Receiver Beacon Transmission Interval Design Using Machine Learning for Multi-Stage Wireless Sensor Networks

Yu-ki Hatada†, Takeo Fujii†
† Advanced Wireless and Communication Research Center (AWCC), The University of Electro-Communications
1–5–1 Chofugaoka, Chofu, Tokyo 182–8585, Japan
Email: {hatada, fujii}@awcc.uec.ac.jp

***Abstract*—U-Bus Air is a protocol that can transmit packets by multi-hop communication, and is utilized in tall buildings and apartments. While U-Bus Air is a time asynchronous wireless sensor network, packet collisions cause degradation of communication quality as the number of terminals increases. A previous research has been proposed, an appropriate transmission cycle control algorithm based on the estimated collision probability by using collected topology information at a Fusion Center (FC). By using this algorithm, the communication performance is improved by feedback from FC to each terminal. However, its communication overhead becomes large due to topology information and feedback packets, and communication efficiency is degraded. In order to solve this problem, we propose a method to autonomously control receiver beacon transmission interval by leveraging decentralized Q-Learning. In this method, each terminal transmits information of a packet holding time at the same time as responding to the beacon. Finally, each terminal utilizes Q-Learning to derive an appropriate transmission interval. In computer simulation, we show an improvement of packet delivery rate (PDR), throughput and power consumption compared with the previous method.**

***Index Terms*—Smart Meter, Machine Learning, Power Saving, Communication Quality**

I. Introduction

Recently, IoT (Internet of Things), which connects various things to the Internet, has attracted attention. In particular, smart meters are expected to visualize consumption of a power, a water, a gas, etc. and reduce labor's costs by leveraging wireless communication to collect infrastructure information. To realize an efficient household energy usage, HEMS (Home Energy Management System) in smart grid is an important technology. Smart grids help to efficiently manage the energy balance between demand and supply by visualizing their usage information. For example, a user can understand the amount of wasted power and a supplier can estimate the power supply quantity [1].

Wi-SUN (Wireless Smart Utility Network) has been standardized as a communication protocol for smart meters [2]. In Japan, JUTA (Japan Utility Telemetering Association) have standardized a low power consumption on demand protocol called U-Bus Air [3] based on Wi-SUN. The communication protocol in U-Bus Air is based on IRDT (Intermittent Receiver-Driven Transmission)-MAC [4], and performs shake hand communication. The basic operation is the same, except that a carrier sense is used for transmission. U-Bus Air constructs a smart meter network by multi-hop communication and collects information packets to Fusion Center (FC). Therefore, U-Bus Air consists of direct communication terminals connected between other terminals and wide area network called FC and other terminals connected each other with multi-hop [5]. Figure 1 shows the system model of U-Bus Air. Each terminal transfers the usage information to FC by the multi-hop packet communication. This protocol, multi-hop communication and power saving realizes by utilizing receiver-based intermittent beacons.

In smart meters, in order to achieve a maintenance-free wireless network, a power-saving in the sleep state of the terminal is important. Therefore, the control of wake-up time to handle a data transfer timing is necessary to realize the sleep function. Currently, two types of protocols have been proposed for the wake-up timing control. The first one is a protocol with receiver-based data transmission control. All terminals which do not hold data, transmit beacon signals intermittently. If a terminal holds a data packet, the terminal shifts to a wait state for the beacon from the surrounding receivable terminal. Therefore, U-Bus Air and IRDT (Intermittent Receiver-Driven Transmission)-MAC are categorized into receiver-based communication systems. On the other hand, a transmitter-based protocol has also been proposed. In the transmitter-based protocol, the terminal transmits a beacon when holding a packet. When the terminal which transmits the beacon receives a response from the terminal which receives the beacon, it sends data packets. X-mac [6] is categorized into the transmitter-based communication system. Both the former and the latter protocols can control the wake-up timing by using such intermittent operation. However, if the intermittent cycle is too long or too short, an overhead increases a power consumption. Therefore, if communication can be administered at appropriate intermittent intervals, it is expected to reduce the power consumption.

