Pure-Exploration Bandits for Channel Selection in Mission-Critical Wireless Communications

Yuan Xue, *Student Member, IEEE*, Pan Zhou, *Member, IEEE*, Shiwen Mao, *Senior Member, IEEE*, Dapeng Wu, *Fellow, IEEE*, and Yingjie Zhou, *Member, IEEE*

***Abstract*—In emergency communications, guaranteeing ultra-reliable and low-latency communication is challenging yet crucial to save human lives and to coordinate the operations of first responders. To address this problem, we introduce a general approach for channel selection in mission-critical communications, i.e., choose channels with the best quality timely and accurately via channel probing. Since the channel conditions are dynamic and initially unknown to wireless users, choosing channels with the best conditions is nontrivial. Thus, we adopt online learning methods to let users probe channels and predict the channel conditions by a restricted time interval of observation. We formulate this problem as an emerging branch of the classic multiarmed bandit (MAB) problem, namely the *pure-exploration bandit* problem, to achieve a tradeoff between sampling time/resource budget and the channel selection accuracy (i.e., the probability of selecting optimal channels). The goal of the learning process is to choose the "optimal subset" of channels after a limited time period of channel probing. We propose and evaluate one learning policy for the single-user case and three learning policies for the distributed multiuser cases. We take communication costs and interference costs into account, and analyze the tradeoff between these costs and the accuracy of channel selection. Extensive simulations are conducted and the results show that the proposed algorithms can achieve considerably higher channel selection accuracy than previous exploration bandit approaches and classic MAB methods.**

***Index Terms*—Online learning, emergency communications, channel probing, exploration multi-armed bandits.**

Manuscript received October 14, 2017; revised April 17, 2018; accepted August 6, 2018. Date of publication August 20, 2018; date of current version November 12, 2018. This work was supported by the National Science Foundation of China under Grants 61401169 and 61801315 and in part by the US NSF under Grants ECCS-1509212, CNS-1247955, and CNS-1702957. The review of this paper was coordinated by Honggang Wang. *(Corresponding authors: Pan Zhou and Yingjie Zhou.)*

Y. Xue was with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, 430074, China. He is now with the Department of Computer Science and Engineering, Lehigh University, Bethlehem, PA 18018 USA (e-mail: yux715@lehigh.edu).

P. Zhou is with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, 430074, China (e-mail: panzhou@hust.edu.cn).

S. Mao is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849 USA (e-mail: smao@ieee.org).

D. Wu is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: dpwu@ieee.org).

Y. Zhou is with the College of Computer Science, Sichuan University, Chengdu 610065, China (e-mail: yjzhou09@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2018.2866198

I. Introduction

EMERGENCY circumstances such as accidents, natural disasters, or terrorist attacks require immediate attention from first responders and can be considered as mission-critical conditions. In such scenarios, guaranteeing ultra-reliable and low-latency communications is challenging yet crucial to save human lives and to coordinate the operations of first responders. Due to the complex wireless environment and the limitation of wireless resources, in many such scenarios, wireless users must be well aware of the wireless channel conditions to select the best channels in a timely manner. However, this is challenging since channel conditions are highly dynamic and users have no prior knowledge of channel changes. Thus, learning-based resource management algorithms are always adopted to solve this problem.

In the channel selection problem, channel selection accuracy represents the probability of the user choosing optimal channels. For critical communications where reliability and latency are of great importance, selecting channels timely and accurately is fundamental for high-quality communications. For instance, in public safety communications [1], choosing a high-rate and ultra-reliable wireless channel can guarantee the quality of communication, and could potentially save lots of lives. In addition, to guarantee low latency for emergency messages, the time it takes for channel selection cannot be too long. Thus, choosing channels with the best quality in a given, the short time duration is clearly a crucial step before the actual communication takes place.

The Defense Advanced Research Projects Agency (DARPA) raises a Spectrum Collaboration Challenge [2] to help to ensure that the exponentially growing number of military and civilian wireless devices will have full access to the increasingly crowded electromagnetic spectrums. Classical approaches manually divide the spectrum into licensed bands and allocate them to primary users, while secondary users share the unused channels. Apparently, those approaches are not very suitable for the increasing spectrum demand and dynamically changing environments. Under such a circumstance, next-generation spectrum access strategies may also require channel sharing among primary users. Especially, in channel access for critical communications, users with high priorities also need to share channels with other users. Thus, channel selection, as an essential step to choose optimal channels before channel access, requires all wireless users to actively probe multiple

channels while these channels are shared among users at the same time. Active channel probing allows a user to collect sufficient channel information to select a bunch of high-quality channels based on probing results; Channel sharing guarantees minimum or no interference between different users. After channel selection, users can follow a uniform channel allocation or a user negotiation strategy to share channels with other users.

Existing works with learning-based methods on channel probing and sharing, such as [3], [4], mainly solve the problem of achieving the highest cumulative throughput for secondary users based on the historical communication information. However, such methods are not applicable to the channel selection problems in critical communications because they only consider cumulative results. Such existing schemes may choose a sub-optimal channel whose quality is close to the best channel, while still achieve good performance. Therefore, relatively good (but not the best) channels can be chosen during the entire process and it can be harmful to emergency communications. Moreover, the existing schemes need to constantly change the target channel during the communication process, leading to potentially high channel switching cost for users. If we divide the process into a channel probing period and a channel access period separately, we could avoid these problems. Consider the channel allocation problem for mobile phone communications illustrated in [5], during a very short time period before the communication starts, a cell phone can explore the set of channels to identify the best one to operate on. Each evaluation of a channel is noisy and there are a limited number of evaluations before the communication starts. The connection is then launched on the channel which is believed to be the best. The cumulative throughput during the exploration phase is irrelevant since the user is only interested in the quality of its communication after the exploration phase. Apparently, using a metric of selection accuracy is more appropriate than cumulative throughput in this scenario.

In this paper, we investigate the problem of channel selection for multiple users of mission-critical communications in both centralized and distributed paradigms. A user as a wireless device is given a time budget to perform channel probing, then it tries to choose the best available channel to access based on its channel probing results . When multiple users coexist in the same area, we try to figure out how to coordinate these users and make the optimal selection of channels. The goal is to maximize the user's chance of choosing optimal channels after the channel probing period.

Since the channel activities are initially unknown to the users, intuitively, if the users spend more time on probing, they will obtain more accurate results. However, in mission-critical communications, users cannot spend a long time on channel selection thus a time budget is needed. As a result, there is a trade-off between completing the probing more quickly versus making a more accurate selection of channels. This non-trivial problem inspires us to formulate a *distributed exploration bandit* or *pure exploration* problem, which is a subclass of the classic multi-armed bandit (MAB) problem [6]. In contrast to standard MAB algorithms such as UCB [7], which are evaluated in terms of *cumulative regret*,[1] pure exploration methods focus on identifying the arm(s) with the maximum expected rewards rather than maximizing arm rewards during the entire learning process.

First, we study the channel selection problem with a single user. Although the user can only probe one channel at a time, it can sample different channels sequentially. Thus, by predicting channel condition after a short time of observation, we could have the user probe multiple channels. We model this problem as a multiple arm identification problem with a *fixed budget* [8]. We propose an elimination-based learning algorithm, which allows the user to reduce the number of channels to be probed during the iterations.

Next, we consider the case where multiple users coexist in the same area. If the users are allowed to communicate with each other, they can share channel information to obtain more accurate probing results. Nevertheless, due to the geographical separation of the users, such cooperation incurs extra communication cost. For such a scenario, we propose a distributed exploration bandit algorithm with limited communications. Specifically, during the channel probing period, when multiple users targeting at the same channel simultaneously, interferences between different users will hurt their ability to get accurate results and cause energy waste. We call this type of interference as *collision* and we want to avoid collisions during channel probing and access.

The main contributions of this work are summarized as follows:

- We address the channel selection problem in critical communications by formulating it as an exploration bandit problem and develop effective solution algorithms. Compared with classic exploration-exploitation methods, our algorithms can achieve better performance in terms of channel selection accuracy. Both single and multiple user algorithms are proposed for different applications in this paper.
- We investigate the distributed exploration bandit problem in the fixed budget setting. This online learning technique has lots of potential applications in wireless communications but few prior works have been done.
- Communication costs and collision losses are taken into consideration in this paper. Several well-designed algorithms are proposed to mitigate the negative impact of collisions and improve channel selection accuracy by exploiting communications among users.

The rest of this paper is organized as follows. In Section II, related works on channel sharing and distributed online learning are discussed. Section III introduces the problem formulation. Algorithms for the single user and multiple users are presented in Section IV and V, respectively. In Section VI we discuss how to access channels after channel selection. Simulation results are presented in Section VII. Finally, we conclude this paper in Section VIII.

[1]Cumulative regret is defined as the cumulative difference between the expected rewards of the optimal strategy made by a *genie* and that of the given policy in the whole process.

