Computer Networks 143 (2018) 176–190

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

SpecWatch: A framework for adversarial spectrum monitoring with unknown statistics

Ming Li [a], Dejun Yang [a,*], Jian Lin [a], Ming Li [b], Jian Tang [c]

[a] Colorado School of Mines, Golden, CO 80401, United States
[b] University of Arizona, Tucson, AZ 85721, United States
[c] Syracuse University, Syracuse, NY 13210, United States

ARTICLE INFO

Article history:
Received 9 December 2017
Revised 31 May 2018
Accepted 9 July 2018
Available online 11 July 2018

Keywords:
Cognitive radio networks
Spectrum monitoring
Multi-armed bandit problem

ABSTRACT

In cognitive radio networks (CRNs), dynamic spectrum access has been proposed to improve the spectrum utilization, but it also generates spectrum misuse problems. One common solution to these problems is to deploy monitors to detect misbehaviors on certain channel. However, in multi-channel CRNs, it is very costly to deploy monitors on every channel. With a limited number of monitors, we have to decide which channels to monitor. In addition, we need to determine how long to monitor each channel and in which order to monitor, because switching channels incurs costs. Moreover, the information about the misuse behavior is not available a priori. To answer those questions, we model the spectrum monitoring problem as a combinatorial adversarial multi-armed bandit problem with switching costs (MAB-SC), propose an effective framework, and design two online algorithms, SpecWatch-II and SpecWatch-III, based on the same framework. To evaluate the algorithms, we use *weak regret*, i.e., the performance difference between the solution of our algorithm and optimal (fixed) solution in hindsight, as the metric. We prove that the *expected* weak regret of SpecWatch-II is $O(T^{2/3})$, where T is the time horizon. Whereas, the *actual* weak regret of SpecWatch-III is $O(T^{2/3})$ with probability $1-\delta$, for any $\delta \in (0, 1)$. Both algorithms guarantee the upper bounds matching the lower bound of the general adversarial MAB-SC problem. Therefore, they are all asymptotically optimal.

1. Introduction

With the proliferation of wireless devices and applications, demand for access to spectrum has been growing dramatically and is likely to continue to grow in the foreseeable future [1]. However, there is a paradoxical phenomenon that usable radio frequencies are exhausted while much of the licensed spectrum lies idle at any given time and location [2]. To improve the radio spectrum utilization efficiency, dynamic spectrum access (DSA) in cognitive radio networks (CRNs) has been proposed as a promising approach. Among various DSA strategies, opportunistic spectrum access (OSA) based on the hierarchical access model has received much attention recently [3–10]. This underlay approach achieves spectrum sharing by allowing secondary (unlicensed) users (SUs) to dynamically search and access the spectrum vacancy while limiting the interference perceived by primary (licensed) users (PUs) [11].

OSA helps to improve the spectrum utilization but also results in spectrum misuse or abuse problems due to the flexibility of spectrum opportunity. For example, an SU may intentionally disobey the interference constraints set by the PU; or some greedy SUs may transmit more aggressively in time and frequency to dominate the spectrum sharing, or even emulate the PU to prevent other SUs from sharing. Through such spectrum access misbehavior, the malicious users (MUs), i.e., the misbehaving SUs, not only harm the spectrum access operations of normal users, but also impede the CRNs to function correctly since there is no incentive to pay for spectrum access [12]. Thus, spectrum monitoring is necessary and imperative.

To address the spectrum misuse problem, different trusted infrastructures have been proposed to detect spectrum misuse and punish MUs [12–15]. In addition, various detection techniques have been designed, including enforcing silence slots [16], publicizing back-off sequences [17,18], exploiting spatial pattern of signal strength [19], measuring detector value [20]. There is also a crowdsourcing-based framework named SpecGuard [21] which explores dynamic power control at SUs to contain the spectrum permit in physical layer signals. Another crowdsourced enforcement

* Corresponding author.
E-mail addresses: mili@mines.edu (M. Li), djyang@mines.edu (D. Yang), jilin@mines.edu (J. Lin), lim@email.arizona.edu (M. Li), jtang02@syr.edu (J. Tang).

https://doi.org/10.1016/j.comnet.2018.07.018
1389-1286/

framework [22] improves the probability of detection while reducing the likelihood of false positives for spectrum misuses and it can detect misuses caused by mobile users. Moreover, applying big data analysis and machine learning to cloud-based radio access networks also provide an appropriate approach to enable long-term spectrum monitoring [23].

However, all these works assume that all channels can be monitored at the same time. But monitoring all channels simultaneously is very energy-consuming and impractical. Therefore, in this paper, we consider spectrum monitoring in multi-channel CRNs with limited monitoring resource. In particular, we deploy one monitor with multiple radios where each radio is in charge of one channel, then make the best choice when choosing channels.

This channel selection problem is challenging because of the following reasons:

1. The information of MUs is unknown a priori. The monitor cannot know whether a channel is under attack or not unless it monitors the channel for a while. As a result, the monitor must balance between exploring new channels and exploiting channels which are frequently attacked.
2. The monitor needs to avoid switching too much since each time the monitor changes its monitoring channels, there are drastic costs in terms of delay, packet loss, and protocol overhead [24]. When switching costs are considered, the balance between exploration and exploitation is more complicated.
3. We consider the smart MUs (details in Section 3 of adaptive adversary in the threat model) which can adaptively change their attacking behaviors. MUs of this type are much more powerful than MUs in existing literatures where there exists some fixed probability distribution on how the MUs will attack the system.
4. The monitor may not be perfect and it may cause detection errors.
5. The monitor in our model can has more than one radios, which enable it to monitor multiple channels at the same time. However, this makes the number of choices become combinatorial.

To conquer all these challenges, we formulate the problem as a combinatorial adversarial (non-stochastic) multi-armed bandit (MAB) problem [25] with switching costs. We then propose an effective spectrum monitoring framework and design two online algorithms based on it.

In summary, we contribute in the following aspects:

1. We study the adversarial spectrum monitoring problem with unknown statistics in multi-channel CRNs, while considering the switching cost. We model this problem as an combinatorial adversarial MAB-SC problem.
2. We propose an online spectrum monitoring framework, SpecWatch. Based on this framework, we design two spectrum monitoring algorithms, SpecWatch-II and SpecWatch-III, which differ in the way of calculating strategy probabilities and updating strategy weights. Our algorithms guarantee the proved performance under any type of adversary settings. In addition, they can work with any spectrum misuse detection techniques in the current literature.
3. We prove that the expected weak regret of SpecWatch-II is $O(T^{2/3})$, which matches the lower bound in [26]. Therefore, SpecWatch-II is asymptotically optimal. Note that the expected value of normalized weak regret is guaranteed to be $O(1/T^{1/3})$, which converges to 0 as time horizon T approaches to ∞.
4. SpecWatch-III select channels more strategically and explore all channels more efficiently. We prove that this algorithm guarantees the actual weak regret to be $O(T^{2/3})$, which is asymptotically optimal as well, with probability $1-\delta$, for any $\delta \in (0, 1)$.

2. Related work

2.1. Spectrum monitoring problem

In this paper, we focus on the algorithm to determine the channels to choose, which is closely related to the sniffer-channel (sniffers are also referred to as monitors) assignment problem [27] in wireless networks.

In [28], Yeo et al. were the first to develop a framework exploiting dedicated sniffers to monitor WiFi networks and identify malicious usages. Cheng et al. [29,30] proposed an infrastructure and modeling techniques to monitor and analyze network behavior. In [31], Shin and Bagchi modeled the channel assignment for monitoring wireless mesh networks as maximum-coverage problem with group budget constraints. They then extended it to the model where monitors may make errors due to poor reception [32]. Along the same line, Nguyen et al. [33] focus on the weighted version of the problem, where users to be covered have weights. To maximize the captured data of interest, Chen et al. [34] utilized support vector regression to guide monitors to intelligently select channels. Considering similar objectives, Shin et al. [35] designed a cost-effective distributed algorithm. With a different approach, Yan et al. [36] solved the problem by predicting secondary users' access patterns. Li et al. [37] further considered the physical restrictions of the sniffers and formulated the problem as a new optimization problem.

However, all the above papers have different objectives and require more prior knowledge on network users.

The closest works to ours were presented in [38–42]. In [38], Arora et al. first modeled the spectrum monitoring problem as an multi-armed bandit problem (MAB) to monitor the maximum number of active users. They designed two algorithms to learn sequentially the user activities while making channel assignment decisions. Observing the above algorithms suffer from high computation cost, Zheng et al. [39] traded off between the rate of learning and the computation cost. They proposed a centralized online approximation algorithm and show that it incurs sub-linear regret bounds over time and a distributed algorithm with moderate message complexity. In [40], Le et al. considered switching costs for the first time and utilized Upper Confident Bound-based (UCB) policy [43] which enjoys a logarithmic regret bound in time that depends sublinearly on the number of arms, while its total switching cost grows in the order of $O(\log(\log T))$. Considering a different objective, Yi et al. [41] used UCB to capture as much as interested user data. However, these works used the stochastic MAB model, where the rewards for playing each arm are generated independently from unknown but *fixed* distributions. In other word, they all assume the user activities are time invariant. Our model, in contrast, does not make such assumptions.

The only work considered the similar problem model to ours is [42,44], where Xu et al. tried to capture packets of target SUs for CRN forensics. However, they did not provide any algorithm whose actual weak regret can be bounded with confidence value.

2.2. Multi-arm bandit problem

The MAB problem first introduced by Robbins [45] has been extensively studied in the literature. The classical MAB problem models the trade-offs faced by a gambler who aims to maximize his rewards over many turns by exploring different arms of slot machines and to exploit arms which have provided him more rewards than others. The gambler has no knowledge about the reward of each arm a priori and only gains knowledge of the arms he has pulled. An MAB algorithm should specify a strategy by which the gambler chooses an arm at each turn. The performance of an algorithm is measured in regret, as will be elaborated in Section 3.

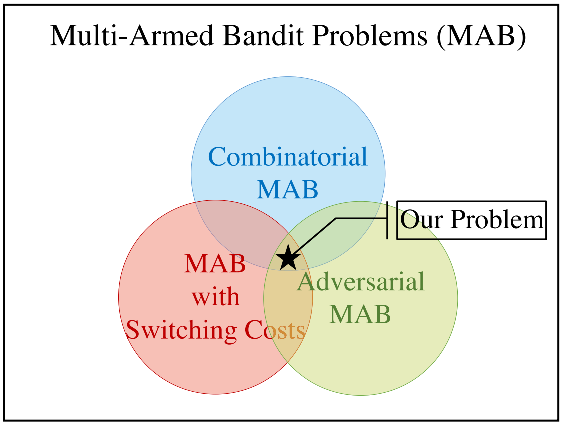


Fig. 1. Our problem considers combinatorial arms, adaptive adversary, and switching costs.

There are mainly two algorithm families based on different formulations of MAB. The upper confidence bounds (UCB) family of algorithms [46] works for stochastic MAB, whose regret can be as small as $O(\ln T)$ where T is the number of turns. However, these algorithms are established with the assumption that there exist fixed (though unknown) probability distributions of different arms to generate rewards, which may not be satisfied in our spectrum monitoring problem and thus not considered by us. The other algorithm family is the EXP3 family [25] for adversarial MAB. Auer et al. [25] has studied MAB with no assumption on the rewards distribution and proposed algorithms with regret of $O(T^{1/2})$. The adversarial MAB is one of the strongest generalizations of the bandit problem [47]. There are also some algorithms considering both stochastic and adversarial adversary [48,49].

However, switching costs are not considered in all above works.

In our model, each time the monitor changes its monitoring channels, there are drastic costs in terms of delay, packet loss, and protocol overhead [24]. These costs must be taken into consideration when designing monitoring algorithms. Although there exists some work on stochastic multi-armed bandit problem with switching cost (MAB-SC) [24], little research has been done on adversarial MAB-SC. Dekel et al. [26] proved the lower bound of the regret for adversarial MAB-SC to be $\tilde{\Omega}(T^{2/3})$. In this paper, the upper bound of regret guaranteed by our algorithms matches this lower bound.

Moreover, different from existing works, the strategy for each turn (or timeslot as in our model) is no longer a single arm because we consider a more general case where multiple channels (called *super arm* in many other papers) can be monitored at the same time. This makes our problem an instance of all combinatorial multi-armed bandit (CMAB) problems (Fig. 1). There are many challenges raised by the combinatorial nature of super arms. First, the exploration of all super arms requires much more heavy computational costs due to the combinatorial exploded number of super arms. Second, the revelation of good super arms becomes harder due to a more complicated reward structure, especially with the existence of detection errors. Therefore, directly treating every super arm as an arm and simply applying the classical MAB framework is not a good approach. A better solution is to utilize the observed information regarding the outcomes of each basic arm, which may be shared by other super arms. However, existing CMAB algorithms [50,51] only work for stochastic MAB problems.

In all, none of existing algorithms can be directly applied to our problem.

3. System model and problem statement

We consider a cognitive radio network which adopts a hierarchical access structure with primary users (PUs) and secondary users (SUs). We assume the spectrum is divided into a set $\mathcal{K} = \{1, 2, \ldots, k, \ldots, K\}$ of K channels. The total time period is discretized into a set $\mathcal{T} = \{1, 2, \ldots, t, \ldots, T\}$ of T timeslots. Ideally, SUs seek spectrum opportunities among K channels in a non-intrusive manner. However, the malicious users (MUs) may perform unauthorized access or selfish access. We consider the scenario where there exists one monitor with l radios and a set $\mathcal{M} = \{1, 2, \ldots, m, \ldots, M\}$ of M MUs. Note that for the case of multiple monitors, if there is a central controller, it is equivalent to one monitor with the same number of radios; otherwise, each monitor can execute our algorithms independently. In the latter case, however, the regret may not be bounded.

