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. 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 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).**

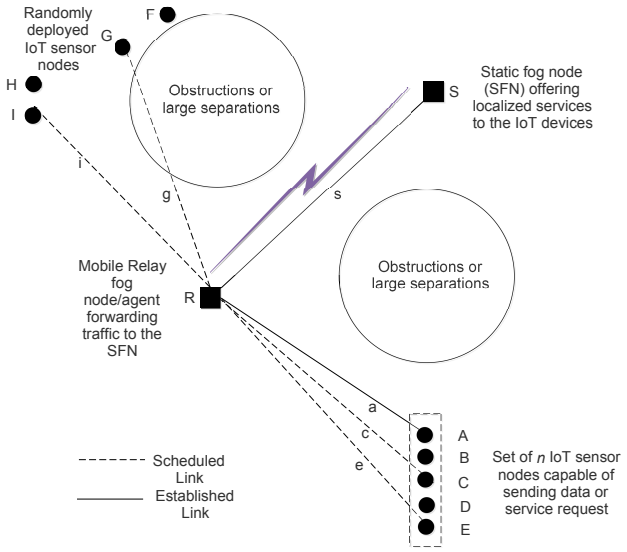

Fig. 1. System model for Idea1.

I. Motivation: Idea 1

For services to be effectively delivered, it is important to ensure that the communication outage within the network is greatly minimized. In Fig. 1, we present a scenario where n number of IoT sensors try to send/receive data/service request to a remote fog device, which has some unique computational/processing capability. Considering the highly-distributed nature of deployed IoT devices, as well as the tendency for increase in the network disruption due to potential obstacles in homes, hospitals, offices, militarised zones, and industries etc., there is a need for some level of intelligence in these devices. In addition to obstacles that obstruct line-of-sight(LOS) communications, the destination device may be to far from the source node, as such, it is possible to leverage fog devices to improve the reliability in the network link. The motivation for this idea was based on some very practical examples, the Industrial IoT (IIoT) and Intelligent monitoring, where industrial robots and surveillance drones could be deployed as relays for meeting stringent quality-of-service(QoS) requirements [3].

II. Related Works: Idea 1

Considering the dynamics and heterogeneity within the ultra-distributed IoT environment, it is important for devices which are mobile to move seamlessly without degradation of the quality of service (QoS) when communicating. Moreover, the communication should be ubiquitous irrespective of technology or domain within the system. As such, the RL techniques will play a very important role in areas of IoT device/service discovery, which will support adaptability and interfacing between devices and sub-networks within the IoT domain. Several works in WSNs [44], [45], [46], [47], focussed on improving some performance metrics such as energy, latency, throughput, within the network.

In [35], 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. The work in [36] proposed an improvement to [35], addressing the issue of scalability in a dynamic network. As such, a novel distributed multi-hop cooperative communication scheme (DMC) was proposed to improve certain QoS metrics such as the communication energy, hop count, and end-to-end delay. However, both works assumed complete prior knowledge of the network environment. In [37], a multi-agent reinforcement learning-based multi-hop mesh cooperative (MRL-CC) mechanism for the improvement of some QoS metrics such as the end-to-

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

end delay, packet delivery ratio in the WSNs. The work outperformed the work in [38], which incurred higher delay, and was distance-based, which is not suitable for a dynamic network environment. 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.

A RL-based Q-routing technique, the Feedback Routing for Optimizing Multiple Sinks (FROMS), was proposed in [44], [45] to enable efficient routing to multiple sinks without overhead. Simulations were done under scenarios where the sink nodes are mobile and with consideration for node failure. However, the mobility pattern and speed was bounded. Moreover, an important cost metric such as the energy consumed by the mobile nodes was not considered. In [46], an evaluation was carried out on the previously proposed FROMS framework using real WSN hardware. The objective was to show that machine learning algorithms could be efficiently deployed on resource-constrained devices. However, during the implementation phase, certain bottlenecks were encountered such as dynamic memory allocation, interference in the wireless channel, poor data gathering, and difficulty in handling lost or corrupted packets. As such, a data clustering and aggregation approach, CLIQUE, was proposed in [47] to minimize the energy expended by the selecting a cluster head using Q-learning. The approach was able to save significant amount of energy when compared to the traditional and random cluster head selection approach.

