NAIST-IS-MT1651005

Master's Thesis

Statistical Coalition Formation for Cooperative Spectrum Sensing Based on the Multi-Armed Bandit Problem

Sho Iizuka

March 15, 2018

Graduate School of Information Science
Nara Institute of Science and Technology

A Master's Thesis
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
MASTER of ENGINEERING

Sho Iizuka

Thesis Committee:

Professor Shoji Kasahara	(Supervisor)
Professor Kazushi Ikeda	(Co-supervisor)
Associate Professor Masahiro Sasabe	(Co-supervisor)
Assistant Professor Jun Kawahara	(Co-supervisor)
Assistant Professor Yuanyu Zhang	(Co-supervisor)

Statistical Coalition Formation for Cooperative Spectrum Sensing Based on the Multi-Armed Bandit Problem*

Sho Iizuka

Abstract

In cognitive radio networks, the detection performance of wireless channel utilization is a key factor to improve the spectral efficiency. Cooperative spectrum sensing, with which a wireless channel is sensed by multiple cooperating secondary users, is a scheme for reducing sensing errors in cognitive radio networks. In order to improve the performance of cooperative spectrum sensing, it is necessary to form an appropriate coalition considering the sensing performance of secondary users. In this thesis, we provide a coalition formation approach for cooperative spectrum sensing based on the multi-armed bandit problem. This approach eliminates the assumption that each secondary user knows not only his/her own sensing performance but also other users' performance. In the proposed method, the sensing performance of secondary users are estimated only from the information that secondary users can obtain and coalition formation is performed considering exploitation-exploration trade-offs. Numerical examples show that the proposed method can reduce miss-detections and false-alarms especially in a scenario where the number of secondary users is large and the performance difference among secondary users is significant.

Keywords:

cognitive radio networks, cooperative spectrum sensing, multi-armed bandit problem, machine learning, variational Bayesian inference

*Master's Thesis, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1651005, March 15, 2018.

多腕バンディット問題に基づく協調スペクトルセンシングのための統計的グループ形成*

飯塚 翔

内容梗概

割当済みの無線チャネルをセンシングし，空き時間にチャネルを二次利用することで周波数帯の利用効率を改善する方法として，コグニティブ無線システムが提案されている．コグニティブ無線システムでは，セカンダリユーザによる周波数チャネルのセンシング性能を改善するために複数台のセカンダリユーザによる協調センシングが提案されているが，センシング性能を改善するためには適切なグループを形成する必要がある．

本稿では，不確実な環境における協調センシングのためのグループ形成の逐次的意思決定について述べる．このアプローチでは，セカンダリユーザが自分自身や他のセカンダリユーザのセンシング性能についての事前知識を持たない場合でも，通信結果から得られるフィードバックをもとにセンシング性能を学習しながらグループ形成の意思決定を行う．提案手法ではこの問題を多腕バンディット問題とみなし，多腕バンディット問題に対する方策のひとつである Thompson sampling に基づいてグループ形成を行う．数値実験による評価によって，セカンダリユーザ間の性能差が大きく，セカンダリユーザの台数が多いような環境において，提案手法が特に効果的に機能することが示された．

キーワード

コグニティブ無線，協調スペクトルセンシング，多腕バンディット問題，機械学習，変分ベイズ推定

*奈良先端科学技術大学院大学 情報科学研究科 修士論文, NAIST-IS-MT1651005, 2018 年 3 月 15 日.

Contents

List of Figures

List of Tables

1. Introduction

In order to utilize limited radio resources, techniques for effectively utilizing frequency bands called cognitive radio networks (CRN) have been proposed [1, 2]. In CRN, the spectrum channel used by a primary user (PU) is sensed by secondary users (SUs). While the PU can preferentially use his/her dedicated channel, SUs can use the channel secondarily only during periods when the PU is not utilizing the channel. It is important for CRN to eliminate the shortage of usable frequency bands by allowing the SUs to opportunistically use the channel of the PU without causing interference to the PU.

When an SU senses a wireless channel, sensing error occasionally occurs due to multipath fading/shadowing. Sensing errors are classified into miss-detection and false-alarm. The miss-detection is a sensing error that an SU mistakenly regards the channel as *idle* despite the PU is utilizing the channel. The false-alarm is that an SU mistakenly regards the channel as *busy* while the PU is not utilizing the channel. Miss-detections cause collisions with the communication of the PU, while false-alarms cause the loss of communication opportunities of SUs.

In order to improve the performance of the spectrum sensing of SUs, cooperative spectrum sensing (CSS) has been proposed, where multiple SUs exchange sensing results of a channel, and the state of the channel is determined based on the exchanged sensing results [3, 4]. In terms of coalition formation for CSS, game theory-based approaches have been proposed [5, 6], where the utility of a coalition to form is calculated based on the sensing performance of SUs in the coalition, and SUs act to maximize its utility. However, these approaches assume that each SU knows not only his/her sensing performance but also other SUs' one.

In this thesis, we propose a statistical coalition formation approach which works without the information of sensing performance for SUs a priori. In this approach, the sensing performance of SUs is dynamically estimated and coalition formation is performed with the aim of reducing miss-detections and false-alarms.

Examples of this approach are as follows. Figure 1 shows SUs placed in an unknown radio wave environment. Some SUs can detect the utilization of the channel by a PU, whereas other SUs cannot detect due to obstacles such as buildings. In this scenario, SUs can perform coalition formation according to the

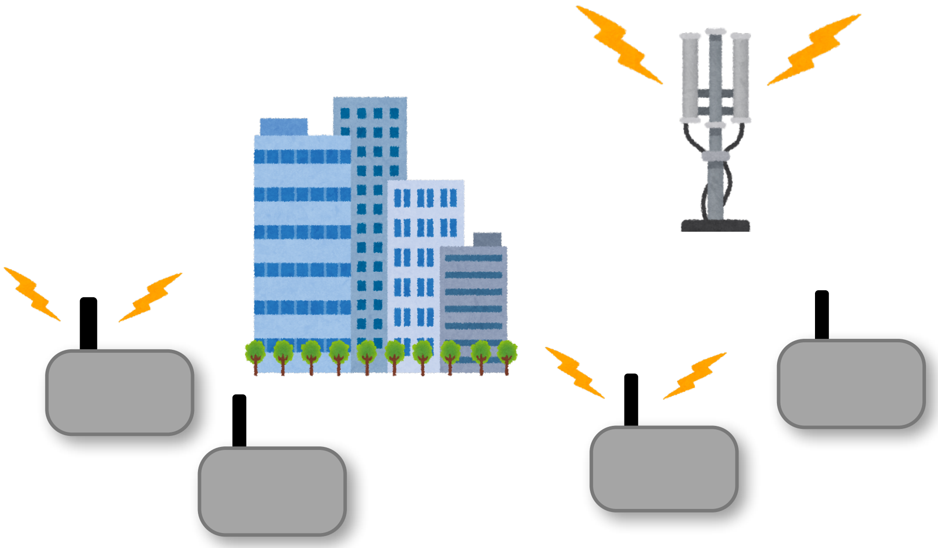

Figure 1: SUs placed in an unknown radio wave environment. Some SUs can detect the utilization of the channel by a PU. However, other SUs can fail to detect due to obstacles such as buildings.

environment of wireless communication by learning the sensing performance of other SUs. Figure 2 shows the mobility of SUs and SUs need to follow the change of a radio wave environment. Also in this scenario, an SU can appropriately perform coalition formation by learning the sensing performance at the place after SU movement. Figure 3 shows SUs placed in a harsh environment. In this scenario, the sensing performance of SUs can be degraded due to breakdowns. Figure 4 shows a situation where some SUs join/leave the system. It is important for the existing SUs to obtain the sensing performance of the new SUs.

The proposed approach is based on the multi-armed bandit problem. In this approach, the sensing performance of SUs is estimated only from the information that SUs can obtain and a coalition is formed using the estimated performance. After a communication is performed by a formed coalition, the estimated performance is updated using the data obtained through the communication. By utilizing the framework of the multi-armed bandit problem, a cycle of estimations, coalition formation, and communications is performed in order to reduce miss-detections and false-alarms. The advantages of the proposed approach are

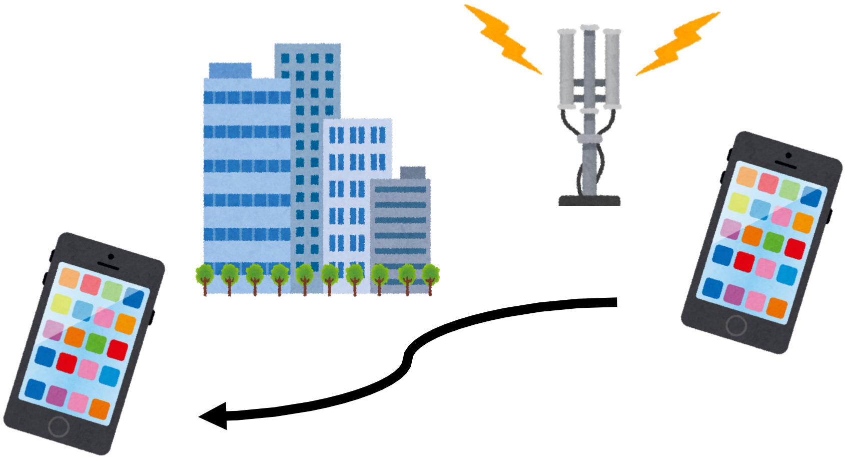

Figure 2: The mobility of SUs. SUs need to follow the change of a wireless communication environment.

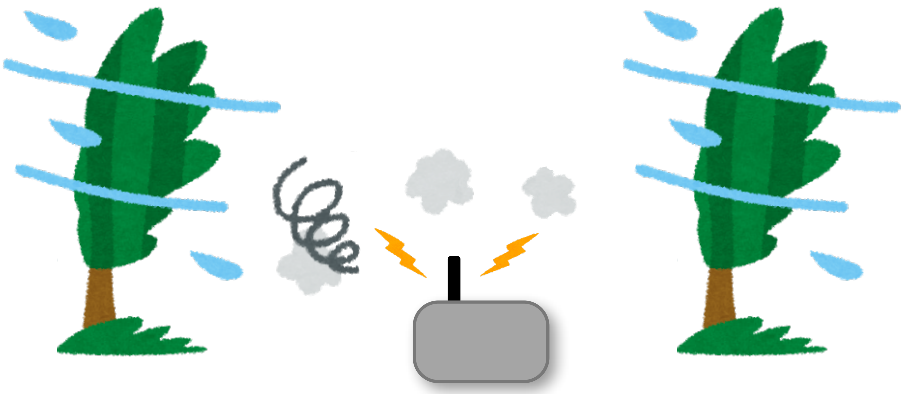

Figure 3: SUs placed in a harsh environment. The detection performance of SUs can be decreased due to breakdowns.

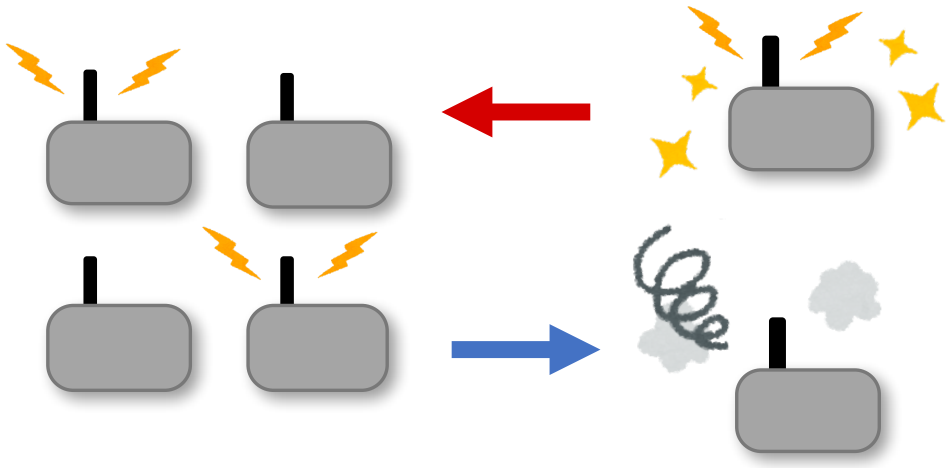

Figure 4: SUs join/leave. The existing SUs need to know the detection performance of the new SUs.

as follows.

- This approach does not need the information of sensing performance of SUs in the system a priori.
- This approach estimates the sensing performance of SUs dynamically and performs coalition formation in order to reduce miss-detections and false-alarms.
- This approach makes decisions based on the combinatorial multi-armed bandit problem. Although the number of combinations of coalitions increases exponentially as the number of SUs increases, the time and space complexity of this approach increases linearly as the number of SUs increases.
- This approach estimates the sensing performance of SUs only from the data that SUs can obtain. The expectation-maximization (EM) algorithm is used for maximum likelihood estimation and the variational Bayes EM is used to estimate the posterior distributions.

We evaluate the proposed method in comparison with a naive baseline approach based on the ε-greedy strategy through three numerical examples with different settings of the number of SUs and performance among SUs. The evaluations are conducted in terms of the cumulative reward, the detection probability, the false-alarm probability, the expected cumulative occurrences of miss-detections and false-alarms, and the expected regret.

The organization of this thesis is as follows. First, in Section 2 we describe related work, and in Section 3 we describe our model and formulation. In Section 4, we propose the coalition formation policy based on the ε-first strategy, which is a baseline method of this problem. Then, we describe the disadvantage of the ε-first strategy and propose the coalition formation policy based on the multi-armed bandit problem. Section 5 presents the numerical examples and Section 6 concludes this thesis.

2. Related Work

Lo and Akyildiz have proposed a cooperative spectrum sensing model for cognitive radio ad hoc networks (CRAHN) based on reinforcement learning [7]. In this model, SUs learn a set of cooperating neighbors to improve the sensing performance under correlated shadowing while considering cooperative sensing delays and energy efficiency. Since SUs make sequential decisions on selecting cooperating neighbors in iterations, the cooperative sensing problem is formulated as a finite-horizon Markov decision process (MDP). Rewards are given when an SU cooperates with other SUs with low cooperative sensing delay and low correlated shadowing. Their model is similar to ours in the sense that both models adopt the idea of adaptive coalition formation. However, their research focuses on reducing cooperative sensing delays and correlated shadowing among a coalition, whereas our model focuses on reducing miss-detections and false-alarms under a situation where SUs do not know the sensing performance of SUs.

Panahi and Ohtsuki have proposed a decision-making scheme for cognitive radio systems based on Partially Observable Markov Decision Process (POMDP) [8]. In this scheme, the channel utilization by a PU at each time step is considered as the *state*, a sensing result by an SU is considered as the *observation*, and data transmission, stop data transmission, and channel switching by an SU at each time step are considered as *actions*. A reward is given to an SU when he/she succeeds in communication, and a penalty is imposed on an SU when a collision occurs with the communication of the PU. Since it is assumed that the state transition of the channel is Markovian and errors may occur in the spectrum sensing, the behavior of an SU is determined by the POMDP to maximize cumulative reward. Notice that their research focuses on the problem of communication control of cognitive radio networks, which is different from the coalition formation problem in this thesis. Our model uses the design of rewards and penalties as described in Section 3.

3. Model and Formulation

In our model, SUs form one coalition and perform CSS and communication utilizing a wireless channel for each time step in $\mathcal{T} = \{1, 2, \ldots, T\}$. Let N be the number of SUs in the system and $\mathcal{N}$ be the set of all the SUs $\{1, 2, \ldots, N\}$. The coalition consists of M SUs from $\mathcal{N}$, where M is a predefined parameter representing the number of the SUs included in the coalition to be formed. We assume that SU 1 must belong to the coalition since SU 1 leads coalition formation and makes global detection decisions. That is, the formed coalition $\mathcal{C}_t$ at the t-th time step is expressed as

$$\mathcal{C}_t = \mathcal{C}' \cup \{1\}, \quad \text{where} \quad \mathcal{C}' \subseteq \{2, \ldots, N\} \text{ and } |\mathcal{C}'| = M - 1.$$

Note that exactly one coalition is formed in the system, and the SUs not belonging to the coalition formed by SU 1 do not form coalitions.