Since the monitor is equipped with l radios, it can monitor up to l channels at the same time. Assume that one radio is tuned to monitor channel k, and there are M_k MUs on that channel, then the detection probability of that radio to successfully detect MUs' presence is $p_d(M_k)$, which is dependent on the monitor's hardware and the detection technique. Any technique in [16–20] can be adopted to detect spectrum misuses. In practice, the detection probability will also be dependent on the presence of PU and other SUs. However, since our algorithms do not require the knowledge of the detection probability, we simplify the notion to $p_d(M_k)$ where it seems that only M_k matters.

Let $\{0, 1, \ldots, l\}^K$ denote the strategy space of the monitor. A *strategy* s is represented as $(a_{s1}, a_{s2}, \ldots, a_{sK})$, where the value of the a_{sk} is 1 if a radio is assigned to monitor channel k, 0 otherwise. Therefore, $\sum_{k \in \mathcal{K}} a_{sk} = l$. For example, considering 4 channels and a monitor with 2 radios, strategy (0, 1, 0, 1) indicates that one radio is tuned to monitor channel 2 and the other radio is tuned to monitor channel 4. For notational simplicity, we will write $k \in s$ instead of $a_{sk} = 1$ to denote that channel k is chosen in strategy s. Since each radio is assigned one out of K channels to monitor, and we have l radios in total, the number of strategies is $S = \binom{K}{l}$. The whole strategy set is represented as $\mathcal{S} = \{1, 2, \ldots, s, \ldots, S\}$. Note that K and l are usually small. For example, the regulated 2.4 GHz band is divided into only 14 channels. The maximum number of radios on each monitor defined by the active IEEE 802.11af standard is set to be 4 [52].

In this paper, we assume both the monitor and MUs are static, i.e., staying at the same location. Note that when mobility is considered as in the pursuit-evasion problems [53], we only need to enlarge the strategy space by including the location dimension.

At the beginning of timeslot $t \in \mathcal{T}$, the monitor selects only one strategy from the strategy set $\mathcal{S}$, and we denote the chosen strategy as X_t. We assume the switching cost $c(X_{t-1}, X_t) \in [0, 1]$, but our algorithm can be generalized to any range $[\underline{c}, \bar{c}]$, $\underline{c} < \bar{c}$ by scaling, where $\underline{c}$ and $\bar{c}$ are the minimum value and the maximum value of the switching cost, respectively. For simplicity, set the switching cost of the first timeslot to be $c(X_0, X_1) = c_0$ regardless of what X_1 is. Clearly, $c(X_{t-1}, X_t) = 0$ if $X_{t-1} = X_t$.

Threat model: At each timeslot $t \in \mathcal{T}$, each MU $m \in \mathcal{M}$ chooses one channel to attack (conduct misuses) according to its attack probability distribution $\mathbf{P}_t^m = \left\{P_{t,1}^m, \ldots, P_{t,K}^m\right\}$ where $P_{t,k}^m$ denotes the probability of MU m attacking channel k in timeslot t. Since MUs may not attack in some timeslot, $\sum_{k \in \mathcal{K}} P_{t,k}^m \leq 1$ for any $m \in \mathcal{M}$ and $t \in \mathcal{T}$. We consider two types of adversary:

1. Oblivious adversary (Stochastic): The MUs keep their attack patterns regardless of how the monitor work. For any MU m, the attack distribution $\mathbf{P}_t^m$ remains the same throughout the time horizon. In this paper, we consider three different adversary settings (elaborated in Section 7): fixed adversary, uniform adversary, and normal adversary.
2. Adaptive adversary (Adversarial): The MUs know every action of the monitor from the beginning to the current timeslot and

adjust their strategies accordingly based on any learning algorithms, i.e., the attack distribution might change with time.

Our framework and algorithms work for both types and the theoretical bounds hold no matter what the adversary type is.

Now we define the reward for the monitor. The *strategy reward* of choosing strategy s in timeslot t is

$$g_{s,t} \overset{\text{def}}{=} \begin{cases} \sum_{k\in s} f_{k,t} & \text{if } s = X_t, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where the *channel reward* $f_{k,\,t}$ is defined as

$$f_{k,t} \overset{\text{def}}{=} \begin{cases} r & \text{if channel } k \in X_t \text{ and misuse is detected,} \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where the *unit reward* r is assumed to be scaled and satisfies $rl \le 1$ for the purpose of mathematical analysis. Note that the probability of at least one MU being detected on monitored channel k is determined by the number of radios on that channel M_k, the detection probability $p_d(M_k)$ and the action of MUs $\{P^m_{t,k}\}^M_{m=1}$. We denote the detection probability on channel k by adopting strategy X_t at timeslot t as $P_d\big(a_{X_t k}, p_d(M_k), \{P^m_{t,k}\}^M_{m=1}\big)$ which is assumed to be a non-decreasing function in $a_{X_t k}$, $p_d(M_k)$, and $P^m_{t,k}$. Thus, the channel reward $f_{k,\,t}$ is r with probability $P_d\big(a_{X_t k}, p_d(M_k), \{P^m_{t,k}\}^M_{m=1}\big)$. Note that the knowledge of this probability is not required.

Assume the monitor follows $X_1, X_2, \ldots, X_T$, which is the strategy sequence generated by any monitoring Algorithm A. At the end of timeslot T, the *cumulative strategy reward* is

$$G_{\mathbb{A}} \overset{\text{def}}{=} \sum_{t=1}^{T} g_{X_t,t}. \tag{3}$$

Meanwhile, the monitor incurs *cumulative switching cost*

$$L_{\mathbb{A}} \overset{\text{def}}{=} \sum_{t=1}^{T} c(X_{t-1}, X_t). \tag{4}$$

Thus, the *utility* of the monitor by choosing Algorithm A is

$$U_{\mathbb{A}} = G_{\mathbb{A}} - L_{\mathbb{A}}. \tag{5}$$

To measure the performance of Algorithm A, we use a special case of the worst-case regret, *weak regret* [25], as the metric.

The weak regret of Algorithm A is the difference between the utility by using *best fixed algorithm* and the actual utility by using Algorithm A. A fixed algorithm chooses only one strategy for all timeslots and never switches. The best fixed algorithm is the one resulting in the highest utility among all fixed algorithms. The strategy chosen in best fixed algorithm is called the *best strategy*, denoted by s_{best}. Formally, $s_{best} \overset{\text{def}}{=} \operatorname{argmax}_{s\in\mathcal{S}} \left(\sum_{t=1}^{T} g_{s,t} - c_0\right)$, and the utility by using the best fixed algorithm is

$$U_{best} \overset{\text{def}}{=} G_{best} - L_{best}, \tag{6}$$

where $G_{best} = \max_{s\in\mathcal{S}} \sum_{t=1}^{T} g_{s,t}$ and $L_{best} = c_0$ since the switch only happens at the first timeslot. Note that the best strategy can only be found in hindsight.

Now we can define the weak regret of Algorithm A as

$$R_{\mathbb{A}} \overset{\text{def}}{=} U_{best} - U_{\mathbb{A}}. \tag{7}$$

Problem statement: Given K channels, time horizon T, and a monitor with l radios, our objective is to design online spectrum monitoring algorithms such that the weak regret is minimized, in the presence of different adversaries. We make no assumption on the knowledge of the probability functions $p_d(M_k)$ and $P_d\big(a_{X_t k}, p_d(M_k), \{P^m_{t,k}\}^M_{m=1}\big)$. In addition, the attack distribution $\mathbf{P}^m_t$ and the reward of choosing a strategy are unknown a priori.

Therefore, any desired algorithm needs to balance not only the trade-off between exploration and exploitation, but also that between strategy rewards and switching costs. This is a very challenging problem.

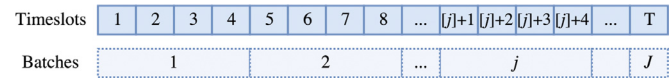


Fig. 2. Example of batching timeslots with batch size $\tau = 4$.

4. Spectrum monitoring framework

In this section, we design SpecWatch, a spectrum monitoring framework, based on the batching version of exponential-weight algorithm for exploration and exploitation, where the idea of batching is inspired by Arora et al. [54].

To control the trade-off between the reward and the switching cost, we group all the timeslots into consecutive and disjoint batches. Within each batch, we stick to the same strategy to avoid the switching cost. Between batches, we reselect a strategy to gain higher rewards. A smaller batch size may result in larger reward but larger switching cost, while a bigger batch size may result in smaller switching cost but smaller reward.

The design details of SpecWatch are illustrated above. We first set the initial weight of each strategy to be 1. Given the batch size τ, the timeslots $1, 2, \ldots, T$ are divided into $J = \lceil T/\tau \rceil$ consecutive and disjoint batches. Let $[j] = (j-1)\tau$ for $1 \le j \le J$. Then the jth batch starts from timeslot $[j]+1$ and ends at timeslot $[j]+\tau$ as shown in Fig. 2. At the beginning of each batch, we calculate the strategy probabilities according to strategy weights. Then we randomly select a strategy based on the probability distribution. During the whole batch, the chosen strategy remains the same. At the end of each batch, strategy weights are updated according to the strategy rewards.

SpecWatch:

Parameter: $\tau \in [1, T]$

Initialization: $w_{s,1} \leftarrow 1$ for all $s \in \mathcal{S}$, $J \leftarrow \lceil T/\tau \rceil$.

for $j \leftarrow 1, \ldots, J$ **do**
- Calculate the strategy probability $p_{s,j}$ for all $s \in \mathcal{S}$
- Choose strategy $Z_j \in \mathcal{S}$ randomly accordingly to the probability distribution $p_{1,j}, \ldots, p_{S,j}$ and incur switching cost $c(Z_{j-1}, Z_j)$.
- Monitor using Z_j for τ timeslots, i.e., $X_{[j]+i} \leftarrow Z_j$ for $1 \le i \le \tau$, and record the reward of each monitored channel, $f_{k,[j]+i}$ for all $k \in Z_j$, $1 \le i \le \tau$.
- Update strategy weight $w_{s,j+1}$ for all $s \in \mathcal{S}$.

end

In the following sections, we design two effective online spectrum monitoring algorithms, SpecWatch-II and SpecWatch-III. These two algorithms are designed based on SpecWatch and SpecWatch$^+$ in our paper [55]. We rename SpecWatch and SpecWatch$^+$ to SpecWatch-I and SpecWatch-III, respectively. We introduce SpecWatch-II in this paper and omit SpecWatch-I because SpecWatch-II has better theoretical performance than SpecWatch-I.

The main notations are summarized in Table 1. All sets are in math calligraphy ($\mathcal{K}, \mathcal{T}, \mathcal{J}, \mathcal{M}, \mathcal{S}, \mathcal{C}$). All timeslot-related notations are corresponding to t, such as X_t, $f_{k,\,t}$. Similarly, all batch-related notations are corresponding to j, such as Z_j, $\bar{f}_{k,j}$. All parameters (τ, γ, η, β) of proposed algorithms are denoted by Greek letters. Except for X_t, Z_j, other notations using capital letters denote summations, e.g., $G_{\mathbb{A}}, R_{\mathbb{A}}, W_j$.

Table 1
Main notations.

Notation	Meaning
$\mathcal{K}=\{1,2,\ldots,k,\ldots,K\}$	set of all channels
$\mathcal{T}=\{1,2,\ldots,t,\ldots,T\}$	set of all **t**imeslots
$\mathcal{J}=\{1,2,\ldots,j,\ldots,J\}$	set of all batches
$\mathcal{M}$	set of **m**alicious users
$\mathcal{S}=\{1,2,\ldots,s,\ldots,S\}$	set of all **s**trategies
$\mathcal{C}=\{1,2,\ldots,c,\ldots,C\}$	**c**overing strategy set
C_k	number of strategies in $\mathcal{C}$ that contains channel k
l	number of radios in the monitor
r	unit channel **r**eward which satisfies $rl\le 1$
$c(X_{t-1},X_t)$	switching **c**ost from strategy X_{t-1} to X_t
$G_{\mathbb{A}}$	cumulative strategy rewards of Algorithm A
$L_{\mathbb{A}}$	cumulative switching costs of Algorithm A
$U_{\mathbb{A}}$	utility of Algorithm A
$R_{\mathbb{A}}$	weak **r**egret of Algorithm A
τ	parameter to determine the batch size
γ	parameter to calculate strategy probabilities
η	parameter to update weights
β	parameter to calculate channel scores
X_t	chosen stratgegy in timeslot t
Z_j	chosen stratgegy in batch j
$f_{k,\,t}$	channel reward of channel k in timeslot t
$g_{s,\,t}$	strategy reward of strategy s in timeslot t
$\bar{f}_{k,j}$	average channel reward of channel k in batch j
$\bar{g}_{s,j}$	average strategy reward of strategy s in batch j
$\bar{f}'_{k,j}$	average channel score of channel k in batch j
$\bar{g}'_{s,j}$	average strategy score of strategy s in batch j
$h_{k,\,j}$	channel weight of channel k in batch j
$w_{s,\,j}$	strategy weight of strategy s in batch j
W_j	total weight of all strategies in batch j
$q_{k,\,j}$	channel probability of channel k in batch j
$p_{s,\,j}$	strategy probability of strategy s in batch j

Different algorithms use different equations to calculate the strategy probabilities (Line 4) and update the strategy weights (Line 7). With carefully chosen parameters, their theoretical performances are shown in Table 2. In summary, SpecWatch-I and SpecWatch-II bound the *expected* weak regrets to $O(T^{2/3})$; while SpecWatch-III bounds the *actual* weak regret to $O(T^{2/3})$ with user-defined probability.

Both algorithms are online algorithm which records the monitoring history every timeslot and update the monitoring strategy every batch (every τ timeslots). Considering each timeslot, the time complexity of the calculation is $O(S)$, where S is the number of all possible strategies. This is the common complexity for all adversarial MAB-based algorithms. Even though $S=\binom{K}{l}$ is not polynomial in K or l, the values of k and l are usually very small in practice. For example, the maximum value of K is 45 according to Google's Spectrum Database [56] and the maximum number of radios on each monitor defined by the active IEEE 802.11af standard is set to be 4 [52].

