ are the set of action spaces for the all UAVs and A^m denotes the action space of the m-*th* UAV. It is obvious that the action should be the resource allocation strategy which including the transmission power ($p_n^{M2M}, p_n^{MTD\_Tx-UAV}$ and $p_m^{UAV-MTD\_Rx}$), spectrum multiplexing factor ($\alpha_{m,n}$), transmission mode (β_n) and relay selection ($\delta_{m,n}$).

4) $\{r^m\}_{m\in M}$ are the immediate rewards for the UAVs with $r^m \triangleq S \times A^1 \times \dots \times A^M \to \mathbb{R}$. For instance, the reward of *m*-th UAV will be obtained after all actions of the UAVs are performed. In this game, we assume that all UAVs are rational and selfish and thus all the UAVs start at an initial state $s_0 \in S$ and select their own actions $a = a^1, \dots, a^M$ non-cooperatively and simultaneously. At the meanwhile, they will receive the immediate rewards with the new observations. In this repeated procedure, all the UAVs try to find their optimal strategies to maximize own long-term rewards. We define the immediate reward function $r_m(t)$ of *m*-th UAV in *t*-th time slot as Eq. (24) and thus the long-term reward can be expressed by Eq. (25).

$$r_m(t) = \begin{cases} \sum_{n=1}^{N} (1-\beta_n) \cdot TR_n^{M2M}(t) + \beta_n \cdot TR_{m,n}^{M-U-M}(t) \\ 0 \end{cases} \quad (24)$$

$$R_m = \sum_{t=0}^{\infty} \gamma_m\, r_t^m\left(s_t^m, \pi_m^*(s_t^m)\right) \quad (25)$$

Eq. (24) indicates that if all constrains in Eq. (23) are satisfied, *m*-th UAV obtains $r_m(t)$, otherwise, the reward is 0. In Eq. (25), $\pi^*$is the equilibrium of the game. Thus, we can define the Nash Q-function of the *m*-th UAV as Eq. (26).

$$Q_m^*(s,a^m,a^{-m}) = r^m(s,a^m,a^{-m}) + V^m(s',\pi_1^*,\dots,\pi_m^*) \quad (26)$$

Where, a^{-m} is the actions of all UAVs except *m*-th UAV. $V^m(s',\pi_1^*,\dots,\pi_m^*)$ is the total discounted reward that all the UAVs follow the equilibrium strategies.

In this work, due to the fact that more historical information will beneficial to the UAVs to learn the environment quickly, we implement LSTM algorithm to derive the approximate value of $Q_m^*(s,a^m,a^{-m})$. In this process, all previous states will be firstly inputted into LSTM and then be stored and predicted to calculate and estimate the optimal strategy. After the approximation, the optimal policy can be presented by Eq. (27).

$$\pi^*(t) = argmax_{a\in A^m{}_{m\in M}}(Q_m^*(s,a^m,a^{-m})) \quad (27)$$

Notably, in this letter, we propose to use Generative Adversarial LSTM Networks to solve the maximization problem. There are two reasons for this adoption: 1) LSTM is an effective Recurrent Neural Network (RNN) architecture for prediction and it enables to capture the temporal variation regularity of resource allocation due to UAVs mobility. Thus, as a type of RNN, LSTM algorithm can efficiently handle sequence problems and it is reasonable to use LSTM to learn the relationship between historical and future states; 2) GAN is normally used while the existing dataset is small for training and it enables to create a virtual environment in which a more comprehensive dataset can be generated by a real dataset with synthetic data.

Specifically, in the proposed Generative Adversarial LSTM Networks framework, there are two inputs for the GANs: 1) a limited set of unlabelled real channel gain and interference from

other M2M and M-U-M links and 2) synthetic data simulated based on the model in [10]. Therefore, we can acquire a large observation dataset to emulate the multi-UAVs-served M2M communications. Explicitly, how to guarantee the similarity of the generated dataset is significant. In order to ensure the output of the generator indistinguishable from discriminator, we assume that the generator's neural network is trained as Eq. (28).

$$\omega_G^* = arg \min_{\omega_G} \max_{\omega_D} f(\omega_G, \omega_D) = arg \min_{\omega_G} f(\omega_G, \omega_D^*(\omega_G)) \quad (28)$$

s.t.

$$\mathbb{E}_{s\sim g_{sim}}\left[\left|\left|F(s;\ \omega_G)\right|\right|\right] < \epsilon_r$$

Where, ω_G is the weight of the neural network in generator that used to generate synthetic data. s is the input of generator and the real-like data $F(s;\ \omega_G)$ is the output of the generator. In our problem, s is the synthetic state variables (channel gain and interference from other M2M and M-U-M links) that we generated to train our LSTM algorithm. ω_D is the weight of the neural network in discriminator. ω_D^* is the optimal weight for the discriminator. g_{sim} is the distribution of the synthetic data. From Eq. (28), we can guarantee the similarity of s and $F(s;\ \omega_G)$.

III. Simulation Results and Analysis

A. Network Setting

We consider a multi-UAVs-served M2M communication scenario in an area of 1000m × 1000m × 40m. $M = 6$ UAVs are deployed flying over the area and $N = 40$ M2M pairs are uniformly deployed in the area. In this simulation, the bandwidth of each subchannel W_{each} is set to 75KHz. The constants a and b in the expression of presence probability of LoS are set to 9.61 and 0.16, respectively [11]. Moreover, the carrier frequency f is set to 2GHz, $\eta^{LoS} = 1$ and $\eta^{NLoS} = 20$. Tensorflow 1.13.1 with Python 3.6.5 is used to build a deep learning network and the OpenAI-Gym library is utilized to build reinforcement learning environment. The LSTM model has 2 hidden layers and each layer has 64 neurons. The output layer is a fully connected layer with only one neuron whose activation function is Rectified Linear Unit (ReLU). The learning rate is set to 0.01 initially and decreases exponentially to 0.001.

