International Journal of Mobile Computing and Multimedia Communications
Volume 14 • Issue 1

Epsilon-Greedy-Based MQTT QoS Mode Selection and Power Control Algorithm for Power Distribution IoT

Xinhong You, Shandong Electric Power Research Institute, State Grid Corporation of China, China*

Pengping Zhang, Shandong Electric Power Research Institute, State Grid Corporation of China, China

Minglin Liu, Shandong Electric Power Research Institute, State Grid Corporation of China, China

Lingqi Lin, Shandong Electric Power Research Institute, State Grid Corporation of China, China

Shuai Li, Shandong Electric Power Research Institute, State Grid Corporation of China, China

ABSTRACT

Employing message queuing telemetry transport (MQTT) in the power distribution internet of things (PD-IoT) can meet the demands of reliable data transmission while significantly reducing energy consumption through the dynamic and flexible selection of three different quality of service (QoS) modes and power control. However, there are still some challenges, including incomplete information, coupling of optimization variables, and dynamic tradeoff between packet-loss ratio and energy consumption. In this paper, the authors propose a joint optimization algorithm named EMMA for MQTT QoS mode selection and power control based on the epsilon-greedy algorithm. Firstly, the joint optimization problem of MQTT QoS mode selection and power control is modeled as a multi-armed bandit (MAB) problem. Secondly, the authors leverage the online learning capability of the epsilon-greedy algorithm to achieve joint optimization of MQTT QoS mode selection and power control. Finally, they verify the superior performance of the proposed algorithm through simulations.

KEYWORDS

Energy Consumption, Epsilon-Greedy Algorithm, Message Queuing Telemetry Transport, Multi-Armed Bandit, Packet-Loss Ratio, Power Control, Power Distribution Internet of Things, QoS Mode Selection

1. INTRODUCTION

1.1 Background

Power distribution internet of things (PD-IoT) realizes comprehensive perception, data fusion, and intelligent applications of the distribution network through the interconnection and intercommunication between power distribution network devices (Lv J. *et al.*, 2018; Liu J. *et al.*, 2018; Liu Z., 2016). With the rapid development of IoT technology in the distribution network and the significantly increasing number of PD-IoT devices, the traditional request/response mechanism is no longer suitable (Kim G. *et al.*, 2019; Stankovic J. A., 2020). Moreover, due to the limited battery capacity and computation

DOI: 10.4018/IJMCMC.306976
*Corresponding Author

resources of PD-IoT devices, it is difficult to meet the quality of service (QoS) requirements of the low packet-loss ratio and low energy consumption in PD-IoT (Zhou Z. *et al.*, 2020; Mumtaz S. *et al.*, 2017; Ding Z. *et al.*, 2019; Liao H. *et al.*, 2020; Sadio O. *et al.*, 2019). Therefore, how to ensure the QoS requirements of PD-IoT devices with limited resources is a difficult problem.

Message queue telemetry transmission (MQTT) provides an effective solution as an IoT transmission protocol based on publish/subscribe mechanism, which has the characteristics of simple implementation, lightweight, and high bandwidth utilization (Sadio O. *et al.*, 2019; Herrero, R. *et al.*, 2019). Specifically, it provides three flexible QoS modes to achieve the reliable transmission of different task data, i.e., “at most once QoS0”, “at least once QoS1”, and “exactly once QoS2”, which are introduced as follows:

At most once QoS0: Named as “QoS0 mode”, sender sends a PUBLISH data packet containing the message to the receiver, and each PUBLISH data packet is sent only once, regardless of whether the receiver successfully receives the PUBLISH data packet. Therefore, although the energy consumption of QoS0 mode is low, the high packet-loss ratio is caused when the channel state is poor or the power control is lacking.

At least once QoS1: Named as “QoS1 mode”. Each PUBLISH data packet is guaranteed to be received successfully at least once. If the feedback confirmation data packet, i.e., the PUBACK data packet is not received by the sender within a period of time, the PUBLISH data packet will be retransmitted. The data deduplication energy consumption is introduced while the packet-loss ratio is zero.

Exactly once QoS2: Named as “QoS2 mode”. The QoS2 mode ensures that each PUBLISH data packet is only successfully received by the receiver once through two interaction processes. If a data packet with the same information header as the previously stored information header is received, the receiver will treat it as a duplicate message and discard it. Therefore, the energy consumption of deduplication is avoided at the cost of increasing the energy consumption of transmission considering the two interaction processes in the QoS2 mode.

In summary, different QoS modes have different performance in terms of packet-loss ratio and energy consumption. In addition, power control also affects the packet-loss ratio and energy consumption simultaneously by adjusting the allocated transmission power. Therefore, in order to achieve a dynamic tradeoff between packet-loss ratio and energy consumption, it is intuitive to jointly optimize the MQTT QoS mode selection and power control.

1.2 Challenge

However, there are still some key technical challenges in the joint optimization of MQTT QoS mode selection and power control. Firstly, the global state information, e.g., channel state information and bandwidth information, is required for the joint optimization problem, which is unavailable for PD-IoT devices in the actual implementation environment considering the prohibitive signaling overhead. Secondly, the optimization problem is a non-deterministic polynomial (NP)-hard problem due to the coupling between MQTT QoS mode selection and power control. Specifically, a larger transmission power is required to ensure the successive transmission when the QoS0 mode is selected. Moreover, this exacerbates the complexity of the optimization problem. Last but not least, it is unrealistic to achieve the lowest energy consumption and the lowest packet-loss ratio at the same time. For example, when the QoS2 mode is selected, although a low packet-loss ratio is achieved, multiple transmissions of task data packet result in higher energy consumption. If the QoS0 mode is selected, although the energy consumption is lower, the packet-loss ratio increase. Therefore, how to achieve a dynamic tradeoff between packet-loss ratio and energy consumption through the joint optimization of MQTT QoS mode selection and power control under incomplete information to meet the different QoS requirements of PD-IoT services is still an open issue.