II. Related Work

In this section, we introduce the key related works on spectrum sharing and distributed online learning. For spectrum sharing [9], many works have been done, such as [3], [4], but the existing algorithms are not suitable for the primary user spectrum sharing problem. For channel access of secondary users, existing algorithms (e.g., [10], [11]) let users keep on probing the channels and switching the target channel according to probing results. However, in critical communications, even licensed users need to share channels during and after the channel selection process. By adopting channel selection techniques in this paper, we guarantee that primary users are able to find a (set of) good channel(s) and they will not suffer high channel switching costs after channel selection.

In critical wireless communications such as public safety applications [12], there are also some efforts on spectrum sharing [13]–[15], but the same flaw as in classic spectrum sharing has been observed. Furthermore, cooperation in spectrum probing has been studied [16] but the communication cost can be high if the algorithm does not take such cost into consideration. Some papers studied collaborative algorithms in channel probing and resource allocation [17], [18]. However, they did not fully consider all the possible conditions such as cooperations may not be allowed in some cases. Tan *et al.* [19] treated the joint process of channel probing and scheduling for communications under delay constraints as a maximal-rate-of-return problem. Authors used a pure threshold policy as optimal distributed opportunistic scheduling method, but they didn't consider cooperations between users. In contrast to these prior works, we take all the possible cases into consideration and propose different algorithms under different circumstances.

In the area of online learning, multi-armed bandit (MAB) problem has drawn a lot of attention in recent years. MAB is a classic example of the tradeoff between exploration and exploitation, aiming to achieve the maximum cumulative sum of rewards in the learning process. Lai & Robbins proposed an index policy in [6] with a logarithmic regret bound and Auer *et al.* [7] introduced the well-known UCB strategy which achieves $O(\log T)$ regret uniformly over time.

Exploration bandit is a new branch of MAB and it can be divided into two main categories: fixed budget setting and fixed confidence setting. In the fixed budget setting, players should seek for a single best arm or a best subset of arms with a fixed time budget. In fixed confidence setting, such as Even-Dar *et al.* in [20] and Kalyanakrishnan *et al.* in [21], players aim at reducing the number of samples (i.e., *simple regret* in [22]) to satisfy the specific constant of finding near-optimal arm(s). Since channel probing time is limited for critical communications, we focus on fixed budget setting in this paper. Compared with classic exploration-exploitation MAB methods, pure exploration bandit methods can achieve more accurate results in channel selection because the cumulative regret evaluation metric is not suitable for lowering the channel selection error probability, especially when time budget is small since classic MAB methods spend significantly more time on sub-optimal channels at the beginning stage. Above all, we develop fixed budget exploration bandit algorithms for channel selections in this paper.

In the exploration bandit problem, many works on the multiple identifications problem (Explore-m in [23]) has been done, such as [8], [21], [24]. Recently, Shahrampour *et al.* [25] have proposed a general unified theory for sequential elimination algorithms in exploration bandit. The authors indicate its important applications for mobile communications where users can explore the set of channels (arms) to find the best one to operate. Nevertheless, none of the existing work has considered subset selection with different players. Different from the existing literature, we study the challenging problem of multiple identifications in the distributed setting.

There have also been considerable efforts on distributed learning techniques for MAB. In the area of distributed learning, Liu and Zhao [3] introduced the time-division fair sharing (TDFS) policy for a centralized time-sharing schedule for multiple users. Tekin and Liu [26] utilized the regenerative property of the Markov chain to solve the problem of rested and restless MAB problems with multiple players. Kalathil *et al.* in [27] proposed an algorithm based on the Bertsekas auction algorithm, which has $O(\log^2 t)$ regret bound due to the communication cost.

Distributed exploration bandit was studied by Hillel *et al.* in [28]. Their work is most related to this paper. However, Hillel *et al.* studied the distributed exploration problem in the fixed confidence setting while we mainly focus on the fixed budget setting. They speed up the learning process via communications among users. In contrast to their work, algorithms requiring different amounts of communications are proposed in our paper to handle different scenarios. Our preliminary work focused on wireless monitoring is published in [29], this paper studies a completely different and more important application in emergency communications, provides more rigorous theoretical analysis, and proposes new collaborative algorithm and channel access policies where cooperations improve channel selection performance significantly.

III. Problem Formulation

A. System Model

Consider single or multiple users try to access some channels among K wireless channels in critical wireless communications. When K is large, the user wants to choose a subset of best channels for high-quality communications. The communication process is consist of two phases: channel selection period and channel access period. For channel selection, channel information is collected by probing K channels, then users select best channels based on channel probing results. In many emergency circumstances, the channel selection task needs to be completed with limited resources (e.g., time, energy, etc.). In channel selection period, to guarantee the lowest communication delay, we focus on the limited time scenario where the user chooses M $(M < K)$ best channels out of K channels within the time budget T. In different scenarios, the time budget varies and users accomplish the channel selection with or without communicating with each other. After channel selection, users access channels based on their channel selection results for communication.

For channel probing, we assume that a user can only probe one channel at a time. In the single user scenario, the user is

TABLE I
SUMMARY OF NOTATION

Notation	Definition
$\mathcal{K}$	set of channels, and $\|\mathcal{K}\| = K$
$\mathcal{M}$	set of best (or worst) channels, and $\|\mathcal{M}\| = M$
$\mathcal{N}$	set of users, and $\|\mathcal{N}\| = n$
T	time budget, number of time slots
t_τ	sampling time for round τ
l	number of total rounds
$\mathcal{A}_\tau$	set of remaining channels in round τ, and $\|\mathcal{A}_\tau\| = A_\tau$
K_τ	number of channels to be eliminated in round τ
μ_j	actual mean rewards for channel j
μ_M	actual mean rewards for Mth best channel
$\hat{\mu}_j$	empirical mean rewards for channel j
e_i^ϕ	user i's probability of wrong selection under policy ϕ
r_i	simple regret of channel selection results for user i
Δ_j^M	difference in mean rewards between channel j and M
H^M	hardness of the channel selection problem
$c_{i,j}$	user i's communication cost of broadcasting channel j
c_0	complexity of communication cost in each time slot
R	number of time slots in channel accessing period

given a time budget T firstly, then it chooses M channels within time T. In most applications, we can use the simplest assumption where one user just need to choose one channel after the probing phase. After finishing the channel selection, the user will access the channel according to the probing results.

In the multi-user scenario, each user will get a complete outcome of M channels ($M = 1$ in most cases) independently. Since users will actively probe channels, interferences will happen if multiple users probe the same channel simultaneously. Thus, we should try to mitigate or get rid of the negative effect of collisions.

Assume there are n users and $n \leq M$. If all users are allowed to communicate with each other, they can exchange information during the probing process to avoid collisions and improve the channel selection accuracy. However, extra communication costs will also affect the probing results achieved by the users, and further hurt the accuracy of the selection. Thus, algorithms with an appropriate amount of communications are necessary for channel selection. After channel selection, n users access channels among M chosen channels based on channel selection result.

Next, we introduce some notations (summarized in Table I) and complexity measurements in this paper.

B. Notations and Complexity Measurements

Consider K channels in a wireless network, where $\mathcal{K} = \{1, \ldots, K\}$ is the channel pool. For simplicity, we assume that each channel j's activity in the wireless network follows an i.i.d. distribution with density function $f(x; \theta_j)$, while the parameter θ_j is a unknown *priori*. Each time the user observes the channels, it will obtain a *reward* which contains the channel information. When there are multiple users in the system, let $\mathcal{N} = \{1, \ldots, n\}$ denote the set of users. Let ϕ be the channel selection policy adopted by the users. During the channel selection process, all users have the same time budget T and their clocks are synchronized. Assume that each channel j has a mean reward μ_j according to its density function $f(x; \theta_j)$, which is the mean of random variable $X_j(t)$. We rank them in descending order, i.e., $\mu_1 > \cdots > \mu_K$. The ground truth of the channel selection result is the channel set $\mathcal{M}$ with mean rewards $\mu_1, \ldots, \mu_M$. After the channel probing phase, users obtain the empirical reward $\hat{\mu}_j$ for channel j, and $\hat{\mu}_j = \sum_t X_j(t)/T$. We also rank these empirical rewards in a descending order. The user chooses the first M channels with the highest empirical rewards as its channel selection result, denoted by channel set $\hat{\mathcal{M}}$. The error probability for channel selection can be defined as

$$e = \mathbb{P}[\hat{\mathcal{M}} \neq \mathcal{M}]. \tag{1}$$

The channel selection accuracy is defined as $1 - e$. We also define the error probability e_i^ϕ to be the probability of choosing sub-optimal channels by user i under policy ϕ. Divide the total probing time into several rounds, in round τ, users will spend t_τ time on its target channels. For user i, our goal is to minimize the error probability within the given time budget, which means

$$\min e_i^\phi \text{ s.t. } \sum_\tau t_\tau < T. \tag{2}$$

In our experiments, to better compare performance between different algorithms under a uniform metric, we use the evaluation metric *simple regret* [22] which represents the difference between true means of the optimal M channels and that of channel chosen by the users. For user i, we define simple regret r_i as

$$r_i = \sum_{j=1}^{M} (\mu_j - \hat{\mu}_j), \tag{3}$$

and the total simple regret for all user is $\sum_{i=1}^n r_i$. We also introduce the notation of hardness H^M. We define the gaps and the complexity measures of the distributed channel selection in mission-critical communication as follows

$$\Delta_j^M = |\mu_M - \mu_j|, \qquad \Delta_{\min} = \min_{1 \leq j < k \leq K} (\mu_k - \mu_j), \tag{4}$$

$$H_1^M = \sum_{j=2}^{K} \frac{1}{\left(\Delta_j^M\right)^2}, \qquad H_2^M = \max_{j \in \mathcal{K}} \frac{j}{\left(\Delta_j^M\right)^2}, \tag{5}$$

where the notation $j \in \{1, \ldots, K\}$ is determined by order of $\Delta_1^M \leq \cdots \leq \Delta_K^M$. Note that from [5] we know H_1^M and H_2^M are equivalent up to a logarithmic factor, and we have $H_2^M \leq H_1^M \leq H_2^M \log 2K$. These notations decide the lower bound on the number of evaluations necessary to identify the best channel, which means the hardness of finding the optimal channels during the channel selection process. We will discuss more details in the following sections.