B. Results Analysis

In the proposed framework, GANs is responsible for generating a comprehensive and effective dataset to train the LSTM networks. Consequently, we firstly evaluate the similarity of the generated data and the real data in the simulation. Fig. 1 compares the GANs-generated dataset and the real dataset. In this evaluation, we take the channel gain between *MTD_Tx* and *MTD_Rx* in M2M mode as an example. It is clear that the GANs-generated dataset has the similar distribution with the real dataset.

Fig. 2 presents the trajectory of each UAV in 3D coordinate from the proposed Generative Adversarial LSTM Networks algorithm to get the maximum sum-rate of M2M communications in the formulated model. From the figure, we can see that the changing of UAVs' location in each time slot makes a significant impact on the sum-rate of M2M communications since the proposed model also considers the location of each UAV to allocate resource for maximizing the sum-rate of M2M communications.

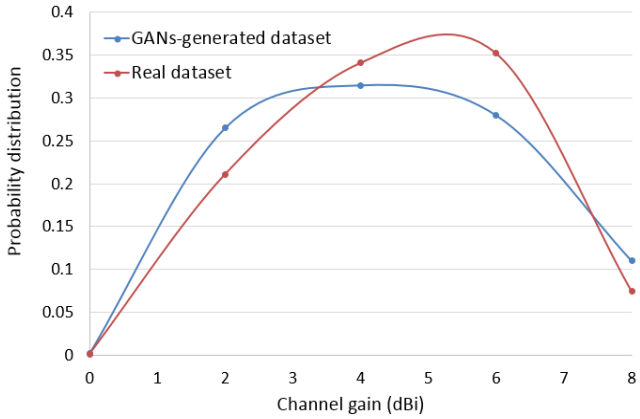

Fig. 1. Comparison of GANs-generated channel gain and the real data

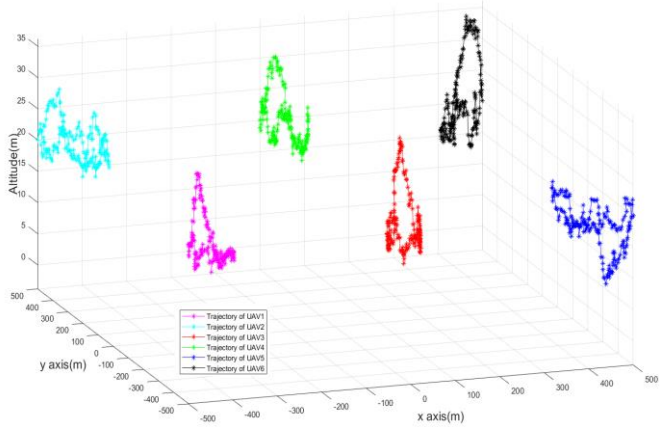

Fig. 2. The trajectory of each UAV under the proposed framework

Fig. 3 gives the comparison of the average sum-rate of M2M communications among three different algorithms. It is obvious that the proposed Generative Adversarial LSTM Networks algorithm has the best performance. However, it is worth noting that the conventional LSTM algorithm has the barely faster convergence speed as compared to Generative Adversarial LSTM Networks algorithm. This is due to the reason that the generator and discriminator in GANs need time to generate data.

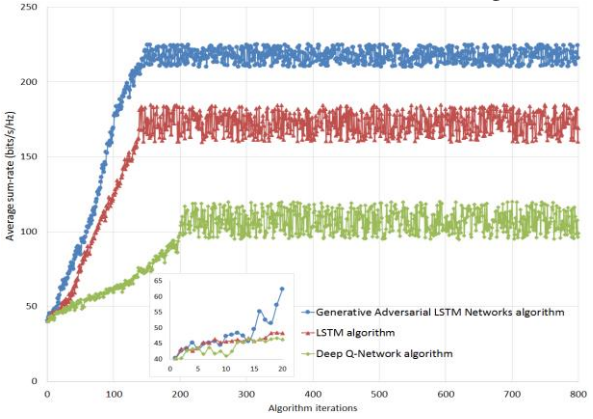

Fig. 3. Comparison for average sum-rate with different algorithms

IV. Conclusion

In this letter, we introduce a resource allocation strategy for multi-UAVs-served M2M communications. We jointly consider the transmission power, transmission mode, frequency

spectrum, relay selection and the trajectory of UAVs to formulate the problem as a Markov game. After that, a Generative Adversarial LSTM Networks framework is proposed to solve the problem of maximizing the sum-rate of M2M communications. In the proposed framework, LSTM is used to learn the relationship between historical and future states and the GANs is responsible for generating more comprehensive dataset for training the LSTM. Finally, simulation results validate the effectiveness of the proposed Generative Adversarial LSTM Networks framework.

REFERENCES

[1] M. Mozaffari, W. Saad, M. Bennis, Y. H. Nam, & M. Debbah, "A tutorial on uavs for wireless networks: applications, challenges, and open problems," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2334-2360, 3rd Quarter, 2019.

[2] L. Xie, J. Xu, and Y. Zeng, Y, "Common throughput maximization for uav-enabled interference channel with wireless powered communications," *IEEE Transactions on Communications,* vol. 68, no. 5, pp. 3197-3212, Feb, 2020.

[3] C. Liu, W. Yuan, Z. Wei, X. Liu, and D. W. K. Ng, (2020). Location-aware predictive beamforming for uav communications: a deep learning approach. *IEEE Wireless Communications Letters,* to be published. DOI: 10.1109/LWC.2020.3045150.

