International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.2 (2016), pp.457-468
http://dx.doi.org/10.14257/ijsip.2016.9.2.39

Blind Spectrum Selection in a Decentralized Cognitive Radio Networks with Heterogeneous Applications

Yongqun Chen[1*], Huaibei Zhou[2], Ruoshan Kong[2], Junyuan Huang[1], Hang Qin[3]

[1]*School of Physics and Technology, Wuhan University, 430072 Wuhan, China*
[2]*International school of software, Wuhan University, 430072 Wuhan, China*
[3]*Computer School, Yangtze University, 430023 Jingzhou, China*
yiyuxiniao@gmail.com

Abstract

In this paper, we consider a cognitive radio network in which the sensing ability of cognitive radio is limited and the channel statistics are not known as a priori information in the opportunistic spectrum access(OSA) framework. It is a special challenge to design a joint spectrum sensing and access strategy for secondary users with diverse service requirements of heterogeneous applications, i.e. the real-time applications and best-effort applications. We formulate the spectrum decision problem as a decentralized multi-armed bandit problem and propose slot structures for cognitive radio network to cope with collisions between heterogeneous applications. The proposed scheme is proved achieving logarithmic regrets in time asymptotically and simulation results show that each user orthogonalizes into their rank-optimal channels according to their pre-allocated priorities, which indicates efficient spectrum utilization while satisfying service requirements.

Keywords: *Cognitive radio, multi-armed bandit, opportunistic spectrum access, distributed algorithms, heterogeneous applications.*

1. Introduction

Cognitive radio (CR) has recently emerged as a promising technique to improve the utilization of the existing statically allocated spectrum [1]. Meanwhile, traditional wireless communication networks like WSNs equipped with cognitive radio will benefit from its potential advantages [2]. One of key challenges in CR is to achieve the coexistence of primary users (PUs) and secondary users (SUs), accessing the same part of the spectrum. PUs have priority in accessing spectrum while SUs must sense spectrum before accessing. It is viable to sense all channels before deciding which channel to access based on accessing strategy in an ideal condition. However, it is difficult to sense the whole operating spectrum band in a given period of time because of wide-band spectrum and hardware constraints. At the same time, the spectrum statistical information as a priori knowledge may not always be securable in a decentralized network, i.e. it is partially observed and priori unknown to the SUs.

The spectrum sensing and accessing problem is a topic of extensive research. Zhao *et al.* [3-4] formulated a partially observable Markov decision process (POMDP) framework under Markovian channel model. Liu *et al.* [5] figured out POMDP framework could also be viewed as a restless MAB process for independent channels and proposed a restless bandit formulation based on Whittle's Index Policy, which built a connection between cognitive medium access and the multi-armed bandit problem. However, above works assumed channel transition probabilities were known to SUs. Meanwhile, the multiple distributed players of this problem regardless of any prior knowledge about channel statistics raised a wide range of interest. Liu *et al.* [6] proposed a family of distributed learning and access policies known as time-division fair share (TDFS). Anandkumar *et al.* [7] proposed the distributed ρ^{PRE} policy under pre-allocation order based on the ε_n-greedy policy. Recently, Gai *et al.* [8] proposed a general SL(K) policy under which SUs will orthogonalize into their rank-optimal channels according to their pre-allocated priorities. Then the prioritized access policy(DLP) and fair access policy(DLF) based on SL(K) can be easily established.

Although all these works enabled CR users to explore and exploit channel availability effectively, the problem in a heterogeneous spectrum environment has not been fully investigated. Generally, CR networks have multiple available spectrum bands over a wide frequency range that show different channel characteristics, and need to support applications with diverse service requirements, such as real-time applications and best-effort applications [9]. For real-time applications, they require minimum delay-based channels and for best-effort applications, maximum capacity-based channels are required. This introduces new critical issues in the above framework. In this paper, we investigate the prioritized access policy based on SL(K) policy in a CRN hybridizing with real-time applications and best-effort applications. We proposed slot structures for both applications to cope with the collisions between them and extended the finding of existing SL(K) policy whose results show that each user orthogonalizes into their rank-optimal channels according to their pre-allocated priorities and achieving logarithmic regrets in time asymptotically.

The rest of this paper is organized as follows: Section II describes the system model with multiple secondary users with different QoS requirements and formulates the problem. In Section III, we present our scheme for real-time applications and best-applications based on SL(K) policy, respectively. Section IV examines the proposed scheme through simulation. Finally, the paper concludes with summary in Section V.

2. System Model and Problem Formulation

We consider a cognitive system with C independent and orthogonal channels licensed to a primary network whose users communicate following a synchronous slot structure illustrated in Figure 1. At the beginning of each slot, the secondary user chooses a channel to sense the availability. Once the sensed result indicates the channel is idle, SUs transmit pilot to receiver to probe the channel state information (CSI). The CSI is fed back through a dedicated error-free feedback channel without delay. After data are transmitted, the receiver acknowledges every successful or unsuccessful transmission as $Z_{i,j}(k)=0$ for collision occurred, otherwise $Z_{i,j}(k)=1$.

