applied sciences

MDPI

Article

Multi-Armed-Bandit Based Channel Selection Algorithm for Massive Heterogeneous Internet of Things Networks

So Hasegawa [1,2], Ryoma Kitagawa [2], Aohan Li [3,*], Song-Ju Kim [2,4], Yoshito Watanabe [1], Yozo Shoji [1] and Mikio Hasegawa [2]

[1] National Institute of Information and Communications Technology, Tokyo 184-8795, Japan; so-hasegawa@nict.go.jp (S.H.); yoshito-watanabe@nict.go.jp (Y.W.); shoji@nict.go.jp (Y.S.)
[2] Department of Electrical Engineering, Tokyo University of Science, Tokyo 125-8585, Japan; r-kitagawa@haselab.ee.kagu.tus.ac.jp (R.K.); kim@sobin.org (S.-J.K.); hasegawa@ee.kagu.tus.ac.jp (M.H.)
[3] Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo 182-8585, Japan
[4] SOBIN Institute LLC, 3-38-7 Keyakizaka, Kawanishi 666-0145, Japan
* Correspondence: aohanli@ieee.org

Abstract: In recent times, the number of Internet of Things devices has increased considerably. Numerous Internet of Things devices generate enormous traffic, thereby causing network congestion and packet loss. To address network congestion in massive Internet of Things systems, an efficient channel allocation method is necessary. Although some channel allocation methods have already been studied, as far as we know, there is no research focusing on the implementation phase of Internet of Things devices while considering massive heterogeneous Internet of Things systems where different kinds of Internet of Things devices coexist in the same Internet of Things system. This paper focuses on the multi-armed-bandit-based channel allocation method that can be implemented on resource-constrained Internet of Things devices with low computational processing ability while avoiding congestion in massive Internet of Things systems. This paper first evaluates some well-known multi-armed-bandit-based channel allocation methods in massive Internet of Things systems. The simulation results show that an improved multi-armed-bandit-based channel selection method called Modified Tug of War can achieve the highest frame success rate in most cases. Specifically, the frame success rate can reach 95% when the numbers of channels and IoT devices are 60 and 10,000, respectively, while 12% channels are suffering traffic load by other kinds of IoT devices. In addition, the performance in terms of frame success rate can be improved by 20% compared to the equality channel allocation. Moreover, the multi-armed-bandit-based channel allocation methods is implemented on 50 Wi-SUN Internet of Things devices that support IEEE 802.15.4g/4e communication and evaluate the performance in frame success rate in an actual wood house coexisting with LoRa devices. The experimental results show that the modified multi-armed-bandit method can achieve the highest frame success rate compared to other well-known frame success rate-based channel selection methods.

Keywords: massive heterogeneous IoT networks; distributed channel selection; reinforcement learning; multi-armed bandit

Citation: Hasegawa, S.; Kitagawa, R.; Li, A.; Kim, S.-J.; Watanabe, Y.; Shoji, Y.; Hasegawa, M. Multi-Armed-Bandit Based Channel Selection Algorithm for Massive Heterogeneous Internet of Things Networks. *Appl. Sci.* **2022**, *12*, 7424. https://doi.org/10.3390/app12157424

Academic Editor: Dimitris Mourtzis

Received: 25 June 2022
Accepted: 22 July 2022
Published: 24 July 2022

1. Introduction

In recent years, the number of Internet-of-Things (IoT) devices has increased. The International Data Corporation (IDC) predicts that more than 40 billion IoT devices will generate 79 zettabytes (ZB;10^{21} bytes) by 2025 [1]. One of the goals of Beyond 5G is to support massive machine-type communications (mMTC). 6G-IoT aims to support the transmission of 10 million connected devices per square kilometer [2]. Various companies and alliances have developed unique low-power wide-area (LPWA) systems and have offered their own devices or network services. The wireless intelligent utility network

(Wi-SUN) spreads across more than 40 countries, Sigfox in more than 70 countries, and LoRa in more than 160 countries, and they have been leading the IoT world [3]. Network congestion would become unavoidable in a "massive" IoT environment. Each country has regulations on available frequencies band and communication channels for IoT devices. For example, the 915 MHz band is assigned in the US, the 433 MHz and 868 MHz band in the EU, and the 920 MHz band in Japan. In addition, the duty ratio, which indicates the communication frequency, is set to 1% or less or 10% or less. Efficient use of frequency resources is essential to ease congestion while complying with regulations. This fact has motivated many researchers to discuss dynamic spectrum access, where IoT devices can access channels dynamically to improve spectrum efficiency.

There are mainly two categories of resource allocation methods: centralized ones and decentralized ones. Regarding the centralized methods, references [4] propose time-slotted channel hopping (TSCH) based channel allocation methods that have been adopted in the IEEE 802.15.4e standard [5] to improve the IoT network performance. Reference [6] proposes a resource allocation method based on deep Q learning performed at the gateway (GW) side. Although the centralized methods can adequately manage the operation of end nodes, increasing the number of nodes brings the burden on the controlling server. In addition, synchronization and constant connection between GW and IoT devices are required, resulting in high power consumption of the end device. Therefore, the centralized methods are unsuitable for avoiding collisions in massive IoT networks.

The IoT device decides the access channels in the autonomous decentralized resource allocation method. Reference [7] proposed a decentralized channel assignment method for cognitive radio networks. In [8], a reinforcement learning method was adopted to dynamically select channels in a complicated communication environment. In [9], a deep learning-based method was proposed to determine the transmission schedule and power. References [10,11] treated channel access by a single user as a multi-armed bandit (MAB) problem [12]. Reference [10] proposed a channel selection protocol based on optimization and reference [11] analyzed the average communication performance of competitive users. Reference [13] formulated the channel access problem as multi-player MAB (MP-MAB) problems. In [14], the channel assignment problem is also formulated as an MP-MAB problem. The proposed MAB-based channel assignment method in this work is implemented on a single-board computer supported by IEEE 802.15.4g/4e communication. The performance is evaluated using 30 IoT devices, verifying that the MAB methods are efficient for the channel assignment in dynamic IoT systems. However, a massive heterogeneous scenario is not well considered. In [15], a distributed learning technique based on bandit algorithms is proposed for LoRa devices to select their access selection. In [16], a distributed channel selection method based on TOW dynamics is proposed for fully decentralized networks. Both reference [15] and reference [16] implement and evaluate their proposed methods on the practical IoT devices.

Contributions of This Paper

This paper focuses on lightweight learning algorithms that can be implemented on IoT devices. MAB algorithms are the simplest reinforcement learning method. In the MAB algorithm-based channel selection methods, the channel can be selected only based on ACK information. The communication performance may be improved by using some other information besides ACK information. However, it takes time to obtain the information, which may increase energy consumption. Since the state information other than ACK information is necessary for the Q learning method or deep reinforcement learning method in related work, which may reduce energy efficiency. Hence, compared to the other reinforcement learning method, the MAB methods considered in this paper may achieve higher energy efficiency and are more suitable for battery powered IoT devices. On the other hand, although there are several works on MAB-based channel selection that are implemented on IoT devices, channel selection under massive heterogeneous IoT networks is not considered in the related work.

To support substantial IoT devices in the next-generation communication systems, the effectiveness of the MAB-based channel assignment methods in massive heterogeneous IoT networks is verified in this paper. Specifically, the effectiveness of the MAB-based channel assignment method in a massive heterogeneous IoT network consisting of 10,000 devices is firstly evaluated by simulations. Subsequently, the performance in frame success rate (FSR) is evaluated using the 50 Wi-SUN IoT devices with IEEE 802.15.4g/4e protocol in the IoT networks coexisting with the LoRa devices. The contributions of this paper can be summarized as follows.

- The channel assignment problem is formulated as a MAB problem and apply MAB algorithms to solve the formulated problem in massive heterogeneous IoT networks.
- The effectiveness of the MAB-based channel assignment methods in FSR is evaluated and verified under a massive IoT network with 10,000 IoT devices.
- The MAB-based channel assignment methods is implemented on actual Wi-SUN IoT devices and evaluates the performance in FSR under the IoT heterogeneous network coexisting with LoRa devices.

The remainder of this paper is organized as follows. Section 2 summarizes the frequency standards of major countries and LPWA networks that are widely deployed worldwide. Section 3 describes the system model and problem formulation. Section 4 presents the MAB-based channel assignment methods. Section 5 evaluates the performance in FSR of the MAB-based channel assignment methods through simulations, assuming a massive heterogeneous IoT networks with 10,000 IoT devices. Section 6 describes the experiments conducted in a real IoT network where Wi-SUN IoT devices coexist with LoRa devices. Finally, Section 7 provides the concluding statement.