In addition, in the previous research, a communication quality of wireless sensor networks such as U-Bus Air has been improved by minimizing a collision probability derived from topology information. In IRDT-MAC, the appropriate intermittent interval has been derived by calculating the collision probability with the topology information. However, there is a

problem that it is difficult to leave and join terminals because the topology is changed. Hence, the protocol with topology estimation has been considered. This protocol collects the topology information at FC, and then the collision probability is derived from the collected topology information (hereinafter, this is called "Feedback-MAC") [7]. The results showed that the communication performance can be improved even when the topology information is unknown. However, a large amount of overhead is required for obtaining the topology information and transmitting feedback packets, leading to degradation of the communication quality.

In this paper, we propose a method for deriving an appropriate intermittent interval of beacons in an autonomous distributed manner by controlling the packet holding time of terminals in lower layers utilizing Q-Learning within the threshold range. Each terminal measures the packet holding time with the time counter when the packet is generated and communicated. Next, a piggyback is performed on the response signal to the beacon, and the packet holding time is transmitted to the transmitting terminal. In addition, each terminal takes an averaging of packet holding time after a certain period from the reception timing of the information. Finally, based on the average value, Q-Learning controls the next intermittent interval so that the packet holding time approaches the threshold range. It is possible to control the beacon transmit interval even if each terminal does not know the topology information, because each terminal employs distributed control autonomously.

The rest of paper is organized as follows. Section Ⅱ explains the U-Bus Air MAC protocol and the packet collision problem. Section Ⅲ presents the learning and search algorithms in this study. Section Ⅳ explains the proposed design method of beacon transmission interval. Section Ⅴ presents a number of numerical simulations. Finally, Section Ⅵ presents the conclusions.

II. U-Bus Air

A. MAC protocol

U-Bus Air is a wireless sensor network protocol with multi-step relay, and the hierarchy can be divided by the number of communications required for communication with FC. Figure 1 shows U-Bus Air data transfer model. It is assumed that the number of hops is determined by the number of layers. For example, if there are 3 layers from the terminal to the FC, 3 steps are required. Figure 2 shows the MAC protocol overview of U-Bus Air. Each terminal continuously repeats a short beacon called RNO (Request NO) transmission period, the short reception waiting period, and the sleep period of several seconds. When the terminal holds a packet, the terminal shifts to a standby state to receive the RNO signal. If the RNO waiting time is completed and the RNO signal from the upper layer terminal is received, the terminal holding the data packet immediately transmits a request signal called SREQ (Send REQuest) signal. After receiving the SREQ signal, the terminal transmits a request response signal called RACK (Request

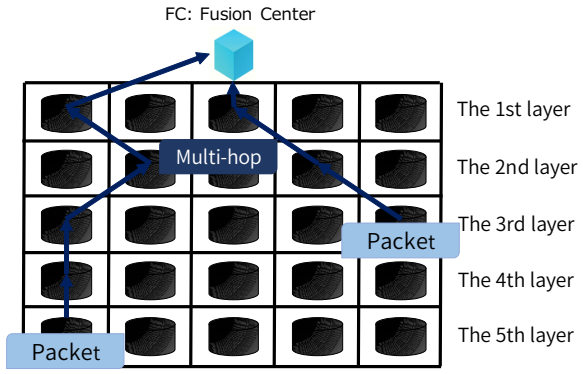

Fig. 1: Data transfer model in U-Bus Air.

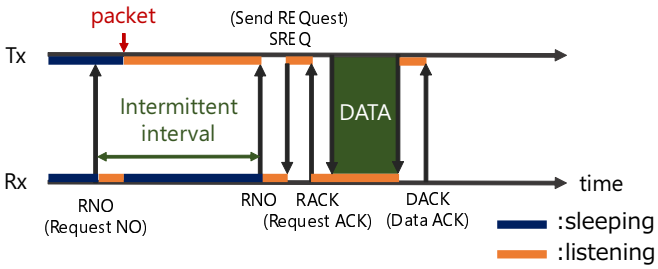

Fig. 2: U-Bus Air protocol.

ACKnowledgement) signal. When the terminal which is holding the data packet receives the RACK signal, a connection link for communication is established to transfer information. Finally, the terminal that receives the data packet transmits a DACK (Data ACKnowledgement) signal. The above operation is repeated, and the data packet is transferred to the FC via relay terminals. Since this protocol can temporarily form a link for communication, it is easy to add a new terminal to the network system of smart meter. Moreover, the power consumption is small because the state of the terminal is changed to the sleep state except the transmission state and the RNO standby state. SREQ and RACK signals before data packet transmission are useful to avoid data packet collisions. The smart meter communication protocol requires high packet transfer success rate at FC because customer data is included in the transmission data.

