Fully-Decentralised Reinforcement Learning in Fog-based Internet-of-Things

XXX YYYY, and XXX ZZZZ

***Abstract*—Reinforcement learning (RL) algorithms offers more insights about the overall futuristic functionalities for intelligent Internet-of-Things (IoT) devices. With the explosive growth in the number of IoT devices, as well as the highly-distributed deployments of these devices today, managing the IoT devices centrally becomes infeasible. As such, several disruptive paradigms have emerged, one of which is the fog computing-based IoT, which aim towards shifting computation, control, and decision-making closer to the network edge. However, mobility and power-constrain of these fog devices remains an issue of concern. In this paper, we aim to minimize the global outage in communication within a fog-based IoT network, by optimizing the power-control parameter of each potential mobile fog-relay agent (MFRA), as well as optimizing the physical position. As such, each MFRA is compelled to take certain actions that may influence its environment. We optimize the communication performance by applying a decentralized q-learning approach, with each MFRA acting independently, but contributing towards a global local-optimal policy. Furthermore, we show that our algorithm is scalable even with very large number of devices.**

***Index Terms*—Reinforcement learning (RL), Internet-of-Things (IoT), Fog-based IoT, Markov decision process (MDP), machine learning (ML).**

I. BACKGROUND

THE fog computing-based IoT paradigm aims at moving computation, control, and decision-making within the IoT ecosystem closer to the network edge [1]. The key driver of this paradigm are fog devices, which may be energy-constrained or not, and can either be mobile or static. The deployment and efficient utilization of these fog devices will contribute to the success of future IoT systems [2], one of which is overcoming communication outages due to obstacles or long distances between a source node and a remote destination node where IoT services may be rendered.

However, in order for devices to communicate efficiently with minimal outage, several bottlenecks may arise, one of which is the efficient utilization of energy by power-constrained IoT devices. Energy can be used up when these devices unnecessarily increase their power-level in order to communicate with neighbouring devices within the network, conversely, energy can be saved when the devices regulate their transmission power, especially in situations when they are relatively close to the communicating parties. For example, an IoT end-device that transmits at high power irrespective of the channel conditions or proximity of its neighbours may drastically deplete its energy and die-out, hence, resulting in communication breakdown due to a point-of-failure within the network. Another challenge to be addressed is the energy consumed due to mobility. In order to deliver on some acceptable quality-of-service (QoS) within the IoT ecosystem, it is imperative for devices to be mobile, however these devices are at risk of draining most of their energy on movement. For instance, energy will be inefficiently used if a power-constrained fog device decides to move closer to the communicating party only to relay very few number of packets, though having a guarantee that all the packets are delivered to the destination. However, if the size of number of packets to be relayed is large, it may be optimal for the fog device to move closer, hence, conserving the energy of the IoT end device, who can transmit no transmit at a lower power-level.

Moreover, since power-control mechanisms and smart mobility are critical in minimizing outages in communication, these devices should be able to learn when to increase, decrease, or maintain their power-levels or to move efficiently in order to increase the long-term performance of the network. More so, near-optimal actions from these devices are required to drive several smart cities applications, most importantly the Industrial IoT (IIoT), where industrial robots are deployed to act intelligible in a dynamic industrial setting, and intelligent monitoring applications, where surveillance drones are deployed in militarised zones to meet stringent quality-of-service (QoS) requirements [3].

The work in [3] used an iterative algorithm based on the steepest descent method to address the problem in a multi-tier fog-based IoT architecture with a fog device which could adjust its position and power-control parameter in order to minimize outage in communication. However, the work did not take into consideration the performance of the approach in a highly-decentralised IoT environment. In [5], a clustering algorithm was used on eight (8) unmanned aerial vehicle (UAVs) to collect data from ground IoT devices with an objective of minimizing the energy consumed by the IoT devices during uplink communications. However, this work considered a centralized network where the locations of every IoT device and UAV were known to the control center. Considering the highly-distributed nature of deployed IoT devices, it becomes infeasible to manage devices centrally [4]. As such, reinforcement learning can be effectively deployed on fog devices to allow them act independently based on their local experiences in the environment, i.e. each fog device should be able to learn independently without a central entity.

A decentralised stateless Q-learning approach was proposed in [4] to improve aggregate throughput in four coexistent wireless networks (WN). Each WN was considered to be an agent running the stateless Q-learning algorithm with

Manuscript received March 14, 2019; revised July X, 2019.