1.3 Contribution

To address the above challenges, in this paper, we propose an epsilon-greedy-based MQTT QoS mode selection and power control algorithm named EMMA for PD-IoT. Firstly, the joint optimization problem of MQTT QoS mode selection and power control is modeled as a multi-armed bandit (MAB) problem, the optimization objective of which is to minimize the weighted sum of packet-loss ratio and energy consumption. Secondly, based on the historical information and local information at the device side, the devices leverage the online learning capability of the epsilon-greedy algorithm to achieve joint optimization of MQTT QoS mode selection and power control. The main contributions of this paper are summarized as follows:

- **Joint optimization of MQTT QoS mode selection and power control under incomplete information:** An MQTT QoS mode selection and power control joint optimization variable is defined. PD-IoT devices learn the historical information and make decisions based on local side information to realize the optimal MQTT QoS mode selection and power control by utilizing the epsilon-greedy algorithm under incomplete information.
- **Tradeoff between packet-loss ratio and energy consumption:** We set the optimization objective as minimizing the weighted sum of packet-loss ratio and energy consumption. Specifically, a nonnegative weight parameter is employed to reflect the relative importance of packet-loss ratio and energy consumption. Therefore, by adjusting the value of the nonnegative weight parameter, we can achieve a dynamic tradeoff between packet-loss ratio and energy consumption.
- **Extensive performance evaluation:** We verify the superior performance of the proposed algorithm through various simulations. The simulation results elaborate that, compared with the existing algorithms, the proposed algorithm can achieve a dynamic tradeoff between packet-loss ratio and energy consumption.

The rest of the paper is organized as follows. Section 2 describes the related work. Section 3 presents the system model and problem formulation. The proposed algorithm is introduced in Section 4. Section 5 provides simulation results. Finally, we conclude in Section 6.

2. RELATED WORKS

There has been some works employing the MQTT protocol in IoT. Specifically, in (Toldinas J., *et al.*, 2017), Todinas *et al.* studied and analyzed the energy consumption of data transmission with different QoS modes of the MQTT protocol. Firstly, the authors obtained the real-time energy consumption values of the MQTT protocol at different QoS modes. Then, the authors proposed an energy consumption model based on the MQTT protocol, which can choose optimal QoS mode by considering reliability, latency, and battery life, and was successfully applied to IoT systems. In (Kim S. J., *et al.*, 2017), Kim *et al.* proposed a centralized control system, which makes the adaptive QoS mode selection decision by analyzing the transmission status of the device and the message received by the receiver. Firstly, the authors analyzed the traffic information transmitted and received between publishers and MQTT brokers based on the considered IoT scenarios. Secondly, based on the information obtained from the above analysis, the authors proposed a control system for adaptive QoS mode selection, which controls the QoS mode of publishers according to information such as traffic changes, and ultimately reduces network energy consumption. In (Jo H. & Jin H., 2015), Jo *et al.* proposed an adaptive framework for periodic N-to-1 communication over MQTT to effectively reduce the time delay. Firstly, the framework completes a closed adaptive formula modeling based on different QoS levels. Secondly, the framework can adaptively adjust the cycle time to minimize the cycle delay by considering factors such as the transmission control protocol (TCP) connection between the publisher and the subscriber. However, the aforementioned works cannot be directly applied to the considered PD-IoT scenario without considering the incomplete information, the joint optimization of the MQTT QoS mode selection and power control, and the dynamic tradeoff between packet-loss ratio and energy consumption.

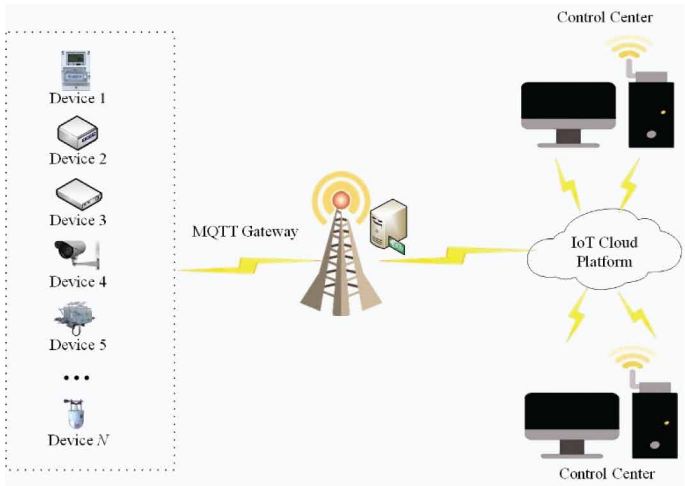
MAB is usually used to deal with a series of multi-stage decision-making problems under incomplete information scenarios (Liao H., *et al.*, 2020). Among many MAB algorithms, the epsilon-greedy algorithm stands out because of its good tradeoff property between exploration and exploitation. There has been a lot of researches utilizing the epsilon-greedy algorithm in the resource allocation of wireless networks (Adeogun R., *et al.*, 2020; Kiran N., *et al.*, 2020; Oksanen J., *et al.*, 2010). In (Adeogun R., *et al.*, 2020), Adeogun *et al.* proposed a channel selection algorithm based on the epsilon-greedy algorithm to minimize the delay. Firstly, the authors studied how to solve the channel selection problem in a distributed scheme for a 6G internal mobile industrial subnet scenario. Secondly, the authors proposed an interference graph through the network topology, which further improves the accuracy of channel allocation. Finally, based on the measured industrial scene parameters, the author verified that the superior performance of proposed algorithm in the network delay. In (Kiran N., *et al.*, 2020), Kiran *et al.* focused on how to realize joint optimization of task offloading and resource allocation between multiple objectives in a distributed way for wireless network. Firstly, the authors proposed a reinforcement learning optimization framework based on software defined network (SDN) and mobile edge computing (MEC). Secondly, a reinforcement learning algorithm based on the epsilon-greedy was proposed to realize the joint optimization of task offloading and resource allocation, which guaranteed the extreme latency and reliability requirements. In (Oksanen J., *et al.*, 2010), Oksanen *et al.* proposed a spectrum sensing and allocation algorithm based on the epsilon-greedy, which can improve the energy efficiency of the system and avoid conflicts between users. Firstly, the authors observed the occupancy of the primary user and estimated the performance of detection for this user based on the epsilon -greedy method. Secondly, an epsilon-greedy-based spectrum sensing and allocation algorithm was proposed to achieve optimal resource allocation by avoiding conflicts between different users, and ultimately maximize the final network throughput and energy efficiency. However, the aforementioned schemes are not suitable for the scenarios considered in this paper because the MQTT protocol and the typical scenarios of PD-IoT are not considered.

3. SYSTEM MODEL

As shown in Figure 1, a PD-IoT scenario is considered, which includes N PD-IoT devices, one base station equipped with an MQTT gateway, one IoT cloud platform, and PD-IoT application control centers. The specific flow of communication is introduced as follows. Firstly, a PD-IoT device sends a communication request to the gateway based on the MQTT protocol, and the gateway forwards relevant configuration information to the PD-IoT device. Secondly, the PD-IoT device sends task data to the IoT cloud platform for processing based on the MQTT protocol. Finally, the IoT cloud platform returns the task processing result to the PD-IoT application control center. In this paper, we focus on the joint optimization of the MQTT QoS mode selection and power control during the process of when PD-IoT devices send task data to the IoT cloud platform.

