Dynamic Spectrum Access in Non-Stationary Environments: A Thompson Sampling Based Method

Shuai Ye, Tianyu Wang, and Shaowei Wang
School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China
E-mail: dz20230027@smail.nju.edu.cn, {tianyu.alex.wang, wangsw}@nju.edu.cn

GLOBECOM 2022 - 2022 IEEE Global Communications Conference | 978-1-6654-3540-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/GLOBECOM48099.2022.10000684

***Abstract*—In dynamic spectrum access (DSA), unlicensed secondary users can estimate the idle probability of each primary channel by using historical sensing results and access the channel with the highest idle probability for opportunistic transmission. Most of the existing works assume that each primary channel is associated with a constant idle probability, which can be accurately estimated by sensing the channel multiple times. However, due to the rapid traffic change and irregular user mobility, primary channels can be highly dynamic and the associated idle probability is generally time-varying. In this paper, we consider DSA in non-stationary environments where the idle probabilities of primary channels vary with time. Specifically, we propose a DSA scheme based on the Thompson sampling method with a change-detection technique, which is capable of detecting the variation of channel statistics and adjusting the channel access strategy accordingly. Numerical results show that the proposed algorithm outperforms the existing algorithms in terms of successful transmission ratio in various settings.**

***Index Terms*—Dynamic spectrum access, multi-armed bandit, non-stationary environments, Thompson sampling.**

I. Introduction

Due to the static spectrum management policy, a large portion of the spectrum is underutilized. While at the same time, there is an increasing demand for spectrum resources due to the emerging wireless applications such as machine-type communications and vehicular communications [1]. Therefore, to improve spectrum utilization, a dynamic spectrum management policy is considered for future wireless networks, where unlicensed secondary users (SUs) can opportunistically access the network when licensed primary users are absent, and it is referred to as the dynamic spectrum access (DSA) [2].

One of the key challenges of DSA is to find the primary channel with the highest idle probability such that the potential transmission opportunities for the SU are maximized. However, due to the hardware limitations, only a limited number of channels can be sensed in each time slot [3], [4]. Therefore, the SU needs to determine the current best primary channel with incomplete sensing observations. In the literature, such an online decision process is formulated by using the multi-armed bandit (MAB) model [5]–[9]. In the MAB model, the SU is seen as a player in front of a bandit machine with multiple arms. Each arm represents a specific primary channel with an unknown idle probability. In each round, the SU pulls an arm and receives a reward if the corresponding channel is idle at the corresponding slot. The goal of the SU is to maximize its total reward within a given number of rounds.

In [5], the time slots are split into staggered exploration and exploitation slots. In the exploration slots, the SU randomly chooses a primary channel to access. In the exploitation slots, the SU accesses the channel having the largest number of idle observations in the previous exploration slots. In [6], the upper confidence bound (UCB) algorithm is introduced to efficiently balance the exploitation of the currently best channel and the exploration of potential better channels. The authors also consider the multi-SU scenario and propose a pre-agreement scheme to schedule the SUs in a round-robin fashion. In [7], a modified UCB algorithm is proposed, where the SU stays on a channel until an idle state is observed and then the UCB index is updated according to the sequence of sensing observations. In [8], the UCB algorithm is introduced for each SU and a rank-based scheduling scheme is proposed to orthogonalize multiple SUs on different channels. In [9], a modified UCB algorithm is proposed to reduce the number of suboptimal accesses, which converges more quickly as compared to the classical UCB algorithm.

The existing works only consider a stationary environment where the idle probability of each channel is time-invariant. However, in practical networks, the primary channels can be highly dynamic due to the rapid change of primary traffic and the irregular mobility of both primary and secondary users. The primary channels are generally non-stationary and the associated idle probabilities are time-variant. Therefore, the existing DSA algorithms based on stationary models may suffer from severe performance degradation when the changes in channel statistics are not detected in time.

In this paper, we consider DSA in a non-stationary environment, where the state of each primary channel follows a non-stationary Bernoulli process with a time-varying idle probability. Specifically, we formulate a non-stationary MAB problem and propose a Thompson sampling with change

This work was partially supported by the National Natural Science Foundation of China 61931023 and U1936202.

detection (TSCD) method to track the network dynamics. The proposed TSCD algorithm does not need prior knowledge of the channel statistics and achieves an efficient tradeoff between the exploration and exploitation of primary channels. Numerical results show that the proposed TSCD algorithm improves the successful transmission ratio as compared to the existing algorithms in various network settings.