IV. SINGLE USER CHANNEL SELECTION

In this section, we introduce a novel single user channel selection algorithm, namely Sequential Multiple Elimination (SME) for channel selection in critical communications. Details of the proposed algorithm are given in Algorithm 1. The general idea for SME is to maintain an active set initialized by K channels, and then to discard channels sequentially within the time budget until there are M channels left. Time budget is divided

Algorithm 1: Sequential Multiple Elimination for Single User (SME).

1: **Input** : K channels, M chosen channels, time budget T, learning rate η.
2: **Initialization** : Let $l = \lceil \log_\eta ((\eta-1)(K-M)+1) \rceil$, $\mathcal{A}_0 = \mathcal{K}$, $A_\tau = |\mathcal{A}_\tau|$, $K_\tau = \left\lceil \frac{(\eta-1)(K-M)+1}{\eta^\tau} \right\rceil$.
3: **for** each $\tau = 1, 2, \ldots, l$ **do**
4: Sample all channels in $\mathcal{A}_\tau$ for $t_\tau = \lceil \frac{T}{lA_\tau} \rceil$ times;
5: Rank these channels according to their empirical rewards, let $\hat{\mu}_1 > \cdots > \hat{\mu}_{A_\tau}$;
6: Eliminate all the channels in $\mathcal{K}_\tau = \{j : \hat{\mu}_j \leq \hat{\mu}_{A_\tau - K_\tau}\}$, $\mathcal{A}_{\tau+1} := \mathcal{A}_\tau / \mathcal{K}_\tau$;
7: **end for**
8: **Output** : $\mathcal{A}_l$.

evenly into several rounds. Different from previous elimination-based algorithms, we allow users to eliminate multiple channels in each round, and they will drop fewer channels in the later rounds. The number of eliminated channels is chosen based on a learning rate η, which is a constant greater than 1, and users will eliminate $1/\eta$ of channels as in the previous round. This policy helps users to observe channels more frequently when it is hard to distinguish "good" channels from "bad" channels, since the reward gap between different channels becomes smaller as exploration time increases. While finding the optimal choice of η is difficult and may vary for different tasks, we try different settings of learning rate in our experiments. More details about the learning rate can be found in Section VII.

First, we divide the channel selection process into l rounds evenly. To ensure that $\sum_{\tau=1}^{l} \eta^{\tau-1} = K - M$, we set $l = \lceil \log_\eta ((\eta-1)(K-M)+1) \rceil$. In round τ, the user will remove $K_\tau = \lceil ((\eta-1)(K-M)+1)/\eta^\tau \rceil$ channels with the lowest empirical rewards from the remaining active channel set $\mathcal{A}_\tau$, and put them into set $\mathcal{K}_\tau$. It follows that

$$\sum_{\tau=1}^{l} K_\tau \geq \frac{\eta}{\eta-1}((\eta-1)(K-M)+1)(1-\eta^{-l}) = K - M. \tag{6}$$

Based on K_τ, we have $A_\tau = |\mathcal{A}_\tau|$ and $A_\tau = \left\lceil \frac{K-M}{\eta^\tau} + \frac{1}{(\eta-1)\cdot\eta^\tau} + M - \frac{1}{\eta-1} \right\rceil$, for all $\tau \leq l$. After l rounds, the user will provide the result of M chosen channels.

SME allows the user to reduce the number of samples constantly during the process of channel selection, and guarantees that each channel will be sufficiently sampled before being dropped. To calculate the error probability of SME, we introduce a lemma first.

Lemma 1: In SME, assume a channel p outside $\mathcal{M}$ is not eliminated before round τ. Then in round τ, for channel $j \in \mathcal{M}$, the probability of $\hat{\mu}_j < \hat{\mu}_p$ satisfies

$$\mathbb{P}[\hat{\mu}_j < \hat{\mu}_p] \leq \exp\left(-2\sum t_\tau (\Delta_j^M + \Delta_p^M)^2\right). \tag{7}$$

Proof: Let $\Delta_{jp} = \Delta_j^M + \Delta_p^M$ when $j \leq M$ and $p > M$. For $\alpha > 0$ and $\beta > 0$, by the Chernoff-Hoeffding inequality, we have

$$\mathbb{P}[\hat{\mu}_p > \mu_p + \alpha\Delta_{jp}] \leq \exp\left(-2\sum t_\tau (\alpha\Delta_{jp})^2\right)$$
$$\mathbb{P}[\hat{\mu}_j < \mu_j - \beta\Delta_{jp}] \leq \exp\left(-2\sum t_\tau (\beta\Delta_{jp})^2\right).$$

Since $\Delta_{jp} = \Delta_j^M + \Delta_p^M = \mu_j - \mu_p$, we have

$$\begin{aligned}\mathbb{P}[\hat{\mu}_j < \hat{\mu}_p] \leq & \exp\left(-2\sum t_\tau ((\alpha\Delta_{jp})^2 + (\beta\Delta_{jp})^2)\right) \\ \leq & \exp\left(-\sum t_\tau (\Delta_j^M + \Delta_p^M)^2\right),\end{aligned}$$

where the last inequality is due to the fact that if $\alpha + \beta \geq 1$, then $\alpha^2 + \beta^2 \geq \frac{1}{2}$. ■

From Lemma 1 we know that under the policy of SME, when the mean reward gap between good channels and bad channels Δ_{jp} is large, the probability for the user to identify the optimal channels becomes high. Furthermore, if the user spends more time on probing, it will achieve a lower probability of choosing bad channels. Then we can derive an upper bound for the error probability of SME.

Theorem 1: The error probability of SME is upper bounded by

$$\mathbb{P}[\mathcal{A}_l \neq \mathcal{M}] \leq \sum_{\tau=1}^{l} M(A_\tau - M)\exp\left(-\frac{\tau T}{lH_2}\right), \tag{8}$$

where $A_\tau = \left\lceil \frac{K-M}{\eta^\tau} + \frac{1}{(\eta-1)\cdot\eta^\tau} + M - \frac{1}{\eta-1} \right\rceil$ and $l = \lceil \log_\eta ((\eta-1)(K-M)+1) \rceil$.

Proof: The probability for user making wrong selections after l rounds is

$$\mathbb{P}[\mathcal{A}_l \neq \mathcal{M}] = \mathbb{P}[\mathcal{M} \cap \cup_{\tau=1}^{l} \mathcal{K}_\tau \neq \emptyset] = \sum_{\tau=1}^{l} \mathbb{P}[\mathcal{K}_\tau \cap \mathcal{M} \neq \emptyset]. \tag{9}$$

Assume that channel $j \in \mathcal{M}$ is not eliminated by SME in the first $\tau - 1$ rounds. Then in round τ, with a union bound, the probability for channel j of being eliminated satisfies

$$\begin{aligned}\mathbb{P}[\mathcal{K}_\tau \cap \mathcal{M} \neq \emptyset] &\leq \sum_{j\in\mathcal{M}} \sum_{A_\tau \leq p \leq K} \mathbb{P}[\hat{\mu}_j < \hat{\mu}_p] \\ &\leq \sum_{j\in\mathcal{M}} \sum_{A_\tau \leq p \leq K} \exp\left(-\sum_\tau t_\tau \Delta_{jp}^2\right) \\ &\leq \sum_{j\in\mathcal{M}} (A_\tau - M)\exp\left(-\sum_\tau t_\tau \Delta_{jA_\tau}^2\right) \\ &\leq M(A_\tau - M)\exp\left(-\frac{\tau T}{lH_2}\right). \end{aligned} \tag{10}$$

Then for all rounds, we have

$$\begin{aligned}\mathbb{P}[\mathcal{A}_l \neq \mathcal{M}] = & \sum_{\tau=1}^{l} \mathbb{P}[\mathcal{K}_\tau \cap \mathcal{M} \neq \emptyset] \\ \leq & \sum_{\tau=1}^{l} M(A_\tau - M)\exp\left(-\frac{\tau T}{lH_2}\right). \end{aligned} \tag{11}$$

■

According to Theorem 1, if we consider the extreme case where $l = 1$ (although l cannot be smaller than 2) and $M = 1$, then we have $\mathbb{P}[\mathcal{A}_l \neq \mathcal{M}] \leq (K-1)\exp(-T/H_2)$. From (11) we can see that large values of l increase τ in terms of $\exp(-\tau T/lH_2)$ and decrease A_τ in each round, but also increase the total number of rounds. Since l is determined by the learning rate η, there is a trade-off in choosing η and the optimal choice of η varies from application to application. More details can be found in Section VII where we compare the performance for different η in different settings via experiments.