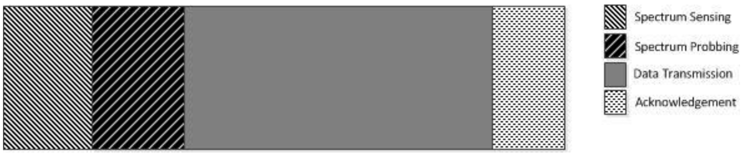


Figure 1. Basic slot structure

The cognitive network is composite of M secondary users of real-time applications and N secondary users of best-effort applications, where $M+N<C$. We assume there exists a network head like the cluster head in WSNs, which is responsible for collecting statistics of the number of users and their application types and then dispatching the pre-allocated ranks to the users. For simplicity, we assume the priority of SU_j with the same application type is ranked by j, i.e. the priority of SU_p is higher than SU_q if $p<q$ for

either type of application. The minimum information is prior to learning and transmission processes and will not be changed afterwards.

We model the channels as Rayleigh fading channels with additive white Gaussian noise (AWGN), whose SNR γ_i is exponentially distributed with mean value $\sigma_i \in \Sigma$ [10]:

$$f(\gamma_i) = \frac{1}{\sigma_i}\exp(-\frac{\gamma}{\sigma_i})$$

and the channel availability W_i is modeled as *i.i.d* Bernoulli process with mean value $\beta_i \in \mathrm{B}$: $W_i \sim B(\beta_i)$. The channel model can be illustrated in Figure 2.

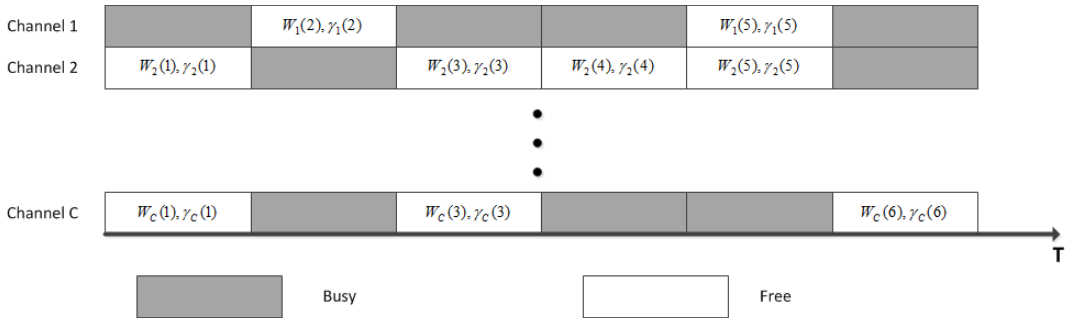


Figure 2. Slotted channel model

Blind spectrum selection for SUs of homogeneous applications can be regarded as a Decentralized-MAB problem [6][7][8][11][12]. Taking the real-time application as an example(in the same way we can formulate the policy performance of best-effort application), we denote π_j as the decentralized policy for user j and $\pi = \{\pi_j, 1 \le j \le M\}$ as the set of policies of all users. Arm i yields reward $X_i(t)$ at slot t according to the distribution of diverse QoS requirement with expectation θ_i (in this paper, the terms "arm" and "channel" are interchangeably used). The performance of the decentralized policies can be defined as the regret of all SUs with real-time applications:

$$R_M^{\pi}(\Theta;T) = T\sum_{i\in \mathrm{O}_M^*}\theta_i - \mathrm{E}^{\pi}[\sum_{t=1}^{T} S_{\pi}(t)] \tag{1}$$

where O_M^* is the set of M arms with M largest expected rewards and $S_{\pi(t)}(t)$ is the sum of the actual reward obtained by all SUs with real-time applications at time t under policy π , which is:

$$S_{\pi}(t) = \sum_{i=1}^{C}\sum_{j=1}^{M} X_i(t)\mathrm{I}_{i,j}(t) \tag{2}$$

where $\mathrm{I}_{i,j}(t)$ is defined to be 1 if user j is the only one to play arm i , and 0 otherwise.

This problem is widely studied and a typical policy named SL(K) which achieves logarithmic regret is proposed [8]. The SL(K) policy is a key subroutine of decentralized learning policies based on UCB1 policy in classical MAB [13]. The advantage of this policy compared to ρ^{PRE} policy [7] is that it does not require prior information about the minimum difference between arms but achieves logarithmic regret in the same way. So we mainly investigated the general SL(K) policy of this problem in a CRN with heterogeneous applications.

Since real-time applications are sensitive to delay and jitter, SUs with these applications have priorities over best-effort applications. In our framework, the objective of real-time applications is to maximize the channel availability according to SUs' priorities, while the objective of best-effort applications is to maximize the channel capacity. That means the reward $X_i(t)$ yielded by channel i for real-time applications is $W_i(t)$ and $W_i(t)\gamma_i(t)$ for best-effort applications. And thus, we denote the expectation of rewards θ_i for real-time applications as $\theta_i^r = \beta_i$ and $\theta_i^b = \beta_i \sigma_i$ for best-effort applications.

