CISTER

Journal Paper

Data-driven Flight Control of Internet-of-Drones for Sensor Data Aggregation using Multi-agent Deep Reinforcement Learning

Kai Li*
Wei Ni
Yousef Emami*
Falko Dressler

Data-driven Flight Control of Internet-of-Drones for Sensor Data Aggregation using Multi-agent Deep Reinforcement Learning

Kai Li*, Wei Ni, Yousef Emami*, Falko Dressler

*CISTER Research Centre

Polytechnic Institute of Porto (ISEP P.Porto)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail: kai@isep.ipp.pt, Wei.Ni@data61.csiro.au, emami@isep.ipp.pt, dressler@ccs-labs.org

https://www.cister-labs.pt

Abstract

Energy-harvesting-powered sensors are increasingly deployed beyond the reach of terrestrial gateways, where there is often no persistent power supply. Making use of the internet of drones (IoD) for data aggregation in such environments is a promising paradigm to enhance network scalability and connectivity. The flexibility of IoD and favorable line-of-sight connections between the drones and ground nodes are exploited to improve data reception at the drones. In this article, we discuss the challenges of online flight control of IoD, where data-driven neural networks can be tailored to design the trajectories and patrol speeds of the drones and their communication schedules, preventing buffer overflows at the ground nodes. In a small-scale IoD, a multi-agent deep reinforcement learning can be developed with long short-term memory to train the continuous flight control of IoD and data aggregation scheduling, where a joint action is generated for IoD via sharing the flight control decisions among the drones. In a large-scale IoD, sharing the flight control decisions in real-time can result in communication overheads and interference. In this case, deep reinforcement learning can be trained with the second-hand visiting experiences, where the drones learn the actions of each other based on historical scheduling records maintained at the ground nodes.

Data-driven Flight Control of Internet-of-Drones for Sensor Data Aggregation using Multi-agent Deep Reinforcement Learning

Kai Li, *Senior Member, IEEE,* Wei Ni, *Senior Member, IEEE,* Yousef Emami, *Student Member, IEEE,* and Falko Dressler, *Fellow, IEEE*

Abstract—Energy-harvesting-powered sensors are increasingly deployed beyond the reach of terrestrial gateways, where there is often no persistent power supply. Making use of the internet of drones (IoD) for data aggregation in such environments is a promising paradigm to enhance network scalability and connectivity. The flexibility of IoD and favorable line-of-sight connections between the drones and ground nodes are exploited to improve data reception at the drones. In this article, we discuss the challenges of online flight control of IoD, where data-driven neural networks can be tailored to design the trajectories and patrol speeds of the drones and their communication schedules, preventing buffer overflows at the ground nodes. In a small-scale IoD, a multi-agent deep reinforcement learning can be developed with long short-term memory to train the continuous flight control of IoD and data aggregation scheduling, where a joint action is generated for IoD via sharing the flight control decisions among the drones. In a large-scale IoD, sharing the flight control decisions in real-time can result in communication overheads and interference. In this case, deep reinforcement learning can be trained with the second-hand visiting experiences, where the drones learn the actions of each other based on historical scheduling records maintained at the ground nodes.

Index Terms—Internet of drones, Multi-agent deep reinforcement learning, Flight control, Data aggregation, Long short-term memory

I. APPLICATIONS OF IoD-ASSISTED SENSOR NETWORKS

Highly mobile and interconnected drones, *a.k.a.* internet of drones (IoD), are increasingly employed as aerial data aggregators in many application scenarios [1], e.g., weather forecast, package delivery in rural areas, or crop monitoring on remote farms. The drones' flight can be coordinated, which enables new data aggregation paradigms and extends the use of drones from a stand-alone aerial platform to an important network component in next-generation wireless networks.

The IoD is ideal to provide connectivity to the ground nodes in remote and hostile areas by maneuvering over the target field [2], where the flight control of a drone determines next waypoints and patrol speeds. The drone can physically approach a ground node to collect the buffered data. A short-distance, line-of-sight (LoS)-dominant, communication link between the drone and a ground node enables high-speed data transmissions. Employing the IoD to collect data can improve the network throughput and extend the coverage range beyond terrestrial gateways.

