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 apply a decentralized q-learning algorithm to minimize the communication outage cost and optimize energy utilization within a fog-based IoT network. Our proposed approach outperforms the results from previous works by guaranteeing reliable delivery of data and minimizing overall energy cost within the network. Our future work will take into consideration fairness and latency within the fog-based IoT system.**

***Index Terms*—Reinforcement learning (RL), Fog-based Internet-of-Things (IoT), q-learning, communication outage, energy management.**

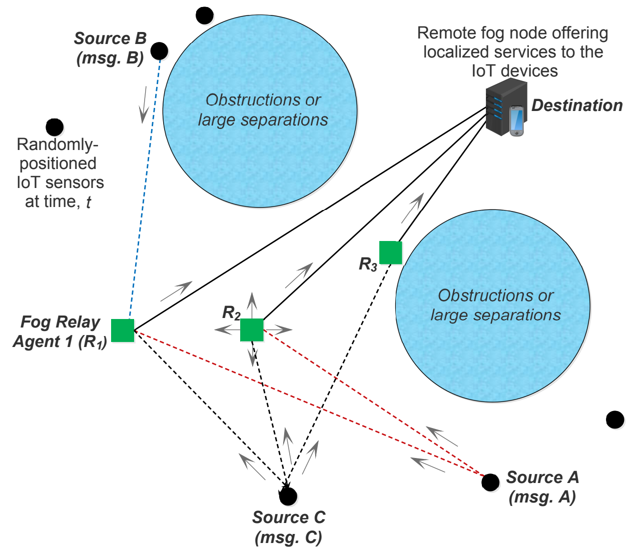

Fig. 1. System model.

I. Background

THE fog/edge computing-based IoT (FECIoT) paradigm aims at moving computation, control, and decision-making within the IoT ecosystem closer to the network edge [1]. Due to the limited range and power of IoT end-devices, mobile fog devices (which are often energy-constrained) will drive the FECIoT paradigm by forwarding sensed IoT data from source to destination. From reducing operational cost to improving channel reliability and load balancing, the mobile fog relays will play an important role in improving the overall network performance [2]. The deployment and efficient utilization of these mobile fog devices as relays will contribute to the success of future IoT systems [3], one of which is to overcome communication outages due to obstacles or long distances between a source (IoT sensor) and a remote destination node (where localized IoT services may be rendered).

However, in order for devices to communicate efficiently with minimal outage (loss of transmitted packets), several bottlenecks needs to be addressed, one of which is the inefficient utilization of energy by power-constrained mobile fog devices. For instance, energy can be used up when these devices actively communicate with neighbouring devices within the network, conversely, energy can be saved when the devices regulate their transmission by entering a passive mode, especially in situations when there are redundant relays that can convey same information. Furthermore, if multiple fog relays are in an active mode and unintelligently keep forwarding same message from an IoT end-device to a remote destination, soon they may run out of energy and die-out, hence, resulting in communication breakdown due to a point-of-failure within the network.

The FECIoT paradigm leverages on mobile fog devices, such as robots, drones, smart phones, smart watches, etc., to offer localized services to IoT end-devices. However, these devices are at risk of draining most of their energy when they move. As such, power-control mechanisms and smart mobility are critical in minimizing outages in communication, as well as optimizing energy utilization. This is crucial to drive several smart cities applications, most importantly the Industrial IoT (IIoT), where industrial robots are deployed 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 [4].

Several relay-based works have tried to address the issue of communication outages [4], [5], [9], energy usage [6], [8], [10], relay selection [4], [5], [7], [9], mobility [4], [7], latency [5], [10] and congestion [10] within the IoT ecosystem. However, these works considered centralized approaches which are prone to several challenges. Such challenges include scalability, failure or downtime in the central entity, overhead resulting from periodic updates and synchronization of nodes with the central controller often leading to inefficient energy utilization and decreased communication

