Reinforcement-Learning-Based Spatial Resource Identification for IoT D2D Communications

Woongsoo Na, Nhu-Ngoc Dao, and Sungrae Cho

Abstract—**The exponential growth of the Internet of Things (IoTs) has led to an increasing demand for intelligent IoT Devices (IoTDs), requiring innovative network capacity expansion. Recently, several research has been conducted on the identification of hidden network resources for network capacity expansion. However, the spatial resource identification scheme through the omni-directional antenna has limitations in terms of frequency efficiency compared to the scheme with the directional antenna. In this paper, we propose a directional spatial-resource identification technique for device-to-device (D2D) communication. To find the optimal identification parameters, we design the objective function and apply a reinforcement learning. The training data used for reinforcement learning is collected in each report phase, and Q-learning is applied to find the optimal beam set. Furthermore, based on the obtained frequency information, we propose a contention-based D2D communication scheme. The proposed contention-based D2D communication scheme can efficiently solve the deafness problem occurring in a directional D2D communication. Finally, we perform a simulation using OPNET to measure the performance and evaluate the effectiveness of the proposed technique. The simulation results show that the proposed schemes realize a better performance than the existing schemes proposed in previous works in terms of energy efficiency, frequency efficiency, aggregate network throughput, and deafness duration.**

Index Terms—**IoT networks, directional identification, D2D communication, Radio Resource Harvesting Edge.**

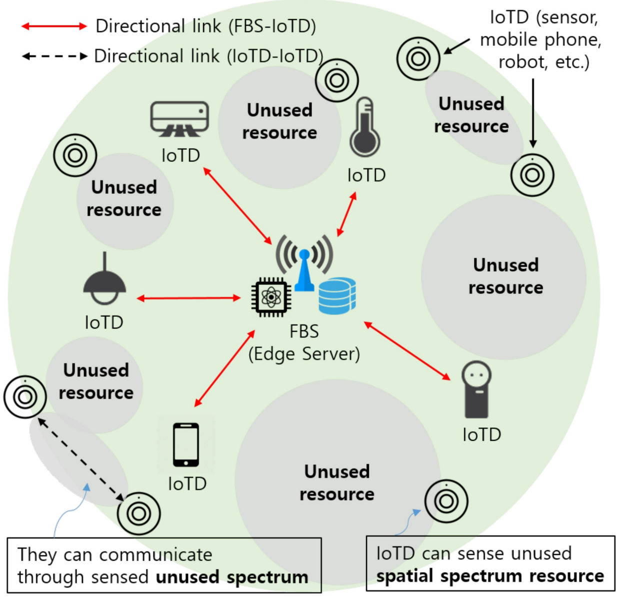

Directional link (FBS-IoTD)
Directional link (IoTD-IoTD)
IoTD (sensor, mobile phone, robot, etc.)
Unused resource
IoTD
Unused resource
IoTD
Unused resource
Unused resource
Unused resource
IoTD
FBS (Edge Server)
Unused resource
IoTD
IoTD
Unused resource
They can communicate through sensed **unused spectrum**
IoTD can sense unused **spatial spectrum resource**

Fig. 1. Example of IoT networks (IoTDs perform D2D communication by identifying unused spatial-frequency resources).

I. Introduction

THE proliferation of IoT devices (IoTDs) has resulted in the exponential growth of wireless data traffic in Internet of Things (IoTs) networks [1], [2]. Existing wireless access methods that employ macro base stations (MBSs) are incapable of accommodating such gargantuan data demands because of the poor signal quality received by IoTDs that are located indoors or at the cell boundaries [3], [4]. As a result, the deployment of femto base stations (FBSs) is considered a viable solution for providing IoTDs with better signal quality [5]. In this architecture, the macro cells offload data through femto-cell networks to the IoTDs [6]; this architecture can facilitate efficient spectrum sharing between IoTDs. In addition, the FBS, which has low deployment cost and flexible configuration ability, can achieve more efficient spectrum sharing by using the spectral information collected from the surrounding IoTDs.

Fig. 1 presents an example of an IoT network model. As shown in the figure, an FBS (or edge server) is deployed in the network topology. In this architecture, the IoTDs transmit data to/from the FBSs over licensed bands. However, certain spatial frequency resources remain unused by the IoTDs, which results in spectrum inefficiency. Identification of these unused spatial resources can improve the spectrum efficiency, and the IoTDs can utilize these frequency resources for infrastructureless Device-to-device (D2D) communication.

Literature reviews have shown that the majority of previous spectrum-identification techniques [7]–[12] used omni-directional antennas in the field of cognitive radio networks (CRNs) or spectrum agile communications. However, it was observed that omni-directional identification is inefficient when compared with directional identification in terms of the spectrum efficiency. Furthermore, as the use of millimeter-wave (mmWave) bands has recently increased to expand network capacity, beamforming using directional antennas has become a necessity [13].

Fig. 2 presents the benefits of directional identification over omni-directional identification. As shown in the figure, the directional-identification-scheme can identify the spectrum over a longer range with the same energy budget and can provide fine-grained identification, which omni-directional identification techniques cannot. As a result, a directional-identification-based scheme can be used to identify the direction of IoTDs as well as hidden spatial-spectrum resources. For instance, in Fig. 2 (a), an IoTD uses omni-directional identification to identify channel 1 and other devices existing within the identification range. Therefore, it cannot use channel 1. In contrast, the directional-identification-based scheme can be used to identify the location of the IoTDs and unused spatial-spectrum resources (indicated by yellow areas in Fig. 2 (b)).

W. Na is with the Division of Computer Science and Engineering, Kongju National University, Cheonan, Korea.

N.-N. Dao is with the Department of Computer Science and Engineering, Sejong University, Seoul, Korea.

S. Cho is with the School of Computer Science and Engineering, Chung-Ang University, Seoul, Korea 156-756.

S. Cho is the corresponding author (E-mail: srcho@cau.ac.kr).

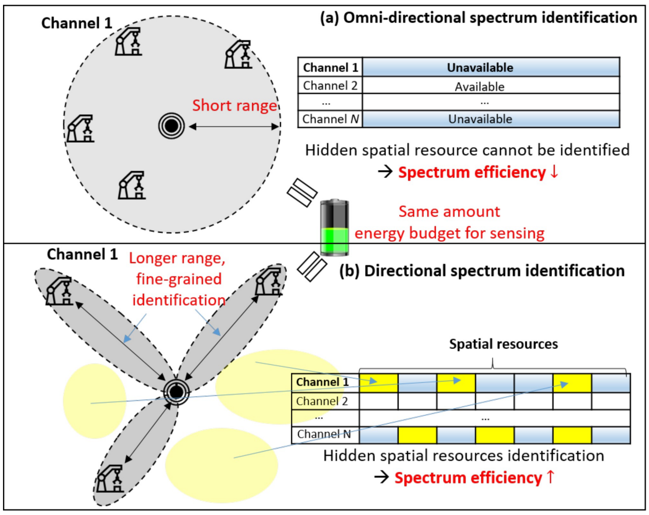


Fig. 2. Omni-directional identification versus directional identification.

Therefore, the IoTDs can communicate through these unused spatial resources.

To overcome the shortcomings of the omni-directional identification technique, a centralized directional spectrum sensing technique was studied in [14]. In [14], secondary users (SUs) performed spectrum sensing with sensing parameters such as the sensing period, power, channel, and beams. The parameters were periodically updated by the central entity of the SUs. Furthermore, a non-linear optimization technique (for sensing periods and power) and a heuristic algorithm (for channels and beams) were used to calculate the optimal sensing parameters. However, although the proposed heuristic-based algorithm efficiently finds the hidden-frequency space, it does not consider the energy consumption for the sensing; thus, the energy consumption of the SUs tended to be unbalanced.

Therefore, we propose a machine-learning-based spatial-resource identification/harvesting scheme[1] for IoT networks. Based on the success of the previous study in terms of efficient cooperative spectrum sensing in the cognitive radio domain, this study is focused on spectrum identification/harvesting for D2D communication with the objective of enhancing the spectrum efficiency. There have been studies that applied reinforcement learning to the cognitive radio domain [15]–[19]. In their studies, reinforcement learning was performed on licensed frequency information such as noise power and primary user resident time in order for SU to access the spectrum in an opportunistic manner [18], [19]. On the other hands, in a few studies, reinforcement learning is used to find optimal identification parameters [15]–[17]. However, since their study assumed an omni-directional antenna environment, the optimal beam selection was not considered. In this paper, using reinforcement learning (RL), the proposed identification/harvesting technique is used to find an energy-efficient beam set. The Radio Resource Harvesting Edge (RRHE) learns the reward according to the beam set identified by each IoTD and selects the beam set that produces the maximum gain for the future. The contributions of this paper can be summarized as follows:

- We propose a directional spectrum identification/harvesting technique for massive IoT networks. Existing resource identification studies were based on omni-directional antennas, the proposed scheme is a study to identify resources using a directional antenna in mmWave environment. The proposed technique is aimed at designing optimal beam-identification parameters through the use of RL. To find the optimal beam-identification parameters, we design the state, action, and reward functions for the Q-learning algorithm.
- To improve communication reliability, we integrated our contention-based D2D communication scheme [21] between the IoTDs through the identified hidden spatial-spectrum resources. In [21], we proposed a D2D communication technique for solving the issue of deafness by using two antennas. However, in this study, we reinforced the D2D communication scheme used to solve the issue of deafness using one data channel by applying non-orthogonal multiple access (NOMA). Furthermore, we analytically proved that the D2D communication scheme does not incur any deafness.
- We showed that our proposed scheme outperforms the existing techniques in terms of various performance indices, including energy efficiency, frequency efficiency, aggregate network throughput, and deafness duration.