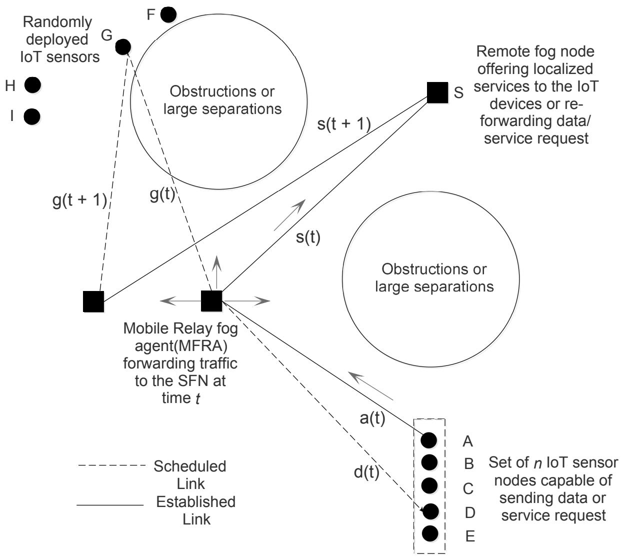

Fig. 1. Micro-view of the proposed fog-based IoT system.

agents having action space as channel number, and transmit power (dBm). A lightweight distributed learning approach was proposed in [7] to increase energy efficiency and reliability of IoT communications. There was significant performance improvement when the proposed algorithm was compared to a centralized optimized strategy. Transmit power, sub-channel, and spreading factor made up the action space. However, the system model in both works were rather hierarchical than distributed, ie. each WN was assumed to be an independent central entity with no specifications to what is learnt within each sub-network. Though IoT is defined as a large-scale network where various sub-networks coexist [1], applying RL to end-devices within sub-network may bring about meaningful performance improvement in the overall IoT network.

The main contribution of this paper is to propose a fully-decentralised multi-agent-based reinforcement learning approach as in [6] that addresses communication performance within a fog-based IoT architecture. First, we assumed a realistic scenario of possible state-action pair for each communication scenario as seen in Fig. 1, where each fog-agent is faced with a unique topology and environment. Next, in order to guarantee energy-efficient communications for the fog devices in the dynamic environment, we apply a decentralised q-learning algorithm, where each agent observes its position with respect to the communicating party, likewise the size of the number of packets to be transmitted and learns to take actions that improve the long-term communication performance. The fog-agents will actively try to communicate with neighbouring IoT devices in a manner where computing and processing are fully-decentralised, as well achieving energy balance within the system.

The remainder of this work is organized as follows. In Section II, we reviewed related works, and present our proposed approach in Section III. In Section IV, we evaluate the proposed fog-based IoT system, and present the results in Section V. Section V concludes the paper and outlines future directions.

II. Problem Definition

A. Problem Formulation

States: The states are defined as a tuple, $\langle$Outage communication cost ($\mathcal{P}_{out}$) /Energy of the fog relay (J) /Energy of the IoT sensor (J)$\rangle$.

- Outage communication cost: This observation will be evaluated based on the cost function in (1), which gives an estimate of the communication outage when there is an action, such as, change in the power level or location, or both.
- Energy of the fog relay: This observation gives the agent insight on how much energy has been expended on mobility.
- Energy of the IoT sensor: This observation gives the agent insight on how much energy the IoT sensor has used up due to high-power transmission.

$$\mathcal{P}_{out} = 1 - (1 + 2\Psi^2 \ln \Psi) \exp\Big(-\frac{N_0 \tilde{\gamma}}{P_I (D_I + \delta)^{-\alpha}}\Big), \quad (1)$$

where $\Psi = \sqrt{(N_0\tilde{\gamma})/(P_R(D_S+\delta)^{-\alpha})}$, and $\mathcal{P}_{out}$ is an expression for the outage probability with values between 0 and 1. We assume a predefined threshold $\tilde{\gamma}$ which determines the outage in communication, P_I is transmit power of the IoT sensor, P_R is transmit power of the fog relay agent, D_I is the distance between IoT sensor and fog relay agent, and D_S is the distance between fog relay agent and destination node. We assume a small change in the position of the fog relay agent, $\delta = 0.1m$, N_0 to be the channel noise, and α to be the path-loss exponent.