After the formation of the coalition $\mathcal{C}_t$ at the t-th time step, CSS and a communication utilizing the channel are performed. The exchanges of control messages and sensing results within the coalition are done via a common control channel. The SUs aim to utilize the channel as many times as possible while reducing collisions with a PU during the T time steps. The PU utilizes his/her dedicated channel according to a Bernoulli process with parameter ρ $(0 \leq \rho \leq 1)$. In other words, the channel is *busy* with probability ρ or *idle* with probability $1 - \rho$ in each time step. Note that ρ is the channel utilization by the PU. It is assumed that the SUs do not know the channel utilization ρ.

The SUs perform CSS within the coalition $\mathcal{C}_t$. If the coalition regards the channel as *idle*, the coalition tries to utilize the channel. On the contrary, if the coalition regards the channel as *busy*, the coalition does nothing during the time step. If the PU does not use the channel when the coalition tries to utilize the channel, the communication of the coalition succeeds. However, if the PU uses the channel when the coalition tries to utilize the channel, a collision occurs. If the coalition does not try to utilize the channel when the channel is actually idle, the SUs lose a communication opportunity at the time step.

We define $\lambda_{\mathrm{D}}^{(i)}$ $(0 \leq \lambda_{\mathrm{D}}^{(i)} \leq 1)$ and $\lambda_{\mathrm{F}}^{(i)}$ $(0 \leq \lambda_{\mathrm{F}}^{(i)} \leq 1)$ as the detection probability and false-alarm one of the i-th SU $(i \in \mathcal{N})$, respectively. Note that $\lambda_{\mathrm{D}}^{(i)}$ is the probability that the i-th SU correctly regards the channel as *busy* given that the channel is used by the PU at a time step. Note also that $\lambda_{\mathrm{F}}^{(i)}$ is the probability that the i-th SU mistakenly regards the channel as *busy* given that the channel is not used by the PU at a time step. It is assumed that miss-detections and false-alarms occur independently in time. Note that i-th SU ($\forall i \in \mathcal{N}$) does not know $\lambda_{\mathrm{D}}^{(j)}$ and $\lambda_{\mathrm{F}}^{(j)}$ for $\forall j \in \mathcal{N}$. That is, each SU does not know the detection probability and the false-alarm one of all SUs a priori.

The coalition estimates the channel state according to the k-out-of-N rule [9, 10, 11, 12] based on the sensing results of the SUs in the coalition. The coalition regards the channel as *busy* if and only if at least k SUs in $\mathcal{C}_t$ regard the channel as *busy*, where k is a predefined parameter for CSS. Assuming that miss-detections and false-alarms occur independently among the SUs, the detection probability $\Lambda_{\mathrm{D}}(\mathcal{C}_t)$ and the false-alarm one $\Lambda_{\mathrm{F}}(\mathcal{C}_t)$ of CSS in the coalition $\mathcal{C}_t$ are expressed as

$$\Lambda_{\mathrm{D}}(\mathcal{C}_t) = \sum_{\mathcal{C} \subseteq \mathcal{C}_t, |\mathcal{C}| \geq k} \prod_{i \in \mathcal{C}} \lambda_{\mathrm{D}}^{(i)} \prod_{i \in \mathcal{C}_t \setminus \mathcal{C}} \left(1 - \lambda_{\mathrm{D}}^{(i)}\right), \tag{1}$$

$$\Lambda_{\mathrm{F}}(\mathcal{C}_t) = \sum_{\mathcal{C} \subseteq \mathcal{C}_t, |\mathcal{C}| \geq k} \prod_{i \in \mathcal{C}} \lambda_{\mathrm{F}}^{(i)} \prod_{i \in \mathcal{C}_t \setminus \mathcal{C}} \left(1 - \lambda_{\mathrm{F}}^{(i)}\right). \tag{2}$$

In the above equations, it is necessary to enumerate subsets of $\mathcal{C}_t$ explicitly to compute the values inside the summation operators one by one. Instead, by using the following recurrence relations, $\Lambda_{\mathrm{D}}(\mathcal{C}_t)$ and $\Lambda_{\mathrm{F}}(\mathcal{C}_t)$ can be calculated in computational complexity $\mathcal{O}(k|\mathcal{C}_t|)$ by a dynamic programming approach.

$$f(\lambda, \mathcal{C}, k) = \begin{cases} 1, & \text{if } |\mathcal{C}| = 0 \\ 0, & \text{if } |\mathcal{C}| > 0 \text{ and } k = 0 \\ \lambda^{(i)} f(\lambda, \mathcal{C} \setminus \{i\}, k-1) & \\ \quad + (1 - \lambda^{(i)}) f(\lambda, \mathcal{C} \setminus \{i\}, k), \ \forall i \in \mathcal{C}, & \text{otherwise.} \end{cases}$$

The goodness of the coalition formation in $\mathcal{T}$ is evaluated by the following utility.

$$U(\mathcal{T}) = \alpha N_{\mathrm{S}}(\mathcal{T}) - (1-\alpha) N_{\mathrm{MD}}(\mathcal{T}), \tag{3}$$

where $N_{\mathrm{S}}(\mathcal{T})$ is the number of successful communications in $\mathcal{T}$, $N_{\mathrm{MD}}(\mathcal{T})$ is the number of collisions with the PU in $\mathcal{T}$, and α ($0 \leq \alpha \leq 1$) is a known parameter for adjusting the balance between $N_{\mathrm{S}}(\mathcal{T})$ and $N_{\mathrm{MD}}(\mathcal{T})$. When α is close to 1, the number of successful communications accounts for a large percentage in the utility. When α is close to 0, on the other hand, the number of collisions with the PU accounts for a large percentage in the utility. This utility is based on the research by Panahi and Ohtsuki [8] mentioned in Section 2. The expectation of the utility $\mathbb{E}[U(\mathcal{T})]$ is represented by

$$\mathbb{E}[U(\mathcal{T})] = \sum_{t \in \mathcal{T}} u(\mathcal{C}_t), \tag{4}$$

where

$$u(\mathcal{C}_t) = \alpha(1-\rho)(1-\Lambda_{\mathrm{F}}(\mathcal{C}_t)) - (1-\alpha)\rho(1-\Lambda_{\mathrm{D}}(\mathcal{C}_t)).$$

If ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ are known, a coalition which maximizes the expected utility at a time step $u(\mathcal{C})$ maximizes $\mathbb{E}[U(\mathcal{T})]$.

4. Proposed Method

In this section, we describe two approaches for coalition formation. The first approach is based on the ε-first strategy, which is regarded as the baseline approach. The second one is based on the multi-armed bandit problem, which is shown to be superior to the first one in Section 5.

4.1 Coalition Formation Policy Based on the ε-first Strategy

In order to maximize the expectation of the utility $U(\mathcal{T})$ of (3), it is sufficient to form a coalition that maximizes the expected utility $u(\mathcal{C}_t)$ in (4). In our model, however, each SU does not know $\lambda_{\mathrm{D}}^{(i)}$ and $\lambda_{\mathrm{F}}^{(i)}$ of all the SUs. Thus, it is necessary to perform coalition formation and the estimation of ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ at the same time.

The policy based on the ε-first strategy [13] is one of the most naive approaches to perform coalition formation and parameter estimation. In this policy, the whole time steps in $\mathcal{T}$ are divided into the exploration steps $\{1, 2, \ldots, T'\}$ and the exploitation steps $\{T'+1, T'+2, \ldots, T\}$. In the exploration steps, the coalitions of M SUs are randomly formed to observe data necessary to estimate ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$. After the exploration steps, the coalition $\mathcal{C}_{\mathrm{opt}}$ that maximizes $u(\mathcal{C}_{\mathrm{opt}})$ is searched by taking into account the resulting ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$, estimated from the obtained data. In the exploitation steps, the coalition $\mathcal{C}_{\mathrm{opt}}$ is formed to maximize $U(\mathcal{T})$.

Algorithm 1 shows the details of the coalition formation based on the ε-first strategy. In this algorithm, Random(N,M) is a function that randomly forms a coalition, where N is the number of all the SUs and M is the number of the SUs included in the coalition to be formed. Collect(i) is a function to collect the sensing result on the channel from the i-th SU. If the i-th SU regards the channel as *busy*, it returns 1; otherwise, it returns 0. The returned values are stored to $s_t = (s_t^{(1)}, s_t^{(2)}, \ldots, s_t^{(N)})$. Communicate($s_t$) is a function that tries to communicate with the channel based on sensing results from the SUs belonging in $\mathcal{C}_t$. If the coalition regards the channel as *idle* and eventually succeeds in communication with the channel, it returns 1. If a collision occurs due to miss-

Algorithm 1 Coalition formation based on the ε-first strategy

procedure CoalitionFormation(N, M, T, T', α)
    Create lists $CS[1, T']$, $S[1, T']$, and $Y[1, T']$ to store $\mathcal{C}_t$, s_t, and y_t
    for $t = 1, 2, \ldots, T'$ **do**
        $\mathcal{C}_t \leftarrow$ Random(N, M)   ▷ Randomly select M SUs from $\{1, 2, \ldots, N\}$
        for all $i \in \mathcal{C}_t$ **do**
            $s_t^{(i)} \leftarrow$ Collect(i)   ▷ Collect sensing result of the i-th SU
        end for
        $y_t \leftarrow$ Communicate(s_t)   ▷ Communicate by taking into account s_t
        $CS[t] \leftarrow \mathcal{C}_t$
        $S[t] \leftarrow s_t$
        $Y[t] \leftarrow y_t$
    end for
    $(\rho, \lambda_{\mathrm{D}}, \lambda_{\mathrm{F}}) \leftarrow$ EstimateEM(CS, S, Y)   ▷ Estimate ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ using the EM algorithm
    $\mathcal{C}_{\mathrm{opt}} \leftarrow$ FindOptimal(N, M, α, ρ, λ_{D}, λ_{F})   ▷ Find the coalition that maximizes $u(\mathcal{C}_{\mathrm{opt}})$ by taking into account estimated ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$
    for $t = T' + 1, T' + 2, \ldots, T$ **do**
        $\mathcal{C}_t \leftarrow \mathcal{C}_{\mathrm{opt}}$
        $s_t \leftarrow$ Collect($\mathcal{C}_t$)
        $y_t \leftarrow$ Communicate(s_t)
    end for
end procedure

detection, it returns 2. If the coalition does nothing during the time step due to the decision of *busy*, it returns 3. The returned value is assigned to y_t.

ESTIMATEEM(CS, S, Y) is a function for the maximum likelihood estimation of ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ by the EM algorithm [14] based on the data observed in the exploration steps. The reason why the EM algorithm is required is that the actual utilization of the channel by the PU at each time step is a hidden variable. To be specific, the actual utilization of the channel by the PU is unknown when $y_t = 3$, i.e., the coalition does not communicate at the t-th time step. There are two possible reasons for the case of $y_t = 3$. The first one is that the coalition successfully detects that the channel is used by the PU. The second possible reason is that a false-alarm occurs when the channel is idle. Details of how to estimate ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ by the EM algorithm are described in Algorithm 2 and Appendix A. Note that, $\mathcal{T}_j$ in Algorithm 2 is defined as $\{t \mid t \in \mathcal{T} \wedge y_t = j\}$ for $j = 1, 2, 3$ and $\mathcal{T}_j^{(i)}$ is defined as $\{t \mid t \in \mathcal{T}_j \wedge i \in \mathcal{C}_t\}$.

FINDOPTIMAL(N, M, α, ρ, λ_{D}, λ_{F}) is a function that finds a coalition $\mathcal{C}$ such that $u(\mathcal{C})$ is maximized with estimated ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$. If N is small and the cost of enumerating all coalitions consisting of M SUs is small, the optimal $\mathcal{C}_{\mathrm{opt}}$ can be found by the brute-force search shown in Algorithm 3. However, if N is large and enumerating cost is large, it is hard to find the optimal $\mathcal{C}_{\mathrm{opt}}$ by the brute-force search. Algorithm 4 shows the heuristics to find $\mathcal{C}_{\mathrm{opt}}$ such that $\mathcal{C}_{\mathrm{opt}}$ is calculated in computational complexity $\mathcal{O}(NM)$. In this heuristics, the search starts with the coalition $\mathcal{C}_{\mathrm{opt}}$ including all the SUs, and repeatedly removes the i-th SU from $\mathcal{C}_{\mathrm{opt}}$ such that $\mathcal{C}_{\mathrm{opt}} \setminus \{i\}$ maximizes $u(\mathcal{C}_{\mathrm{opt}} \setminus \{i\})$.

Algorithm 2 Maximum likelihood estimation based on the EM algorithm.

function ESTIMATEEM(CS, S, Y)
  $\rho \leftarrow 0.5$
  for $i = 1, 2, \ldots, N$ **do**
    $\lambda_{\mathrm{D}}^{(i)} \leftarrow 0.5$, $\lambda_{\mathrm{F}}^{(i)} \leftarrow 0.5$
  end for
  while ρ, $\lambda_{\mathrm{D}}^{(i)}$, or $\lambda_{\mathrm{F}}^{(i)}$ change due to updating **do**
    for $t = 1, 2, \ldots, T$ **do**
      $a \leftarrow \log \rho + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{D}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{D}}^{(i)})$
      $b \leftarrow \log(1 - \rho) + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{F}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{F}}^{(i)})$
      $r_t \leftarrow 1/(1 + \exp(b - a))$
    end for
    $a[\rho] \leftarrow |\mathcal{T}_2| + \sum_{t \in \mathcal{T}_3} r_t$
    $b[\rho] \leftarrow |\mathcal{T}_1| + \sum_{t \in \mathcal{T}_3} (1 - r_t)$
    $\rho \leftarrow a[\rho]/(a[\rho] + b[\rho])$
    for $i = 1, 2, \ldots, N$ **do**
      $a[\lambda_{\mathrm{D}}^{(i)}] \leftarrow \sum_{t \in \mathcal{T}_2^{(i)}} s_t^{(i)} + \sum_{t \in \mathcal{T}_3^{(i)}} r_t s_t^{(i)}$
      $b[\lambda_{\mathrm{D}}^{(i)}] \leftarrow \sum_{t \in \mathcal{T}_2^{(i)}} (1 - s_t^{(i)}) + \sum_{t \in \mathcal{T}_3^{(i)}} r_t (1 - s_t^{(i)})$
      $\lambda_{\mathrm{D}}^{(i)} \leftarrow a[\lambda_{\mathrm{D}}^{(i)}]/(a[\lambda_{\mathrm{D}}^{(i)}] + b[\lambda_{\mathrm{D}}^{(i)}])$
      $a[\lambda_{\mathrm{F}}^{(i)}] \leftarrow \sum_{t \in \mathcal{T}_1^{(i)}} s_t^{(i)} + \sum_{t \in \mathcal{T}_3^{(i)}} (1 - r_t) s_t^{(i)}$
      $b[\lambda_{\mathrm{F}}^{(i)}] \leftarrow \sum_{t \in \mathcal{T}_1^{(i)}} (1 - s_t^{(i)}) + \sum_{t \in \mathcal{T}_3^{(i)}} (1 - r_t)(1 - s_t^{(i)})$
      $\lambda_{\mathrm{F}}^{(i)} \leftarrow a[\lambda_{\mathrm{F}}^{(i)}]/(a[\lambda_{\mathrm{F}}^{(i)}] + b[\lambda_{\mathrm{F}}^{(i)}])$
    end for
  end while
  return $(\rho, \lambda_{\mathrm{D}}, \lambda_{\mathrm{F}})$
end function

Algorithm 3 The brute-force search for $\mathcal{C}_{\mathrm{opt}}$ which maximizes $u(\mathcal{C}_{\mathrm{opt}})$.

function FINDOPTIMAL(N, M, α, ρ, λ_{D}, λ_{F})
  $CS \leftarrow \{\mathcal{C}' \cup \{1\} \mid \mathcal{C}' \subseteq \{2, \ldots, N\} \wedge |\mathcal{C}'| \neq M - 1\}$
  $\mathcal{C}_{\mathrm{opt}} \leftarrow \arg\max_{\mathcal{C} \in CS} u(\mathcal{C})$
  return $\mathcal{C}_{\mathrm{opt}}$
end function

Algorithm 4 Heuristics to find $\mathcal{C}_{\text{opt}}$ which maximizes $u(\mathcal{C}_{\text{opt}})$.

function FINDOPTIMAL(N, M, α, ρ, λ_{D}, λ_{F})

  $\mathcal{C}_{\text{opt}} \leftarrow \{1, 2, \dots, N\}$

  while $|\mathcal{C}_{\text{opt}}| > M$ **do**

    $CS \leftarrow \{\mathcal{C}_{\text{opt}} \setminus \{i\} \mid i \in \mathcal{C}_{\text{opt}} \setminus \{1\}\}$

    $\mathcal{C}_{\text{opt}} \leftarrow \arg\max_{\mathcal{C} \in CS} u(\mathcal{C})$

  end while

  return $\mathcal{C}_{\text{opt}}$

end function

4.2 Coalition Formation Policy Based on the Multi-Armed Bandit Problem

The coalition formation based on the ε-first strategy has two disadvantages. The first one is that it is necessary to appropriately set T' which represents the length of the exploration steps in advance. If the length of the exploration steps is long, the estimation of ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ becomes accurate, while the coalitions with low $u(\mathcal{C}_t)$ are selected many times during the exploration steps. On the other hand, if the length of the exploration steps is short, it is possible to perform coalition formation many times in the exploitation steps with the coalition whose $u(\mathcal{C}_t)$ is estimated to be the maximum. However, the estimation of ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ becomes incorrect due to the shortage of the observed data. The second disadvantage is that coalitions are randomly formed in the exploration steps. In order to reduce miss-detections and false-alarms occurring in $\mathcal{T}$, it is no longer necessary to accurately estimate the sensing performance of SUs which are expected to have low sensing performance.