B. Problem of U-Bus Air

U-Bus Air has two problems in packet collision. In order to avoid packet collisions, the transmitting terminal conducts carrier sense before transmitting each signal. However, the collision avoidance utilizing the carrier sense causes a hidden terminal problem. As shown in Fig. 3, terminal A and B are out of the carrier sense detection area. At terminal C, DATA packets and RNO signal from terminal A collide. Even if the terminal A performs carrier sense, the signal of hidden terminal cannot be detected. In U-Bus Air protocol, the collision caused by RNO signal is dominant because many RNO signals are transmitted as compared with other

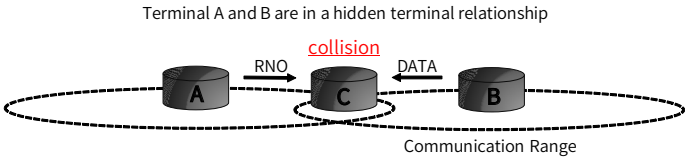

Fig. 3: Collision caused by RNO.

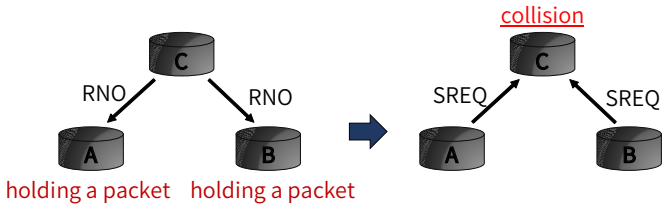

Fig. 4: Collision caused by SREQ.

signals. In particular, a packet loss is caused by the collisions between DATA and RNO signals, since data packets have a long transmission time. Therefore, when the intermittence interval of RNO signal is shorter, the problem of hidden terminal occurs and the collision increases. On the other hand, the transmitted beacons at intermittent intervals causes the other problem in U-Bus Air. As shown in Fig. 4, the terminal immediately responds to the SREQ signal when it receives the RNO signal. However, if there are multiple terminals are in the RNO standby state, such as terminal A and B, they respond to one RNO signal of terminal C. At this time, collisions between SREQ signals are caused. Therefore, the longer intermittent interval from the RNO signal, the more collisions between the SREQ signals and the packet loss increases. As a result, hidden terminal problems and specific collisions of U-Bus Air are in a trade-off relationship.

III. Reinforcement Learning

Markov Decision Process (MDP) [8] sequentially observes and makes decisions of the next action. After that, it is the process of being rewarded for reaching particular states. The dynamic environment can be solved by reinforcement learning that is divided into five components, such as learning agent, environment, policy, reward function, and value function [9]. The learning agent generates a pair of current state and action, and selects the action corresponding to the policy. When the current state is S_t and the action is a_t at time t, action based on the policy is selected from the current state. At an early stage, since the learning agent does not have environmental information, action is selected with randomness corresponding to the search method. As time passes, the learning agent learns the surrounding wireless environment and makes the action selection based on the previous learned policy.

A. Q-Learning

The Q-Learning algorithm [10] is a form of temporal difference learning predicting a quantity that depends on future values of a given information. At each time step, an agent in state, S_t, selects an action, a, in the next state, S_{t+1}, while acquiring reward, R_t, and transitions. The goal of Q-Learning is to maximize state action value function, $Q(S, a)$, rewarded by utilizing experience, (S_t, a_t, R_t, S_{t+1}), to learn. The quantity, $Q(S, a)$, is a measure of the total expected reward over the future if the agent chooses an action, a, in state, S, and then follows the policy. The Q-Learning updating rule in its general form is:

$$Q(S_t, a_t) \leftarrow Q(S_t, a_t) + \alpha \left[R_t + \gamma \max_a Q(S_{t+1}, a_t) - Q(S_t, a_t) \right]. \quad (1)$$

$\alpha\,(0 \leq \alpha \leq 1)$ represents the learning rate, which affects how much the Q value is updated. Next, $\gamma\,(0 \leq \gamma \leq 1)$ represents the discount rate, which determines the degree to which future rewards affect Q value updating.