II. CHALLENGES OF DATA-DRIVEN FLIGHT CONTROL

A. Flight control and data aggregation scheduling

Figure 1 presents the case study of a typical IoD-assisted sensor network, where agriculture sensors are used to monitor the crop growth and manage the growing environment [3]. Ground nodes deployed beyond the reach of terrestrial gateways, often have no persistent power supply, and may have to be powered by renewable energy harvested from ambient environments. The ground nodes are increasingly equipped with energy harvesting devices, such as solar panels or wind generators. A ground node can buffer sensory data awaiting to be transmitted. The sensory data often has a random packet arrival because the sensors typically generate data only when sensed values (e.g., temperature and humidity of the environment) change for the reduction of computation and transmission redundancy [4].

The flight control (i.e., velocities and trajectories) and the data collection schedule of the drones need to be appropriately coordinated to reduce the data loss resulting from the buffer overflows and transmission failures of the nodes. The reason is that sensory data of a ground node is often event-driven and generated when the environment changes, e.g., a change in the temperature and humidity of the soil. A ground node can undergo dynamic data arrivals at its buffer. When a drone approaches in an attempt to collect data from a ground node with an empty data buffer, the other ground nodes may already have full buffers and suffer from buffer overflows upon the arrivals of new data. On the other hand, when a drone is far away from a ground node scheduled for data aggregation, the poor link quality is prone to data reception errors [5]. The link quality between the drone and the ground node can change drastically when the drone flies at a high speed due to Doppler-induced fast fading. In contrast, a slow patrol speed of the drone can give rise to buffer overflows at the ground nodes, as newly arrived data cannot be promptly delivered.

Despite the communication range, memory and storage capacity of the ground nodes have been continuously improving, the data buffer can still (and always) overflow in practice. First, a queue's size will grow rapidly if the incoming data rate of the queue is greater than its outgoing rate counterpart, hence leading to a buffer overflow. This is likely to take place when there are a large number of ground nodes or the trajectories of the IoD is inadequately
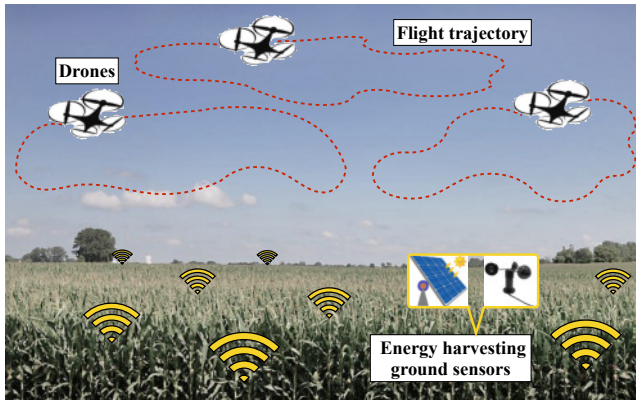
Fig. 1: Energy-harvesting-powered agriculture sensors are deployed in a remote farmland for crops monitoring. IoD manoeuvres over the target field following their trajectories, while collecting and/or processing the real-time sensory data.

designed. Second, emerging sensing platforms, comprising a notably large number of compact integrated sensors that are characterized by larger data analytics capability, generate enormous data volumes. For instance, the size of a high-definition picture and audio/video clip can be a few megabytes, while most commercial-off-the-shelf sensors have small data storage due to considerations of cost and energy budgets.

B. Data-driven decision-making on flight control

It is important to jointly design the flight control of the IoD and its data aggregation schedule by taking the battery levels and propulsion energy consumption of the drones into account [6]. This is because different patrol velocities of the drone can incur different energy consumptions of the flight control, which may deplete the onboard battery. Since a sensor node cannot perform data transmission when its battery level is flat, newly arrived data can potentially result in the buffer overflow of the ground node. The ground nodes are equipped with rechargeable batteries to harvest renewable energy from ambient sources, such as solar and wind, and power their sensing and transmission operations. With a finite capacity, the batteries can be overcharged. In this sense, the environmental conditions have a strong impact on the energy that can be harvested. The battery levels of the sensor nodes are an indispensable part of the environment parameters, and can substantially affect the operation of the drones.