2. Low-Power Wide-Area Networks

Applications such as smart cities and smart meters need to cover a large area with low power consumption communication to realize long-term operation without maintenance. Traditional wireless local area networks (WLANs) such as Wi-Fi and cellular networks are not suitable to meet this requirement. To meet the requirement, various standards for IoT networks have already been developed over the last decade, such as LoRa, Sigfox, and Wi-SUN. This section introduces these major standards of the IoT networks.

LoRa is a unique chirp spread spectrum modulation technique optimized for long-range low-power communications. The data rate of the LoRa devices mainly depends on the used bandwidth, spreading factor (SF), and forward error correction (FEC) rate. The bandwidth is typically set to 125 kHz or 250 kHz for the uplink and 500 kHz for the downlink. The SF values can range from 7 to 12, and the FEC rate can vary from 4/8 to 4/5. Setting a larger SF value can improve receiver sensitivity and wider coverage; however, the data rate is consequently reduced. LoRaWAN is the most widely used protocol stack for LoRa networks and has 240 million devices in 170 countries [17], End LoRa devices connect to one or more gateways through a single hop. A LoRa gateway can process up to nine channels in parallel by combining different sub-bands, and SF [18]. LoRa has a capture effect that makes it possible to recover a LoRa signal, provided that the desired signal is at least one dB above the interference level.

Sigfox is a standard originating in France and currently has 75 regions and countries with more than 19 million devices [19]. Sigfox utilizes unlicensed ISM bands and differential-BPSK (D-BPSK) modulation. The message is sent with a fixed bandwidth of 100 Hz and a speed of 100 bps for the uplink, and 200 Hz and 600 bps for the downlink, respectively. This modulation technique belongs to the ultra-narrow band (UNB) modulation. The advantages of using D-BPSK modulation are its high efficiency in the spectrum medium access and ease of implementation. A low bit-rate enables the use of low-cost transceiver components. Sigfox transmits data by changing the frequency three times for each data to ensure data transmission. Sigfox technology has a duty cycle whose restrictions vary within the transmission band from 0.1% to 10%, depending on regional regulations [20].

Wi-SUN is a standard based on IEEE 802.15.4g/4e and deployed in 46 countries, and has more than 100 million devices [21]. IEEE 802.15.4g is an amendment to the IEEE 802.15.4 standard, focusing on SUN communications that play an essential role in the smart grid [22]. The standard specifies several modes that operate in different bands, including the sub-GHz industrial science and medical (ISM) bands. Multirate frequency-shift keying (MR-FSK) with 2-FSK or 4-FSK is the main modulation technique used in Wi-SUN devices. The data rate varies from 2.4 to 200 kbps, depending on the region and frequency band. The mandatory configuration for all regions is 2-FSK, which operates at 50 kbps, implying a channel spacing of 200 kHz. More than 50 million smart meters in Japan that can collect electricity consumption data using this standard have already been deployed. Their number is expected to increase in the future.

The IoT standards described above coexist in the same frequency band called the ISM band [23]. Thus, an increase in the number of IoT devices leads to a significant decrease in network performance because devices following each standard may affect other devices, which will bring collisions to the massive heterogeneous IoT networks [24]. By the MAB-based channel assignment methods that will be present in this paper, approximately 10,000 IoT devices can be accommodated in such coexistence IoT networks.

3. System Model

This section describes the system model and problem formulation. Figure 1 illustrates an IoT network environment where one or more gateways with multiple asynchronous IoT devices are distributed in each of the m IoT networks. In each network, the IoT devices send data to the gateway according to their standards regarding the network configuration, frequency channel to be used, transmission timing, and so on. Each IoT network could not know the standards and the locations of the IoT devices and gateways of the other IoT networks. Therefore, it is difficult to avoid collisions between heterogeneous IoT networks.

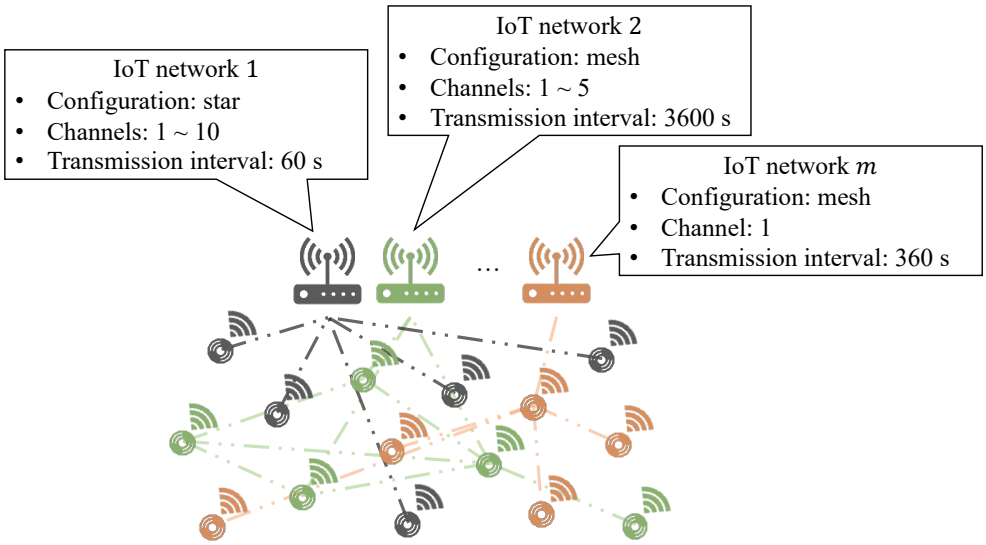


Figure 1. System model of the heterogeneous IoT networks.

Figure 2 shows the channel selection problem of one IoT network in such heterogeneous IoT networks. In the network, K channels are available. A gateway and M IoT devices are associated with the star topology. IoT device sends data to the gateway, and when the gateway receives the data properly, the IoT device will receive an acknowledgment (ACK). This paper defines communication as success when the IoT device receives ACK information. Communication is defined as failure otherwise. If the communication fails, NACK information will be obtained on the IoT device side. For example, node 1 sends data to the gateway using channel 1; communication is successful since no other IoT device is accessing that channel. Hence, an ACK can be received from the gateway. Meanwhile,

node two and node $M-1$ transmit data using channel three simultaneously. The transmissions of node two and node $M-1$ are failures because of the collision between them. In addition, node M transmits using channel k also fails because the node interferes with other IoT networks. In summary, the transmission will be assumed as a failure if two or more than two IoT devices access the same channel simultaneously, no matter which IoT network the devices belong to.

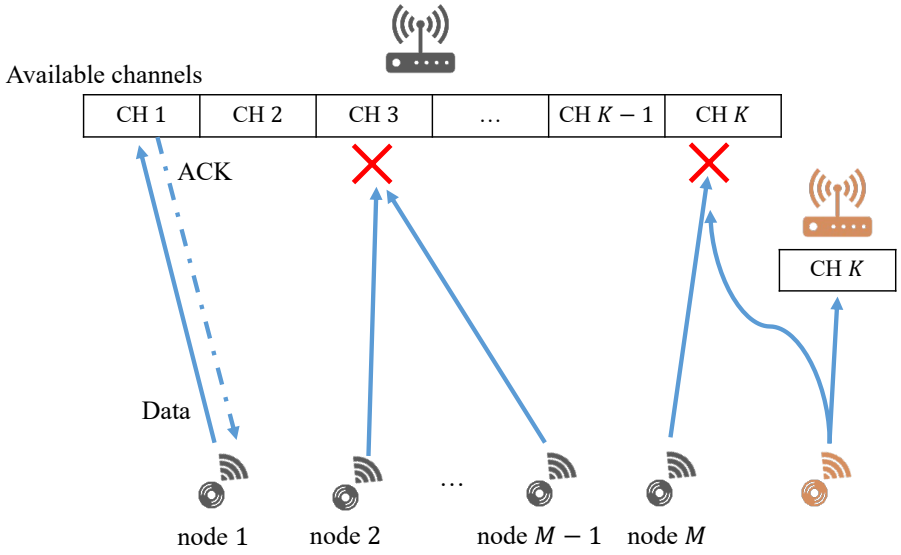


Figure 2. Channel access model.