Actions: The actions are a tuple, $\langle$Move and transmit/power-level of IoT end-device$\rangle$.

- Mobility: Move by $\pm\delta$ and transmit, where $\delta = 0.1m$ and mobility range = [-20m, 20m]
- Power-level: Choose power-level and transmit, ρ, where transmit power ranges = [0.1W, 0.2W, 0.3W, 0.4W, 0.5W, 0.6W, 0.7W, 0.8W, 0.9W, 1W]

Goal: The goal is for the agent to learn to minimize cost, $\langle$Outage communication cost ($\mathcal{P}_{out}$) /Energy of the fog relay (J) /Energy of the IoT sensor (J)$\rangle$.

Rewards: We may consider the reward function in (2)

$$R = 100\Big[(1 - \mathcal{P}_{out}) + (1 - \frac{EU_{fog}}{C_{fog}}) + (1 - \frac{EU_{IoT}}{C_{IoT}})\Big], \quad (2)$$

where EU_{fog} and EU_{IoT} are the energy used up by the fog agent and the IoT sensor, C_{fog} and C_{IoT} are the initial capacity of the fog agent and the IoT sensor, respectively.

Metrics: Outage probability, i.e. the ratio of the number of packet lost to those transmitted, and Energy in Joules.

B. Related Works: Idea 1

Considering the dynamics and heterogeneity within the ultra-distributed IoT environment, it is important for communication devices which are mobile to move seamlessly without degrading quality of service (QoS). More so, the constrained devices should communicate efficiently without

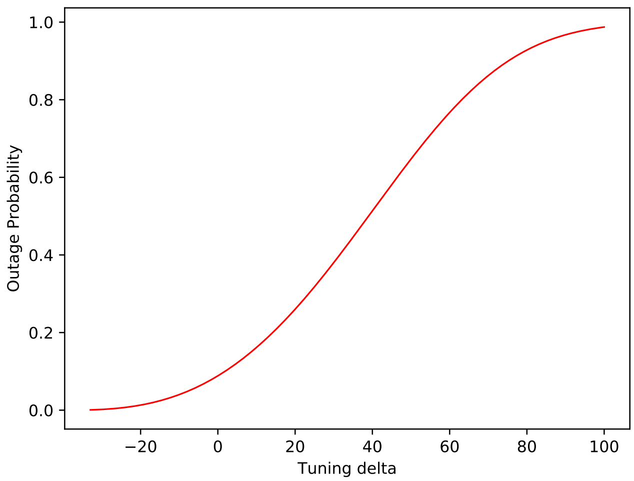

Fig. 2. Model relationship showing the outage observation w.r.t. change in position of fog agent.

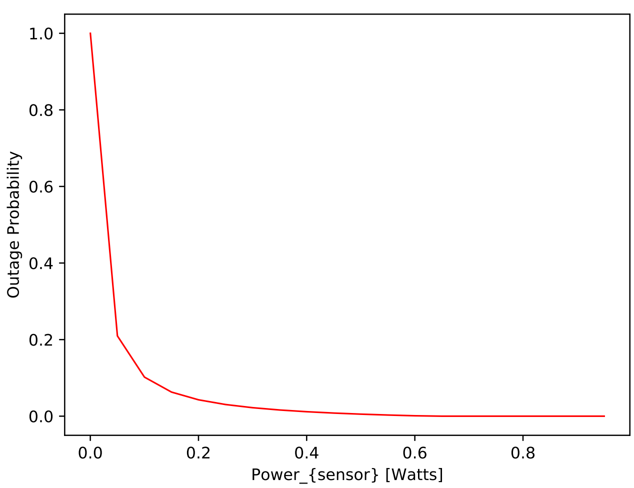

Fig. 3. Model relationship showing the outage observation w.r.t. change in power level of the IoT end-device.

depleting all their energy by transmitting at high-power, which may have long-term consequences to the network. Several non-RL-based approaches have been proposed to optimize the communication performance in IoT based networks, however, when some parameters in the network are changed, these approaches may fail. The work in [3] considered a multi-tier fog-based IoT architecture where a mobile/static fog node acts as an amplify and forward relay that transmits received information from a IoT sensor node to a higher hierarchically-placed fog device, which offers some localized services. In order to minimize the outage in communication, an iterative algorithm based on the steepest descent method (SDM) was proposed to jointly optimize the mobility pattern and power-control parameters. However, the single-agent based work did not take into consideration the performance of the approach in a highly-decentralised IoT environment. Furthermore, each time the topology is changed, the agent will need to recompute the model in other to act optimally.