II. System Model

We consider a slotted DSA network with $\mathcal{K} = \{1, 2, \ldots, K\}$ independent channels. The entire time horizon is given by T. In any slot $t \in [1, T]$, we denote the instant state of channel k as $s_k(t) \in \{0, 1\}$, where $s_k(t) = 1$ represents the channel is idle and $s_k(t) = 0$ represents the channel is occupied by PUs. We assume that there is only one single SU and the SU seeks transmission opportunities in every slot.

Considering the hardware capabilities and resource constraints, we assume the SU can sense only one channel in each slot and the sensing result is always correct. We denote the sensing action of the SU at slot t as $a_t \in \mathcal{K}$ and the corresponding sensing result is given by $s_{a_t}(t)$. If $s_{a_t}(t) = 0$, the SU does not transmit and waits for the next slot. If $s_{a_t}(t) = 1$, the SU transmits on channel a_t.

A. Non-Stationary Environment

We assume the time horizon is split into V piecewise-stationary segments by slots $\phi_1, \phi_2, \ldots, \phi_{V-1}$. In each segment $v \in [1, V]$, the idle probability of channel k is unchanged and denoted by $p_{k,v}$. For simplicity, we set $\phi_0 = 0$ and $\phi_V = T$. Thus, for any channel $k \in \mathcal{K}$ in segment v, we have

$$s_k(t) = \begin{cases} 1, & \text{with probability} \quad p_{k,v} \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $\phi_{v-1} < t \leq \phi_v$. Note that $p_{k,v}$ and ϕ_v are unknown to the SU. Also, we define the average idle probability as $\lambda = \sum_{k=1}^{K} \sum_{v=1}^{V} p_{k,v}/KV$, which reflects the average traffic load of all primary channels.

B. Problem Formulation

Since $s_{a_t}(t)$ determines whether the SU can successfully transmit at slot t, the successful transmission ratio (STR) of the SU is then given by

$$\text{STR} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[s_{a_t}(t)]. \tag{2}$$

An ideal DSA policy is to sense and access the channel with the highest $p_{k,v}$ in each segment v. However, due to the lack of information of $p_{k,v}$ and ϕ_v, no practical policy can achieve the ideal performance. Here, we aim to maximize the STR while considering the practical constraints of the SU.

We note that the considered DSA problem can be formulated as a non-stationary MAB problem. The SU is the player that pulls an arm a_t and gets a reward $r(t) = s_{a_t}(t)$ in round t. To evaluate the performance of practical online algorithms, we introduce the regret metric $R(T)$, which is defined as the total reward gap between the ideal policy and the considered DSA policy, i.e.,

$$R(T) = \sum_{v=1}^{V} \sum_{t=\phi_{v-1}+1}^{\phi_v} p_{k_v^*,v} - \sum_{v=1}^{V} \sum_{t=\phi_{v-1}+1}^{\phi_v} \mathbb{E}[p_{a_t,v}], \tag{3}$$

where $k_v^* = \text{argmax}_{k \in \mathcal{K}} p_{k,v}$ is the best channel in segment v. Therefore, in each slot t, the SU needs to minimize its regret by choosing its access channel a_t online based on its historical sensing results.

III. Proposed Algorithm

A. Change Detection

The change detection (CD) technique aims to detect the change of the idle probability by using a sequence of historical sensing observations. We denote by $\Delta_{k,v}$ the actual change of the idle probability of channel k at the end of the v-th segment, i.e., $\Delta_{k,v} = |p_{k,v+1} - p_{k,v}|, v \in [1, V-1]$. A successful detection should raise an alarm about the change of $p_{k,v}$ in the $v+1$-th segment as early as possible.

We denote the total access times of channel k as n_k. The corresponding observations are denoted by $\mathcal{H}_k = \{h_{k,1}, h_{k,2}, \ldots, h_{k,n_k}\}$. The CD algorithm detects the change of $p_{k,v}$ by using the latest $2w$ observations in $\mathcal{H}_k$. The CD statistic of channel k is defined as the difference between the average values of the first half and the last half of the $2w$ observations, which is given by

$$\text{D}_{k,w} = \frac{\left| \sum_{i=n_k-w+1}^{n_k} h_{k,i} - \sum_{i=n_k-2w+1}^{n_k-w} h_{k,i} \right|}{w}. \tag{4}$$

In the stationary environments, the $2w$ observations follow the same distribution. The corresponding statistic $\text{D}_{k,w}$ is a zero-mean random variable and its variance decreases as the observation window $2w$ increases. However, in the non-stationary environments, a part of the $2w$ observations follow the distribution with parameter $p_{k,v}$ and the rest observations follow the distribution with parameter $p_{k,v+1}$, which results in a positive drift of $\text{D}_{k,w}$. Thus, we can set a threshold δ and the CD algorithm raises an alarm when $\text{D}_{k,w} > \delta$.