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

TABLE I
CLASSIFICATIONS OF SOME PROBLEMS ADDRESSED IN RELATED WORKS

Reference	*Energy*	*Outage*	*Selection*	*Mobility*	*Latency*	*Traffic*	*Sources*	*Relays*	*Destination*	*Objective*	*Approach*
Omoniwa *et al.* [4]	-	✓	✓	✓	-	-	1	M	1	QoS	Steepest Descent method
Simiscuka *et al.* [5]	-	✓	✓	-	✓	-	M	M	M	QoS	Re-clustering Algorithm
Alsharoa *et al.* [6]	✓	-	-	-	-	-	M	M	1	Energy, Relay planning	Genetic Algorithm
Manzoor *et al.* [7]	-	-	✓	✓	-	-	M	1	1	Relay selection	Prototype design
Lv *et al.* [8]	✓	-	-	-	-	-	1	M	M	Energy	Numerical
Kawabata *et al.* [9]	-	✓	✓	-	-	-	M	M	M	QoS	Stochastic geometry
Behdad *et al.* [10]	✓	-	-	-	✓	✓	1	1	1	Energy, Latency, Conjestion	Analytical
Our approach	✓	✓	✓	✓	-	-	M	M	M	QoS, Energy	Decentralized Q-learning

performance within the network.

The closest RL-based works were [12] and [13], however these works focused on improving communication performance using the multi-armed bandit approach, which does not consider state transitions. On the other hand, our work model a realistic scenario of the defined problem and most importantly, applied a decentralized Q-learning approach with well-defined states, actions, and rewards.

The main contribution of our paper are summarized below:

1) To apply a decentralised reinforcement learning approach as in [11] to optimize the communication performance within a fog-based IoT architecture.
2) To optimize energy utilization within the network, taking into consideration the death of agents within the IoT network, hence making our decentralized approach robust to failure as compared to the centralized approaches.
3) Taking into account mobility of the fog relays, we propose an efficient RL-based selection strategy where an active fog relay is selected for a particular transmission phase from a set of available potential fog relays.

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 VI concludes the paper and outlines future directions.

II. RELATED WORKS

In Table I, we highlight some related works that address some of the key challenges within a relay-based IoT network. Several works have tried to address the issue of communication outages [4], [5], [9], energy usage [6], [8], [10], relay selection [4], [5], [7], [9], mobility [4], [7], latency [5], [10] and congestion [10] within the IoT ecosystem.

a) *Outage in communication*: An iterative algorithm based on the steepest descent method was proposed in [4] to optimize communication performance in a multi-tier fog-based IoT architecture, where a fog device acts as an amplify and forward relay of received information from an IoT sensor to a higher hierarchically-placed destination fog device, which offers some localized services. In [5], a relay and mobility scheme for IoT (REMOS-IoT) was proposed to improve the network QoS by effective resource management and decision-making. The centralized REMOS-IoT algorithm introduced also considered the mobility of fog gateways/relays in improving the network throughput without considering the efficient utilization of energy of these power-constrained IoT end-devices.
b) *Energy utilization*: An energy-efficient relaying scheme for IoT communications was presented in [6] to minimize energy consumption within an IoT network by solving the relay planning and QoS problems. A genetic algorithm-based approach was used to arrive at a sub-optimal low-complexity solution. Using numerical methods, the authors in [8] considered the energy-efficient design of a multi-pair decode-and-forward relay-based IoT network, in which multiple sources simultaneously transmit their information to the corresponding destinations via a relay equipped with a large array.
c) *Relay selection*: The authors in [7] designed a prototype, a mobile relay architecture for low-power IoT devices, which exploits third-party unknown mobile relays for the forwarding of medical data generated by BLE sensors to some central server in the cloud. In [9], a relay selection scheme for large-scale energy-harvesting IoT networks was proposed. The work applied a stochastic geometric approach and attempts to minimize the outage probability using a novel energy harvesting (EH) relay selection scheme.

However, most of these works considered centralized approaches, which may not be suitable within the IoT domain due to several challenges. These challenges are listed below.

1) A possible operational failure in the central controller may have drastic consequences on the entire network.
2) It takes time for the central entity to become aware of unavailable relay nodes due to mobility, death or departure of these nodes.
3) Managing energy usage of devices centrally is a complex task.
4) Signaling overhead between the central controller and network devices may result in increased energy cost within the IoT network.

The work in [12] proposed a decentralized stateless variation of Q-learning to improve aggregate throughput in four coexistent wireless networks (WN). The approach used in the work was a variant of a multi-armed bandit approach as only actions an rewards were defined, however, the Q-learning update function used in the work was modified, and excluded state transitions. Similarly, a MAB approach was presented in [13], (which is a single-state RL-based algorithm with no state transitions) where the agent only observes rewards based on actions taken. Both works may not be suitable to model a realistic IoT environment as defined in our paper, which takes into consideration mobility, death, energy utilization and communication performance in the system.