The remainder of this paper is organized as follows. Section II presents the system model of IoT networks. Section III describes the proposed RL-based spectrum identification/harvesting technique. Section IV presents the description and analysis of the deafness-free D2D communication scheme. Section V presents an evaluation of the performance of the proposed algorithms using simulations. Finally, the conclusions of this study are presented in Section VI.

II. System Model

A. Basic Assumption

The overall system model is illustrated in Fig. 3. We consider a network model that consists of IoTDs, and a RRHE. In addition, we consider that IoTDs have three phases: *identification phase*, *report phase*, and *transmission phase*. In each phase, the IoTD performs the following actions:

- *Identification phase*: Each IoTD detects signals within the sensing range of IoTD based on an energy-detection scheme and identifies empty spatial resources. The proposed identification scheme used the optimized identification parameters and stores the location of the detected IoTD in a local connectivity matrix.
- *Report phase*: Each IoTD sends a local connectivity matrix containing the location information of an IoTD detected during the identification period to the RRHE. The report frame is delivered via the underlying control channel (or it can be forwarded in a multi-hop manner via wireless backhaul routing). The RRHE identifies the optimal identification parameters based on the information and delivers this information to each IoTD in the next report phase.
- *Transmission phase*: During this phase, an IoTD can communicate by utilizing unused spatial-frequency resources.

Fig. 3 presents the timeline for the proposed scheme. In the proposed scheme, each IoTD identifies an unused spectrum resource with its identification parameters in each identification phase. After the identification phase, each IoTD transmits a report frame containing the identification results to the RRHE and receives a command frame from the RRHE.

[1] Spatial-resource identification/harvesting is different from spectrum sensing in that identification/harvesting is performed for the same communication resource pool, while whereas sensing ([14]) is performed for the resources of the primary users.

Table I: Key notation descriptions.

Notations	Description
N	The number of IoTDs
M	The number of antennas
B_i, $\|B_i\|$	The set of identification beams for ith IoTD and the number of selected beams for identification
$\mathbf{P}$	The IoTD detection probability
$\mathbf{B}$	The identification overhead
t_s	The identification time
L_r	The length of the report phase
ρ	the energy consumption per unit time
s, a, r	State, action, and reward

After receiving the command frame, each IoTD blocks the beams mentioned in the received command frame. In the transmission phase, the IoTDs transmit/receive data frames to/from other IoTDs using the unblocked beams. In this phase, the IoTDs operate in a contention-based manner. The associated communication techniques are discussed in Section IV. Major notations used in this paper are listed in Table I.

B. Antenna Model

IoTDs performing D2D communication are equipped with directional antennas, and they communicate directionally. Directional antennas in the mmWave band are classified as 1) switched-beam antennas and 2) beam-steering antennas. Switched-beam antennas are designed to cover a certain area per fixed beam, and one beam is activated to perform the communication. In beam-steering antennas, the main beam is controlled by the phase shifters in the desired direction to transmit and receive information.

The former has the advantage of convenient implementation and low cost, but has the disadvantage of attenuating the signal strength during switching between beams. The latter has the advantage of high signal quality realized through sophisticated control, but its implementation is expensive and complex. We assume that switched-beam antennas are used in IoTD systems operating under limited available energy and computing power. We also assume that the IoTDs including RRHEs are equipped with a switched-beam array antenna with M beam patterns, and each beam pattern is ideally non-overlapped. During transmission, only one direction of each sector is activated to transmit the signals, and the other sectors are blocked. During reception, multiple sectors can be activated simultaneously, or only a specific direction can be activated. An antenna controller is assumed to be used to keep track of the direction from which the maximum signal power is received.

C. Radio Resource Harvesting Edge

We assume that the RRHE collects the identification results from the IoTDs. Based on the identification results, the RRHE determines the optimal identification parameters of each IoTD. Furthermore, the RRHE sends the identified spatial frequency information to the IoTDs for D2D communication. The RRHE can comprise any form or combination of an MBS, FBS, WiFi AP, or dedicated RRHE. It should be noted that even if the FBS/MBS is aware of the frequency information in its cell area, it cannot know the local information of the IoTDs. Therefore, the RRHE can maximize the frequency resource efficiency by allocating spatial-frequency resources to the IoTDs based on their local identification information.

D. Channel Model

We assume that the IoTDs use C data channels and NOMA [22]-based wireless networks. As in [22], the NOMA scheme is implemented by combining orthogonal-frequency division multiplexing access and multi-carrier code division multiple access. As in their scheme, we assume a single physical data channel with S subcarriers. The data channel is divided into two subcarrier groups (SCGs) as follows:

- *SCG 1* ($\mathbb{S}_1$): the minimum number of subcarriers for data transmission. These subcarriers occupy only a small portion of the bandwidth of the data channel.
- *SCG 2* ($\mathbb{S}_2$): The set of remaining subcarriers excluding $\mathbb{S}_1$ in the data channel.

We assume that a full subcarrier is allocated for each beam of the IoTD i. Let $\mathbb{C}_j^i$ denote the data channel of the jth beam direction of the IoTD i. $\mathbb{C}_j^i$ can be interpreted as the geographical transmission and reception coverage area of the IoTD i when it exploits the jth beam. Furthermore, $\mathbb{S}_{1,j}^i$ and $\mathbb{S}_{2,j}^i$ denote SCGs 1 and 2 for the jth beam of the IoTD i, respectively. We designed the system to have a wider spectrum for $\mathbb{S}_2$ than for $\mathbb{S}_1$. We assume that $\mathbb{S}_1$ and $\mathbb{S}_2$ are sufficiently separated to negate any interference between them[2].

In the proposed scheme, we also assume an underlying control channel that utilizes all $\mathbb{S}_1$ of the C channels. The implementation of an underlying control channel for CRNs was validated by [24]. Each IoTD reports its identification results (presence/absence and position of the IoTD) to the RRHE via the control channel. Similarly, the RRHE calculates the optimal identification parameters for each IoTD by utilizing the reported information and disseminates the identification parameters to all the IoTDs via the control channel.

E. IoTD Detection

Several classical spectrum identification techniques have been proposed, including matched filters, feature detection, and energy detection. In our study, we assume that the IoTDs utilize an energy-detection technique to determine the presence or absence of an IoTD based on the amount of energy received. The received signal is integrated over the observation interval. Finally, the output of the integrator divided by the noise power (i.e., the signal-to-noise ratio, SNR) is compared with a certain threshold (or IoTD detection sensitivity) to determine the presence of an IoTD. An IoTD is present in the identified channel if the SNR is greater than the IoTD-detection sensitivity.

III. Centralized Spectrum Identification/Harvesting

A. Directional Spectrum Identification/Harvesting

In the cooperative directional identification/harvesting scheme, the IoTDs share the identification area with each other, which prevents the detection of overlapping areas. Furthermore, the RRHE forms multiple clusters. Each IoTD identifies a specific assigned direction using its identification beams. These identification beams for the IoTDs are assigned

[2] Inter-carrier interference between subcarriers can be reduced to a level where data communication is possible through an interference cancellation technique. [23]

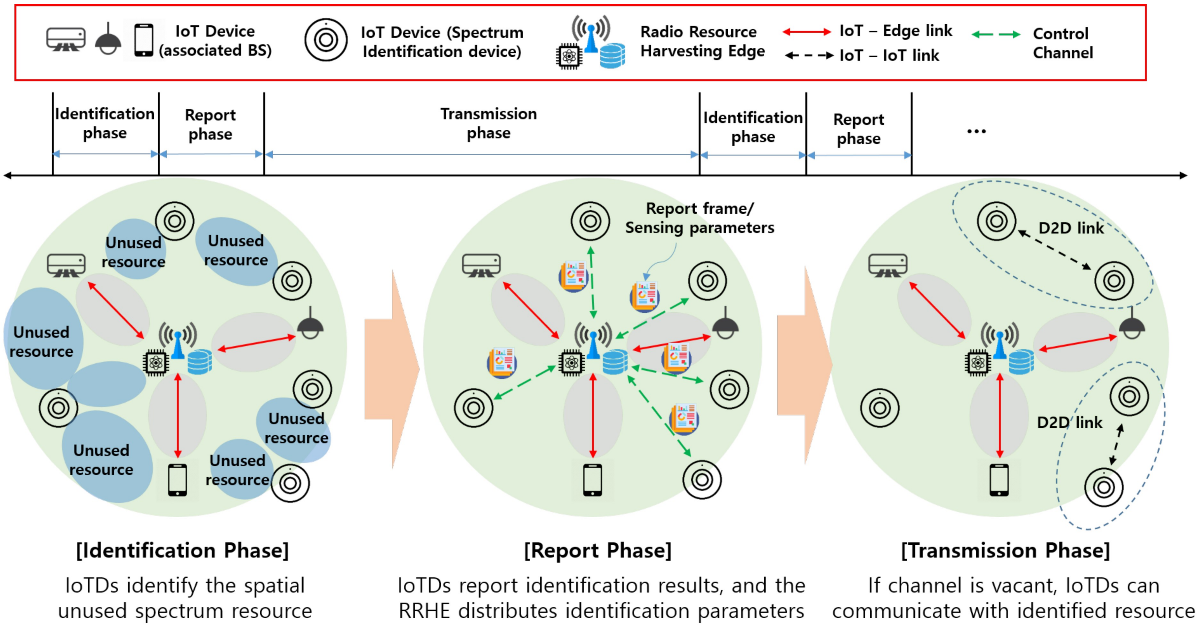

IoT Device (associated BS)
IoT Device (Spectrum Identification device)
Radio Resource Harvesting Edge
IoT – Edge link
IoT – IoT link
Control Channel
Identification phase
Report phase
Transmission phase
Identification phase
Report phase
...
Unused resource
Unused resource
Unused resource
Unused resource
Unused resource
Unused resource
Report frame/ Sensing parameters
D2D link
D2D link
[Identification Phase]
IoTDs identify the spatial unused spectrum resource
[Report Phase]
IoTDs report identification results, and the RRHE distributes identification parameters
[Transmission Phase]
If channel is vacant, IoTDs can communicate with identified resource

Fig. 3. Timeline for the proposed directional spectrum-sensing technique.

by the RRHE, which attempts to maximize the detection probability of the IoTD while reducing the overall identification overhead.