We assume that the actual change $\Delta_{k,v}$ always exceeds the threshold δ, i.e., $\Delta_{k,v} \geq \delta$. For any $\Delta_{k,v} = \delta + c$ $(c > 0)$, we have [10]

$$\begin{aligned} \mathbb{P}(\text{D}_{k,w} > \delta) &\geq 1 - 2\text{exp}(-\frac{w(\Delta_{k,v} - \delta)^2}{2}) \\ &= 1 - 2\text{exp}(-\frac{wc^2}{2}), \end{aligned} \tag{5}$$

where the first inequality is derived by using the McDiarmids inequality. Therefore, when the last $2w$ observations crosses two segments, the CD algorithm raises an alarm with probability at least $1 - 2\text{exp}(-wc^2/2)$.

It has been shown that if idle probability $p_{k,v}$ changes by amount $\Delta_{k,v} \geq \delta$, then $w \geq 1/2\delta^2$ is sufficient to detect the change [11]. There exists a tradeoff between detection timeliness and detection accuracy. If δ is large, the CD algorithm can detect the changes with fewer observations at

the expense of neglecting the minimum changes. If δ is small, the CD algorithm can detect smaller changes but requires more observations, which can result in a large detection delay. For the considered DSA problem, a small change of the idle probability probably does not change the channel with the highest idle probability, and even if it does, such a suboptimal access decision only leads to a limited regret. Therefore, we tend to set a relatively large value for δ such that a large change of idle probability can be detected quickly.

B. Thompson Sampling Based Channel Access

Thompson sampling (TS) is a randomized Bayesian algorithm, which is asymptotically optimal for the stationary MAB problem and achieves a lower regret than the classical UCB algorithm [12]. Therefore, in each piecewise stationary segment, we apply TS to decide the channel a_t to sense. In each segment v detected by the CD algorithm, the true idle probability $p_{k,v}$ of channel k is approximated by a random variable Θ_k following a beta distribution Beta(S_k, F_k), the probability density function of which is given by

$$P(\Theta_k) = \frac{\Gamma(S_k + F_k)}{\Gamma(S_k)\Gamma(F_k)} \Theta_k^{S_k - 1}(1 - \Theta_k)^{F_k - 1}, \tag{6}$$

where Γ is the Gamma function, $S_k > 0$ and $F_k > 0$ are distribution parameters. For beta distribution Beta(S_k, F_k), the mean is given by $S_k/(S_k + F_k)$ and the variance is given by $S_k F_k/[(S_k + F_k + 1)(S_k + F_k)^2]$.

At slot t in segment v, the SU draws a sample $\theta_k(t)$ from Beta(S_k, F_k) for each channel $k \in \mathcal{K}$, which is assumed to be an approximation of $p_{k,v}$. To maximize the opportunity for successful transmission, the SU chooses the channel with the highest sampling value, i.e.,

$$a_t = \underset{k \in \mathcal{K}}{\operatorname{argmax}}\, \theta_k(t). \tag{7}$$

Upon observing the channel state $s_{a_t}(t)$, the SU updates the parameters of beta distributions as,

$$S_{a_t} = S_{a_t} + s_{a_t}(t), \tag{8}$$

$$F_{a_t} = F_{a_t} + 1 - s_{a_t}(t). \tag{9}$$

We note that S_{a_t} represents the total number of idle slots, and F_{a_t} represents the total number of busy slots. $S_{a_t} + F_{a_t} = n_{a_t}$ represents the total access times of channel a_t. Thus, the mean of the beta distribution represents the frequency of idle slots. According to the law of large numbers, the mean converges to the true idle probability $p_{a_t,v}$ as the number of observations increases to infinity

$$\lim_{n_{a_t} \to \infty} \frac{S_{a_t}}{S_{a_t} + F_{a_t}} = p_{a_t,v}. \tag{10}$$