Some works [3], [10], [33], [25] have addressed pertinent issues on cooperative communications within the IoT environment. A Q-learning relay selection algorithm was introduced in [10] to maximize the throughput in a typical cooperative network with relays supporting communication between source and destination. The relays were assumed to be the states, with three possible actions of remaining in the present state with relay r or choosing $r+1$ or $r-1$. However, the outage analysis was carried out without a closed-form expression for the outage probability within the network. Furthermore, in an attempt to reduce the complexity of the formulated problem, the problem algorithm did not consider the relays (states) which do not satisfy the minimum mutual information constraint. A scalable parallel Q-learning algorithm was introduced in [33] to minimize communication cost. The work considered distributed and resource-constrained environment. A Q-learning-based duty cycle control technique was introduced in [25] to provide improved performance and reliable M2M communication for IoT applications. However, the proposed Q-learning based duty cycle control only considered a two-hop cluster tree network in evaluating the performance of the system. The work in [3] built on the concept presented in [4], 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 static fog device, which offers some localized services. An iterative algorithm based on the steepest descent method (SDM) was proposed to jointly optimize the mobility pattern and the transmit power of the fog relay. However, this work the scalability of the approach was not defined. Furthermore, the cost function was unrealistically assumed to be known to the fog relay.

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 [14]. 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 [42], and can easily be represented in a tabular form. However, it is difficult to represent continuous space using Q-learning tables. The work in [48] considered a RL agent that explores continuous state and action space using Gaussian unit search behaviour. Other works [14], [50] 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 [49]. 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 RN may lower/increase its power level to save energy or increase it 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.

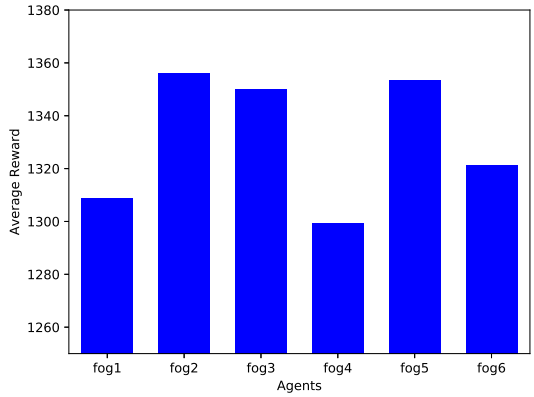

Fig. 2. Average reward for each agent over 2000 episodes.

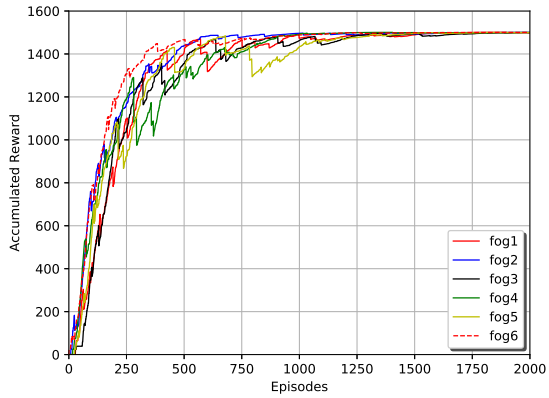

Fig. 3. Accumulated reward for each agent over 2000 episodes.

5) Each MFRA independently tries to optimize power usage and moves in a direction that maximizes the communication outage.

A. Problem Formulation

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 potential relay node. However, there may be change in the topology of the network or even in the environment, making it difficult for data collection/service provision. Moreover, the environment may be too hazardous for human control. Some level of intelligence is expected from the relay node/agent (RN). The RN learns to optimize its position and adjust it's power-level to minimize the communication outage.