RL can be applied to a new environment, since the it allows the agent to learn, by take actions that will improve its long-term return. Furthermore, RL can be applied to multiple agents. In [4], a decentralised, multi-agent-based stateless Q-learning approach was proposed, where no information about neighbouring nodes are available to agents. Though the paper was able to improve aggregate throughput in the network, by allowing the networks modify both the transmission power and the channel used, however, high variability was observed in the throughput of the individual networks. Only eight agents were observed using a stateless Q-learning approach. The work did not adequately depict the distributed nature of the IoT network. Moreover, mobility of agents was completely ignored.

Several works in WSN have applied the decentralised RL technique. In [41], a multi-agent reinforcement learning-based multi-hop mesh cooperative (MRL-CC) mechanism for the improvement of some QoS metrics in the WSNs. Though the mobility of the cooperative nodes were taken into account when learning the optimal policy, the MRL-CC failed to consider the power dissipated by the power-constrained devices, as well as the overall outage in communication.

In [39], a reliable and energy-efficient routing (REER) protocol was proposed using a geographic routing approach. The work considered the idea of a central entity, called a reference node (RN), which is assumed to be situated at an ideal location between source and destination. Several other cooperative nodes, which contend to relay data, are assumed to be situated around the RN. The work was able to examine the trade-off between reliability and energy-efficiency when the distances between RNs was adjusted.

We present a fully-decentralised multi-agent reinforcement learning approach that adequately addresses some performance issues in IoT. The main task of our work is to minimize global outage in communication within a fog-based IoT network, by optimizing the power-control parameter of the potential mobile fog-relay agent (MFRA), as well as optimizing the position of each relaying agents in the network. As such, each MFRA is compelled to take certain actions that may influence its environment. However, the duration it takes the MFRA to learn is significantly influenced by the state space, as well as the possible set of actions [18]. The variables for the state, action and reward of an agent may be discrete or continuous, with the former represented as small interval of values which imply distinct levels [46], and can easily be represented in a tabular form. However, it is difficult to represent continuous space using Q-learning tables. The work in [52] considered a RL agent that explores continuous state and action space using Gaussian unit search behaviour. Other works [18], [54] considered the reduction of states by eliminating states that are unlikely to occur. However, this may pose a big risk especially in a highly dynamic environment. RL can be effective for learning action policies in discrete stochastic environments, but its efficiency can decay exponentially with increasing state space [53]. Our proposed problem can be observed to have continuous state-action pairs, and is approached by discretizing the state and action space.

It is noteworthy that agents in a multi-agent system (MAS) may take actions that can have direct consequences on neighbouring agents, which may have further impact on other agents within the network. For instance, if a relaying agent decides to increase the transmit power beyond some threshold value, in order to boost its communication capabilities, its action may result in channel interference to its immediate neighbours, and worst, it may deplete its energy fast, and die-out, leading to link failure that can affect the performance of the entire network. Issues like this may arise in a typical multi-agent IoT network, as such, we present a fully-decentralised MAS where each agents learn to follow a local policy that improves the global objective. To the best of our knowledge, this is the first approach that employs a decentralised RL technique to minimize global outage in communication within a fog-based IoT network, taking into consideration the wireless channel conditions, by jointly optimizing the location of each relaying agent and the power-control parameter.

In this paper, we assume the following.

1) The MFRA is completely oblivious of its environment, and as such, has no prior knowledge of the overall cost function.
2) The MRFA may change its position in order to ensure better communication. Also, the MFRA may change its position (2D/3D) depending on the scenario considered.
3) The MFRA has an objective of learning to make actions that yield better outcomes within its local view of the environment.
4) The states are be divided into discrete levels to overcome the exponential decay in the efficiency of the proposed approach due infinite state space.
5) Each MFRA independently tries to optimize power usage and moves in a direction that maximizes the communication outage.

