future internet

MDPI

Article

Decentralized Blind Spectrum Selection in Cognitive Radio Networks Considering Handoff Cost

Yongqun Chen [1,*], Huaibei Zhou [2], Ruoshan Kong [2], Li Zhu [3] and Huaqing Mao [3]

[1] School of Physics and Technology, Wuhan University, Wuhan 430072, China
[2] International School of Software, Wuhan University, Wuhan 430072, China; bzhou@whu.edu.cn (H.Z.); krs1024@126.com (R.K.)
[3] Oujiang College, Wenzhou University, Chashan University Town, Wenzhou 325035, China; yeah_1397118@hotmail.com (L.Z.); oldmao_2001@163.com (H.M.)
* Correspondence: cyq321@whu.edu.cn or yiyuxiniao@gmail.com; Tel.: +86-27-6877-5482

Academic Editor: Francisco Javier Falcone Lanas
Received: 15 February 2017; Accepted: 28 March 2017; Published: 31 March 2017

Abstract: Due to the spectrum varying nature of cognitive radio networks, secondary users are required to perform spectrum handoffs when the spectrum is occupied by primary users, which will lead to a handoff delay. In this paper, based on the multi-armed bandit framework of medium access in decentralized cognitive radio networks, we investigate blind spectrum selection problem of secondary users whose sensing ability of cognitive radio is limited and the channel statistics are a priori unknown, taking the handoff delay as a fixed handoff cost into consideration. In this scenario, secondary users have to make the choice of either staying foregoing spectrum with low availability or handing off to another spectrum with higher availability. We model the problem and investigate the performance of three representative policies, i.e., ϱ^{PRE}, SL(K), kth-UCB1. The simulation results show that, despite the inclusion of the fixed handoff cost, these policies achieve the same asymptotic performance as that without handoff cost. Moreover, through comparison of these policies, we found the kth-UCB1 policy has better overall performance.

Keywords: cognitive radio networks; opportunistic spectrum access; multi-armed bandit; handoff cost; distributed algorithms

1. Introduction

In cognitive radio networks (CRNs), secondary users (SUs) may access a potentially large number of frequency bands or channels that are not occupied by primary users (PUs) at given time and space. Therefore, the coexistence of PUs and SUs becomes one of the key challenges while accessing the same part of the spectrum [1]. In an ideal condition, SUs must sense all channels before deciding which channel to access based on accessing strategy. However, in actuality, because of wide-band spectrum and hardware constraints, it is difficult for SUs to sense the entire operating spectrum band (300GHz) in a given period of time. Although compressive sensing is adopted as a wideband spectrum sensing technology for CRNs to solve this problem [2,3], little research has been done to implement feasible wide band spectrum sensing, as it is especially difficult to perform compressive sensing when prior knowledge of the primary signals is lacking. In fact, spectrum statistical information as a priori knowledge may not always be securable in a decentralized cognitive radio network. Hence, the blind sub-Nyquist wideband sensing is still an open issue in the field of compressive sensing for CRNs [4]. Some efforts [5] have been made to solve this problem. Cognitive compressive sensing has been formulated as a restless multi-armed bandit (rMAB) problem, which makes compressive sensing adaptive and cognitive.

In this paper, we investigate the blind spectrum selection problem for classical narrow-band spectrum sensing technology considering the handoff cost. In a decentralized CRN, prior knowledge of spectrum statistical information maybe not acquirable; in this context, many scholars have developed the multi-armed bandit (MAB) framework for opportunistic spectrum access (OSA) of CRN [6–9]. Anandkumar et al. [8] proposed a distributed algorithm named ϱ^{PRE} policy based on the ε_n-greedy policy [10]. Gai et al. [9] proposed a SL(K) subroutine and then established the prioritized access policy (DLP) and fair access policy (DLF) based on SL(K) and a pre-allocation order. Chen et al. [11] then proposed *k*th-UCB1 policy combined the ε_n-greedy and UCB1 policy and evaluated its performance both for real-time applications and best-effort applications. All of them achieve logarithmic regret in the long run.

Due to the spectrum varying nature of CRN, SUs are required to perform proactive spectrum handoffs when the spectrum band is occupied by PUs in the MAB framework, which results in a handoff delay consisting of RF reconfiguration or negotiation between transceiver In the above work [9–11], the handoff delay is not taken into consideration in their paper, i.e., the spectrum handoff is assumed to be costless. In this paper, by including a fixed handoff delay, SUs have to make the choice of either staying foregoing spectrum with low availability or handing off to a spectrum with higher availability and tolerating the handoff delay. We formulate this problem and investigate the performance of the above policies, i.e., ϱ^{PRE}, SL(K), and *k*th-UCB1. To the best of our knowledge, the influence of the handoff delay on the above policies in the MAB framework has not been investigated yet.