Also, the variance of beta distribution decreases to zero as the number of observations increases to infinity, i.e., $\lim_{n_{a_t} \to \infty} \frac{S_{a_t} F_{a_t}}{(S_{a_t} + F_{a_t} + 1)(S_{a_t} + F_{a_t})^2} = 0$, as we have

$$\begin{aligned} 0 &\leq \lim_{n_{a_t} \to \infty} \frac{S_{a_t} F_{a_t}}{(S_{a_t} + F_{a_t} + 1)(S_{a_t} + F_{a_t})^2} \\ &\leq \lim_{n_{a_t} \to \infty} \frac{1}{4(S_{a_t} + F_{a_t} + 1)} = 0. \end{aligned} \tag{11}$$

Therefore, the sample $\theta_k(t)$ drawn from the time-varying beta distribution converges to the true idle probability $p_{k,v}$, i.e., $\lim_{t \to \infty} \theta_k(t) = p_{k,v}$.

On the one hand, the TS method ensures that the sample $\theta_k(t)$ in each slot t converges to the authentic idle probability. Thus, the channel with the highest $p_{k,v}$ is chosen by the SU with probability 1, which maximizes the total reward in the long term. On the other hand, during the limited piecewise-stationary segment, the sampling method based on the beta distribution implies that each channel has a chance to be selected. Thus, the channel characteristics can be explored. Therefore, the TS method achieves an efficient tradeoff between exploitation and exploration of primary channels.

Next, we demonstrate that the TS is asymptotically optimal for the DSA problem in each detected piecewise-stationary segment. Consider there is a piecewise-stationary segment with segment length τ. Thus, the superscript v can be omitted without any confusion. The idle probability of channel k is p_k and the best channel is given by k^*. The regret of a DSA algorithm is given by

$$R(\tau) = \sum_{t=1}^{\tau} p_{k^*} - \sum_{t=1}^{\tau} \mathbb{E}[p_{a_t}]. \tag{12}$$

An algorithm is asymptotically optimal if its regret $R(\tau)$ satisfies [13]

$$\lim_{\tau \to \infty} \frac{R(\tau)}{\ln \tau} = \sum_{k \in \mathcal{K}} \frac{p_{k^*} - p_k}{d(p_{k^*} || p_k)}, \tag{13}$$

where $d(p_{k^*}||p_k) = p_{k^*} \ln(\frac{p_{k^*}}{p_k}) + (1 - p_{k^*}) \ln(\frac{1 - p_{k^*}}{1 - p_k})$ is the Kullback-Leibler divergence between the two Bernoulli distributions with parameters p_{k^*} and p_k.

To bound the regret of TS, we rewrite the regret as

$$\begin{aligned} R(\tau) &= \sum_{t=1}^{\tau} p_{k^*} - \sum_{t=1}^{\tau} \mathbb{E}[p_{a_t}] \\ &= \sum_{k \in \mathcal{K}} (p_{k^*} - p_k) \mathbb{E}[n_k], \end{aligned} \tag{14}$$

where $\mathbb{E}[n_k]$ is the expected number of selections of channel k and $p_{k^*} - p_k$ is the performance gap between the best channel k^* and a suboptimal channel k. Thus, $(p_{k^*} - p_k)\mathbb{E}[n_k]$ represents the performance loss due to the selection of channel k. The regret is the sum of performance loss of all suboptimal channels. For simplicity, we denote the performance gap as $\zeta_k = p_{k^*} - p_k$. Since the gap ζ_k is fixed, we only need to bound the expected number of selections $\mathbb{E}[n_k]$. For each suboptimal channel $k \in \mathcal{K}$ and $\epsilon > 0$, there exists a constant $Q(\epsilon, p_{k^*}, p_k)$ such that [14]

$$\mathbb{E}[n_k] \leq (1 + \epsilon) \frac{\ln \tau + \ln\ln \tau}{d(p_{k^*} || p_k)} + Q(\epsilon, p_{k^*}, p_k). \tag{15}$$

Substituting in Equation (14), we get

$$R(\tau) \leq (1 + \epsilon) \sum_{k \in \mathcal{K}} \frac{\zeta_k (\ln \tau + \ln\ln \tau)}{d(p_{k^*} || p_k)} + C(\epsilon, p_1, \ldots, p_K), \tag{16}$$