The multi-armed bandit problem is an optimal-policy problem of considering exploitation-exploration trade-offs [15]. In the simplest case called the Bernoulli bandit [16], there are K slot machines each of which has one arm, and when a player pulls the arm of the k-th slot machine, the given reward is 1 (resp. 0) with probability λ_k (resp. $1 - \lambda_k$). He/she does not know λ_k for each k. He/she can pull arms T times and his/her aim is to maximize the sum of the rewards.

UCB-1 [17] and Thompson sampling [18, 19] have been proposed as policies against the Bernoulli bandit, and the performance of them is theoretically analyzed. In particular, Thompson sampling is developed for the case where the reward follows a distribution belonging to exponential family [20], including normal distribution [21].

The procedure of Thompson sampling for Bernoulli bandit is as follows. Since the conjugate prior distribution of the Bernoulli distribution is the beta distribution [14], for each decision making, the posterior distribution of λ_k for each k is given by

$$\begin{aligned} p(\lambda_k \mid D) &= \frac{p(D \mid \lambda_k) p(\lambda_k)}{P(D)} \\ &= \frac{\lambda_k^{S[k]}(1-\lambda_k)^{F[k]}}{\int_0^1 t^{S[k]}(1-t)^{F[k]} \mathrm{d}t} \\ &= \beta(\lambda_k; S[k]+1, F[k]+1), \end{aligned}$$

where

$$\beta(x; a, b) = \frac{x^{a-1}(1-x)^{b-1}}{\int_0^1 t^{a-1}(1-t)^{b-1} \mathrm{d}t}.$$

In the above equations, D is the observed data, $S[k]$ (resp. $F[k]$) is the number of times that the reward obtained from the k-th arm is 1 (resp. 0), and $\beta(x; a, b)$ is the beta distribution [22]. It is assumed that the prior distribution of λ_k is the uniform distribution.

After calculating the posterior distribution of λ_k for each arm, a sample is generated from the posterior distribution for each k. Then, the arm which generates the largest sample is pulled. Finally, the reward obtained by pulling the arm is stored in $S[k]$ and $F[k]$. The pseudocode of the procedure of Thompson sampling is shown in Algorithm 5.

In order to apply the multi-armed bandit problem to coalition formation, a selection of coalitions to form is regarded as an arm selection in the multi-armed bandit problem. We can regard each coalition as an arm and simply apply the framework of the multi-armed bandit problem. However, there are several issues.

The first problem is that the number of possible coalitions increases exponentially since patterns of coalitions are generated as combinations of the SUs. The second problem is that despite the fact that data on other arms is observed partially when an arm is pulled, that observed data is not utilized. For instance, when a coalition that consists of SUs $\{1, 2, 3\}$ performs CSS and a communication, not only data on a coalition consists of SUs $\{1, 2, 3\}$ but also information on other coalitions sharing some SUs such as $\{2, 3, 4\}$ and $\{2, 3, 5\}$ can also be obtained partially. In a naive approach which regards each coalition as an independent arm, data observed when another arm is pulled cannot be utilized. This is an explanation of why it is considered that the classical framework of the multi-armed bandit problem is less effective. Since one super arm is expressed as a combination of simple arms, this problem is considered to be one of the

Algorithm 5 Procedure of Thompson sampling for Bernoulli bandit.

```
procedure ThompsonSampling(T, K)
    Create lists S[1, K] and F[1, K] to store the rewards gained from each arm
    for k = 1, 2, ..., K do
        S[k] ← 0, F[k] ← 0
    end for
    for t = 1, 2, ..., T do
        for k = 1, 2, ..., K do
            Sample λ_k ~ β(λ_k; S[k] + 1, F[k] + 1)
        end for
        k_opt = arg max_k λ_k
        r_k ← Pull(k_opt)                ▷ pull the k_opt-th arm and obtain reward r_k
        (S[k], F[k]) ← (r_k, 1 − r_k)
    end for
end procedure
```

families of the combinatorial multi-armed bandit problems [23]. Several studies show that Thompson sampling also exhibits superior performance experimentally even in the combinatorial multi-armed bandit problem, and its performance is theoretically analyzed in simple cases [24, 25].

In the proposed method, a procedure to calculate the posterior distributions of ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$, to generate samples from the posterior distribution, to find $\mathcal{C}_{\mathrm{opt}}$ which maximizes $u(\mathcal{C}_{\mathrm{opt}})$ and to form $\mathcal{C}_{\mathrm{opt}}$ is done in order at each time step. Algorithm 6 shows the details of the coalition formation based on the multi-armed bandit problem.

In the calculation of the posterior distributions, we have to consider that the actual utilization of the channel by the PU at each time step is a hidden variable, as described in Section 4.1. The calculation of the posterior distributions of ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ is not as simple as in the case of Bernoulli bandit. For this reason, the posterior distributions of ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ are calculated using the variational Bayes EM [26, 27]. The details of how to estimate the posterior distributions of ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ by the variational Bayes EM are described in Algorithm 7 and Appendix B. $\psi(\cdot)$ in Algorithm 7 is called the digamma function [22].

Algorithm 6 Coalition formation based on the multi-armed bandit problem

procedure COALITIONFORMATION(N, M, T, T', α)
  Create lists $CS[1,T]$, $S[1,T]$, and $Y[1,T]$ to store $\mathcal{C}_t$, s_t, and y_t
  for $t = 1, 2, \ldots, T$ **do**
    $(a[\cdot], b[\cdot]) \leftarrow$ ESTIMATEVB(CS, S, Y)
    Sample $\rho \sim \beta(\rho; a[\rho], b[\rho])$
    for $i = 1, 2, \ldots, N$ **do**
      Sample $\lambda_{\mathrm{D}}^{(i)} \sim \beta(\lambda_{\mathrm{D}}^{(i)}; a[\lambda_{\mathrm{D}}^{(i)}], b[\lambda_{\mathrm{D}}^{(i)}])$
      Sample $\lambda_{\mathrm{F}}^{(i)} \sim \beta(\lambda_{\mathrm{F}}^{(i)}; a[\lambda_{\mathrm{F}}^{(i)}], b[\lambda_{\mathrm{F}}^{(i)}])$
    end for
    $\mathcal{C}_{\mathrm{t}} \leftarrow$ FINDOPTIMAL(N, M, α, ρ, λ_{D}, λ_{F}) ▷ Find the coalition that maximizes $u(\mathcal{C}_t)$ by taking into account sampled ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$
    for all $i \in \mathcal{C}_t$ **do**
      $s_t^{(i)} \leftarrow$ COLLECT(i) ▷ Collect sensing result of the i-th SU
    end for
    $y_t \leftarrow$ COMMUNICATE(s_t) ▷ Communicate by taking into account s_t
    $CS[t] \leftarrow \mathcal{C}_t$
    $S[t] \leftarrow s_t$
    $Y[t] \leftarrow y_t$
  end for
end procedure

Algorithm 7 Estimation of the posterior distribution of ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ using the variational Bayes EM.

function ESTIMATEVB(CS, S, Y)
    $a[\rho] \leftarrow 1$, $b[\rho] \leftarrow 1$
    for $i = 1, 2, \ldots, N$ **do**
        $a[\lambda_{\mathrm{D}}^{(i)}] \leftarrow 1$, $b[\lambda_{\mathrm{D}}^{(i)}] \leftarrow 1$
        $a[\lambda_{\mathrm{F}}^{(i)}] \leftarrow 1$, $b[\lambda_{\mathrm{F}}^{(i)}] \leftarrow 1$
    end for
    while $a[\rho]$, $b[\rho]$, $a[\lambda_{\mathrm{D}}^{(i)}]$, $b[\lambda_{\mathrm{D}}^{(i)}]$, $a[\lambda_{\mathrm{F}}^{(i)}]$, or $b[\lambda_{\mathrm{F}}^{(i)}]$ change due to updating **do**
        for $t = 1, 2, \ldots, T$ **do**

$$
\begin{aligned}
\eta_t \leftarrow\ & (\psi(a[\rho]) - \psi(b[\rho])) \\
& + \textstyle\sum_{i \in \mathcal{C}_t} s_t^{(i)} \left(\psi\left(a[\lambda_{\mathrm{D}}^{(i)}]\right) - \psi\left(a[\lambda_{\mathrm{D}}^{(i)}] + b[\lambda_{\mathrm{D}}^{(i)}]\right)\right) \\
& + \textstyle\sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \left(\psi\left(b[\lambda_{\mathrm{D}}^{(i)}]\right) - \psi\left(a[\lambda_{\mathrm{D}}^{(i)}] + b[\lambda_{\mathrm{D}}^{(i)}]\right)\right) \\
& - \textstyle\sum_{i \in \mathcal{C}_t} s_t^{(i)} \left(\psi\left(a[\lambda_{\mathrm{F}}^{(i)}]\right) - \psi\left(a[\lambda_{\mathrm{F}}^{(i)}] + b[\lambda_{\mathrm{F}}^{(i)}]\right)\right) \\
& - \textstyle\sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \left(\psi\left(b[\lambda_{\mathrm{F}}^{(i)}]\right) - \psi\left(a[\lambda_{\mathrm{F}}^{(i)}] + b[\lambda_{\mathrm{F}}^{(i)}]\right)\right)
\end{aligned}
$$

            $r_t \leftarrow 1/(1 + \exp(-\eta_t))$
        end for
        $a[\rho] \leftarrow |\mathcal{T}_2| + \sum_{t \in \mathcal{T}_3} r_t + 1$
        $b[\rho] \leftarrow |\mathcal{T}_1| + \sum_{t \in \mathcal{T}_3} (1 - r_t) + 1$
        for $i = 1, 2, \ldots, N$ **do**
            $a[\lambda_{\mathrm{D}}^{(i)}] \leftarrow \sum_{t \in \mathcal{T}_2^{(i)}} s_t^{(i)} + \sum_{t \in \mathcal{T}_3^{(i)}} r_t s_t^{(i)} + 1$
            $b[\lambda_{\mathrm{D}}^{(i)}] \leftarrow \sum_{t \in \mathcal{T}_2^{(i)}} (1 - s_t^{(i)}) + \sum_{t \in \mathcal{T}_3^{(i)}} r_t (1 - s_t^{(i)}) + 1$
            $a[\lambda_{\mathrm{F}}^{(i)}] \leftarrow \sum_{t \in \mathcal{T}_1^{(i)}} s_t^{(i)} + \sum_{t \in \mathcal{T}_3^{(i)}} (1 - r_t) s_t^{(i)} + 1$
            $b[\lambda_{\mathrm{F}}^{(i)}] \leftarrow \sum_{t \in \mathcal{T}_1^{(i)}} (1 - s_t^{(i)}) + \sum_{t \in \mathcal{T}_3^{(i)}} (1 - r_t)(1 - s_t^{(i)}) + 1$
        end for
    end while
    return $(a[\cdot], b[\cdot])$
end function

Similar to the coalition formation based on the ε-first strategy, it is necessary to find $\mathcal{C}_{\text{opt}}$ that maximizes $u(\mathcal{C}_{\text{opt}})$ in the coalition formation based on the multi-armed bandit problem. We can use the brute-force search shown in Algorithm 3 if N is small. Otherwise the heuristics shown in Algorithm 4 can be utilized.

5. Numerical Examples

In this section, we evaluate the two approaches described in Section 4 by simulation experiments. In these experiments, SUs form only one coalition in total and perform CSS according to the policy for each time step in $\mathcal{T}$. After the formation of the coalition $\mathcal{C}_t$, CSS and a communication utilizing the channel are performed. The SUs aim to maximize the utility defined in (3) and to utilize the channel as many times as possible while reducing collisions with the PU in time steps $\mathcal{T}$.

We conduct simulation experiments under three scenarios. In the first scenario, the number of all the SUs is small and the performance difference of the miss-detection probability and the false-alarm one among the SUs is significant. In the second scenario, the number of all the SUs is small and the performance difference among the SUs is small compared with the first one. In the third scenario, the number of all the SUs is large compared with the first one. In each experiment, we performed 70 simulation runs by changing the seed value of the random number generator. The evaluation measures defined below are the average values of the 70 runs.

The performance measures of the cumulative reward, the detection probability and the false-alarm one of the coalition, the expected cumulative occurrences of miss-detections and false-alarms, and the expected regret are used to evaluate the two approaches. The cumulative reward is defined in (3) and is calculated from the number of successful communications and the number of collisions with the PU until the t-th time step. The greater the cumulative reward is, the better the policy is. The detection probability and the false-alarm one of the coalition at the t-th time step are defined in (1) and (2), respectively. The greater the detection probability is, the better the policy is and the less the false-alarm probability is, the better the policy is. The expected cumulative occurrences of miss-detections and false-alarms until the t-th time step are calculated using the detection probability and the false-alarm probability of the coalition until the t-th time step, respectively. The less the values are, the better the policy is. The expected regret is one of the evaluation measures of the multi-armed bandit problem and defined as

$$\mathbb{E}[\mathrm{regret}(t)] = \mathbb{E}\left[\sum_{t'=1}^{t} X_{i^*}(t') - \sum_{t'=1}^{t} X_{\pi(t')}(t')\right],$$

where $X_{i^*}(t')$ is the reward obtained by pulling the optimal arm at the t'-th step and $X_{\pi(t')}(t')$ is the reward obtained by pulling the arm selected by the policy at the t'-th time step. In other words, the expected regret represents the difference between the cumulative reward obtained by pulling the optimum arm for t times and the cumulative reward obtained by pulling the arm selected by the policy until the t-th time step.

5.1 The Small Number of The SUs and The Large Performance Difference

In this section, we consider the scenario where the number of all the SUs is small and the performance difference among the SUs is significant. Table 1 shows the configuration of the parameters. In this configuration, the detection probability and the false-alarm one of the SUs $\{1, 2, 3\}$ are superior to those of the SUs $\{1, 4, 5\}$. Therefore, the optimal coalition which maximizes the expected utility in $\mathcal{T}$ is $\{1, 2, 3\}$. The coalition of the SUs $\{1, 2, 3\}$ gives $\Lambda_{\mathrm{D}}(\{1, 2, 3\}) = 0.969$ and $\Lambda_{\mathrm{F}}(\{1, 2, 3\}) = 0.0310$. Besides, the coalition of the SUs $\{1, 4, 5\}$ gives $\Lambda_{\mathrm{D}}(\{1, 4, 5\}) = 0.784$ and $\Lambda_{\mathrm{F}}(\{1, 4, 5\}) = 0.216$. Since N is small, the brute-force search shown in Algorithm 3 is used as $\textsc{FindOptimal}(\cdot)$, and we set T' equal to 10, 20, and 30. Since the probability of occurrence of collision with the PU, represented by $\rho(1 - \Lambda_{\mathrm{D}}(\cdot))$, is smaller than the probability of successful communication, represented by $(1 - \rho)(1 - \Lambda_{\mathrm{F}}(\cdot))$, we set α to a small value (0.2) to punish the collision with PU.

Figure 5 shows how the cumulative rewards for two policies change with time. This figure shows that the policy based on the multi-armed bandit problem can achieve more cumulative reward than the policy based on the ε-first strategy. In the policy based on the ε-first policy, the increasing speed of the cumulative reward changes at the T'-th time step. This is because the coalition $\mathcal{C}_{\mathrm{opt}}$ that maximizes $u(\mathcal{C}_{\mathrm{opt}})$ is estimated and formed immediately after T' from the data obtained until T'. By contrast, in the policy based on the multi-armed bandit

Table 1: The configuration for Experiment 1. This configuration assumes that the number of all the SUs is small and that the performance difference among the SUs is significant.