[4] Y. Lin, M. Wang, X. Zhou, G. Ding, and S. Mao, "Dynamic spectrum interaction of uav flight formation communication with priority: a deep reinforcement learning approach," *IEEE Transactions on Cognitive Communications and Networking,* vol. 6, no. 3, pp. 892-903, Feb, 2020.

[5] T. Spadini, G. S. I. Aldeia, G. Barreto, K. Alves, and K. Nose-Filho, "On the application of SEGAN for the attenuation of the ego-noise in the speech sound source localization problem," presented at the *4th Workshop on Communication Networks and Power Systems (WCNPS 2019),* Brasilia, Brazil, Oct. 3-4, 2019.

[6] F. Wang, C. Xu, L. Song, and Z. Han, "Energy-efficient resource allocation for device-to-device underlay communication," *IEEE Trans. Wireless Commun.,* vol. 14, no. 4, pp. 2082-2092, Dec, 2014.

[7] Y. Zeng, R. Zhang, and T. J. Lim, "Throughput maximization for uav-enabled mobile relaying systems," *IEEE Trans. Commun.,* vol. 64, no. 12, pp.4983-4996, Dec, 2016.

[8] S. Chandrasekharan, K. Gomez, A. Al-Hourani, S. Kandeepan, *et al.,* "Designing and implementing future aerial communication networks," *IEEE Communications Magazine,* vol. 54, no. 5, pp.26-34, May, 2016.

[9] M. L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, New York, NY, USA:Wiley, 2014

[10] Y. H. Xu, Y. B. Tian, P. K. Searyoh, G. Yu, Y. T. Yong, "Deep reinforcement learning-based resource allocation strategy for energy harvesting-powered cognitive machine-to-machine networks," *Computer Communications,* vol.160, pp.706-717, July, 2020.

[11] M. Alzenad, A. El-Keyi, F. Lagum, and H. Yanikomeroglu, "3-d placement of an unmanned aerial vehicle base station (uav-bs) for energy-efficient maximal coverage," *IEEE Wireless Communications Letters*, vol. 6, no. 4, pp. 434-437, May, 2017.

Generative Adversarial LSTM Networks Learning for Resource Allocation in UAV-served M2M Communications

Yi-Han Xu[1,2], Xin Liu[1], Wen Zhou[1] and Gang Yu[3]

***Abstract*—This letter investigates the resource allocation problem for multiple Unmanned Aerial Vehicles (UAVs)-served Machine-to-Machine (M2M) communications. Our goal is to maximize the sum-rate of UAVs-served M2M communications by jointly considering the transmission power, transmission mode, frequency spectrum, relay selection and the trajectory of UAVs. In order to model the uncertainty of stochastic environments, we formulate the resource allocation problem to be a Markov game, which is the generalization of Markov Decision Process (MDP) for the case of multiple agents. However, owning to the UAVs mobility poses the difficulty of perceiving the environment, we propose a Long Short-Term Memory (LSTM) with Generative Adversarial Networks (GANs) framework to better track and forecast the UAVs mobility and improving the network reward. Numerical results demonstrate that the proposed framework outperforms the conventional LSTM and Deep Q-Network (DQN) algorithms.**

***Index Terms*—Unmanned aerial vehicles, M2M communications, Resource allocation, Long short-term memory, Generative adversarial networks**

I. Introduction

MACHINE-to-Machine (M2M) communication emerges as a facilitator for Internet of Things (IoTs), has currently attracted extensive interests from both industry and academia. Owning to its inherent nature of massive and pervasiveness of Machine-Type Devices (MTDs) connectivity, most existing M2M communications rely on cellular infrastructure since the Base Stations (BSs) are capable of providing centralized, Quality of Service (QoS) guaranteed and secured services. However, in contrast to Human-to-Human (H2H) communications, parts of the MTDs in M2M communications are environmental-oriented which are typically deployed in remote areas. In such situation, the utilization of conventional cellular infrastructure is impracticable.

Fortunately, Unmanned Aerial Vehicles (UAVs) have been recently proposed to serve as aerial BSs for providing cost-effective and on-demand wireless coverage services in future wireless communications, attributed to its flexible deployment [1]. In addition, the channel provided by UAV-served communications are probably Line-of-Sight (LoS), which is also beneficial to the performance of wireless communications [2]. Consequently, in this letter, we intend to investigate multiple UAVs served as aerial BSs to facilitate the M2M communications in the area, where the conventional cellular infrastructure is unavailable. The goal of this work is to maximize the sum-rate of UAVs-served M2M communications from the perspective of resource allocation. More specifically, in order to model the uncertainty of stochastic environments, we formulate the resource allocation problem to be a Markov game by jointly considering the transmission power, transmission mode, frequency spectrum, relay selection and the trajectory of UAVs. In the game, each UAV acts as a learning agent and enables to effectively learn from the environment to make the allocation decision. However, owning to the fact that each UAV has different mobility pattern and the conventional Reinforcement Learning (RL) algorithms have shed little light on the possible influence of UAVs mobility on the perceived demand of resource [3]. Therefore, we propose a Long Short-Term Memory (LSTM) [4] with Generative Adversarial Networks (GANs) [5] framework, ca-called Generative Adversarial LSTM Networks to better track and forecast UAVs' mobility and thus improving the network reward. Numerical results demonstrate that the proposed framework is superior to the conventional LSTM and Deep Q-Network (DQN) algorithms.

This work was supported by National Natural Science Foundation of China under Grant of 61601275.

Yi-Han Xu is with the College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037 China and School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney 2052, Australia (e-mail: xuyihan@njfu.edu.cn).

Xin Liu is with the College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037 China (e-mail: 572860097@qq.com).

Wen Zhou is with the College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037 China (e-mail: wenzhou@ustc.edu).

Gang Yu is with the Department of Electronic and Electrical Engineering, The University of Sheffield, Sheffield S10 2TN, UK (e-mail: 1781895340@qq.com).