Algorithm 1 The proposed TSCD for DSA

initialize $S_k = 1$, $F_k = 1$, $n_k = 0$ and $\mathcal{H}_k = \varnothing$ for all $k \in \mathcal{K}$
for $t = 1, 2, \ldots, T$ **do**
  for each $k \in \mathcal{K}$ **do**
    Draw $\theta_k(t) \sim \text{Beta}(S_k, F_k)$
  end for
  sense channel a_t given by (7) and observe state $s_{a_t}(t)$
  update S_{a_t} and F_{a_t} according to (8) and (9)
  $n_{a_t} \leftarrow n_{a_t} + 1$, $h_{a_t, n_{a_t}} = s_{a_t}(t)$
  $\mathcal{H}_{a_t} \leftarrow \mathcal{H}_{a_t} \cup \{s_{a_t}(t)\}$
  if $n_{a_t} \geq 2w$ & $\mathrm{D}_{a_t,w} > \delta$ **then**
    $\mathcal{H}_k \leftarrow \varnothing$, $\forall k \in \mathcal{K}$
    $S_k = F_k = 1$, $n_k = 0$, $\forall k \in \mathcal{K}$
  end if
end for

where $C(\epsilon, p_1, \ldots, p_K) = \zeta_1 Q(\epsilon, p_{k^*}, p_1) + \zeta_2 Q(\epsilon, p_{k^*}, p_2) + \ldots + \zeta_K Q(\epsilon, p_{k^*}, p_K)$ is a problem depend constant. The fact that the inequality (16) holds for every $\epsilon > 0$ proves the asymptotically optimality of TS.

C. TSCD

The proposed TSCD algorithm consists of two components, i.e., the CD part and the TS part. The CD algorithm is responsible for detecting the changes in the idle probability of each channel in non-stationary environments. The TS algorithm is responsible for balancing the exploration and exploitation of channel selection in each piecewise-stationary segment.

In each slot t, we first apply TS to decide the channel a_t to access and update the corresponding parameters S_{a_t} and F_{a_t} based on the sensing observation $s_{a_t}(t)$. The total number of observations n_{a_t} is increased by 1 and the historical observations set $\mathcal{H}_{a_t}$ is updated by adding current observation $s_{a_t}(t)$, i.e., $\mathcal{H}_{a_t} \leftarrow \mathcal{H}_{a_t} \cup \{s_{a_t}(t)\}$. Then we apply CD to detect the variation of $p_{a_t,v}$. If a change is detected, i.e., $\mathrm{D}_{a_t,w} > \delta$, we reset the historical observations $\mathcal{H}_k \leftarrow \varnothing, \forall k$ and reinitialize the corresponding parameters $S_k = F_k = 1, n_k = 0, \forall k$. Details of the TSCD algorithm are summarized in Algorithm 1.

IV. Numerical Results

In this section, we present the numerical results in two typical cases. In case 1, the idle probabilities of successive segments are independent from each other, which represents the scenario when the SU moves from one primary cell to another. In case 2, the idle probability varies with time gradually, which represents the variations of primary traffic.

In case 1, we set the entire time horizon $T = 20000$ and the number of piecewise-stationary segments $V = 6$ with $\phi_1 = 2000, \phi_2 = 5000, \phi_3 = 10000, \phi_4 = 12000, \phi_5 = 15000$. For each channel k in segment v, the idle probability $p_{k,v}$ is sampled uniformly between 0 and 1.

In case 2, we assume the idle probability $p_{k,v}$ changes at every slot, i.e., $V = T = 20000$ and $\phi_n = n, 1 \leq n < T$. For each channel k, we set the initial idle probability as $p_{k,1} = 0.5$, and for any slot $t \geq 2$, the idle probability $p_{k,v}$ is statistically given by

$$p_{k,v} = f(p_{k,v-1} + 0.02\mu(v)), \tag{17}$$

where $\mu(v)$ is a random variable following a uniform distribution $U[-0.5, 0.5]$, and $f(x)$ is given by

$$f(x) = \begin{cases} x, & x \in [0,1], \\ 0, & x < 0, \\ 1, & x > 1. \end{cases} \tag{18}$$

We compare the proposed TSCD algorithm with the modified UCB (MUCB) [9], the TS [14] and the sliding window TS (SWTS) [15] algorithms. Parameters of these algorithms are tuned as suggested in the corresponding papers. Specifically, the sliding window of the SWTS algorithm is set as the optimal value $2\sqrt{T \log T/(V-1)}$ for the segment number V, which however is generally unknown in practice. In addition, we also present the performance of an oracle as the upper bound, which always knows the channel with the highest idle probability in each segment. The CD parameters of the proposed TSCD algorithm are set as $w = 156$ and $\delta = 0.08$ based on a series of numerical experiments. All the numerical results are averaged by 1000 Monte Carlo simulations.