B. Boltzmann selection

This section describes solution to the search and use dilemma in learning. Dilemma of search and profits is a problem in reinforcement learning. Focusing on the use of knowledge increases the possibility of erroneous selection, and focusing on search delays the recovery of profits. In this study, we adopt Boltzmann selection [11], which is effective for solving local optimal problems, and can find the entire extremum by searching. The transition probability equation in Boltzmann selection is:

$$\pi(S, a) = \frac{\exp \frac{Q(S,a)}{T}}{\sum_{b \in A} \exp \frac{Q(S,b)}{T}}. \quad (2)$$

T is the temperature parameter, and A is the number of selectable actions.

In Boltzmann selection, the degree of freedom of action selection can be adjusted in the process of learning. In the initial stage, the randomness is set large. After learning has progressed, we can make a better search by reducing the randomness. In this study, the number of states is large and the full search is difficult. Therefore, we use an approach to combine the action selection based on the packet holding time, which is explained in detail in Sect. 4, and Boltzmann selection.

IV. Proposed Design Method with Machine Learning of Intermittent Interval

In this section, we explain how to determine the beacon interval based on the packet holding time of the lower layer terminal leveraging Q-Learning. Each terminal starts counting the time when it generates a packet and obtains a packet by communication. When the terminal holding the packet receives the RNO signal from the terminal of the upper layer, the SREQ signal is transmitted as the response. At that time, the information of the packet holding time is piggybacked to the SREQ signal and transmitted. After a constant time has passed, each terminal averages the packet holding time from collected

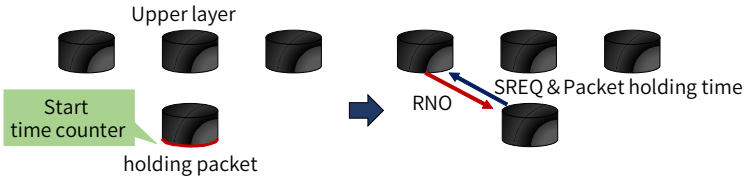

Fig. 5: Time counter and Transmission of packet holding time.

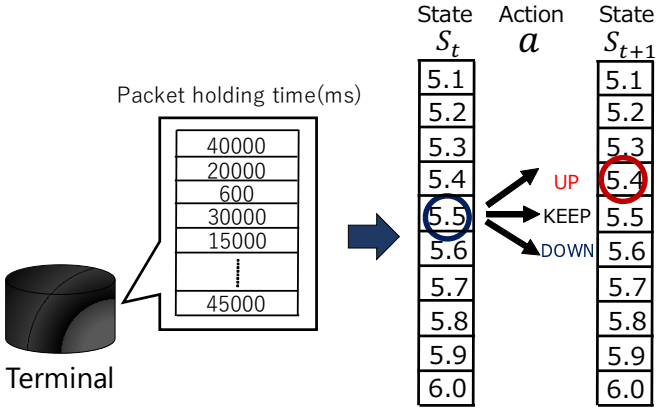

Fig. 6: Action determination by Q-Learning.

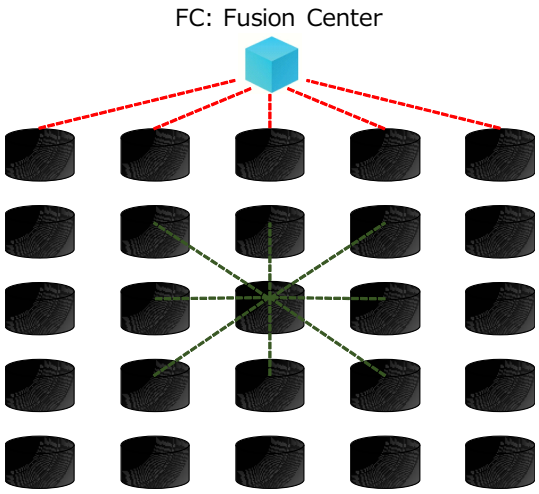

Fig. 7: Simulation model.

TABLE I: Simulation Parameters.