In Fig. 1, we present a situation where different IoT end-devices, inclusive of smart things embedded with sensors (smart-meters, smart-watches, traffic lights, washing machine, dish-washers, herds, or even a sick patient being monitored, etc.), which can send data/service request to a remote fog service provider via a MFRA. However, there may be change in the topology of the network or even in the environment, making it difficult for data collection or service provision. Furthermore, the environment may be too hazardous for human control and as such, some level of intelligence is expected from the MFRA. The MFRA learns to optimize its position and adjust it's power-level to minimize the communication outage. In Fig. **??**, we go further to define a finite state space where the MFRA acts.

However, the dynamic IoT environment increases the complexity in system design, as such, many classical approaches will fail when domain knowledge of the environment mismatches the actual environment [46].

In Fig. 1, we present micro-view of randomly deployed IoT sensors trying to send/receive data/service request to a remote fog device via a potential mobile fog-relay agent (MFRA). Intuitively, if a fog device does not have the required computational/processing capability, it may decide to forward the data/service request to other available fog devices.

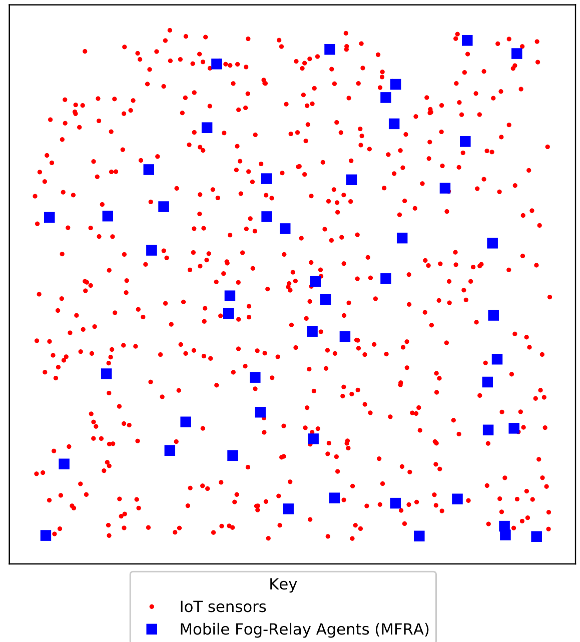

Fig. 4. A pictorial view of how MFRA and IoT sensors are randomly deployed.

III. IDEA 2

A. Problem Formulation

Devices/agents in an IoT network are resource-constrained, as such, it will be proper to deploy lightweight RL-based techniques that will improve the performance of the network. Furthermore, we may have to experiment on a very dynamic environment considering factors that depict a realistic IoT scenario to meet strict quality-of-service requirements.

In this work, a finite-horizon MDP is considered with continuous state and action spaces defined by the tuple $\langle \mathcal{S}, \mathcal{A}, p, p_0, \mathcal{P}_{out}, \gamma \rangle$, where $\mathcal{S}$ is the set of states, $\mathcal{A}$ is the set of actions, $p : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{G}^+$ is the conditional probability density over successor states given the current state and action, $p_0 : \mathcal{S} \to \mathbb{G}^+$ is the probability density over initial states, $\mathcal{P}_{out}$ is a function that maps state to cost, and the discount factor is $\gamma \in (0, 1]$. In the RL techniques, the agent has a choice to take certain actions in each time step, causing the environment to respond with new conditions, and consequently, the agent receives reward for that action as a form of feedback. The reward could be positive, negative or even zero, and the main objective of the agent is to maximize the positive reward or minimize the negative reward (often the cost) over the entire time step N.

Our objective is to learn a stochastic policy $\pi^* : \mathcal{S} \times \mathcal{A} \to \mathbb{G}^+$, which is a conditional probability density over the present state, in such a way as to minimize the expected cumulative cost.

$$\pi^* = \arg\min_{\pi} \mathbb{E}_{s_0,a_0,s_1,a_1,...,s_N}\Big[\sum_{i=0}^{N}\gamma^i \mathcal{P}_{out}(s_i)\Big], \quad (3)$$

We take the expectation over the joint distribution of all state-action pairs, with the density give as,

$$q(s_0, a_0, s_1, a_1, ..., s_N) = p_0(s_0)\prod_{i=0}^{N-1}\pi(a_i|s_i)p(s_{i+1}|s_i, a_i). \quad (4)$$

Fig. **??** shows a dynamic IoT environment with an IoT sensor attempting to request some services from a remote target fog agent/node through randomly deployed fog devices. The devices act as relays to forward traffic from the source to the destination. However, based on their position, line-of-sight(LoS) obstruction, which affect the conditions of the wireless channel, some degree of communication outage may occur. Our aim is to ensure that the agents are able to learn the optimal route to take through their experience with the environment.