We present a decentralised reinforcement learning approach that adequately addresses some key performance issues within the IoT domain. The focus of our work is to minimize outage in communication in a fog-based IoT network as shown in Fig. 1, minimize energy cost by active fog relays, and present an efficient RL-based selection strategy where an active fog relay is selected for a particular transmission phase from a set of available potential mobile fog relays. Unlike the centralized approaches considered in [4] – [10], we present a decentralized autonomous system that is robust to failure, allowing agents to take independent actions and decisions. Furthermore, we applied the Q-learning algorithm on the defined problem. To the best of our knowledge, this is the first work that applies Q-learning to a fog-based IoT network.

III. System Model

In this section, we provide a full description of the system model, as well the RL approach used to address the problem. In Fig. 1, we show a network topology at time t where some randomly deployed IoT sensors have data/service request to send to remote destinations via some mobile fog relay agents (MFRA). We depict a realistic scenario where sensors may at some point in time be unable to reach the destination via all available MFRA due to long separation or obstructions between source and destination. We observe that source A can reach the local server via two MFRA, source B via a MFRA, and source C via three MFRA. As such, we try as much as possible to capture the features of the MFRA and its environment.

A. MFRA environment

States: The states are defined as a tuple, $\langle$Outage communication cost ($\mathcal{P}_{out}$) /Energy status of the fog relay (J) /Neighbour potential to relay message (Availability of redundant nodes)$\rangle$.

- *Outage communication cost*: Outage observations from the environment is estimated using (1) from [4], which gives an estimate of the communication outage when the agent takes an action, such as changing power levels or location, or both.
- *Energy expended by fog relay*: This observation gives the agent insight on how much energy by the fog agent when following policy $\omega_i \in \omega_{fog}$. If the fog agent continues to take sub-optimal actions, it depletes its energy and dies out.
- *Neighbour potential to relay message (Availability of redundant nodes)*: This observation gives the agent insight on the availability of redundant nodes that can help in relaying same type of message emanating from a particular IoT sensor. If there exist no potential relay agent to convey message from an IoT sensor to a remote destination, then the agent should learn to remain active for that transmission phase. However, if there exist one or more potential relays agents, the agent should learn to take no action to help conserve energy and improve the longevity of the network.

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

where $\Psi = \sqrt{(N_0\tilde{\kappa})/(P_R(D_S+\delta)^{-\sigma})}$, and $\mathcal{P}_{out}$ is an expression for the outage probability with values between 0 and 1. We assume a predefined threshold $\tilde{\kappa}$ 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, $\delta = \pm 0.25m$, N_0 to be the channel noise, and σ to be the path-loss exponent.

B. MRFA agent

We apply a the Q-Learning algorithm, an RL approach which requires no prior knowledge of the environment by the agent. In Q-learning, the agent interacts with the environment over periods of time according to a policy ω. At every time-step $k \in N$, the environment produces an observation $s_k \in \mathbb{R}^{D_s}$. By sampling, the agent then picks an action a_k over $\omega(s_k)$, $a_k \in \mathbb{R}^{D_a}$, which is applied to the environment. The environment consequently produces a reward $r(s_k, a_k)$ and may end the episode at state s_N or transits to a new state s_{k+1}. The agent's goal is to minimize the expected cumulative cost, $\min_\omega \mathbb{E}_{s_0,a_0,s_1,a_1,\ldots,s_N}\Big[\sum_{i=0}^{N} \gamma^i \mathcal{C}(s_i)\Big]$, where $0 \leq \gamma \leq 1$ is the discount factor, and $\mathcal{C}$ is the overall cost function of our model.

First, the agent takes an initial random action a_k and gets observations from the environment which corresponds to that action, as well as a reward. It then discretizes the continuous observations emanating from the environment into a $3 \times 3 \times 3$ state space corresponding to the tuple, $\langle$Outage communication cost ($\mathcal{P}_{out}$) /Energy status of the fog relay (J) /Neighbour potential to relay message (Availability of redundant nodes)$\rangle$. The agent then updates it's Q-values at each time-step k following (2).

$$\begin{aligned} Q(s_k, a_k) &:= Q(s_k, a_k) \\ &+ \alpha\Big[r_{k+1} + \gamma \max_a Q(s_{k+1}, a) - Q(s_k, a_k)\Big], \end{aligned} \quad (2)$$

where α is the learning rate, which determines the impact of new experience on the Q-value, r_{k+1} is the reward the agent receives by being in s_{k+1} from s_k. Based on the policy followed by the agent, it gets observations and rewards from the environment. The action space, goal, reward and performance metrics considered in this work are given below.