II. System Model

In this letter, we consider a time-slotted multi-UAVs-served M2M communications scenario. We deploy M UAVs denoted as U_m $(m \in \{1, 2, \dots, M\})$ and N M2M pairs denoted as D_n $(n \in \{1, 2, \dots, N\})$. The ground MTDs in M2M pairs are randomly distributed in a square area of $R \times R$. Each M2M pair has a transmitter (*MTD_Tx*) and a receiver (*MTD_Rx*). In this model, W_{total} bandwidth is available and which is equally divided into F orthogonal sub-channels as $W_{each} = \{\frac{W_{total}}{F}, \frac{2 \cdot W_{total}}{F}, \dots, \frac{(F-1) \cdot W_{total}}{F}, W_{total}\}$. Without loss of generality, we assume that $F = M$ and each sub-channel is pre-assigned to the UAV. Meanwhile, M2M links are allowed to reuse the pre-occupied UAV sub-channel to enhance the spectrum efficiency. Thus, we define a binary indicator $\alpha_{m,n} \in \{0, 1\}, \forall m \in (1, 2, \dots, M)$ and $\forall n \in (1, 2, \dots, N)$ to denote if the *n-th* M2M link currently reuses the *m-th* UAV spectrum. It is reasonable to note that each M2M link can reuse no more than one spectrum of UAV link at a time. Therefore, one constraint can be derived as Eq. (1).

$$\sum_{m=1}^{M} \alpha_{m,n} \leq 1, (\forall n \in N) \qquad (1)$$

In addition, we assume that the Channel State Information (CSI) between *MTD_Tx* and *MTD_Rx* and that between MTDs and UAV are known by UAV [6]. Therefore, in every time slot, the M2M pairs can be categorized into two transmission modes: 1) M2M mode and 2) M-U-M mode. If the channel gain between *MTD_Tx* and *MTD_Rx* is better than that between MTDs and UAV, the M2M pair communicates in M2M mode. Otherwise, it works in M-U-M mode. Hence, we define another indicator $\beta_n \in \{0, 1\}, \forall n \in (1, 2, \dots, N)$ to denote which transmission mode is utilized by n-th M2M pair. We set $\beta_n = 1$ indicates the M-U-M mode is currently used. Otherwise, M2M pair communicates in M2M mode. Notably, each M2M pair can only operate in one transmission mode and each UAV can only serve as a relay for one M2M pair in a time slot. Therefore, we can get another constraint as Eq. (2).

$$\sum_{n=1}^{N} \beta_n \leq M, (\forall n \in N) \qquad (2)$$

Moreover, it is worth noting that which UAV is served as a relay for a certain M2M pair in M-U-M mode is critical to the performance of network. Therefore, we define another parameter $\delta_{m,n} \in \{0, 1\}, \forall m \in (1, 2, \dots, M)$ and $\forall n \in (1, 2, \dots, N)$ as an indicator of the m-th UAV is served as the relay for the n-th M2M pair.

$$\sum_{n=1}^{N} \delta_{m,n} \leq 1, (\forall n \in N) \qquad (3)$$

In M2M mode, we can obtain the instantaneous transmission rate of n-th M2M pair in t-th time slot based on Shannon's theorem.

$$TR_n^{M2M}(t) = W_{each} \cdot log_2(1 + SINR_n^{M2M}(t)) \qquad (4)$$

Where, $SINR_n^{M2M}(t)$ can be given by Eq. (5).

$$SINR_n^{M2M}(t) = \frac{p_n^{M2M} \cdot g_n^{Tx\_Rx}}{I_n^{M2M}(t) + n_0} \qquad (5)$$

in which,

$$I_n^{M2M}(t) = \sum_{\substack{n,i,j \in N \\ M2M_j \neq M2M_n \\ M2M_j \neq M2M_i}} \frac{\delta_{m,j} \cdot (p_j^{MTD\_Tx-UAV} \cdot g_m^{MTD\_Tx-UAV_j} + p_m^{UAV-MTD\_Rx} \cdot g_m^{UAV-MTD\_Rx_j})}{2} + \sum_{\substack{n,i \in N \\ M2M_i \neq M2M_n}} \alpha_{m,i} \cdot p_i^{M2M} \cdot g_i^{Tx\_Rx} \qquad (6)$$

The expression of $I_n^{M2M}(t)$ includes two items, the first item indicates the interference from M-U-M link and the second item is the interference from other M2M links which recently uses the same pre-assigned spectrum of m-th UAV.

In M-U-M mode, we consider a 3-Demensional (3D) Cartesian coordinate, in which $C_{UAV,m}(t) = [x_{UAV}, y_{UAV}, z_{UAV}]^T \in \mathbb{R}^{3\times1}$ and $C_{MTD,n}(t) = [x_{MTD}, y_{MTD}, z_{MTD}]^T \in \mathbb{R}^{3\times1}$ denote the coordinate of m-th UAV and n-th M2M pair in t-th time slot, respectively. We assume that all MTDs are located on the ground. For explicitly, we use $C_{MTD\_Tx,n}(t) = [x_{MTD\_Tx}, y_{MTD\_Tx}, z_{MTD\_Tx}]^T$ and $C_{MTD\_Rx,n}(t) = [x_{MTD\_Rx}, y_{MTD\_Rx}, z_{MTD\_Rx}]^T$ to denote the coordinates of the n-th *MTD_Tx* and *MTD_Rx* in t-th time slot, respectively. Therefore, we can obtain the distance between m-th UAV and n-th *MTD* in t-th time slot as Eq. (7).