5. Spectrum monitoring algorithm with bounded expected weak regret

In this section, we design one spectrum monitoring algorithm, SpecWatch-II, whose expected weak regret is theoretically bounded.

5.1. SpecWatch-II

5.1.1. Algorithm design

Compared to SpecWatch-I, the input parameters of SpecWatch-II are τ and η, where τ determines the batch size and η is used to calculate strategy weights; the strategy probabilities in SpecWatch-II are proportion to the strategy weights only, and the strategy weights are updated smaller for the next timeslot.

Calculating strategy probability $p_{s,\,j}$. The probability of choosing strategy $s\in\mathcal{S}$ is calculated using

$$p_{s,j}=\frac{w_{s,j}}{W_j}, \tag{8}$$

where $w_{s,\,j}$ is the strategy weight of strategy s in batch j, and $W_j=\sum_{s\in\mathcal{S}}w_{s,j}$. Recall that in SpecWatch-I, we use the weighted average of two terms, $p_{s,j}=(1-\gamma)\frac{w_{s,j}}{W_j}+\frac{\gamma}{S}$, where the first is to exploit strategies with good reward history, and the second guarantees the exploration over all strategies. γ controls the balance between them. Now we only have one term, but the balance still exists. This is because a strategy not chosen before is guaranteed to have a higher weight, due to the way we update the strategy weight as discussed below.

Choosing strategyZ_j and monitoring spectrum. We select a strategy $Z_j\in\mathcal{S}$ randomly according to the probabilities calculated above. The monitor keeps using Z_j for all τ timeslots in batch j, i.e., $X_{[j]+i}=Z_j$ for $1\le i\le\tau$. Therefore, the monitor only incurs switching cost $c(Z_{j-1},Z_j)$ once for the whole batch j.

Depending on the misuse behavior of MUs (discussed in Section 3), the monitor receives rewards on monitored channels accordingly. The monitor keeps records of $f_{k,[j]+i}$ for all $k\in Z_j$ and $1\le i\le\tau$. The strategy reward gained by the monitor is the summation of rewards over all monitored channels.

Updating strategy weight $w_{s,j+1}$. At the end of each batch, we update strategy weights in the following steps.

First, we calculate the average channel reward of each channel $k\in\mathcal{K}$ in batch j,

$$\bar{f}_{k,j}=\frac{1}{\tau}\sum_{i=1}^{\tau}f_{k,[j]+i}. \tag{9}$$

By (2), $\bar{f}_{k,j}\in[0,r]$. We also calculate the probability of choosing channel $k\in Z_j$ by summing up the probabilities of strategies containing that channel,

$$q_{k,j}=\sum_{s:k\in s}p_{s,j}. \tag{10}$$

where s: $k\in s$ denotes any strategy s containing channel k. Based on (9) and (10), we calculate the *average channel score*,

$$\bar{f}'_{k,j}=\frac{1/l-\bar{f}_{k,j}}{q_{k,j}}\mathbb{1}_{k\in Z_j}, \tag{11}$$

where $\mathbb{1}_{k\in Z_j}$ is an indicator function which has the value 1 if $k\in Z_j$; 0 otherwise. This score is always nonnegative since $\bar{f}_{k,j}\le r\le 1/l$. Then we update each *channel weight* by

$$h_{k,j+1}=h_{k,j}\exp\left(-\eta\bar{f}'_{k,j}\right), \tag{12}$$

where $h_{k,1}=1$ for all $k\in\mathcal{K}$. Note that the channel weight is non-increasing from batch to batch.

Finally, we give the formal definition of *strategy weight*, which is defined as

$$w_{s,j}\overset{\textbf{def}}{=}\prod_{k\in s}h_{k,j}. \tag{13}$$

Combining (12) and (13), we can directly update the strategy weight for each $s\in\mathcal{S}$ by

$$w_{s,j+1}=w_{s,j}\exp(-\eta\bar{g}'_{s,j}), \tag{14}$$

where $\bar{g}'_{s,j}$ is the *average strategy score* for each $s\in\mathcal{S}$, i.e.,

$$\bar{g}'_{s,j}=\sum_{k\in s}\bar{f}'_{k,j}. \tag{15}$$

Table 2
Algorithm comparison.

Algorithm	Core functions	Theoretical bound of weak regret
SpecWatch-I	$p_{s,j} = (1-\gamma)\frac{w_{s,j}}{w_j} + \frac{\gamma}{S}$ $w_{s,j+1} = w_{s,j}\exp\left(\frac{\gamma}{S}\sum_{k\in s}\frac{\frac{1}{\tau}\sum_{i=1}^{\tau} f_{k,[j]+i}}{\sum_{s:k\in s} p_{s,j}}\right)$	$\mathbf{E}[R_{\mathrm{I}}] \le 3((e-1)S\ln S)^{\frac{1}{3}}T^{\frac{2}{3}}$
SpecWatch-II	$p_{s,j} = \frac{w_{s,j}}{w_j}$ $w_{s,j+1} = w_{s,j}\exp\left(-\eta\sum_{k\in s}\frac{\mathbb{1}_{k\in Z_j}(1/l - \frac{1}{\tau}\sum_{i=1}^{\tau} f_{k,[j]+i})}{\sum_{s:k\in s} p_{s,j}}\right)$	$\mathbf{E}[R_{\mathrm{II}}] \le 3\left(\frac{1}{2}S\ln S\right)^{\frac{1}{3}}T^{\frac{2}{3}}$
SpecWatch-III	$p_{s,j} = (1-\gamma)\frac{w_{s,j}}{w_j} + \frac{\gamma}{C}\mathbb{1}_{s\in\mathcal{C}}$ $w_{s,j+1} = w_{s,j}\exp\left(\eta\sum_{k\in s}\frac{\frac{1}{\tau}\sum_{i=1}^{\tau} f_{k,[j]+i}+\beta}{(1-\gamma)\frac{\sum_{s:k\in s} w_{s,j}}{w_j} + \frac{\gamma C_k}{C}}\right)$	$\Pr\left[R_{\mathrm{III}} \le 2\left(4\sqrt{lC\ln S} + 2\sqrt{lK\ln\frac{K}{\delta}}\right)^{\frac{2}{3}}T^{\frac{2}{3}}\right]$ $\le 1-\delta$, where $C \le K$

Note that by combining (9), (10) and (11), we can directly calculate $\bar{g}'_{s,j}$ directly by

$$\bar{g}'_{s,j} = \sum_{k\in s}\frac{\mathbb{1}_{k\in Z_j}(1/l - \bar{f}_{k,j})}{q_{k,j}} = \sum_{k\in s}\frac{\mathbb{1}_{k\in Z_j}(1/l - \frac{1}{\tau}\sum_{i=1}^{\tau} f_{k,[j]+i})}{\sum_{s:k\in s} p_{s,j}}. \quad (16)$$

Remark. We do not update strategy weights based on strategy rewards, but instead calculate channel weights (12) first. This is because the rewards of monitored channels provide useful information on those *unchosen strategies* containing these channels.

5.1.2. Performance analysis

To analyze the performance of SpecWatch-II, we first bound the difference between the reward gained by SpecWatch-II and that by the best fixed algorithm (Lemma 1), and then prove the upper bound of the expected weak regret (Theorem 1). For a better understanding of Theorem 1, we present a specific bound obtained by a particular choice of parameters η and τ (Corollary 1).

Recalling (5), (6) and (7), the weak regret of SpecWatch-II is

$$R_{\mathrm{II}} = (G_{best} - L_{best}) - (G_{\mathrm{II}} - L_{\mathrm{II}}). \quad (17)$$

Since the best fixed algorithm never switches the strategies, and SpecWatch-II only switches between batches for at most J times, their cumulative switching costs are

$$L_{best} = c_0 \in [0, 1] \quad \text{and} \quad L_{\mathrm{II}} = \sum_{j=1}^{J} c(Z_{j-1}, Z_j) \le J.$$

Thus, we have

$$L_{\mathrm{II}} - L_{best} \le J. \quad (18)$$

Now it suffices to only consider the difference between rewards.

An important observation is that we group the timeslots into batches, calculate the strategy probabilities only at the beginning of each batch, and use the average value of entire batch to update weight. Therefore, each batch can be considered as a round in conventional MAB. With this consideration, we introduce the notations below for our proofs,

$$\bar{g}_{s,j} = \mathbb{1}_{s=Z_j}\sum_{k\in s}\bar{f}_{k,j},\ \bar{G}_{\mathrm{II}} \overset{\text{def}}{=} \sum_{j=1}^{J}\bar{g}_{Z_j,j} \text{ and } \bar{G}_{best} \overset{\text{def}}{=} \max_{s\in\mathcal{S}}\sum_{j=1}^{J}\bar{g}_{s,j}. \quad (19)$$

Note that

$$\begin{aligned}\bar{g}_{s,j} &= \frac{1}{\tau}\mathbb{1}_{s=Z_j}\sum_{k\in s} f_{k,j} = \frac{1}{\tau}\mathbb{1}_{s=Z_j}\sum_{k\in s}\sum_{i=1}^{\tau} f_{k,[j]+i} \\ &= \frac{1}{\tau}\mathbb{1}_{s=Z_j}\sum_{i=1}^{\tau}\sum_{k\in s} f_{k,[j]+i} = \frac{1}{\tau}\sum_{i=1}^{\tau} g_{s,[j]+i}. \end{aligned} \quad (20)$$

Thus we have

$$G_{\mathrm{II}} = \sum_{t=1}^{T} g_{X_t,t} = \tau\sum_{j=1}^{J}\bar{g}_{Z_j,j} = \tau\bar{G}_{\mathrm{II}}. \quad (21)$$

Similarly,

$$G_{best} = \tau\bar{G}_{best}. \quad (22)$$

We now provide the bound of the expected difference between $\bar{g}_{\mathrm{II}}$ and $\bar{g}_{best}$.

Lemma 1. *For any type of adversaries, any $T > 0$, and any $\eta > 0$, we have*

$$\mathbf{E}\left[\bar{G}_{best} - \bar{G}_{\mathrm{II}}\right] \le \frac{\eta JS}{2} + \frac{\ln S}{\eta}, \quad (23)$$

where ln *is the natural logarithm function.*

Proof. The proof of this lemma is based on [48, Theorem 3.1] with necessary modifications and extensions.

First we analyze $\frac{W_{j+1}}{W_j}$. For any sequences $Z_1, \ldots, Z_j$ generated by SpecWatch-II, we have

$$\begin{aligned}\frac{W_{j+1}}{W_j} &= \sum_{s\in\mathcal{S}}\frac{w_{s,j+1}}{W_j} \\ &= \sum_{s\in\mathcal{S}}\frac{w_{s,j}}{W_j}\exp\left(-\eta\bar{g}'_{s,j}\right) \\ &= \sum_{s\in\mathcal{S}} p_{s,j}\exp\left(-\eta\bar{g}'_{s,j}\right) \end{aligned} \quad (24)$$

$$\le \sum_{s\in\mathcal{S}} p_{s,j}\left(1 - \eta\bar{g}'_{s,j} + \frac{1}{2}\left(\eta\bar{g}'_{s,j}\right)^2\right) \quad (25)$$

$$\le 1 - \eta\sum_{s\in\mathcal{S}}\left(p_{s,j}\bar{g}'_{s,j}\right) + \frac{\eta^2}{2}\sum_{s\in\mathcal{S}} p_{s,j}\left(\bar{g}'_{s,j}\right)^2, \quad (26)$$

where (24) uses the definition of $p_{s,\,j}$, and (25) holds by the fact that for $x \ge 0$, $e^{-x} \le 1 - x + \frac{1}{2}x^2$.

Next, we bound (26) by bounding $\sum_{s\in\mathcal{S}} p_{s,j}\bar{g}'_{s,j}$ and $\sum_{s\in\mathcal{S}} p_{s,j}\left(\bar{g}'_{s,j}\right)^2$.

$$\begin{aligned}\sum_{s\in\mathcal{S}} p_{s,j}\bar{g}'_{s,j} &= \sum_{s\in\mathcal{S}}\left(p_{s,j}\sum_{k\in s}\bar{\mathrm{f}}'_{k,j}\right) \\ &= \sum_{k\in\mathcal{K}}\left(\bar{\mathrm{f}}'_{k,j}\sum_{s:k\in s} p_{s,j}\right) \\ &= \sum_{k\in\mathcal{K}}\left(\bar{\mathrm{f}}'_{k,j} q_{k,j}\right) \\ &= \sum_{k\in Z_j}\left(1/l - \bar{\mathrm{f}}_{k,j}\right) \\ &\ge 1 - \bar{g}_{Z_j,j}, \end{aligned} \quad (27)$$

$$
\begin{aligned}
\sum_{s\in\mathcal{S}} p_{s,j}\left(\bar{g}'_{s,j}\right)^2 &= \sum_{s\in\mathcal{S}}\left(p_{s,j}\left(\sum_{k\in s}\bar{f}'_{k,j}\right)\right)^2\\
&\le \sum_{s\in\mathcal{S}}\left(p_{s,j}\cdot l\cdot\sum_{k\in s}\left(\bar{f}'_{k,j}\right)^2\right)\\
&= l\cdot\sum_{k\in\mathcal{K}}\left(\left(\bar{f}'_{k,j}\right)^2\sum_{s:k\in s}p_{s,j}\right)\\
&= l\cdot\sum_{k\in\mathcal{K}}\left(\left(\bar{f}'_{k,j}\right)^2 q_{k,j}\right)\\
&= l\cdot\sum_{k\in\mathcal{K}}\left(\bar{f}'_{k,j}\cdot\frac{(1/l-\bar{f}_{k,j})}{q_{k,j}}\cdot q_{k,j}\right)\\
&\le l\cdot\sum_{k\in\mathcal{K}}\left(\bar{f}'_{k,j}\cdot\frac{1}{l}\right)\\
&= \sum_{k\in\mathcal{K}}\bar{f}'_{k,j}\\
&= \sum_{k\in Z_j}\bar{f}'_{k,j}+\sum_{k\in\mathcal{K}\setminus Z_j}\bar{f}'_{k,j}\\
&= \bar{g}'_{Z_j,j},
\end{aligned}
\tag{28}
$$

where (28) holds as a special case of the Cauchy–Schwarz inequality,

Thus, we have

$$
\sum_{s\in\mathcal{S}} p_{s,j}\bar{g}'_{s,j}\ge 1-\bar{g}_{Z_j,j}\ \text{and}\ \sum_{s\in\mathcal{S}} p_{s,j}\left(\bar{g}'_{s,j}\right)^2\le \bar{g}'_{Z_j,j}. \tag{29}
$$

Combining (26) and (29), we have

$$
\frac{W_{j+1}}{W_j}\le 1-\eta(1-\bar{g}_{Z_j,j})+\frac{\eta^2}{2}\bar{g}'_{Z_j,j}. \tag{30}
$$

Taking the log of both sides and using $1+x\le e^x$ gives

$$
\ln\frac{W_{j+1}}{W_j}\le -\eta(1-\bar{g}_{Z_j,j})+\frac{\eta^2}{2}\bar{g}'_{Z_j,j}. \tag{31}
$$

Summing over j, we then get

$$
\ln\frac{W_{J+1}}{W_1}\le -\eta(J-\bar{G}_{\mathbb{I}})+\frac{\eta^2}{2}\sum_{j=1}^{J}\bar{g}'_{Z_j,j}. \tag{32}
$$

Now we consider the lower bound of $\ln\frac{W_{J+1}}{W_1}$.