Figure 3 shows the occupied state of the channel in the time domain. If multiple nodes send data using the same channel simultaneously, a collision will occur, and the communications are assumed to fail. Assume that an IoT battery-powered node will be driven for an extended period without charging. All nodes repeat the wake-up and sleep modes. Assume the sleep time for each IoT device is t. After t. time sleep, a node sends data to the gateway by utilizing the selected channel k based on the MAB methods from the available K channels. It does not matter if the node performs carrier sensing before transmission as long as it complies with the communication standard. The contents of the data and data size depend on the requirement of IoT devices. A node will be in sleep mode after sending and saving the result of receiving the ACK/NACK information.

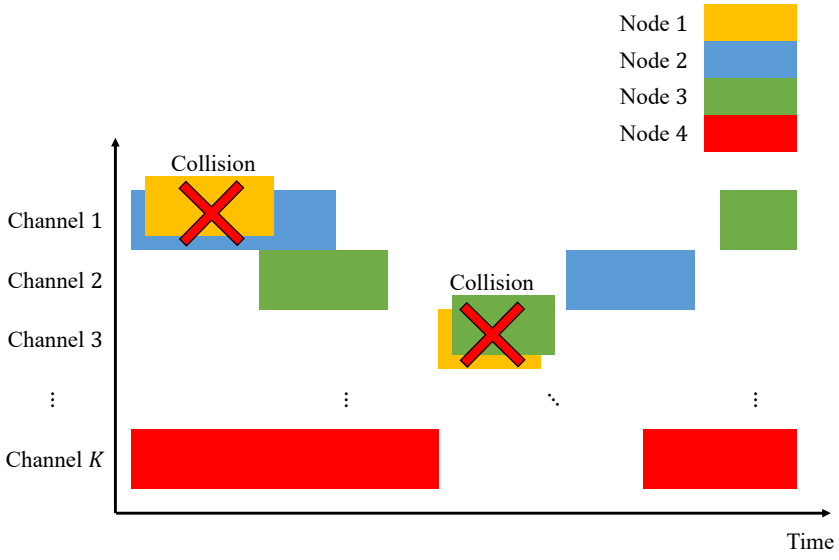


Figure 3. Occupied state of the channels in time domain.

4. Channel Selection Algorithm Based on Multi-Armed-Bandit Algorithm

This section describes channel selection algorithms based on major MAB methods. That is the ε-greedy algorithm, UCB1-tunned algorithm, TOW dynamics algorithm, and the MTOW algorithm. The reason that we investigate these four algorithms is summarized as follows. The epsilon greedy method is the easiest to implement among the multi-armed-bandit (MAB) methods, while the UCB1-tunned can almost get the highest performance among the MAB methods. Moreover, TOW and MTOW proposed by the co-author of this paper can achieve higher performance than UCB1-tuned under certain environments. Hence, we investigated the effects of these four methods. Generally, the performance of the other MAB algorithm is between the UCB1-tunned method and the epsilon greedy method. In the rest of this section, the relationship between the MAB problem and the channel selection problem is presented first. Then, the details of the channel selection method are given.

4.1. Formulation of the Channel Selection as a Multi-Armed-Bandit Problem

The MAB problem [12] is a gambler model that plays multiple slot machines. This problem aims to obtain a strategy to decide which slot machines to be played to earn maximum rewards. Initially, the gambler had no prior information regarding the reward probability of any of the machines. The player gathers information about each slot machine when playing a slot machine. In the MAB problem, resolving the tradeoff between "exploration" to search and identify good slot machines and "exploitation" to play the optimal slot machine and obtain rewards is essential. The MAB problem has various applications, such as in-network advertising, medical fields, network routing, and channel selection. Various algorithms have been proposed that include (among others) ε-greedy [25], Softmax [26], upper confidence bounds (UCB) [27] and UCB1-tuned [28], which is a champion algorithm improved UCB. Besides those well-known MAB methods, a MAB method based on Tug-of-War (TOW) dynamics has been proposed, and the high performance with a small computational cost is shown in [29–33]. Ref. [29] proposes a decision-making mechanism inspired by the behavior of unicellular organisms. Ref. [30] proposes the application of decision-making method based TOW to channel access selection in cognitive radio. Ref. [31] analyzes how to give optimal reward and punishment values in the learning process of TOW. Ref. [32] analyzes individual rewards and social rewards in a competitive environment. Ref. [33] proposes a model in which TOW works on atomic switch.

A channel access model aided by cognitive radio has been proposed in [10,11]. In that model, the authors assume that there are K available channels, and each frame is separated into time slots in the network. The cognitive user selects one channel from $K = \{1, \ldots, k\}$ channels to send data. In this model, the availability probability p_k, which denotes the degree of channel congestion, is defined. Figure 4 illustrates the correspondence between the channel selection problem and the MAB problem when the ACK frame from the gateway is used as a reward. First, the IoT device has no information about the IoT networks, including the state of the available channels. As a node accesses some channels to send data, it gradually learns the communication conditions depending on whether or not ACK frames can be received. This channel selection model is consistent with the MAB problem and can be considered a problem that maximizes the number of successful data transmissions to the gateway.

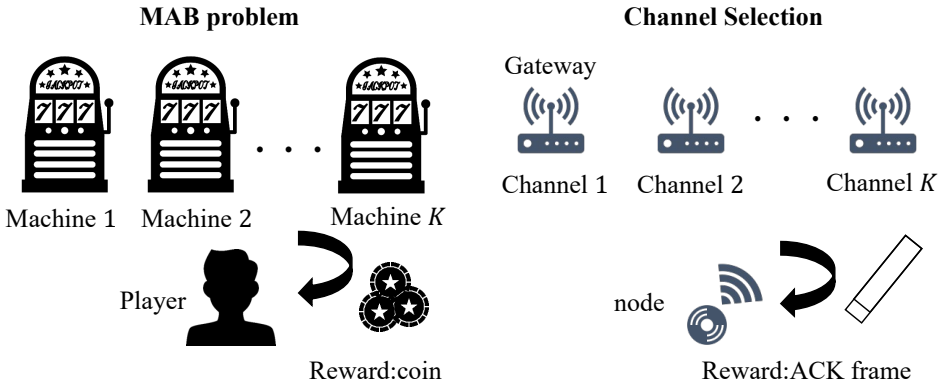


Figure 4. MAB problem vs. channel selection problem.

4.2. Learning Rules of the Multi-Armed-Bandit Algorithms

Figure 5 illustrates a series of flows from the determination of the transmission channel to the data transmission based on the MAB algorithm when each node treats the ACK frame from the gateway as a reward for the MAB problem. The node periodically repeats the wakeup mode for data transmission and learning and the sleep mode for suppressing power consumption. The process shown in the red frame in this flow is the channel-selection process. A node updates the learning parameters such as the number of data transmission attempts N and the number of successful messages R according to whether ACK can be received or not after data transmission. The following sections introduce some major MAB algorithms that can be used as the channel-selection algorithm in this process.

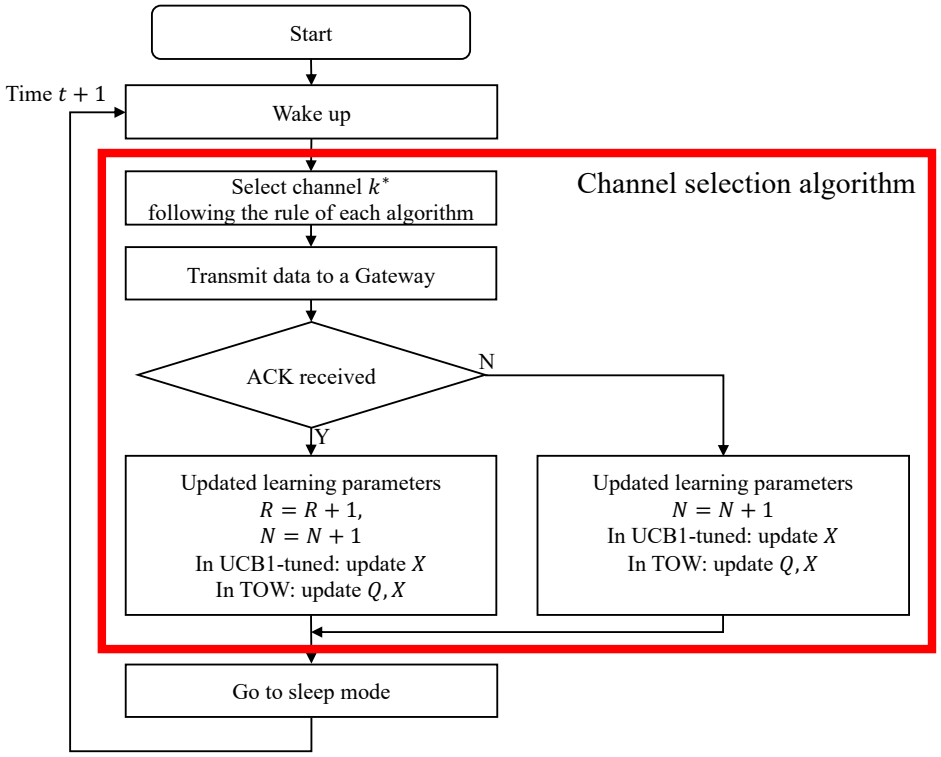


Figure 5. Flowchart of the channel selection.