Theorem 1 shows that the error probability of SME is $O(e^{-T})$ with respect to time budget T. So as the time budget grows, the error probability of SME decreases exponentially. Also, SME needs at most $O(H_2 \log K)$ times of observation with respect to the number of channels K to identify the optimal channels, and has a smaller factor than existing algorithms such as SAR [8].

V. Distributed Channel Selection

In this section, we examine the scenarios of multiple users probing multiple channels in the same area. Assume there are n users probing K channels simultaneously, while each channel's information is initially unknown to all users. Each user will determine a complete set of M chosen channels after a given time budget T.

In the channel selection period, an important issue is that when multiple users are probing the same channel, interference between them will hurt the results observed by each user. This type of collision should not be neglected in the design of a multiple user scheme. To completely avoid collisions among different users, users could either communicate with other users to avoid the same choice of channels, or stay idle for a while and yield to other users to prevent collisions from happening. Communication will cost energy and may hurt the information obtained by users; Concession will sacrifice channel sampling time and also affect the channel selection results. Although in most cases algorithms with communication have better performance in terms of selection accuracy, in some specific applications (such as military tasks) where communication can be very costly or dangerous, a yielding algorithm will be a better choice.

Based on our single user algorithm, we introduce three algorithms that require limited communications. The first two algorithms are collision-free and each user works independently; The third one exploits the cooperation of users to improve the channel selection accuracy.

First, we introduce two algorithms with no collisions. We assume that users can communicate with each other to avoid the collision. However, as discussed before, the communication cost will degrade the accuracy of the results gained by users. So it would be costly for users to keep on exchanging information with each other. During the learning process, each user has to make their own decision with limited help from other users. We propose two distributed exploration bandit algorithms in this section. In the proposed algorithms, communication cost for each user is taken into account.

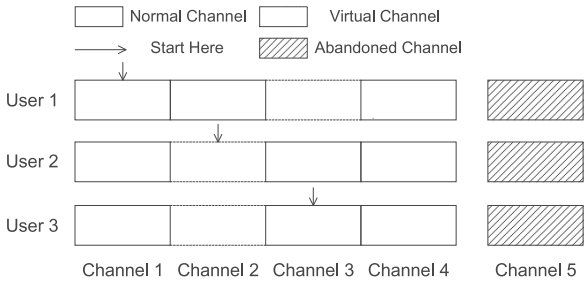

Fig. 1. Channel sampling strategy in Algorithm 2.

For the policy that allows concession, we propose an algorithm called Distributed Sequential Multiple Elimination with Virtual Channels (DSME-VC) presented in Algorithm 2. In DSME-VC, we evenly divide the time budget T into l rounds. The elimination process is the same as the SME. Next, the user will add its chosen channels into a channel set $\mathcal{V}$, which we call the *virtual channel* set. The user will stay on virtual channels but it will not collect any channel information. Meanwhile, the user will broadcast its chosen channels to all other users. If one channel is chosen by all the users, then the user will remove this channel from $\mathcal{V}$.

As illustrated in Fig. 1, with a round-robin fashion time allocation policy, if the user encounters a channel outside its virtual channel set, it will probe it as usual. When user 2 is assigned to probe a virtual channel 2, it won't collect any information about channel 2 and yield to other users. In other words, the user will spend time t_τ on a void channel. This strategy can completely avoid the potential collisions among users. For different users, they may choose different channels in the same round. So we calculate the probability for a channel of being chosen by all the users. Then we can compute the expectation of the remaining channels in each round.

In DSME-VC, we completely avoid the potential collisions among users by introducing the virtual channels. A user never really drops a channel unless it believes the channel is selected by all users simultaneously. Compared with the single user algorithm, DSME-VC wastes some time on virtual channels, which is an inevitable cost for avoiding collisions. At round τ, assume the communication cost for each user is c_τ. Since we divide T into l rounds, the total communication cost of l rounds is $C = \sum c_\tau$. Actually, the communication cost will hurt the results observed by users, and will affect the accuracy of channel selection results.

Step 6 of Algorithm 2 shows the round-robin fashion channel assignment policy. Channel $[(i + \tau + k) \bmod K - 1]$ is assigned to user i so that different users could aim at different channels in each round. If some user joins this probing activity halfway, existing users do not need to change their activities and the new user will also follow the channel assignment policy. However, the new user will not be able to have a channel selection result after this batch of probing. It should wait for the next batch and continue its probing until the number of rounds reaches l. Similarly, when some user leaves halfway, it will not affect other users' activity. Thus, this algorithm works well for the dynamic scenarios in real life.

Algorithm 2: Distributed Sequential Multiple Elimination with Virtual Channels (DSME-VC).

Input : K channels, M chosen channels, n users, time budget T, learning rate η.
Initialization : Let $l = \lceil \log_\eta ((\eta-1)(K-M)+1) \rceil$, $A_\tau = |\mathcal{A}_\tau|$, $\mathcal{A}_0 = \mathcal{K}$, $K_\tau = \left\lceil \frac{(\eta-1)(K-M)+1}{\eta^\tau} \right\rceil$, $k = 0$, $\mathcal{V} = \varnothing$.
for each $\tau = 1, 2, \ldots, l$ **do**
  for each user $i \in \mathcal{N}$ **do**
    while $k \leq K$ **do**
      if the channel $[(i+\tau+k) \bmod K - 1]$ belongs to $\mathcal{A}_\tau$ **then**
        Sample it for $t_\tau = \frac{T}{l|\mathcal{A}_\tau \bigcup \mathcal{V}|}$ times;
      end if
      $k := k + 1$;
    end while
    Let $k := 0$, rank channels according to their empirical rewards where $\hat{\mu}_1 > \cdots > \hat{\mu}_{A_\tau}$;
    Choose all the channels in $\mathcal{K}_\tau = \{j : \hat{\mu}_j \leq \hat{\mu}_{A_\tau - K_\tau}\}$, $\mathcal{A}_{\tau+1} := \mathcal{A}_\tau / \mathcal{K}_\tau$. Broadcast chosen channels to other users;
    Eliminate channels chosen by all users, and add others back to $\mathcal{V}$ as virtual channels;
  end for
end for
Output : $\mathcal{A}_l$.

To calculate the communication cost of DSME-VC, we first define the complexity of communication cost as

$$c_0 = \max_{i \in \mathcal{N}, j \in \mathcal{K}} \{c_{i,j}\}, \tag{12}$$

where $c_{i,j}$ is the communication cost for user i of broadcasting information about channel j to other users and c_0 here refers to the maximum communication cost for a single user and single channel.

Consider n users probing multiple channels simultaneously. Then user i's communication cost in round τ is $c_{i,\tau} = n \sum_{j=1}^{K_\tau} c_{i,j}$. Thus, the total communication cost for user i is upper bounded as

$$C_i = \sum_{\tau=1}^{l} c_{i,\tau} = n \sum_{\tau=1}^{l} \sum_{j=1}^{K_\tau} c_{i,\tau} \leq n \sum_{\tau=1}^{l} K_\tau c_0 = n(K-M)c_0. \tag{13}$$

In the following, we derive an upper bound for the probability for any channel j to be selected in round τ. Then we prove the expected number of channels chosen by all users.

Lemma 2: In DSME-VC, The probability for channel j where j is smaller than K_τ being chosen by one user in round τ satisfies

$$\mathbb{P}[j \in \mathcal{A}_{i,\tau}] \leq K_{\tau-1} \exp(-\Delta_{\min}^2 T_{\tau-1}) - K_\tau \exp(-\Delta_{\min}^2 T_\tau). \tag{14}$$

Proof: Based on Lemma 1 and a union bound, we have

$$\begin{aligned}
\mathbb{P}[j \in \mathcal{A}_{i,\tau}] &\leq \mathbb{P}[\bigcup\nolimits_{A_\tau < k \leq A_{\tau-1}} (\hat{\mu}_j < \hat{\mu}_k)] \\
&\leq \sum_{k > A_\tau}^{A_{\tau-1}} \mathbb{P}[\hat{\mu}_j < \hat{\mu}_k] \\
&\leq \sum_{k > A_\tau}^{A_{\tau-1}} \exp(-(\Delta_{jk})^2 t_\tau) \\
&\leq (A_{\tau-1} - A_\tau) \exp\left(-(\Delta_{jA_\tau})^2 t_\tau\right).
\end{aligned}$$

For any $j < K_\tau$, we have

$$\begin{aligned}
&\mathbb{P}[j < K_\tau | j \in \mathcal{A}_{i,\tau}] \\
&\leq (A_{\tau-1} - A_\tau) \exp(-(\Delta_{jA_\tau})^2 t_\tau) \\
&\leq (A_{\tau-1} - A_\tau) \exp(-(\Delta_{A_{\tau-1}} - \Delta_{A_\tau})^2 t_\tau) \\
&\leq (A_{\tau-1} - A_\tau) \exp(-\Delta_{\min}^2 t_\tau) \\
&\leq A_{\tau-1} \exp(-\Delta_{\min}^2 t_{\tau-1}) - A_\tau \exp(-\Delta_{\min}^2 t_\tau).
\end{aligned} \tag{15}$$

■

With Lemma 2, we can now prove the expected number of channels chosen by all users.