To improve the detection probability, the IoTDs may use as many beams as possible for identification. However, this approach results in various side effects such as significant energy consumption by the IoTDs in addition to time and network resources consumed for the identification. The following are the main overhead issues caused by the identification process:

1) Energy overhead caused by identification: An IoTD that is performing identification consumes additional energy.
2) Energy overhead caused by reporting: An IoTD that is transmitting an identification result to the RRHE consumes energy to transmit the report frames.

Therefore, the optimization technique should identify the identification parameters that can detect the IoTD and reduce the identification overhead simultaneously. The relevant optimization techniques are described in subsection III-B.

B. Identification Parameter Optimization

To maximize the efficiency of our spectrum identification process, we optimize the set of beams to identify (B_i) for the IoTDs while using an objective function similar to that used in [14]. B_i can be further described by B_i=[b_i^0, b_i^1, b_i^2, $\ldots$, and $b_i^{M-1}]^T$, where b_i^j is 1 if the jth beam of the IoTD $_i$ is used for the identification. Otherwise, b_i^j is zero.

Here, the objective function is written as follows:

$$\min_{\mathbf{B}} \alpha\left(1-\mathbf{P}(\mathbf{B})\right)+(1-\alpha)\mathbf{O}(\mathbf{B}), \tag{1}$$

where $\mathbf{P}(\mathbf{B})$ and $\mathbf{O}(\mathbf{B})$ denote the IoTD detection probability and identification overhead in terms of $\mathbf{B}$, respectively and α denotes the weight factor. Furthermore, $\mathbf{B}$ denotes the set of beams for all IoTDs i.

Here, the IoTD detection probability $\mathbf{P}$ is modeled in a manner similar to that in [14]. The IoTD detection probability $\mathbf{P}$ is given as

$$\begin{aligned} P(\mathbf{B}) &= P(\text{ At least one IoTD detect any other IoTD}) \\ &= 1-P(N \text{ nodes cannot find any other IoTD}) \\ &= 1-\prod_{i=0}^{N-1}\left\{1-\mathbf{P}_i(B_i)\right\}, \end{aligned} \tag{2}$$

where N denotes the number of IoTDs, and $\mathbf{P}_i(B_i)$ denotes the ith IoTD detection probability in terms of the identified beams. Then, $\mathbf{P}_i(B_i)$ is given as

$$\mathbf{P}_i(B_i)=\frac{\text{Number of beams to sense}}{\text{Number of entire beams}}=\frac{\sum_{j=0}^{M-1} b_i^j}{M}. \tag{3}$$

In the previous study, the identification overhead was modeled in terms of the time overhead for the identification. However, herein, the identification overhead ($\mathbf{O}$) has been modeled based on the energy consumed for the identification.

The identification overhead $\mathbf{O}$ is defined as

$$\mathbf{O}(\mathbf{B})=\sum_{z\in\{s,r\}}\mathbf{O}_z(\mathbf{B}), \tag{4}$$

where $\mathbf{O}_s(\mathbf{B})$ denotes the overhead due to spectrum identification. $\mathbf{O}_r(\mathbf{B})$ denotes the overhead due to the transmission of the identification results and reception of the identification parameters from the RRHE.

An IoTD consumes energy while identifying a channel. Let ρ denote the energy consumption per unit time. Then, $\mathbf{O}_s(\mathbf{B})$ is expressed as

$$\mathbf{O}_s(\mathbf{B})=\sum_{i=0}^{N-1}|B_i| * t_s * \rho \tag{5}$$

where t_s and $|B_i|$ denote the identification time and number of selected beams for the spectrum identification, respectively.

Similarly, IoTDs also consume energy while transmitting identification results and receiving the identification parameters. Hence, the report overhead $\mathbf{O}_r(\mathbf{B})$ is given as

$$\mathbf{O}_r(\mathbf{B}) = \sum_{i=0}^{N-1} |B_i| * L_r * \rho, \tag{6}$$

where L_r denotes the length of the report phase.

In the previous work, the identification beams with the highest SNR values were used for improving the identification accuracy and minimizing the overlapped identification area to reduce the overhead. However, in the previous algorithm, a beam with a good channel condition was always selected, resulting in an energy-consumption inequality between the IoTDs. Furthermore, because the identification overhead was modeled as a wasted time opportunity, the energy consumed for the identification was relatively high. To address this problem, we propose a beam selection algorithm based on a machine learning technique.

C. *Optimal Beam Selection based on Reinforcement Learning*

A standard RL model comprises a finite set of possible states of an environment $\mathbf{S} = \{s_1, s_2, \ldots, s_n\}$, a set of possible actions $\mathbf{A} = \{a_1, a_2, \ldots, a_m\}$ of a learning agent, a scalar reinforcement signal r, and an agent policy π. At each time step, the agent perceives the state $s \in \mathbf{S}$ of the environment and selects an action $a \in \mathbf{A}$ based on its current policy π. Time is represented by a sequence of time steps $t = 0, 1, \cdots$. At each time step, a controller observes the system's current state and selects an action. Correspondingly, the environment makes a transition to the new state $s' \in \mathbf{S}$ and generates a reinforcement signal c_t, which is called an immediate reward. This is given to the agent. The learning agent then updates its policy and the next round of iteration is begins [25].

The objective of the agent is to find an optimal policy $\pi^*(s)$ for each state that minimizes the total expected discounted reward over an infinite time horizon. This reward is defined as

$$V^*(s) = \min_{\pi} \mathbb{E}\left(\sum_{t=0}^{\infty} \gamma^t c_t\right), \tag{7}$$

where $\mathbb{E}$ represents the expectation of the operator, and $\gamma \in (0, 1)$ is a discount factor. Note that a reinforcement learning algorithm is considered to converge when the learning curve becomes flat and no longer increases. In theory, Q-Learning has been proven to converge towards the optimal solution [26]. Thus, the optimality condition can be defined by

$$V(s)^{(t+1)} - V(s)^{(t)} < e \approx 0, \tag{8}$$

where t denotes the iteration step and e denotes the small size threshold. As per Bellman's optimality criterion, the optimal policy π^* satisfies

$$V^*(s) = \min_{a}\left(C(s,a) + \gamma \sum_{s' \in \mathbf{S}} P_{s,s'}(a) V^*(s')\right) \tag{9}$$

where $C(s, a)$ denotes the expected cost $C(s, a) = \mathbb{E}\{c(s, a)\}$, and $P_{s,s'}$ denotes the transition probability for the change from s to s'.

Given the optimal value function, we can specify the optimal policy as

$$\pi^*(s) = \arg\min_{a}\left(C(s,a) + \gamma \sum_{s' \in \mathbf{S}} P_{s,s'}(a) V^*(s')\right). \tag{10}$$

For each agent i, we define an evaluation function, denoted by $Q(s, a)$, as the expected discounted reinforcement of taking action a at the state s and then counting by optimally selected action.

$$Q(s,a) = \mathbb{E}\left\{\sum_{t=0}^{\infty} \gamma^t c(s_t, \pi(s)) | s_0 = s\right\} \tag{11}$$

For each agent i, Q(s,a) can be rewritten as

$$Q(s,a) = C(s,a) + \gamma \sum_{s' \in \mathbf{S}} P_{s,s'}(a) Q(s', a'). \tag{12}$$

To apply Bellman's criterion, we must find an intermediate minimal value of $Q(s, a)$, denoted as $Q^*(s, a)$, where the intermediate evaluation function for every possible subsequent state–action pair is minimized, and the optimal action is performed with respect to each subsequent state. $Q^*(s, a)$ is given as

$$Q^*(s,a) = C(s,a) + \gamma \sum_{s' \in \mathbf{S}} P_{s,s'}(a) \min_{a' \in \mathbf{A}} Q^*(s', a'). \tag{13}$$

We can then determine the action a^* with respect to the current state s. In other words, we can determine π^*. Therefore, $Q^*(s, a^*)$ is minimal and can be expressed as

$$Q^*(s, a^*) = \min_{a \in \mathbf{A}} Q^*(s, a). \tag{14}$$

In the Q-learning process, attempts are made to find $Q^*(s, a)$ in a recursive manner by utilizing the available information (s,a,s',a'), where s and s' are the states at times t and $t+1$, respectively, and a and a' are the actions taken at time t and $t+1$, respectively. The Q-learning rule for updating the Q values relative to agent i is given as

$$Q(s,a) = Q(s,a) + \alpha\left[c + \gamma \min_{a} Q(s', a') - Q(s,a)\right], \tag{15}$$

where α denotes the learning rate.