According to [7], each of the drones in the IoD can independently determine an action for its flight control, namely, flight trajectories and patrol speeds, and the ground node selection, to minimize the data loss resulting from buffer overflows and communication failures. Not only can an action determine the future network states, but also have non-negligible influence on the actions of the other drones. Such a data-driven flight control can be interpreted as a multi-agent Partially Observable Markov Decision Process (POMDP) [8], where each drone plays the role

of an agent. The network state can consist of the battery levels and data queue lengths of the ground nodes, the link qualities, and the locations and battery levels of the drones. Dynamic programming algorithms, such as value iteration or policy iteration, are typically used to solve POMDPs offline, provided that the a-priori knowledge of the state transition probabilities of the system is available. The optimal action-value function of the value/policy iteration is estimated and updated based on the Bellman optimality equation. The action-value function can be reinforced once the value/policy iteration converges.

C. Sharing observations for training the IoD

In practical IoD-assisted sensor networks, state transitions in POMDP are often unknown since the complete network state information, such as the battery levels and data backlogs of the ground nodes, is not instantaneously observable to the drones. This is because the ground nodes are geo-distributed over a large area (e.g., a remote farmland or a human-unfriendly rainforest). A drone can barely maintain real-time wireless connections with all the ground nodes. Deep reinforcement learning (DRL) [9] can be adopted to solve the multi-agent POMDP, and learn the best actions of each drone in the IoD online. Specifically, each of the drones individually conducts DRL to find its best strategy, e.g., to minimize the data loss of the ground nodes, adapting to its local observations of the network state and the actions of the other drones [10]. Moreover, the drone can conduct vision-based techniques or utilize event cameras to adjust the flight attitudes for collision or obstacle avoidance.

For multi-agent DRL, the actions have to be cooperatively trained by sharing the actions with each other in the IoD, where the action of a drone needs to be determined by taking into account the other drones' flight control and ground node selection. Otherwise, multiple drones can generate the same flight trajectory while selecting the same ground node for data aggregation, reducing the service efficiency of the IoD. In practice, it is formidably challenging for many drones to share their actions, i.e., flight trajectories, patrol velocities, and the ground node selection, with each other in real-time due to the limited radio coverage and fast movement of the drone. Given a large number of drones, online sharing in the IoD can result in high signaling overhead and strong interference to the transmission of the ground nodes.

III. DATA-DRIVEN MULTI-AGENT DEEP REINFORCEMENT LEARNING WITH IOD

A. Data-driven flight control based on LSTM-DDPG

Flight control of the IoD is typically conducted with a large number of continuous real numbers, e.g., instantaneous coordinates and velocities of the drones, link qualities, and battery levels of the ground nodes. This would immensely extend the state and action space in the multi-agent DRL. Thus, the multi-agent DRL for learning the network states and actions in a discrete domain [11], e.g.,

(a) multi-agent LSTM-DDPG with online sharing.

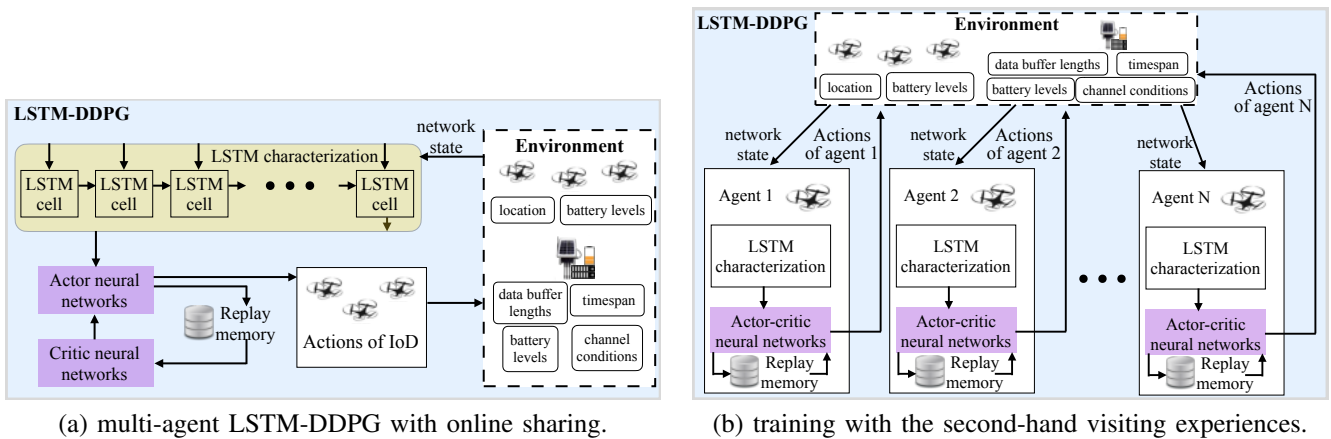(b) training with the second-hand visiting experiences.