4.2.1. ϵ-Greedy Based Algorithm

The ϵ-greedy algorithm is a simple method that determines the ratio of exploration to exploitation using the variable ϵ. Therefore, it is widely used as an algorithm for solving MAB problems. In each trial $t = 1, 2, \ldots$, among the K channels, the channel that seems to have the highest reward probability from past experience is selected with a probability

of $1-\epsilon$. Subsequently, another slot is randomly selected with a probability of ϵ. In the ϵ-greedy algorithm, ϵ must be set or adjusted correctly. If the value of ϵ is too low, it will be difficult to find the optimal behavior. On the contrary, the behavior will be close to random and the rewards that can be obtained will be unstable if the ϵ is too high. The estimated reward probability of machine k is expressed as follows:

$$p_k(t) = \frac{R_k(t)}{N_k(t)}, \tag{1}$$

$$k^* = \begin{cases} \arg\max_{k \in K} p_k(t) & \text{if } 1-\epsilon, \\ \text{Randomly selected} & \text{otherwise}, \end{cases} \tag{2}$$

where $N_k(t)$ is the number of transmissions by channel k until time t, that is, the number of times channel k is selected. $R_k(t)$ is the number of rewards gained until time t, that is, the number of successful transmissions and ACKs received. k^* is an index of the channel to be selected at the next time $t+1$, with the highest reward probability at time t. Algorithm 1 describes the ϵ-greedy based channel selection process.

Algorithm 1 ϵ-greedy based channel selection algorithm.

Initialize $R_k(0), N_k(0)$
Set $\epsilon \in [0,1]$
while wake time $t = 1, 2, \ldots,$ is not expired **do**
  Generate a normal random number $x \in [0,1]$
  if $x > \epsilon$ **then**
    Select channel $k^* = \arg\max p_k(t) (\forall k \in K)$
    Transmit the data frame
    Update $N_k(t) + 1$
    if the data frame is transmitted and the corresponding ACK frame is received **then**
      Data transmission succeed
      Update $R_k(t) + 1$
    else
      Data transmission failed
      Update $p_k(t)$ given by Equation (1)
    end if
  else
    Select channel k^* from K
  end if
  $t = t + 1$
end while

4.2.2. Upper Confidence Bounds1-Tuned Based Algorithm

The UCB1-tuned algorithm is the best MAB algorithm with no parameter. It is also widely used in various applications. The characteristic of this algorithm is that it considers the mean value of the reward and the variance value V_k of the number of selections for each channel, as shown in Equation (3). After selecting all channels and transmitting once, to select channel k^* in $t+1$ trial is with the highest UCB value X_k, as represented by (4). Algorithm 2 describes the UCB1-tuned based channel selection process.

$$X_k(t) = \frac{R_k(t)}{N_k(t)} + \sqrt{\frac{\ln t}{N_k(t)} \min(1/4, V_k)}, \tag{3}$$

$$k^* = \arg\max_{k \in K} X_k(t). \tag{4}$$

Algorithm 2 UCB1-tuned based channel selection algorithm.

```
Initialize Q_k(0), R_k(0), N_k(0)
while wake time t = 1, 2, ..., is not expired do
    if t < K then
        Select channel k* = t
    else
        Select channel k* = arg max X_k(t)(∀k ∈ K)
    end if
    Transmit the data frame
    Update N_k(t) + 1
    if the data frame is transmitted and the corresponding ACK frame is received then
        Data transmission succeed
        Update R_k(t) + 1
    else
        Data transmission failed
    end if
    Calculate variance V_k
    Update X_k(t) given by Equation (3)
    t = t + 1
end while
```

4.2.3. Tug-of-War Dynamics-Based Algorithm

TOW dynamics is a reinforcement learning algorithm that rules reward estimates Q_k obtained by (5):

$$Q_k(t) = N_k(t) - (1 + \omega) L_k(t), \tag{5}$$

where $L_k(t)$ is the number of times channel k is selected and transmitted without receiving ACKs until time t, and ω is a weight parameter. Furthermore, (5) can be expressed as follows:

$$Q_k(t) = Q_k(t-1) + \Delta Q_k(t), \tag{6}$$

where ΔQ_k follows the following rule.

$$\Delta Q_k(t) = \begin{cases} +1 & k = k^* \text{and if receiving ACK,} \\ -\omega & k = k^* \text{ and if not receiving ACK,} \\ 0 & k \neq k^*. \end{cases} \tag{7}$$

The displacement $X_k(t)$ at time t, which is used for the decision of the TOW, is expressed using $Q_k(t)$ in (5) and (6) as follows:

$$X_k(t) = Q_k(t-1) - \frac{1}{K-1} \sum_{k' \neq k}^{K} Q_k(t) \tag{8}$$

Reference [31] describes the way to obtain the optimal weight parameter ω. Let us consider the expected value of the reward estimate of machine k from Equation (9).

$$E[Q_k(t)] = \{p_k(t) - \omega(1 - p_k(t))\}N_k(t). \tag{9}$$

This paper describes the highest and second highest reward probabilities of the machine at time t as $p_{1st}(t)$ and $p_{2nd}(t)$, respectively. To ensure that the machine with the highest reward probability is always selected, both of the following two expressions should be satisfied:

$$p_{1st}(t) - \omega(1 - p_{1st}(t)) > 0, \tag{10}$$

$$p_{2nd}(t) - \omega(1 - p_{2nd}(t)) < 0. \tag{11}$$

These equations can be written as follows:

$$p_{1st}(t) < \frac{\omega}{1+\omega} < p_{2nd}(t). \tag{12}$$

It can be confirmed that (13) satisfies (12), and the optimal ω is derived.

$$\frac{\omega}{1+\omega} = \frac{p_{1st}(t) + p_{2nd}(t)}{2}, \tag{13}$$

$$\omega = \frac{p_{1st}(t) + p_{2nd}(t)}{2 - (p_{1st}(t) + p_{2nd}(t))}. \tag{14}$$

The channel to be selected at time $t+1$ is the channel k^*, corresponding to Equation (15).

$$k^* = \arg\max_{k \in K} X_k(t+1). \tag{15}$$

This algorithm has a small calculation cost compared with the champion algorithm UCB, which uses a complicated calculation methodology, such as the square-root calculation.

4.2.4. Modified Tug-of-War

As the channel quality changes dynamically owing to the emergence of mobile nodes or the deployment of new networks, a forgetting parameter α $(0 < \alpha \leq 1)$ is introduced to reduce the influence of past experiences. Subsequently, Equation (5) is modified to (16). $X_k(t)$ is calculated and selects the channel k^* is selected using Equation (15), after it wakes up at the next cycle.

$$Q_k(t) = \alpha Q_k(t-1) + \Delta Q_k(t). \tag{16}$$

The smaller α is, the lesser the effect of past experiences on the current learning state. Algorithm 3 describes the learning process for TOW dynamics and MTOW-based channel selection.

Algorithm 3 TOW dynamics or MTOW-based channel selection algorithm.