For any strategy s,

$$
\begin{aligned}
\ln\frac{W_{J+1}}{W_1} &\ge \ln\frac{w_{s,J+1}}{W_1}\\
&= \ln\frac{w_{s,1}\exp\left(-\eta\sum_{j=1}^{J}\bar{g}'_{s,j}\right)}{Sw_{s,1}}\\
&= -\eta\sum_{j=1}^{J}\bar{g}'_{s,j}-\ln S.
\end{aligned}
$$

Since the above inequality holds for any strategy s, we get

$$
\ln\frac{W_{J+1}}{W_1}\ge -\eta\min_{s\in\mathcal{S}}\sum_{j=1}^{J}\bar{g}'_{s,j}-\ln S. \tag{33}
$$

Combining (32) and (33), we have

$$
-\eta(J-\bar{G}_{\mathbb{I}})+\frac{\eta^2}{2}\sum_{j=1}^{J}\bar{g}'_{Z_j,j}\ge -\eta\min_{s\in\mathcal{S}}\sum_{j=1}^{J}\bar{g}'_{s,j}-\ln S.
$$

Dividing both sides by $-\eta$ and moving terms, we have

$$
(J-\bar{G}_{\mathbb{I}})-\min_{s\in\mathcal{S}}\sum_{j=1}^{J}\bar{g}'_{s,j}\le\frac{\eta}{2}\sum_{j=1}^{J}\bar{g}'_{Z_j,j}+\frac{\ln S}{\eta}. \tag{34}
$$

Next, we take expectations of both sides in (34) with respect to the distribution of $Z_1, Z_2, \ldots, Z_j$.

For the conditional expected value of each $\bar{g}'_{s,j}$, since the expected value of a discrete random variable is the probability-weighted average of all possible values, we have

$$
\begin{aligned}
&\mathbf{E}\left[\bar{g}'_{s,j}|Z_1,Z_2,\ldots,Z_{j-1}\right]\\
=&p_{s,j}\cdot\mathbf{E}\left[\bar{g}'_{s,j}|s=Z_j\right]+(1-p_{s,j})\cdot\mathbf{E}\left[\bar{g}'_{s,j}|s\ne Z_j\right].
\end{aligned}
\tag{35}
$$

Note that

$$
\begin{aligned}
\mathbf{E}\left[\bar{g}'_{s,j}|s\ne Z_j\right] &= \sum_{k\in s}\frac{\mathbb{1}_{k\in Z_j}(1/l-\bar{f}_{k,j})}{q_{k,j}}\\
&= \sum_{k\in s\wedge k\in Z_j}\frac{\mathbb{1}_{k\in Z_j}(1/l-\bar{f}_{k,j})}{q_{k,j}}\\
&\le \sum_{k\in Z_j}\frac{\mathbb{1}_{k\in Z_j}(1/l-\bar{f}_{k,j})}{q_{k,j}}\\
&= \mathbf{E}\left[\bar{g}'_{s,j}|s=Z_j\right],
\end{aligned}
\tag{36}
$$

where (36) is based on (16).

Therefore, from (35) we have

$$
\begin{aligned}
\mathbf{E}\left[\bar{g}'_{s,j}\right] &\le \mathbf{E}\left[\bar{g}'_{s,j}|s=Z_j\right]\\
&= \mathbf{E}\left[\sum_{k\in s}\frac{\mathbb{1}_{k\in Z_j}(1/l-\bar{f}_{k,j})}{q_{k,j}}|s=Z_j\right]\\
&= \sum_{k\in Z_j}(1/l-\bar{f}_{k,j})
\end{aligned}
\tag{37}
$$

$$
\begin{aligned}
&= 1-\bar{g}_{Z_j,j}\\
&\le 1-\bar{g}_{s',j},\quad \text{for any } s'\in\mathcal{S},
\end{aligned}
\tag{38}
$$

where (37) is based on $\mathbb{1}_{k\in Z_j}=1, q_{k,j}=1$ given $k\in s\wedge s=Z_j$, and (38) is based on the definition of strategy reward (1) and the definition of the average strategy reward (20). More clearly, $\bar{g}_{s',j}=\frac{1}{\tau}\sum_{i=1}^{\tau}g_{s',[j]+i}=\frac{1}{\tau}\sum_{i=1}^{\tau}0=0$ if $s'\ne Z_j$ for any $s'\in\mathcal{S}$. The notation s' stands for an arbitrary monitoring strategy, this strategy can be or not be the same as s.

Then we get

$$
\mathbf{E}\left[\min_{s\in\mathcal{S}}\sum_{j=1}^{J}\bar{g}'_{s,j}\right]\le J-\max_{s\in\mathcal{S}}\sum_{j=1}^{J}\bar{g}_{s,j}=J-\bar{G}_{best} \tag{39}
$$

and

$$
\begin{aligned}
\mathbf{E}\left[(J-\bar{G}_{\mathbb{I}})-\min_{s\in\mathcal{S}}\sum_{j=1}^{J}\bar{g}'_{s,j}\right] &\ge \mathbf{E}\left[(J-\bar{G}_{\mathbb{I}})-(J-\bar{G}_{best})\right]\\
&= \mathbf{E}\left[\bar{G}_{best}-\bar{G}_{\mathbb{I}}\right].
\end{aligned}
\tag{40}
$$

Moreover, from (38) we also know that

$$
\mathbf{E}\left[\bar{g}'_{Z_j,j}\right]\le 1-\bar{g}_{Z_j,j}\le 1. \tag{41}
$$

Therefore, combing (34), (40), and (41), we have

$$
\mathbf{E}\left[\bar{G}_{best}-\bar{G}_{\mathbb{I}}\right]\le\frac{\eta}{2}JS+\frac{\ln S}{\eta}.
$$

□

Now taking into consideration the bound of switching costs, (18), we have the following theorem:

Theorem 1. *For any type of adversaries, the expected weak regret of SpecWatch-II is* $O(T^{\frac{2}{3}})$ *with parameters* $\eta = A_\eta T^{-\frac{1}{3}} > 0$ and $\tau = A_\tau T^{\frac{1}{3}} \in [1, T]$, *where* A_γ *and* A_τ *are constants. Specifically,*

$$\mathbf{E}[R_{\mathrm{II}}] \le \left(\frac{A_\eta S}{2} + \frac{A_\tau \ln S}{A_\eta} + \frac{1}{A_\tau}\right) T^{\frac{2}{3}}. \tag{42}$$

Proof.

$$\begin{aligned}\mathbf{E}[R_{\mathrm{II}}] &= \mathbf{E}[G_{best} - L_{best} - G_{\mathrm{II}} + L_{\mathrm{II}}]\\ &\le \mathbf{E}[G_{best} - G_{\mathrm{II}}] + J\\ &= \mathbf{E}\left[\tau \bar{G}_{best} - \tau \bar{G}_{\mathrm{II}}\right] + J\\ &= \frac{\eta J S \tau}{2} + \frac{\tau \ln S}{\eta} + J\\ &= \frac{\eta S T}{2} + \frac{\tau \ln S}{\eta} + T/\tau\\ &= \left(\frac{A_\eta S}{2} + \frac{A_\tau \ln S}{A_\eta} + \frac{1}{A_\tau}\right) T^{\frac{2}{3}}.\end{aligned}$$

□

For better understanding of Theorem 1, we now give a specific bound by choosing particular parameters.

Corollary 1. *For any type of adversaries, when* $T \ge \frac{1}{2} S \ln S$, *with parameters* $\eta = \sqrt[3]{\frac{4 \ln S}{S^2 T}}$ *and* $\tau = \sqrt[3]{\frac{2T}{S \ln S}}$, *the expected weak regret of SpecWatch-II is*

$$\mathbf{E}[R_{\mathrm{II}}] \le 3\left(\frac{1}{2} S \ln S\right)^{\frac{1}{3}} T^{\frac{2}{3}}. \tag{43}$$

Proof. Substituting the parameters in (42), we have the immediate result. □

Remark. The expected weak regret of SpecWatch-II is bounded by $O(T^{2/3})$. This upper bound matches the lower bound of the MAB-SC problem proved in [26] which is used to model the spectrum monitoring problem in this paper. Thus, SpecWatch-II is asymptotically optimal. If we calculate the normalized weak regret $\frac{R_{\mathrm{II}}}{T}$, i.e., amortizing the regret to every timeslot, then it is clear that the expected value of normalized weak regret converges to 0 as T approaches to ∞.

6. Spectrum monitoring algorithm with bounded weak regret

We have already proved that SpecWatch-II is an effective online spectrum monitoring algorithms with expected normalized regret converging to 0. Though the *expectation* provides a quite legitimate estimate on the performance of SpecWatch-I and SpecWatch-II, the actual value of weak regret may sometimes deviate a lot from the expected bound as expectation just represents the mean. In this section, we propose the improved algorithm, SpecWatch-III, whose actual weak regrets are bounded by $O(T^{2/3})$ with any user-defined confidence level.

Moreover, by introducing a new concept, *covering strategy set*, we reduce the coefficient of the weak regret's bound from $O(\sqrt[3]{S \ln S})$ to $O(\sqrt[3]{C \ln S})$, where $C \le K$. When $S \gg K$, this algorithm is highly recommended than SpecWatch-II.

6.1. SpecWatch-III

6.1.1. Algorithm design

SpecWatch-III is designed similar to SpecWatch-II. However, it takes four parameters τ, γ, β and η as input, where γ is used to calculate strategy probabilities, β is used to calculate average channel scores, and η is used to update strategy weights.

Calculating strategy probability $p_{s,j}$. For calculating strategy probabilities, we introduce a new concept called *covering strategy set*. A covering strategy set $\mathcal{C} \subset \mathcal{S}$ is a set of strategy that *covers* all channels $\mathcal{K}$, where a channel $k \in \mathcal{K}$ is covered by $\mathcal{C}$ if there is a strategy $s \in \mathcal{C}$ such that $k \in s$. In SpecWatch-III, we randomly construct a minimal covering strategy set whose size $C \stackrel{\mathbf{def}}{=} |\mathcal{C}|$ is less than or equal to K. The probability of each strategy s is calculated by

$$p_{s,j} = (1-\gamma)\frac{w_{s,j}}{w_j} + \frac{\gamma}{C}\mathbb{1}_{s \in \mathcal{C}}, \tag{44}$$

where $\mathbb{1}_{s \in \mathcal{C}}$ is an indicator function which has the value 1 if $s \in \mathcal{C}$; 0 otherwise. In this way, the strategies in the covering set are more likely to be chosen than others. As a result, SpecWatch-III can explore all channels more quickly, and thus reveal the best channels sooner, which expedites the exploration for the best strategy.

Choosing strategy Z_j and monitoring spectrum. This part is the same as SpecWatch-II.

Updating strategy weight $w_{s,j+1}$. For calculating average channel scores, we introduce a new parameter β and have $\bar{f}'_{k,j} = \frac{\bar{f}_{k,j} + \beta}{q_{k,j}}$, where $q_{k,j}$ is the channel probability of $k \in \mathcal{K}$ in batch j. By (44), we have is $q_{k,j} = \sum_{s: k \in s} p_{s,j} = (1-\gamma)\frac{\sum_{s:k\in s} w_{s,j}}{w_j} + \frac{\gamma C_k}{C}$, where C_k is the number of strategies in the covering strategy set and containing channel k, i.e., $C_k = |\{s | s \in \mathcal{C} \wedge k \in s\}|$.

The average channel score use $q_{k,j}$ as the denominator in order to compensate the rewards of channels with low probabilities. Among the channels receiving rewards, those with lower probabilities can obtain higher average channel scores, and therefore higher channel weights. Note that we could also gain rewards on an unmonitored channel if we had monitored it, which indicates that the average channel score of that channel should be positive. With this concern, we use parameter β to reduce the bias between monitored and unmonitored channels.