Fig. 2: The architectures of multi-agent LSTM-DDPG in small or large IoD-assisted sensor networks.

deep Q-learning (DQL), is hardly to be used for the flight control of the IoD. Moreover, it is difficult for an individual drone to learn the data packet arrival pattern, battery level fluctuation (due to solar charging), and the link qualities (due to channel fading randomness) of all the ground nodes. In other words, the drones cannot train the multi-agent DRL with a complete observation of all the network states. Such incomplete network state observation can compromise the efficiency and accuracy of the flight control and data aggregation schedule.

In Figure 2(a), we study a new multi-agent DRL based on deep deterministic policy gradient (DDPG) in small-scale IoD-assisted sensor networks, which can achieve the flight control of the IoD with the continuous states and actions. Specifically, the drones can share with each other about their real-time decisions of flight control and communication schedules. The new multi-agent DRL can be implemented at the drones, which leverages actor-critic neural networks in DDPG to train a joint action of the agents, i.e., instantaneous headings and patrol velocities of the drones, as well as the selection of the ground node. In particular, the selection of the ground node, which is given as a positive number in the training of DDPG, can be discretized to a natural number. With the increase of the training time, the joint action in the continuous domain is sufficiently trained, which can reduce the discretization error regarding the selection of the ground node.

For enhancing the network state observations in the training environment of the multi-agent DDPG, the new multi-agent DRL in Figure 2(a) develops a long short-term memory (LSTM) to predict the time-varying battery levels, data queue backlogs, and link qualities of the ground nodes [12]. The LSTM utilizes memory cells in deep neural networks to process input data sequentially and embrace the hidden state over time. The memory cell captures long-term (often unknown) dependency among sequential time-varying data. The LSTM characterization helps address the incompleteness of the network state observation at the drones in the sense that the obscure network state is approximated for the subsequent joint action training with

DDPG.

Each of the drones can be equipped with a replay memory to store its training experience at every learning epoch to facilitate evaluating the joint action in the critic neural networks. Mini-batches of the learning experiences can be randomly sampled from the replay memory to train the action of the drone, along with the shared actions of the other drones and the network state of the environment.

B. Training with the second-hand knowledge

Another multi-agent DRL architecture of the IoD is that each of the drones (i.e., agents) can learn independently based on its own observation (or in other words, the observable network information). This is particularly important in the situation where the coverage of the IoD is vast and the drones do not have reliable connectivity with each other. None of the drones can have the complete, instantaneous knowledge of the entire network state.

To train the multi-agent LSTM-DDPG in a large-scale IoD, sharing the cruise control information in real-time among the drones becomes not possible. To circumvent this impasse, Figure 2(b) illustrates a potential solution that allows the drones to train their DRL from second-hand (typically outdated) network state information. Specifically, the drones can receive the visitor log from each ground node along with the sensing data, upon the polling of the node by the drone. The visitor log of the ground node records the visits that all drones have made, including the drones' IDs, flight control, offloaded data size, and consumed energy, since the last visit that the currently polling drone made to the ground node [13].

The multi-agent LSTM-DDPG allows the drone to store those second-hand visiting experiences of the visitor log to its onboard replay memory. The local visitor log enables each of the drones to train its onboard DDPG and LSTM for the action of flight control and communication schedules with the consideration of the other drones' decisions. This can prevent the same flight trajectory from being generated by multiple drones and the same ground node from being repeatedly selected for data aggregation.

Every time a ground node is scheduled by a drone, the ground node only needs to update the recent historical record of visits for the drone. Thus, the local visitor log typically has a small size. Consider 100 drones and 100 ground nodes, the size of the historical record at a ground node is just 5 kilobits, where the network state has 30 bits (10 bits for battery levels and data buffer lengths of the ground nodes, link qualities) and 20 bits for the action (7 bits for node ID and 13 bits for waypoints and speeds). Given the data rate of 256 kbits/s, it only takes 0.02 second to transfer the whole local visitor log to the drone. Moreover, the time that the ground node takes to update the record is negligible. Therefore, the multi-agent LSTM-DDPG with the second-hand visiting experiences requires a small amount of computation and communication overhead at the ground nodes, which is feasible in the large-scale IoD-assisted sensor networks.