The rest of this paper is organized as follows: Section 2 describes the system model, which is similar to related works [9,12] except that the handoff delay is included as a handoff cost. In Section 3, we formulate the problem and present the three policies. In Section 4, we examine the proposed scheme through simulation. Finally, the paper concludes with a summary in Section 5.

2. System Model

The channel model of a cognitive radio network with $C, C \geq 2$ independent and orthogonal channels that are licensed to a primary network following a synchronous slot structure is illustrated in Figure 1. We model the channel availability W_i as an i.i.d. Bernoulli process with mean value $\beta_i \in$ B: $W_i \sim B(\beta_i)$, in which $W_i(t)$ denotes the "free"(denoted by 1) or "busy"(denoted by 0) state at time t of channel i. The SUs can access the free slot, which will yield a handoff delay if they access a channel that they did not access in the previous slot.

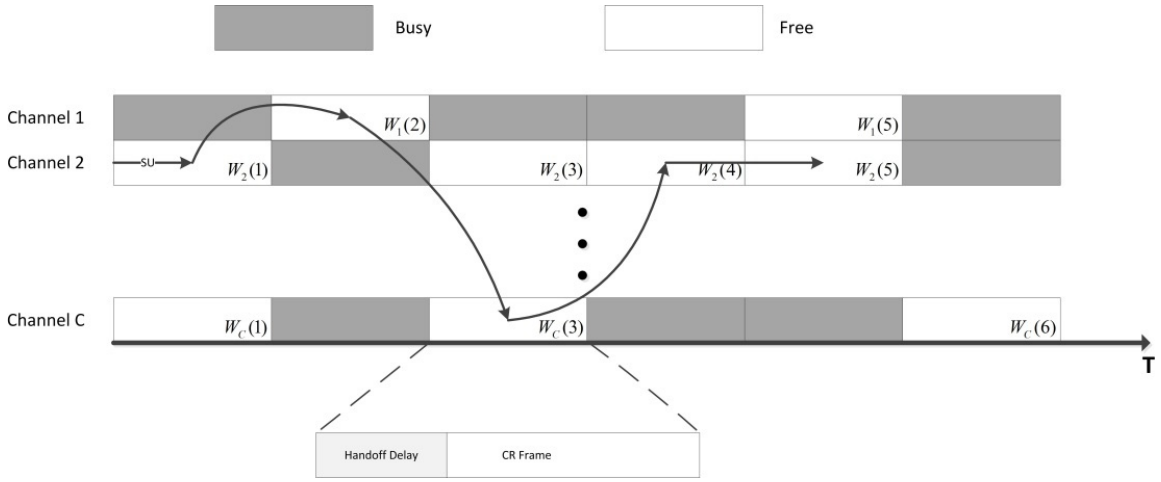


Figure 1. Slotted channel model.

The cognitive radio network is a composite of M secondary users. They access the channels in a decentralized way, i.e., there is no centralized entity collecting channel availability and channel state information (CSI) and then dispatching channels to SUs. In a slot cognitive radio network, a centralized entity will heavily impact the network performance. Regardless, we assume each SU has a pre-allocated rank that is dispatched by the network when the network forming or SU joins the network. For simplicity, we assume the priority of SU_j is ranked by j, i.e., the priority of SU_p is higher than SU_q if $p < q$ for either type of application. This rank is prior to learning and transmission processes and will not be changed afterwards.

The SUs behave in a proactive way to access the channels: They record past channel access histories and then utilize them to make predictions on future spectrum availability following a given policy. In addition, because of the inclusion of the handoff delay, SUs have to make the choice of either staying foregoing spectrum with relatively low availability or handing off to a spectrum with higher availability and tolerating the handoff delay. We denote the proportion of the handoff delay to the entire slot as fixed handoff cost H for simplicity.

The cognitive radio frame structure is shown in Figure 2. At the beginning of the frame, an SU chooses a channel to sense. Once the sensing result indicates that the channel is idle, the SU transmits pilot to receiver to probe CSI. The CSI is fed back through a dedicated error-free feedback channel without delay. The length of the data transmission is scalable and will be adopted by a transceiver according to the handoff delay or the data length in this scheme. At the end of the frame, the receiver acknowledges every successful or unsuccessful transmission as $Z_{i,j}(k) = 0$ for a collision that occurs; otherwise, it is 1.

Figure 2. Cognitive radio frame structure.