Symbol	Value	Description
N	5	The number of all the SUs
M	3	The number of the SUs in the coalition to formed
k	2	The predefined parameter for the k-out-of-N rule
ρ	0.5	The channel utilization by the PU
$\lambda_{\mathrm{D}}^{(i)}$	$(0.7, 0.95, 0.95, 0.7, 0.7)$	The detection probability of the SUs
$\lambda_{\mathrm{F}}^{(i)}$	$(0.3, 0.05, 0.05, 0.3, 0.3)$	The false-alarm probability of the SUs
α	0.2	The parameter for adjusting the balance between N_{S} and N_{MD}

problem, increasing speed of the cumulative reward grows the fastest among the four curves. This trend suggests that the policy based on the multi-armed bandit problem learns a coalition $\mathcal{C}_t$ which has a high $u(\mathcal{C}_t)$ as the time step increases.

Figure 6 shows how the detection probability of $\mathcal{C}_t$ changes with time. In this figure, for the policy based on the multi-armed bandit problem, the detection probability of $\mathcal{C}_t$ grows as the time step increases. By contrast, for the policy based on the ε-first strategy, the detection probability of $\mathcal{C}_t$ is low until T', and then increases after T'. This is because the coalition $\mathcal{C}_{\mathrm{opt}}$ that maximizes $u(\mathcal{C}_{\mathrm{opt}})$ is estimated and formed immediately after T' from the data obtained until T'. Moreover, the reason of the high detection probability for large T' is that many observation data can be obtained when T' is large. Figure 7 shows how the false-alarm probability of $\mathcal{C}_t$ changes with time. In this figure, for the policy based on the multi-armed bandit problem, the false-alarm probability of $\mathcal{C}_t$ declines as the time step increases. On the contrary, for the policy based on the ε-first strategy, the false-alarm probability of $\mathcal{C}_t$ is high until T', and then decreases after T'. It can be explained in the same manner as the detection probability.

In both Figures 6 and 7, we observe that the ε-first strategy provides better

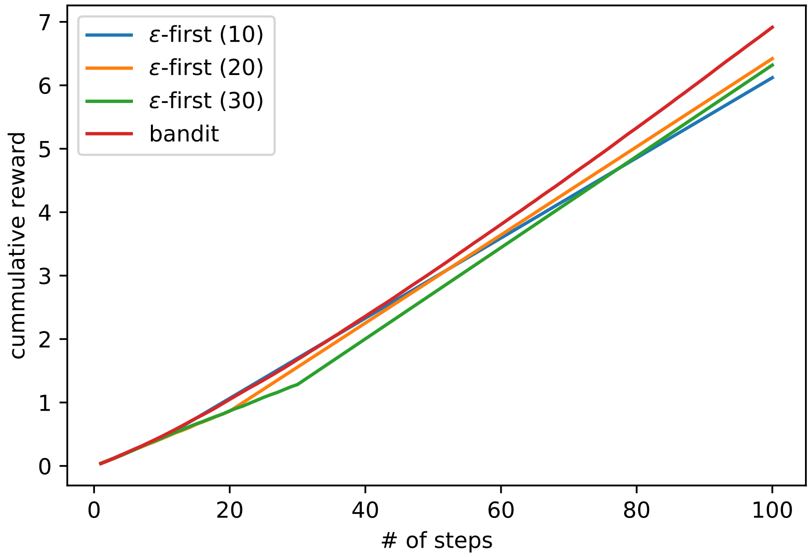

Figure 5: The cumulative rewards against the time step in Experiment 1.

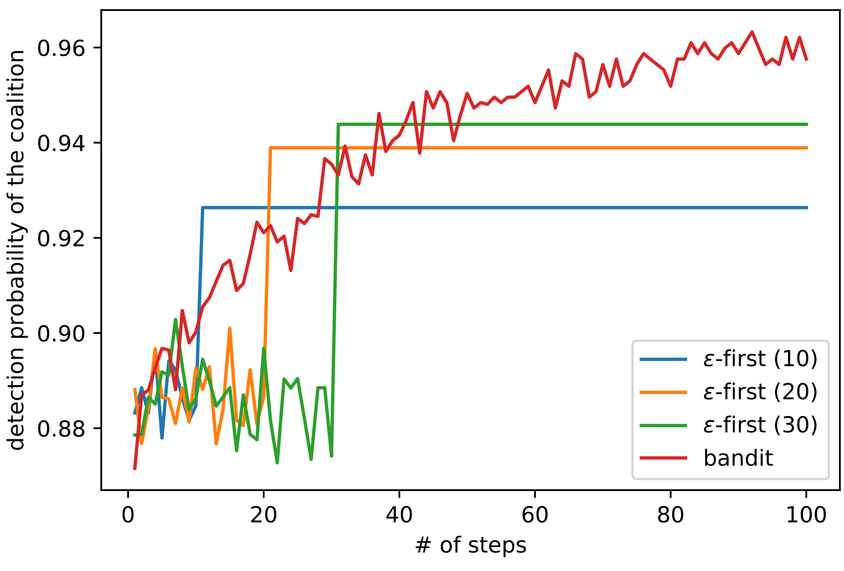

Figure 6: The detection probability of the coalition against the time step in Experiment 1.

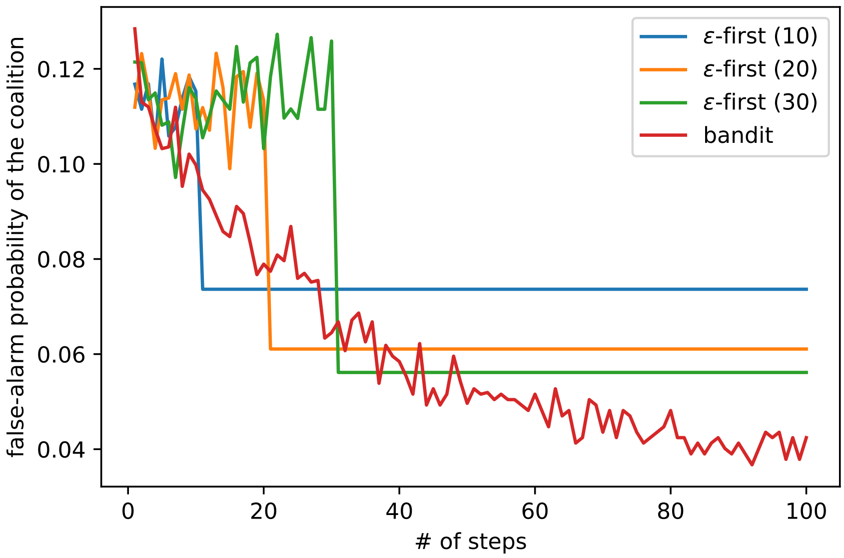

Figure 7: The false-alarm probability of the coalition against the time step in Experiment 1.

performance than the policy based on the multi-armed bandit problem at the T'-th time step. This is because the policy based on the ε-first strategy ignores miss-detections and false-alarms occurring in the exploration steps. By contrast, the policy based on the multi-armed bandit problem aims to maximize $U(\mathcal{T})$ while considering exploitation-exploration trade-offs. Figures 8 and 9 show the expected cumulative occurrences of miss-detections and false-alarms, respectively. These figures show that the policy based on the multi-armed bandit problem is superior to the policy based on the ε-first strategy in terms of miss-detections and false-alarms.

Figure 10 shows the expected regret against the time step. In this figure, the regret of the policy based on the multi-armed bandit problem increases logarithmically after 100 steps. We can consider that the number of times a coalition except for the optimal coalition is formed declines as the time step increases.

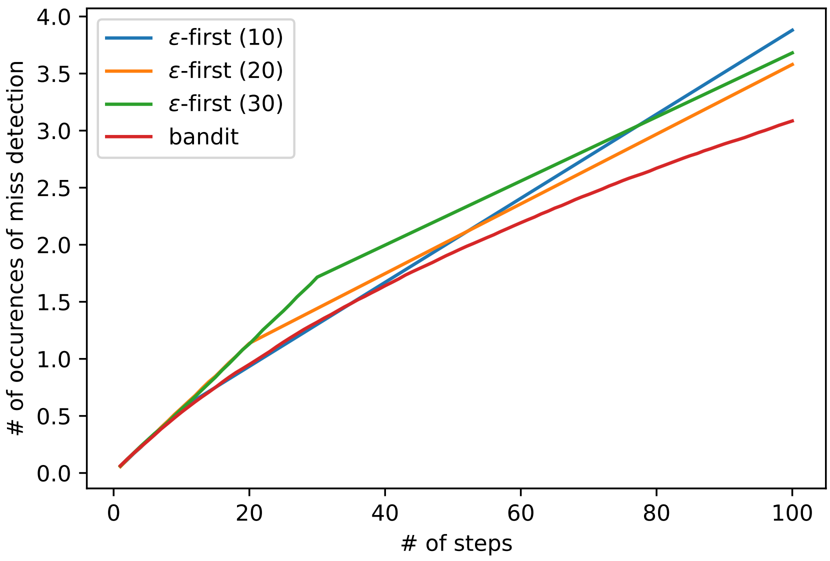

Figure 8: The expected cumulative occurrences of miss-detections against the time step in Experiment 1.

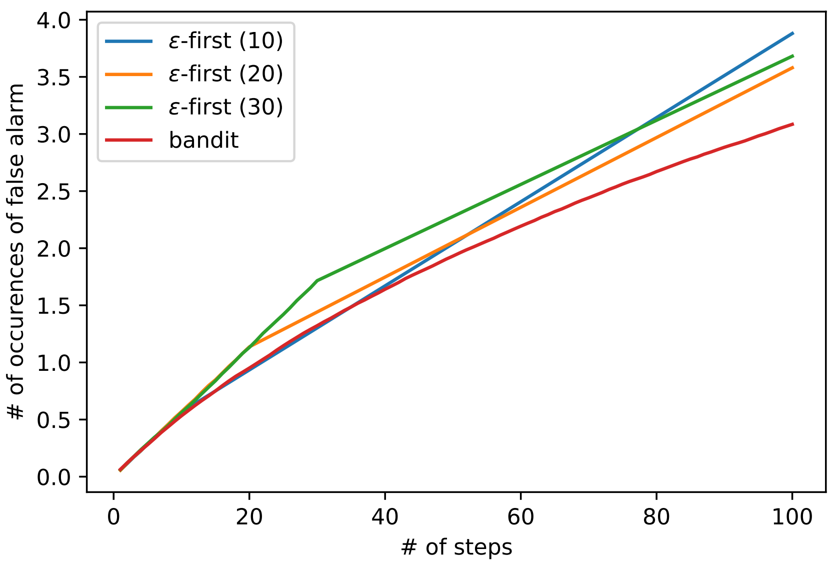

Figure 9: The expected cumulative occurrences of false-alarms against the time step in Experiment 1.

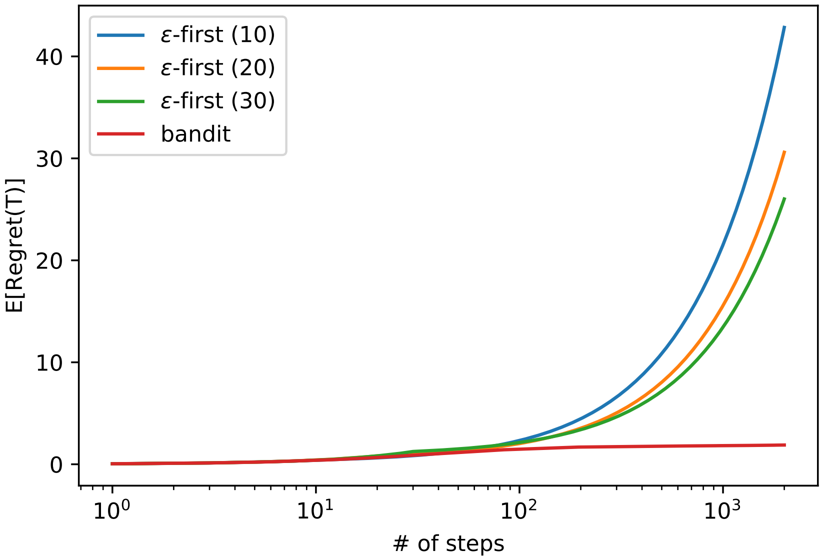

Figure 10: $\mathbb{E}[\text{Regret}(t)]$ against the time step in Experiment 1.

Table 2: The configuration for Experiment 2. This configuration assumes that the number of the SUs is small and that the performance difference among the SUs is smaller than that of Experiment 1.

Symbol	Value	Description
N	5	The number of all the SUs
M	3	The number of the SUs in the coalition to formed
k	2	The predefined parameter for the k-out-of-N rule
ρ	0.5	The channel utilization by the PU
$\lambda_{\mathrm{D}}^{(i)}$	$(0.9, 0.95, 0.95, 0.9, 0.9)$	The detection probability of the SUs
$\lambda_{\mathrm{F}}^{(i)}$	$(0.1, 0.05, 0.05, 0.1, 0.1)$	The false-alarm probability of the SUs
α	0.2	The parameter for adjusting the balance between N_{S} and N_{MD}

5.2 The Small Number of The SUs and The Small Performance Difference

In this subsection, we consider the scenario where the number of the SUs is small and the performance difference among the SUs is smaller than that of Experiment 1. Table 2 shows the configuration of the parameters. In this configuration, the detection probability and the false-alarm one of the SUs $\{2, 3\}$ are superior to those of the SUs of $\{1, 4, 5\}$. Therefore, the optimal coalition which maximizes the expected utility in $\mathcal{T}$ is $\{1, 2, 3\}$. On the one hand, the coalition of the SUs $\{1, 2, 3\}$ gives $\Lambda_{\mathrm{D}}(\{1, 2, 3\}) = 0.988$ and $\Lambda_{\mathrm{F}}(\{1, 2, 3\}) = 0.0120$. On the other hand, the coalition of the SUs $\{1, 4, 5\}$ gives $\Lambda_{\mathrm{D}}(\{1, 4, 5\}) = 0.972$ and $\Lambda_{\mathrm{F}}(\{1, 4, 5\}) = 0.028$. Since N is small, the brute-force search shown in Algorithm 3 is used as $\textsc{FindOptimal}(\cdot)$. T' is set to 100, 200, and 300. The reason why the values of T' is larger than those in Experiment 1 is that many observed data are required for estimation due to small performance difference among the SUs.

Figure 11 shows how the cumulative rewards for two policies change with time.

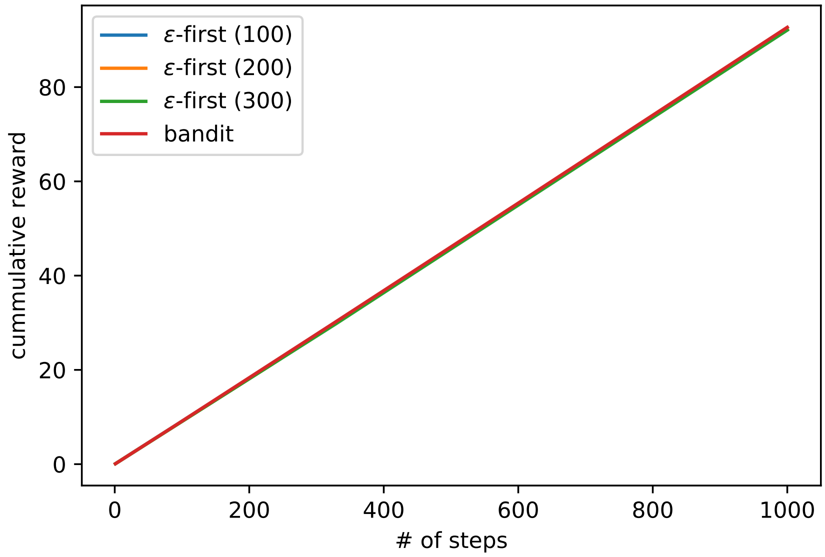

Figure 11: The cumulative rewards against the time step in Experiment 2.

We observe no difference between the two approaches. This trend is different from Figure 5. The reason of this is that the performance difference among the SUs is smaller than that of Experiment 1 and that the SU selection does not improve the detection and false-alarm probabilities for the resulting coalition.