- *Action space*: The actions for the fog relay agent are move closer and transmit, move farther and transmit, and do nothing (become passive).
- *Goal*: The goal of the agent first, is to be alive during the transmission phase and relay message received from source to destination at a reasonably QoS by ensuring that the packets received in each transmission phase do not fall below some pre-defined threshold, which was set at 95%, and endeavour to be active when there exist no potential relays to convey same message from IoT sensor to remote destination.
- *Reward*: The reward function used is given in (3) as

$$R = \begin{cases} 100, & \text{if } goal == Reached \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

- *Metrics*: Outage probability, i.e. the ratio of the number of packet lost to those transmitted, which we measure in percentages, and energy status of the fog relay agent, i.e. the ratio of depleted energy to the initial capacity in Joules, which we measure as a percentage.

The MFRA's learning process is summarized in Algorithm 1. A new learning episode is terminated when the agent attains the pre-defined goal of minimizing the communication outage in the link, or when either the fog relay agent die out, which may be due to taking sub-optimal actions without getting to the goal. When a fog relay agent decides to do nothing, it should imply that there exist other available agents more capable of transmitting during that transmision phase, as such the agent learns to conserve energy. On the contrary, if the fog relay agent continues to move and transmit even when there exist sufficient redundancy to relay same message, it may die out soon, hereby causing a point-of-failure to the network. In this work, an episode is completed either when the agent reaches its goal, when the agent dies, or when the maximum step for an episode is reached. The reward is updated in the Q-learning table, with environmental information updated as well.

Algorithm 1 MFRA Learning Process

Initialize: $\delta = \pm 0.25m$ and mobility range (m) = [-30, 30]
top:
ResetEnvironment()
state ← MapLocalObservationToState(*env*)
action ← QLearning.SelectAction(*state*)
if *action* == "move close and Tx"||"move away and Tx" **then**
    Env.EstimateOutage using (1)
    Env.EstimateFogEnergyUsed
    Env.CheckAvailableActiveFogNeighbour
else if *action* == "do nothing" **then**
    100% ← Env.Outage
    0J ← Env.FogEnergyUsed
    Env.CheckAvailableActiveFogNeighbour
endif
InvokePolicy(ExponentialDecay)
UpdateQLearningProcedure() (2)
CurrentState ← NewState
if *goal* == "Reached" **then return** Reward = 100,
else if *goal* != "Reached" or Agent == "Death" **then return** Reward = 0
endif
EndEpisode **goto** *top*.

IV. Experimental Setup

Experimentation was carried out using the Python IDE 3.7.2, and the performance metrics considered in evaluating our proposed fog-based IoT network are: (i) packets delivered, and (ii) energy consumed by fog relay agents. The packets delivered is defined as the ratio of received packets at the destination node to that transmitted by the IoT sensor via some fog-relay agents, while the energy consumed by fog relay agents is defined as ratio of the energy drained by the fog devices in Joules to the initial capacity fog devices.

A. Baselines

In this work, we consider three existing baselines, namely:

1) Rule-based + centralised selection (RB-CS): The central controller selects a potential MFRA that satisfies the highest QoS requirement to become an active MFRA for that transmission phase.
2) Rule-based + randomized selection (RB-R): Active MFRAs are randomly selected from a set of potential MFRAs that are within the neighbourhood of the IoT sensor and the destination device to forward traffic.
3) Rule-based + round-robin selection (RB-RR): Active MFRAs are chosen in a round-robin manner from a set of potential MFRAs that are within the neighbourhood of the IoT sensor and the destination device.

B. Indicators

To evaluate our proposed approach, we examine the convergence of our approach in minimizing the steps it takes the agent to get to its goal, the energy utilization of the active MFRA in the IoT network.

V. Results and Discussions

In this section, we present the results of our fog-based IoT system. Table II shows a summarises the parameters used in the experiments. For the sake of evaluation, we considered the last forty episodes to ensure convergence in the learning process. We compare the proposed approach with some baseline by evaluating the percentage of packets successfully transmitted via potential fog relay agents, then we examine energy utilization of our proposed approach with the baseline. We compared our proposed RL approach with existing baselines.