C. Heads and tails of a coin

In Table I, we compare the pros and cons of several typical multi-agent learning techniques. The key difference between the multi-agent LSTM-DDPG with online sharing (in Figure 2(a)) and the one with the second-hand visiting experiences (in Figure 2(b)) is that the online sharing can enable the DDPG to instantly train each drone's action according to the instantaneous decisions of all other drones and the time-varying network state. This can improve the learning accuracy and also accelerate the learning convergence. Nevertheless, the multi-agent LSTM-DDPG with online sharing suffers from large signaling overheads with the increase of agents. In the large-scale IoD-assisted sensor networks, the second-hand visiting experiences can strengthen the experience replay in the DDPG and the state prediction in LSTM. It is worth noting that the multi-agent LSTM-DDPG may sacrifice the learning accuracy and convergence time since the drones train their real-time flight control based on the historical outdated knowledge in the local visitor log. It is also worth mentioning that the new multi-agent LSTM-DDPG can be potentially repurposed to support different objective functions. For example, it can be potentially repurposed to improve the energy efficiency, which is the ratio of network throughput to the energy consumption.

IV. PERFORMANCE EVALUATION

A. Implementation of the multi-agent LSTM-DDPG

The multi-agent LSTM-DDPG can be implemented in Python on Google TensorFlow or Pytorch, which are the two most widely used machine learning platforms. In this paper, TensorFlow is set up on a 64-bit Ubuntu 18.04 workstation running on an INSYS CorporateWorkstation (equipped with 4-core Intel i7-6700K 4GHz CPUs and 16G memory). The training and performance evaluation on the machine learning platforms is critical to further develop or extend the IoD-assisted sensor networks in real world. Additionally, the multi-agent LSTM-DDPG model can account for generic collision or obstacle avoidance via

the offline training and can adapt to the specific real-world application scenario via online refinement. The drones can also be equipped with event cameras or utilize vision-based techniques to avoid collisions and adjust the flight behavior.

Data aggregation of the IoD can consist of a number of time frames. Each time frame that contains a number of time slots can be allocated for the drones' flight and data collection. Moreover, each of the drones can determine which ground node to collect data and then fly to the selected sensor. Next, the drone broadcasts a short beacon message, which contains the ID of the selected ground node, to initialize the data aggregation. Upon the receipt of the beacon message, the selected ground node starts the data offloading to the drone. In particular, the hardware information of the ground node, e.g., battery level and data buffer length, can be involved in the control segment of the data packet.

The overhead of this control segment is small. For example, consider battery level of 100 and 100 packets in the buffer, the overhead is only 12 bits, much smaller than the size of the data packet. Therefore, the transmission time and the energy consumption of the control segment can be negligible. Once the data is correctly received, the drone can send an acknowledgment to the ground node.

For training the LSTM characterization, the network state, i.e., battery levels and data buffer lengths of the ground nodes, and air-ground link qualities, can be measured offline with the energy-harvesting-powered sensor [14], [15]. By utilizing the offline datasets, the future network state can be predicted and imported to enrich the training environment. In addition, The LSTM can be implemented in Keras (the Python deep learning library) while DDPG can be configured in TensorFlow to minimize the training loss.

B. Numerical analysis

100 ground sensors are randomly deployed in a square area with a size of 1,000 m × 1,000 m. Each of the ground sensors can buffer 20 packets at most in its queue. The onboard replay memory of the drone can store 10,000 samples of the learning experience at every step. The performance of the multi-agent LSTM-DDPG is compared with two state-of-the-art multi-agent approaches using DQL and channel-priority assignment (CPA):

- Multi-agent DQL (MA-DQL). The trajectories of the drones are predetermined, where MA-DQL is trained to schedule the data transmission of the ground sensors by learning the changes of their battery levels, buffer lengths, and channels.
- Multi-agent CPA (MA-CPA). The drones move at their lowest speed and send beacons to the ground sensors. The drones measure the air-ground link qualities according to the ground sensors' replies to the beacons. The ground sensor with the highest link quality is scheduled for data aggregation.