Figures 12 and 13 show the detection probability and the false-alarm one of $\mathcal{C}_t$ against the time step, respectively. These figures show that both the policies improve the detection probability and false-alarm one of the coalition over time. However, the fact that the detection probability changes from 0.980 to 0.986 and the false-alarm probability changes from 0.020 to 0.014 shows the improvement is small. This is simply due to a small performance difference among the SUs.

Figures 14 and 15 show the expected cumulative occurrences of miss-detections and false-alarms, respectively. We observe no difference among the two approaches in terms of the expected rewards. On the other hand, these figures show that the policy based on the multi-armed bandit problem is superior to the policy based on the ε-first strategy in terms of miss-detections and false-alarms. Although there is only a small effect, we can claim that the policy based on the multi-armed bandit problem succeeds in reducing miss-detections and false-

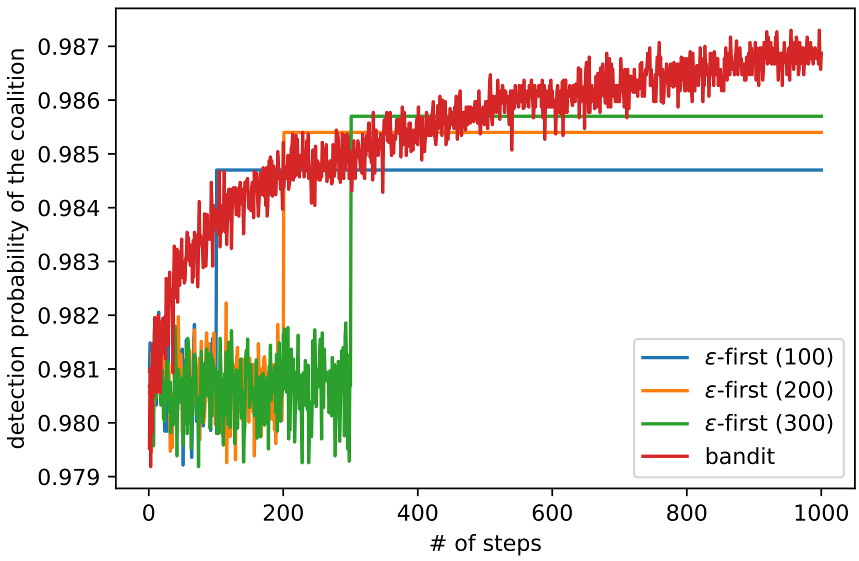

Figure 12: The detection probability of the coalition against the time step in Experiment 2.

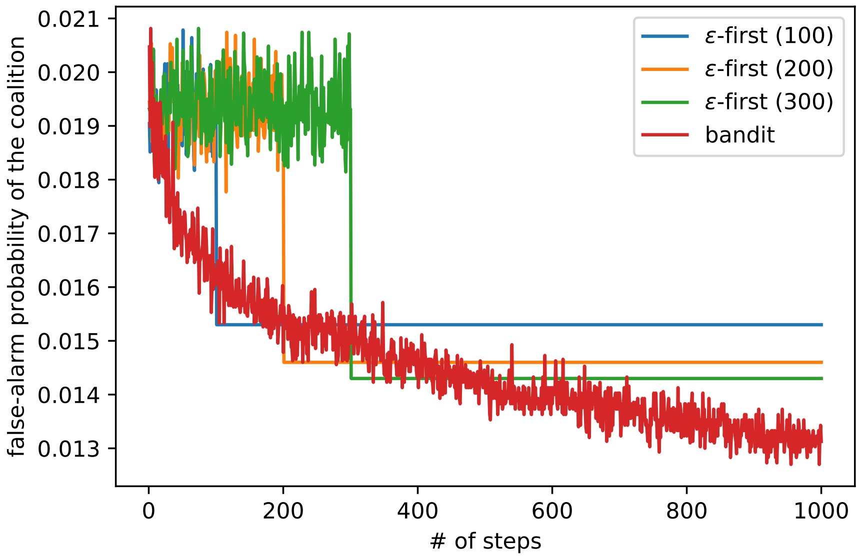

Figure 13: The false-alarm probability of the coalition against the time step in Experiment 2.

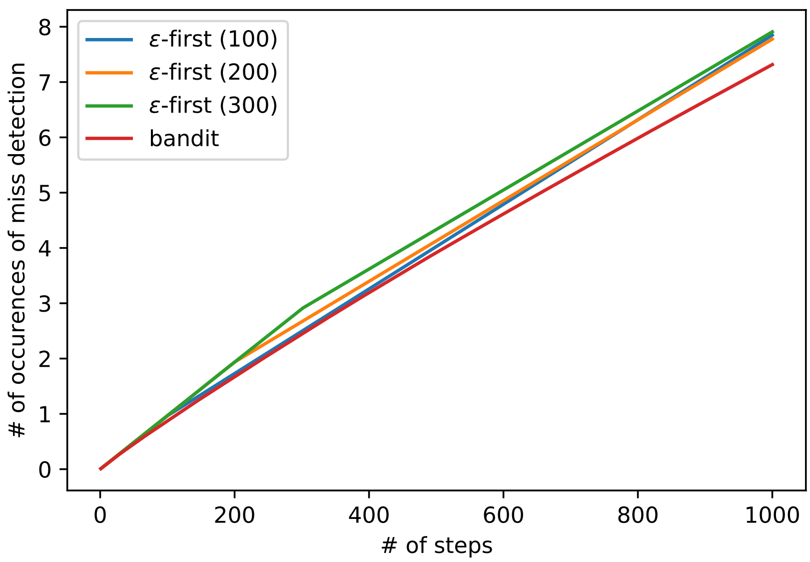

Figure 14: The expected cumulative occurrences of miss-detections against the time step in Experiment 2.

alarms since it considers exploitation-exploration trade-offs.

Figure 16 shows the expected regret against the time step. In this figure, the regret of the policy based on the multi-armed bandit problem increases logarithmically after 300 steps. The number of times a coalition except for the optimal coalition is formed declines as the time step increases.

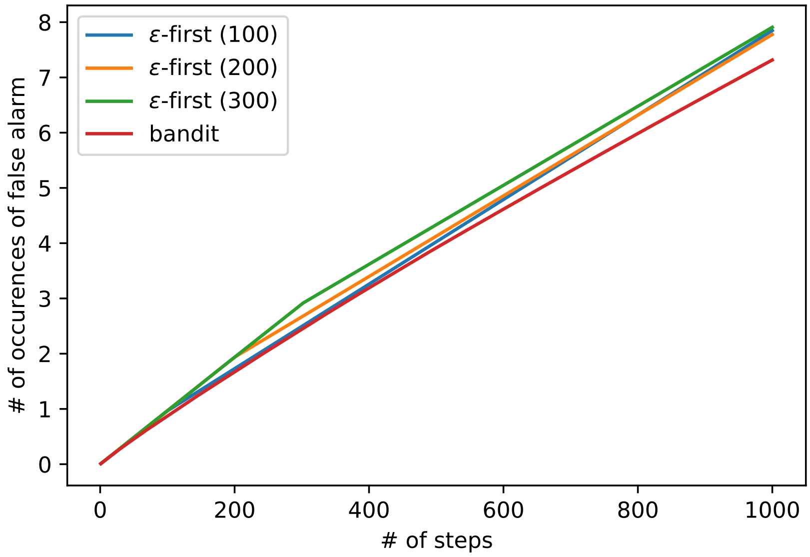

Figure 15: The expected cumulative occurrences of false-alarms against the time step in Experiment 2.

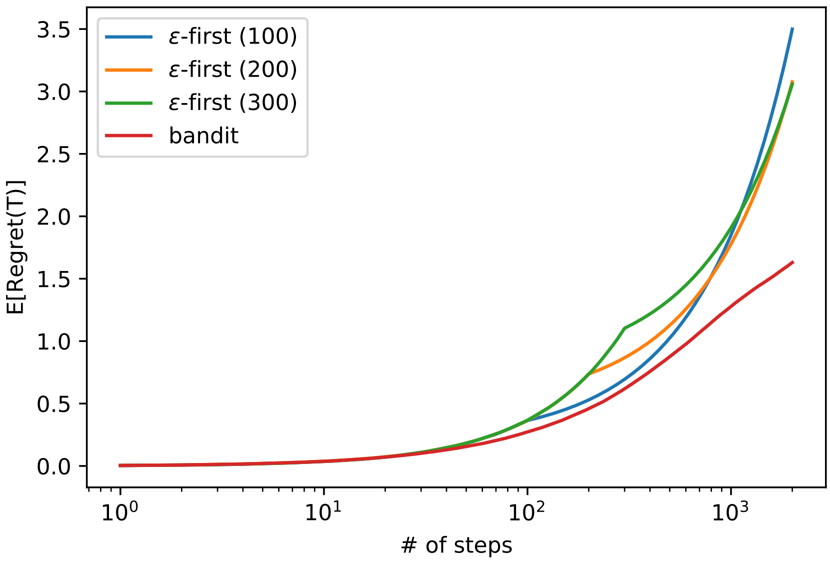

Figure 16: Regret(t) against the time step in Experiment 2.

Table 3: The configuration for Experiment 3. This configuration assumes that the number of all the SUs is larger than that of Experiment 1 and Experiment 2.

Symbol	Value	Description
N	50	The number of all the SUs
M	5	The number of the SUs in the coalition to formed
k	3	The predefined parameter for the k-out-of-N rule
ρ	0.5	The channel utilization by the PU
$\lambda_{\mathrm{D}}^{(i)}$	$(0.7, 0.95, 0.95, 0.95, 0.95, 0.7 \times 45)$	The detection probability of the SUs
$\lambda_{\mathrm{F}}^{(i)}$	$(0.3, 0.05, 0.05, 0.05, 0.05, 0.3 \times 45)$	The false-alarm probability of the SUs
α	0.2	The parameter for adjusting the balance between N_{S} and N_{MD}

5.3 The Large Number of The SUs

In this subsection, we consider the scenario where the number of SUs is larger than those of the first and the second numerical examples. Table 3 shows the configuration of parameters. In this configuration, the detection probability and the false-alarm one of the SUs $\{2, 3, 4, 5\}$ are superior to those of the SUs of $\{1, 6, 7, \ldots, 50\}$. Therefore, the optimal coalition which maximizes the expected utility in $\mathcal{T}$ is $\{1, 2, 3, 4, 5\}$. The coalition of the SUs $\{1, 2, 3, 4, 5\}$ gives $\Lambda_{\mathrm{D}}(\{1, 2, 3, 4, 5\}) = 0.995$ and $\Lambda_{\mathrm{F}}(\{1, 2, 3, 4, 5\}) = 0.00454$, while the coalition of the SUs $\{1, 6, 7, 8, 9\}$ gives $\Lambda_{\mathrm{D}}(\{1, 6, 7, 8, 9\}) = 0.837$ and $\Lambda_{\mathrm{F}}(\{1, 6, 7, 8, 9\}) = 0.163$. Since N is large, the heuristics shown in Algorithm 4 is used as FindOptimal($\cdot$). In this experiment, we set T' equal to 200, 300, and 400. The reason why the values of T' is larger than those in Experiment 1 is that the number of SUs is greater than that of Experiment 1.

Figure 17 shows how the cumulative rewards for two policies change with

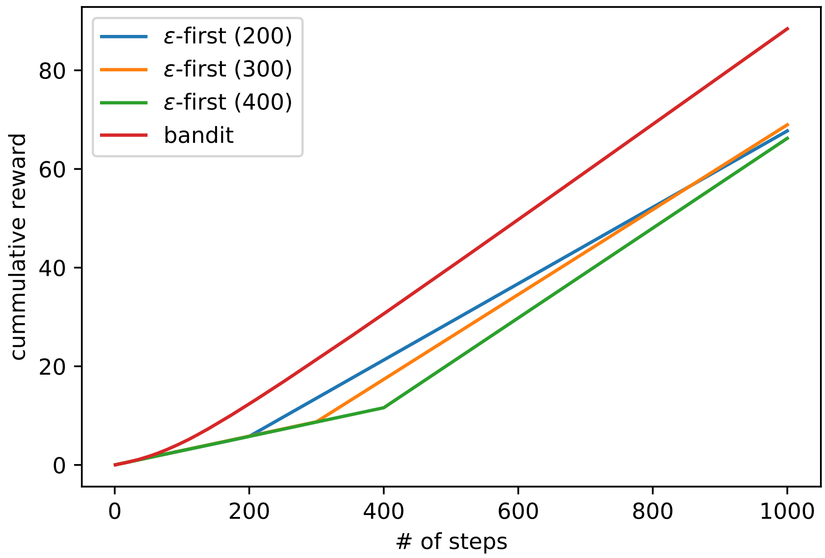

Figure 17: The cumulative rewards against the time step in Experiment 3.

time. Compared with Experiments 1 and 2, there is a large difference between the two policies. The reason is that the number of SUs is larger than that of Experiments 1 and 2. Although it is not necessary to accurately estimate the sensing performance of all the SUs as in the second disadvantage described in Section 4.2, the policy based on the ε-first strategy tries to estimate them. Hence, it requires long exploration steps and many miss-detections and false-alarms occur due to coalition formation with low $u(\mathcal{C}_t)$.

Figures 18 and 19 show the detection probability and the false-alarm one of $\mathcal{C}_t$ against time step, respectively. These figures show that both policies improve the detection probability and false-alarm one of the coalition over time. Unlike Experiment 1, the detection probability and the false-alarm one of $\mathcal{C}_t$ at the T'-th time step by the policy based on the multi-armed bandit problem are superior to those by the policy based on the ε-greedy strategy. This is because the policy based on the ε-greedy strategy forms a coalition consisting of randomly selected SUs in the exploration steps. By contrast, the policy based on the multi-armed bandit problem forms the coalition consisting of SUs which are expected to maximize $u(\mathcal{C}_t)$.

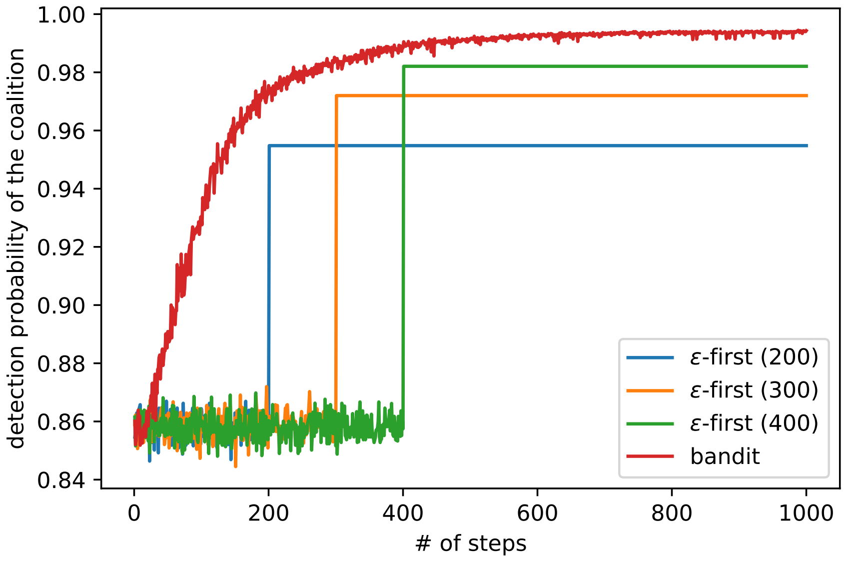

Figure 18: The detection probability of the coalition against the time step in Experiment 3.

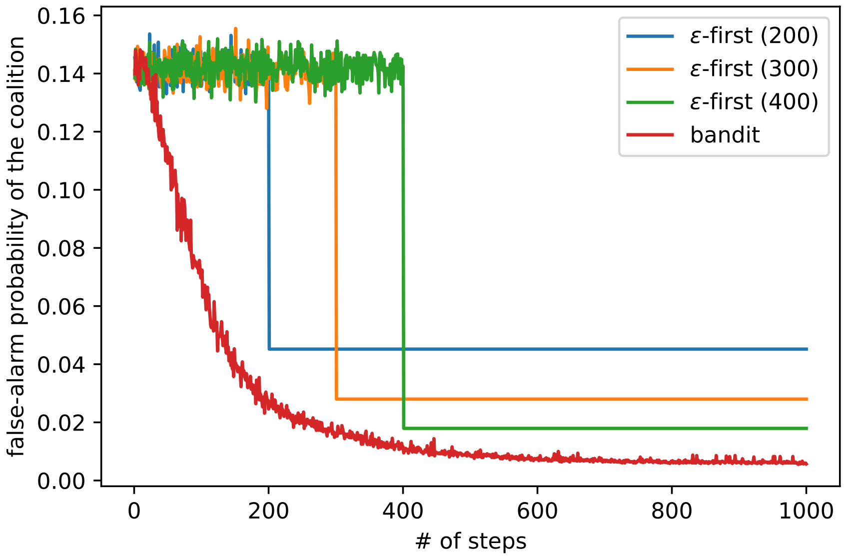

Figure 19: The false-alarm probability of the coalition against the time step in Experiment 3.