Parameter	Value
Terminal wake-up timing	0-5[sec]
Packet occur timing	5[min]
Default intermittent interval	5[sec]
Number of layers	1,2,3,4,5
Number of terminals per a layer	5
Running time	48[hour]
Search time of initial stage	24[hour]

terminals in lower layer, and then the terminal decides whether the beacon interval is changed or kept by using Q-Learning.

A. Counting and sending packet holding time

In U-Bus Air, the next terminal for transmission of a packet is determined by obtaining the RNO signal from other terminals. If the terminal that sent the RNO signal is the terminal of the upper layer, the terminal holding packet selects the next communication terminal and sends the SREQ signal. Therefore, each terminal can get an opportunity to communicate with terminals in lower layers. In our assumption, the terminal which is the reception state counts up by time. As shown in Fig. 5, the terminal holding the packet is in the receiving state and activates the time counter. Next, if the RNO signal arrives from the upper terminal, the information of the packet holding time is piggybacked on the SREQ signal and transmit to the upper layer terminal which makes transfer the data packet. When each terminal holds a packet, the above operation is performed intermittently.

B. The method of beacon transmission interval determination utilizing Q-Learning

At first, each terminal performs averaging of packet holding time after a certain time has elapsed. Based on the average value obtained by this method, the terminal leverages Q-Learning to select to maintain, shorten or extend the transmission interval. As shown in Fig. 6, in this study, we set the threshold range by focusing on the packet holding time of the terminal in the lower layer, and controlled by Q-Learning so as to carry out the intermittent interval of beacon within that range. Therefore, if the average packet holding time is within the threshold range, the reward is given. In addition, the control of the intermittent interval is performed every 0.1 [s] at a beacon interval of 1.0-10.0 [s], which is a feature of this study that the number of states is extremely numerous because there are 91 states. In order to make the search more appropriate, we divided the search into two forms. First, in order to search for the correct range, we acted to make the packet holding time of the lower layer terminal approach the threshold range. This is to avoid unnecessary searches in situations where there are many search ranges. This search method leads to the approximate range. For example, if the average a packet holding time is much higher than the threshold, the beacon transmission interval is shorten. However, if this search method is utilized alone, the correct intermittent interval cannot be converged. Therefore, we utilize Boltzmann selection as the next search method to determine the appropriate intermittent interval. By combining the two searches, it is possible to learn the appropriate range in the initial search, and to perform more accurate learning in Boltzmann selection.

V. Simulation Results and Performance Evaluations

In this section, we utilize computer simulation to evaluate the proposed Q-Learning based transmission control method with without topology information.

TABLE II: Frame Size.

Parameter	Value
RNO	18bytes (1.44ms)
SREQ	17bytes (1.36ms)
RACK	17bytes (1.36ms)
DATA	130bytes (10.4ms)
DACK	17bytes (1.36ms)

TABLE III: Power Parameters.

Parameter	Value
Voltage	1.8[V]
Tx current	18.0[mA]
Rx current	10.0[mA]
Sleep current	40[nA]

TABLE IV: Learning Parameters.

Parameter	Value
Number of states	91
Number of behavioral patterns	3(UP,KEEP,DOWN)
Learning rate	0.1
Discount rate	0.9
Reward	10
Range of reward condition	500-1000[ms]

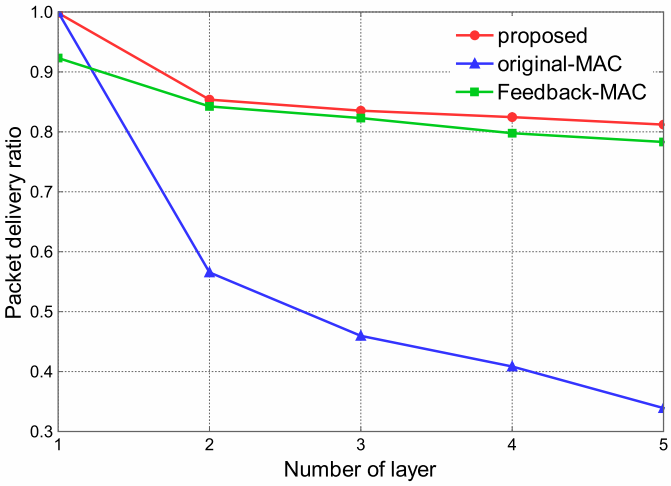

Fig. 8: Average Packet delivery ratio (Compared).