Initialize $Q_k(0), R_k(0), N_k(0)$,
while wake time $t = 1, 2, \ldots,$ is not expired **do**
  Select channel $k^* = \arg\max X_k(t) (\forall k \in K)$
  Transmit the data frame
  if the data frame is transmitted and the corresponding ACK frame is received **then**
    Data transmission succeed
    Update $R_k(t) + 1$
    Set $\Delta Q_{k^*}(t) = +1$
  else
    Data transmission failed
    Update $p_k(t)$ given by Equation (1)
    Update $\omega(t)$ given by Equation (14)
    Set $\Delta Q_{k^*}(t) = -\omega(t)$
  end if
  Update $Q_k(t)$ given by Equation (6) in TOW or Equation (16) in MTOW
  Update $X_k(t+1)$ given by Equation (8)
  $t = t + 1$
end while

5. Performance Evaluation

In this section, the performance of the MAB-based channel selection methods is evaluated by simulation. In the simulation, there are two kinds of massive IoT networks where ALOHA communication without channel sensing is adapted. IoT devices in one network transmit data using the channel selected by the MAB-based methods. IoT devices in the other network transmit data using the allocated fixed channel following a two-state Markov model [34]. The two states are ON state and OFF state. During the ON state, IoT devices operate regularly, including wake-up and sleep modes. During the wake-up mode, the IoT device transmits data using the allocated fix channel. During the sleep mode and OFF state, IoT devices keep silent without transmitting data. The state transition probability of the two-state Markov model can be expressed as:

$$P = \begin{pmatrix} \text{ON-ON} & \text{ON-OFF} \\ \text{OFF-ON} & \text{OFF-OFF} \end{pmatrix} = \begin{pmatrix} \frac{1+\lambda_k}{2} & \frac{1-\lambda_k}{2} \\ \frac{1-\lambda_k}{2} & \frac{1+\lambda_k}{2} \end{pmatrix}, \tag{17}$$

where λ_k is a parameter that indicates the ease of transition of the state transitions in channel k and its range is from -1 to +1. This formulation means that the network continues the past state (i.e., ON-ON or OFF-OFF) with a probability of $\frac{1+\lambda_k}{2}$ and transitions from the past state to a different state (i.e., ON-OFF or OFF-ON) with a probability of $\frac{1-\lambda_k}{2}$. The transmission is a failure if two or more IoT devices simultaneously access the same channel. This paper verifies the effectiveness of applying the MAB algorithm such as ϵ-greedy, UCB1-tuned, and basic TOW and MTOW to the channel selection problem in a massive heterogeneous IoT network and evaluates the network performance in terms of FSR. The common simulation settings are summarized in Table 1. Note that the value of ϵ is set to 0.1 in the simulation.

Table 1. Simulation settings.

Simulation time [s]	10,000
Number of nodes M	100, 1000, 10,000
Number of available channels K	15, 30, 60
Duty cycle	0.01, 0.1
Number of load channels	1/5, 2/5, 3/5, 4/5 of K
Load duty cycle [%]	0.1, 0.5, 0.9
Duration of each state [s]	100
λ	0.8
Forgetting parameter α	0.95
ϵ	0.10

Figure 6 shows an example of channel selection based on the TOW method. In the simulation, the numbers of IoT nodes and channels are set to 10 and 15, respectively. Among the 15 channels, 12 channels are loaded by the other IoT network. The orange line in Figure 6 indicates the loaded channel and the loaded time by the IoT devices of the other IoT network. The channel access situation of the 10 IoT devices in the IoT network accessing channels based on TOW algorithms are shown at different points. Simulation results show that the IoT devices can avoid accessing the channels loaded by other IoT networks. The reason is that IoT devices can select the channel with the highest probability based on the TOW method. The communication fails when IoT device selects the channel that are loaded by other IoT networks, which reduces the probability of the selected channel. By iterative selection and the update of the probability parameter corresponding to each channel, IoT devices can select the channels without load.

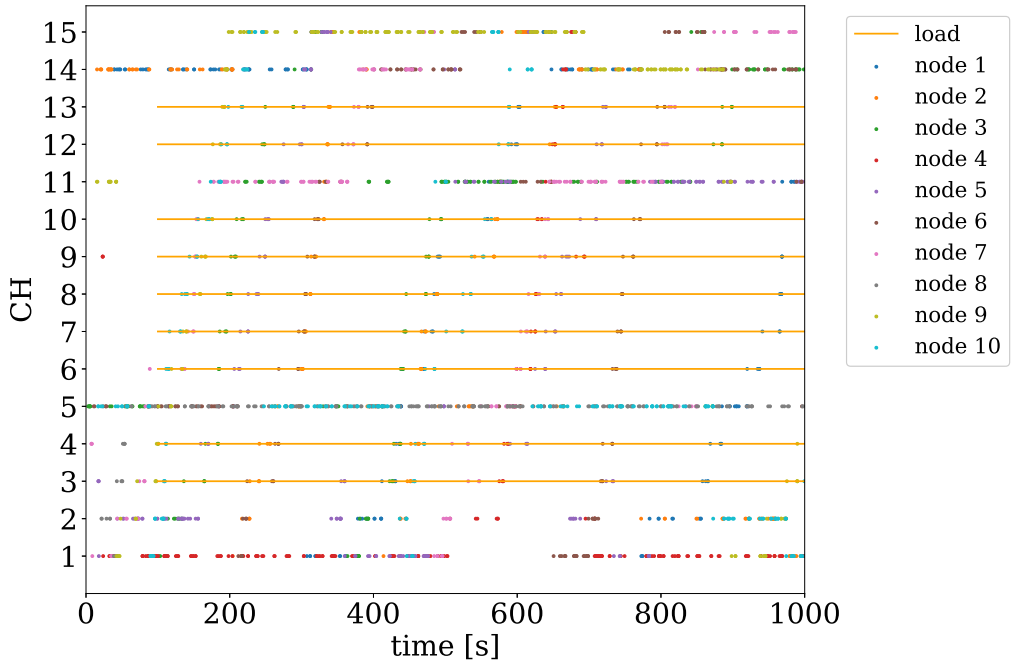


Figure 6. Example of channel selection based on TOW method in the heterogeneous IoT networks.

Figure 7 illustrates the relation between the FSR and the number of IoT nodes. In this simulation, the number of IoT devices varies from 10^2 to 10^4. The number of channels is set to 20. The parameter related to the state transition probability of the IoT devices in the network where IoT devices transmit data using the allocated fixed channels λ is set as 0.8. The duty ratio is set to 0.5. From the simulation results, it can be observed that as the number of nodes increases, FSR decreases. The reason is that with the increase of the IoT devices, the collisions among IoT devices in the same IoT network increase. Even though the FSR decreases with the increase of the number of IoT devices, the FSR is higher than 90% for the TOW/MTOW channel selection method when the number of IoT devices is 10^4. The reason is that the TOW/MTOW-based channel selection method can select the channel with highest available probability. Hence, the TOW/MTOW is effective for the massive heterogeneous IoT networks.

Figure 8 illustrates the relation between the FSR and the duty cycle of the IoT nodes. In this simulation, the numbers of IoT nodes and channels are set to 10,000 and 30. Among the 30 available channels, 12 channels may be loaded by the IoT devices in the other network. λ and the duty ratio are set to 0.8 and 0.5, respectively. The simulation results show that the FSR decreases with the increase of the duty ratio. The reason is that the transmission interval becomes shorter with the increased duty ratio, which will increase the collision probability among IoT devices. Moreover, the MTOW can obtain the highest FSR. The reason is that channel selection-based MTOW can select the channel with the highest available probability. In addition, the MTOW method is more adaptable to dynamic environments due to the introduction of the forgetting parameter, which is introduced to reduce the influence of past experiences.

Figure 9 illustrates the relation between the FSR and the number of available channels. In this simulation, the number of IoT nodes is set to 10,000. The number of channels that may be loaded by the IoT devices in the other IoT network is set to 12. The transition parameter λ and the duty ratio are set to 0.8 and 0.5. Simulation results show that the FSR increases with the number of available channels. The reason is that with the increase of the number of available channels, the average number of IoT nodes assigned to each channel is decreased, which reduces the probability of collisions. In addition, the MTOW can get the highest FSR no matter how many available channels there are for the IoT network. The reason is that the introduction of the forgetting parameter can make the method more adaptable to a dynamic environment.

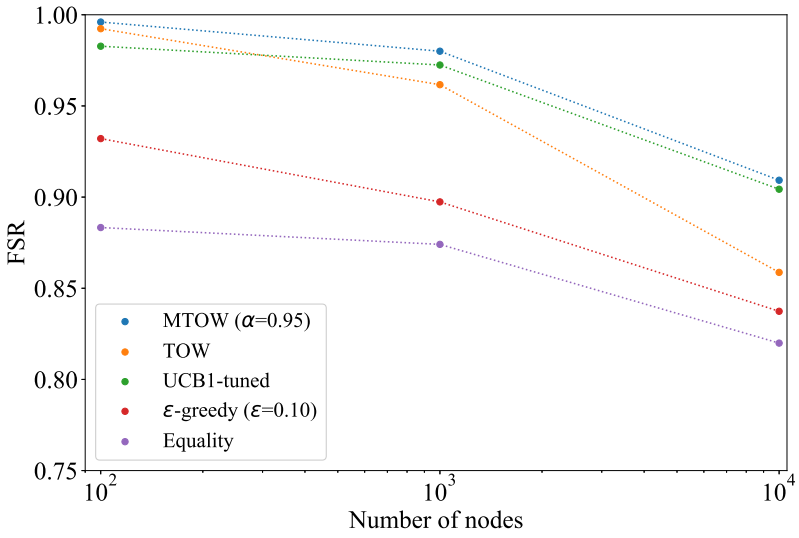


Figure 7. Number of nodes vs. FSR in the IoT network, where 30 available channels and 12 channels may be occupied by the IoT devices from the other network with $\lambda = 0.8$ and a duty ratio of 0.5.