Then the channel weight of k is

$$h_{k,j+1} = h_{k,j} \exp\left(\eta \bar{f}'_{k,j-1}\right), \tag{45}$$

where $h_{k,1} = 1$ for all $k \in \mathcal{K}$. Thus, the strategy weight of s is updated by

$$w_{s,j+1} = w_{s,j} \exp(\eta \bar{g}'_{s,j}), \tag{46}$$

where $\bar{g}'_{s,j}$ is the average strategy score of s in batch j and can be calculated directly by

$$\bar{g}'_{s,j} = \sum_{k \in s} \frac{\frac{1}{\tau}\sum_{i=1}^{\tau} f_{k,[j]+i} + \beta}{(1-\gamma)\frac{\sum_{s:k\in s} w_{s,j}}{w_j} + \frac{\gamma C_k}{C}}. \tag{47}$$

6.1.2. Performance analysis

Since SpecWatch-III only update the monitoring strategy across batches, each batch can be regarded as a round in conventional MAB. We define $\bar{g}_{\mathrm{III}} \stackrel{\mathbf{def}}{=} \sum_{j=1}^{J} \bar{g}_{Z_j,j}$, where Z_j is SpecWatch-III's chosen strategy for each batch. Then we have the following lemma.

We first introduce the following two notations,

$$\bar{F}_{k,n} \stackrel{\mathbf{def}}{=} \sum_{j=1}^{n} \bar{f}_{k,j} \quad \text{and} \quad \bar{F}'_{k,n} \stackrel{\mathbf{def}}{=} \sum_{j=1}^{n} \bar{f}'_{k,j} \quad \text{for} \quad k \in \mathcal{K}, \tag{48}$$

where n is an arbitrary batch, $\bar{f}_{k,j} = \frac{1}{\tau}\sum_{i=1}^{\tau} f_{k,[j]+i}$ and $\bar{f}'_{k,j} = \frac{\bar{f}_{k,j} + \beta}{q_{k,j}}$.

We then prove the following lemma.

Lemma 2. *For any type of adversaries, for any $\delta \in (0, 1)$, $\beta \in (0, 1)$ and $k \in \mathcal{K}$ in SpecWatch-III, we have*

$$\Pr[\bar{F}_{k,n} \geq \bar{F}'_{k,n} + \frac{1}{\beta}\ln\frac{K}{\delta}] \leq \frac{\delta}{K}. \tag{49}$$

Proof. We prove (49) based on [57, Lemma 2]. Note that $\bar{f}'_{k,j}$ for different batches are generated independently and $\bar{F}_{k,n}$ is the sum of these independent random variables. By the Chernoff bound, we have

$$\Pr\left[\bar{F}_{k,n} \geq \bar{F}'_{k,n} + u\right] \leq \exp(-uv)\mathbf{E}\left[\exp\left(v\left(\bar{F}_{k,n} - \bar{F}'_{k,n}\right)\right)\right],$$

for any $k \in \mathcal{K}$, any $u > 0$, and any $v > 0$. Let $u = \frac{1}{\beta}\ln\frac{K}{\delta}$ and $v = \beta$, then we have

$$\begin{aligned}
&\exp(-uv)\mathbf{E}\left[\exp\left(v\left(\bar{F}_{k,n} - \bar{F}'_{k,n}\right)\right)\right] \\
&= \exp(-\ln\frac{K}{\delta})\mathbf{E}\left[\exp\left(\beta\left(\bar{F}_{k,n} - \bar{F}'_{k,n}\right)\right)\right] \\
&= \frac{\delta}{K}\mathbf{E}\left[\exp\left(\beta\left(\bar{F}_{k,n} - \bar{F}'_{k,n}\right)\right)\right].
\end{aligned}$$

Thus, to prove (49), it suffices to prove that for all n,

$$\mathbf{E}\left[\exp\left(\beta\left(\bar{F}_{k,n} - \bar{F}'_{k,n}\right)\right)\right] \leq 1.$$

Define

$$D_n \stackrel{\text{def}}{=} \exp\left(\beta\left(\bar{F}_{k,n} - \bar{F}'_{k,n}\right)\right).$$

We first show that $\mathbf{E}_n[D_n] \leq D_{n-1}$ for $n \geq 2$, where $\mathbf{E}_n$ denotes the conditional expectation $\mathbf{E}[\cdot|Z_1, Z_2, \ldots, Z_{n-1}]$. Note that

$$D_n = D_{n-1}\exp\left(\beta\left(\bar{f}_{k,n} - \bar{f}'_{k,n}\right)\right).$$

Taking conditional expectations, we obtain

$$\begin{aligned}
&\mathbf{E}_n[D_n] \\
&= D_{n-1}\mathbf{E}_n\left[\exp\left(\beta\left(\bar{f}_{k,n} - \bar{f}'_{k,n}\right)\right)\right] \\
&= D_{n-1}\mathbf{E}_n\left[\exp\left(\beta\left(\bar{f}_{k,n} - \frac{\mathbb{1}_{k\in Z_n}\bar{f}_{k,n} + \beta}{q_{k,n}}\right)\right)\right] \\
&= D_{n-1}\exp\left(-\frac{\beta^2}{q_{k,n}}\right) \\
&\quad \cdot \mathbf{E}_n\left[\exp\left(\beta\left(\bar{f}_{k,n} - \frac{\mathbb{1}_{k\in Z_n}\bar{f}_{k,n}}{q_{k,n}}\right)\right)\right] \\
&\leq D_{n-1}\exp\left(-\frac{\beta^2}{q_{k,n}}\right)\mathbf{E}_n\Bigg[1 + \beta\left(\bar{f}_{k,n} - \frac{\mathbb{1}_{k\in Z_n}\bar{f}_{k,n}}{q_{k,n}}\right) \\
&\quad + \beta^2\left(\bar{f}_{k,n} - \frac{\mathbb{1}_{k\in Z_n}\bar{f}_{k,n}}{q_{k,n}}\right)^2\Bigg] \qquad (50) \\
&= D_{n-1}\exp\left(-\frac{\beta^2}{q_{k,n}}\right)\mathbf{E}_n\left[1 + \beta^2\left(\bar{f}_{k,n} - \frac{\mathbb{1}_{k\in Z_n}\bar{f}_{k,n}}{q_{k,n}}\right)^2\right] \qquad (51) \\
&\leq D_{n-1}\exp\left(-\frac{\beta^2}{q_{k,n}}\right)\mathbf{E}_n\left[1 + \beta^2\left(\frac{\mathbb{1}_{k\in Z_n}\bar{f}_{k,n}}{q_{k,n}}\right)^2\right] \\
&\leq D_{n-1}\exp\left(-\frac{\beta^2}{q_{k,n}}\right)\left(1 + \frac{\beta^2}{q_{k,n}}\right) \\
&\leq D_{n-1}, \qquad (52)
\end{aligned}$$

where $\mathbb{1}_{k\in Z_n} = 1$ if $k \in Z_n$ and 0 otherwise; (50) holds because $\beta < 1$, $\bar{f}_{k,n} - \frac{\mathbb{1}_{k\in Z_n}\bar{f}_{k,n}}{q_{k,n}} \leq 1$ and $e^x \leq 1 + x + x^2$ for $x \leq 1$; (51) follows from $\mathbf{E}_t\left[\frac{\mathbb{1}_{k\in Z_n}\bar{f}_{k,n}}{q_{k,n}}\right] = \bar{f}_{k,n}$; (52) holds by the inequality $1 + x \leq e^x$. Taking expectations on both sides proves

$$\mathbf{E}[D_n] \leq \mathbf{E}[D_{n-1}].$$

A similar approach shows that $\mathbf{E}[D_1] \leq 1$, which implies $\mathbf{E}[D_n] = \mathbf{E}\left[\exp\left(\beta\left(\bar{F}_{k,n} - \bar{F}'_{k,n}\right)\right)\right] \leq 1$ as desired. □

Lemma 3. *For any type of adversaries, any $T > 0$, $\gamma \in (0, 1/2)$, $\tau \in [1, T]$, $\beta \in (0, 1)$, and $\eta > 0$ satisfying $2\eta lC \leq \gamma$, we have*

$$\bar{G}_{best} - \bar{G}_{\mathrm{III}} \leq \gamma J + 2\eta lCJ + \frac{l}{\beta}\ln\frac{K}{\delta} + \frac{\ln S}{\eta} + \beta KJ.$$

with probability at least $1 - \delta$ for any $\delta \in (0, 1)$.

Proof. First, we show $0 \leq \eta\bar{g}_{s,j} \leq 1$. It is easy to notice that $\eta\bar{g}'_{s,j} \geq 0$. In addition, we have

$$\begin{aligned}
\eta\bar{g}'_{s,j} &= \eta\sum_{k\in s}\bar{f}'_{k,j} \leq \eta\sum_{k\in s}\frac{\bar{f}_{k,j} + \beta}{\bar{q}_{s,j}} \\
&\leq \eta\sum_{k\in s}\frac{1+\beta}{\frac{\gamma}{C}} \leq \frac{(1+\beta)\eta lC}{\gamma} \leq 1,
\end{aligned}$$

where the second inequality is due to $q_{k,j} \geq \frac{\gamma}{C}$ for all $k \in \mathcal{K}$, and the last inequality is due to $2\eta lC \leq \gamma$.

Then we analyze $\frac{W_{j+1}}{W_j}$. For any sequence $Z_1, \ldots, Z_j$ generated by SpecWatch-III, we have

$$\begin{aligned}
\frac{W_{j+1}}{W_j} &= \sum_{s\in\mathcal{S}}\frac{w_{s,j+1}}{W_j} \\
&= \sum_{s\in\mathcal{S}}\frac{w_{s,j}}{W_j}\exp\left(\eta\bar{g}'_{s,j}\right) \\
&= \sum_{s\in\mathcal{S}}\frac{p_{s,j} - \frac{\gamma}{C}\mathbb{1}_{s\in\mathcal{C}}}{1-\gamma}\exp\left(\eta\bar{g}'_{s,j}\right) \qquad (53) \\
&\leq \sum_{s\in\mathcal{S}}\frac{p_{s,j} - \frac{\gamma}{C}\mathbb{1}_{s\in\mathcal{C}}}{1-\gamma}\left(1 + \eta\bar{g}'_{s,j} + \eta^2\left(\bar{g}'_{s,j}\right)^2\right) \qquad (54) \\
&\leq \frac{1-\gamma}{1-\gamma} + \sum_{s\in\mathcal{S}}\frac{p_{s,j}}{1-\gamma}\left(\eta\bar{g}'_{s,j} + \eta^2\left(\bar{g}'_{s,j}\right)^2\right) \\
&\leq 1 + \frac{\eta}{1-\gamma}\sum_{s\in\mathcal{S}}p_{s,j}\bar{g}'_{s,j} + \frac{\eta^2}{1-\gamma}\sum_{s\in\mathcal{S}}p_{s,j}\left(\bar{g}'_{s,j}\right)^2, \qquad (55)
\end{aligned}$$

where (53) uses the definition of $p_{s,j}$ in (44), and (54) holds by the fact that $e^x \leq 1 + x + x^2$ for $0 \leq x \leq 1$.

Next we bound (55). For the second term, we have

$$\begin{aligned}
\sum_{s\in\mathcal{S}}p_{s,j}\bar{g}'_{s,j} &= \sum_{s\in\mathcal{S}}\left(p_{s,j}\sum_{k\in s}\bar{f}'_{k,j}\right) \\
&= \sum_{k\in\mathcal{K}}\left(\bar{f}'_{k,j}\sum_{s:k\in s}p_{s,j}\right) \\
&= \sum_{k\in\mathcal{K}}\left(\bar{f}'_{k,j}q_{k,j}\right) \\
&= \sum_{k\in\mathcal{K}}\left(\bar{f}_{k,j} + \beta\right) \\
&= \sum_{k\in Z_j}\left(\bar{f}_{k,j} + \beta\right) + \sum_{k\in\mathcal{K}\setminus Z_j}\left(\bar{f}_{k,j} + \beta\right) \\
&= \bar{g}_{Z_j,j} + K\beta, \qquad (56)
\end{aligned}$$

where (56) uses the definition of average strategy reward and the fact that $\bar{f}_{k,j}$ is 0 when $k \notin Z_j$.

For the second sum,

$$\sum_{s\in\mathcal{S}} p_{s,j}\left(\bar{g}'_{s,j}\right)^2 = \sum_{s\in\mathcal{S}}\left(p_{s,j}\left(\sum_{k\in s}\bar{f}'_{k,j}\right)^2\right)$$

$$\leq \sum_{s\in\mathcal{S}}\left(p_{s,j}\cdot l\cdot\sum_{k\in s}\left(\bar{f}'_{k,j}\right)^2\right) \tag{57}$$

$$\begin{aligned}
&= l\cdot\sum_{k\in\mathcal{K}}\left(\left(\bar{f}'_{k,j}\right)^2\sum_{s:k\in s}p_{s,j}\right)\\
&= l\cdot\sum_{k\in\mathcal{K}}\left(\left(\bar{f}'_{k,j}\right)^2 q_{k,j}\right)\\
&= l\cdot\sum_{k\in\mathcal{K}}\left(\bar{f}'_{k,j}\cdot\frac{\mathbb{1}_{k\in Z_j}\bar{f}_{k,j}+\beta}{q_{k,j}}\cdot q_{k,j}\right)\\
&\leq l\cdot(1+\beta)\cdot\sum_{k\in\mathcal{K}}\bar{f}'_{k,j}\\
&\leq l\cdot(1+\beta)\cdot\sum_{s\in\mathcal{C}}\bar{g}'_{s,j},
\end{aligned} \tag{58}$$

where (57) holds as a special case of the Cauchy–Schwarz Inequality $\left(\sum_{i=1}^{n} a_i\cdot 1\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} 1^2\right)$, and (58) holds because covering strategy set $\mathcal{C}$ covers each channel at least once.