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

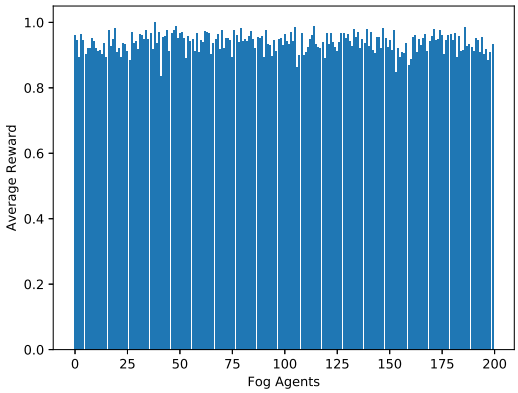

Fig. 4. Average reward for 200 agent over 1000 episodes.

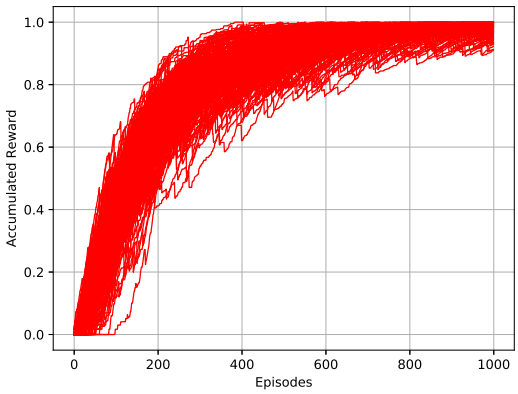

Fig. 5. Accumulated reward for 200 agent over 1000 episodes.

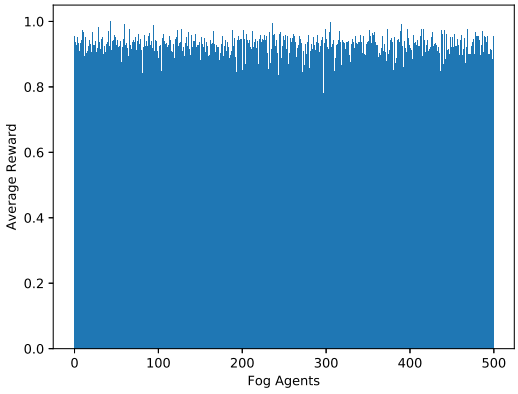

Fig. 6. Average reward for 500 agent over 1000 episodes.

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

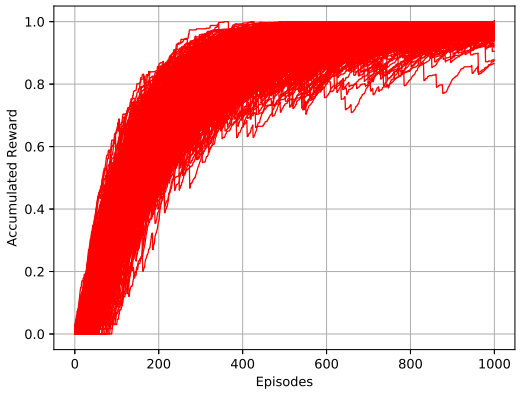

Fig. 7. Accumulated reward for 500 agent over 1000 episodes.

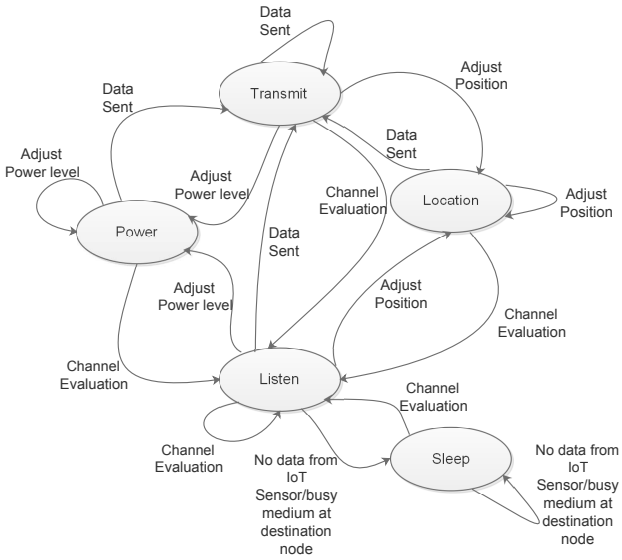

Fig. 8. MDP.