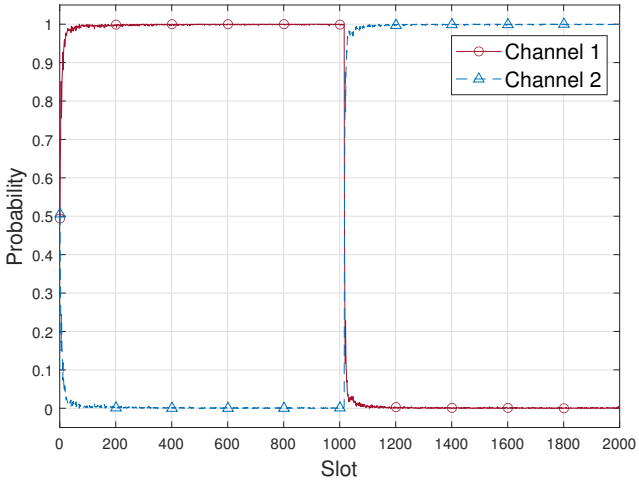
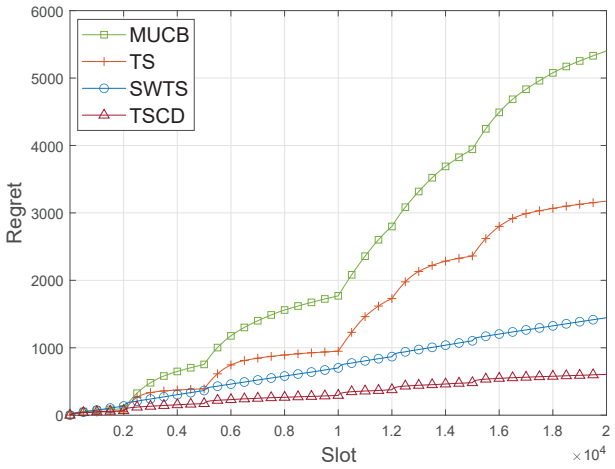


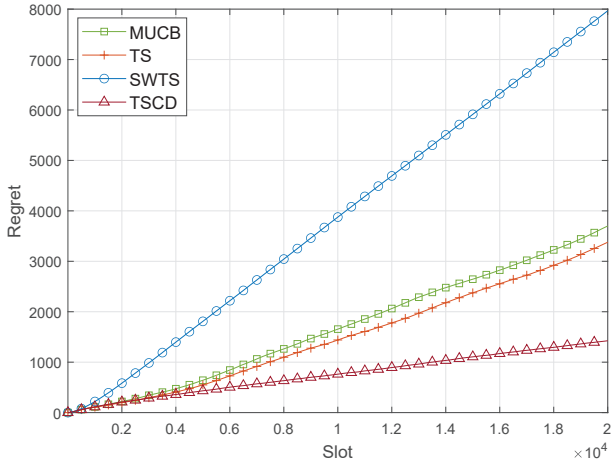
Fig. 1: Selection probability as a function of time in networks with two channels.

A. Example

We first give an example to show the ability of the proposed TSCD algorithm to track the best channel. In the considered network, there are $K = 2$ channels and the time horizon is split into 2 segments of length 1000. Channel 1 is the best channel in the first segment with $p_{1,1} = 0.9$ and $p_{2,1} = 0.3$. In the second segment, channel 2 is the best channel with $p_{1,2} = 0.3$ and $p_{2,2} = 0.9$. Fig. 1 shows the selection probability of each channel as a function of time. The proposed TSCD algorithm accesses the best channel in each segment with a probability of almost 1. As we can see, TSCD can detect the change of