Therefore, combining (55), (56), and (58), we have

$$\frac{W_{j+1}}{W_j} \leq 1+\frac{\eta}{1-\gamma}\left(\bar{g}_{Z_j,j}+K\beta\right)+\frac{\eta^2 l(1+\beta)}{1-\gamma}\sum_{s\in\mathcal{C}}\bar{g}'_{s,j}.$$

Taking the log of both sides and using $1+x\leq e^x$ gives

$$\ln\frac{W_{j+1}}{W_j} \leq \frac{\eta}{1-\gamma}\left(\bar{g}_{Z_j,j}+K\beta\right)+\frac{\eta^2 l(1+\beta)}{1-\gamma}\sum_{s\in\mathcal{C}}\bar{g}'_{s,j}.$$

Summing over j we then get

$$\ln\frac{W_{J+1}}{W_1} \leq \frac{\eta}{1-\gamma}\left(\bar{G}_{\mathrm{III}}+JK\beta\right)+\frac{\eta^2 l(1+\beta)}{1-\gamma}\sum_{j=1}^{J}\sum_{s\in\mathcal{C}}\bar{g}'_{s,j}.$$

Note that

$$\sum_{j=1}^{J}\sum_{s\in\mathcal{C}}\bar{g}'_{s,j} \leq C\max_{s\in\mathcal{S}}\sum_{j=1}^{J}\bar{g}'_{s,j} \leq C\max_{s\in\mathcal{S}}\bar{G}'_{s,J}.$$

We have

$$\ln\frac{W_{J+1}}{W_1} \leq \frac{\eta}{1-\gamma}\left(\bar{G}_{\mathrm{III}}+JK\beta\right)+\frac{\eta^2 l(1+\beta)C}{1-\gamma}\max_{s\in\mathcal{S}}\bar{G}'_{s,J}. \tag{59}$$

Now we consider the lower bound of $\ln\frac{W_{J+1}}{W_1}$.

For any strategy s,

$$\begin{aligned}
\ln\frac{W_{J+1}}{W_1} &\geq \ln\frac{w_{s,J+1}}{W_1}\\
&= \ln\frac{w_{s,1}\exp\left(\eta\sum_{j=1}^{J}\bar{g}'_{s,j}\right)}{Sw_{s,1}}\\
&= \eta\sum_{j=1}^{J}\bar{g}'_{s,j}-\ln S\\
&= \eta\bar{G}'_{s,J}-\ln S.
\end{aligned}$$

Since the above inequality holds for any strategy s, we get

$$\ln\frac{W_{J+1}}{W_1} \geq \eta\max_{s\in\mathcal{S}}\bar{G}'_{s,J}-\ln S. \tag{60}$$

Combining (59) and (60), we have

$$\bar{G}_{\mathrm{III}} \geq (1-\gamma-\eta l(1+\beta)C)\max_{s\in\mathcal{S}}\bar{G}'_{s,J}-\frac{1-\gamma}{\eta}\ln S-JK\beta. \tag{61}$$

Note that

$$\bar{G}'_{s,J}=\sum_{j=1}^{J}\bar{g}'_{s,j}=\sum_{j=1}^{J}\sum_{k\in s}\bar{f}'_{k,j}=\sum_{k\in s}\bar{F}'_{k,n},$$

and that

$$\bar{G}_{s,J}=\sum_{j=1}^{J}\bar{g}_{s,j}=\sum_{j=1}^{J}\sum_{k\in s}\bar{f}_{k,j}=\sum_{k\in s}\bar{F}_{k,n}.$$

By using Lemma (49) and applying Boole's inequality, we obtain that, with probability at least $1-\delta$,

$$\begin{aligned}
\bar{G}_{\mathrm{III}} &\geq (1-\gamma-\eta l(1+\beta)C)\left(\max_{s\in\mathcal{S}}\bar{G}_{s,J}-\frac{l}{\beta}\ln\frac{K}{\delta}\right)\\
&\quad -\frac{1-\gamma}{\eta}\ln S-JK\beta\\
&\geq (1-\gamma-\eta l(1+\beta)C)\left(\bar{G}_{best}-\frac{l}{\beta}\ln\frac{K}{\delta}\right)\\
&\quad -\frac{1-\gamma}{\eta}\ln S-JK\beta,
\end{aligned}$$

where $1-\gamma-\eta l(1+\beta)C>0$ because $\eta l(1+\beta)C\leq 2\eta lC\leq\gamma<1/2$.

Therefore,

$$\begin{aligned}
\bar{G}_{best}-\bar{G}_{\mathrm{III}} \leq\ & (\gamma+\eta l(1+\beta)C)\bar{G}_{best}\\
&+(1-\gamma-\eta l(1+\beta)C)\frac{l}{\beta}\ln\frac{K}{\delta}+\frac{1-\gamma}{\eta}\ln S+JK\beta\\
\leq\ & \gamma J+2\eta lCJ+\frac{l}{\beta}\ln\frac{K}{\delta}+\frac{\ln S}{\eta}+\beta KJ,
\end{aligned}$$

where the last inequity is due to the fact that $\bar{G}_{best}\leq J$. □

Next, we bound the difference between the cumulative strategy reward of the best fixed algorithm and that of SpecWatch-III.

Theorem 2. *For any type of adversaries, with probability at least $1-\delta$, the weak regret of SpecWatch-III is bounded by $O(T^{\frac{2}{3}})$. In particular, choosing $\tau=B_\tau T^{\frac{1}{3}}\in[1,T]$, $\gamma=B_\gamma T^{-\frac{1}{3}}\in(0,\frac{1}{2})$, $\beta=B_\beta T^{-\frac{1}{3}}\in(0,1)$, and $\eta=\frac{B_\gamma}{2lC}T^{-\frac{1}{3}}$, where B_τ, B_γ, and B_β are constants, we have*

$$R_{\mathrm{III}} \leq \left(2B_\gamma+B_\beta K+B_\tau\left(\frac{l\ln\frac{K}{\delta}}{B_\beta}+\frac{\ln S}{B_\eta}\right)+\frac{1}{B_\tau}\right)T^{\frac{2}{3}}. \tag{62}$$

Proof. Similar to the proof of Theorem 1, we have

$$\begin{aligned}
R_{\mathrm{II}} &= G_{best}-L_{best}-G_{\mathrm{II}}+L_{\mathrm{II}}\\
&\leq G_{best}-G_{\mathrm{II}}+J\\
&= \tau\bar{G}_{best}-\tau\bar{G}_{\mathrm{II}}+\frac{T}{\tau}\\
&\leq \tau\left(\gamma J+2\eta lCJ+\frac{l}{\beta}\ln\frac{K}{\delta}+\frac{\ln S}{\eta}+\beta KJ\right)+\frac{T}{\tau},\\
&\leq (\gamma+2\eta lC+\beta K)T+\tau\left(\frac{l}{\beta}\ln\frac{K}{\delta}+\frac{\ln S}{\eta}\right)+\frac{T}{\tau},
\end{aligned} \tag{63}$$

with probability at least $1-\delta$. The last inequality follows from Lemma 3. Plugging in the value of parameters finishes the proof. □

We now provide an example choice of parameters to reach a specific bound.

Corollary 2. *For any type of adversaries, under the condition of*

$$T \geq \max\left\{B^2, \frac{8(lC\ln S)^{3/2}}{B}, \frac{\left(\frac{l}{K}\ln\frac{K}{\delta}\right)^{3/2}}{B}\right\},$$

using parameters $\tau = B^{-\frac{2}{3}}T^{\frac{1}{3}}, \gamma = \sqrt{lC\ln S}\cdot B^{-\frac{1}{3}}T^{-\frac{1}{3}}, \beta = \sqrt{\frac{l}{K}\ln\frac{K}{\delta}}\cdot B^{-\frac{1}{3}}T^{-\frac{1}{3}}$*, and* $\eta = \sqrt{\frac{\ln S}{4lC}}\cdot B^{-\frac{1}{3}}T^{-\frac{1}{3}}$*, where* $B = 4\sqrt{lC\ln S} + 2\sqrt{lK\ln\frac{K}{\delta}}$*, we have*

$$R_{\mathrm{III}} \leq 2\left(4\sqrt{lC\ln S} + 2\sqrt{lK\ln\frac{K}{\delta}}\right)^{\frac{2}{3}} T^{\frac{2}{3}}, \tag{64}$$

with probability at least $1-\delta$*.*

Proof. Substituting the parameters in (62), we have the immediate result. □

Remark. Note that it is not guaranteed that SpecWatch-III always outperforms SpecWatch-II. The improvement over SpecWatch-II is the fact that SpecWatch-III guarantees the actual weak regret to be bounded with any predefined confidence level. Moreover, when choosing parameters as those in Corollaries 1 and 2, SpecWatch-III has a much better bound than SpecWatch-II since the term $S\ln S$ is removed. Recall that

$$\mathrm{E}[R_{\mathrm{II}}] \leq 3\left(\frac{1}{2}S\ln S\right)^{\frac{1}{3}} T^{\frac{2}{3}}$$

while

$$\Pr\left[R_{\mathrm{III}} \leq 2\left(4\sqrt{lC\ln S} + 2\sqrt{lK\ln\frac{K}{\delta}}\right)^{\frac{2}{3}} T^{\frac{2}{3}}\right] \leq 1-\delta.$$

This could be a huge improvement when $S \gg K$ because S is exponential to K while C is no larger than K.

7. Performance evaluation

We conduct extensive simulations to demonstrate the performance of our proposed online spectrum monitoring algorithms, SpecWatch-I, SpecWatch-II, and SpecWatch-III. We first show the convergence of normalized weak regrets of all three algorithms and then study and compare their performances under different adversary settings. We also discuss how algorithm parameters impact the performance of proposed algorithms. Last but not least, we demonstrate the impact of the detection probability, the number of radios, the number of MUs, and adversary settings, on the algorithm performance.

In the simulation setting, we consider $K = 10$ channels, and we deploy a monitor with $l = 2$ radios. We set the unit reward of successfully detecting on a single channel to be $r = 0.3$ and the unit switching cost of tuning one radio to be $c = 0.03$. If not specified, the detection probability of each radio is set to be $p_d = 0.9$ as it is the recommended detection accuracy in consistent with [58]. The parameters of all algorithms are chosen as in the corollaries. If the monitor uses SpecWatch-III, we set $\delta = 0.5$ so that the weak regret is relatively small with an acceptable confidence level.

We assume there are $m = 2$ MUs attacking channels either obliviously or adaptively. Specifically, we consider four adversary settings,

- Fixed adversary (*Fixed*): Each MU selects a fixed channel and never switches throughout the time horizon T.
- Uniform adversary (*Uniform*): In every timeslot, each MU selects a channel uniformly at random.
- Normal adversary (*Normal*): In every timeslot, each MU selects a channel following the same normal distribution.
- Adaptive adversary (*Adaptive*): Each MU adopts modified SpecWatch-I, where the actual channel reward is r if the MU is not captured on that channel, and 0 otherwise.

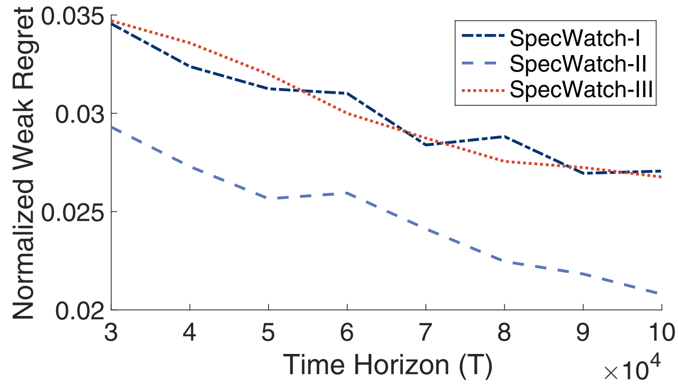


Fig. 3. Convergence of weak regrets.

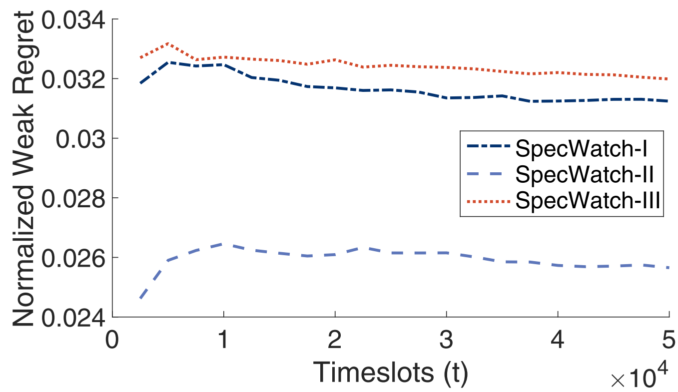


Fig. 4. Algorithm comparison.

The simulation results are shown below, and each of them is averaged over 100 trials.

Weak regret. Fig. 3 shows the normalized weak regrets of all algorithms decrease with time horizon T, which supports our theoretical analysis that the normalized weak regret converges to 0 as $T \to \infty$. In all following simulations, we fix the time horizon to be $T = 50,000$. Here we only show the result with adaptive adversary since in other adversary settings, the results are similar. Note that SpecWatch-II outperforms other algorithms as shown in the figure, but it only means that SpecWatch-II's way to trade-off between exploration and exploitation is more appropriate under our current simulation setting.

Fig. 4 plots how the normalized weak regret decreases with timeslots under adaptive adversary. At the beginning, there is apparent fluctuation of the normalized weak regret. As time goes by, the monitor and the adaptive adversary enter a relatively stable stage, but we can still see the decreasing trend of the normalized weak regret, which indicates our algorithms are learning from monitoring history to make smart decisions.