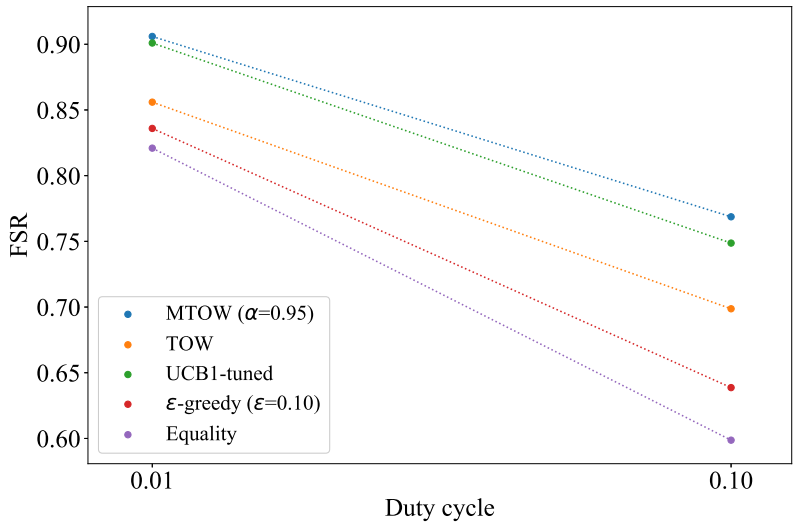


Figure 8. Duty cycle vs. FSR in the IoT network where there are 10,000 nodes and 30 available channels, while 12 channels may be loaded by the IoT devices of the other kind of network with a communication frequency of $\lambda = 0.8$ and a duty ratio of 0.5.

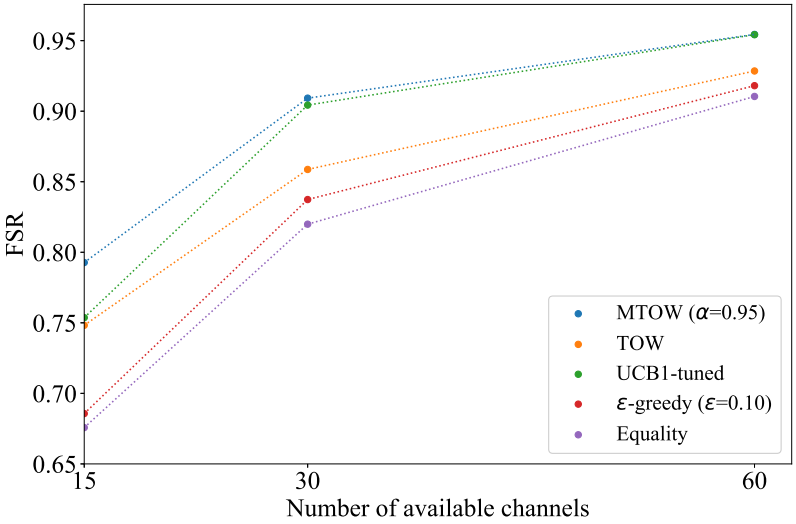
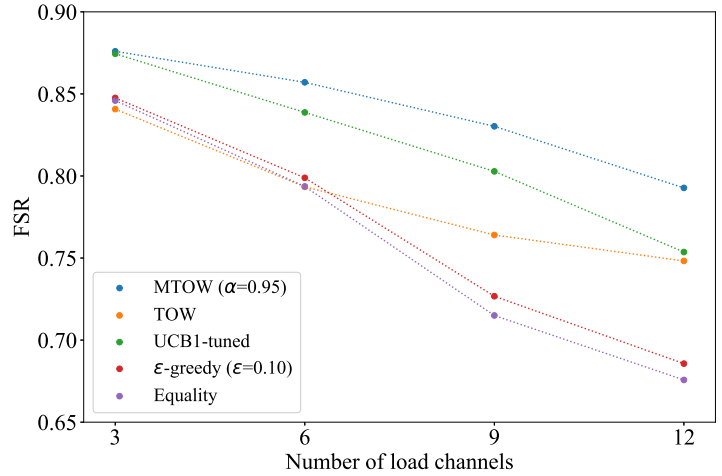
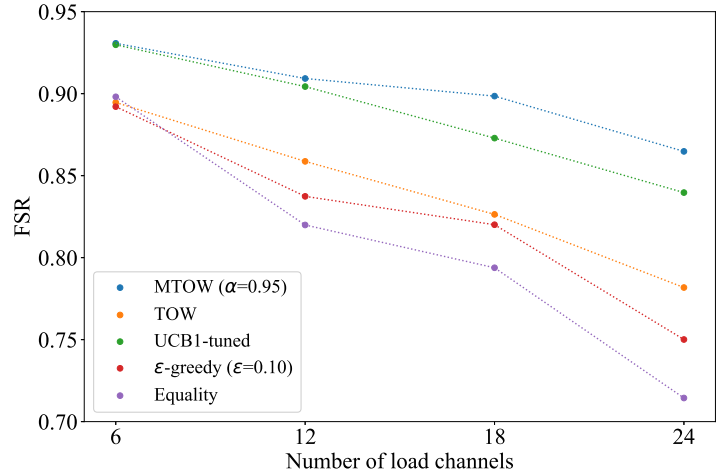


Figure 9. Number of available channels vs. FSR in the IoT network where 10,000 nodes and 12 channels may be loaded.

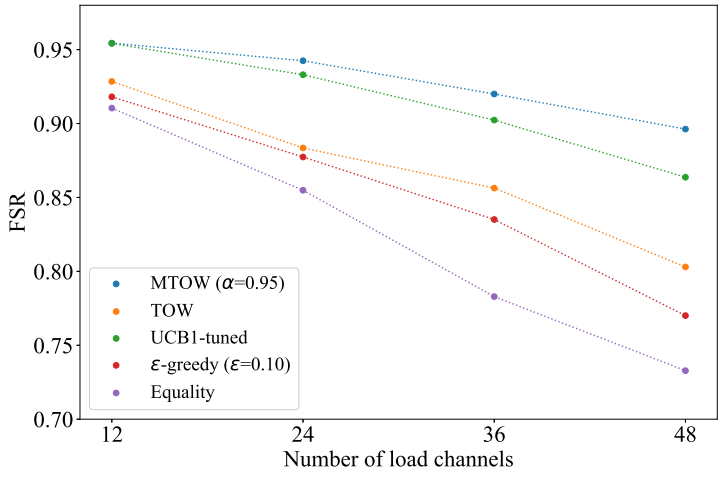
Figure 10 illustrates the relation between the FSR and the numbers of available and loaded channels. Figure 10a–c show the FSR when the number of available channels are set to 15, 20, and 60, respectively. The number of channels that may be loaded varies from 20% to 80% of the available channels. The numbers of duty ratio and IoT nodes are set to 0.5 and 10,000, respectively. The transition parameter λ is set to 0.8. Simulation results show that the FSR decreases with the increases of the loaded channels. The reason is that collisions among different IoT networks increase when the number of load channels becomes larger. In addition, MTOW can get much better FSR than the other methods when the number of loaded channels increases, which shows that MTOW may be more effective for heterogeneous IoT networks with a larger number of IoT devices. Hence, MTOW could be applied for the heterogeneous IoT network with a higher congestion degree, while UCB1-tuned/MTOW could be used for the heterogeneous IoT network with a lower congestion degree.



(**a**)



(**b**)



(**c**)

Figure 10. Number of load channels vs. FSR in the IoT network where the number of IoT nodes is 10,000 with $\lambda = 0.8$ and a duty ratio of 0.5. (**a**) Number of load channels vs. FSR when the number of available channels is 15. (**b**) Number of load channels vs. FSR when the number of available channels is 30. (**c**) Number of load channels vs. FSR when the number of available channels is 60.