Generally, if the SL(K) policy is applied to the two types of applications separately, they cannot guarantee the collisions avoidance in the long term between the two applications because of the common channel availability, i.e.$O_M^* \cap O_N^* \neq \varnothing$. Hence, we proposed a scheme to ensure $O_M^* \cap O_N^* = \varnothing$, which is illustrated in the next section.

3. Heterogeneous Applications Framework

First, based on above discussion, we assume the channel capacity always meet the requirement of real-time applications because real-time applications require low channel capacity and bad channels can be excluded in spectrum sensing stage. And the real-time applications always have priority over best-effort applications. Based on these assumptions, we design the slot structures for real-time applications as showed in Figure 3 and for best-effort applications as showed in Figure 4. SUs with best-effort applications is silent while the SUs with real-time applications sensing, which ensures the priority of the real-time applications. Noting that the access of the real-time applications alters the channel statistics for best-effort applications, which means the rewards distribution will be non-stationary, we examine the suitability of SL(K) policy for best-effort applications in subsection 2.

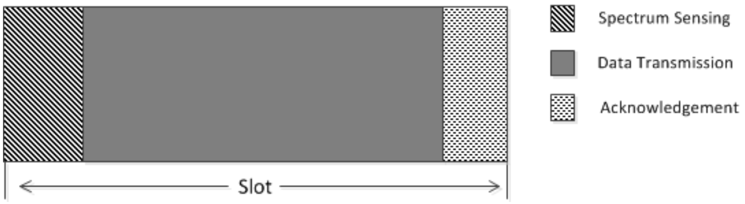


Figure 3. slot structure of real-time applications

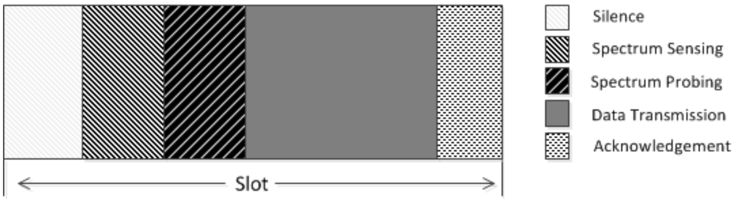


Figure 4. slot structure of best-effort applications

3.1. CRN with Real-time Applications

The SL(K) can be applied to the CRN with real-time applications directly, in which each user selects an channel with the K-th largest expected channel availabilities. The policy is described as **Algorithm** 1.

According to Theorem 1 in [8], the expected number of times SU_K of real-time applications access any channel $i \neq A_K$ after T time slots $E[n_i(T)]$ is at most:

$$\frac{8\ln T}{\Delta_{K,i}^{2}}+1+\frac{2\pi^{2}}{3} \qquad (3)$$

where A_K is the arm of the K-th largest expected reward and $\Delta_{K,i}=|\theta_K^r-\theta_i^r|$, θ_K^r is the K-th largest expected reward for real-time applications. Hence, the expected regret $R_M^{\pi}(\Theta;T)$ grows as $O(M(C+M-2)\ln T)$ according to Theorem 2 in [8].

Algorithm 1: SL(K) policy for the user with rank K, $K=1,\cdots,M$

```
// Init: play each arm once
For $t=1$ to $C$
    Play arm $i=t$ and let $n_i(t)=1$, $\hat{\theta}_i(t)=X_i(t)$
EndFor
// Main loop
For $t=N+1$ to $T$
    Step1: Select the set $O_K$ contains arms with the $K$ highest index values.
        $\hat{\theta}_i(t-1)+\sqrt{\frac{2\ln t}{n_i(t-1)}}$    (4)
    Step2: Play the arm with the minimal index value in $O_K$ according to.
        $\hat{\theta}_i(t-1)-\sqrt{\frac{2\ln t}{n_i(t-1)}}$    (5)
    Step3: Update $n_k(t)=n_k(t-1)+1$ and $\hat{\theta}_k(t)=\frac{\hat{\theta}_k(t-1)n_k(t-1)+X_k(t)}{n_k(t)}$
EndFor
```

3.2. CRN with Best-effort Applications

Since the SUs with best-effort applications always sense the channel after the SUs with real-time applications do according to the slot structures we design above, the statistics of channel capacities have been altered. Hence, the rewards to them should be remodeled.

According to Eq.(3), the expected number of times the SUs with real-time applications access any channel $i\notin O_M^*$ after T time slots is at most:

$$\sum_{k=1}^{M}\left(\frac{8\ln T}{\Delta_{K,i}^{2}}+1+\frac{2\pi^{2}}{3}\right) \qquad (6)$$

which alters the statistics of channel state of CRN with best-effort applications:

$$\theta_i^b(T)=[\beta_i-\frac{1}{T}\sum_{k=1}^{M}(\frac{8\ln T}{\Delta_{K,i}^{2}}+1+\frac{2\pi^{2}}{3})]\cdot\sigma_i \qquad (7)$$

which means the reward to SUs with best-effort applications is non-stationary. It has the form:

$$\theta_i^b(T)=\theta_i^b+\frac{a_i\ln T+b_i}{T} \qquad (8)$$

where a_i, b_i are constants.