We assume that there are a total of I large data packets at the PD-IoT device, which are indexed by $\mathcal{I} = \{1,2,\ldots,i,\ldots,I\}$. Each large data packet contains J small packets of equal size, which are indexed by $\mathcal{J} = \{1,2,\ldots,j,\ldots,J\}$. We assume that when the data transmission of each large packet starts, the MQTT QoS mode selection and power control policy is decided at the PD-IoT device, which remains unchanged until a large data packet transmission is completed. We define $m(i) \in \left\{0,1,2\right\}$ as the MQTT QoS mode selection variable of the i-th large packet, where $m(i) = 0$ means selecting the QoS0 mode, $m(i) = 1$ means selecting the QoS1 mode, and $m(i) = 2$ means selecting the QoS2 mode. We define $p(i)$ as the transmission power of the i-th large packet. Specifically, we adopt a

Figure 1. System model



discrete power control model, where the set of K power control modes is defined as $p(i) \in \left\{p_1, p_2, \cdots, p_k, \cdots, p_K\right\}$.

The channel gain of the n-th transmission of the j-th small packet of the i-th large packet is given by:

$$g_{i,j,n} = \frac{\left|H_{i,j,n}\right|^2}{n_0} \tag{1}$$

where $H_{i,j,n}$ represents the channel frequency response at the n-th transmission of the j-th small packet of the i-th large packet, and n_0 is the noise power. In particular, we assume that during any small packet task data transmission process, the channel state remains unchanged. Since QoS0 mode does not involve retransmission, we define $n = 0$ in the QoS0 mode.

Considering the differences among the three QoS modes, the models of the packet-loss ratio and energy consumption corresponding to the QoS0 mode, the QoS1 mode, and the QoS2 mode are introduced as follows.

3.1 The QoS0 Mode

3.1.1 Packet-Loss Ratio Model

We define $a_{i,j}^{0,p(i)}$ as the packet-loss indicator variable of the j-th small packet of the i-th large packet in the QoS0 mode. Particularly, $a_{i,j}^{0,p(i)} = 1$ means that the j-th small packet of the i-th large packet is lost, and $a_{i,j}^{0,p(i)} = 0$ otherwise. If the current signal-to-noise ratio is lower than the threshold, it means packet loss has occurred. Therefore, the expression is given by:

$$a_{i,j}^{0,p(i)} = \begin{cases} 1, & p(i) g_{i,j,1} < G_{th} \\ 0, & otherwise \end{cases} \tag{2}$$

where G_{th} is the signal-to-noise ratio threshold required for successful task data transmission. Therefore, in the QoS0 mode, the packet-loss ratio of the i-th large packet is given by:

$$Q_i^{0,p(i)} = \frac{a_i^{0,p(i)}}{J} \tag{3}$$

where $a_i^{0,p(i)}$ represents the total number of small packets lost in the i-th large packet, which is given by:

$$a_i^{0,p(i)} = \sum_{j=1}^{J} a_{i,j}^{0,p(i)} \tag{4}$$

3.1.2 Energy Consumption Model

Since the QoS0 mode does not include the energy consumption of deduplication, the total energy consumption of the i-th large packet equals to the transmission energy consumption, which is given by:

$$E_i^{0,p(i)} = \sum_{j=1}^{J} \frac{S}{B\log_2\left(1 + p(i)g_{i,j,1}\right)} p(i) \tag{5}$$

where S represents the data size of each small packet, and B represents the bandwidth. Each PUBLISH data packet will not be retransmitted even if the transmission fails.

3.2 The QoS1 Mode

3.2.1 Packet-Loss Ratio Model

Since the QoS1 mode adopts the retransmission to ensure successful task data transmission, the packet-loss ratio of the i-th large packet is zero, which is given by:

$$Q_i^{0,p(i)} = 0 \tag{6}$$

3.2.2 Energy Consumption Model

In the QoS1 mode, $a_{i,j,n}^{1,p(i)}$ is defined as the transmission indicator variable of the PUBLISH data packet of the j-th small packet of the i-th large packet, where $a_{i,j,n}^{1,p(i)} = 1$ indicates that the n-th transmission of the PUBLISH data packet fails, and $a_{i,j,n}^{1,p(i)} = 0$ otherwise. If the current signal-to-noise ratio is lower than the threshold, it means packet loss has occurred. Therefore, the expression is given by:

$$a_{i,j,n}^{1,p(i)} = \begin{cases} 1, & p(i)g_{i,j,n} < G_{th} \\ 0, & p(i)g_{i,j,n} \geq G_{th} \end{cases} \tag{7}$$

$b_{i,j,n}^{1,p(i)}$ is defined as the transmission indicator variable of the PUBACK data packet, where $b_{i,j,n}^{1,p(i)} = 1$ indicates that the n-th transmission of the PUBACK data packet fails, and $b_{i,j,n}^{1,p(i)} = 0$ otherwise. $b_{i,j,n}^{1,p(i)}$ is given by:

$$b_{i,j,n}^{1,p(i)} = \begin{cases} 1, & p(i) g_{i,j,n} \geq G_{th} \text{ and } p_{back} g_{i,j,n,back} < G_{th} \\ 0, & p(i) g_{i,j,n} \geq G_{th} \text{ and } p_{back} g_{i,j,n,back} \geq G_{th} \end{cases} \tag{8}$$

where p_{back} represents the transmission power of the data packets transmitted from the IoT cloud platform to the PD-IoT device, i.e., the PUBACK data packet in the QoS1 mode, and the PUBREC data packet as well as the PUBCOMP data packet in the QoS2 mode. $g_{i,j,n,back}$ represents the channel gain when the PUBACK data packet is transmitted from IoT cloud platform to the PD-IoT device. There are two reasons for the failure of the PUBLISH data packet transmission. One reason is the loss of the PUBLISH data packet caused by the poor channel gain, the other reason is that although the PUBLISH data packet is successfully sent, the PUBACK data packet is lost during the transmission. Under this condition, the sender needs to retransmit the PUBLISH data packet until the PUBACK packet is successfully transmitted back.