Impact of algorithm parameters. Among all parameters of the three algorithms, the most important one is the batch size τ, which controls the trade-off between cumulative reward and cumulative switching cost. As shown in Fig. 5, we conducted simulations where the batch size τ was set to be exactly $T^{1/\Delta}$ and plotted how the cumulative utility changes with Δ, under the adaptive adversary. It is shown that all algorithms achieve highest cumulative utility ratio when τ is around $T^{1/3}$, which is in consistency with our theoretical analysis. The performances of all algorithms are almost the same because they are designed based on the same

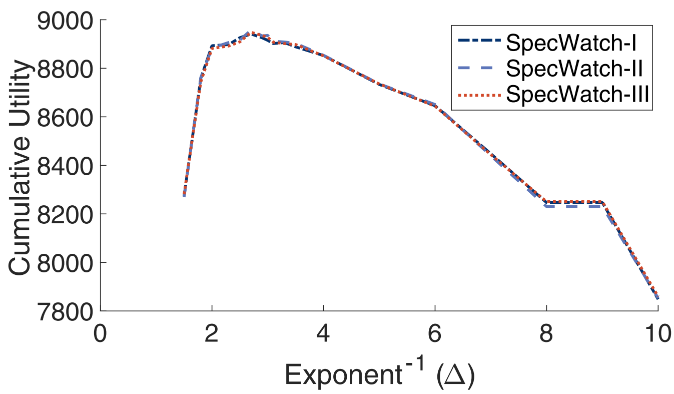


Fig. 5. Impact of batch size on cumulative utility.

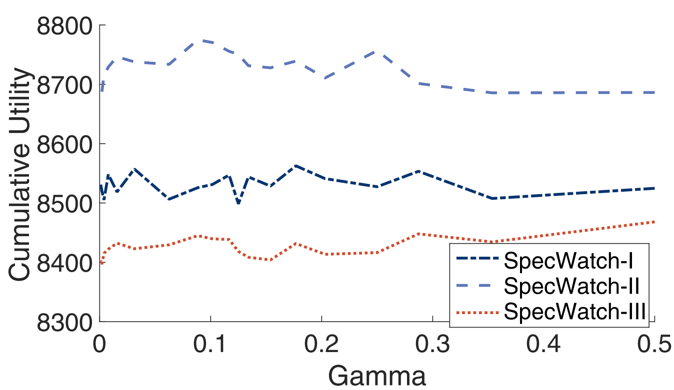


Fig. 6. Impact of parameter γ on cumulative utility.

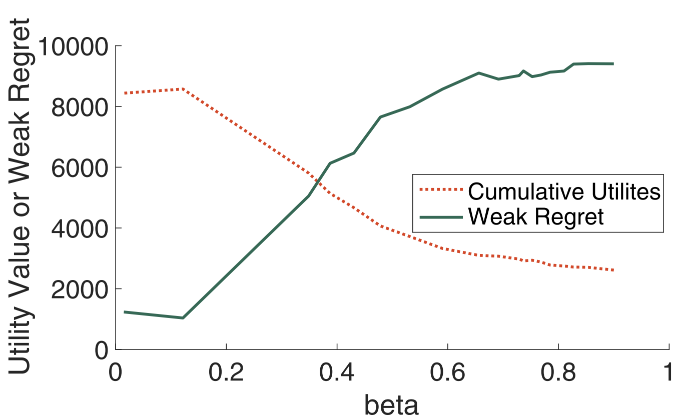


Fig. 7. Impact of parameter β on cumulative utility.

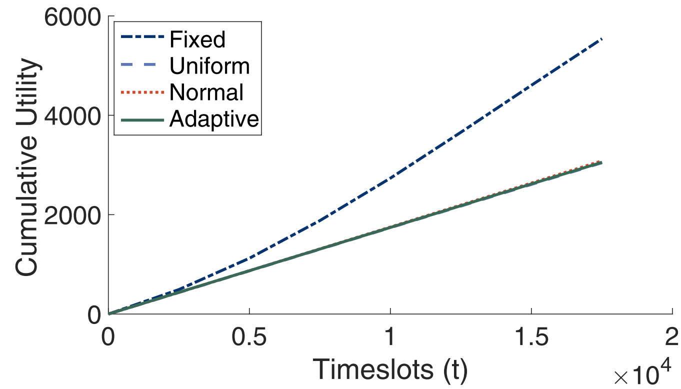


Fig. 8. Cumulative utility under different adversary settings.

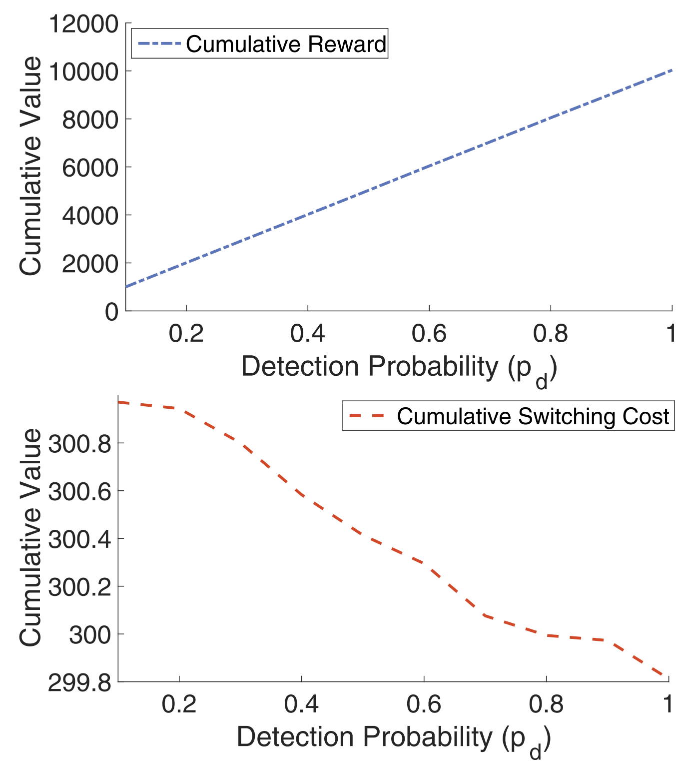


Fig. 9. Impact of detection probability on cumulative reward and switching cost.

framework and using same way to trade-off between rewards and switching costs.

Fig 6 shows unstable cumulative utilities when changing the parameter γ. This implies that there is no universally applicable choice of γ which could balance the exploration and exploitation in the best way. Note that our theorem requires $\eta = \frac{\gamma}{2IC}$, so we do not show how η influences the algorithm performance separately. In addition, it is shown that SpecWatch-III does not achieve better utilities than SpecWatch-II. This is possible because the parameters are not chosen as in the corollaries, and the corollaries only bound the weak regret, but not the cumulative utility.

Fig 7 shows how β, the parameter exclusively used in SpecWatch-III, affects the cumulative utilities and weak regrets. Recall that β is used to make up for the channels not monitored. For all unmonitored channels, we assume their channel rewards are β instead of 0. As shown in the figure, this value is the smaller the better for most of the cases. However, if it is very close to 0, such benefits will diminish.

Cumulative utility. Fig. 8 plots the actual utilities gained by SpecWatch-III. The figures for the other two algorithms are very similar and thus omitted. We observe that the cumulative utilities under fixed adversary greatly exceed the other three settings, and the other three settings have similar results.

Impact of system parameters and adversary settings. In simulations, we fix the time horizon to be $T = 50000$. Since the impacts on all algorithms are similar, we only present results of SpecWatch-III.

Fig. 9 shows the impact of detection probability p_d on the cumulative rewards and the cumulative switching cost.

As expected, the cumulative reward grows with decreasing slope as the detection probability increases. The cumulative switching cost, however, has a decreasing trend. This is because the larger the detection probability, the more accurate for the monitor to evaluate each strategy; thus the best strategy is revealed

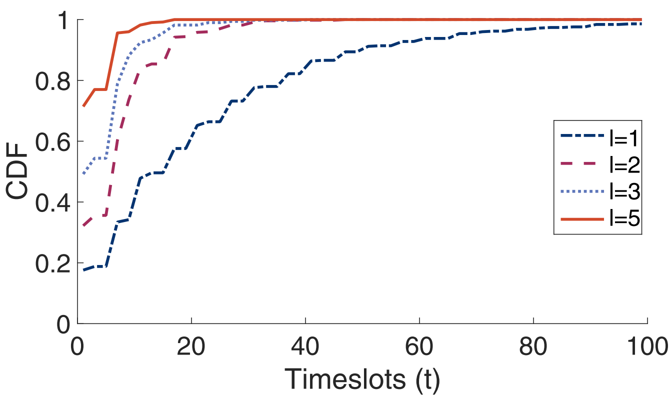


Fig. 10. Number of timeslots to detect the first misuse with different number of radios.

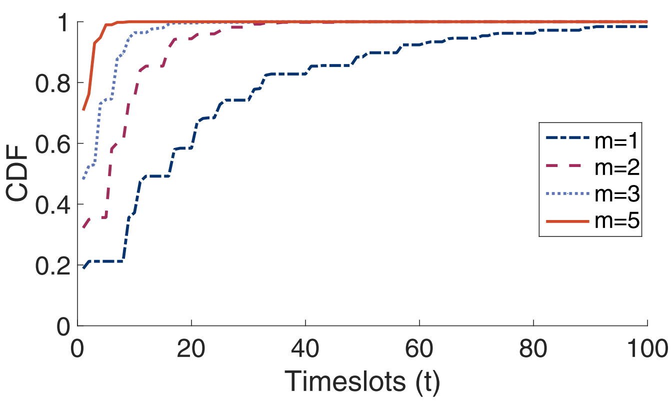


Fig. 11. Number of timeslots to detect the first misuse with different number of MUs.

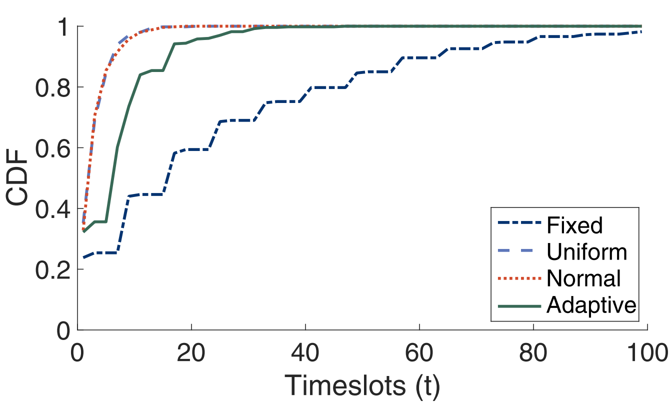


Fig. 12. Number of timeslots to detect the first misuse under different adversary settings.

more quickly, avoiding unnecessary switches and reducing cumulative switching cost.

We also study the impact the number of radios l, the number of MUs m, and the types of adversary on the performance of our algorithms. Figs. 10–12 illustrate the cumulative distribution function (CDF) of expected number of timeslots to detect the first misuse. In general, more radios or more MUs make it sooner for the monitor to detect successfully. In Fig. 12, the monitor takes the longest time to detect the first misuse under fixed adversary setting, which is because the monitor sticks to the same strategy for the whole batch to prevent switching costs. If the monitor does not choose the channels attacked by MUs at the first timeslot in a batch, it will not detect misuse for the following $\tau - 1$ timeslots. As a result, it takes longer time for the monitor to detect the first misuse under fixed adversary setting.

8. Conclusion

In this paper, we studied the adversarial spectrum monitoring problem with unknown statistics by formulating it as an combinatorial adversarial multi-armed bandit problem with switching costs. As far as we know, we are the first to study such a problem. We proposed an online spectrum monitoring framework named SpecWatch and designed two effective online algorithms, SpecWatch-II and SpecWatch-III. The framework generally accomplish the challenge of considering switching costs while different algorithms accomplish the challenges of unknown MUs, imperfection detections, combinatorial strategies and adaptive adversary. We rigorously proved that the weak regrets of two algorithms are bounded by $O(T^{2/3})$, which matches the lower bound of the general MAB-SC problem. Thus, they are asymptotically optimal.

Despite that the two algorithms perform well both in theory and in simulations, they are not perfect. The time complexity of them is exponential of the number of channels. However, if we provide every strategy with a chance to be chosen, then this exponential complexity is inevitable. But future works could be done to strike the balance between the complexity and the overall performance. This is a valuable research direction as our framework and algorithms can be applied to spectrum patrolling problems. Like Li et al. [37], regarding different locations as another type of channel, our algorithms will output the patrolling path instead of sequences of chosen channels. In this case, we must address the combinatorial problem, which is one of our future research directions.

Another research direction could be changing the parameters adaptively. Currently, all parameters will be chosen ahead of some fixed time and fix them until the full time horizon is reached. However, this may not be optimal because the MUs' attacks could be very short, concentrated, and severe. Therefore adaptively changing the parameters could be a potential solution for these scenarios.

Acknowledgments

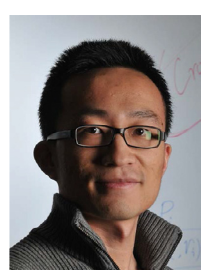
This is an extended and enhanced version of the paper [55] that appeared in INFOCOM 2016. This research was supported in part by NSF grants 1443966, 1444059, 1619728, and 1717315. The information reported here does not reflect the position or the policy of the federal government.

References

[1] F. S. P. T. Force, Report of the Spectrum Efficiency Working Group, Technical Report, 2002.
[2] Q. Zhao, B.M. Sadler, A survey of dynamic spectrum access, IEEE Signal Process. Mag. 24 (3) (2007) 79–89.
[3] C. Santivanez, R. Ramanathan, C. Partridge, R. Krishnan, M. Condell, S. Polit, Opportunistic spectrum access: challenges, architecture, protocols, in: Proc. of WICON, 2006.
[4] A. Anandkumar, N. Michael, A. Tang, Opportunistic spectrum access with multiple users: learning under competition, in: Proc. of INFOCOM, 2010.
[5] C. Tekin, M. Liu, Online learning in opportunistic spectrum access: a restless bandit approach, in: Proc. of INFOCOM, 2011, pp. 2462–2470.
[6] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, Y.-D. Yao, Opportunistic spectrum access in unknown dynamic environment: a game-theoretic stochastic learning solution, IEEE Trans. Wirel. Commun. 11 (4) (2012) 1380–1391.
[7] O. Altrad, S. Muhaidat, A. Al-Dweik, A. Shami, P.D. Yoo, Opportunistic spectrum access in cognitive radio networks under imperfect spectrum sensing, IEEE Trans. Veh. Technol. 63 (2) (2014) 920–925.
[8] R.N. Yadav, R. Misra, U. Gupta, S. Bhagat, Opportunistic spectrum access in CR network in licensed and unlicensed channels, in: Proc. of ICDCN, 2015.
[9] Q. Wang, K. Ren, P. Ning, S. Hu, Jamming-resistant multiradio multichannel opportunistic spectrum access in cognitive radio networks, IEEE Trans. Veh. Technol. 65 (10) (2016) 8331–8344.
[10] G.I. Tsiropoulos, O.A. Dobre, M.H. Ahmed, K.E. Baddour, Radio resource allocation techniques for efficient spectrum access in cognitive radio networks, IEEE Commun. Surv. Tutorials 18 (1) (2016) 824–847.