Now we provide an expected upper bound of the SL(K) policy for best-effort applications described below:

Theorem 1: If policy SL(K) running on arbitrary arms with non-stationary rewards has the form as Eq.(8), and denoting $N_{i,j}$ as the solution of $\theta_i^b(t) = \theta_j^b(t)$, then the expected number of times that we pick any arm $i \neq \mathrm{A}_K$ after $T \geq N_{i,j}^*, N_{i,j}^* = \max_j\{N_{i,j}\}$ time slots $\mathrm{E}[n_i(T)]$ is at most: $\frac{8\ln T}{\min_j \Delta_{ij}^2} + C$.

Proof: see Appendix.

Corollary 1: The expected regret under policy SL(K) for SUs with best-effort applications in our scheme grows logarithmically in time slots.

Proof: for each user k, the regret arises due to two cases: (1) user k plays arm $i \notin \mathrm{O}_N^*$ and (2) other user $l \neq k$ plays arm A_k. Hence, the regret of user k is upper bounded according to Theorem 1:

$$R_k^\pi(\Theta;T) \leq \sum_{i \notin \mathrm{O}_N^*} \mathrm{E}[n_{k,i}(T)]\theta_{ik} + \sum_{l \neq k} \mathrm{E}[n_{l,k}(T)]\theta_k \tag{9}$$

Then the regret for SUs with best-effort applications is bounded:

$$\begin{aligned} R_N^\pi(\Theta;T) &= \sum_{k=1}^{N} R_k^\pi(\Theta;T) \\ &\leq \sum_{k=1}^{N} (\sum_{i \notin \mathrm{O}_N^*} \mathrm{E}[n_{k,i}(T)]\theta_{ik} + \sum_{l \neq k} \mathrm{E}[n_{l,k}(T)]\theta_k) \\ &\leq N(C + N - 2)(\frac{8\ln T}{\min_{i,j} \Delta_{ij}^2} + C)\theta_{\max} \end{aligned}$$

where $\theta_{max} = \max_i \theta_i$. Hence Corollary 1 is proved.

The above result states that in our framework, the SL(K) policy running for best-effort applications can also achieve logarithmical regret in time slots and each user orthogonalizes into different channels from real-time applications.

4. Numerical Results

In this section, we present simulation results for the scheme proposed in this work. In the simulations, we assume $C = 9$ channels with channel availabilities B=[0.5, 0.2, 0.8, 0.6, 0.9, 0.03, 0.4, 0.1, 0.7], channel capacities Σ= [0.45,0.15,0.6,0.55,0.1,0.25,0.35,0.05,0.25] and $M = N = 3$ SUs.

Figure 5 shows the simulation results of the three secondary users with real-time applications averaged over 50 runs. As expected, the expected regret of the policy grows logarithmically in time slots. And the actions of all the users converge to their rank-optimal channels, i.e. $\mathrm{O}_M^* = \{5,3,9\}$.

Figure 6 shows the average channel capacities over 50 runs for best-effort applications after the access of real-time applications. The expected rewards of channel $i \in \mathrm{O}_M^*$

decrease and the others increase to the maximum values, which is consistent with the results of rewards statistical analysis for best-effort applications, i.e. Eq.(7).

Figure 7 shows the results of the three secondary users with best-effort applications. The expected regret grows logarithmically generally. Furthermore, $O_N^* = \{4,1,7\}$ has no intersection with $O_M^* = \{5,3,9\}$, which proves the feasibility of the proposed scheme. Meanwhile, the results also indicate the policy has enough stability as well as the original application of SL(K) (**Figure 5**) from the percentiles' curves.