In the QoS1 mode, the transmission energy consumption of the j-th small packet of the i-th large packet is given by:

$$\begin{aligned} E_{i,j}^{1,p(i),TX} = & \sum_{n=1}^{N_{i,j}} a_{i,j,n}^{1,p(i)} \frac{S}{B \log_2\left(1 + p(i) g_{i,j,n}\right)} p(i) \\ & + \sum_{n=1}^{N_{i,j}} b_{i,j,n}^{1,p(i)} \left(\frac{S}{B \log_2\left(1 + p(i) g_{i,j,n}\right)} p(i) + \frac{S_{BACK}}{B \log_2\left(1 + p_{back} g_{i,j,n,back}\right)} p_{back} \right) \\ & + \frac{S}{B \log_2\left(1 + p(i) g_{i,j,N_{i,j}}\right)} p(i) + \frac{S_{BACK}}{B \log_2\left(1 + p_{back} g_{i,j,N_{i,j},back}\right)} p_{back} \end{aligned} \tag{9}$$

where $N_{i,j}+1$ represents the total number of transmissions of the j-th small packet of the i-th large packet, and S_{BACK} represents the size of the PUBACK data packet. The first term of (9) represents the transmission energy consumption in the case of PUBLISH packet transmission failure. The second term of (9) represents the PUBLISH data packet transmission energy consumption caused by the PUBACK data packet feedback failure and the transmission energy consumption of the PUBACK data packet. The third and fourth terms of (9) indicate the transmission energy consumption when the PUBLISH data packet is successfully transmitted and the PUBACK packet is also fed back successfully. The transmission of the PUBLISH data packet is completed when both the PUBLISH data packet and the PUBACK data packet are transmitted successfully.

In the QoS1 mode, the IoT cloud platform may receive duplicate PUBLISH data packets and need to perform deduplication. Repeated transmission of the PUBLISH data packets is caused by the numbers of successful PUBLISH data packet transmissions and PUBCOMP data packet feedback failures, and the deduplication energy consumption is given by:

$$E_i^{1,O} = \sum_{j=1}^{J} \sum_{n=1}^{N_{i,j}} b_{i,j,n}^{1} E_0 \tag{10}$$

where E_0 is the deduplication energy consumption of single PUBLISH data packet. The deduplication energy consumption of any small packet is uniformly defined as E_0.

Hence, the total energy consumption in the QoS1 mode is given by:

$$E_i^{1,p(i)} = \sum_{j=1}^{J} E_{i,j}^{1,p(i),TX} + E_i^{1,O} \tag{11}$$

3.3 The QoS2 Mode

3.3.1 Packet-Loss Ratio Model

Similar to the QoS1 mode, the QoS2 mode also adopts the retransmission to ensure successful task data transmission, the packet-loss ratio of the i -th large packet is zero, which is given by:

$$Q_i^{0,p(i)} = 0 \tag{12}$$

3.3.2 Energy Consumption Model

The QoS2 mode includes the following processes, i.e., the transmission of the PUBLISH data packet, the PUBREC data packet, the PUBREL data packet, and the PUBCOMP data packet, respectively. The PUBREL data packet will be transmitted only after the PUBREC data packet is successfully fed back. For ease of presentation, we collectively refer to the first two transmission processes as the first transmission process under the QoS2 mode, and collectively refer to the latter two transmission processes as the second transmission process under the QoS2 mode.

In the first transmission process, $a_{i,j,n}^{2,p(i)}$ and $b_{i,j,n}^{2,p(i)}$ are defined as the transmission indicator variable of the PUBLISH data packet and the return indicator variable of the PUBREC data packet, respectively. In the second transmission process, $a_{i,j,n.REL}^{2,p(i)}$ and $b_{i,j,n,COMP}^{2,p(i)}$ are defined as the transmission indicator variable of the PUBREL data packet and the return indicator variable of the PUBCOMP data packet. The expressions of the above indicator variables are similar to the QoS1 mode.

Therefore, in the QoS2 mode, the transmission energy consumption of the j -th small packet of the i -th large packet in the first transmission process is given by:

$$\begin{aligned} E_{i,j,first}^{2,p(i),TX} &= \sum_{n=1}^{N_{i,j}^{1st}} a_{i,j,n}^{2,p(i)} \frac{S}{B\log_2\left(1+p(i)g_{i,j,n}^{1st}\right)} p(i) \\ &+ \sum_{n=1}^{N_{i,j}^{1st}} b_{i,j,n}^{2,p(i)} \left(\frac{S}{B\log_2\left(1+p(i)g_{i,j,n}^{1st}\right)} p(i) + \frac{S_{REC}}{B\log_2\left(1+p_{back}g_{i,j,n,back}^{1st}\right)} p_{back} \right) \\ &+ \frac{S}{B\log_2\left(1+p(i)g_{i,j,N_{i,j}}^{1st}\right)} p(i) + \frac{S_{REC}}{B\log_2\left(1+p_{back}g_{i,j,N_{i,j},back}^{1st}\right)} p_{back} \end{aligned} \tag{13}$$

where $g_{i,j,n,REL}^{2,p(i)}$ and $g_{i,j,n,back}^{1st}$ represent the channel gain when the PUBLISH data packet and the PUBREC data packet are transmitted, respectively. $N_{i,j}^{1st}+1$ represents the total number of transmissions of the PUBLISH data packet and the PUBREC data packet. The first item of (13) represents the transmission energy consumption in the case of PUBLISH data packet transmission failure. The second item of (13) represents the PUBLISH data packet transmission energy consumption caused by the PUBREC data packet feedback failure and the transmission energy consumption of the PUBREC data packet. The third and fourth items of (13) represent the transmission energy consumption when the PUBLISH data packet is successfully transmitted and the PUBREC packet is also fed back successfully:

$$E_{i,j,\sec ond}^{2,p(i),TX} = \sum_{n=1}^{N_{i,j}^{2nd}} a_{i,j,n,REL}^{2,p(i)} \frac{S_{REL}}{B\log_2\left(1+p(i)g_{i,j,n}^{2nd}\right)}p(i) + \sum_{n=1}^{N_{i,j}^{2nd}} b_{i,j,n,COMP}^{2,p(i)} \left(\frac{S_{REL}}{B\log_2\left(1+p(i)g_{i,j,n}^{2nd}\right)}p(i) + \frac{S_{COMP}}{B\log_2\left(1+p_{back}g_{i,j,n,back}^{2nd}\right)}p_{back} \right) + \frac{S_{REL}}{B\log_2\left(1+p(i)g_{i,j,N_{i,j}}^{2nd}\right)}p(i) + \frac{S_{COMP}}{B\log_2\left(1+p_{back}g_{i,j,N_{i,j},back}^{2nd}\right)}p_{back} \tag{14}$$

where $g_{i,j,n}^{2nd}$ and $g_{i,j,back}^{2nd}$ represent the channel gain when the PUBREL data packet and the PUBCOMP data packet are transmitted, respectively. S_{REL} and S_{COMP} represent the data size of each PUBREL data packet and PUBCOMP data packet, respectively. $N_{i,j}^{2nd}+1$ represents the total number of transmissions of the PUBREL data packet and the PUBCOMP data packet. The first item of (14) represents the transmission energy consumption in the case of the PUBREL data packet transmission failure. The second item of (14) represents the PUBREL data packet transmission energy consumption caused by the PUBCOMP data packet transmission failure and the transmission energy consumption of the PUBCOMP data packet transmission. The third and fourth items of (14) represent the transmission energy consumption when the PUBREL data packet and the PUBCOMP packet are both transmitted. Each small packet transmission process is completed when both the PUBREL data packet and the PUBCOMP data packet are transmitted successfully.