$$D_{UAV,m-MTD,n}(t) = \sqrt{\| C_{UAV,m}(t) - C_{MTD,n}(t) \|^2} \qquad (7)$$

Meanwhile, we suppose that all the UAVs will fly back to the base after complete the mission, so the Eq. (8) should be satisfied. Similarly, the trajectories of each UAV also constrained to its mobility speed and the minimum distance with other UAVs. Therefore, Eq. (9) and Eq. (10) should be satisfied.

$$C_{UAV,m}(1) = C_{UAV,m}(T) \qquad (8)$$

$$\| C_{UAV,m}(t+1) - C_{UAV,m}(t) \| \leq V_{max} \qquad (9)$$

$$\| C_{UAV,m\in M}(t) - C_{UAV,\forall j \neq m, \forall j \in M}(t) \| \geq D_{max} \qquad (10)$$

Despite the flying UAVs are capable of providing a LoS link with ground MTDs, it also depends on the practical environment. Therefore, the randomness appearance of the LoS and Non-LoS should be considered. One commonly used expression can be

given as Eq. (11) [7, 8].

$$Prb_{UAV,m-MTD,n}^{LoS}(\mathrm{t}) = \frac{1}{1 + a \cdot exp[-b \cdot (\theta_{UAV,m-MTD,n}(t) - a)]} \quad (11)$$

Where, a and b are the constants that depend on the carrier frequency and environment. $\theta_{UAV,m-MTD,n}(t)$ is the elevation angle, which is expressed as Eq. (12).

$$\theta_{UAV,m-MTD,n}(t) = \frac{180}{\pi \cdot \sin(z_{UAV}/D_{UAV,m-MTD,n}(t))} \quad (12)$$

Herein, we can obtain the average path loss between m-th UAV and n-th MTD in t-th time slot as expressed by Eq. (13).

$$PL_{m,n}(t) = Prb_{UAV,m-MTD,n}^{LoS}(t) \cdot PL_{UAV,m-MTD,n}^{LoS}(t) + (1 - Prb_{UAV,m-MTD,n}^{LoS}(t)) \cdot PL_{UAV,m-MTD,n}^{NLoS}(t) \quad (13)$$

where,

$$PL_{UAV,m-MTD,n}^{LoS}(t) = PL_{UAV,m-MTD,n}^{FS}(t) + \eta^{LoS} \quad (14)$$

$$PL_{UAV,m-MTD,n}^{NLoS}(t) = PL_{UAV,m-MTD,n}^{FS}(t) + \eta^{NLoS} \quad (15)$$

Here, $PL_{UAV,m-MTD,n}^{FS}(t)$ is the free space path loss between m-th UAV and n-th MTD in t-th time slot. η^{LoS} and η^{NLoS} are the mean additional losses for LoS and NLoS, respectively.

After above discussion, we can get the instantaneous transmission rate of n-th M2M link in t-th time slot in M-U-M mode. However, different from M2M mode, M-U-M link consists of two parts: *MTD_Tx*-UAV and UAV-*MTD_Rx* links.

$$TR_{m,n}^{MTD\_Tx-UAV}(t) = W_{each} \cdot log_2\left(1 + SINR_{m,n}^{MTD\_Tx-UAV}(t)\right) \quad (16)$$

$$TR_{m,n}^{UAV-MTD\_Rx}(t) = W_{each} \cdot log_2\left(1 + SINR_{m,n}^{UAV-MTD\_Rx}(t)\right) \quad (17)$$

where,

$$SINR_{m,n}^{MTD\_Tx-UAV}(t) = \frac{p_n^{MTD\_Tx-UAV} \cdot g_n^{MTD\_Tx-UAV} \cdot PL_{m,n}^{-1}(t)}{I_{m,n}^{MTD\_Tx-UAV}(t) + n_0} \quad (18)$$

$$SINR_{m,n}^{UAV-MTD\_Rx}(t) = \frac{p_m^{UAV-MTD\_Rx} \cdot g_n^{UAV-MTD\_Rx} \cdot PL_{m,n}^{-1}(t)}{I_{m,n}^{UAV-MTD\_Rx}(t) + n_0} \quad (19)$$

in which,

$$I_n^{MTD\_Tx-UAV}(t) = I_n^{UAV-MTD\_Rx}(t) = \sum_{\substack{n,i \in N \\ M2M_i \neq M2M_n}} \frac{\alpha_{m,i} \cdot p_i^{M2M} \cdot g_i^{Tx\_Rx}}{2} \quad (20)$$

It is worth noting that in M-U-M mode, the first half time slot is used for transmission of *MTD_Tx*-UAV link and the second half time slot is used by UVA for forwarding the data to the *MTD_Rx*. Hence, the effective transmission rate of the M-U-M mode is given by Eq. (21).

$$TR_{m,n}^{M-U-M}(t) = \frac{1}{2} \cdot \min\left(TR_{m,n}^{MTD\_Tx-UAV}(t), TR_{m,n}^{UAV-MTD\_Rx}(t)\right) \quad (21)$$

Therefore, we can get the sun-rate of the UAV-served M2M Communication systems as Eq. (22).

$$SR = \sum_t \sum_{m=1}^{M} \sum_{n=1}^{N} (1 - \beta_n) \cdot TR_n^{M2M}(t) + \beta_n \cdot TR_{m,n}^{M-U-M}(t) \quad (22)$$

To this end, the resource allocation problem can be formulated as:

$$\max_{\substack{\{\alpha_{m,n},\ \beta_n,\ \delta_{m,n}\} \\ \{p_n^{M2M},\ p_n^{MTD\_Tx-UAV}, p_m^{UAV-MTD\_Rx}\}}} SR \quad (23)$$

s.t.