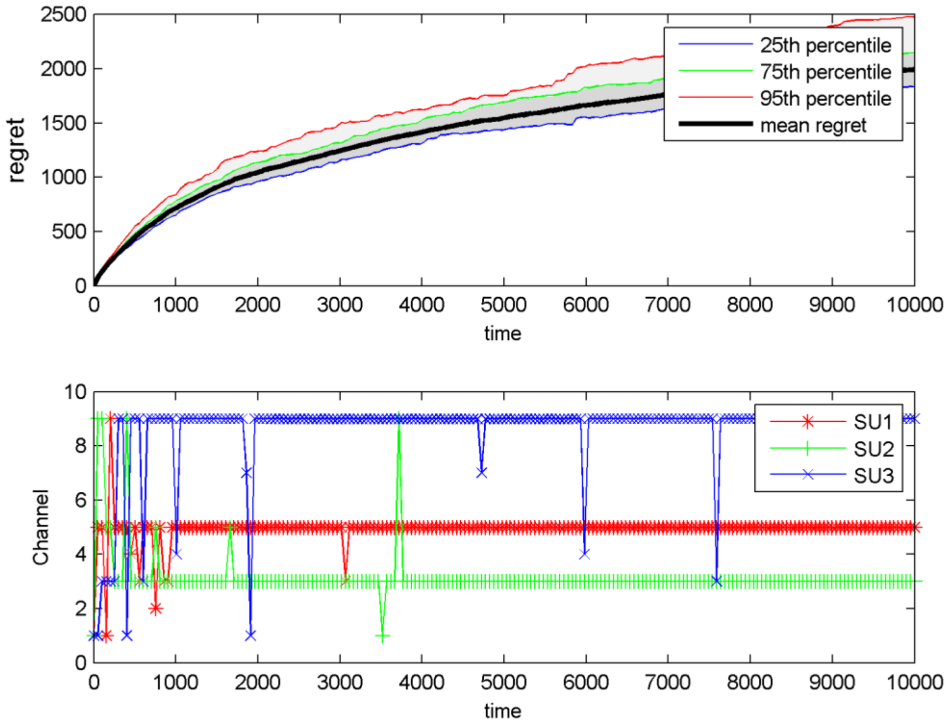


Figure 5. The regrets and actions of SUs with real-time applications

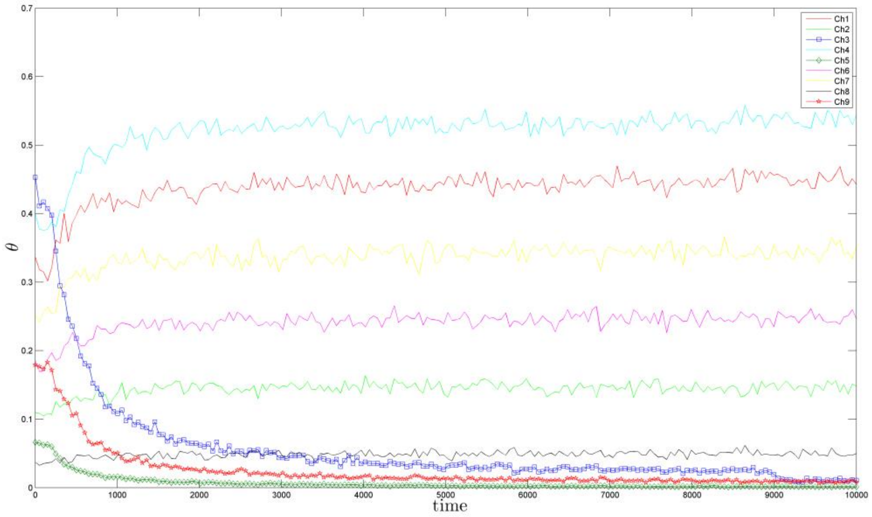


Figure 6 The non-stationary rewards for best-effort applications

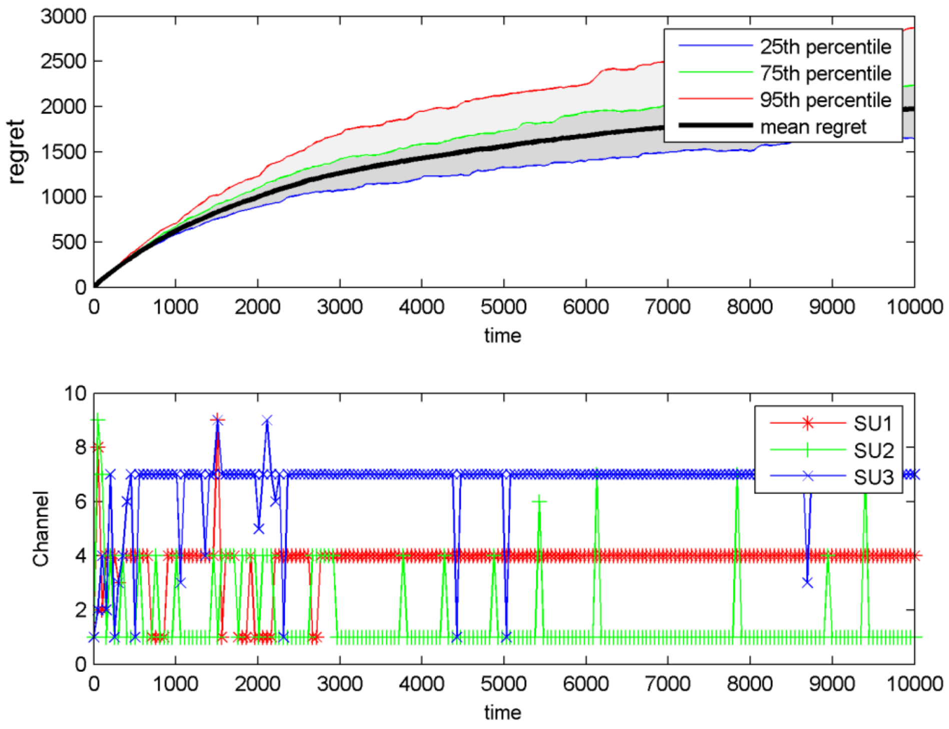


Figure 7. The regrets and actions of SUs with best-effort applications

5. Conclusion

In this work, we study the spectrum sensing and accessing problem of cognitive radio under coexistence of two types of applications, the real-time applications and the best-effort applications, which have different QoS requirements. We have made two key contributions to this problem. First, we design the slot structures of both applications for collision avoidance. Second, we prove the SL(K) policy can work under the non-stationary rewards whose expectation has the form of Eq.(8). Through simulations, the results show that the proposed scheme achieves logarithmical regrets for both applications.