Hence, the total energy consumption in the QoS2 mode consists of the transmission energy consumption of the two processes without deduplication, which is given by:

$$E_i^{2,p(i)} = \sum_{j=1}^{J}\left(E_{i,j,first}^{2,p(i),TX} + E_{i,j,\sec ond}^{2,p(i),TX}\right) \tag{15}$$

3.4 Problem Formulation

The objective of the joint optimization problem is to balance the tradeoff between the packet-loss ratio and the energy consumption through the optimization of the MQTT QoS mode selection and power control based on local information and historical information. Hence, the optimization problem is formulated as:

$$\begin{aligned} P1: &\min_{\{m(i),p(i)\}} \sum_{i=1}^{I}\left(Q_i^{m(i),p(i)} + VE_i^{m(i),p(i)}\right) \\ s.t.\quad & C_1: \forall m(i) \in \{0,1,2\} \\ & C_2: \forall p(i) \in \{p_1,\ p_2,\ \cdots,\ p_k,\ \cdots\ p_K\} \end{aligned} \tag{16}$$

where V is a nonnegative weight parameter used to weigh the importance of packet-loss ratio and energy consumption. C_1 and C_2 represent the constraints of MQTT QoS mode selection and power control variables.

4. EPSILON-GREEDY-BASED MQTT QOS MODE SELECTION AND POWER CONTROL ALGORITHM

In this section, the design and implementation process of the proposed EMMA algorithm are elaborated, which includes the MAB model, the principle of the proposed EMMA algorithm, and the implementation process.

4.1 MAB Model

MAB is an effective method to solve a series of multi-stage decision-making problems under incomplete information scenarios. The MAB model is mainly composed of players, arms, rewards, actions, and policies. The MAB model assumes that each gambling machine has multiple arms, and the return of each arm obeys a certain probability distribution. The player chooses an action through strategy in each round without the information, i.e., pull the arm, and the pulled arm will give the player a reward value. The purpose of the player is to maximize the total reward by constantly pulling the arm. Specifically, we transform the optimization problem **P1** into an MAB problem, which is described below.

Player: PD-IoT devices are defined as players.

Arm: All possible combinations of MQTT QoS mode selection and power control strategies are defined as arms, i.e.:

$$\left\{m(i), p(i) \mid \forall m(i) \in \left\{0,1,2\right\}, \forall p(i) \in \left\{p_1, p_2, \cdots, p_k, \cdots p_K\right\}\right\}$$

There are a total of $3K$ arms.

Reward: The opposite number of the weighted sum of the packet-loss ratio and energy consumption is defined as the reward of selecting QoS m mode and transmission power p_k, which is given by:

$$Reward_i^{m(i),p(i)} = -\left(Q_i^{m(i),p(i)} + VE_i^{m(i),p(i)}\right) \tag{17}$$

Action: Selecting MQTT QoS mode selection and power control is defined as action, and the action indicator variable is defined as $x_{i,m,k}$ ($\forall i \in \mathcal{I}$, $\forall m \in \{0,1,2\}$), $\mathcal{J} = \{1,2,\ldots,j,\ldots,J\}$ $\forall k \in \{1,2,\ldots,k,\ldots,K\}$, which is given by:

$$x_{i,m,k} = \begin{cases} 1, & m(i) = m \ and \ p(i) = p_k \\ 0, & otherwise \end{cases} \tag{18}$$

Policy: The strategy of selecting the MQTT QoS mode selection and power control is defined as a policy, and we design a policy based on the epsilon-greedy algorithm, which is elaborated in the following subsection.

Accordingly, P1 can be transformed into P2, which is given by:

$$\begin{aligned} P2 : & \max_{\left\{x_{i,m,k}\right\}} \sum_{i=1}^{I} \sum_{m=0}^{2} \sum_{k=1}^{K} x_{i,m,k} Reward_i^{m,p_k} \\ s.t. \quad & C_3 : x_{i,m,k} \in \left\{0,1\right\}, \forall m \in \left\{0,1,2\right\}, \forall k \in \left\{1,2,\cdots,K\right\} \\ & C_4 : \sum_{m=0}^{2} \sum_{k=1}^{K} x_{i,m,k} = 1, \forall m \in \left\{0,1,2\right\}, \forall k \in \left\{1,2,\cdots,K\right\} \end{aligned} \tag{19}$$

where constraints C_3 and C_4 correspond to constraints C_1 and C_2 in P1, which means that each large packet can only select one MQTT QoS mode and power control strategy, and the MQTT QoS mode selection and power control strategy remains unchanged during a large data packet transmission.

4.2 The Proposed EMMA Algorithm

The proposed algorithm named EMMA is developed based on the epsilon-greedy algorithm, which provides effective method to solve the MAB problem, and balance the tradeoff between exploration and exploitation performance. Specifically, the policy based on the proposed EMMA algorithm is given by:

$$\phi_i = \begin{cases} \arg\max\limits_{x_{i,m,k}} \overline{Reward_{i-1}^{m,p_k}}, & u > \varepsilon \\ random\ \ selection, & otherwise \end{cases} \tag{20}$$

where ϕ_i represents the MQTT QoS mode selection and power control strategy of the i -th large package. $\overline{Reward_{i-1}^{m,p_k}}$ represents the empirical average reward of the strategy that selects QoS m mode and transmission power p_k, which is given by:

$$\overline{Reward_i^{m,p_k}} = \frac{\overline{Reward_{i\text{-}1}^{m,p_k}}\, r_{i-1,m,k} + x_{i,m,k} Reward_i^{m,p_k}}{r_{i-1,m,k} + x_{i,m,k}} \tag{21}$$

where $r_{i,m,\mathrm{k}}$ represents the total number of times that QoS mode m and transmission power p_k are selected when the i -th large packet is transmitted, which is given by:

$$r_{i,m,\mathrm{k}} = r_{i-1,m,k} + x_{i,m,k} \tag{22}$$

The principle of the proposed EMMA algorithm is to set the value of $\varepsilon \in \left(0,1\right)$. Then EMMA explores with the probability of ε each time, i.e., randomly selects an arm, and exploits with the probability of $1-\varepsilon$, i.e., selects the arm with the largest empirical average reward. The closer ε is to 0, the more conservative EMMA is towards exploration.