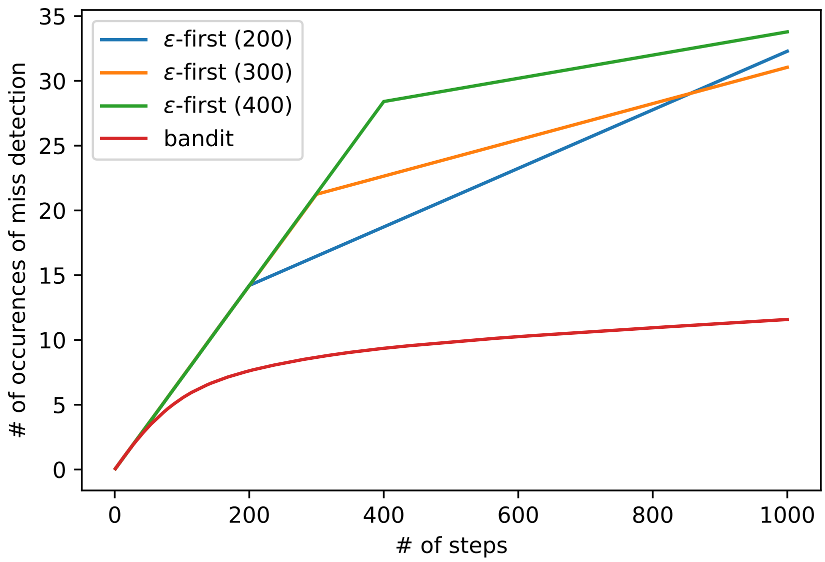

Figure 20: The expected cumulative occurrences of miss-detections against the time step in Experiment 3.

Figures 20 and 21 show the expected cumulative occurrences of miss-detections and false-alarms, respectively. These figures show that the policy based on the multi-armed bandit problem reduces the expected cumulative occurrences of miss-detections and false-alarms to one third as compared with the policy based on the ε-first strategy.

Figure 22 shows the expected regret against time step. In this figure, the regret of the policy based on the multi-armed bandit problem increases logarithmically after 400 steps. We can claim that the policy based on the multi-armed bandit problem works effectively even when the number of SUs is large.

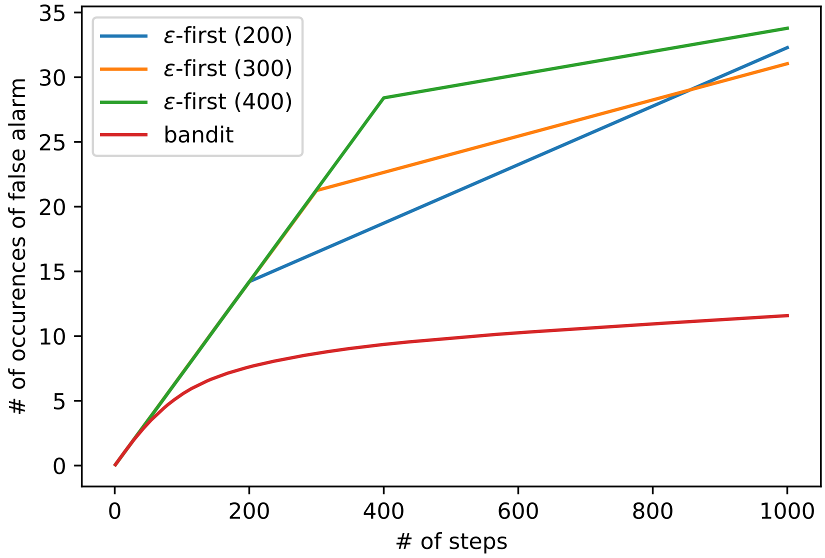

Figure 21: The expected cumulative occurrences of false-alarms against the time step in Experiment 3.

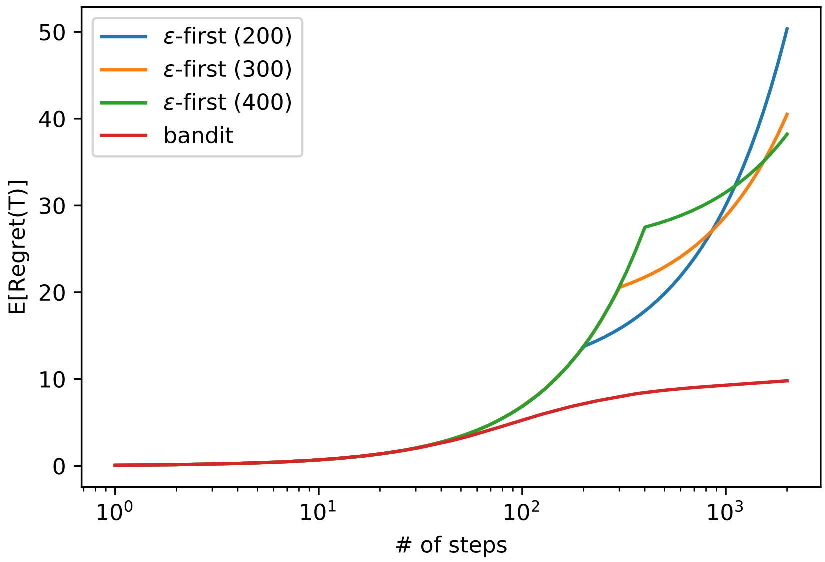

Figure 22: Regret(t) against the time step in Experiment 3.

6. Conclusion

In this thesis, we considered the decision making about how to perform coalition formation under uncertain environment in cognitive radio networks. In our model, each SU does not know not only the channel utilization by a PU but also the detection probability and the false-alarm one of all SUs. The SUs aim to utilize the channel as many times as possible while reducing collisions with the PU.

We proposed two approaches for coalition formation. The first approach is based on the ε-first strategy, which is a baseline. The second one is based on the multi-armed bandit problem. In order to adapt to an uncertain environment, both policies collect data and perform coalition formation and the estimation of unknown information at the same time.

In order to compare the performance of the two policies, we conducted numerical experiments, and calculated the performance measures of the cumulative reward, the detection probability, the false-alarm probability, the expected cumulative occurrences of miss-detections and false-alarms, and the expected regret. In the scenario where the number of all the SUs is small and the performance difference among the SUs is significant, we observed that the policy based on the multi-armed bandit problem succeeds in reducing the cumulative occurrences of miss-detections and false-alarms since it considers exploitation-exploration trade-offs. In the scenario where the number of all the SUs is small and the performance difference among the SUs is small compared with the first one, we observed no difference between the two approaches in terms of the cumulative rewards since the SU selection does not improve the detection and false-alarm probabilities for the resulting coalition. In the scenario where the number of all the SUs is large compared with the first one, we observed that the policy based on the multi-armed bandit problem works effectively even when the number of SUs is large. Numerical examples show that the proposed method can decrease miss-detections and false-alarms especially in a scenario where the number of SUs is large and the performance difference among SUs is significant.

Future work is to extend the model so that SUs can form multiple coalitions. In our model, SU 1 leads coalition formation and all other SUs passively cooperate. In order to allow SUs to form multiple coalitions, issues on the selection of leaders to lead coalition formation, information sharing about the sensing performance of SU inferred, and assignment of SUs to improve overall throughput should be considered.

References

[1] J. Mitola and G. Maguire, "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[3] A. Ghasemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *Proc. of 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, 2005, pp. 131–136.

[4] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical Communication*, vol. 4, no. 1, pp. 40–62, 2011.

[5] W. Saad, Z. Han, T. Basar, M. Debbah, and A. Hjorungnes, "Coalition formation games for collaborative spectrum sensing," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 276–297, 2011.

[6] T. Nishida, M. Sasabe, and S. Kasahara, "Maximizing communication opportunity for collaborative spectrum sensing in cognitive radio networks," in *Proc. of 27th International Telecommunication Networks and Applications Conference (ITNAC)*, 2017, pp. 1–6.

[7] B. F. Lo and I. F. Akyildiz, "Reinforcement learning for cooperative sensing gain in cognitive radio ad hoc networks," *Wireless Networks*, vol. 19, no. 6, pp. 1237–1250, 2013.

[8] F. H. Panahi and T. Ohtsuki, "Optimal channel-sensing scheme for cognitive radio systems based on fuzzy Q-learning," *IEICE Transactions on Communications*, vol. E97.B, no. 2, pp. 283–294, 2014.

[9] C. Sun, W. Zhang, and K. Ben Letaief, "Cooperative spectrum sensing for cognitive radios under bandwidth constraints," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, 2007, pp. 1–5.

[10] E. C. Y. Peh, Y.-C. Liang, Y. L. Guan, and Y. Zeng, "Optimization of cooperative sensing in cognitive radio networks: A sensing-throughput tradeoff view," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 9, pp. 5294–5299, 2009.

[11] S. Althunibat and F. Granelli, "Novel energy-efficient reporting scheme for spectrum sensing results in cognitive radio," in *Proc. of IEEE International Conference on Communications (ICC)*, 2013, pp. 2438–2442.

[12] S. Althunibat, M. Di Renzo, and F. Granelli, "Optimizing the K-out-of-N rule for cooperative spectrum sensing in cognitive radio networks," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, 2013, pp. 1607–1611.

[13] J. Vermorel and M. Mohri, "Multi-armed bandit algorithms and empirical evaluation," in *Proc. of 16th European Conference on Machine Learning (ECML)*, 2005, pp. 437–448.

[14] S. J. D. Prince, *Computer vision: models, learning, and inference.* Cambridge University Press, 2012.

[15] H. Robbins, "Some aspects of the sequential design of experiments," *Bulletin of the American Mathematical Society*, vol. 58, no. 5, pp. 527–536, 1952.

[16] D. A. Berry, "A Bernoulli two-armed bandit," *The Annals of Mathematical Statistics*, vol. 43, no. 3, pp. 871–897, 1972.

[17] P. Auer, N. Cesa-bianchi, and P. Fischer, "Finite time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, no. 2–3, pp. 235–256, 2002.

[18] O. Chapelle and L. Li, "An empirical evaluation of Thompson sampling," in *Advances in Neural Information Processing Systems 24 (NIPS)*, 2011, pp. 2249–2257.

[19] S. Agrawal and N. Goyal, "Analysis of Thompson sampling for the multi-armed bandit problem," in *Proc. of 25th Annual Conference on Learning Theory (COLT)*, vol. 23, 2012, pp. 39.1–39.26.

[20] N. Korda, E. Kaufmann, and R. Munos, "Thompson sampling for 1-dimensional exponential family bandits," in *Advances in Neural Information Processing Systems 26 (NIPS)*, 2013, pp. 1448–1456.

[21] J. Honda and A. Takemura, "Optimality of Thompson sampling for Gaussian bandits depends on priors," in *Proc. of 17th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2014, pp. 375–383.

[22] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Dover, 1972.

[23] W. Chen, Y. Wang, and Y. Yuan, "Combinatorial multi-armed bandit: General framework and applications," in *Proc. of 30th International Conference on Machine Learning (ICML)*, 2013, pp. 151–159.

[24] A. Gopalan, S. Mannor, and Y. Mansour, "Thompson sampling for complex online problems," in *Proc. of 31st International Conference on Machine Learning (ICML)*, 2014, pp. 100–108.

[25] J. Komiyama, J. Honda, and H. Nakagawa, "Optimal regret analysis of Thompson sampling in stochastic multi-armed bandit problem with multiple plays," in *Proc. of 32nd International Conference on Machine Learning (ICML)*, 2015, pp. 1152–1161.

[26] K. P. Murphy, *Machine learning: a probabilistic perspective.* MIT Press, 2012.

[27] M. J. Beal and Z. Ghahramani, "The variational bayesian EM algorithm for incomplete data: with application to scoring graphical model structures," *Bayesian Statistics*, vol. 7, pp. 453–464, 2003.

Appendix

A. Maximum Likelihood Estimation of ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ by the EM Algorithm

In this appendix, we summarize the maximum likelihood estimation by the EM algorithm in order to estimate ρ (the channel utilization by the PU), $\lambda_{\mathrm{D}}^{(i)}$ (the detection probability of each SU) and $\lambda_{\mathrm{F}}^{(i)}$ (the false-alarm probability of each SU). In the following, $(\rho, \lambda_{\mathrm{D}}^{(1)}, \ldots, \lambda_{\mathrm{D}}^{(N)}, \lambda_{\mathrm{F}}^{(1)}, \ldots, \lambda_{\mathrm{F}}^{(N)})$ is denoted by $\boldsymbol{\theta}$.

The observed data that can be used for estimation are $\mathcal{C}_t$, $s_t^{(i)}$ and y_t. $\mathcal{C}_t$ is the coalition formed at the t-th time step, and $s_t^{(i)}$ is the sensing result of the i-th SU at the t-th time step. When the SU regards the channel as *busy*, $s_t^{(i)} = 1$. When the SU regards the channel as *idle*, $s_t^{(i)} = 0$. y_t is the communication result of the coalition at the t-th time step. If the coalition tries to utilize the channel and the communication is successful, $y_t = 1$. If the coalition tries to utilize the channel and a collision occurs, $y_t = 2$. If the coalition does not try to utilize the channel, $y_t = 3$. In the following, $(\mathcal{C}_t, s_t^{(i)}, y_t)$ at the t-th time step is denoted by $\boldsymbol{x}_t$ and $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ is denoted by $\boldsymbol{X}$. In addition, $\{t \mid t \in \mathcal{T} \wedge y_t = j\}$ is expressed as $\mathcal{T}_j$ for $j = 1, 2, 3$ and $\{t \mid t \in \mathcal{T}_j \wedge i \in \mathcal{C}_t\}$ is expressed as $\mathcal{T}_j^{(i)}$.

We introduce a hidden variable z_t as the actual utilization of the channel by the PU at the t-th time step. This is because when $y_t = 3$, i.e., the coalition does not communicate at the t-th time step, there are two possible reasons. The first one is that the coalition successfully detects that the channel is used by the PU. The second possible reason is that a false-alarm occurs when the channel is actually idle. Therefore, if the channel is actually *busy* at the t-th time step, $z_t = 1$; otherwise $z_t = 0$. In the following, $(z_1, z_2, \ldots, z_T)$ is denoted by $\boldsymbol{Z}$.