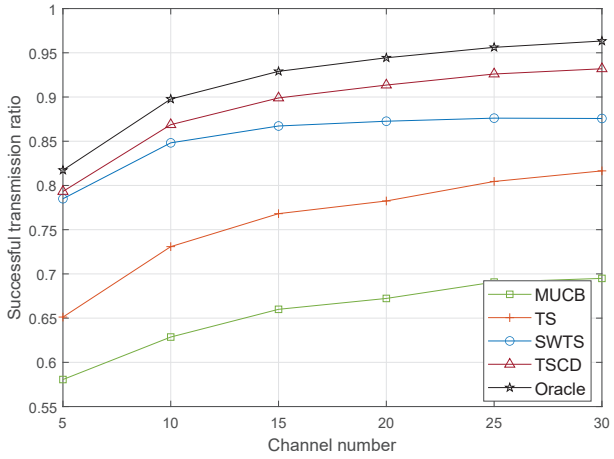


(a) Case 1

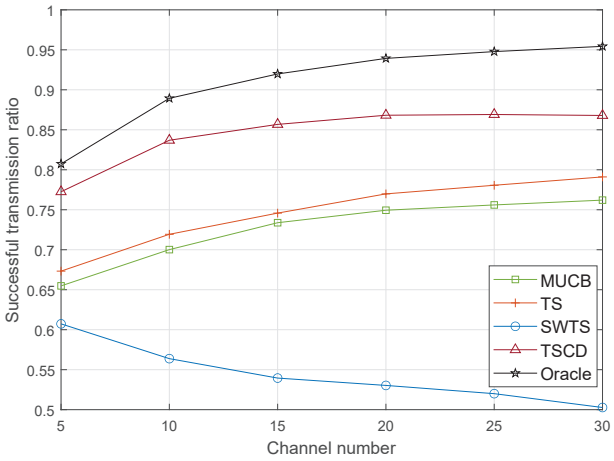


(b) Case 2

Fig. 2: Regret as a function of time in the considered two cases.



(a) Case 1



(b) Case 2

Fig. 3: STR as a function of channel number K in the considered two cases.

$p_{1,1}$ within a short period and quickly converge to channel 2 in the second segment.

B. Regret Performance

Fig. 2a shows the regrets of different DSA algorithms as a function of time in case 1 for $K = 20$. Since the regret of the oracle is zero, we don't show the oracle curve in the figure. The regrets of other algorithms increase with t and the proposed TSCD algorithm achieves the lowest regret during the entire time horizon.

As we can see, the proposed TSCD algorithm can quickly detect the change in idle probability and converge to the best channel in each segment. The SWTS algorithm can also adapt to the changes of channel statistics. However, it only uses recent observations in the sliding window, which results in information loss and reduces its capability to exploit the best channel. The MUCB and TS algorithms cannot track the changes in channel statistics and their regrets increase significantly with time.

Fig. 2b shows the regrets of different DSA algorithms as a function of time in case 2 for $K = 20$. Due to the highly non-stationary environments, no algorithms can achieve a sublinear regret. The proposed TSCD algorithm again achieves the lowest regret. When the cumulative change of idle probability exceeds the CD threshold, the proposed TSCD algorithm can detect the change and reset the past outdated observations. Thus, the proposed TSCD algorithm achieves a lower regret than the other algorithms. For the SWTS algorithm, the window size is $2\sqrt{\log T} \approx 6$, which is even smaller than the number of channels K. Thus, it cannot track the channel statistics at all with such a small window size, and generates the highest regret among all considered algorithms.

C. Successful Transmission Rate

Fig. 3a shows the STRs of different DSA algorithms as a function of channel number K in case 1. The proposed TSCD algorithm achieves the highest STR among all considered algorithms except for the oracle. Based on the first-order statistics, the maximal idle probabilities of all channels increase with the number of channels. Thus, the STR of the SU should increase with the number of channels. However, the increasing

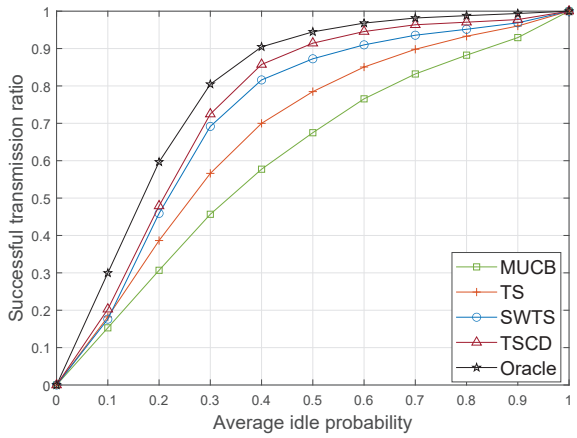


Fig. 4: STR as a function of average idle probability λ.

number of channels also increases the difficulty of tracking the best primary channel, as the sensing capability of the SU stays the same. As we can see, the STR of the SU increases with the channel number for all algorithms, but the curve flattens when the channel number is large. Note that the sliding window of SWTS $2\sqrt{T\log T/(V-1)}$ is independent from the channel number. The difficulty introduced by extra channels offsets the additional transmission opportunities. Thus, the performance of SWTS keeps unchanged when the channel number increases.