A. Communication performance for 2 MFRAs

We measure the communication performance in terms of the ratio of packets delivered to that transmitted. Fig. 3 (a)

TABLE II
SIMULATION PARAMETERS

Parameter	*Values*
Simulation space	80×80 metres
P_I	[0.001, 0.3] Watts
P_R	0.3 Watts
δ	± 0.25 metres
MFRA mobility bound	[-30, 30] metres
Noise power N_0	2×10^{-7} Watts
Path-loss exponent σ	3
Pre-defined threshold κ	1
Discount factor γ	0.9
Learning rate α	0.1
Episodes N	100
Maximum iteration runs	100000
Policy ϵ	$e^{-0.0015N}$

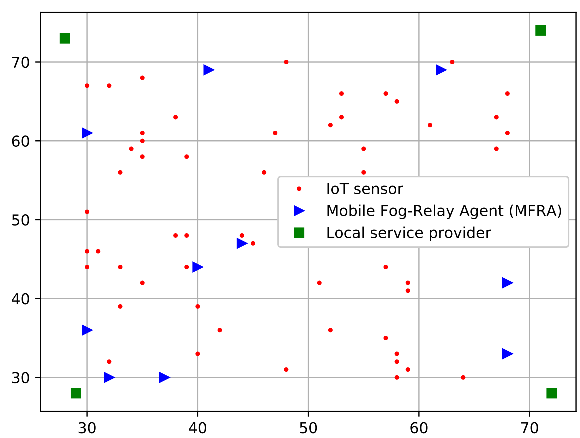

Fig. 2. Simulation instance using 60 IoT sensors, 10 MFRAs, and 4 local service providers.

shows the percentage of packets successfully transmitted by the active fog relay agents when conveying messages from IoT sensors to a remote destination. We observe that about 95% packets in our RL proposed approach is successfully transmitted. The RB-CS strategy perform closely to the RL approach with values ranging between 93.9% - 95% of packets successfully transmitted and a median of about 94.9%. The good performance is based on the assumption that a central controller selects the best performing agent in each transmission phase. However, the RB-CS strategy performs poorly in few instances, which may be due to the several reasons, such as, the controller not taking into account death or departure of selected agent.

We observe higher variability in the RB-R and RB-RR strategies, with the percentage of successfully transmitted packets ranging between 89.8% - 94.5% and 89.3% - 94.6%, and median of 92.2% and 92.3%, respectively. In general, we observe variability mostly in the baselines as compared to our proposed approach.

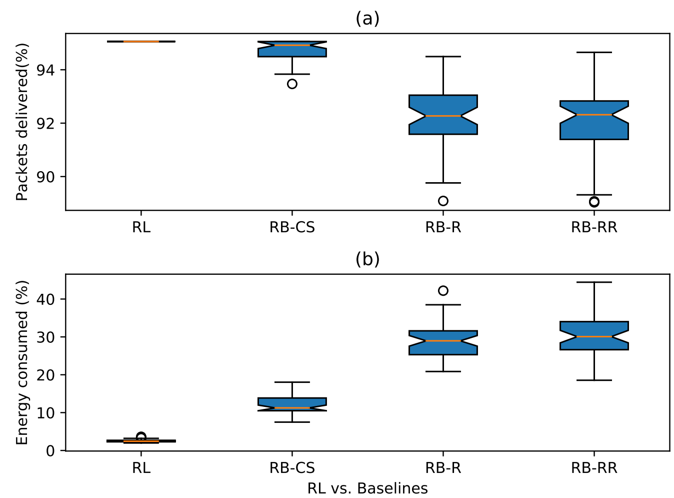

Fig. 3. Comparison of RL with baselines over 50 runs in a 2 agents environment (a) Percentage of packets successfully transmitted by the active fog relay agents, (b) Percentage of energy consumed by the active fog relay agents.

B. Energy utilization for 2 MFRAs

Fig. 3 (b) depicts the percentage of energy consumed by the active fog relay agents in the network. Our proposed approach out-performs the baseline by efficiently minimizing energy cost within the network. The RB-CS strategy was able to achieve a range between 7.5% - 17.8%, and a median of 11.2%, performing better than the RB-R and RB-RR strategies, which achieved a range between 20.9% - 38.6% and 18.6% - 44.6%, and median of 28.9% and 30.0%, respectively. The proposed RL approach has the least variation with a median of about 2.3%. This shows that the decentralized learning by agents within the system significantly improves its energy utilization. The agents can learn when not to be active in order to minimize energy cost. Overall, the performance of the proposed approach yields better results than the baseline.