4.3 The Implementation Process

The implementation process of the proposed algorithm is shown in Algorithm 1, which includes three phases, i.e., the initialization phase, the decision-making phase, and the learning phase.

In the initialization phase, all indicator variables are initialized as zero, i.e.:

$$x_{i,m,k} = 0, a_{i,j}^{0,p(\mathrm{i})} = 0, a_{i,j,n}^{1,p(\mathrm{i})} = 0, b_{i,j,n}^{1,p(\mathrm{i})} = 0, a_{i,j,n}^{2,p(\mathrm{i})} = 0, b_{i,j,n}^{2,p(\mathrm{i})} = 0, a_{i,j,n.REL}^{2,p(\mathrm{i})} = 0, b_{i,j,n,COMP}^{2,p(\mathrm{i})} = 0$$

We assume that when a PD-IoT device transmits the first $3K$ large packets, each mode is selected once.

In the decision-making stage, a random number $\mu \in (0,1)$ is firstly generated, and the MQTT QoS mode and power control strategy ϕ_i is selected according to (20).

In the learning phase, the PD-IoT device firstly observes the channel states and obtains the packet loss of each small packet. Then the packet-loss ratio and energy consumption of the i -th large packet are calculated. Finally, $\overline{Reward_i^{m,p_k}}$ and $r_{i,m,\mathrm{k}}$ are updated according to (20) and (21), respectively.

Algorithm 1. The Proposed EMMA Algorithm

Input: $I, J, V, B, G_{th}, S, S_{BACK}, S_{REC,} S_{REL}, S_{COMP}, p_{back}, t_0, E_0, \varepsilon, n_0, p_k, \forall k \in \{1, 2, \&, K\}$.
Output: ϕ_i .
Phase 1: Initialization
Set $x_{i,m,k} = 0, a_{i,j}^{0,p(\mathrm{i})} = 0, a_{i,j,n}^{1,p(\mathrm{i})} = 0, b_{i,j,n}^{1,p(\mathrm{i})} = 0, a_{i,j,n}^{2,p(\mathrm{i})} = 0, b_{i,j,n}^{2,p(\mathrm{i})} = 0, a_{i,j,n.REL}^{2,p(\mathrm{i})} = 0, b_{i,j,n,COMP}^{2,p(\mathrm{i})} = 0$
$\forall i \in \mathcal{I}, \forall m \in \{0,1,2\}, \forall k \in \{1,2,\ldots,K\}, \forall j \in \mathcal{J}, \forall n \in \{0,1,\ldots,N_{i,j}\}$
When the first $3K$ large packets are transmitted, each mode is selected once.
for each large packet $\forall i \in \mathcal{I}$ **do**
    Phase 2: Decision-making
    Generate a random number $\mu \in (0,1)$.
    Select the optimal MQTT QoS mode and power control strategy ϕ_i according to (20).
    Phase 3: Learning
    for each small packet j **do**
        Observe the channel states and obtains the packet loss of each small packet.
    end for
        Calculate the packet-loss ratio and energy consumption.
        Update $\overline{Reward}_i^{m,p_k}$ according to (21).
        Update $r_{i,m,\mathrm{k}}$ according to (22).
end for

5. SIMULATION RESULTS

In this section, the performances of the proposed algorithm is evaluated through simulations. We assume that the channel gain is randomly distributed within (4, 9) in the first 300 large packet transmissions, and randomly distributed within (6, 11) in the next 500 large packet transmissions (Zheng R., *et al.*, 2013; Guamán Y., *et al.*, 2020). Six fixed MQTT QoS mode selection and power control strategies are used for comparison. The genetic-based MQTT QoS mode selection and power control algorithm named GMMA is utilized as a comparison algorithm.

Figure 2 shows the weight sum of packet-loss ratio and energy consumption versus the number of large packet transmissions. It can be seen that the QoS1/ p_1 strategy performs the best during the first 300 large packet transmissions. After 300 large packet transmissions, all the curves show the downward trend, and QoS0/ p_1 strategy decreases fastest and performs best. The reason is that after 300 large packet transmissions, the channel quality is improved, thereby decreasing the packet-loss ratio significantly under the QoS0 mode. Compared with the fixed strategies QoS0/ p_1 and QoS1/ p_1 , the proposed algorithm can reduce the weighted sum of the packet-loss ratio and energy consumption by 7.41% and 49.49%. The reason is that the proposed algorithm can dynamically learn the optimal QoS mode selection and power control strategy after the channel changes, and it switches from QoS1/ p_1 to QoS0/ p_1 . However, GMMA cannot carry out dynamic learning, resulting in poor performance after channel changes.

Figure 3 and Figure 4 show the packet-loss ratio and energy consumption versus the number of large packet transmissions. It can be seen that in all large packet transmissions, the energy

Figure 2. The weight sum of packet-loss ratio and energy consumption versus the number of large packet transmissions

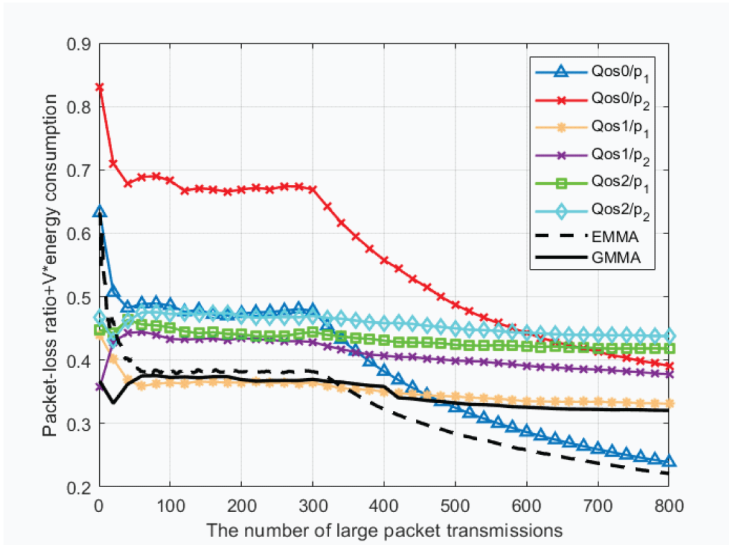
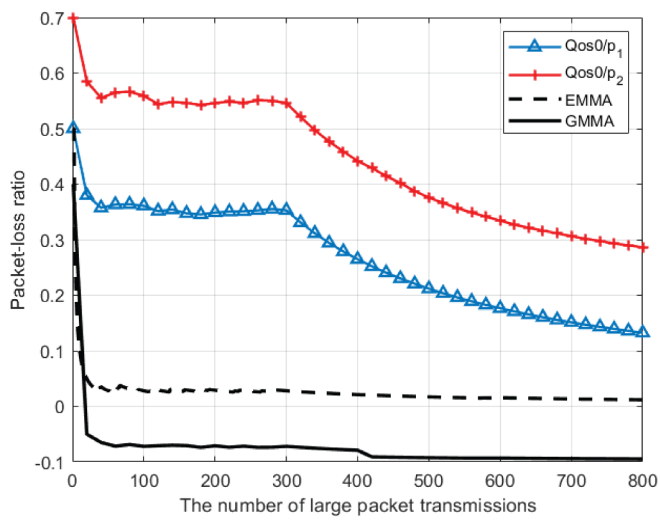