The experimental results described above show that MTOW can achieve the highest-performance in FSR compared to the other three methods. These results have also been evaluated in [16]. Moreover, the theoretical analysis in [30] also shows that regret, an

indicator of how much loss was made from the appropriate choices, is smaller for the TOW-based algorithm than that for the UCB1-tuned algorithm. From the experimental results of this paper, it is clarified that MTOW-based channel selection algorithm works properly in a dynamic environment with a huge number of IoT nodes where competition caused by other coexistence IoT exists.

6. Implementation and Performance Evaluation of the Multi-Armed-Bandit-Based Channel Selection Methods on Internet of Things Devices

In this section, the performance of the channel selection-based MAB algorithms simulated in Section 5 is implemented and evaluated using the actual IoT devices. Wi-SUN IoT devices and LoRa are used as the evaluation IoT network and the interference IoT network, respectively. For the Wi-SUN IoT devices, Lazurite 920J, which supports IEEE 802.15.4g/4e standard and can communicate using the 920 MHz band, is used. Table 2 lists the specification of the Lazurite 920J.

Table 2. The specification of the Lazurite 920J.

Operating voltage [V]	1.8–3.3
Operating frequency	16 MHz (operation) 32.768 kHz (sleep)
Standby current [μA]	7
Operating current [mA]	4 (not using radio) 25 (using radio)
RAM	6 KB
ROM	64 KB

Figure 11 shows the settings of the IoT devices in the experiment. The heterogeneous IoT network is constructed in a 13 m × 7 m square area. A Wi-SUN IoT network consists of 50 Wi-SUN devices (i.e., Lazurite) and a gateway is deployed in this area. Meanwhile, a LoRa network composed of 5 LoRa devices is also deployed in the area. Wi-SUN devices select the access channels based on the MAB algorithms. A Raspberry Pi controls each Wi-SUN device. The MAB algorithms are implemented on the Raspberry Pi to calculate the access channel and control the Wi-SUN device to transmit data using the selected channel. LoRa devices transmit data using channel 34 every 20 min. The transmission interval for the LoRa device is 10 s.

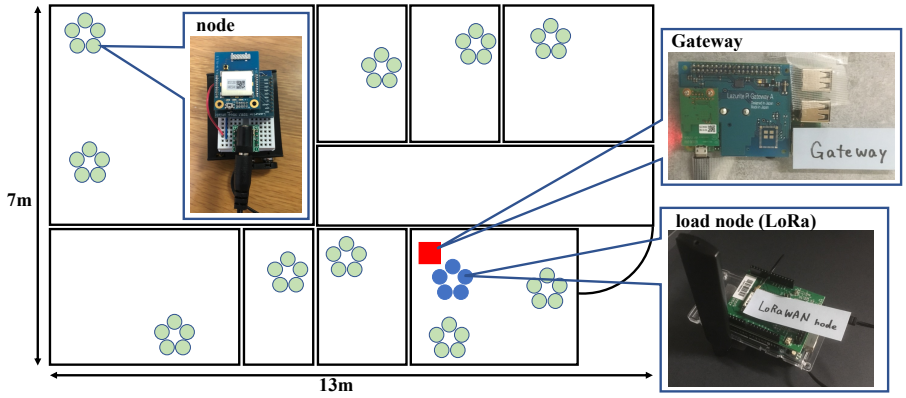


Figure 11. The layout of the IoT devices in the experiment.

Table 3 summarizes the experimental settings.

Table 3. The experimental settings.

Experiment time [s]	7200
Transmission power [mW]	20
Bit rate [kbps]	50
Channel (center frequency [MHz])	34 (922.6), 37 (923.2), 40 (923.8)
Number of IoT node M	50
Number of available channels K	3
Sleep interval of IoT node [s]	1.0
Number of load node (LoRa)	5
Load channel (center frequency [MHz])	34 (922.6)
Sleep interval of load node [s]	10.0

Figure 12 shows the experimental results in FSR of the MAB-based channel selection methods, i.e., ϵ-greedy based, UCB-1tuned based, TOW-based, and MTOW-based channel selection method. Wi-SUN devices select their access channel among three channels, i.e., CH34, CH37, and CH40, using the MAB-based channel selection methods. Figure 12 shows that the MTOW can achieve the highest FSR. In addition, the fairness index (FI) values for ϵ-greedy based, UCB-1tuned based, TOW-based, and MTOW-based channel selection method are 0.997, 0.995, 0.974, and 0.998, respectively. Hence, The MTOW-based channel selection method is superior to the other MAB-based channel selection methods in FI and FSR under both simulation and implementation. The experimental results also verify the effectiveness of the MTOW algorithm in dynamic environments, which introduces the forgetting parameter to adapt to the dynamic environment.

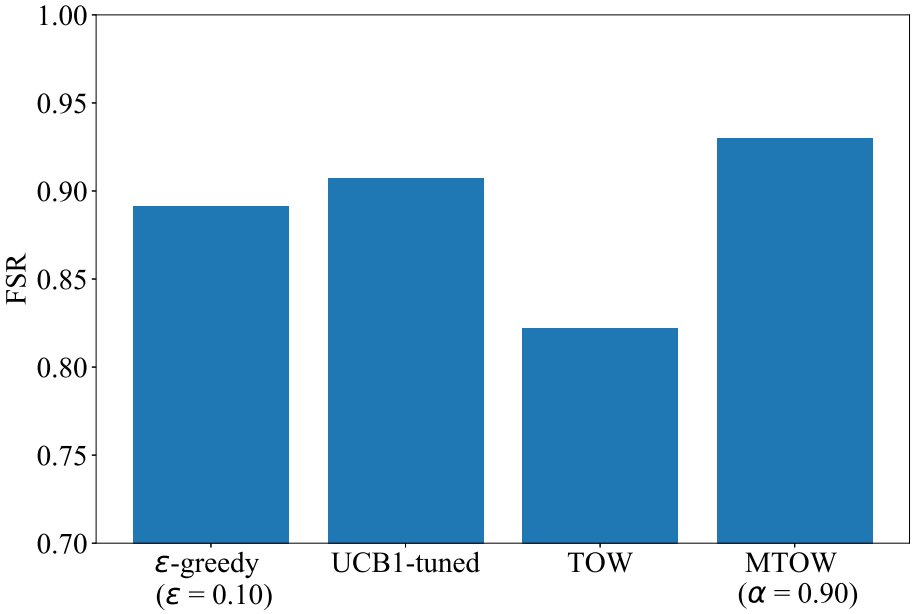


Figure 12. Implementation results in terms of FSR for the MAB-based channel selection methods.

7. Conclusions

In this paper, the effectiveness of the MAB-based autonomous decentralized channel selection methods for massive heterogeneous IoT networks is evaluated. Specifically, the FSR can reach 95% when the numbers of channels and IoT devices are 60 and 10,000, respectively, while 12% channels are suffering traffic load by other kinds of IoT devices.

In addition, the performance in terms of FSR improves by 20% compared to the equality channel allocation. In addition, the MAB algorithm, termed MTOW, can achieve the highest FSR in any setting. Moreover, the performance of the MAB-based channel selection methods is implemented and evaluated on IoT devices. Experimental results show that the MTOW-based channel selection method can achieve the highest FSR and FI, i.e., around 0.95, and 0.998, respectively. In summary, the MTOW-based channel selection method can get high FSR either in simulation or experiments. As described above, due to the high FSR and FI, and easy application concerning the practical IoT devices, the MTOW-based channel selection method may become an efficient technique to support massive IoT applications, such as smart city, personal IoT, Smart grid, industrial assets monitoring, agriculture, and many other applications [35] in the next generation of wireless communication networks. In our further work, we will investigate the impact of the parameters in the MAB algorithm on the performance for massive heterogeneous IoT networks. For instance, the ϵ value in ϵ-greedy algorithm, the α in MTOW algorithm, and so on. Moreover, the time consumption of the evaluated algorithms in this paper will be experimentally analyzed later.