We attempt to find an optimal beam set for all the IoTDs in different environment states such that the objective function is minimized. Our reasons for applying RL to find the optimal beam selection are as follows:

- The advantage of RL is that, over time, reward-based learning results in increasing number of optimal results. In the network environment we assume the IoTDs are distributed, and they opportunistically detect unused spectrum resources for D2D communication. Therefore, fixed IoTDs can detect a great number of unused spectrum at a lower cost over time.
- There are frequent network topology changes. The status of the IoTDs changes from time to time. In some cases, the battery wears out, thus causing the IoTD to power down or a new IoTD to participate. Therefore, in a topology where the network conditions change frequently, RL techniques that yield the optimum result with relatively few computations may be suitable.

We consider the RRHE as the learning agent, and the IoTDs being served and their identification beam set as the agent's environment. Correspondingly, we define the basic RL elements as follows:

- **State**: The selection of a state space is a basic step of Q-learning. The selected state variables should comprise the features that are knowable and have no aftereffects. In this scheme, we define the state as the set of beams that each IoTD utilizes for identification. As there are

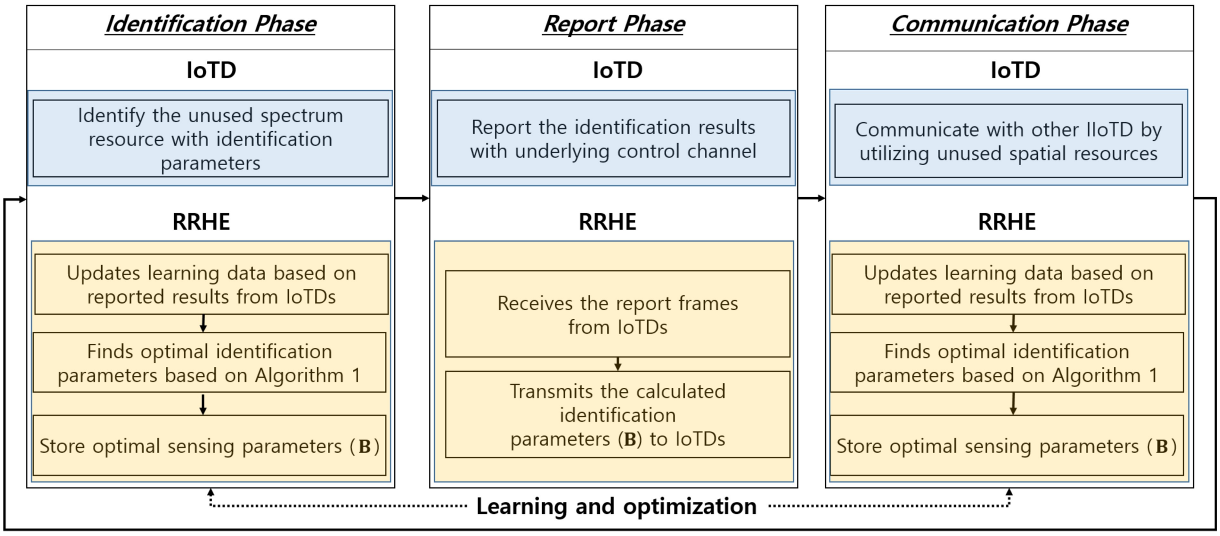

Identification Phase
IoTD
Identify the unused spectrum resource with identification parameters
RRHE
Updates learning data based on reported results from IoTDs
Finds optimal identification parameters based on Algorithm 1
Store optimal sensing parameters (B)
Report Phase
IoTD
Report the identification results with underlying control channel
RRHE
Receives the report frames from IoTDs
Transmits the calculated identification parameters (B) to IoTDs
Communication Phase
IoTD
Communicate with other IIoTD by utilizing unused spatial resources
RRHE
Updates learning data based on reported results from IoTDs
Finds optimal identification parameters based on Algorithm 1
Store optimal sensing parameters (B)
Learning and optimization

Fig. 4. Flow chart of the proposed RL-based identification scheme for D2D communications.

2^M total combinations of beams utilized, the number of states for each IoTD is 2^M, and the state of IoTD i (s_i) is defined as follows:

$$s_i = \left\{b_i^0, b_i^1, b_i^2, \dots, b_i^{M-1}\right\}, \forall i \tag{16}$$

where b_i^j is 1 if the jth beam of the IoTD i is utilized for identification. Otherwise, b_i^j is zero.

- **Action:** The set of possible actions is determined based on the selected beams for each IoTD. The action of the IoTD i (a_i) is defined as

$$a_i = \left\{\bar{b}_i^0, \bar{b}_i^1, \bar{b}_i^2, \dots, \bar{b}_i^{S-1}\right\} \subset s_i, \forall i \tag{17}$$

where $\bar{b}_i^j$ denotes the member of the selected beam set and S denotes the number of selected beams.

- **Reward:** In our RL-based Q-learning, each state is defined as the set of beams that each IoTD utilizes for identification, and the action is defined as the beam set to be activated for the identification of all the nodes. When determining new beam sets for all the IoTDs, the overall network reward is defined as the probability of finding spectrum resources against the energy consumed for the identification. In other words, the reward function $C(s,a)$ can be designed with the objective function (1). Therefore, $C(s,a)$ is defined as

$$C(s,a) = \alpha\left(1 - \mathbf{P}(\mathbf{B})\right) + (1-\alpha)\mathbf{O}(\mathbf{B}). \tag{18}$$

In the proposed identification algorithm, the RRHE determines the identification parameters of all the IoTDs. As each IoTD has a different channel environment, location, etc., the RRHE should learn about all of the IoTDs individually. Algorithm 1 outlines the Q-learning-based B_i selection algorithm for the spectrum identification. For the beam selection (action) by each IoTD, the RRHE utilizes an ϵ-greedy algorithm, where ϵ is a random factor that is used to find the optimal values to avoid the local minima.

RRHE works as follows with our proposed RL technique.

- The RRHE collects the IoTD's spectrum identification information in the report phase.
- The RRHE performs the RL algorithm until the next report phase and calculates the reward for determining the optimal beam set.

Algorithm 1 Q-learning algorithm for B_i selection

```
Initialize:
Initialize $Q(s,a)$ with random weights;
Evaluate the starting state $s_t^i$;

Learning:
loop
    Generate a random number $r$, $0 \le r \le 1$;
    if $r < \epsilon$ then
        Select an action randomly;
    else
        Select the action $a$ characterized by the minimum Q value;
    end if
    Evaluate an immediate cost $c_t^i$ based on (18);
    Observe the next state $s_{t+1}^i$;
    $Q(s,a) = Q(s,a) + \alpha\left[c + \gamma \min_a Q(s',a') - Q(s,a)\right]$;
    $s_t^i \leftarrow s_{t+1}^i$;
end loop
```

- The RRHE shares the identification beam parameters with the IoTDs based on previously obtained values in the next report phase and retrains and finds the optimal values based on the reported results.

It should be noted that the IoTDs can report manipulated malicious information for reducing energy consumption. That is, the entire system may be vulnerable owing to selfish IoTD behaviors. However, some research contributions present a view of trust computing by demonstrating why devices inside a networks system should act honestly [20]. In [20], the total network utilities of each participant are maximized when the devices are truthful. Therefore, we also assumed that none of the IoTDs exhibited dishonest behavior.

D. Complexity Analysis

In our algorithm, the RRHE is a learning agent that should manage the identification parameters of all the IoTDs. Therefore, the algorithm complexity depends on the number of IoTDs and number of antennas. In this section, we analyze the complexity of our algorithm to find the optimal identification beam set based on our RL.

In our model, each IoT has a total of 2^M states, and there are 2^M candidate actions. Therefore, the algorithm for finding the optimal beam set based on the proposed RL has a worst-case time complexity of $O(N \cdot 2^M)$, assuming a total of N IoTDs. In the algorithm for finding the optimal beam set, the RRHE finds the optimal beam sets for all the IoTDs based on RL and informs each IoTD of the optimal beam set in the next report phase. Therefore, the calculation should be completed within the interval of the report phase.

Assuming an RRHE implementation with the Raspberry Pi 4B as a representative IoTD, the CPU generates 1.5 million clocks per second (1.5 GHz). If we consider approximately 100 clocks for calculating the reward, approximately 1.5×10^7 operations are performed [14]. When $M = 16$ (the number of antennas of the IoTD = 16) and the identification cycle is 1 s, we can find the optimal identification beams for 228 IoTDs ($228 \cdot 2^{16} \approx 1.5 \times 10^7$). Thus, by assuming that the identification period is T_s, we can find the identification beams for $228 \cdot T_s$ IoTDs.