$$\sum_{m=1}^{M} \alpha_{m,n} \leq 1, (\forall n \in N)$$

$$\sum_{n=1}^{N} \beta_n \leq M, (\forall n \in N)$$

$$\sum_{n=1}^{N} \delta_{m,n} \leq 1, (\forall n \in N)$$

$$\alpha_{m,n}, \beta_n, \delta_{m,n} \in \{0, 1\}\ \forall m \in (1, 2, \dots, M), \forall n \in (1, 2, \dots, N)$$

$$C_{UAV,m}(1) = C_{UAV,m}(T)$$

$$\| C_{UAV,m}(t+1) - C_{UAV,m}(t) \| \leq V_{max}$$

$$\| C_{UAV,m \in M}(t) - C_{UAV,\forall j \neq m, \forall j \in M}(t) \| \geq D_{max}$$

$$0 \leq p_n^{M2M} \leq p_{max}^{MTD}$$

$$0 \leq p_n^{MTD\_Tx-UAV} \leq p_{max}^{MTD}, 0 \leq p_m^{UAV-MTD\_Rx} \leq p_{max}^{UAV}$$

From Eq. (23), we found that the problem is a mixed integer nonlinear programming problem and owning to the reward of each UAV is only dependent on the current state and action that is satisfied by the properties of Markov chain [9]. Therefore, we formulate the problem as a Markov game, which is defined by a 4-elements tuple (*UAV, S,* $\{A^m\}_{m\in M}$, $\{r^m\}_{m\in M}$) that is described as below:

1) $UAV \triangleq \{1, \dots, m, \dots, M\}$ is a set of agents, which are the UAVs in the work.

2) $S \triangleq \{S^1, \dots, S^m, \dots, S^M\}$ is the global state spaces for the all UAVs and the S^m denotes the state space of the *m*-th UAV. In this work, the state of each UAV $S_m^k = (g_k^{Tx\_Rx}, g_k^{MTD\_Tx-UAV}, g_k^{UAV-MTD\_Rx}, I_k^{M2M}, I_k^{MTD\_Tx-UAV}, I_k^{UAV-MTD\_Rx})$, where, k is the number of time steps.

3) $\{A^m\}_{m\in M}$ are the set of action spaces for the all UAVs and A^m denotes the action space of the m-*th* UAV. It is obvious that the action should be the resource allocation strategy which including the transmission power ($p_n^{M2M}, p_n^{MTD\_Tx-UAV}$ and $p_m^{UAV-MTD\_Rx}$), spectrum multiplexing factor ($\alpha_{m,n}$), transmission mode (β_n) and relay selection ($\delta_{m,n}$).

4) $\{r^m\}_{m\in M}$ are the immediate rewards for the UAVs with $r^m \triangleq S \times A^1 \times \dots \times A^M \to \mathbb{R}$. For instance, the reward of *m*-th UAV will be obtained after all actions of the UAVs are performed. In this game, we assume that all UAVs are rational and selfish and thus all the UAVs start at an initial state $s_0 \in$ S and select their own actions $a = a^1, \dots, a^M$ non-cooperatively and simultaneously. At the meanwhile, they will receive the immediate rewards with the new observations. In this repeated procedure, all the UAVs try to find their optimal strategies to maximize own long-term rewards. We define the immediate reward function $r_m(t)$ of *m*-th UAV in *t*-th time slot as Eq. (24) and thus the long-term reward can be expressed by Eq. (25).

$$r_m(t) = \begin{cases} \sum_{n=1}^{N} (1-\beta_n) \cdot TR_n^{M2M}(t) + \beta_n \cdot TR_{m,n}^{M-U-M}(t) \\ 0 \end{cases} \quad (24)$$

$$R_m = \sum_{t=0}^{\infty} \gamma_m \, r_t^m\left(s_t^m, \pi_m^*(s_t^m)\right) \quad (25)$$

Eq. (24) indicates that if all constrains in Eq. (23) are satisfied, *m*-th UAV obtains $r_m(t)$, otherwise, the reward is 0. In Eq. (25), $\pi^*$is the equilibrium of the game. Thus, we can define the Nash Q-function of the *m*-th UAV as Eq. (26).

$$Q_m^*(s, a^m, a^{-m}) = r^m(s, a^m, a^{-m}) + V^m(s', \pi_1^*, \dots, \pi_m^*) \quad (26)$$

Where, a^{-m} is the actions of all UAVs except *m*-th UAV. $V^m(s', \pi_1^*, \dots, \pi_m^*)$ is the total discounted reward that all the UAVs follow the equilibrium strategies.

In this work, due to the fact that more historical information will beneficial to the UAVs to learn the environment quickly, we implement LSTM algorithm to derive the approximate value of $Q_m^*(s, a^m, a^{-m})$. In this process, all previous states will be firstly inputted into LSTM and then be stored and predicted to calculate and estimate the optimal strategy. After the approximation, the optimal policy can be presented by Eq. (27).

$$\pi^*(t) = argmax_{a\in A^m{}_{m\in M}}(Q_m^*(s, a^m, a^{-m})) \quad (27)$$

Notably, in this letter, we propose to use Generative Adversarial LSTM Networks to solve the maximization problem. There are two reasons for this adoption: 1) LSTM is an effective Recurrent Neural Network (RNN) architecture for prediction and it enables to capture the temporal variation regularity of resource allocation due to UAVs mobility. Thus, as a type of RNN, LSTM algorithm can efficiently handle sequence problems and it is reasonable to use LSTM to learn the relationship between historical and future states; 2) GAN is normally used while the existing dataset is small for training and it enables to create a virtual environment in which a more comprehensive dataset can be generated by a real dataset with synthetic data.