Acknowledgments

The work presented in this paper was supported by the International S&T cooperation Program of China under grand No. 2013DFA12460 and the work of Hang Qin was supported by the Natural Science Foundation of China under grant No. 61440023.

Appendix

Proof of Theorem 1:

Denote the index value of UCB1 Eq.(4) as $I_i(t) = \hat{\theta}_i(t-1) + C_{t,n_i}$ and Eq.(5) as $I_i'(t) = \hat{\theta}_i(t-1) - C_{t,n_i}$, where $C_{t,n_i} = \sqrt{\frac{2\ln t}{n_i(t-1)}}$ and $n_i(t)$ is the number of times that arm i was played in time slot t. We note the solution of $\theta_i(t) = \theta_j(t), \mathrm{t} \geq 0$ is

$N_{i,j} = \frac{a_{ij}}{\theta_{ij}} W_0(\frac{\theta_{ij}}{a_{ij}} \exp(-b_{ij} / a_{ij}))$ if exists or otherwise we just let $N_{i,j} = 1$, where $W_0(\cdot)$ is the Lambert W function[14] and $\theta_{ij} = \theta_i - \theta_j, a_{ij} = a_i - a_j, b_{ij} = b_i - b_j$. Obviously, the horizon $T > N^*_{i,j}, N^*_{i,j} = \max_j \{N_{i,j}\}$ should be considered. Then, for any arm $i \neq \mathrm{A}_K$:

$$\begin{aligned} \mathrm{E}[n_i(T)] &= 1 + \sum_{t=C+1}^{T} \{A(t) = i\} \\ &= 1 + \sum_{t=C+1}^{T} \{A(t) = i \mid \theta_i < \theta_K\} + \{A(t) = i \mid \theta_i > \theta_K\} \end{aligned} \tag{10}$$

where $A(t)$ denotes the action of SU_K at time t under the policy SL(K).

For $\theta_i < \theta_K$, arm i is picked at time t means one of the two cases happened: $t \leq N_{i,K}$ or there exists an arm $j \in \mathrm{O}^*_K$ but $j \notin \mathrm{O}_K$ at time t in the Step 1 in Algorithm 1. The former will never happen when sufficient plays, and to the latter, the following inequality holds: $I_j(t) \leq I_i(t)$. Then we have:

$$\begin{aligned} \mathrm{E}[n_i(T)] &\leq l + \sum_{t=l}^{T} \{A(t) = i \mid \theta_i < \theta_K, n_i(t-1) \geq l, l > N^*_{i,j}\} \\ &\leq l + \sum_{t=l}^{T} \Pr\{I_j(t) \leq I_i(t) \mid n_i(t-1) \geq l, l \geq N^*_{i,j}\} \\ &\leq l + \sum_{t=l}^{\infty} \sum_{n_j=1}^{t-1} \sum_{n_i=l}^{t-1} \Pr\{\hat{\theta}_j(t) + C_{t,n_j} \leq \hat{\theta}_i(t) + C_{t,n_i}, t \geq N^*_{i,j}\} \end{aligned} \tag{11}$$

$\hat{\theta}_j(t) + C_{t,n_j} \leq \hat{\theta}_i(t) + C_{t,n_i}$ implies that at least one of the following must be true:

$$\hat{\theta}_j(t) \leq \theta_j(t) - C_{t,n_j} \tag{12}$$

$$\hat{\theta}_i(t) \geq \theta_i(t) + C_{t,n_i} \tag{13}$$

$$\theta_j(t) < \theta_i(t) + 2C_{t,n_i} \tag{14}$$

According to the Chernoff-Hoeffding bound, we can find the upper bound of Eq.(12) and Eq.(13):

$$\begin{aligned} \Pr\{\hat{\theta}_j(t) \leq \theta_j(t) - C_{t,n_j}\} &= \Pr\{\hat{\theta}_j(t) \leq \theta_j + \frac{a_j \ln t + b_j}{t} - \sqrt{\frac{2 \ln t}{n_j(t)}}\} \\ &\approx \Pr\{\hat{\theta}_j(t) \leq \theta_j - \sqrt{\frac{(2 - \delta_j(t)) \ln t}{n_j(t)}}\} \leq t^{-(2-\delta_j(t))^2} \end{aligned} \tag{15}$$

$$\begin{aligned} \Pr\{\hat{\theta}_i(t) \geq \theta_i(t) + C_{t,n_j}\} &= \Pr\{\hat{\theta}_i(t) \geq \theta_i + \frac{a_i \ln t + b_i}{t} + \sqrt{\frac{2 \ln t}{n_i(t)}}\} \\ &\approx \Pr\{\hat{\theta}_i(t) \geq \theta_i + \sqrt{\frac{(2 + \delta_i(t)) \ln t}{n_i(t)}}\} \leq t^{-(2+\delta_i(t))^2} \end{aligned} \tag{16}$$

where $\delta_i(t) = 2(a_i \ln t + b_i) / t$.