C. Communication performance for 5 MFRAs

In Fig. 4 (a), we show the percentage of packets successfully transmitted by the active fog relay agents when conveying messages from IoT sensors to a remote destination. We observe that both our RL approach and RB-CS successfully transmitted about 95% packets. The improved performance of RB-CS strategy may be due to the increase in redundant relays within the network. However, both the RB-R and RB-RR strategies perform poorly with the percentage of successfully transmitted packets ranging between 90.4% - 94.8% and 89.0% - 94.0%, and median of 92.3% and 91.8%, respectively. Furthermore, we observe higher variability in the RB-R and RB-RR strategies, as compared to the RL approach and the RB-CS strategy.

D. Energy utilization for 5 MFRAs

Fig. 4 (b) shows the percentage of energy consumed by the active fog relay agents in a 5 MFRA-based IoT network. Our proposed approach out-performs the baseline by efficiently

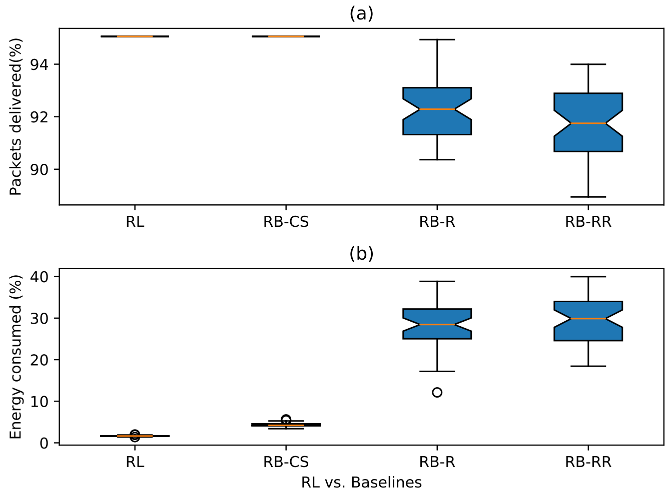

Fig. 4. Comparison of RL with baselines over 50 runs in a 5 agents environment (a) Percentage of packets successfully transmitted by the active fog relay agents, (b) Percentage of energy consumed by the active fog relay agents.

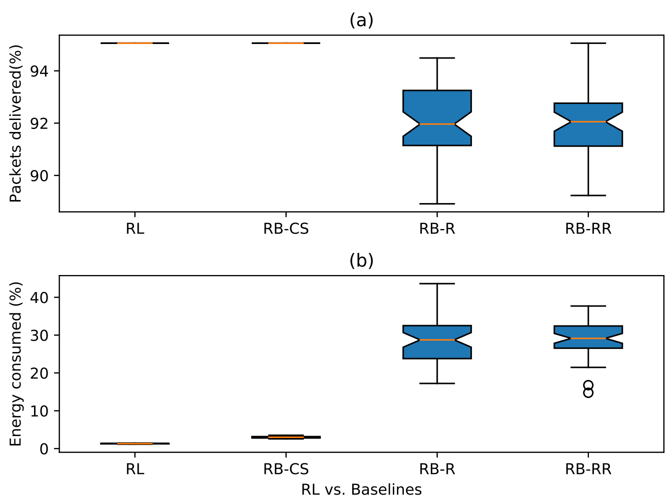

Fig. 5. Comparison of RL with baselines over 50 runs in a 10 agents environment (a) Percentage of packets successfully transmitted by the active fog relay agents, (b) Percentage of energy consumed by the active fog relay agents.

minimizing energy cost within the network. This is due to the fact that the agent learns to enter a passive mode and do nothing when there exist potential MFRA that can tranmit same message at a better QoS. However, the RB-CS strategy was able to achieve a good energy utilization with a median of 5.3%, performing better than the RB-R and RB-RR strategies, ranging between 17.3% - 39.0% and 18.8% - 40.4%, and median of 28.9% and 28.7%, respectively. The proposed RL approach has the least variation with a median of about 1.7%. This shows that the decentralized learning by 5 agents significantly improves its energy utilization. The agents can learn when not to be active in order to minimize energy cost. Overall, the performance of the proposed approach yields better results than the baseline.