Let $\ell(\boldsymbol{X}, \boldsymbol{z} \mid \boldsymbol{\theta})$ be the log likelihood for complete data. Since z_t is observed if $y_t = 1$ or $y_t = 2$, it is given by

$$\begin{aligned}\ell(\boldsymbol{X}, \boldsymbol{z} \mid \boldsymbol{\theta}) &= \sum_{t \in \mathcal{T}} \log p(\boldsymbol{x}_t, z_t \mid \boldsymbol{\theta}) \\ &= \sum_{t \in \mathcal{T}_1} \log p(\boldsymbol{x}_t, z_t = 0 \mid \boldsymbol{\theta}) \\ &\quad + \sum_{t \in \mathcal{T}_2} \log p(\boldsymbol{x}_t, z_t = 1 \mid \boldsymbol{\theta}) \\ &\quad + \sum_{t \in \mathcal{T}_3} \log p(\boldsymbol{x}_t, z_t \mid \boldsymbol{\theta}).\end{aligned}$$

The first and second terms in the right hand side of the above equation are yielded as

$$\begin{aligned}&\sum_{t \in \mathcal{T}_1} \log p(\boldsymbol{x}_t, z_t = 0 \mid \boldsymbol{\theta}) \\ &= \sum_{t \in \mathcal{T}_1} \log \left[(1-\rho) \prod_{i \in \mathcal{C}_t} \left(\lambda_{\mathrm{F}}^{(i)}\right)^{s_t^{(i)}} \left(1 - \lambda_{\mathrm{F}}^{(i)}\right)^{1-s_t^{(i)}} \right] \\ &= \sum_{t \in \mathcal{T}_1} \left[\log(1-\rho) + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{F}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{F}}^{(i)}) \right],\end{aligned} \tag{5}$$

$$\begin{aligned}&\sum_{t \in \mathcal{T}_2} \log p(\boldsymbol{x}_t, z_t = 1 \mid \boldsymbol{\theta}) \\ &= \sum_{t \in \mathcal{T}_2} \log \left[\rho \prod_{i \in \mathcal{C}_t} \left(\lambda_{\mathrm{D}}^{(i)}\right)^{s_t^{(i)}} \left(1 - \lambda_{\mathrm{D}}^{(i)}\right)^{1-s_t^{(i)}} \right] \\ &= \sum_{t \in \mathcal{T}_2} \left[\log \rho + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{D}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{D}}^{(i)}) \right].\end{aligned} \tag{6}$$

When $y_t = 3$, noting that z_t is unknown, it is necessary to consider the case of

$z_t = 0$ and $z_t = 1$, and we obtain

$$
\begin{aligned}
&\sum_{t \in \mathcal{T}_3} \log p(\boldsymbol{x}_t, z_t \mid \boldsymbol{\theta}) \\
&= \sum_{t \in \mathcal{T}_3} \log \left[p(\boldsymbol{x}_t, z_t = 0 \mid \boldsymbol{\theta})^{1-z_t} \, p(\boldsymbol{x}_t, z_t = 1 \mid \boldsymbol{\theta})^{z_t} \right] \\
&= \sum_{t \in \mathcal{T}_3} (1 - z_t) \left[\log(1-\rho) + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{F}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{F}}^{(i)}) \right] \\
&\quad + \sum_{t \in \mathcal{T}_3} z_t \left[\log \rho + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{D}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{D}}^{(i)}) \right].
\end{aligned}
$$

Hence, the log likelihood for complete data is given by

$$
\begin{aligned}
&\ell(\boldsymbol{X}, \boldsymbol{z} \mid \boldsymbol{\theta}) \\
&= \sum_{t \in \mathcal{T}_1} \left[\log(1-\rho) + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{F}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{F}}^{(i)}) \right] \\
&\quad + \sum_{t \in \mathcal{T}_2} \left[\log \rho + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{D}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{D}}^{(i)}) \right] \\
&\quad + \sum_{t \in \mathcal{T}_3} (1 - z_t) \left[\log(1-\rho) + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{F}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{F}}^{(i)}) \right] \\
&\quad + \sum_{t \in \mathcal{T}_3} z_t \left[\log \rho + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{D}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{D}}^{(i)}) \right].
\end{aligned}
$$

We use the EM algorithm [14] to find $\boldsymbol{z}$ and $\boldsymbol{\theta}$ to maximize the log likelihood. First, each element of $\boldsymbol{\theta}'$ is initialized with an arbitrary value which satisfies the constraint. In the expectation step (E step), $p(z_t \mid \boldsymbol{x}_t, \boldsymbol{\theta}')$ is calculated by

$$
p(z_t \mid \boldsymbol{x}_t, \boldsymbol{\theta}') = \frac{p(\boldsymbol{x}_t, z_t \mid \boldsymbol{\theta}')}{\sum_{k \in \{0,1\}} p(\boldsymbol{x}_t, z_t = k \mid \boldsymbol{\theta}')},
$$

where $p(\boldsymbol{x}_t, z_t \mid \boldsymbol{\theta}')$ can be calculated with (5) and (6). Let r_t be $p(z_t = 1 \mid \boldsymbol{x}_t, \boldsymbol{\theta}')$. Note that $1 - r_t = p(z_t = 0 \mid \boldsymbol{x}_t, \boldsymbol{\theta}')$. Then, the expected value of the log likelihood with respect to $\boldsymbol{z}$ is given by

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{z}}\left[\ell(\boldsymbol{X}, \boldsymbol{z} \mid \boldsymbol{\theta}) \mid \boldsymbol{X}, \boldsymbol{\theta}'\right] \\
&= \sum_{t \in \mathcal{T}_1} \left[\log(1-\rho) + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{F}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{F}}^{(i)}) \right] \\
&\quad + \sum_{t \in \mathcal{T}_2} \left[\log \rho + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{D}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{D}}^{(i)}) \right] \\
&\quad + \sum_{t \in \mathcal{T}_3} (1 - r_t) \left[\log(1-\rho) + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{F}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{F}}^{(i)}) \right] \\
&\quad + \sum_{t \in \mathcal{T}_3} r_t \left[\log \rho + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{D}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{D}}^{(i)}) \right].
\end{aligned}
$$

From here, $\mathbb{E}_{\boldsymbol{z}}\left[\ell(\boldsymbol{X}, \boldsymbol{z} \mid \boldsymbol{\theta}) \mid \boldsymbol{X}, \boldsymbol{\theta}'\right]$ is expressed as $L(\boldsymbol{\theta})$. In the maximization step (M step), $\boldsymbol{\theta}$ which maximizes $L(\boldsymbol{\theta})$ is calculated by the method of Lagrangian multiplier. First, ρ is calculated as follows.

$$
\frac{\partial}{\partial \rho} L(\boldsymbol{\theta}) = -\sum_{t \in \mathcal{T}_1} \frac{1}{1-\rho} + \sum_{t \in \mathcal{T}_2} \frac{1}{\rho} - \sum_{t \in \mathcal{T}_3} \frac{1 - r_t}{1-\rho} + \sum_{t \in \mathcal{T}_3} \frac{r_t}{\rho} = 0,
$$

$$
\rho = \frac{a[\rho]}{a[\rho] + b[\rho]},
$$

where

$$
\begin{aligned}
a[\rho] &= |\mathcal{T}_2| + \sum_{t \in \mathcal{T}_3} r_t, \\
b[\rho] &= |\mathcal{T}_1| + \sum_{t \in \mathcal{T}_3} (1 - r_t).
\end{aligned}
$$

When the number of times the PU uses the channel is n and the number of times the PU does not use the channel is m, the maximum likelihood estimator of ρ is $n/(n+m)$. This is an explanation of why we can interpret $a[\rho]$ as the expected value of the number of times the PU uses the channel and $b[\rho]$ as the expected value of the number of times the PU does not use the channel.

Next, $\lambda_{\mathrm{D}}^{(i)}$ is calculated as follows.

$$\begin{aligned}\frac{\partial}{\partial \lambda_{\mathrm{D}}^{(i)}} L(\boldsymbol{\theta}) = & \sum_{t \in \mathcal{T}_2^{(i)}} \left(s_t^{(i)} \frac{1}{\lambda_{\mathrm{D}}^{(i)}} - (1 - s_t^{(i)}) \frac{1}{1 - \lambda_{\mathrm{D}}^{(i)}} \right) \\ & + \sum_{t \in \mathcal{T}_3^{(i)}} r_t \left(s_t^{(i)} \frac{1}{\lambda_{\mathrm{D}}^{(i)}} - (1 - s_t^{(i)}) \frac{1}{1 - \lambda_{\mathrm{D}}^{(i)}} \right) \\ = & \, 0,\end{aligned}$$

$$\lambda_{\mathrm{D}}^{(i)} = \frac{a[\lambda_{\mathrm{D}}^{(i)}]}{a[\lambda_{\mathrm{D}}^{(i)}] + b[\lambda_{\mathrm{D}}^{(i)}]},$$

where

$$\begin{aligned}a[\lambda_{\mathrm{D}}^{(i)}] &= \sum_{t \in \mathcal{T}_2^{(i)}} s_t^{(i)} + \sum_{t \in \mathcal{T}_3^{(i)}} r_t s_t^{(i)}, \\ b[\lambda_{\mathrm{D}}^{(i)}] &= \sum_{t \in \mathcal{T}_2^{(i)}} (1 - s_t^{(i)}) + \sum_{t \in \mathcal{T}_3^{(i)}} r_t (1 - s_t^{(i)}).\end{aligned}$$

Similarly, $\lambda_{\mathrm{F}}^{(i)}$ is calculated as follows.

$$\begin{aligned}\frac{\partial}{\partial \lambda_{\mathrm{F}}^{(i)}} L(\boldsymbol{\theta}) = & \sum_{t \in \mathcal{T}_1^{(i)}} \left(s_t^{(i)} \frac{1}{\lambda_{\mathrm{F}}^{(i)}} - (1 - s_t^{(i)}) \frac{1}{1 - \lambda_{\mathrm{F}}^{(i)}} \right) \\ & + \sum_{t \in \mathcal{T}_3^{(i)}} (1 - r_t) \left(s_t^{(i)} \frac{1}{\lambda_{\mathrm{F}}^{(i)}} - (1 - s_t^{(i)}) \frac{1}{1 - \lambda_{\mathrm{F}}^{(i)}} \right) \\ = & \, 0,\end{aligned}$$

$$\lambda_{\mathrm{F}}^{(i)} = \frac{a[\lambda_{\mathrm{F}}^{(i)}]}{a[\lambda_{\mathrm{F}}^{(i)}] + b[\lambda_{\mathrm{F}}^{(i)}]},$$

where

$$\begin{aligned}a[\lambda_{\mathrm{F}}^{(i)}] &= \sum_{t \in \mathcal{T}_1^{(i)}} s_t^{(i)} + \sum_{t \in \mathcal{T}_3^{(i)}} (1 - r_t) s_t^{(i)}, \\ b[\lambda_{\mathrm{F}}^{(i)}] &= \sum_{t \in \mathcal{T}_1^{(i)}} (1 - s_t^{(i)}) + \sum_{t \in \mathcal{T}_3^{(i)}} (1 - r_t)(1 - s_t^{(i)}).\end{aligned}$$

After the M step, $\boldsymbol{\theta}'$ is updated by $\boldsymbol{\theta}$ obtained in the M step. A local optimal solution can be obtained by alternately repeating the calculation of $\mathbb{E}_{\boldsymbol{z}} \left[\ell(\boldsymbol{X}, \boldsymbol{z} \mid \boldsymbol{\theta}) \mid \boldsymbol{X}, \boldsymbol{\theta}' \right]$ at the E step and the calculation of $\boldsymbol{\theta}$ at the M step.

B. Estimating the Posterior Distributions of ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ by the Variational Bayes EM

The purpose of this section is to develop the estimation scheme for the posterior distributions of ρ (the channel utilization by the PU), $\lambda_{\mathrm{D}}^{(i)}$ (the detection probability of each SU) and $\lambda_{\mathrm{F}}^{(i)}$ (the false-alarm probability of each SU).

Let $q(\rho)$, $q(\lambda_{\mathrm{D}}^{(i)})$, and $q(\lambda_{\mathrm{F}}^{(i)})$ be the posterior distributions we want to derive. Since ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ are parameters of the Bernoulli distributions and the conjugate prior distribution of the Bernoulli distribution is the beta distribution, we assume that $q(\rho)$, $q(\lambda_{\mathrm{D}}^{(i)})$, and $q(\lambda_{\mathrm{F}}^{(i)})$ are the beta distributions. That is, $q(\rho)$, $q(\lambda_{\mathrm{D}}^{(i)})$, and $q(\lambda_{\mathrm{F}}^{(i)})$ are $\beta(\rho; a\,[\rho]\,, b\,[\rho])$, $\beta(\lambda_{\mathrm{D}}^{(i)}; a[\lambda_{\mathrm{D}}^{(i)}], b[\lambda_{\mathrm{D}}^{(i)}])$, and $\beta(\lambda_{\mathrm{F}}^{(i)}; a[\lambda_{\mathrm{F}}^{(i)}], b[\lambda_{\mathrm{F}}^{(i)}])$, respectively. Here, $a\,[\cdot]$ and $b\,[\cdot]$ represent parameters of the beta distribution followed by variables written in square brackets. The probability density function of the beta distribution is given by

$$\beta(x; a, b) = \frac{x^{a-1}(1-x)^{b-1}}{\mathrm{B}(a, b)},$$

where

$$\begin{aligned}
\mathrm{B}(a, b) &= \int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t \\
&= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \\
\Gamma(x) &= \int_0^\infty t^{x-1}\exp(-t)\,\mathrm{d}t.
\end{aligned}$$

$\mathrm{B}(a, b)$ and $\Gamma(x)$ in the above equation are called the beta function and the gamma function, respectively [22].

Let $p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})$ be a joint probability for complete data. The log likelihood of

$p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})$ is defined as

$$
\begin{aligned}
&\log p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}) = \log p(\boldsymbol{X}, \boldsymbol{Z} \mid \boldsymbol{\theta}) + \log p(\boldsymbol{\theta}) \\
&= \sum_{t \in \mathcal{T}_1} \left[\log(1-\rho) + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{F}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{F}}^{(i)}) \right] \\
&\quad + \sum_{t \in \mathcal{T}_2} \left[\log \rho + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{D}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{D}}^{(i)}) \right] \\
&\quad + \sum_{t \in \mathcal{T}_3} (1 - z_t) \left[\log(1-\rho) + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{F}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{F}}^{(i)}) \right] \\
&\quad + \sum_{t \in \mathcal{T}_3} z_t \left[\log \rho + \sum_{i \in \mathcal{C}_t} s_t^{(i)} \log \lambda_{\mathrm{D}}^{(i)} + \sum_{i \in \mathcal{C}_t} (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{D}}^{(i)}) \right].
\end{aligned}
$$

It is assumed that a prior probability distribution for parameters $P(\boldsymbol{\theta})$ is a uniform distribution.

We use the variational Bayes EM [26, 27] to find $q(\boldsymbol{Z}, \boldsymbol{\theta})$ that minimizes Kullback–Leibler divergence with $p(\boldsymbol{Z}, \boldsymbol{\theta} \mid \boldsymbol{X})$. By using a coordinate descent method, an equation for updating the probability distribution we want to obtain is given by

$$
\log q(x_j) = \mathbb{E}_{q\left(x_{i \neq j}\right)}[\log p(\boldsymbol{x})] + \mathrm{const},
$$

where $x_{i \neq j}$ is a collection of random variables except for x_i. The sum of all terms which are not related to the random variable that we are focusing on is expressed as *const* in the equation.

Initially, $a\,[\cdot]$ and $b\,[\cdot]$ are initialized with arbitrary random positive values. In the variational E step, we calculate $q(\boldsymbol{Z})$ using $q(\boldsymbol{\theta})$ as

$$
\begin{aligned}
\log q(z_t) &= \mathbb{E}_{q(\boldsymbol{\theta})} \left[\log p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}) \right] + \mathrm{const} \\
&= \mathbb{E}_{q(\rho)} \left[(1 - z_t) \log(1-\rho) + z_t \log \rho \right] \\
&\quad + \sum_{i \in \mathcal{C}_t} \mathbb{E}_{q(\lambda_{\mathrm{F}}^{(i)})} \left[(1 - z_t) \left\{ s_t^{(i)} \log \lambda_{\mathrm{F}}^{(i)} + (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{F}}^{(i)}) \right\} \right] \\
&\quad + \sum_{i \in \mathcal{C}_t} \mathbb{E}_{q(\lambda_{\mathrm{D}}^{(i)})} \left[z_t \left\{ s_t^{(i)} \log \lambda_{\mathrm{D}}^{(i)} + (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{D}}^{(i)}) \right\} \right] + \mathrm{const}.
\end{aligned}
$$

In order to calculate the above equation, the expected values of $\log x$ and $\log(1-x)$ where x follows the beta distribution $\beta(x; a, b)$ are required. These expected values are calculated by

$$\begin{aligned}\int_0^1 \frac{x^{a-1}(1-x)^{b-1}}{\mathrm{B}(a,b)} \log x \, \mathrm{d}x &= \frac{1}{\mathrm{B}(a,b)} \int_0^1 \frac{\partial}{\partial a} x^{a-1}(1-x)^{b-1} \mathrm{d}x \\ &= \frac{1}{\mathrm{B}(a,b)} \frac{\partial}{\partial a} \mathrm{B}(a,b) \\ &= \frac{\partial}{\partial a} \log \mathrm{B}(a,b) \\ &= \frac{\partial}{\partial a} (\log \Gamma(a) + \log \Gamma(b) - \log \Gamma(a+b)) \\ &= \psi(a) - \psi(a+b),\end{aligned}$$

$$\begin{aligned}\int_0^1 \frac{x^{a-1}(1-x)^{b-1}}{\mathrm{B}(a,b)} \log(1-x) \, \mathrm{d}x &= \frac{1}{\mathrm{B}(a,b)} \int_0^1 \frac{\partial}{\partial b} x^{a-1}(1-x)^{b-1} \mathrm{d}x \\ &= \frac{1}{\mathrm{B}(a,b)} \frac{\partial}{\partial b} \mathrm{B}(a,b) \\ &= \frac{\partial}{\partial b} \log \mathrm{B}(a,b) \\ &= \frac{\partial}{\partial b} (\log \Gamma(a) + \log \Gamma(b) - \log \Gamma(a+b)) \\ &= \psi(b) - \psi(a+b).\end{aligned}$$

The logarithmic derivative of the gamma function is called the digamma function [22] and is defined by

$$\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x} \log \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}.$$

Specific values of the digamma function can be calculated using numerical libraries such as `boost/math/special_functions/digamma.hpp` for C++ and `scipy.special.digamma` for Python.