Specifically, in the proposed Generative Adversarial LSTM Networks framework, there are two inputs for the GANs: 1) a limited set of unlabelled real channel gain and interference from other M2M and M-U-M links and 2) synthetic data simulated based on the model in [10]. Therefore, we can acquire a large observation dataset to emulate the multi-UAVs-served M2M communications. Explicitly, how to guarantee the similarity of the generated dataset is significant. In order to ensure the output of the generator indistinguishable from discriminator, we assume that the generator's neural network is trained as Eq. (28).

$$\omega_G^* = arg \min_{\omega_G} \max_{\omega_D} f(\omega_G, \omega_D) = arg \min_{\omega_G} f(\omega_G, \omega_D^*(\omega_G)) \quad (28)$$

s.t.

$$\mathbb{E}_{s\sim g_{sim}}\left[\left|\left|F(s;\, \omega_G)\right|\right|\right] < \epsilon_r$$

Where, ω_G is the weight of the neural network in generator that used to generate synthetic data. s is the input of generator and the real-like data $F(s;\, \omega_G)$ is the output of the generator. In our problem, s is the synthetic state variables (channel gain and interference from other M2M and M-U-M links) that we generated to train our LSTM algorithm. ω_D is the weight of the neural network in discriminator. ω_D^* is the optimal weight for the discriminator. g_{sim} is the distribution of the synthetic data. From Eq. (28), we can guarantee the similarity of s and $F(s;\, \omega_G)$.

III. Simulation Results and Analysis

A. Network Setting

We consider a multi-UAVs-served M2M communication scenario in an area of 1000m × 1000m × 40m. $M = 6$ UAVs are

deployed flying over the area and $N = 40$ M2M pairs are uniformly deployed in the area. In this simulation, the bandwidth of each subchannel W_{each} is set to 75KHz. The constants a and b in the expression of presence probability of LoS are set to 9.61 and 0.16, respectively [11]. Moreover, the carrier frequency f is set to 2GHz, $\eta^{LoS} = 1$ and $\eta^{NLoS} = 20$. Tensorflow 1.13.1 with Python 3.6.5 is used to build a deep learning network and the OpenAI-Gym library is utilized to build reinforcement learning environment. The LSTM model has 2 hidden layers and each layer has 64 neurons. The output layer is a fully connected layer with only one neuron whose activation function is Rectified Linear Unit (ReLU). The learning rate is set to 0.01 initially and decreases exponentially to 0.001.

B. Results Analysis

In the proposed framework, GANs is responsible for generating a comprehensive and effective dataset to train the LSTM networks. Consequently, we firstly evaluate the similarity of the generated data and the real data in the simulation. Fig. 1 compares the GANs-generated dataset and the real dataset. In this evaluation, we take the channel gain between *MTD_Tx* and *MTD_Rx* in M2M mode as an example. It is clear that the GANs-generated dataset has the similar distribution with the real dataset.

Fig. 2 presents the trajectory of each UAV in 3D coordinate from the proposed Generative Adversarial LSTM Networks algorithm to get the maximum sum-rate of M2M communications in the formulated model. From the figure, we can see that the changing of UAVs' location in each time slot makes a significant impact on the sum-rate of M2M communications since the proposed model also considers the location of each UAV to allocate resource for maximizing the sum-rate of M2M communications.

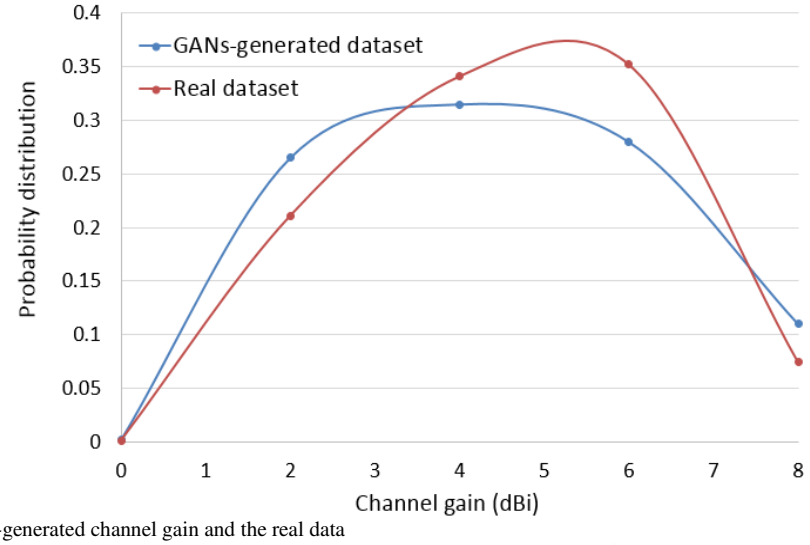

Fig. 1. Comparison of GANs-generated channel gain and the real data

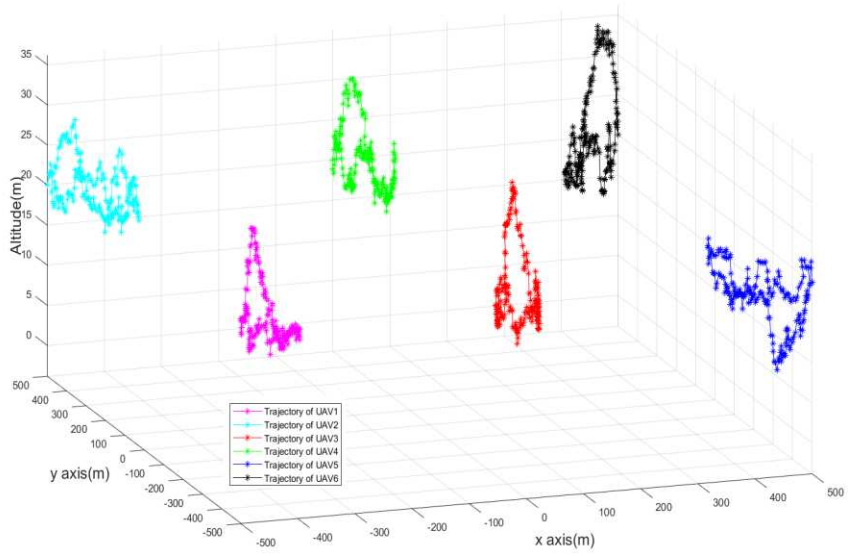

Fig. 2. The trajectory of each UAV under the proposed framework

Fig. 3 gives the comparison of the average sum-rate of M2M communications among three different algorithms. It is obvious that the proposed Generative Adversarial LSTM Networks algorithm has the best performance. However, it is worth noting that the conventional LSTM algorithm has the barely faster convergence speed as compared to Generative Adversarial LSTM Networks algorithm. This is due to the reason that the generator and discriminator in GANs need time to generate data.