Although we have performed a preliminary analysis on the time consumption of the Q-learning algorithm in this case based on an IoT system, the applicability of the proposed scheme to a real system is unclear. To evaluate the applicability of the proposed RL-based harvesting method, we built a simple IoT system and measured its computational time. In this system, the Raspberry Pi 3 model plays the role of the RRHE, and RL is used to find the optimal beam of the surrounding 200 IoTDs. In our experimental result, the reward is converged between approximately 150 and 200 epochs, as shown in Fig. 5. As 20 epochs consumed approximately 1 s, it was analyzed that approximately 7.5 s would be required in the actual experimental environment. Therefore, the optimization should be completed within 7.5 s to ensure accurate sensing results on this system, i.e., the sum of the identification period (T_s) and communication period (T_c) should be longer than 7.5 s. Note that an RRHE with high computing power may have shorter delay requirements in a real system.

E. Workflow of the Proposed RL-based Identification Scheme

Fig. 4 presents the entire workflow of the proposed RL-based identification scheme. As mentioned in the previous section, the proposed technique comprises three phases. In the identification phase, each IoTD finds unused spatial resources using the identification parameters selected from the RRHE. Simultaneously, the RRHE finds the optimal identification parameters through RL. In the report phase, each IoTD sends the identification results to the RRHE through the control channel. The RRHE receives the identification results and sends the identification parameters found in the RL in the previous phase to the IoTD. In the communication phase, each IoTD performs D2D communication through the surrounding unused spectrum resources, and the RRHE updates the new data (identification results) and starts identifying the identification parameters using RL. It should be noted that the RRHE performs the same step as in the identification phase, which can be combined into the learning and optimization phase.

IV. Contention-Based Communication Scheme

A. Deafness-Free D2D communication

In the transmission phase, each IoTD can communicate opportunistically through an empty channel/beam. Each IoTD can perform ad hoc communication with other IoTDs (D2D link). As this D2D link should not interfere with the communication within the femto-cell, it can perform D2D transmission by performing carrier sensing before the transmission and sending the communication request frame.

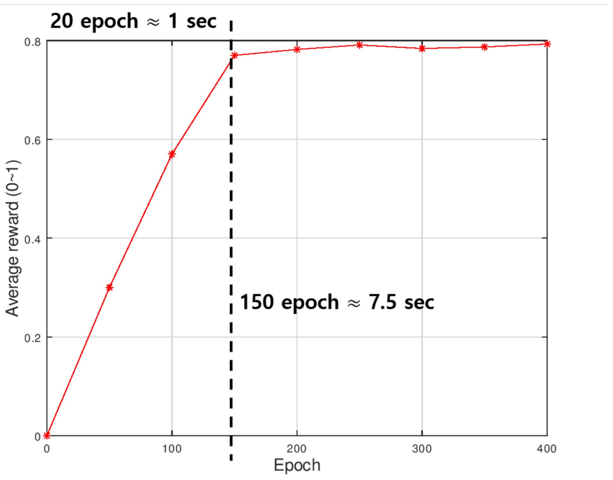


Fig. 5. Variation of reward according to the learning epoch in the experiment (Number of IoTDs: 200, RRHE: Raspberry PI 3).

However, a deafness problem occurs when the transmitting IoTD transmits a communication request frame at the same time when the receiving IoTD is communicating, wherein the frame cannot be heard. To overcome this problem, there have been research about contention-based communication schemes such as circular-request/response-based MAC [27], [28], advanced-notice-based MAC [29], [30], and tone-based MAC [31], [32]. The characteristics of each contention-based MAC scheme can be summarized as follows:

- Circular-request/response-based MAC [27], [28]: A technique to prevent deafness by notifying that the node is communicating by transmitting the control frame in a circular manner in all directions.
- Advanced-notice-based MAC [29], [30]: A technique to prevent deafness by checking its transmission waiting queue and notifying the next communication target node in advance.
- Tone-based MAC [31], [32]: A technique to prevent deafness by notifying that a node is communicating by sending a tone around during communication.

The above studies are not a fundamental solution because there are scenarios in which deafness can occur, but the proposed technique can identify deafness and avoid it.

In the proposed scheme, the transmitting IoTD can transmit a communication request frame with both SCGs to the receiving IoTD. In [21], we studied the problem of deafness in an ad hoc network environment. We dealt with the deafness problem using two channels. In this work, we modified the previous scheme using the NOMA technique with one data channel. For our scheme, we can consider the following scenarios.

- **1) Normal case**: If the receiving IoTD is idle, it receives two communication request frames from $\mathbb{S}_1$ and $\mathbb{S}_2$. It then sends a reply frame with both SCGs. After some time, it starts to transmit the data frame only on $\mathbb{S}_2$.
- **2) Deafness case**: If the receiving IoTD is communicating with another IoTD, it receives only one content request frame from $\mathbb{S}_1$. The transmitting IoTD then receives only one reply frame from $\mathbb{S}_1$ and is aware that the receiving

IoTD is a deaf node. Thus, it attempts to connect with another IoTD.

- **3) Collision case**: If the receiving IoTD does not receive any frame on $\mathbb{S}_1$ and $\mathbb{S}_2$ because two communication request frames collide in both SCG, then the sender's (more than two IoTDs) retransmission timer expires, and the sender retransmits the content request frames on both SCGs after a back-off period.

The transmitting IoTD can detect an actual network failure by distinguishing deafness from a collision using the above technique. If the sender is aware that the receiver is in a deaf state, the IoTD can start communicating with another IoTD, thereby improving the aggregate throughput.

B. Analysis of Deafness-Free D2D Communication

In this section, we analytically demonstrate how the deafness-free Medium Access Control (MAC) overcomes the deafness problem. First, we introduce a few notations used in the analysis.

Let C_i denote the set of all SCGs of node i. C_i is defined as

$$C_i = \left\{C^i_{k,j} \mid k \in \{1,2\} \text{ and } 0 \leq j < M\right\}. \tag{19}$$

We denote T_i^j as the set of SCGs utilized for transmission from node i to node j, and R_j^i as the set of SCGs used for the reception from node j to node i. For instance, if node i transmits a request to send an RTS frame to j using i's third beam, and node j receives the RTS frame with j's second beam, then $T_i^j = \left\{C^i_{c,3}, C^i_{d,3}\right\}, R_j^i = \left\{C^j_{c,2}, \text{and} C^j_{d,2}\right\}$. Because we consider that the SCGs are geographical transmission and reception coverage areas shared by a transmitter and receiver, respectively, (i.e., $C^i_{k,3} = C^j_{k,2}, \forall k$), it is sufficient to state that $T_i^j = R_j^i$ in the above example.

If we denote A_i as the set of available (or idle) SCGs and U_i as the set of unavailable SCGs for node i, then $A_i \cap U_i = \phi$. Here, unavailability results from two scenarios: (1) the SCGs are busy communicating, or (2) the SCGs are blocked because they are overhearing nearby ongoing communications. Furthermore, because the data and control channels are sufficiently separated, we obtain $C^i_{d,j} \cap C^i_{c,j} = \phi$.

Definition 1. If $\exists i$ such that node i has data for node j and $A_j \cap R_j^i = \phi$, then node j is a deaf node.

Here, $A_j \cap R_j^i = \phi$ reflects the situation wherein node j cannot respond to i because j is either busy with communicating with another node or because the beams from i are blocked. Moreover, we exclude the case wherein the SCGs from i to j are unavailable owing to deafness (i.e., the case of $A_i \cap T_i^j = \phi$ (or $A_i \cap R_j^i = \phi$)).

Definition 2. A collision occurs if a random combination of nodes i_k's ($0 \leq k < N, N > 1$) transmits frames at the same time and $\left\{\bigcap_{k=0}^{N-1} R_j^{i,k}\right\} \neq \phi$ for all nodes k.

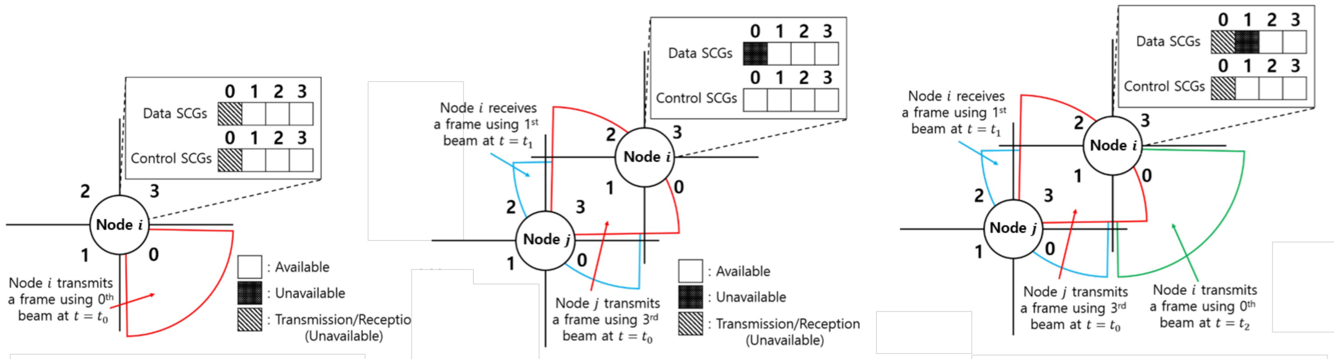
Let us consider a general node (e.g., source node i) and destination node j. The SCGs between nodes i and j may be in one of the following possible states:

- *Case 1) All the SCGs are idle:* In this case, there is no communication nearby. Therefore, node i is ready for transmission or reception.
- *Case 2) All the SCGs are blocked:* The DF-DMAC protocol is designed to utilize the control SCGs to transmit request frames and response frames only, which means that at most one control SCG can be blocked while i transmits or receives a frame. Therefore, the other $M-1$ control SCGs are always idle. Because our DF-DMAC protocol operates with $M > 1$, this case is automatically excluded from the proof.
- *Case 3) Multiple SCGs are unavailable:* In this case, multiple SCGs are blocked owing to their overhearing of other ongoing communication or they are busy with transmission or reception. Therefore, we further divide this case as follows.
  - *Case 3.1) Only a pair of data and control SCGs is utilized for communication:* Node i is engaged in communication. Node i does not listen to any other communication. Therefore, only one data and one control SCG are unavailable, and all the other SCGs are available. For example, on the red line in Fig. 6 (a), node i is engaged in communication and is utilizing beam zero. Therefore, the data and control SCGs that are affected by beam zero are unavailable, whereas the other SCGs are available.
  - *Case 3.2) Multiple data SCGs are blocked by ongoing overheard communication:* Although node i is not engaged in communication, it can overhear other nearby communications and set its directional network allocation vector (DNAV) accordingly[3]. It means that multiple data SCGs may be unavailable. All the control SCGs are available because they are blocked only when the nodes are engaged in communication. For example, on the red line in Fig. 6 (b), node i listens to the communication while using beam 1. Therefore, the data SCG that is affected by beam 1 is unavailable.
  - *Case 3.3) Combination of Cases 3.1 and 3.2:* In this case, node i is in communication and overhears other communications, which means that it updates its DNAV. Therefore, some data SCGs and one control SCG are unavailable. For example, on the red line in Figure 6 (c), node i is engaged in communication and is using beam zero. Therefore, the data and control SCGs that are affected by beam zero are unavailable. Furthermore, node i listens to other communications while using beam 1. Therefore, the data SCG that is affected by beam 1 is unavailable.
- *Case 4) Others:* Other cases besides the aforementioned cases cannot occur, for example, a case wherein multiple control SCGs are blocked. However, more than two control SCGs of node i cannot be blocked for the reason discussed in (*Case* 2). Therefore, we do not need to consider this case. In DF-DMAC, a management entity is exploited to report errors in the system.

Lemma 1. In deafness-free D2D communication, if a source node transmits an RTS to a destination that is not engaged in communication, then destination can reply to the source node (no deafness).

Proof. For the proof of this lemma, we consider a destination node that is not in communication (the opposite case is proved in Lemma 2). Therefore, we do not need to consider *Cases 3.1* and *3.3*. If the source node i wishes to communicate with the destination node j, then node i transmits an RTS over $T_i^j = \left\{C^i_{d,m'}, C^i_{c,m}\right\}$, and node j replies with a clear-to-send

[3] A node that is listening to any other communication updates its corresponding DNAV. Therefore, the corresponding data SCG is unavailable.


Data SCGs
0 1 2 3
Control SCGs
0 1 2 3
2
3
Node i
1
0
Node i transmits a frame using 0th beam at $t = t_0$
: Available
: Unavailable
: Transmission/Receptio (Unavailable)
Data SCGs
0 1 2 3
Control SCGs
0 1 2 3
Node i receives a frame using 1st beam at $t = t_1$
2
3
Node i
1
0
2
3
Node j
1
0
Node j transmits a frame using 3rd beam at $t = t_0$
: Available
: Unavailable
: Transmission/Reception (Unavailable)
Data SCGs
0 1 2 3
Control SCGs
0 1 2 3
Node i receives a frame using 1st beam at $t = t_1$
2
3
Node i
1
0
2
3
Node j
1
0
Node j transmits a frame using 3rd beam at $t = t_0$
Node i transmits a frame using 0th beam at $t = t_2$

(a) Example of case 3.1

(b) Example of case 3.2

(c) Example of case 3.3

Fig. 6. Example of possible SCG states ($M = 4$).

message over $R_j^i = \left\{C_{d,N'}^j, C_{c,n}^j\right\}$ (m is beam i's index for transmission from i to j; and n is beam j's index for reception from i to j, where $0 \le m < M$)). As mentioned previously, $C_{k,m}^i = C_{k,n'}^j$, $\forall k$.

In *Case 1)*,

$$U_j = \phi, A_j = \left\{C_{k,n}^j | k \in c, d, 0 \le n < M\right\}. \tag{20}$$

Therefore, we obtain the following result:

$$\begin{aligned} A_j \cap R_j^i &= \left\{C_{d,n}^j, C_{c,n}^j\right\} \\ &= \left\{C_{d,m}^i, C_{c,m}^i\right\} \neq \phi. \end{aligned} \tag{21}$$

In this case, node j can reply to i because j is not a deaf node according to Definition 1.

In *Case 3.3)*,

$$\begin{aligned} U_j &= \left\{C_{d,n}^j | n \subset \Omega_i, n \neq \phi\right\} \\ A_j &= C_j - U_i. \end{aligned} \tag{22}$$

Therefore, we obtain the following result:

$$A_j \cap R_j^i = \begin{cases} \left\{C_{c,n}^j\right\} & \text{if } C_{d,n}^j \notin A_j \\ \left\{C_{d,n}^j, C_{c,n}^j\right\} & \text{otherwise} \end{cases}. \tag{23}$$

Although the data SCG of node j is unavailable in the former case, node j can reply to i by using a control SCG. In the latter case, node j can reply to i with both the SCGs that are pointed toward i. □

Lemma 2. In the DF-DMAC protocol, a source node that transmits an RTS frame to a destination node that is engaged in communication can distinguish between collision and deafness.

Proof. Lemma 2 is based on the same assumptions as Lemma 1, and we do not need to consider *case 1* because in Lemma 2, it is assumed that the destination node is engaged in communication. For one source node, if node i wishes to communicate with node j, then node i transmits an RTS frame with $T_i^j = \left\{C_{d,m'}^i, C_{c,m}^i\right\}$ (m is a used beam index when node i transmits to node j $(0 \le m < M)$).

In *Case 3.1)*,

$$\begin{aligned} U_j &= \left\{C_{c,n}^j, C_{d,n}^j\right\}, (0 \le n < M) \\ A_j &= \left\{C_{k,a}^j | k \in \{c, d\}, (0 \le a < n, n < a < M\right\}. \end{aligned} \tag{24}$$

Therefore, we obtain the following result:

$$A_j \cap R_j^i = \begin{cases} \phi & \text{if } R_j^i = U_j \\ \left\{C_{d,n}^j, C_{c,n}^j\right\} & \text{if } R_j^i \neq U_j \end{cases}. \tag{25}$$

In the former case, node j cannot reply to i. However, node j is transmitting a frame in this case.

From Definition 2, we can obtain

$$\begin{aligned} R_j^i \cap R_j^k &= U_j \cap R_j^k \\ &= \left\{C_{c,n}^j, C_{d,n}^j\right\} \neq \phi. \end{aligned} \tag{26}$$

Therefore, this is a collision case, and node j does not reply to node i. In the latter case, node j can reply to i because both SCGs that are pointed toward i are available.

In *Case 3.2)*,

$$\begin{aligned} U_i &= \left\{C_{c,n}^j, C_{d,n}^j\right\} \cup \left\{C_{d,a}^j | 0 \le a < n, n < a < M\right\} \\ A_j &= C_j - U_i. \end{aligned} \tag{27}$$

Therefore, we obtain the following result:

$$A_j \cap R_j^i = \begin{cases} \phi & \text{if } R_j^i \in U_j \\ \left\{C_{d,n}^j\right\} & \text{if else } C_{d,n}^j \in U_j \\ \left\{C_{d,n}^j, C_{c,n}^j\right\} & \text{otherwise} \end{cases}. \tag{28}$$

In the first case, node j cannot reply to i. However, node j is transmitting a frame in this case. From Definition 2, we can obtain the following:

$$R_j^i \cap R_j^k = \left\{C_{c,n}^j, C_{d,n}^j\right\} \neq \phi. \tag{29}$$

Therefore, this is a collision case, and node j does not reply to node i. In the second and third cases, node j can reply to i because more than one SCG is available in these two cases. Therefore, if the source node does not receive a reply, the transmitted frame has encountered a collision. Otherwise, the source node can react appropriately to the destination reply. □

Table II: Simulation parameters.

Parameter	Value
Topology Size	1 km x 1 km
Data Frame Size	1024 Bytes
Inter-Arrival Time of IoTDs	0.1 s
Energy Consumption	0.4 mJ
Number of Beams M	2∼16
Operating Frequency	28 GHz
Number of IoTDs	24
Bandwidth W	1 kHz
The Length of Report Phase L_r	0.04 sec
The Length of Sensing Phase t_s	0.01 sec
Discount Factor γ	0.9

Theorem 1. There is no deafness problem in the DF-DMAC protocol.

Proof. Owing to Lemma 1 and 2, there is no deafness problem in the DF-DMAC protocol. □

V. Performance Evaluation

For the performance evaluation, we developed a simulator based on OPNET modeler 14.5. The evaluation performance was measured using the following metrics:

- *Energy Consumption for Spectrum Identification (mJ)*: The total energy consumption from all the IoTDs for the spectrum identification.
- *Channel Utilization (%)*: The percentage of time for which a channel is occupied by the IoTDs over the entire time period.
- *Aggregate Throughput of IoTDs (Mbps)*: The total amount of traffic successfully transmitted and received through the unused spatial resources among the IoTDs.
- *Deafness Duration (ms)*: The time between the first transmission of the request frame and its corresponding response frame.