Figure 3 presents the training of the new multi-agent LSTM-DDPG in regards to the episodes, where the number

TABLE I: Comparison of the typical multi-agent DRL methodologies in IoD-assisted sensor networks.

| Multi-agent learning methodologies	Key features	Heads	Tails
LSTM-DDPG with online sharing	Drones can share the real-time actions.	Improved learning accuracy and accelerated learning convergence.	Large signaling overheads with the increasing number of agents.
LSTM-DDPG with the second-hand visiting experiences	Drones train their DRL based on the visitor log of ground nodes.	Reduced overheads with the large number of agents.	Extended training time for the convergence.
DDPG	Enabling network state observations and training of actions in the continuous domain.	Instantaneous headings and patrol velocities of the drones can be trained.	Each of the drones trains the action with incomplete network state observations.
DQL	Actions of the drones are trained in the discrete domain.	Can be used for training discrete actions, e.g., way-points planning or communication scheduling.	Cannot be used to train the continuous flight control of the drones.
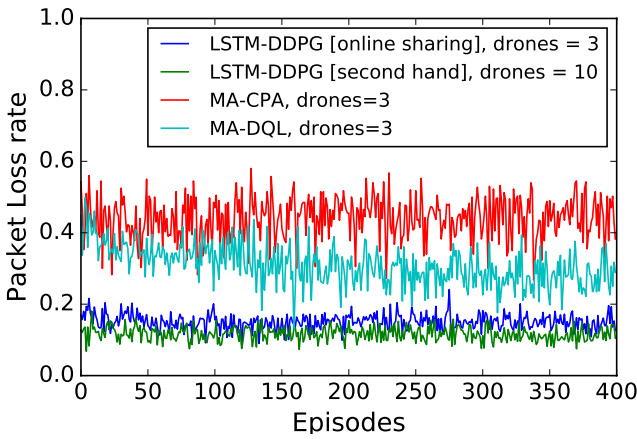
Fig. 3: Training of the multi-agent LSTM-DDPG, where the IoD contains 3 or 10 drones.

Fig. 4: Trajectories of the drones are designed by the multi-agent LSTM-DDPG model, where the number of drones is 2 or 10.

of iterations for training the LSTM characterization is set to 500. Given 3 drones, the LSTM-DDPG model conducts the online sharing. Once the LSTM-DDPG model is sufficiently trained and the performance converges to 18.5%, which achieves the lowest packet loss rate, as compared to MA-DQL and MA-CPA. The reason can be twofold. First, the LSTM characterization of the LSTM-DDPG model, which is trained by the offline datasets, effectively predicts the time-varying network states in the learning environment of the DDPG. Therefore, the DDPG can train the actions of the drones with both the observed and predicted state information of all the ground sensors to minimize the packet loss. Second, the flight control of the IoD and data aggregation scheduling can be adapted by the DDPG in the continuous action space. This can lead to more degrees of freedom on the trajectory design and speed control while scheduling more potential ground sensors to minimize the packet loss.

When the number of drones increases to 10, the LSTM-DDPG model takes advantage of the second-hand visiting experiences, and the packet loss rate drops to around 9.6%. This is because increasing the number of drones allows more ground sensors to be scheduled in parallel, hence
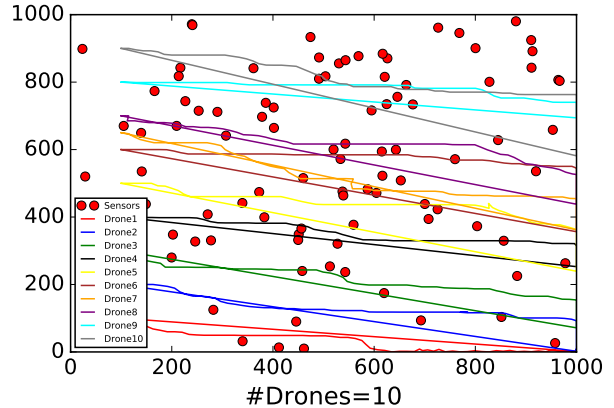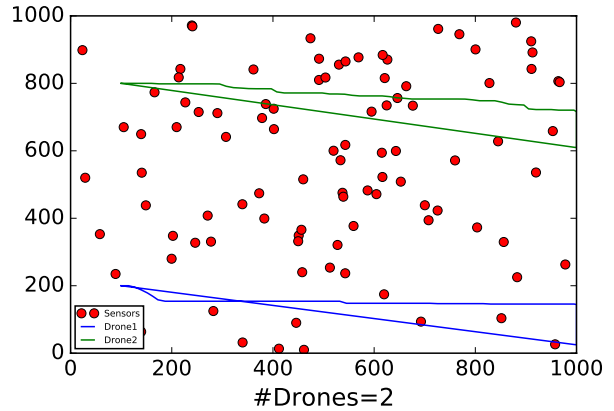
reducing the buffer overflow. Despite increasing the number of drones can extend the data aggregation coverage and reduce the packet loss, the learning accuracy of the LSTM-DDPG model may decrease. Since the drones may not schedule the same ground node, some records in the visitor log of the ground node are not updated. As a result, the drone is not able to timely learn the others' flight control decisions.