Thereby, the expected value of $\log p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})$ with respect to $q(\rho)$ is calculated as

$$\begin{aligned}&\mathbb{E}_{q(\rho)} \left[(1-z_t) \log(1-\rho) + z_t \log \rho \right] \\ &= (1-z_t)(\psi(b[\rho]) - \psi(a[\rho] + b[\rho])) + z_t(\psi(a[\rho]) - \psi(a[\rho] + b[\rho])) \\ &= z_t(\psi(a[\rho]) - \psi(b[\rho])) + \text{const.}\end{aligned}$$

Next, the expected value of $\log p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})$ with respect to $q(\lambda_{\mathrm{D}}^{(i)})$ is calculated as

$$
\begin{aligned}
&\mathbb{E}_{q(\lambda_{\mathrm{D}}^{(i)})}\left[z_t\left\{s_t^{(i)}\log\lambda_{\mathrm{D}}^{(i)}+(1-s_t^{(i)})\log(1-\lambda_{\mathrm{D}}^{(i)})\right\}\right]\\
&=z_t\left\{s_t^{(i)}\left(\psi\left(a[\lambda_{\mathrm{D}}^{(i)}]\right)-\psi\left(a[\lambda_{\mathrm{D}}^{(i)}]+b[\lambda_{\mathrm{D}}^{(i)}]\right)\right)\right.\\
&\qquad\left.+(1-s_t^{(i)})\left(\psi\left(b[\lambda_{\mathrm{D}}^{(i)}]\right)-\psi\left(a[\lambda_{\mathrm{D}}^{(i)}]+b[\lambda_{\mathrm{D}}^{(i)}]\right)\right)\right\}.
\end{aligned}
$$

Similarly, the expected value of $\log p(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})$ with respect to $q(\lambda_{\mathrm{F}}^{(i)})$ is calculated as

$$
\begin{aligned}
&\mathbb{E}_{q(\lambda_{\mathrm{F}}^{(i)})}\left[(1-z_t)\left\{s_t^{(i)}\log\lambda_{\mathrm{F}}^{(i)}+(1-s_t^{(i)})\log(1-\lambda_{\mathrm{F}}^{(i)})\right\}\right]\\
&=(1-z_t)\left\{s_t^{(i)}\left(\psi\left(a[\lambda_{\mathrm{F}}^{(i)}]\right)-\psi\left(a[\lambda_{\mathrm{F}}^{(i)}]+b[\lambda_{\mathrm{F}}^{(i)}]\right)\right)\right.\\
&\qquad\left.+(1-s_t^{(i)})\left(\psi\left(b[\lambda_{\mathrm{F}}^{(i)}]\right)-\psi\left(a[\lambda_{\mathrm{F}}^{(i)}]+b[\lambda_{\mathrm{F}}^{(i)}]\right)\right)\right\}.
\end{aligned}
$$

As a result, $q(z_t)$ is expressed as follows.

$$\log q(z_t)=\eta_t z_t+\mathrm{const},$$

$$q(z_t)=\frac{\exp(\eta_t z_t)}{\sum_{k\in\{0,1\}}\exp(\eta_t k)},$$

where

$$
\begin{aligned}
\eta_t=&\left(\psi(a[\rho])-\psi(b[\rho])\right)\\
&+\sum_{i\in\mathcal{C}_t}s_t^{(i)}\left(\psi\left(a[\lambda_{\mathrm{D}}^{(i)}]\right)-\psi\left(a[\lambda_{\mathrm{D}}^{(i)}]+b[\lambda_{\mathrm{D}}^{(i)}]\right)\right)\\
&+\sum_{i\in\mathcal{C}_t}(1-s_t^{(i)})\left(\psi\left(b[\lambda_{\mathrm{D}}^{(i)}]\right)-\psi\left(a[\lambda_{\mathrm{D}}^{(i)}]+b[\lambda_{\mathrm{D}}^{(i)}]\right)\right)\\
&-\sum_{i\in\mathcal{C}_t}s_t^{(i)}\left(\psi\left(a[\lambda_{\mathrm{F}}^{(i)}]\right)-\psi\left(a[\lambda_{\mathrm{F}}^{(i)}]+b[\lambda_{\mathrm{F}}^{(i)}]\right)\right)\\
&-\sum_{i\in\mathcal{C}_t}(1-s_t^{(i)})\left(\psi\left(b[\lambda_{\mathrm{F}}^{(i)}]\right)-\psi\left(a[\lambda_{\mathrm{F}}^{(i)}]+b[\lambda_{\mathrm{F}}^{(i)}]\right)\right).
\end{aligned}
$$

In the variational M step, we calculate $q(\boldsymbol{\theta})$ using $q(\boldsymbol{Z})$. Let r_t be $\mathbb{E}_{q(z_t)}[z_t]=q(z_t=1)$. Note that $1-r_t=\mathbb{E}_{q(z_t)}[1-z_t]=q(z_t=0)$. Then, $q(\rho)$ is calculated as follows.

$$
\begin{aligned}
\log q(\rho) &= \mathbb{E}_{q(\boldsymbol{Z})}\left[\log p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})\right] + \text{const} \\
&= \sum_{t \in \mathcal{T}_1} \log(1-\rho) + \sum_{t \in \mathcal{T}_2} \log \rho \\
&\quad + \mathbb{E}_{q(\boldsymbol{Z})}\left[\sum_{t \in \mathcal{T}_3} (1 - z_t) \log(1-\rho) + \sum_{t \in \mathcal{T}_3} z_t \log \rho\right] + \text{const} \\
&= \left(|\mathcal{T}_2| + \sum_{t \in \mathcal{T}_3} r_t\right) \log \rho \\
&\quad + \left(|\mathcal{T}_1| + \sum_{t \in \mathcal{T}_3} (1 - r_t)\right) \log(1-\rho) + \text{const},
\end{aligned}
$$

$$
\begin{aligned}
q(\rho) &= \frac{\exp((a[\rho]-1)\log\rho + (b[\rho]-1)\log(1-\rho))}{\int_0^1 \exp((a[\rho]-1)\log t + (b[\rho]-1)\log(1-t))\,\mathrm{d}t} \\
&= \frac{\rho^{a[\rho]-1}(1-\rho)^{b[\rho]-1}}{\int_0^1 t^{a[\rho]-1}(1-t)^{b[\rho]-1}\mathrm{d}t} \\
&= \beta(\rho; a[\rho], b[\rho]),
\end{aligned}
$$

where

$$
\begin{aligned}
a[\rho] &= |\mathcal{T}_2| + \sum_{t \in \mathcal{T}_3} r_t + 1, \\
b[\rho] &= |\mathcal{T}_1| + \sum_{t \in \mathcal{T}_3} (1 - r_t) + 1.
\end{aligned}
$$

Next, $q(\lambda_{\mathrm{D}}^{(i)})$ is calculated as follows.

$$
\begin{aligned}
\log q(\lambda_{\mathrm{D}}^{(i)}) &= \mathbb{E}_{q(\boldsymbol{Z})}\left[\log p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})\right] + \mathrm{const} \\
&= \sum_{t \in \mathcal{T}_2^{(i)}} \left(s_t^{(i)} \log \lambda_{\mathrm{D}}^{(i)} + (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{D}}^{(i)}) \right) \\
&\quad + \mathbb{E}_{q(\boldsymbol{Z})}\left[\sum_{t \in \mathcal{T}_3^{(i)}} z_t \left(s_t^{(i)} \log \lambda_{\mathrm{D}}^{(i)} + (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{D}}^{(i)}) \right) \right] + \mathrm{const} \\
&= \left(\sum_{t \in \mathcal{T}_2^{(i)}} s_t^{(i)} + \sum_{t \in \mathcal{T}_3^{(i)}} r_t s_t^{(i)} \right) \log \lambda_{\mathrm{D}}^{(i)} \\
&\quad + \left(\sum_{t \in \mathcal{T}_2^{(i)}} (1 - s_t^{(i)}) + \sum_{t \in \mathcal{T}_3^{(i)}} r_t (1 - s_t^{(i)}) \right) \log(1 - \lambda_{\mathrm{D}}^{(i)}) + \mathrm{const},
\end{aligned}
$$

$$
\begin{aligned}
q(\lambda_{\mathrm{D}}^{(i)}) &= \frac{\exp\left((a[\lambda_{\mathrm{D}}^{(i)}] - 1) \log \lambda_{\mathrm{D}}^{(i)} + (b[\lambda_{\mathrm{D}}^{(i)}] - 1) \log(1 - \lambda_{\mathrm{D}}^{(i)}) \right)}{\int_0^1 \exp\left((a[\lambda_{\mathrm{D}}^{(i)}] - 1) \log t + (b[\lambda_{\mathrm{D}}^{(i)}] - 1) \log(1 - t) \right) \mathrm{d}t} \\
&= \frac{(\lambda_{\mathrm{D}}^{(i)})^{a[\lambda_{\mathrm{D}}^{(i)}] - 1} (1 - \lambda_{\mathrm{D}}^{(i)})^{b[\lambda_{\mathrm{D}}^{(i)}] - 1}}{\int_0^1 t^{a[\lambda_{\mathrm{D}}^{(i)}] - 1} (1 - t)^{b[\lambda_{\mathrm{D}}^{(i)}] - 1} \mathrm{d}t} \\
&= \beta(\lambda_{\mathrm{D}}^{(i)}; a[\lambda_{\mathrm{D}}^{(i)}], b[\lambda_{\mathrm{D}}^{(i)}]),
\end{aligned}
$$

where

$$
\begin{aligned}
a[\lambda_{\mathrm{D}}^{(i)}] &= \sum_{t \in \mathcal{T}_2^{(i)}} s_t^{(i)} + \sum_{t \in \mathcal{T}_3^{(i)}} r_t s_t^{(i)} + 1, \\
b[\lambda_{\mathrm{D}}^{(i)}] &= \sum_{t \in \mathcal{T}_2^{(i)}} (1 - s_t^{(i)}) + \sum_{t \in \mathcal{T}_3^{(i)}} r_t (1 - s_t^{(i)}) + 1.
\end{aligned}
$$

Similarly, $q(\lambda_{\mathrm{F}}^{(i)})$ is calculated as follows.

$$
\begin{aligned}
\log q(\lambda_{\mathrm{F}}^{(i)}) &= \mathbb{E}_{q(\boldsymbol{Z})}\left[\log p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})\right] + \text{const} \\
&= \sum_{t \in \mathcal{T}_1^{(i)}} \left(s_t^{(i)} \log \lambda_{\mathrm{F}}^{(i)} + (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{F}}^{(i)}) \right) \\
&\quad + \mathbb{E}_{q(\boldsymbol{Z})}\left[\sum_{t \in \mathcal{T}_3^{(i)}} (1 - z_t) \left(s_t^{(i)} \log \lambda_{\mathrm{F}}^{(i)} + (1 - s_t^{(i)}) \log(1 - \lambda_{\mathrm{F}}^{(i)}) \right) \right] + \text{const} \\
&= \left(\sum_{t \in \mathcal{T}_1^{(i)}} s_t^{(i)} + \sum_{t \in \mathcal{T}_3^{(i)}} (1 - r_t) s_t^{(i)} \right) \log \lambda_{\mathrm{F}}^{(i)} \\
&\quad + \left(\sum_{t \in \mathcal{T}_1^{(i)}} (1 - s_t^{(i)}) + \sum_{t \in \mathcal{T}_3^{(i)}} (1 - r_t)(1 - s_t^{(i)}) \right) \log(1 - \lambda_{\mathrm{F}}^{(i)}) + \text{const},
\end{aligned}
$$

$$
\begin{aligned}
q(\lambda_{\mathrm{F}}^{(i)}) &= \frac{\exp\left((a[\lambda_{\mathrm{F}}^{(i)}] - 1) \log \lambda_{\mathrm{F}}^{(i)} + (b[\lambda_{\mathrm{F}}^{(i)}] - 1) \log(1 - \lambda_{\mathrm{F}}^{(i)}) \right)}{\int_0^1 \exp\left((a[\lambda_{\mathrm{F}}^{(i)}] - 1) \log t + (b[\lambda_{\mathrm{F}}^{(i)}] - 1) \log(1 - t) \right) \mathrm{d}t} \\
&= \frac{(\lambda_{\mathrm{F}}^{(i)})^{a[\lambda_{\mathrm{F}}^{(i)}] - 1} (1 - \lambda_{\mathrm{F}}^{(i)})^{b[\lambda_{\mathrm{F}}^{(i)}] - 1}}{\int_0^1 t^{a[\lambda_{\mathrm{F}}^{(i)}] - 1} (1 - t)^{b[\lambda_{\mathrm{F}}^{(i)}] - 1} \mathrm{d}t} \\
&= \beta(\lambda_{\mathrm{F}}^{(i)}; a[\lambda_{\mathrm{F}}^{(i)}], b[\lambda_{\mathrm{F}}^{(i)}]),
\end{aligned}
$$

where

$$
\begin{aligned}
a[\lambda_{\mathrm{F}}^{(i)}] &= \sum_{t \in \mathcal{T}_1^{(i)}} s_t^{(i)} + \sum_{t \in \mathcal{T}_3^{(i)}} (1 - r_t) s_t^{(i)} + 1, \\
b[\lambda_{\mathrm{F}}^{(i)}] &= \sum_{t \in \mathcal{T}_1^{(i)}} (1 - s_t^{(i)}) + \sum_{t \in \mathcal{T}_3^{(i)}} (1 - r_t)(1 - s_t^{(i)}) + 1.
\end{aligned}
$$

The posterior distributions of ρ, $\lambda_{\mathrm{D}}^{(i)}$, and $\lambda_{\mathrm{F}}^{(i)}$ that minimize Kullback–Leibler divergence with $p(\boldsymbol{Z}, \boldsymbol{\theta} \mid \boldsymbol{X})$ can be calculated by alternately repeating the calculation of $q(\boldsymbol{Z})$ at the variational E step and the calculation of $q(\boldsymbol{\theta})$ at the variational M step.

Acknowledgements

本研究に取り組むにあたって，笠原正治教授，川原純助教には定期的な個人ミーティングで親身かつ熱心なご指導を賜りましたことを心からお礼申し上げます．数理情報学研究室の池田和司教授には副指導としてご指導，ご指摘をいただきました．笹部昌弘准教授，張元玉助教には研究室ミーティングで多くの有意義なご助言をいただきましたことを感謝いたします．秘書の橋本洋子様，林英子様には研究生活に関するさまざまなお力添えを賜りました．研究を通じて数多くの知識や助言をいただいた大規模システム管理研究室の皆様にも感謝の意を表します．最後に，この２年間様々な面で関わった皆様に心より感謝致します．

Publication List

1. Sho Iizuka, Jun Kawahara, and Shoji Kasahara, "Coalition Formation Approach for Cooperative Spectrum Sensing in Cognitive Radio Networks by Multi-Armed Bandit Problem", Technical Report of IEICE, vol. 116, no. 484, NS2016-242, pp. 487–492 (Mar. 2017).

2. Sho Iizuka, Jun Kawahara, and Shoji Kasahara, "Performance Inference for Cooperative Spectrum Sensing with the k-out-of-N Rule: An MCMC-based Approach", Technical Report of IEICE, vol. 117, no. 204, NS2017-82, pp. 67–72 (Sep. 2017).

3. Sho Iizuka, Jun Kawahara, and Shoji Kasahara, "Group Formation for Cooperative Spectrum Sensing for Minimizing Miss-Detection and False-Alarm Under Uncertainty", IEEE COMSOC Kansai Chapter Student Workshop (Oct. 2017).