To evaluate the effectiveness of our proposed scheme, we compared the proposed scheme, omni-directional identification scheme, and directional identification scheme [14]. In addition, we analyzed the deafness-free D2D communication scheme with other contention-based communication schemes such as circular-request/response-based MAC [27], [28], advanced-notice-based MAC [29], [30], and tone-based MAC [31], [32] in terms of the medium access efficiency from the identification/harvesting perspective.

The details of the simulation parameters are listed in Table II. The simulation system consists of 24 IoTDs and one RRHE. The IoTDs are uniformly distributed in the topology, where the RRHE is located at the center. Let us assume that each IoTD consumes 0.4 mJ of energy per spectrum identification. In this experiment, we assume that each IoTD enters the idle mode to save energy after the operating time is complete. Then, each IoTD can wake up from the idle mode for communication or identification. We assume that the awake time interval of the IoTDs follows an exponential distribution with a mean of 5.0 s. In addition, we adopt the inter-arrival-time of the data frame from an upper layer as 0.1 s. The data frame size is set as 1024 bytes. To measure the value of each graph, we generated a total of 100 samples for each point and calculated the average value. In addition, each sample is generated through a random seed value and the total simulation time of each sample is 1 h. The location of the IoTDs for each iteration is random within the topology and the on-off duration follows a normal Distribution with mean 10 and standard deviation 2.

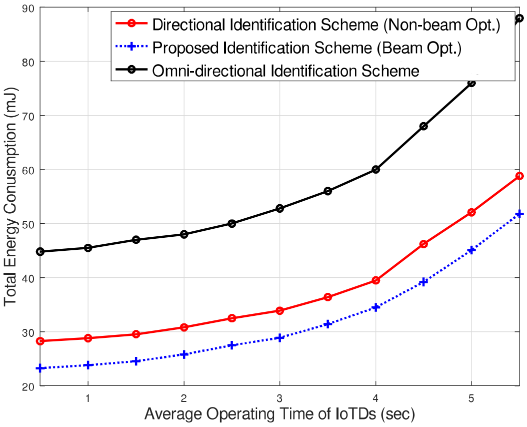


Fig. 7. Average operating time of IoTDs versus total energy consumption for identification.

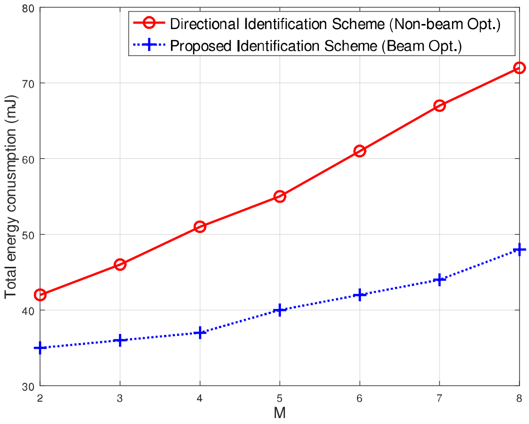


Fig. 8. Number of antennas versus total energy consumption of IoTDs for identification.

Fig. 7 presents the simulation results of the energy con-

Table III: Awake time interval distribution of IoTDs versus their total energy consumption (mJ)

	Omni-directional Identification Scheme	Directional Identification Scheme (Non-beam opt.)	Proposed Identification Scheme (Beam opt.)
Exponential Distribution ($\lambda = 0.2$)	76	52.08	45.08
Uniform Distribution ($u(1, 5)$)	81.4	57.32	42.16
Normal Distribution ($N(5, 1)$)	77.68	53.11	44.81

Table IV: Awake time interval distribution of IoTDs versus their channel utilization (%)

	Omni-directional Identification Scheme	Directional Identification Scheme (Non-beam opt.)	Proposed Identification Scheme (Beam opt.)
Exponential Distribution ($\lambda = 0.2$)	29	53	62
Uniform Distribution ($u(1, 5)$)	32	57	66
Normal Distribution ($N(5, 1)$)	31	55	65

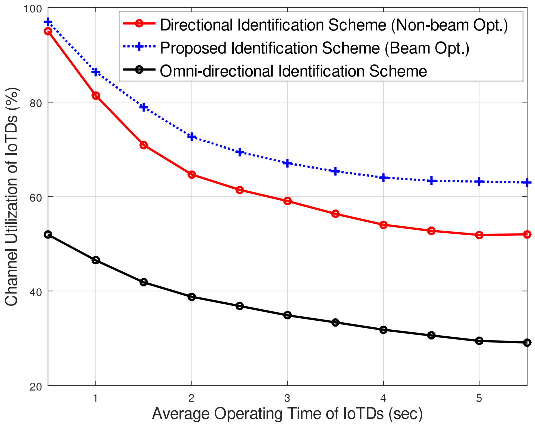


Fig. 9. Average operating time of IoTDs versus their channel utilization.

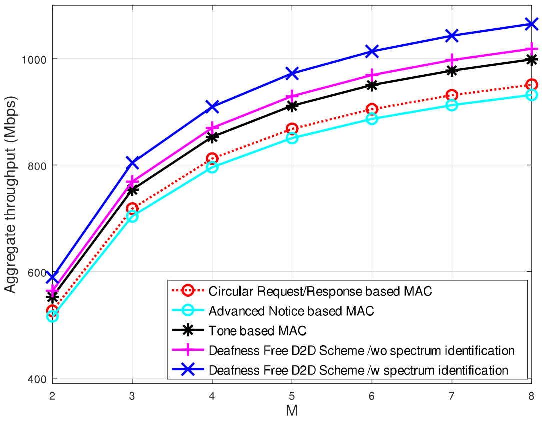


Fig. 10. Number of antennas versus aggregate throughput of IoTDs.

sumption of the IoTDs for identification versus the average operating time of the IoTDs[4]. It is observed that the energy consumption increases as the operating time of the IoTDs increases because the IoTDs perform spectrum identification more frequently. Moreover, we observed that the omni-directional identification technique consumes more energy for the identification than the directional identification technique. This is because the identification range in the omni-directional identification technique is smaller than that in the directional identification scheme. In addition, our proposed RL-based spectrum identification technique shows an approximately 18% performance improvement over that of the directional identification scheme. This is because identification in the directional identification scheme is performed by specific IoTDs, whereas the identification work is uniformly distributed in our proposed scheme. Table III shows the awake time interval distribution of IoTDs versus their total energy consumption, where the average operating time of the IoTDs is 5 s. As shown in the table, the proposed scheme demonstrates the best performance regardless of the distribution. This is because it adapts well to changes in the environment topology, which is one of the advantages of our proposed RL-based identification scheme. Fig. 8 presents the simulation results of the energy consumption of the IoTDs for identification versus their number of antennas. As shown in the figure, as the number of antennas increases, the energy consumed for sensing increases, and thus, the total energy consumption increases. However, the slope of the proposed technique is less than that of the non-optimized technique. Similar to Fig. 7, the proposed technique with beam optimization results in approximately 50% greater energy savings when compared with the non-optimized technique. Note that the proposed method consumes more energy than the non-sensing technique owing to transmission of sensing data and network signaling. However, energy consumption in the identification and report phases is less than 5% of the total energy consumption. On the other hand, the energy consumption in transmission phase is the highest. Therefore, our technique can realize performance improvement in terms of throughput by utilizing network resources while generating less energy overhead.

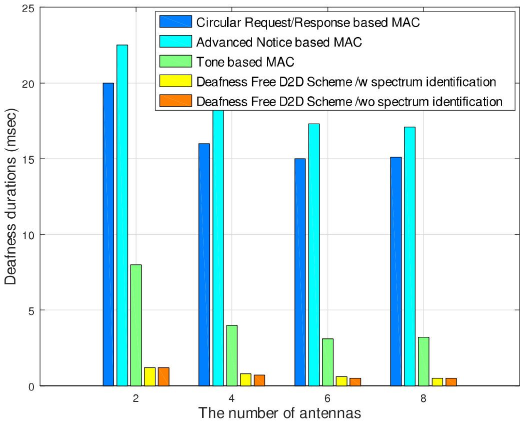


Fig. 11. Average deafness duration of IoTDs.

Fig. 9 presents the channel utilization versus the average operating time of the IoTDs. It is observed that the channel utilization decreases as the average operating time of the IoTDs increases, and the time used by the IoTDs is greater because they cannot transmit data frames, as at least one IoTD occupies the channel. As in the aforementioned results, directional identification efficiently finds spatial frequency resources, which means that the channel efficiency is better than that in omni-directional identification. In addition, the proposed RL-based identification technique demonstrates better efficiency than the directional identification scheme. Table IV shows the channel utilization of the IoTDs in the form of the awake time interval distribution. As shown in the table, the proposed scheme demonstrates the best performance regardless of distribution. From the results of Tables III and IV, we can observe that the proposed technique consumes less energy and activates more hidden space resources regardless of the network environment.