Theorem 2: In DSME-VC, the expectation of number of channels chosen by all users in the round τ satisfies

$$\begin{aligned}
&\mathbb{E}\left[\text{\# of channels chosen by all users}\right] \\
&\geq (A_{\tau-1} - A_\tau) \left(\frac{1}{A_{\tau-1} - A_\tau} - A_\tau \exp\left(-\Delta_{\min}^2 \frac{T}{Kl} \right) \right)^n \\
&\triangleq N_\tau.
\end{aligned} \tag{16}$$

Proof: For a channel inside $\mathcal{K}_\tau$, the probability of being chosen by one user is at least $1 - \sum_{j=1}^{A_\tau} \mathbb{P}[j \in \mathcal{A}_{i,\tau}]$. Then we have

$$\begin{aligned}
&\mathbb{E}\left[\text{\# of channels chosen by all users}\right] \\
&\geq (A_{\tau-1} - A_\tau) \left(\frac{1 - A_\tau \mathbb{P}[j \in \mathcal{A}_{i,\tau}]}{A_{\tau-1} - A_\tau} \right)^n \\
&\geq (A_{\tau-1} - A_\tau) \left(\frac{1}{A_{\tau-1} - A_\tau} - A_\tau \exp(-\Delta_{\min}^2 t_\tau) \right)^n,
\end{aligned} \tag{17}$$

where (17) follows from the last inequality in (15). Since $t_\tau = T/l(K - \sum_{r=1}^{\tau} N_r) \leq T/Kl$, we have Theorem 2. ■

In Theorem 2, the probability for all users choosing the same channel grows exponentially as the increase of the number of users n. This fact guarantees that when there are many users in the same area, DSME-VC will not have too many virtual channels. Next, we derive an upper bound on each user's error probability for Algorithm 2.

Theorem 3: The error probability of DSME-VC for each user satisfies

$$\mathbb{P}[\mathcal{A}_l \neq \mathcal{M}] \leq \sum_{\tau=1}^{l} M(A_\tau - M) \times \exp\left(-\frac{\tau T M}{l H_2(K - N_1' + N_\tau')}\right), \quad (18)$$

where N_τ' is $A_\tau \left(1/(A_{\tau-1} - A_\tau) - A_\tau \exp(-\Delta_{\min}^2 T/Kl)\right)^n$.

Proof: First, we modify the time allocation policy used in Algorithm 4 as

$$T_\tau = \sum_{r=1}^{\tau} t_r \leq \sum_{r=1}^{\tau} \frac{T}{l\left(K - \sum_{r \leq \tau} N_r\right)}, \quad (19)$$

from Theorem 2 we have that

$$\sum_{r \leq \tau} N_r \geq K\left(\frac{4}{3(K-M)+1} - A_1 \exp\left(-\Delta_{\min}^2 \frac{T}{Kl}\right)\right)^n - A_\tau \left(\frac{1}{A_{\tau-1} - A_\tau} - A_\tau \exp\left(-\Delta_{\min}^2 \frac{T}{Kl}\right)\right)^n.$$

With (10) in Theorem 1, the error probability in round τ is upper bounded as

$$\mathbb{P}[\mathcal{K}_\tau \cap \mathcal{M} \neq \emptyset] \leq M(A_\tau - M) \exp\left(-\frac{\tau T A_\tau}{l H_2(K - \sum_{r \leq \tau} N_r)}\right) \leq M(A_\tau - M) \exp\left(-\frac{\tau T M}{l H_2(K - (N_1' - N_\tau'))}\right). \quad (20)$$

Then, the total error probability of Algorithm 2 is

$$\mathbb{P}[\mathcal{A}_l \neq \mathcal{M}] = \sum_{\tau=1}^{l} \mathbb{P}[\mathcal{K}_\tau \cap \mathcal{M} \neq \emptyset] \leq \sum_{\tau=1}^{l} M(A_\tau - M) \exp\left(-\frac{\tau T M}{l H_2(K - N_1' + N_\tau')}\right).$$

■

Compared with Theorem 1, Theorem 3 shows that the upper bound on the error probability of DSME-VC is bigger than that of SME. However, it completely eliminates the collisions among users and when K is not very large, the error probability of DSME-VC is close to that of SME. So the DSME-VC algorithm is applicable for distributed channel selection when the number of users is relatively small.

We also propose a distributed algorithm without using virtual channels in Algorithm 3, named Distributed Auction-based Channel Assignment (DACA). DSME-VC solves the potential collision problem and only needs very little communications, however, when K becomes very large, DSME-VC may waste too much time on virtual channels. So Algorithm 2 is not efficient enough when the channel pool is very large. DACA solves this problem without using the virtual channels.

The basic idea of Algorithm 3 is based on an auction process among different users. Assume there is an undirected bipartite graph $\mathcal{G}(\mathcal{S}, \mathcal{U}, \mathcal{E})$, where $\mathcal{S}$ and $\mathcal{U}$ are the set of users and channels, respectively. $\mathcal{E}$ stands for the connection between users and channels. When user i eliminates channel j, the edge $E(i,j)$ will also be removed from set $\mathcal{E}$. So the user will only provide a price to the channel in its active set and each user's price is set to be the channel's empirical rewards observed by the user. This setting is intuitively reasonable since users prefer to choose channels that seem "good" to them. When communicating with others, the user will decide whether to probe the channel or not. If user i is not the highest bidder for channel j, it will choose another channel randomly.

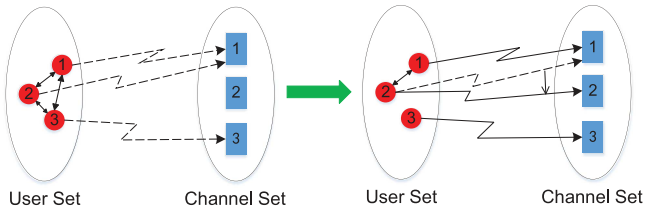

Fig. 2. The negotiation process between users in Algorithm 3.

Algorithm 3: Distributed Auction-based Channel Assignment Algorithm (DACA).

Input : K channels, M chosen channels, n users, time budget T, learning rate η.
Initialization : Let $l = \lceil \log_\eta ((\eta-1)(K-M)+1) \rceil$, $A_\tau = |\mathcal{A}_\tau|$, $\mathcal{A}_0 = \mathcal{K}$, $K_\tau = \left\lceil \frac{(\eta-1)(K-M)+1}{\eta^\tau} \right\rceil$. For a random channel j, user i provides a price p_{ij} randomly, broadcast p_{ij} to other users.
for each $\tau = 1, 2, \ldots, l$ **do**
  while $k < A_\tau$ **do**
    For user i, let $j = \arg\max_{j \in \mathcal{S}_{k,\tau}} \hat{\mu}_{i,j}$;
    if $i = \arg\max_{i \in \mathcal{N}} p_{ij}$ **then**
      User i samples channel j for $t_\tau = \lceil \frac{T}{lK_\tau} \rceil$ times, broadcast to all other users that it has finished this round, then waits for other users;
    else
      Move to another channel $s \in \mathcal{S}_{k,\tau}$ randomly. Let $p_{is} = \hat{\mu}_{i,s}$, communicate with other remaining users, and go back to step 6;
    end if
    Let $\mathcal{S}_{k,\tau} := \mathcal{S}_{k,\tau}/j, k := k+1$;
  end while
  Update the empirical reward $\hat{\mu}_{i,j}$ for all $j \in \mathcal{A}_\tau$. Eliminate channels in $\mathcal{K}_\tau = \{j : \hat{\mu}_j \leq \hat{\mu}_{A_\tau - K_\tau}\}$, $\mathcal{A}_{\tau+1} := \mathcal{A}_\tau/\mathcal{K}_\tau, k := 1, \mathcal{S}_{k,\tau} := \mathcal{A}_\tau$;
end for
Output : $\mathcal{A}_l$.

For example, in Fig. 2, there are three users probing three channels. In the first communication round, both users 1 and 2 bid for channel 1 and user 3 bids for channel 3. After communication with each other, users 1 and 3 finds them to be the highest bidder for channel 1 and 3, respectively. In the second round, user 2 bids for channel 2. After communicating with user 1, user

Algorithm 4: Collaborative Sequential Multiple Elimination for Channel Selection (CSME).