IV. Ideas2

V. Ideas3

References

[1] B. Omoniwa, R. Hussain, M. A. Javed, S. H. Bouk and S. A. Malik, "Fog/Edge Computing-based IoT (FECIoT): Architecture, Applications, and Research Issues," in IEEE Internet of Things Journal.

[2] M. Chiang and T. Zhang, “Fog and IoT: An Overview of Research Opportuinities,” IEEE Internet of Things, vol. 3, no. 6, pp. 854-864, Dec. 2016.

[3] B. Omoniwa et al., "An Optimal Relay Scheme for Outage Minimization in Fog-based Internet-of-Things (IoT) Networks," in IEEE Internet of Things Journal.

[4] F. Wilhelmi, B. Bellalta, C. Cano and A. Jonsson, "Implications of decentralized Q-learning resource allocation in wireless networks," 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, 2017, pp. 1-5.

[5] M. Mozaffari, W. Saad, M. Bennis and M. Debbah, "Mobile Internet of Things: Can UAVs Provide an Energy-Efficient Mobile Architecture?," 2016 IEEE Global Communications Conference (GLOBECOM), Washington, DC, 2016, pp. 1-6.

[6] M. Gueriau and I. Dusparic, "SAMoD: Shared Autonomous Mobility-on-Demand using Decentralized Reinforcement Learning," 1558-1563. 10.1109/ITSC.2018.8569608.

[7] A. Azari and C. Cavdar, "Self-organized low-power IoT networks: A distributed learning approach," arxiv

[8] A. H. Ngu, M. Gutierrez, V. Metsis, S. Nepal and Q. Z. Sheng, "IoT Middleware: A Survey on Issues and Enabling Technologies," in IEEE Internet of Things Journal, vol. 4, no. 1, pp. 1-20, Feb. 2017.

[9] S. Earley, “Analytics, Machine Learning, and the Internet of Things,” *IT Professional,* vol. 17, no. 1, pp. 10-13, Jan.-Feb. 2015.

[10] D. Zhang, X. Han and C. Deng,“Review on the research and practice of deep learning and reinforcement learning in smart grids,” *CSEE Journal of Power and Energy Systems,* vol. 4, no. 3, pp. 362-370, September 2018.

[11] C. J. C. H. Watkins and P. Dayan, Machine Learning (1992), Kluwer Academic Publishers, vol. 8: 279.

[12] Z. Wen, D. O’Neill and H. Maei, “Optimal Demand Response Using Device-Based Reinforcement Learning,” *IEEE Transactions on Smart Grid,* vol. 6, no. 5, pp. 2312-2324, Sept. 2015.

[13] R. S. Sutton, and A. G. Barto, *Reinforcement learning - an introduction.* Cambridge, MA: MIT Press, 1998.

[14] M. A. Jadoon and S. Kim, "Relay selection Algorithm for wireless cooperative networks: a learning-based approach," in IET Communications, vol. 11, no. 7, pp. 1061-1066, 11 5 2017.

[15] J. Zhu, Y. Song, D. Jiang and H. Song, “A New Deep-Q-Learning-Based Transmission Scheduling Mechanism for the Cognitive Internet of Things,” *IEEE Internet of Things Journal,* vol. 5, no. 4, pp. 2375-2385, Aug. 2018.

[16] J. Zhu, Z. Peng and F. Li, “A transmission and scheduling scheme based on W-learning algorithm in wireless networks,” *2013 8th International Conference on Communications and Networking in China (CHINACOM),* Guilin, 2013, pp. 85-90.

[17] K. Gai, M. Qiu, “Optimal resource allocation using reinforcement learning for IoT content-centric services,” *Applied Soft Computing,* vol. 70, pp. 12-21, Sept. 2018.

[18] I. Dusparic and V. Cahill, “Distributed W-Learning: Multi-Policy Optimization in Self-Organizing Systems,” 2009 Third IEEE International Conference on Self-Adaptive and Self-Organizing Systems, San Francisco, CA, 2009, pp. 20-29.