Figure 4 plots the trajectories of the drones, which are optimized by the multi-agent LSTM-DDPG model. Every drone of the IoD is employed to maneuvere over part of the target field where a large number of ground nodes are sparsely deployed. Given two drones, the multi-agent LSTM-DDPG applies online sharing for the flight control and data aggregation, where the drones are aware of the real-time actions of each other. Given ten drones, the multi-agent LSTM-DDPG is trained with the second-hand visiting experiences, where the trajectory is designed according to the historical records in the local visitor log of the ground node. The multi-agent LSTM-DDPG is trained at each of the drones, and then, the drones separately move to different areas for data aggregation. Therefore, Figure 4 validates the feasibility of the multi-agent LSTM-DDPG in small and large IoD-assisted sensor networks.

V. Conclusions and Future Directions

We discussed the need for data-driven flight control in the IoD. As shown in this article, deep learning approaches help to significantly improve the performance. In particular, we presented a multi-agent LSTM-DDPG model for flight control and data aggregation scheduling. Given a small number of drones, the joint action of the multi-agent LSTM-DDPG model can be trained with the shared online decisions of the drones to minimize buffer overflows at the ground nodes and communication failures. When increasing the number of drones, sharing the online flight control and scheduling decisions can result in large communication overhead and interference in the IoD. In this case, the multi-agent LSTM-DDPG model can be trained with the second-hand visiting experiences, where the drones learn the actions of each other based on the records at the ground nodes.

In future research, the multi-agent LSTM-DDPG model could be extended to improve the IoD-assisted data aggregation while satisfying heterogeneous quality-of-service requirements. For example, considering a stringent delay requirement of the aggregated data, the training time of the multi-agent LSTM-DDPG model at the drones has to be reduced to ensure the data freshness. In addition, more datasets will be used to train the LSTM-DDPG model to validate the stability and optimality in diverse IoD-assisted applications.

References

[1] Z. Kaleem, M. Yousaf, A. Qamar, A. Ahmad, T. Q. Duong, W. Choi, and A. Jamalipour, "UAV-empowered disaster-resilient edge architecture for delay-sensitive communication," *IEEE Network*, vol. 33, no. 6, pp. 124–132, 2019.

[2] Y. Xu, Z. Liu, C. Huang, and C. Yuen, "Robust resource allocation algorithm for energy harvesting-based D2D communication underlaying UAV-assisted networks," *IEEE Internet of Things Journal*, 2021.

[3] P. Spachos and S. Gregori, "Integration of wireless sensor networks and smart UAVs for precision viticulture," *IEEE Internet Computing*, vol. 23, no. 3, pp. 8–16, 2019.

[4] M. Z. A. Bhuiyan, J. Wu, G. Wang, T. Wang, and M. M. Hassan, "e-sampling: Event-sensitive autonomous adaptive sensing and low-cost monitoring in networked sensing systems," *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, vol. 12, no. 1, pp. 1–29, 2017.

[5] J. Cui, Z. Ding, Y. Deng, and A. Nallanathan, "Model-free based automated trajectory optimization for UAVs toward data transmission," in *IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6.

[6] Q. Zhang, W. Saad, M. Bennis, X. Lu, M. Debbah, and W. Zuo, "Predictive deployment of UAV base stations in wireless networks: Machine learning meets contract theory," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 637–652, 2020.