Fig. 3b shows the STRs of different DSA algorithms as a function of channel number K in case 2. The STRs of the TSCD, TS and MUCB algorithms increase with K as the maximal idle probability increases with K. The proposed TSCD algorithm achieves the highest STR due to its capability to better track channel statistics. For the SWTS algorithm with sliding window $2\sqrt{\log T} \approx 6$, it cannot track the channel statistics at all. Its performance is further aggravated by the large number of candidate channels. Thus, the STR of the SWTS algorithm decreases with the number of channels.

Fig. 4 shows the STRs of different DSA algorithms as a function of average idle probability λ in case 1. As we can see, the STRs increase with λ for all considered algorithms, since a larger λ implies more transmission opportunities. When the primary channels are fully occupied, i.e., $\lambda = 0$, or fully unoccupied, i.e., $\lambda = 1$, there is no difference between all DSA algorithms. Thus, all considered algorithms achieve the same performance in those extreme scenarios. When the primary channels are partially occupied, the proposed TSCD algorithm outperforms the other algorithms as it can detect the changes in channel statistics and keep a balance between the exploitation of the currently best channel and the exploration of potential better channels.

V. CONCLUSION

In this paper, we considered the DSA problem in non-stationary environments and proposed the TSCD algorithm to track the dynamics of primary channels. Specifically, the proposed TSCD algorithm can effectively detect the change of the idle probability of each primary channel and achieve an efficient tradeoff between the exploitation of the currently best channel and the exploration of potential better channels. Therefore, the proposed TSCD algorithm enables the SU to quickly converge to the best primary channel in non-stationary environments. Numerical results show that the proposed TSCD algorithm improves the successful transmission ratio as compared to the existing algorithms in various settings.

REFERENCES

[1] F. Li *et al.*, "Advances and emerging challenges in cognitive Internet-of-Things," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5489–5496, Aug. 2020.
[2] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
[3] M. Zandi, M. Dong, and A. Grami, "Distributed stochastic learning and adaptation to primary traffic for dynamic spectrum access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1675–1688, Mar. 2016.
[4] J. Dai and S. Wang, "Clustering-based spectrum sharing strategy for cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 1, pp. 228–237, Jan. 2017.
[5] H. Liu, K. Liu, and Q. Zhao, "Learning in a changing world: Restless multiarmed bandit with unknown dynamics," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1902–1916, Mar. 2013.
[6] 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.
[7] 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.
[8] A. Anandkumar *et al.*, "Distributed algorithms for learning and cognitive medium access with logarithmic regret," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 731–745, Apr. 2011.
[9] J. Oksanen and V. Koivunen, "An order optimal policy for exploiting idle spectrum in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1214–1227, Mar. 2015.
[10] Y. Cao *et al.*, "Nearly optimal adaptive procedure with change detection for piecewise-stationary bandit," in *Proc. Int. Conf. Artif. Intell. Stat.*, Naha, Okinawa, Japan, Apr. 2019.
[11] P. Auer, P. Gajane, and R. Ortner, "Adaptively tracking the best bandit arm with an unknown number of distribution changes," in *Proc. Annu. Conf. Learn. Theory*, Phoenix, Arizona, Jun. 2019.
[12] Z. Kuai and S. Wang, "Thompson sampling-based antenna selection with partial CSI for TDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7533–7546, Dec. 2020.
[13] T. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Adv. Appl. Math.*, vol. 6, no. 1, pp. 4–22, Mar. 1985.
[14] S. Agrawal and N. Goyal, "Analysis of Thompson sampling for the multi-armed bandit problem," in *Proc. Annu. Conf. Learn. Theory*, Edinburgh, Scotland, Jun. 2012.
[15] F. Trovo *et al.*, "Sliding-window Thompson sampling for non-stationary settings," *J. Artif. Intell. Res.*, vol. 68, pp. 311–364, May 2020.