E. Communication performance for 10 MFRAs

In Fig. 5 (a), we show the percentage of packets successfully transmitted by the active fog relay agents when conveying messages from IoT sensors to a remote destination. We observe that both our RL approach and RB-CS successfully transmitted about 95% packets. The improved performance of RB-CS strategy may be due to the increase in redundant relays within the network as in the 10 MFRA network. However, both the RB-R and RB-RR strategies perform poorly with the percentage of successfully transmitted packets ranging between 89.9% - 94.5% and 89.2% - 95%, and median of 91.9% and 92.1%, respectively. Furthermore, we observe higher variability in the RB-R and RB-RR strategies, as compared to the RL approach and the RB-CS strategy, which converged.

F. Energy utilization for 10 MFRAs

Fig. 5 (b) shows the percentage of energy consumed by the active fog relay agents in a 10 MFRA-based IoT network. Our proposed approach out-performs the baseline by efficiently minimizing energy cost within the network. This is due to the fact that the agent learns to enter a passive mode and do nothing when there exist potential MFRA that can tranmit same message at a better QoS. However, the RB-CS strategy was able to achieve a good energy utilization with a median of 3.5%, performing better than the RB-R and RB-RR strategies, ranging between 17.2% - 43.7% and 21.5% - 38.0%, and median of 28.9% and 29.2%, respectively. The proposed RL approach has the least variation with a median of about 1.5%. This shows that the decentralized learning by 10 agents significantly improves its energy utilization. The agents can learn when not to be active in order to minimize energy cost. Overall, the performance of the proposed approach yields better results than the baseline.

VI. Conclusion and Future works

In this paper, we apply a decentralized q-learning algorithm on a multi-agent fog-based IoT system where multiple MFRAs compete to relay information in a highly dynamic environment, where long source to destination distances and obstacles considerably contribute to communication outages within the network. Furthermore, in an attempt to improve the communication performance, the MFRAs change positions which increases energy cost. As such, we carry out a comparative study and perform 50 experiments which show that our RL-based approach outperforms the three baselines (both in QoS and energy utilization) with lesser variations. Moreover, there was significant improvement in communication performance and energy utilizations as more MFRAs were deployed, which can be justified by addition of redundant agents to improve overall QoS. Intuitively, for the other approaches to match the performance of our proposed approach, more MFRAs must be deployed in the network. Overall, the performance of the proposed approach yields better results than the baselines. Our future work will take into consideration fairness and latency within the fog-based IoT system.

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] A. BenMimoune and M. Kadoch, "Relay Technology for 5G Networks and IoT Applications," Internet of Things: Novel Advances and Envisioned Applications, vol. 25, pp. 3-26, April 2017.

[3] 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.

[4] 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, vol. 6, no. 2, pp. 3044-3054, April 2019.

[5] A. A. Simiscuka and G. Muntean, "A Relay and Mobility Scheme for QoS Improvement in IoT Communications," 2018 IEEE International Conference on Communications Workshops (ICC Workshops), Kansas City, MO, 2018, pp. 1-6.

[6] A. Alsharoa, X. Zhang, D. Qiao and A. Kamal, "An Energy-Efficient Relaying Scheme for Internet of Things Communications," 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, 2018, pp. 1-6.

[7] A. Manzoor, P. Porambage, M. Liyanage, M. Ylianttila and A. Gurtov, "DEMO: Mobile Relay Architecture for Low-Power IoT Devices," 2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), Chania, 2018, pp. 14-16.

[8] T. Lv, Z. Lin, P. Huang and J. Zeng, "Optimization of the Energy-Efficient Relay-Based Massive IoT Network," in IEEE Internet of Things Journal, vol. 5, no. 4, pp. 3043-3058, Aug. 2018.

[9] H. Kawabata, K. Ishibashi, S. Vuppala and G. T. F. de Abreu, "Robust Relay Selection for Large-Scale Energy-Harvesting IoT Networks," in IEEE Internet of Things Journal, vol. 4, no. 2, pp. 384-392, April 2017.

[10] Z. Behdad, M. Mahdavi and N. Razmi, "A New Relay Policy in RF Energy Harvesting for IoT Networks—A Cooperative Network Approach," in IEEE Internet of Things Journal, vol. 5, no. 4, pp. 2715-2728, Aug. 2018.

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

[12] 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.

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