[7] K. Li, W. Ni, and F. Dressler, "Continuous maneuver control and data capture scheduling of autonomous drone in wireless sensor networks," *IEEE Transactions on Mobile Computing*, 2021.

[8] H. Bayerlein, M. Theile, M. Caccamo, and D. Gesbert, "Multi-UAV path planning for wireless data harvesting with deep reinforcement learning," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1171–1187, 2021.

[9] K. Li, W. Ni, E. Tovard, and A. Jamalipour, "Online velocity control and data capture of drones for the internet of things: An onboard deep reinforcement learning approach," *IEEE Vehicular Technology Magazine*, vol. 16, no. 1, pp. 49–56, 2020.

[10] Y. Emami, B. Wei, K. Li, W. Ni, and E. Tovar, "Deep Q-networks for aerial data collection in multi-UAV-assisted wireless sensor networks," in *International Wireless Communications and Mobile Computing (IWCMC)*. IEEE, 2021, pp. 669–674.

[11] A. Feriani and E. Hossain, "Single and multi-agent deep reinforcement learning for AI-enabled wireless networks: A tutorial," *IEEE Communications Surveys & Tutorials*, 2021.

[12] K. Li, W. Ni, and F. Dressler, "LSTM-characterized deep reinforcement learning for continuous flight control and resource allocation in UAV-assisted sensor network," *IEEE Internet of Things Journal*, 2021.

[13] Y. Emami, B. Wei, K. Li, W. Ni, and E. Tovard, "Joint communication scheduling and velocity control in multi-uav-assisted sensor networks: A deep reinforcement learning approach," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2021.

[14] S. Yang, Y. Tahir, P.-y. Chen, A. Marshall, and J. McCann, "Distributed optimization in energy harvesting sensor networks with dynamic in-network data processing," in *IEEE INFOCOM*. IEEE, 2016, pp. 1–9.

[15] N. Ahmed, S. S. Kanhere, and S. Jha, "Utilizing link characterization for improving the performance of aerial wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 8, pp. 1639–1649, 2013.

VI. About the Authors
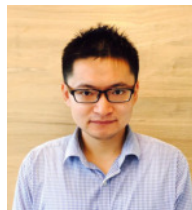
Kai Li [S'09-M'14-SM'20] (kai@isep.ipp.pt) is a senior research scientist and project leader at CISTER Research Centre, Portugal. He is a research fellow with Carnegie Mellon Portugal Research Program, which is jointly supported by Carnegie Mellon University, U.S., and FCT Portugal. He serves as an Associate Editor for Elsevier Ad Hoc Networks Journal, Elsevier Computer Communications Journal, and IEEE Access Journal.

Wei Ni [M'09-SM'15] (wei.ni@data61.csiro.au) received his B.E. and Ph.D. degrees from Fudan University, Shanghai, China, in 2000 and 2005, respectively. He is a Principal Research Scientist, Information Security and Privacy Group, CSIRO, Australia, and an Adjunct Professor at the University of Technology Sydney and Honorary professor at Macquarie University. He serves as the Chair of the IEEE VTS NSW Chapter, and an Editor for IEEE Transactions on Wireless Communications and IEEE Transactions on Vehicular Technology.

Yousef Emami [S'20] (emami@isep.ipp.pt) received the B.Sc. degree in computer engineering from Islamic Azad University of Dezful, Dezful, Iran, in 2007, and the M.Sc. degree in computer networks from Shiraz University of Technology, Shiraz, Iran, in 2015. He is currently Ph.D. student at University of Porto and researcher with Real-Time and Embedded Computing Systems Research Centre (CISTER). His research is focused on UAV networks.

Falko Dressler [S'02-M'04-SM'08-F'17] (dressler@ccs-labs.org) is full professor and Chair for Telecommunication Networks at the School of Electrical Engineering and Computer Science, TU Berlin. He is a member of the German National Academy of Science and Engineering (acatech). He has been an IEEE Distinguished Lecturer as well as an ACM Distinguished Speaker. His research objectives include adaptive wireless networking (sub-6GHz, mmWave, visible light, molecular communication) and wireless-based sensing with applications in ad hoc and sensor networks, the Internet of Things, and Cyber-Physical Systems.