A. Simulation Model

Figure 7 shows the simulation model. Information packets are collected by FC via terminals of each layer. It is assumed that each terminal can detect signals of other terminals with a carrier sense only in adjacent terminals as shown in Fig. 7 and can link to only terminals in upper layers. Also, it is assumed that FC is linked to the terminal belonging to the highest layer. Table I shows simulation parameters. The number of terminals per layer is set to 5 and FC is set to 1, and communication performance is evaluated by increasing the number of multi-hop stages. Table II shows the frame size of U-Bus Air. Table III shows the power parameters of U-Bus Air. Finally, Table IV shows the learning parameters used in Q-Learning. In the simulation, the original protocol (hereinafter, this is called "original-MAC") utilizes a fixed intermittent interval and Feedback-MAC proposed in Ref.[7] are compared with the proposed method. The interval of original-MAC is fixed at 5 seconds. Except for Feedback-MAC, a link table and a feedback packet are not generated. We also assume that the power of terminal is not consumed in Q-Learning.

B. Packet Delivery Rate

Figure 8 shows a packet delivery rate (PDR) from each terminal to the FC. The proposed method can improve the degradation except in the case of layer 1 only as compared with the original-MAC. This reason is that the proposed method enables highly reliable communication by appropriately changing the intermittent interval for each terminal compared to the original-MAC.

Compared to Feedback-MAC, the proposed method can improve the PDR performance. In particular, in the case of one layer, significant improvement can be confirmed. The degradation of PDR performance is affected by the increase of traffic region due to the overhead of unnecessary transmit packets. Therefore, the performance in Feedback-MAC degrades. The proposed method can solve the problem caused by the overhead because there is no communication other than the necessary packet communication. However, the performance improvement is small even though the proposed method eliminates the overhead. There are two possible reasons. The first one is the profit loss by learning search. The proposed method tries many intermittent intervals for learning. Thus the PDR performance of the proposed method did not improve significantly. The second one is that the proposed method solves a quasi - optimum solution because all the information of surrounding terminals cannot be collected. The proposed method learns that the packet holding time of the lower layer terminals is within the threshold, but includes many dynamic factors. Therefore, the intermittent interval obtained by learning is the quasi - optimum solution.

C. Throughput Characteristics

Figure 9 shows the results of the throughput performance. The throughput performance also have the same tendency as PDR results. The improvement of each methods can be confirmed as the number of layers increases. Figure 9 confirms that the performance of the proposed method improves in all layers compared to the other methods.

D. Power Consumption

Figure 10 shows the results of power consumption per transmission packet to FC. It can be seen that the power consumption increases as the number of layer increases in the original-MAC. On the other hand, in the proposed method, since the power consumption is 7 J/packet or less, the performance in multiple stages can improve. This is because each terminal does not transmit an extra RNO signal. For example, the proposed method is modified to reduce the number of the RNO signal at the lowest terminal. As a result, it is not

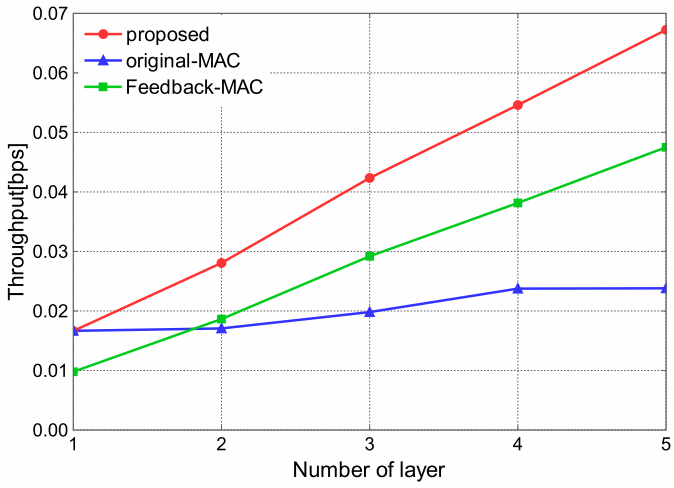

Fig. 9: Average throughput (Compared).

