# Reinforcement Learning in Fog-based Internet-of-Things: Applications and Research Issues

XXX YYYY, and XXX ZZZZ

***Abstract*—Reinforcement learning (RL) algorithms are the platform for connecting intelligent Internet-of-Things (IoT) devices. The data provided by these intelligent devices can offer more insights about the overall futuristic functionalities in terms of usage and preferences. Machine learning algorithms have been widely used for predictions, with several optimization algorithms using reinforcement learning techniques to improve the performance and interaction of devices within a dynamic IoT ecosystem. In this paper, we present a comprehensive review of state-of-the-art RL techniques that have been applied to solve trending problems in the IoT and cyber-physical systems domain. Furthermore, we give an in-depth analysis into its applications in the emerging fog-based IoT networks. Finally, we propose alternative research directions that may yield better outcomes.**

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

## I. IDEAS1

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 data/service request to a remote fog device, which has some unique computational/processing capability. Considering the nature of deployment of IoT devices, as well as the tendency for high level of network disruption due to potential obstacles in homes, hospitals and offices, etc., there is a need for online learning of devices to ensure interrupted communication. In addition to obstacles, the destination device may be to far from the sending 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.

### A. Problem Formulation

In Fig. 1, we present a situation where different IoT end-devices, inclusive of smart things embedded with sensors (smart meters, 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.

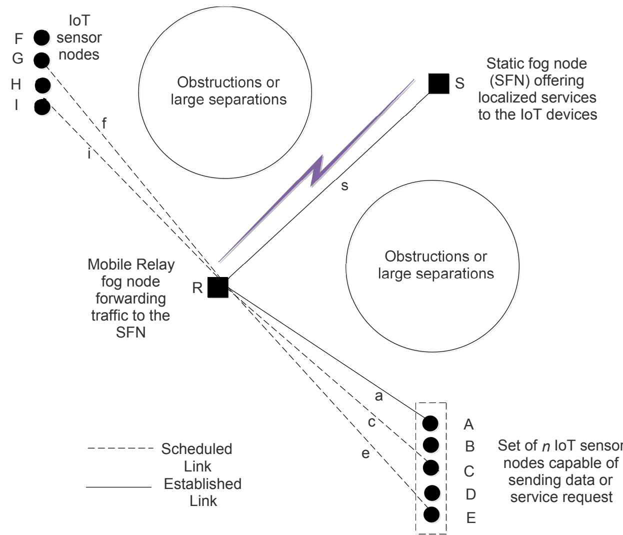


Fig. 1. System model for Idea1.

1) Actions: The RN may lower/increase its power level to save energy or increase it to ensure better communication. Also, the RN may change its position (2D/3D) depending on the scenario considered.
2) Goal of the RN is to effectively minimize the communication cost by carrying out some actions. We may also consider optimizing fairness by effective scheduling.
3) State: The states may be divided into time steps. At each time step, the RN transmits and is in a different position and power-control level.
4) Reward: When the RN uses maximum power or moves in a direction that maximizes the communication outage, then a negative reward is received, otherwise, a positive reward is received.

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

## II. IDEAS2A

### 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, $p : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{G}^+$ 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.

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

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.

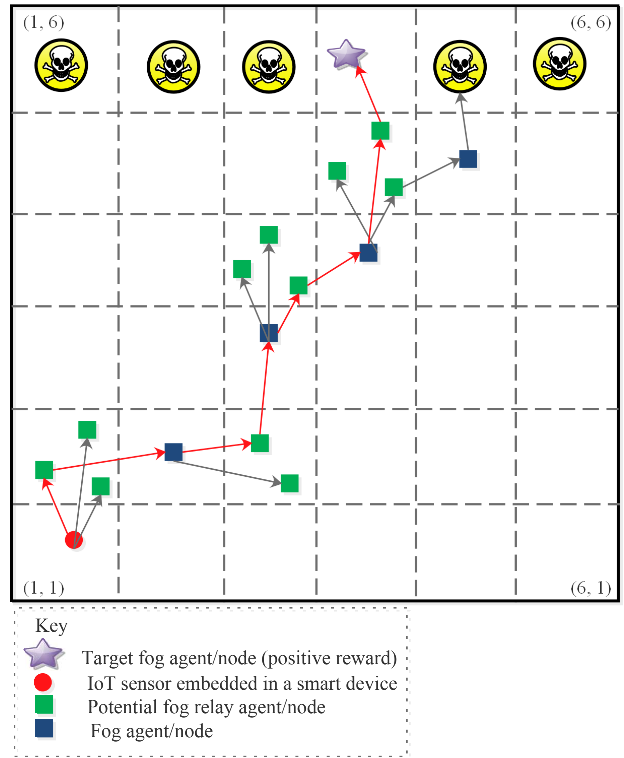


Fig. 2. System model depicting the role of fog agents in a dynamic and heterogenous IoT environment with a route policy in red.

## III. IDEAS2

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

[3] 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.

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

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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

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

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.