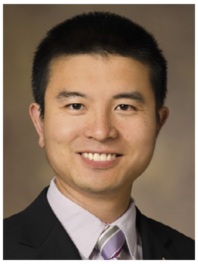
[11] Q. Wang, K. Ren, P. Ning, Anti-jamming communication in cognitive radio networks with unknown channel statistics, in: Proc. of ICNP, 2011, pp. 393–402.
[12] L. Yang, Z. Zhang, B.Y. Zhao, C. Kruegel, H. Zheng, Enforcing dynamic spectrum access with spectrum permits, in: Proc. of MobiHoc, 2012, pp. 195–204.
[13] M.B. Weiss, M. Altamimi, M. McHenry, Enforcement and spectrum sharing: a case study of the 1695–1710 Mhz band, in: Proc. of CROWNCOM, 2013, pp. 7–12.
[14] C. Sorrells, P. Potier, L. Qian, X. Li, Anomalous spectrum usage attack detection in cognitive radio wireless networks, in: Proc. of HST, 2011, pp. 384–389.
[15] G.P. Kumar, D.K. Reddy, Frequency domain techniques for void spectrum detection in cognitive radio network for emulation attack prevention, in: Proc. of CCPCT, 2016, pp. 1–6.
[16] G. Atia, A. Sahai, V. Saligrama, Spectrum enforcement and liability assignment in cognitive radio systems, in: Proc. of DySPAN, 2008, pp. 1–12.
[17] P. Kyasanur, N.H. Vaidya, Detection and handling of MAC layer misbehavior in wireless networks, in: Proc. of DSN, 2003, pp. 173–182.
[18] Y. Zhang, L. Lazos, Countering selfish misbehavior in multi-channel MAC protocols, in: Proc. of INFOCOM, 2013, pp. 2787–2795.
[19] S. Liu, L.J. Greenstein, W. Trappe, Y. Chen, Detecting anomalous spectrum usage in dynamic spectrum access networks, Ad Hoc Netw. 10 (5) (2012) 831–844.
[20] J. Tang, Y. Cheng, Selfish misbehavior detection in 802.11 based wireless networks: an adaptive approach based on Markov decision process, in: Proc. of INFOCOM, 2013, pp. 1357–1365.
[21] X. Jin, J. Sun, R. Zhang, Y. Zhang, C. Zhang, SpecGuard: spectrum misuse detection in dynamic spectrum access systems., in: Proc. of INFOCOM, 2015, pp. 172–180.
[22] A. Dutta, M. Chiang, 'See something, say something' crowdsourced enforcement of spectrum policies, IEEE Trans. Wireless Commun. 15 (1) (2016) 67–80.
[23] P. Baltiiski, I. Iliev, B. Kehaiov, V. Poulkov, T. Cooklev, Long-term spectrum monitoring with big data analysis and machine learning for cloud-based radio access networks, Wirel. Pers. Commun. 87 (3) (2016) 815–835.
[24] L. Chen, S. Iellamo, M. Coupechoux, Opportunistic spectrum access with channel switching cost for cognitive radio networks, in: Proc. of ICC, 2011, pp. 1–5.
[25] P. Auer, N. Cesa-Bianchi, Y. Freund, R.E. Schapire, The nonstochastic multiarmed bandit problem, SIAM J. Comput. 32 (1) (2002) 48–77.
[26] O. Dekel, J. Ding, T. Koren, Y. Peres, Bandits with switching costs: $T^{2/3}$ regret, in: Proc. of STOC, 2014, pp. 459–467.
[27] Y. Song, X. Chen, Y.-A. Kim, B. Wang, G. Chen, Sniffer channel selection for monitoring wireless LANs, in: Proc. of WASA, 2009, pp. 489–498.
[28] J. Yeo, M. Youssef, A. Agrawala, A framework for wireless LAN monitoring and its applications, in: Proc. of WiSe, 2004, pp. 70–79.
[29] Y.-C. Cheng, J. Bellardo, P. Benkö, A.C. Snoeren, G.M. Voelker, S. Savage, Jigsaw: solving the puzzle of enterprise 802.11 analysis, in: Proc. of SIGCOMM, 2006, pp. 39–50.
[30] Y.-C. Cheng, M. Afanasyev, P. Verkaik, P. Benkö, J. Chiang, A.C. Snoeren, S. Savage, G.M. Voelker, Automating cross-layer diagnosis of enterprise wireless networks, in: Proc. of SIGCOMM, 2007, pp. 25–36.
[31] D.-H. Shin, S. Bagchi, Optimal monitoring in multi-channel multi-radio wireless mesh networks, in: Proc. of MobiHoc, 2009, pp. 229–238.
[32] D.-H. Shin, S. Bagchi, C.-C. Wang, Toward optimal sniffer-channel assignment for reliable monitoring in multi-channel wireless networks, in: Proc. of SECON, 2013, pp. 203–211.
[33] H. Nguyen, G. Scalosub, R. Zheng, On quality of monitoring for multichannel wireless infrastructure networks, IEEE Trans. Mobile Comput. 13 (3) (2014) 664–677.
[34] S. Chen, K. Zeng, P. Mohapatra, Efficient data capturing for network forensics in cognitive radio networks, IEEE/ACM Trans. Netw. 22 (6) (2014) 1988–2000.
[35] D.-H. Shin, S. Bagchi, C.-C. Wang, Distributed online channel assignment toward optimal monitoring in multi-channel wireless networks, in: Proc. of INFOCOM, 2012, pp. 2626–2630.
[36] Q. Yan, M. Li, F. Chen, T. Jiang, W. Lou, Y.T. Hou, C.-T. Lu, SpecMonitor: towards efficient passive traffic monitoring for cognitive radio networks, IEEE Trans. Wirel. Commun. 13 (10) (2014) 5893–5905.
[37] J. Li, J. Xu, W. Liu, S. Gong, K. Zeng, Robust optimal spectrum patrolling for passive monitoring in cognitive radio networks, in: Proc. of CIT, 2017, pp. 63–68.
[38] P. Arora, C. Szepesvári, R. Zheng, Sequential learning for optimal monitoring of multi-channel wireless networks, in: Proc. of INFOCOM, 2011, pp. 1152–1160.
[39] R. Zheng, T. Le, Z. Han, Approximate online learning algorithms for optimal monitoring in multi-channel wireless networks, IEEE Trans. Wirel. Commun. 13 (2) (2014) 1023–1033.
[40] T. Le, C. Szepesvari, R. Zheng, Sequential learning for multi-channel wireless network monitoring with channel switching costs, IEEE Trans. Signal Process. 62 (22) (2014) 5919–5929.
[41] S. Yi, K. Zeng, J. Xu, Secondary user monitoring in unslotted cognitive radio networks with unknown models, in: Proc. of WASA, 2012, pp. 648–659.
[42] J. Xu, Q. Wang, R. Jin, K. Zeng, M. Liu, Secondary user data capturing for cognitive radio network forensics under capturing uncertainty, in: Proc. of MILCOM, 2014, pp. 935–941.
[43] P. Auer, N. Cesa-Bianchi, P. Fischer, Finite-time analysis of the multiarmed bandit problem, Mach. Learn. 47 (2–3) (2002) 235–256.
[44] J. Xu, Q. Wang, K. Zeng, M. Liu, W. Liu, Sniffer channel assignment with imperfect monitoring for cognitive radio networks, IEEE Trans. Wirel. Commun. 15 (3) (2016) 1703–1715.
[45] H. Robbins, Some aspects of the sequential design of experiments, Bull. Am. Math. Soc. 58 (5) (1952) 527–535.
[46] T.L. Lai, H. Robbins, Asymptotically efficient adaptive allocation rules, Adv. Appl. Math. 6 (1) (1985) 4–22.
[47] G. Burtini, J. Loeppky, R. Lawrence, A survey of online experiment design with the stochastic multi-armed bandit, Comput. Sci. (2015).
[48] S. Bubeck, N. Cesa-Bianchi, et al., Regret analysis of stochastic and nonstochastic multi-armed bandit problems, Found. Trends® Mach. Learn. 5 (1) (2012) 1–122.
[49] P. Auer, C.-K. Chiang, An algorithm with nearly optimal pseudo-regret for both stochastic and adversarial bandits, in: Proc. of COLT, 2016, pp. 116–120.
[50] G. Yi, B. Krishnamachari, M. Liu, On the combinatorial multi-armed bandit problem with markovian rewards, in: Global Telecommunications Conference, 2011.
[51] S. Tang, Y. Zhou, K. Han, Z. Zhang, J. Yuan, W. Wu, Networked stochastic multi-armed bandits with combinatorial strategies, in: Proc. of ICDCS, 2017, pp. 786–793.
[52] IEEE standard for information technology–telecommunications and information exchange between systems local and metropolitan area networks–specific requirements - part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, IEEE Std 802.11-2016 (2016) 1–3534.
[53] Q. Wang, M. Liu, Learning in hide-and-seek, IEEE/ACM Trans. Netw. 24 (2) (2016) 1279–1292.
[54] R. Arora, O. Dekel, A. Tewari, Online bandit learning against an adaptive adversary: from regret to policy regret, in: Proc. of ICML, 2012, pp. 1503–1510.
[55] M. Li, D. Yang, M. Li, J. Lin, J. Tang, SpecWatch: adversarial spectrum usage monitoring in CRNs with unknown statistics, in: Proc. of INFOCOM, 2016, pp. 1–9.
[56] Google spectrum database, (https://www.google.com/get/spectrumdatabase/channel/). Accessed: 2018-04-13.
[57] A. Gyorgy, T. Linder, G. Lugosi, G. Ottucsak, The on-line shortest path problem under partial monitoring, J. Mach. Learn. Res. 8 (2007) 2369–2403.
[58] J.-K. Choi, S.-J. Yoo, Time-constrained detection probability and sensing parameter optimization in cognitive radio networks, EURASIP J. Wireless Commun. Netw. 2013 (1) (2013) 1–12.

Ming Li (S'15) received the B.S. degree in geochemistry from Peking University, Beijing, China, in 2013 and the M.S. degree in computer science from Colorado School of Mines, Golden, CO, USA, in 2015.She is currently working toward the Ph.D. degree at Colorado School of Mines. Her main research interests include game theory, contract theory, crowdsourcing, smartphone-based sensing systems, visible light communication technologies.

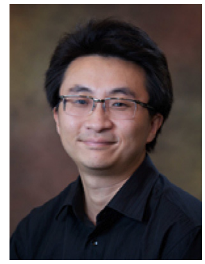


Dejun Yang is the Ben L. Fryrear Assistant Professor of Computer Science at Colorado School of Mines. He received his PhD degree in computer science from Arizona State University in 2013 and BS degree in computer science from Peking University, Beijing, China, in 2007. His research interests include economic and optimization approaches to networks, crowdsourcing, smart grid, and security and privacy. He received Best Paper Awards at IEEE GLOBECOM'2015, IEEE MASS'2011, ICC'2011, 2012, and a Best Paper Award Runner-up at IEEE ICNP'2010. He also served as a TPC member for many conferences including IEEE INFOCOM, ICC and GLOBECOM.

Jian Lin (S'15) received the M.S. and B.S. degree in computer science from Southwest Jiaotong University, Chengdu China, in 2014 and 2011, respectively. He has been working toward the Ph.D. degree in computer science at Colorado School of Mines, Golden, CO, USA, since 2014. His research interests include economic and optimization approaches to networks, mobile crowdsensing and mobile computing.



Ming Li (M'11-SM'17) is an Associate Professor in the Department of Electrical and Computer Engineering of University of Arizona. He was an Assistant Professor in the Computer Science Department at Utah State University from 2011 to 2015. He received his Ph.D. in ECE from Worcester Polytechnic Institute in 2011. His main research interests are wireless and cyber security, with current emphases on cross-layer optimization and machine learning in wireless networks, wireless physical layer security, privacy enhancing technologies, and cyber-physical system security. He received the NSF Early Faculty Development (CAREER) Award in 2014, and the ONR Young Investigator Program (YIP) Award in 2016. He is a senior member of IEEE, and a member of ACM.



Jian Tang (M'06-SM'13) received the Ph.D degree in Computer Science from Arizona State University in 2006. Currently he is a professor in the Department of Electrical Engineering and Computer Science at Syracuse University, Syracuse, NY, USA. His research interests lie in the areas of Wireless Networking, Deep Learning, Big Data and Cloud Computing. Dr. Tang has published over 100 papers in premier journals and conferences. He received an NSF CAREER award in 2009, the 2016 Best Vehicular Electronics Paper Award from IEEE Vehicular Technology Society, and Best Paper Awards from ICC (2014) and GLOBECOM (2015) respectively. He has been an editor for IEEE Transactions on Network Science and Engineering since 2017, for IEEE Transactions on Wireless Communications since 2016, for IEEE Internet of Things Journal since 2013, and for IEEE Transactions on Vehicular Technology in 2010–2017. He served as a TPC co-chair for IEEE International Conference on Internet of Things (iThings) (2015) and for ICNC (2016).