And if $l \geq \left\lceil 8 \ln T \, / \, (\theta_{ij} + \frac{a_{ij} \ln T + b_{ij}}{T})^2 \right\rceil$,

$\theta_j(t) - \theta_i(t) - 2C_{t,n_i} \geq \theta_{ji} + \frac{a_{ji} \ln T + b_{ji}}{T} - 2C_{t,l} = 0$, which means Eq.(14) not holds for $l \geq \left\lceil 8 \ln T \, / \, (\theta_{ij} + \frac{a_{ij} \ln T + b_{ij}}{T})^2 \right\rceil$.

So we get

$$\begin{aligned}
\mathrm{E}[n_i(T)] &\leq l + \sum_{t=l}^{\infty} \sum_{n_j=1}^{t-1} \sum_{n_i=l}^{t-1} \Pr\{\hat{\theta}_j(t) + C_{t,n_j} \leq \hat{\theta}_i(t) + C_{t,n_i}\} \\
&\leq l + \sum_{t=l}^{\infty} \sum_{n_j=1}^{t-1} \sum_{n_i=l}^{t-1} (t^{-(2-\delta_j(t))^2} + t^{-(2+\delta_i(t))^2}) \\
&\leq 8 \ln T \, / \, (\theta_{ij} + \frac{a_{ij} \ln T + b_{ij}}{T})^2 + C \\
&\leq 8 \ln T \, / \, \min_j \theta_{ij}^2 + C
\end{aligned}$$

where $C = \sum_{t=l}^{\infty} \sum_{n_j=1}^{t-1} \sum_{n_i=l}^{t-1} (t^{-(2-\delta_j(t))^2} + t^{-(2+\delta_i(t))^2})$ is a constant according to Riemann zeta function.

For $\theta_i > \theta_K$, arm i is picked at time t means one of the two cases happened: $\mathrm{O}_{\bar{K}} = \mathrm{O}_K^*$ or $\mathrm{O}_{\bar{K}} \neq \mathrm{O}_K^*$. Here we assume $t > N_{i,j}^*$ according to above derivation.

If $\mathrm{O}_{\bar{K}} = \mathrm{O}_K^*$, it implies the Step 2 in Algorithm 1 picks arm i wrongly at time t which gives : $I_i'(t) \leq I_K'(t)$.

If $\mathrm{O}_{\bar{K}} \neq \mathrm{O}_K^*$, it implies at least one arm $j \in \mathrm{O}_K^*$ but $j \notin \mathrm{O}_K$. So we have $I_i'(t) \leq I_j'(t)$.

So, we can conclude both cases for $\theta_i > \theta_K$: denote $\mathrm{O}_{K-1}^* = \mathrm{O}_K^* - \mathrm{A}_K$, there exists one arm $j \notin \mathrm{O}_{K-1}^*$ but $j \in \mathrm{O}_K$, such that $I_i'(t) \leq I_j'(t)$.

Similar analysis can be conducted to $I_i'(t) \leq I_j'(t)$, and we have

$$\begin{aligned}
\mathrm{E}[n_i(T)] &\leq l + \sum_{t=l}^{\infty} \sum_{n_j=1}^{t-1} \sum_{n_i=l}^{t-1} \Pr\{\hat{\theta}_j(t) - C_{t,n_j} \leq \hat{\theta}_i(t) - C_{t,n_i}\} \\
&\leq l + \sum_{t=l}^{\infty} \sum_{n_j=1}^{t-1} \sum_{n_i=l}^{t-1} (t^{-(2+\delta_j(t))^2} + t^{-(2-\delta_i(t))^2}) \\
&\leq 8 \ln T \, / \, (\theta_{ij} + \frac{a_{ij} \ln T + b_{ij}}{T})^2 + C \\
&\leq 8 \ln T \, / \, \min_j \theta_{ij}^2 + C
\end{aligned}$$

Hence, theorem 1 is proved.

References

[1] Q. Zhao and B. M. Sadler, “A Survey of Dynamic Spectrum Access,” Signal Processing Magazine, IEEE, vol. 24, no. 3. pp. 79–89, May-2007.

[2] O. Akan, O. Karli, and O. Ergul, “Cognitive radio sensor networks,” IEEE Netw., vol. 23, no. August, pp. 34–40, 2009.

[3] L. Wu, W. Wang, and Z. Zhang, “A POMDP-based optimal spectrum sensing and access scheme for cognitive radio networks with hardware limitation,” Wireless Communications and Networking Conference (WCNC), 2012 IEEE. Ieee, pp. 1281–1286, Apr-2012.