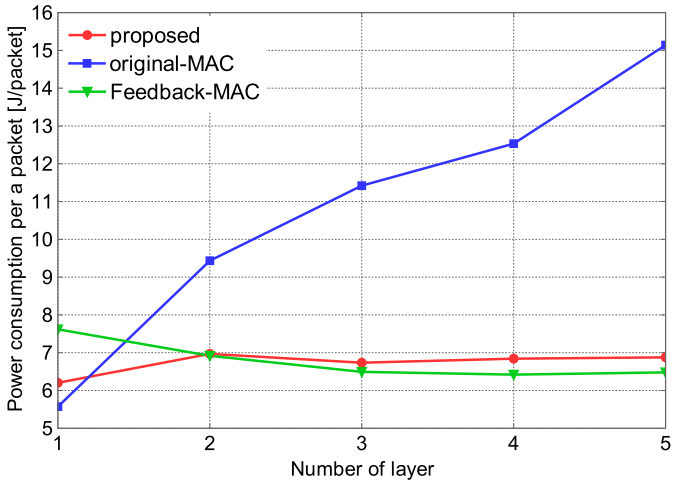

Fig. 10: Average power cunsumption per packet (Compared).

necessary to communicate the extra RNO signal, and then the total transmission power can be reduced.

Compared to Feedback-MAC, the proposed method consumes a little more power in 3 layers and later. However, despite the increase in the number of layers, the power consumption per packet can be maintained a low value.

VI. Conclusion

In this paper, we propose an appropriate receiver beacon transmission interval determination method utilizing Q-Learning in multiple-stage relay wireless network. In multiple-stage relay wireless network, the method of determining the communication frequency of each terminal is very important in terms of a power saving and communication performance. The performance evaluation of the proposed method is evaluated by computer simulation. From the results, the improvement of packet delivery rate, throughput and power consumption of each terminal are confirmed. In particular, the proposed method shows the highest performances except the power saving at all layers in the multistage topology. In the power consumption performance, the proposed method can be maintained at a low value.

Acknowledgment

A part of this work is results of joint research with Tokyo Gas Co., Ltd..

References

[1] Z. M. Fadlullah and N. Kato, Evolution of Smart Grids, ser. Springer Briefs in Electrical and Computer Engineering. Springer International Publishing, 2015.

[2] National Institute of Information and Communication Technology (NICT), "Sun (smart utility networks)," http://www2.nict.go.jp/wireless/smartlab/swl ja/project/sun.html.

[3] H. Hayashi, "Evolution of next-generation gas metering system in japan," Microwave Symposium (IMS), 2014 IEEE MTT-S International, pp. 1–4, June 2014.

[4] D. Kominami, M. Sugano, M. Murata, and T. Hatauchi, "Energy-efficient receiver-driven wireless mesh sensor networks," Sensors, vol. 11, no. 1, pp. 111–137, Dec. 2011.

[5] K. Suzawa, J. Fujiwara, M. Yasui, Y. Fujii, A. Asada, R. Yamashita, Y. Masuda, K. Nishiguchi, K. Fukushima, and T. Ichikawa, "Development of gas smart metering system," International Gas Union Research Conference (IGRC) 2014, Sep. 2014.

[6] M. Buettner, G. V. Yee, E. Anderson, and R. Han, "X-mac: A short preamble mac protocol for duty-cycled wireless sensor networks," Proceedings of the 4th International Conference on Embedded Networked Sensor Systems (SenSys) 06, pp. 307–320, Nov. 2006.

[7] T. Moriyama, T. Nakayama, T. Fujii, "Receiver beacon transmission interval design for multi-stage wireless sensor networks," International Conference on Ubiquitous and Future Networks (ICUFN) 2016.

[8] R. Bellman, "A markovian decision process," Indiana Univ. Math. J, vol. 6, pp. 679–684, 1957.

[9] R.Sutton, A. Barto, Reinforcement Learning: An Introduction, MIT Press, 1998.

[10] C. Watkins, P. Dayan, "Q-learning," Machine learning, vol. 8, no. 3-4, pp. 279–292, 1992.

[11] David E. Goldberg, "A note on Boltzmann tournament selection for Genetic algorithms and population-oriented simulated anneaing, "Complex Systems, vol. 4, pp. 445–460, 1990.