the set of actions, $p : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{G}^+$ is the conditional probability density over successor states given the current state and action, $p_0 : \mathcal{S} \rightarrow \mathbb{G}^+$ 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 $\pi^* : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{G}^+$, which is a conditional probability density over the present state, in such a way as to minimize the expected cumulative cost.

$$\pi^* = \arg\min_{\pi} \mathbb{E}_{s_0,a_0,s_1,a_1,...,s_N}\Big[\sum_{i=0}^{N} \gamma^i \mathcal{P}_{out}(s_i)\Big], \quad (1)$$

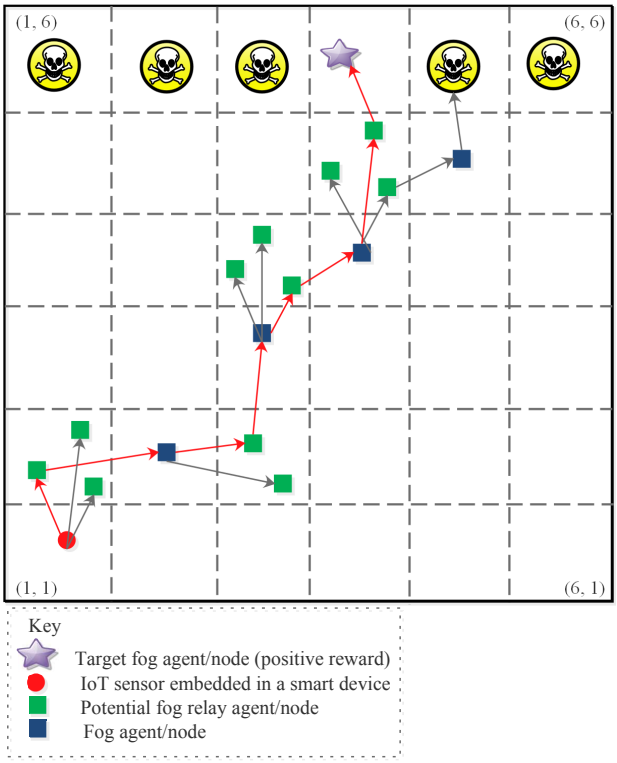

Fig. 9. System model depicting the role of fog agents in a dynamic and heterogenous IoT environment with a route policy in red.

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

$$q(s_0, a_0, s_1, a_1, ..., s_N) = p_0(s_0) \prod_{i=0}^{N-1} \pi(a_i|s_i)p(s_{i+1}|s_i, a_i). \quad (2)$$

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] 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.

[2] 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.

[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] 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.

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

[6] 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.

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

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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

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

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] M. Camelo and J. F. a. S. LatrÃl', "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.

[34] A. T. Nassar, and Y. Yilmaz, "Reinforcement-Learning-Based Resource Allocation in Fog Radio Access Networks for Various IoT Environments," Arxiv

[35] 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.

[36] M. Chen, M. Qiu, L. Liao, J. Park and J. Ma, "Distributed multi-hop cooperative communication inÂădense wireless sensor networks", *The Journal of Supercomputing,* vol. 56, no. 3, pp. 353-369, June 2011.

[37] 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.

[38] 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.

[39] 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

[40] 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.

[41] 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.

[42] 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.

[43] 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. 1045âĂŞ1075, Nov. 2015.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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. 547âĂŞ565, 2006.