Input : $i \in \mathcal{N}$, channel set $\mathcal{K}^i$, K channels, M chosen channels, n users, time budget T, learning rate η.
Initialization : Let $l = \lceil \log_\eta ((\eta-1)(K-M)+1) \rceil$, $A_\tau^i = |\mathcal{A}_\tau^i|$, $\mathcal{A}_0^i = \mathcal{K}^i$, $A_\tau = |\mathcal{A}_\tau|$, $\mathcal{A}_0 = \mathcal{K}$, $K_\tau = \left\lceil \frac{(\eta-1)(K-M)+1}{\eta^\tau} \right\rceil$, $k = 0$.
for each $\tau = 1, 2, \ldots, l$ **do**
  for each user $i \in N$ **do**
    while $k \le A_\tau$ **do**
      if channel $[(i+\tau+k) \bmod A_\tau - 1]$ belongs to $\mathcal{A}_\tau^i$ **then**
        Sample it for $t_\tau = \frac{T}{l|\mathcal{A}_\tau^i|}$ times;
      end if
      $k := k + 1$;
    end while
    Let $k := 0$, and broadcast results to all other users;
    Update the empirical reward $\hat{\mu}_j$ for all $j \in \mathcal{A}_\tau$. Eliminate channels in $\mathcal{K}_\tau = \{j : \hat{\mu}_j \le \hat{\mu}_{A_\tau - K_\tau}\}$, $\mathcal{A}_{\tau+1} := \mathcal{A}_\tau / \mathcal{K}_\tau$, $\mathcal{A}_{\tau+1}^i := \mathcal{A}_\tau^i / (\mathcal{K}_\tau \cap \mathcal{A}_\tau^i)$;
  end for
end for
Output : $\mathcal{A}_l$.

2 starts to collect information of channel 2. This process lasts until every user finds a channel. After a channel is observed by some user, it will be removed from set $\mathcal{E}$ temporarily until the next round.

Since the channel selection process of DACA is the same as SME, the error probability of Algorithm 3 will be the same as Algorithm 4 and users could spend more time on the channels than in DSME-VC. Thus, the performance of DACA will be better than DSME-VC. However, to better evaluate DACA, we should also take communication cost into consideration.

Then we will derive the communication cost of DACA. First, we introduce a lemma to bound the number of communications in DACA.

Lemma 3: The communication cost for each user in DACA satisfies

$$C_i \le \left(K - M + \left(M - \frac{1}{\eta - 1}\right) l\right) \frac{n^3 - n}{6} c_0.$$

Proof: Consider the worst case of Algorithm 3. When user i communicates with other users in the negotiation phase of the τth round, if it always fails to provide the highest price, it has to keep on communicating with all the remaining users. Then the number of communications is at most

$$((n-1)(n-2)\cdots 1) A_\tau = \frac{n^3 - n}{6} A_\tau. \tag{21}$$

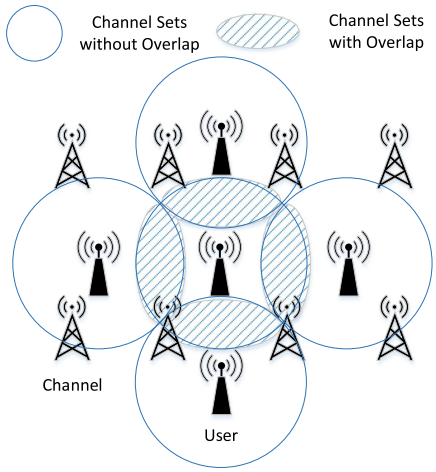

Fig. 3. Channel assignment model of CSME.

The total communication cost satisfies

$$C_i = \sum_{\tau=1}^{l} c_{i,\tau} \le \sum_{\tau=1}^{l} \frac{n^3 - n}{6} A_\tau c_0 = \left(K - M + \left(M - \frac{1}{\eta - 1}\right) l\right) \frac{n^3 - n}{6} c_0$$

■

With Lemma 3, we obtain an upper bound for the communication cost of DACA. Lemma 3 also shows that when the number of users becomes very large, the DACA algorithm has a higher communication cost than DSME-VC. The numerical results on the communication cost of DACA and DSME-VC can be found in Section VII.

As mentioned before, the error probability of DACA is the same as SME in Theorem 1. Compared with Theorem 3, DACA has a lower error probability bound than DSME-VC. The reason is that DACA does not use virtual channels, so the user can sample each channel with longer time than DSME-VC, and thus it generates more accurate results. Meanwhile, DACA requires much more communications than DSME-VC. When the communication cost is high, the low communication cost of DSME-VC might outweigh the better accuracy of DACA. This is a choice of horses for courses.

After the discussion of different users probing the same amount channels, we then consider improving the accuracy and efficiency of channel selection by allowing user collaboration. Compared to multi-user channel selection algorithms without collaboration, if different users could focus on different sets of channels and broadcast their channel probing results to other users, every user will have an identical channel selection result after the channel probing period with high accuracy. Assume each user is only in charge of probing a portion of K channel; the channel set for user i is $\mathcal{K}^i$, and its cardinality is denoted by K^i. As illustrated in Fig. 3, each user has its own channel set to be probed but every channel will be probed by at least one user. There are potential overlapping areas among different

users' channel sets. The channel assignment policy guarantees that no channel will be missed in the channel probing process.

In our proposed algorithm, termed Collaborative Sequential Multiple Elimination (CSME), users will follow almost the same round-robin fashion channel assignment policy as in DSME-VC in case there are some overlap among different users' channel sets. The only difference is that there are no virtual channels during the channel probing. After one probing round, each user will broadcast its probing results to all other users and update all remaining channel's empirical rewards accordingly. In round τ, let $\hat{\mu}_j^i$ represent user i's probing result for channel j and t_τ^i be the time spent on each channel, we have

$$\hat{\mu}_j := \frac{\sum_{k=1}^{\tau}\sum_{i=1}^{N}\hat{\mu}_j^i t_k^i}{\sum_{k=1}^{\tau}\sum_{i=1}^{N} t_k^i}. \tag{22}$$

The elimination process is based on all users' probing results and the user will only eliminate channels in its own channel set. Let's consider the general case for CSME. Since the channel assignment policy avoids the collision, in round τ, each channel will be sampled by at least $T/(\max_i K^i l)$ times. Combined with (10) and (11), we have that

$$\begin{aligned}\mathbb{P}[\mathcal{A}_l \neq \mathcal{M}] &= \sum_{\tau=1}^{l}\mathbb{P}[\mathcal{K}_\tau \cap \mathcal{M} \neq \emptyset] \\ &\leq \max_i \sum_{\tau=1}^{l} M(A_\tau - M)\exp\left(-\frac{\tau T}{lK^i}\Delta_{\min}^2\right).\end{aligned} \tag{23}$$

In the best case, if there are no overlapping between different users' channel sets, we have $\sum_i K^i = K$. In this case, each user only needs to communicate with others once in each probing round for broadcasting channel probing results. Thus, the communication cost will be the same as DSME-VC.

The bound (23) indicates that if all channels are allocated evenly to all users and there are no overlapping among different channel sets, i.e., $K^i = K/n$ for any i, we have

$$\mathbb{P}[\mathcal{A}_l \neq \mathcal{M}] \leq \sum_{\tau=1}^{l} M(A_\tau - M)\exp\left(-\frac{n\tau T}{lK}\Delta_{\min}^2\right). \tag{24}$$

Now we consider the worst case for CSME. If some user i probes all K channels and some channels are only probed by user i, then we have

$$\mathbb{P}[\mathcal{A}_l \neq \mathcal{M}] \leq \sum_{\tau=1}^{l} M(A_\tau - M)\exp\left(-\frac{\tau T}{lK}\Delta_{\min}^2\right), \tag{25}$$

which is the error probability bound for CSME.

If every user probes all K channels, which means $\forall i \in \mathcal{N}$: $K_i = K$, then in each probing round, users do one more broadcasting than in DSME-VC. We can show that the communication cost of CSME satisfies

$$C_i = \sum_{\tau=1}^{l} c_{i,\tau} \leq 2n(K-M)c_0. \tag{26}$$

The number of communications in CSME is between the number in DSME-VC and DACA. In the following section, we consider the channel access after users get their channel selection results.

VI. Channel Access after Channel Selection

Although we mainly focus on channel selection for mission-critical communications in this paper, we need to consider channel access after the channel probing period for the multi-user scenario. With our proposed channel selection algorithms, each user will have a result of M best channels after the probing time T. For mission-critical communications, how to coordinate different users to access the channels and avoid collisions among them remains to be unsolved.

First, we consider the case where each user has the same channel selection results (such as in CSME). We adopt a round-robin fashion channel allocation policy to let users loop through these M channels. Assume each user only needs one channel for communication. Let t denotes the total communication time and we evenly divide it into R time slots. In time slot r, user i is allocated with channel $[(i+r) \mod M]$. This policy guarantees that no collision will happen in the channel access period and the user does not need to communicate with other users.