Author Contributions: conceptualization, S.H., A.L., S.-J.K., Y.S. and M.H.; methodology, S.H., A.L., S.-J.K. and M.H.; software, S.H., R.K.; validation, S.H.; formal analysis, S.H., A.L., S.-J.K., M.H.; investigation, S.H., A.L., M.H.; resources, S.H.; data curation, S.H., R.K.; writing—original draft preparation, S.H.; writing—review and editing, S.H., A.L., S.-J.K., Y.W., Y.S. and M.H.; visualization, A.L., M.H.; supervision, A.L., S.-J.K., Y.S. and M.H.; project administration, M.H.; funding acquisition, M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by JSPS KAKENHI Grant Numbers JP22H01493 to M.H. and JP22K14263 to A.L.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. International Data Corporation (IDC), IDC Media Center. Available online: https://www.idc.com/getdoc.jsp?containerId=prUS45213219 (accessed on 19 November 2019).
2. Nguyen, D.C.; Ding, M.; Pathirana, P.N.; Seneviratne, A.; Li, J.; Niyato, D.; Dobre, O.; Poor, H.V. 6G Internet of Things: A Comprehensive Survey. *IEEE Internet Things J.* **2022**, *9*, 359–383. [CrossRef]
3. IoT Telekom, NB-IoT, LoRaWAN, Sigfox: An Up-to-Date Comparison. Available online: https://dt.iotsolutionoptimizer.com/LoadDocument/3047/NB-IoT,LoRaWAN,Sigfox20-20An20Up-to-date20Comparison.pdf (accessed on 4 February 2021).
4. Chincoli, M.; Boef, P.D.; Liotta, A. Cognitive channel selection for wireless sensor communications. In Proceedings of the 2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC), Calabria, Italy, 16–18 May 2017.
5. *IEEE Std 802.15.4-2011*; IEEE Standard for Local and Metropolitan Area Networks—Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs) Amendment 1: MAC Sublayer. IEEE: Piscatvey, NJ, USA, 2012.
6. Aihara, N.; Adachi, K.; Takyu, O.; Ohta, M.; Fujii, T. Q-Learning Aided Resource Allocation and Environment Recognition in LoRaWAN With CSMA/CA. *IEEE Access* **2019**, *7*, 152126–152137. [CrossRef]
7. Wu, C.M.; Wu, M.S.; Yang, Y.J.; Sie, C.Y. Cluster-Based Distributed MAC Protocol for Multichannel Cognitive Radio Ad Hoc Networks. *IEEE Access* **2019**, *7*, 65781–65796. [CrossRef]
8. Macaluso, I.; Finn, D.; Ozgul, B.; DaSilva, L.A. Complexity of Spectrum Activity and Benefits of Reinforcement Learning for Dynamic Channel Selection. *IEEE J. Sel. Areas Commun.* **2013**, *31*, 2237–2248. [CrossRef]
9. Zhu, J.; Song, Y.; Jiang, D.; Song, H. A New Deep-Q-Learning-Based Transmission Scheduling Mechanism for the Cognitive Internet of Things. *IEEE Internet Things J.* **2018**, *5*, 2375–2385. [CrossRef]
10. Lai, L.; Jiang, H.; Poor, H.V. Medium access in cognitive radio networks: A competitive multi-armed bandit framework. In Proceedings of the IEEE 42nd Asilomar Conference on Signals, System and Computers, Pacific Grove, CA, USA, 26–29 October 2008; pp. 98–102.
11. Lai, L.; Gamal, H.E.; Jiang, H.; Poor, H.V. Cognitive Medium Access: Exploration, Exploitation, and Competition. *IEEE Trans. Mob. Comput.* **2011**, *10*, 23–253.
12. Robbins, H. Some aspects of the sequential design of experiments. *Bull. Am. Math. Soc.* **1952**, *58*, 527–535. [CrossRef]
13. Shi, C.; Shen, C. Multi-player Multi-armed Bandits with Collision-Dependent Reward Distributions. *IEEE Trans. Signal Process.* **2021**, *69*, 4385–4402. [CrossRef]
14. Ma, J.; Hasegawa, S.; Kim, S.-J.; Hasegawa, M. A Reinforcement-Learning-Based Distributed Resource Selection Algorithm for Massive IoT. *Appl. Sci.* **2019**, *9*, 3730. [CrossRef]
15. Abdelghany, A.; Uguen, B.; Moy, C.; Lemur, D. Decentralized Adaptive Spectrum Learning in Wireless IoT Networks based on Channel Quality Information. *IEEE Internet Things J.* **2022**. [CrossRef]

16. Yamamoto, D.; Furukawa, H.; Li, A.; Ito, Y.; Sato, K.; Oshima, K.; Hasegawa, S.; Watanabe, Y.; Shoji, Y.; Kim, S.J.; et al. Performance Evaluation of Reinforcement Learning Based Distributed Channel Selection Algorithm in Massive IoT Networks. *IEEE Access* **2022**, *10* 67870–67882. [CrossRef]
17. LoRa Alliances. Available online: https://lora-alliance.org/ (accessed on 25 April 2022).
18. Centenaro, M.; Vangelista, L.; Zanella, A.; Zorzi, M. Long-range communications in unlicensed bands: The rising stars in the IoT and smart city scenarios. *IEEE Wirel. Commun.* **2016**, *23*, 60–67. [CrossRef]
19. Sigfox, OUR STORY. Available online: https://www.sigfox.com/en/sigfox-story (accessed on 25 April 2022).
20. Lavric, A.; Petrariu, A.I.; Popa, V. Long Range SigFox Communication Protocol Scalability Analysis Under Large-Scale, High-Density Conditions. *IEEE Access* **2019**, *7*, 35816–35825. [CrossRef]
21. Wi-SUN Alliances. Available online: https://wi-sun.org/ (accessed on 25 April 2022).
22. *IEEE Std 802.15.4-2012*; IEEE Standard for Local and Metropolitan Area Networks—Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs) Amendment 3: Physical Layer (PHY) Specifications for Low-Data-Rate, Wireless, Smart Metering Utility Networks. IEEE: Piscatvey, NJ, USA, 2012.
23. Chen, M.; Miao, Y.; Jin, X.; Wang, X.; Humar, I. Cognitive-LPWAN: Towards intelligent wireless services in hybrid low power wide area networks. *IEEE Trans. Green Commun. Netw.* **2019**, *3*, 409–417. [CrossRef]
24. Lin, F.; Chen, C.; Zhang, N.; Guan, X.; Shen, X. Autonomous channel switching: Toward efficient spectrum sharing for industrial wireless sensor networks. *IEEE Internet Things J.* **2016**, *4*, 231–243. [CrossRef]
25. Sutton, R.; Barto, A. *Reinforcement Learning: An Introduction*; The MIT Press: Cambridge, MA, USA, 1998.
26. Vermorel, J.; Mohri, M. Multi-armed Bandit Algorithms and Empirical Evaluation. In Proceedings of the 16th European Conference on Machine Learning, Porto, Portugal, 3–7 October 2005; pp. 437–448.
27. Lai, T.L.; Robbins, H. Asymptotically efficient adaptive allocation rules. *Adv. Appl. Math.* **1985**, *6*, 4–22. [CrossRef]
28. Auer, P.; Cesa-Bianchi, N.; Fischer, P. Finite-time Analysis of the Multiarmed Bandit Problem. *Mach. Learn.* **2002**, *47*, 235–256. [CrossRef]
29. Kim, S.-J.; Aono, M.; Hara, M. Tug-of-war model for the two-bandit problem: Nonlocally-correlated parallel exploration via resource conservation. *BioSystems* **2010**, *101*, 29–36. [CrossRef]
30. Kim, S.-J.; Aono, M. Amoeba-inspired algorithm for cognitive medium access. *Nonlinear Theory Its Appl.* **2014**, *5*, 198–209. [CrossRef]
31. Kim, S.-J.; Aono, M.; Nameda, E. Efficient decision-making by volume-conserving physical object. *New J. Phys.* **2015**, *17*, 083023. [CrossRef]
32. Kim, S.-J.; Naruse, M.; Aono, M. Harnessing the Computational Power of Fluids for Optimization of Collective Decision Making. *Philosophies* **2016**, *1*, 245–260. [CrossRef]
33. Kim, S.-J.; Tsuruoka, T.; Hasegawa, T.; Terabe, K.; Aono, M. Decision maker based on atomic switches. *AIMS Mater. Sci.* **2016**, *3*, 245–259. [CrossRef]
34. Sengottuvelan, S.; Ansari, J.; Mähönen, P.; Venkatesh, T.G.; Petrova, M. Channel Selection Algorithm for Cognitive Radio Networks with Heavy-Tailed Idle Times. *IEEE Trans. Mob. Comput.* **2017**, *16*, 1258–1271. [CrossRef]
35. Raza, U.; Kulkarni, P.; Sooriyabandara, M. Low power wide area networks: An overview. *IEEE Commun. Surv. Tutor.* **2017**, *19* 855–873. [CrossRef]