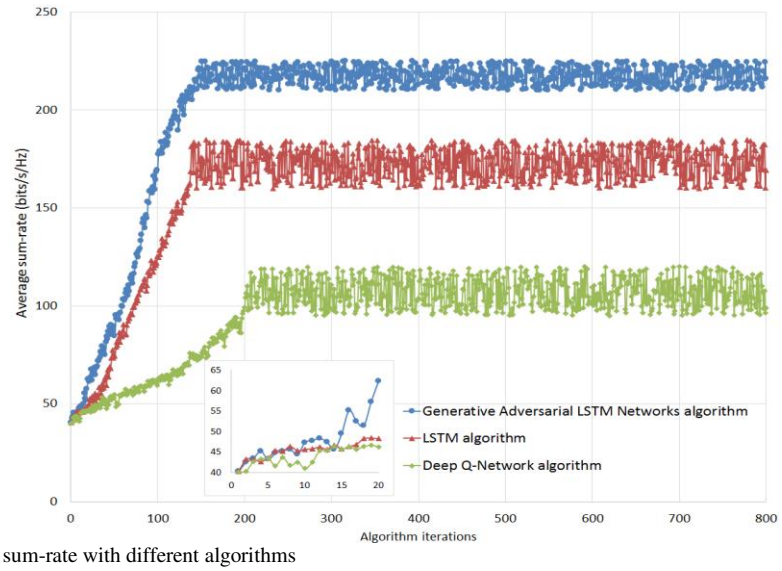

Fig. 3. Comparison for average sum-rate with different algorithms

IV. Conclusion

In this letter, we introduce a resource allocation strategy for multi-UAVs-served M2M communications. We jointly consider the transmission power, transmission mode, frequency spectrum, relay selection and the trajectory of UAVs to formulate the problem as a Markov game. After that, a Generative Adversarial LSTM Networks framework is proposed to solve the problem of maximizing the sum-rate of M2M communications. In the proposed framework, LSTM is used to learn the relationship between historical and future states and the GANs is responsible for generating more comprehensive dataset for training the LSTM. Finally, simulation results validate the effectiveness of the proposed Generative Adversarial LSTM Networks framework.

References

[1] M. Mozaffari, W. Saad, M. Bennis, Y. H. Nam, & M. Debbah, "A tutorial on uavs for wireless networks: applications, challenges, and open problems," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2334-2360, 3rd Quarter, 2019.

[2] L. Xie, J. Xu, and Y. Zeng, Y, "Common throughput maximization for uav-enabled interference channel with wireless powered communications," *IEEE Transactions on Communications,* vol. 68, no. 5, pp. 3197-3212, Feb, 2020.

[3] C. Liu, W. Yuan, Z. Wei, X. Liu, and D. W. K. Ng, (2020). Location-aware predictive beamforming for uav communications: a deep learning approach. *IEEE Wireless Communications Letters,* to be published. DOI: 10.1109/LWC.2020.3045150.

[4] Y. Lin, M. Wang, X. Zhou, G. Ding, and S. Mao, "Dynamic spectrum interaction of uav flight formation communication with priority: a deep reinforcement learning approach," *IEEE Transactions on Cognitive Communications and Networking,* vol. 6, no. 3, pp. 892-903, Feb, 2020.

[5] T. Spadini, G. S. I. Aldeia, G. Barreto, K. Alves, and K. Nose-Filho, "On the application of SEGAN for the attenuation of the ego-noise in the speech sound source localization problem," presented at the *4th Workshop on Communication Networks and Power Systems (WCNPS 2019),* Brasilia, Brazil, Oct. 3-4, 2019.

[6] F. Wang, C. Xu, L. Song, and Z. Han, "Energy-efficient resource allocation for device-to-device underlay communication," *IEEE Trans. Wireless Commun.,* vol. 14, no. 4, pp. 2082-2092, Dec, 2014.

[7] Y. Zeng, R. Zhang, and T. J. Lim, "Throughput maximization for uav-enabled mobile relaying systems," *IEEE Trans. Commun.,* vol. 64, no. 12, pp.4983-4996, Dec, 2016.

[8] S. Chandrasekharan, K. Gomez, A. Al-Hourani, S. Kandeepan, *et al.,* "Designing and implementing future aerial communication networks," *IEEE Communications Magazine,* vol. 54, no. 5, pp.26-34, May, 2016.

[9] M. L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, New York, NY, USA:Wiley, 2014

[10] Y. H. Xu, Y. B. Tian, P. K. Searyoh, G. Yu, Y. T. Yong, "Deep reinforcement learning-based resource allocation strategy for energy harvesting-powered cognitive machine-to-machine networks," *Computer Communications,* vol.160, pp.706-717, July, 2020.

[11] M. Alzenad, A. El-Keyi, F. Lagum, and H. Yanikomeroglu, "3-d placement of an unmanned aerial vehicle base station (uav-bs) for energy-efficient maximal coverage," *IEEE Wireless Communications Letters*, vol. 6, no. 4, pp. 434-437, May, 2017.