If different users have different channel selection results after the channel probing period, extra steps are taken to make sure no users will access the same channel at the same time. We divide t into R time slots and user i is allocated with channel $[(i+r) \mod M]$ initially, but it needs to communicate with other users in case they choose the same channel. Same as the negotiation phase in DCAC, the user provides a price for its target channel according to the empirical rewards from the channel probing period. If the user is not the highest bidder of that channel, it will choose another channel which has not been chosen by other users to bid.

Since each user needs to communicate with others once in each time slot, the communication cost in the channel access period is bounded as

$$C_i = \sum_{r=1}^{R} c_{i,r} \leq nRc_0. \tag{27}$$

Note that to completely avoid collisions, the users have to communicate with each other. If users want to avoid communications with other users in channel access period, they should adopt algorithms such as CSME which requires extra communications during the channel selection period; If users prefer less or no communication costs during channel probing and selection, they have to except additional communication cost in the channel access.

VII. Simulation Results

In this section, we present the simulation results on the proposed channel selection algorithms. We use simple regret of channel selection results (defined in Section III) to compare performance between different algorithms. First, we compare our single user algorithm with the SAR algorithm in [8] and discuss the different choices of learning rate. Then we illustrate the performance of the proposed distributed al-

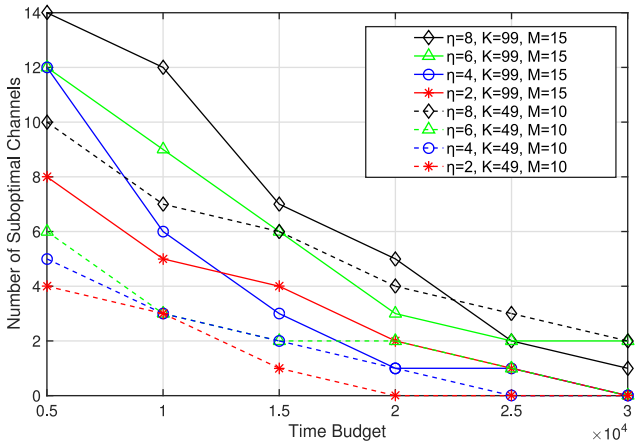

Fig. 4. Error probability performances comparison for different learning rates.

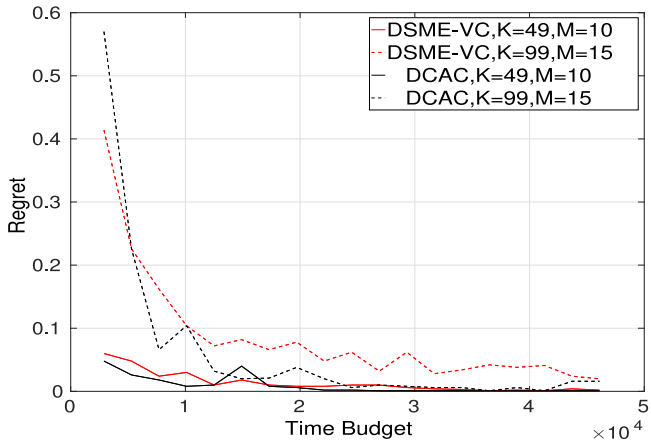

Fig. 6. Regret performances comparison for multiple users with communications.

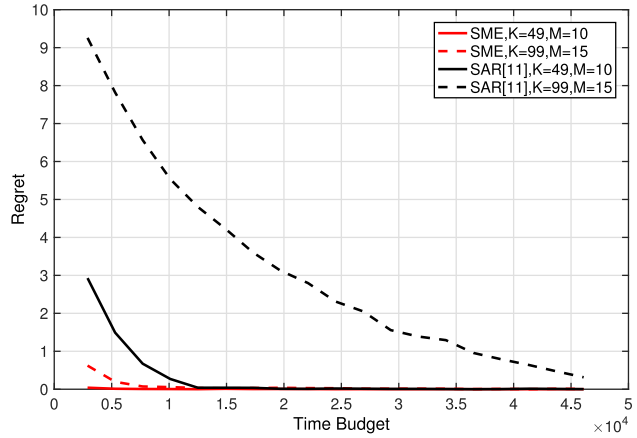

Fig. 5. Regret performances comparison for single user.

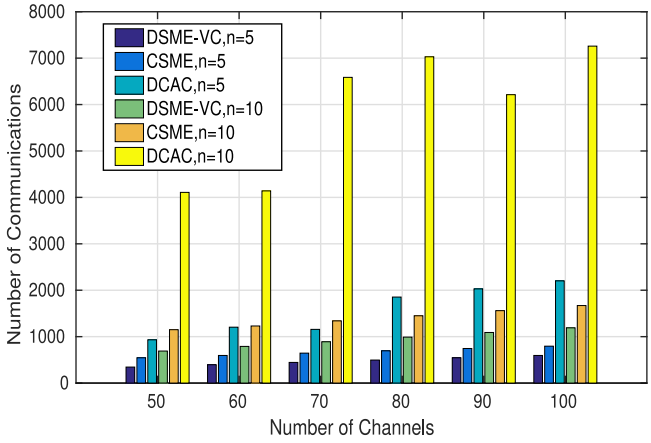

Fig. 7. Expected communication cost for each user.

gorithms and compare them with the distributed UCB algorithm proposed in [3]. We consider a few different setups where number of channels and users varies. With out loss of generality, we assume that each channel's reward is associated with an i.i.d. Bernoulli distribution as in [4], [27]. When $K = 49, M = 10$ and $K = 99, M = 15$, the parameters of each distribution are $\Theta = (0.02, 0.04, 0.06, \ldots, 0.98)$ and $\Theta = (0.01, 0.02, 0.03, \ldots, 0.99)$, respectively. All simulation results are averaged over 50 runs.

In Fig. 4, we compare error probability with different settings of learning rate η in SME, where the time budget is the number of time slots. Note that a less number of suboptimal channels indicates a better performance in channel selection. In both scenarios for $K = 99$ and $K = 49$, $\eta = 4$ and $\eta = 2$ have better performance than $\eta = 6$ and $\eta = 8$, and they have similar performance. When $K = 49$, $\eta = 2$ has slightly lower error probability than $\eta = 4$. However, we can see the trend that $\eta = 4$ is getting better when K becomes large. Although the optimal choice of η varies for different applications, for simplicity, we choose $\eta = 4$ in our simulations .

In Fig. 5, SME has much smaller regret when compared with SAR, especially when K is large. When $K = 99$, compared with SAR, SME improves the accuracy and efficiency significantly. Even when $K = 49$, the proposed SME algorithm is ten times better than SAR.

For channel selection with communications, from Fig. 6, we can see that the DACA algorithm has lower regret bound than DSME-VC in both scenarios. When $K = 99$, compared with DSME-VC, DACA improves the accuracy of channel selection for more than 30%. Although DACA has higher accuracy, we need to consider its high communication cost. We compare the communication cost for three algorithms in Fig. 7.

In Fig. 7, we set $K = 99$, $M = 15$, $T = 5000$ and randomly assign $K/2$ channels to each user for CSME. One can see that when the number of users is 5, the difference between three algorithms is not very large, which indicates that the communication cost of DACA is acceptable for relatively small number of users. However, when n equals to 10, compared with DSME-VC and CSME, the communication frequency of DACA becomes high, which also demonstrates the advantage of DSME-VC/CSME with massive users.

We also compare the performance of CSME with a classic distributed UCB algorithm with the TDFS policy proposed in [3] in Fig. 8. For CSME, we set the number of users to be 10 and randomly assign $K/2$ channels to each user. We observe that CSME improves the channel selection performance significantly

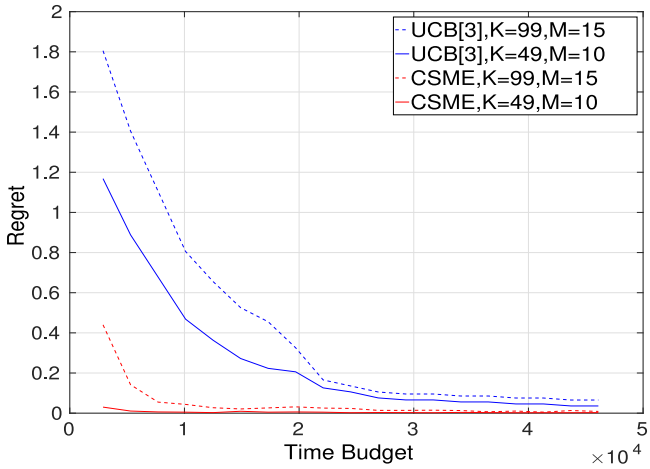

Fig. 8. Regret performances comparison for multiple users without communications.

TABLE II
SUMMARY OF THE PROPOSED ALGORITHMS

Algorithm	*Distributed?*	*Communications?*	*Collisions?*
SME	×	×	×
DSME-VC	✓	low rate	×
DACA	✓	high rate	×
CSME	✓	medium rate	×

compared to conventional UCB, especially when time budget is small. Although the performance of CSME depends heavily on the channel assignment result, this simulation result shows that collaboration indeed helps users make the better selection.