[19] L. Zhou, A. Swain and A. Ukil, “Q-learning and Dynamic Fuzzy Q-learning Based Intelligent Controllers for Wind Energy Conversion Systems,” *2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia),* Singapore, 2018, pp. 103-108.

[20] T. Park, N. Abuzainab and W. Saad, “Learning How to Communicate in the Internet of Things: Finite Resources and Heterogeneity,” *IEEE Access,* vol. 4, pp. 7063-7073, 2016.

[21] A. Hans and S. Udluft, “Ensembles of Neural Networks for Robust Reinforcement Learning,” *2010 Ninth International Conference on Machine Learning and Applications,* Washington, DC, 2010, pp. 401-406.

[22] D. D. Nguyen, H. X. Nguyen and L. B. White, “Reinforcement Learning With Network-Assisted Feedback for Heterogeneous RAT Selection,” *IEEE Transactions on Wireless Communications,* vol. 16, no. 9, pp. 6062-6076, Sept. 2017.

[23] M. Mohammadi, A. Al-Fuqaha, M. Guizani and J. Oh, “Semisupervised Deep Reinforcement Learning in Support of IoT and Smart City Services,” *IEEE Internet of Things Journal,* vol. 5, no. 2, pp. 624-635, April 2018.

[24] S. Alletto et al., “An Indoor Location-Aware System for an IoT-Based Smart Museum,” *IEEE Internet of Things Journal,* vol. 3, no. 2, pp. 244-253, April 2016. doi: 10.1109/JIOT.2015.2506258

[25] K. Kolomvatsos, and C. Anagnostopoulos, “Reinforcement Learning for Predictive Analytics in Smart Cities,” *Informatics,* vol. 3, no. 16, June 2017.

[26] S. Conti, G. Faraci, R. Nicolosi, S. A. Rizzo and G. Schembra, “Battery Management in a Green Fog-Computing Node: a Reinforcement-Learning Approach,” *IEEE Access,* vol. 5, pp. 21126-21138, 2017.

[27] L. Mai, N.-N. Dao, and M. Park, “Real-time task assignment approach leveraging reinforcement learning with evolution strategies for long-term latency minimization in fog computing,” *Sensors,* vol. 18, no. 2830, Aug. 2018.

[28] C. Kwok and D. Fox, “Reinforcement learning for sensing strategies,” *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566),* Sendai, 2004, pp. 3158-3163 vol.4.

[29] Y. Li, K. K. Chai, Y. Chen and J. Loo, “Smart duty cycle control with reinforcement learning for machine to machine communications,” *2015 IEEE International Conference on Communication Workshop (ICCW),* London, 2015, pp. 1458-1463.

[30] Y. Li, K. K. Chai, Y. Chen and J. Loo, “Optimised delay-energy aware duty cycle control for IEEE 802.15.4 with cumulative acknowledgement,” *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC),* Washington, DC, 2014, pp. 1051-1056.

[31] M. Grammatopoulou, A. Kanellopoulos, and K. G. Vamvoudakis, “A multi-step and resilient predictive Q-learning algorithm for IoT: a case study in water supply networks”, *ACM Proceedings of the 8th International Conference on the Internet of Things,* Santa Barbara, California, 2018, pp. 1-8.

[32] Y. Debizet, G. Lallement, F. Abouzeid, P. Roche and J. Autran, "Q-Learning-based Adaptive Power Management for IoT System-on-Chips with Embedded Power States," 2018 IEEE International Symposium on Circuits and Systems (ISCAS), Florence, 2018, pp. 1-5.

[33] M. I. Khan, M. M. Alam, Y. Le Moullec and E. Yaacoub, "Cooperative reinforcement learning for adaptive power allocation in device-to-device communication," 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), Singapore, 2018, pp. 476-481.

[34] S. K. Routray and Sharmila K. P., “Routing in dynamically changing node location scenarios: A reinforcement learning approach,” 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), Chennai, 2017, pp. 458-462.