Fig. 10 presents the aggregate throughput versus the number of antennas. As shown in the figure, the aggregate throughput increases as the number of beams increases because a narrower beam stimulates the spatial reuse in the directional contention-based MAC protocol. As a result, the nodes have a greater opportunity to transmit simultaneously. The aggregate throughput of the deafness-free D2D communication scheme achieves the best performance among the compared schemes. This is because other directional contention-based MAC pro-

[4] Average operating time is defined as the average time from when the IoTD starts its identification to its return to the idle mode.

tocols suffer from deafness problems and cannot cope with them efficiently. In our simulation result, the omni-directional-identification-based MAC protocol (CSMA/CA) has a low throughput (approximately 520 Mbps). Because the simulation environment has a high-operating-frequency bandwidth, it has a short transmission distance, and the spatial frequency resources are not maximized even if a spatial frequency is found through the spectrum identification. On comparing the deafness-free MAC with and without spectrum identification, the performance of the former is found to be better because it harnesses the hidden spatial spectrum resources.

Fig. 11 presents the deafness duration for 24 IoTDs. The bar graph indicates that the deafness-free D2D communication scheme significantly reduces the deafness duration when compared with the other schemes. In particular, the deafness-free D2D communication scheme reduces the response time to 80% of the tone-based MAC protocol and 96% of the advanced-notice-based MAC protocol. This is because a sender that identifies a deaf node can immediately attempt to identify another idle node. Thus, the deafness duration of the proposed D2D scheme is similar to that of the omni-directional MAC. It should be noted that that no deafness problem occurs between IoTDs with omni-directional antennas. For D2D communication, the neighboring nodes of a transceiver can recognize that the nodes are in communication status because the communication request and response frames are transmitted in an omni-directional manner.

VI. Conclusion and Future Work

In this paper, we introduced mmWave spectrum identification/harvesting techniques for D2D communication with the use of conventional networks. The proposed spectrum identification/harvesting technique can find the optimal identification beam set and D2D links can be utilized by using identified spatial resources. In addition, we used a deafness-free D2D communication scheme that solves the deafness problem in D2D communication, and proved that the D2D communication scheme does not experience any deafness. Through simulation, we showed that our proposed scheme outperforms the existing conventional networks in terms of energy consumption, channel utilization, overall network throughput, and deafness duration. However, our computer simulations can overlook the computational plane elements. In our future works, we intend to reference other emulation-based tools such as Mininet and EmuEdge to evaluate the proposed scheme for a more realistic computation complexity.

References

[1] M. K. Afzal, Y. B. Zikria, S. Mumtaz, A. Rayes, A. Al-Dulaimi, and M. Guizani, "Unlocking 5G spectrum potential for intelligent IoT: Opportunities, challenges, and solutions," *IEEE Communications Magazine*, vol. 56, no. 10, pp. 92-93 2018.

[2] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26-35, May 2014.

[3] X. Wang, P. Sun, and Z. Wang, "A 3-D Self-Calibration Method for Multiple Base Stations in Large Complex Indoor Environment," *in Proc. of IEEE Wireless Communications and Networking Conference (WCNC)* pp. 1-6, April 2019.

[4] M. Polese, M. Giordani, T. Zugno, , A. Roy, S. Goyal, D. Castor, and M. Zorzi, "Integrated Access and Backhaul in 5G mmWave Networks: Potential and Challenges. IEEE Communications Magazine," vol. 58, no. 3, pp. 62-68, March 2019.

[5] N. N. Dao, M. Park, J. Kim, J. Paek, and S. Cho, "Resource-aware relay selection for inter-cell interference avoidance in 5G heterogeneous network for Internet of Things systems," *Future Generation Computer Systems*. (in press).

[6] H. R. Cheon, and J. H. Kim, "Social context-aware mobile data offloading algorithm via small cell backhaul networks," *IEEE Access*, vol. 7, pp. 39030-39040, 2019.

[7] C. Liu, J. Wang, X. Liu, and Y. C. Liang, "Maximum eigenvalue-based goodness-of-fit detection for spectrum sensing in cognitive radio," *IEEE Transactions on Vehicular Technology*, vol. 68, no.8, pp. 7747-7760, 2019.

[8] O. H. Toma, M. López-Benítez, D. K. Patel, and K. Umebayashi, "Estimation of primary channel activity statistics in cognitive radio based on imperfect spectrum sensing," *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2016-2031 2020.

[9] A. Bhowmick, M. K. Das, J. Biswas, S. D. Roy, and S. kundu, "Throughput Optimization with Cooperative Spectrum Sensing in Cognitive Radio Network," *in Proc. of IEEE IACC*, pp. 329-332, 2014.

[10] D. Treeumnuk, S. L. Macdonald, and D. C. Popescu, "Optimizing Performance of Cooperative Sensing for Increased Spectrum Utilization in Dynamic Cognitive Radio Systems," *in Proc. of IEEE ICC*, 2013.

[11] P. Cheng, R. Deng, and J. Chen, "Energy-efficient Cooperative Spectrum Sensing in Sensor-aided Cognitive Radio Networks," *IEEE Wireless Communications*, vol. 19, no. 6, pp. 100-105, 2012.

[12] H. Li, X. Cheng, K. Li, X. Xing, and T. Jing, "Utility-Based Cooperative Spectrum Sensing Scheduling in Cognitive Radio Networks," *in Proc. of IEEE INFOCOM*, pp. 165-169, 2013.

[13] Z. Lin, M. Lin, W. P. Zhu, J. B. Wang, and J. Cheng, "Robust secure beamforming for wireless powered cognitive satellite-terrestrial networks," *IEEE Transactions on Cognitive Communications and Networking*, 2020.

[14] W. Na, J. Yoon, S. Cho, D. Griffith, and N. Golmie, "Centralized Cooperative Directional Spectrum Sensing for Cognitive Radio Networks," *IEEE Transactions on Mobile Compting,* vol. 17, no. 6, pp. 1260-1274, 2018.

[15] B. F. Lo and I. F. Akildiz, "Reinforcement Learning for Cooperative Sensing Gain in Cognitive Radio Ad Hoc Networks," *in Proc. of IEEE PIMRC*, pp. 2244-2249, 2010.

[16] V. Raj, I. Dias, T. Tholeti, and S. Kalyani, "Spectrum access in cognitive radio using a two-stage reinforcement learning approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 20-34, 2018.

[17] P. Zhou, Y. Chang, and J. A. Copeland, "Reinforcement learning for repeated power control game in cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 1, pp. 54-69, 2011.

[18] Y. Wang, Z. Ye, P. Wan, and J. Zhao, "A survey of dynamic spectrum allocation based on reinforcement learning algorithms in cognitive radio networks," *Artificial Intelligence Review*, vol. 51 , no. 3, pp. 493-506, 2019.

[19] Y. Lin, C. Wang, J. Wang, and Z. Dou, "A novel dynamic spectrum access framework based on reinforcement learning for cognitive radio sensor networks," *Sensors*, vol. 16, no. 10, 2016.

[20] S. Jeong, W. Na, J. Kim, and S. Cho, "Internet of Things for Smart Manufacturing System: Trust Issues in Resource Allocation," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4418-4427, 2018.

[21] W. Na, L. Park, and S. Cho, "Deafness-aware MAC protocol for directional antennas in wireless ad hoc networks," *Elsevier Ad hoc Networks Journal*, vol. 24, Part A, pp. 121-134, January 2015.

[22] A. Maatouk, E. Calıskan, M. Koca, M. Assaad, G. Gui, and H. Sari, "Frequency-Domain NOMA With Two Sets of Orthogonal Signal Waveforms," *IEEE Communications Letters*, vol. 22, no. 5, pp. 906-909, 2018.

[23] Y. Li, M. Zhang, X. Cheng, M. Wen, and L. Q. Yang, "Index modulated OFDM with intercarrier interference cancellation," *in Proc. of IEEE ICC*, pp. 1-6, 2016.

[24] D. L. Wasden, H. Moradi, and B. Farhang-Boroujeny, "Design and implementation of an underlay control channel for cognitive radios," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 10, pp. 1875–1889, 2012.

[25] A. G.-S. and L. Giupponi "Distributed Q-Learning for Aggregated Interference Control in Cognitive Radio Networks," *IEEE Transactions on Vehicular Technology,* vol. 59, no. 4, pp. 1823-1834, 2010.

[26] C. J. Watkins, and P. Dayan, "Q-learning", *Machine learning*, vol. 8, no. 3-4, pp. 279-292, 1992.

[27] A. Akhtar and S. C. Ergen, "Directional MAC Protocol for IEEE 802.11ad based Wireless Local Area Networks," *Ad Hoc Networks*, vol. 69, pp. 49-64, 2018.

[28] T. Korakis, G. Jakllari, and L. Tassiulas, "A MAC protocol for full exploitation of directional antennas in ad hoc wireless networks", *in Proc. of Mobihoc*, 2003.

[29] J. Feng, P. Ren, and S. Yan, "A deafness free MAC protocol for ad hoc networks using directional antennas", *in Proc. of ICIEA*, 2009.

[30] M. Takata, M. Bandai, and T. Watanabe, "RI-DMAC: a receiver-initiated directional MAC protocol for deafness problem," *Int. J. Sensor Networks* vol. 5 pp. 79–89, 2009.

[31] H.-N. Dai, K.-W. Ng, and M.-Y. Wu, "On Busy-tone based MAC Protocol for Wireless Networks with Directional Antennas," *Wireless Personal Communications,* vol. 73, no. 3, pp. 611-636, 2013.

[32] A.A. Abdullah, L. Cai, F. Gebali, "DSDMAC: dual sensing directional MAC protocol for ad hoc networks with directional antennas", *Trans. Veh. Technol.* vol. 61, no. 3, 2012.