Figure 3. The packet-loss ratio versus the number of large packet transmissions



consumption of mode QoS0/ p_1 and QoS0/ p_2 are the lowest, but the packet-loss ratio are both higher than the proposed algorithm. The reason is that the QoS0 mode sacrifices packet-loss ratio performance to reduce the energy consumption. GMMA has a high packet-loss rate due to its poor channel status recognition ability. After 300 large packet transmissions, the energy consumption of each strategy and the packet-loss ratio decreases. Simulation results demonstrate that the proposed algorithm can balance the tradeoff between packet-loss ratio and energy consumption when the channel state is changing.

Figure 4. The energy consumption versus the number of large packet transmissions

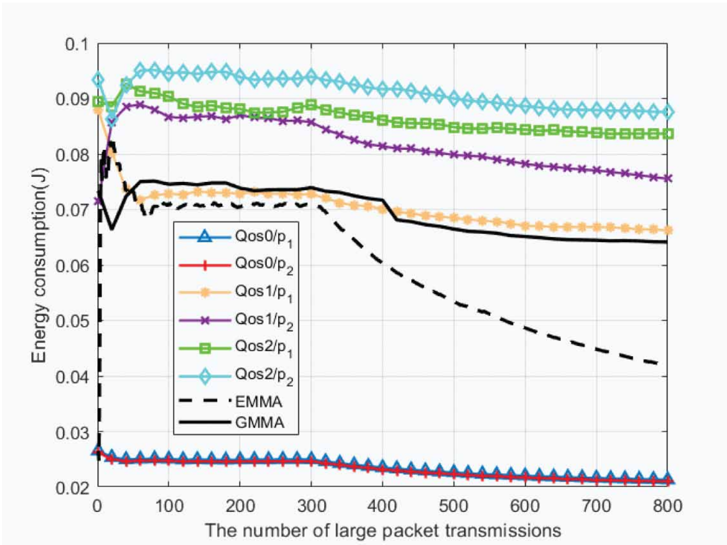


Figure 5 shows the impact of G_{th} on the weight sum of packet-loss ratio and energy consumption. It can be seen that the weight sum of packet-loss ratio and energy consumption increases with the increase of the value of G_{th}. The reason is that as G_{th} increases, the packet-loss ratio of the QoS0 mode gradually increases, and the energy consumption for data packet retransmission also increases. In the QoS1 mode and the QoS2 mode, to guarantee the packet-loss ratio as zero, the number of

Figure 5. The impact of G_{th} on the weight sum of packet-loss ratio and energy consumption

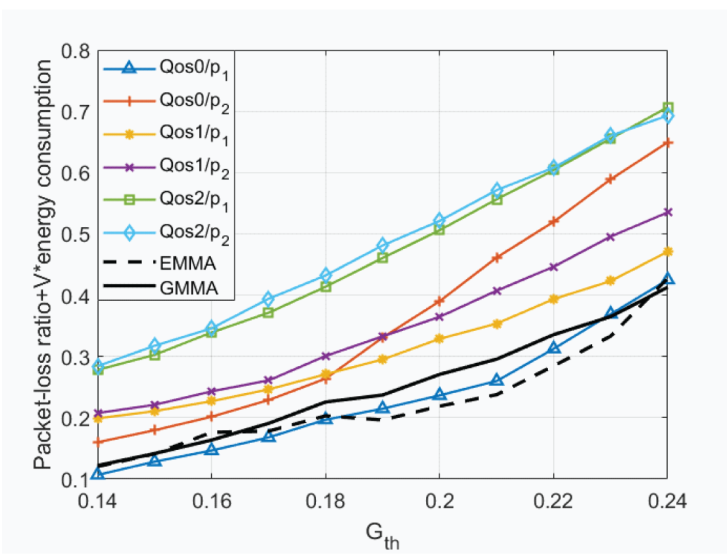
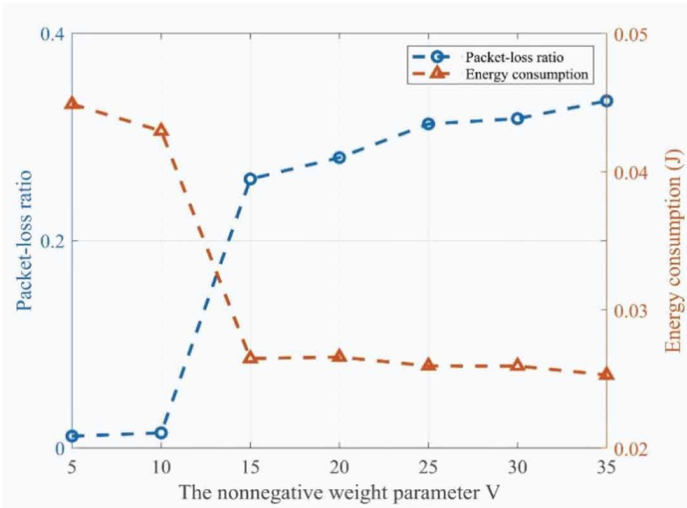


Figure 6. The impact of the nonnegative weight parameter V on packet-loss ratio and energy consumption



packet retransmissions increases so that the weight sum of packet-loss ratio and energy consumption increases. GMMA performs inferiorly to EMMA due to its poor dynamic learning ability. When G_{th} =(0.18, 0.24), the proposed algorithm performs the best since it can dynamically select the optimal strategy according to the channel state.