[35] Y. Liu, L. Liu and W. Chen, "Intelligent traffic light control using distributed multi-agent Q learning," 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, 2017, pp. 1-8.
[36] G. M. Dias, M. Nurchis and B. Bellalta, "Adapting sampling interval of sensor networks using on-line reinforcement learning," 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT), Reston, VA, 2016, pp. 460-465.
[37] M. Camelo and J. F. a. S. Latré, "A Scalable Parallel Q-Learning Algorithm for Resource Constrained Decentralized Computing Environments," 2016 2nd Workshop on Machine Learning in HPC Environments (MLHPC), Salt Lake City, UT, 2016, pp. 27-35.
[38] A. T. Nassar, and Y. Yilmaz, "Reinforcement-Learning-Based Resource Allocation in Fog Radio Access Networks for Various IoT Environments," Arxiv
[39] M. Chen, T. Kwon, S. Mao, Y. Yuan and V. C. M. Leung, "Reliable and energy-efficient routing protocol in dense wireless sensor networks," Int. J. Sen. Netw. vol. 4, no. 1/2, pp. 104-117, July 2008.
[40] M. Chen, M. Qiu, L. Liao, J. Park and J. Ma, "Distributed multi-hop cooperative communication in dense wireless sensor networks", *The Journal of Supercomputing,* vol. 56, no. 3, pp. 353-369, June 2011.
[41] Xuedong Liang, Min Chen, Yang Xiao, I. Balasingham and V. C. M. Leung, "A novel cooperative communication protocol for QoS provisioning in wireless sensor networks," 2009 5th International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities and Workshops, Washington, DC, 2009, pp. 1-6.
[42] M. Chen, Xuedong Liang, V. Leung and I. Balasingham, "Multi-hop mesh cooperative structure based data dissemination for wireless sensor networks," 2009 11th International Conference on Advanced Communication Technology, Phoenix Park, 2009, pp. 102-106.
[43] M. Mihaylov and Y. L. Borgne, "Decentralised reinforcement learning for energy-efficient scheduling in wireless sensor networks," Int. J. Communication Networks and Distributed Systems, Vol. 9, Nos. 3/4, 2012
[44] K. Shah and M. Kumar, "Distributed Independent Reinforcement Learning (DIRL) Approach to Resource Management in Wireless Sensor Networks," 2007 IEEE International Conference on Mobile Adhoc and Sensor Systems, Pisa, 2007, pp. 1-9.
[45] K. Shah, M. D. Francesco and M. Kumar, "Distributed resource management in wireless sensor networks using reinforcement learning," Wireless Networks, vol. 19, no. 5, pp. 705-724, July 2013.
[46] K.-L. A. Yau, P. Komisarczuk, P. D. Teal, "Reinforcement learning for context awareness and intelligence in wireless networks: Review, new features and open issues," Journal of Network and Computer Applications, vol. 35, no. 1, pp. 253-267, Jan. 2012.
[47] K.-L. A. Yau, H. G. Goh, D. Chieng and K. H. Kwong, "Application of reinforcement learning to wireless sensor networks: models and algorithms," Computing, vol. 97, no. 11, pp. 1045–1075, Nov. 2015.
[48] A. Egorova-Forster and A. L. Murphy, "Exploiting Reinforcement Learning for Multiple Sink Routing in WSNs," 2007 IEEE International Conference on Mobile Adhoc and Sensor Systems, Pisa, 2007, pp. 1-3.
[49] A. Forster and A. L. Murphy, "FROMS: Feedback Routing for Optimizing Multiple Sinks in WSN with Reinforcement Learning," 2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information, Melbourne, Qld., 2007, pp. 371-376.
[50] A. Forster, A. L. Murphy, J. Schiller and K. Terfloth, "An Efficient Implementation of Reinforcement Learning Based Routing on Real WSN Hardware," 2008 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, Avignon, 2008, pp. 247-252.
[51] A. Forster and A. L. Murphy, "CLIQUE: Role-Free Clustering with Q-Learning for Wireless Sensor Networks," 2009 29th IEEE International Conference on Distributed Computing Systems, Montreal, QC, 2009, pp. 441-449.
[52] N. Vucevic, J. Perez-Romero, O. Sallent and R. Agusti, "Reinforcement Learning for Active Queue Management in Mobile All-IP Networks," 2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications, Athens, 2007, pp. 1-5.
[53] W. T. B. Uther and M. M. Veloso, "Tree Based Discretization for Continuous State Space Reinforcement Learning," *Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence,* 1998, Madison, Wisconsin, USA, pp. 769-774.
[54] H. Cuayahuitl, S. Renals, O. Lemon, and H. Shimodaira. "Learning multi-goal dialogue strategies using reinforcement learning with reduced state-action spaces," In Int. Journal of Game Theory, pp. 547–565, 2006.