[4] Q. Zhao, L. Tong, A. Swami, and Y. Chen, “Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework,” Selected Areas in Communications, IEEE Journal on, vol. 25, no. 3. pp. 589–600, Apr-2007.

[5] K. Liu and Q. Zhao, “A Restless Bandit Formulation of Opportunistic Access: Indexablity and Index Policy,” Sensor, Mesh and Ad Hoc Communications and Networks Workshops, 2008. SECON Workshops ’08. 5th IEEE Annual Communications Society Conference on. Ieee, pp. 1–5, Jun-2008.

[6] K. Liu and Q. Zhao, “Decentralized multi-armed bandit with multiple distributed players,” Information Theory and Applications Workshop (ITA), 2010. pp. 1–10, 2010.

[7] A. Anandkumar, N. Michael, and A. Tang, “Opportunistic Spectrum Access with Multiple Users: Learning under Competition,” INFOCOM, 2010 Proceedings IEEE. Ieee, pp. 1–9, Mar-2010.

[8] Y. Gai and B. Krishnamachari, “Distributed Stochastic Online Learning Policies for Opportunistic Spectrum Access,” Signal Processing, IEEE Transactions on, vol. 62, no. 23. pp. 6184–6193, 2014.

[9] W.-Y. Lee and I. F. Akyldiz, “A Spectrum Decision Framework for Cognitive Radio Networks,” Mobile Computing, IEEE Transactions on, vol. 10, no. 2. pp. 161–174, 2011.

[10] S. Qu and Y. Xin, “Distribution of SNR and Error Probability of a Two-Hop Relay Link in Rayleigh Fading,” Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th. Ieee, pp. 1–4, Sep-2009.

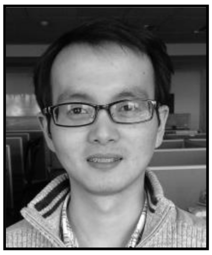
[11] D. Kalathil, N. Nayyar, and R. Jain, “Decentralized learning for multiplayer multiarmed bandits,” IEEE Trans. Inf. Theory, vol. 60, no. 4, pp. 2331–2345, 2014.

[12] A. Anandkumar, N. Michael, S. Member, A. K. Tang, and A. Swami, “Distributed Algorithms for Learning and Cognitive Medium Access with Logarithmic Regret,” Sel. Areas Commun. IEEE J., vol. 29, no. 4, pp. 731–745, 2011.

[13] P. Auer, N. Cesa-Bianchi, and P. Fischer, “Finite-time analysis of the multiarmed bandit problem,” Mach. Learn., vol. 47, no. 2–3, pp. 235–256, 2002.

[14] T. C. Scott, R. Mann, R. E. Martinez II, and R. E. Martinez, “General relativity and quantum mechanics: towards a generalization of the Lambert W function A Generalization of the Lambert W Function,” Appl. Algebr. Eng. Commun. Comput., vol. 17, no. 1, pp. 41–47, 2006.

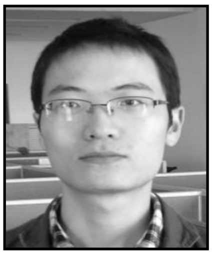
Authors



Yongqun Chen received the B.S. degree from School of Physics and Technology, Wuhan University, Wuhan, China, in 2009. He is now working on his M.S. and Ph.D. degrees in Electronical and Information Engeering in Wuhan University. His Current research interests include machine learning, cognitive radio and WLAN based indoor positioning.

Huaibei Zhou received the B.S. and M.S. degrees in physics from Wuhan University, Wuhan, China, in 1984 and 1987, respectively, and the Ph.D. degree in physics from the University of Maryland, College Park, in 1994. From 1994 to 1996, he was a Postdoctoral Fellow with the National Institute of Standard and Technology, Gaithersburg, MD. From 1996 to 1999, he was a Senior Engineer with General Electric. From 1999 to 2002, he was a Senior Technical Manager with NEXTEL. He is currently the Dean of the International School of Software Engineering, Wuhan University. His research interests include cognitive radio an agation.

Ruoshan Kong received the B.S. and Ph.D. degrees in computer science from Wuhan University, China in 2002 and 2007, respectively. He is currently an associate professor in the International School of Software, Wuhan University, China. His research interests include mobile communication protocols, IPv6 networking protocols, and cognitive network protocols.



Junyuan Huang received the B.S. degree in physics from Wuhan University, Wuhan, China, in 2009. Now he is working on his M.S. and Ph.D. degrees in electronics in Wuhan University. His research interests include cognitive radio and WLAN based indoor positioning.

Hang Qin received his B.S. Degree in Central China Normal University in 2001, M.S. Degree in Huazhong University of Science and Technology in 2004, Ph. D. Degree in Wuhan University in 2010, all in computer science. He is now an associate professor in Computer School, Yangtze University. His current research interests include cognitive radio networks, big data, machine learning, and smart grid.