Figure 6 shows the impact of the nonnegative weight parameter V on packet-loss ratio and energy consumption. It can be observed that as the nonnegative weight parameter V increases, the energy consumption decreases while the packet-loss ratio increases. The reason is that with the increase of V, the proposed algorithm is more concerned about the energy consumption and less concerned about packet-loss ratio. The proposed algorithm can balance the tradeoff between packet-loss ratio and energy consumption by adjusting the value of the nonnegative weight parameter V.

6. CONCLUSION

In this paper, we proposed an epsilon-greedy-based MQTT QoS mode selection and power control algorithm named EMMA. Firstly, the weighted sum of packet-loss ratio and energy consumption was minimized by jointly optimizing MQTT QoS mode selection and power control, which s modeled as an MAB problem. Secondly, the epsilon-greedy algorithm with efficient online learning capacity is leveraged to make decisions based on the historical information and local information at the PD-IoT device side. Finally, we verify the superior performance of the proposed algorithm through simulations. The simulation results elaborate that, compared with the fixed strategies QoS0/ p_1 and QoS1/ p_1, the proposed algorithm can reduce the weighted sum of the packet-loss ratio and energy consumption by 7.41% and 49.49%, respectively. Furthermore, when the threshold of signal-to-noise ratio is adjusted, the proposed algorithm can always choose the optimal strategy. In the future, we will study how to integrate the energy harvesting into the proposed framework.

ACKNOWLEDGMENT

This research was supported by Science and Technology Project of State Grid Shandong Electric Power Company (grant number 52062619002C).

FUNDING AGENCY

The publisher has waived the Open Access Processing fee for this article.

REFERENCES

Adeogun, R., Berardinelli, G., Rodriguez, I., & Mogensen, P. (2020). Distributed Dynamic Channel Allocation in 6G in-X Subnetworks for Industrial Automation. *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 1-7.

Ding, Z., Xu, J., Dobre, O. A., & Poor, H. V. (2019). Joint Power and Time Allocation for NOMA-MEC Offloading. *IEEE Transactions on Vehicular Technology*, *68*(6), 6207–6211. doi:10.1109/TVT.2019.2907253

Guamán, Y., Ninahualpa, G., Salazar, G., & Guarda, T. (2020). Comparative Performance Analysis between MQTT and CoAP Protocols for IoT with Raspberry PI 3 in IEEE 802.11 Environments. *Proc. 2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*, 1-6.

Herrero, R. (2019). Dynamic CoAP Mode Control in Real Time Wireless IoT Networks. *IEEE Internet Things of Journal*, *6*(1), 801–807.

Jo, H., & Jin, H. (2015). Adaptive Periodic Communication over MQTT for Large-Scale Cyber-Physical Systems. *Proc. 2015 IEEE 3rd International Conference on Cyber-Physical Systems, Networks, and Applications,* 66-69.

Kim, G., Kang, S., Park, J., & Chung, K. (2019). An MQTT-Based Context-Aware Autonomous System in one M2M Architecture. *IEEE Internet of Things Journal*, *6*(7), 8519–8528. doi:10.1109/JIOT.2019.2919971

Kim, S. J., Cho, K. W., Lee, M. E., & Oh, C. H. (2017). A Study on Adaptive QoS Control System based on MQTT for Reducing Network Traffic. *Proc. International Conference On Future Information & Communication Engineering*, 223-226.

Kiran, N., Pan, C., Wang, S., & Yin, C. (2020). Joint Resource Allocation and Computation Offloading in Mobile Edge Computing for SDN based Wireless Networks. *Journal of Communications and Networks (Seoul)*, *22*(1), 1–11.

Liao, H. (2020). Blockchain and Learning-Based Secure and Intelligent Task Offloading for Vehicular Fog Computing. *IEEE Transactions on Intelligent Transportation Systems*, *22*(7), 1–13.

Liao, H., Zhou, Z., Zhao, X., Zhang, L., Mumtaz, S., Jolfaei, A., Ahmed, S. H., & Bashir, A. K. (2020). Learning-Based Context-Aware Resource Allocation for Edge-Computing-Empowered Industrial IoT. *IEEE Internet of Things Journal*, *7*(5), 4260–4277. doi:10.1109/JIOT.2019.2963371

Liu, J., Zhao, Z., & Ji, X. (2018). Research and Application of Internet of Things in Power Transmission and Distribution System. *Chinese Journal on Internet of Things*, *1*(4), 88–102.

Liu, Z. (2016). Research of Global Clean Energy Resource and Power Grid Interconnection. *Zhongguo Dianji Gongcheng Xuebao*, *36*(41), 5103–5110.

Lv, J., Luan, W., Liu, R., Wang, P., & Lin, J. (2018). Architecture of Distribution Internet of Things based on Widespread Sensing & Software Defined Technology. *Power System Technology*, *42*(3), 3108–3115.

Mumtaz, S., Alsohaily, A., Pang, Z., Rayes, A., Tsang, K. F., & Rodriguez, J. (2017). Massive Internet of Things for Industrial Applications: Addressing wireless IIoT Connectivity Challenges and Ecosystem Fragmentation. *IEEE Industrial Electronics Magazine*, *11*(1), 28–33. doi:10.1109/MIE.2016.2618724

Oksanen, J., Lundén, J., & Koivunen, V. (2010). Reinforcement Learning Method for Energy Efficient Cooperative Multiband Spectrum Sensing. *Proc. 2010 IEEE International Workshop on Machine Learning for Signal Processing*, 59-64.

Sadio, O., Ngom, I., & Lishou, C. (2019). Lightweight Security Scheme for MQTT/MQTT-SN Protocol. *Proc. 2019 Sixth International Conference on Internet of Things: Systems, Management and Security (IOTSMS)*, 119-123. doi:10.1109/IOTSMS48152.2019.8939177

Stankovic, J. A. (2020). Research Direction for the Internet of Things. *IEEE Internet of Things Journal*, *1*(1), 3–9. doi:10.1109/JIOT.2014.2312291

Toldinas, J., Lozinskis, B., Baranauskas, E., & Dobrovolskis, A. (2019). MQTT Quality of Service versus Energy Consumption. *Proc. 2019 23rd International Conference Electronics*, 1-4. doi:10.1109/ELECTRONICS.2019.8765692

Zheng, R., Le, T., & Han, Z. (2013). Approximate Online Learning for Passive Monitoring of Multi-Channel Wireless Networks. *Proceedings - IEEE INFOCOM*, 3111–3119.

Zhou, Z., Wang, Z., Yu, H., Liao, H., Mumtaz, S., Oliveira, L., & Frascolla, V. (2020). Learning-Based URLLC-Aware Task Offloading for Internet of Health Things. *IEEE Journal on Selected Areas in Communications*, *39*(2), 396–410. doi:10.1109/JSAC.2020.3020680