The numerical results suggest that as the time budget increases, error probabilities for all algorithms decrease exponentially, which is completely in conformity with our theoretical analysis. They beat both classic MAB algorithm and previous exploration bandit algorithm in different cases. Moreover, each algorithm has its own advantages in specific scenarios. In summary, simulation results prove the advantages of the proposed algorithms in the channel selection for mission-critical communications.

VIII. CONCLUSION

In this paper, we studied the problem of channel selection for mission-critical communications. We considered both cases of a single user and multiple users with a pure-exploration bandit problem formulation. As illustrated in Table II, a few single or distributed channel selection algorithms were proposed for different settings. By applying the proposed channel selection algorithms, users could select a set of good channels via a short period of channel probing, which guarantees the ultra-reliable and low-latency communication in emergency circumstances. The performance of the proposed algorithms was analyzed, simulations were conducted and the results illustrated the performance of the proposed algorithms for the multiple channel selection. Both theoretical analysis and simulation results showed that the well-designed algorithms proposed in this paper have impressive performances for channel selection in different scenarios. Moreover, the proposed pure-exploration bandits algorithms are not limited to channel selection for emergency communications. The proposed schemes are quite general and can apply to general wireless communications scenarios (e.g., cognitive radio networks), where user QoS/QoE is considered.

REFERENCES

[1] M. Ulema, A. Kaplan, K. Lu, N. Amogh, and B. Kozbe, "Critical communications and public safety networks part 1: Standards, spectrum policy, and economics," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 12–13, Mar. 2016.
[2] DARPA. (2016). Spectrum collaboration challenge. [Online]. Available: https://spectrumcollaborationchallenge.com/
[3] K. Liu and Q. Zhao, "Distributed learning in cognitive radio networks: Multi-armed bandit with distributed multiple players," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2010, pp. 3010–3013.
[4] Y. Gai and B. Krishnamachari, "Distributed stochastic online learning policies for opportunistic spectrum access," *IEEE Trans. Signal Process.*, vol. 62, no. 23, pp. 6184–6193, Dec. 2014.
[5] J.-Y. Audibert and S. Bubeck, "Best arm identification in multi-armed bandits," in *Proc. 23th Conf. Learn. Theory*, 2010, pp. 1–13.
[6] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Elsevier Adv. Appl. Math.*, vol. 6, no. 1, pp. 4–22, 1985.
[7] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, nos. 2/3, pp. 235–256, May 2002.
[8] S. Bubeck, T. Wang, and N. Viswanathan, "Multiple identifications in multi-armed bandits," in *Proc. 30th Int. Conf. Mach. Learn.*, Atlanta, GA, USA, Jun. 2013, pp. 258–265.
[9] D. B. Rawat, M. Song, and S. Shetty, *Dynamic Spectrum Access for Wireless Networks*. New York, NY, USA: Springer, 2015.
[10] A. Anandkumar, N. Michael, and A. Tang, "Opportunistic spectrum access with multiple users: Learning under competition," in *Proc. IEEE INFOCOM*, 2010, pp. 1–9.
[11] D. B. Rawat, C. Bajracharya, and S. Grant, "nroar: Near real-time opportunistic spectrum access and management in cloud-based database-driven cognitive radio networks," *IEEE Trans. Netw. Serv. Manage.*, vol. 14, no. 3, pp. 745–755, Sep. 2017.
[12] T. Doumi *et al.*, "LTE for public safety networks," *IEEE Commun. Mag.*, vol. 51, no. 2, pp. 106–112, Feb. 2013.
[13] S. D. Jones *et al.*, "Characterization of spectrum activities in the us public safety band for opportunistic spectrum access," in *Proc. 2nd IEEE Int. Symp. New Frontiers Dyn. Spectrum Access Netw.*, 2007, pp. 137–146.
[14] Q. Wang and T. X. Brown, "Public safety and commercial spectrum sharing via network pricing and admission control," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 622–632, Apr. 2007.
[15] M. M. Sohul, M. Yao, X. Ma, E. Y. Imana, V. Marojevic, and J. H. Reed, "Next generation public safety networks: A spectrum sharing approach," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 30–36, Mar. 2016.
[16] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Phys. Commun.*, vol. 4, no. 1, pp. 40–62, 2011.
[17] Y. Zou, Y.-D. Yao, and B. Zheng, "Cooperative relay techniques for cognitive radio systems: Spectrum sensing and secondary user transmissions," *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 98–103, Apr. 2012.
[18] S. Wang, Z.-H. Zhou, M. Ge, and C. Wang, "Resource allocation for heterogeneous cognitive radio networks with imperfect spectrum sensing," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 464–475, Mar. 2013.
[19] S.-S. Tan, D. Zheng, J. Zhang, and J. Zeidler, "Distributed opportunistic scheduling for ad-hoc communications under delay constraints," in *Proc. INFOCOM*, 2010, pp. 1–9.
[20] E. Even-Dar, S. Mannor, and Y. Mansour, "Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems," *J. Mach. Learn. Res.*, vol. 7, pp. 1079–1105, Dec. 2006.
[21] S. Kalyanakrishnan, A. Tewari, P. Auer, and P. Stone, "Pac subset selection in stochastic multi-armed bandits," in *Proc. 29th Int. Conf. Mach. Learn.*, Edinburgh, Scotland, Jun. 2012, pp. 655–662.
[22] S. Bubeck, R. Munos, and G. Stoltz, "Pure exploration in multi-armed bandits problems," in *Algorithmic Learning Theory*. New York, NY, USA: Springer, 2009, pp. 23–37.

[23] S. Kalyanakrishnan and P. Stone, "Efficient selection of multiple bandit arms: Theory and practice," in *Proc. 27th Int. Conf. Mach. Learn.*, Haifa, Israel, Jun. 2010, pp. 511–518.
[24] S. Chen, T. Lin, I. King, M. R. Lyu, and W. Chen, "Combinatorial pure exploration of multi-armed bandits," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 379–387.
[25] S. Shahrampour, M. Noshad, and V. Tarokh, "On sequential elimination algorithms for best-arm identification in multi-armed bandits," *IEEE Trans. Signal Process.*, vol. 65, no. 16, pp. 4281–4292, Aug. 2017.
[26] C. Tekin and M. Liu, "Online learning of rested and restless bandits," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5588–5611, Aug. 2012.
[27] D. Kalathil, N. Nayyar, and R. Jain, "Decentralized learning for multiplayer multiarmed bandits," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2331–2345, Apr. 2014.
[28] E. Hillel, Z. S. Karnin, T. Koren, R. Lempel, and O. Somekh, "Distributed exploration in multi-armed bandits," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 854–862.
[29] Y. Xue, P. Zhou, T. Jiang, S. Mao, and X. Huang, "Distributed learning for multi-channel selection in wireless network monitoring," in *Proc. 13th Annu. IEEE Int. Conf. Sens., Commun., Netw.*, 2016, pp. 1–9.

Yuan Xue (S'15) received the B.S. degree from the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, Hubei, China in 2015. He is currently working toward the Ph.D. degree in computer science with Lehigh University, Bethlehem, PA, USA. His current research interests include computer vision and machine learning.

Pan Zhou (S'07–M'14) received the B.S. degree in the *advanced class* of Huazhong University of Science and Technology (HUST), Wuhan, China, and the M.S. degree from the Department of Electronics and Information Engineering, HUST, in 2006 and 2008, respectively. He received the Ph.D. degree from the School of Electrical and Computer Engineering, Georgia Institute of Technology (Georgia Tech), Atlanta, GA, USA, in 2011. He is currently an Associate Professor with the School of Electronic Information and Communications, HUST. He held honorary degree in his bachelor and merit research award of HUST in his master study. He was a senior technical member at Oracle Inc, America during 2011 to 2013, and worked on Hadoop and distributed storage system for big data analytics at Oralce Cloud Platform. His current research interests include big data analytics and machine learning, security and privacy, and information networks.

Shiwen Mao (S'99–M'04–SM'09) received the Ph.D. degree in electrical and computer engineering from Polytechnic University, Brooklyn, NY, USA, in 2004. He is the Samuel Ginn Distinguished Professor, and Director of the Wireless Engineering Research and Education Center, Auburn University, Auburn, AL, USA. His research interests include wireless networks, IoT, and Smart Grid. He is a Distinguished Speaker of the IEEE Vehicular Technology Society. He is on the Editorial Board of IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE INTERNET OF THINGS JOURNAL, IEEE MULTIMEDIA, *ACM GetMobile*, among others.

Dapeng Wu (S'98–M'04–SM'06–F'13) received the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2003. He is a Professor with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA. His research interests include the areas of networking, communications, signal processing, computer vision, machine learning, smart grid, and information and network security.

Yingjie Zhou (M'14) received the Ph.D. degree from the School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2013. He is currently an Assistant Professor with the College of Computer Science, Sichuan University, Chengdu, China. He was a visiting scholar with the Department of Electrical Engineering, Columbia University, New York, NY, USA. His current research interests include network measurement, behavioral data analysis, resource allocation, and neural networks.