3. Problem Formulation and Policies

Blind spectrum selection in decentralized cognitive radio networks can be formulated as a decentralized MAB problem for multiple distributed SUs [8,9,12–14], and in this paper the terms "channel" and "arm" are used interchangeably. Denote π_j as the decentralized policy for SU j and $\pi = \{\pi_j, 1 \leq j \leq M\}$ as the set of homogeneous policies of all users. Arm i yields reward $X_i(t)$ at slot t according to its distribution, whose expectation is θ_i, $\theta_i \in \Theta$. Thus, the sum of the actual reward obtained by all users after T slots following policy π is

$$\sum_{t=1}^{T} S^{\pi}(t) = \sum_{t=1}^{T} \sum_{i=1}^{C} \sum_{j=1}^{M} X_i(t) \mathbb{I}_{i,j}(t) \tag{1}$$

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

In the ideal scenario where the availability statistics Θ are known, the SUs are orthogonally allocated to the $\mathcal{O}_M^*$ channels, where $\mathcal{O}_M^*$ is the set of M arms with M largest expected rewards. Then, the expected reward after the t slots is

$$\sum_{t=1}^{T} S^*(t) = T \sum_{i \in \mathcal{O}_M^*} \theta_i. \tag{2}$$

Then, we can define the performance of the policy π as regret $R_M^{\pi}(\Theta; T)$:

$$R_M^{\pi}(\Theta; T) = T \sum_{i \in \mathcal{O}_M^*} \theta_i - \mathbb{E}^{\pi}[\sum_{t=1}^{T} S_{\pi}(t)] \tag{3}$$

where $\mathbb{E}[\cdot]$ is the expectation operator.

We call a policy π uniformly good if for every configuration Θ, the regret satisfies

$$R_M^{\pi}(\Theta, T) = o(T^a) \text{ for } a > 0. \tag{4}$$

Such policies do not allow the total regret to increase rapidly for any Θ.

This problem is widely studied and several representative policies that are uniformly good are proposed: the distributed ϱ^{PRE} policy [8], the SL(K) policy [9], and the kth-UCB1 policy [11]. The distributed ϱ^{PRE} policy based on the ε_n-greedy policy, which prescribes to play the highest average reward arm with probability $1 - \varepsilon_n$ and a random arm with probability ε_n, and ε_n decreases as the experiment proceeds. However, one parameter of the policy requires prior evaluation of the arm reward means. To avoid this problem, the SL(K) policy is proposed based on the classical UCB1 policy of the MAB problem. Through it guarantees logarithm regret in the long run, it leads to a larger leading constant in the logarithmic order. The kth-UCB1 policy makes a good tradeoff on both policies.

Algorithm 1: ϱ^{PRE} policy for the user with rank K.

1. //Define:

   $n_i(t)$: the number of arm i is played after t slots.

   $\hat{\theta}_i(t)$: sample mean availabilities after t slots.

   $\varepsilon_t := \min[\frac{\beta}{t}, 1]$, where decay rate β is prior valuated according to the arm reward means
2. //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
3. //Main loop

   For $t = N + 1$ to T

   Step1: play the arm of the Kth highest index values in $\{\hat{\theta}_i(t)\}$ with probability $1 - \varepsilon_t$ and play a channel uniformly at random with probability ε_t

   Step2: Update $n_i(t)$, $\hat{\theta}_i(t)$ and ε_n

   EndFor

Algorithm 2: SL(K) policy for the user with rank K.

1. //Define:

   $n_i(t)$: the number of arm i is played after t slots.

   $\hat{\theta}_i(t)$: sample mean availabilities after t slots.
2. // 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
3. // Main loop

   For $t = N + 1$ to T

   Step1: Select the set $\mathcal{O}_K$ contains arms with the K highest index values:

$$\hat{\theta}_i(t-1) + \sqrt{\frac{2\ln t}{n_i(t-1)}}.$$

   Step2: Play the arm with the minimal index value in $\mathcal{O}_K$ according to

$$\hat{\theta}_i(t-1) - \sqrt{\frac{2\ln t}{n_i(t-1)}}.$$

   Step3: Update $n_i(t)$ and $\hat{\theta}_i(t)$.

   EndFor

Algorithm 3: kth-UCB1 policy for the user with rank K.

1. //Define:

   $n_i(t)$: the number of arm i is played after t slots.

   $\hat{\theta}_i(t)$: sample mean availabilities after t slots.

   $\varepsilon_t := \min[\frac{\beta}{t}, 1]$, where decay rate β is prior valuated according to the arm reward means
2. // 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
3. // Main loop

   For $t = N + 1$ to T

   Step1: Select the set $\mathcal{O}_K$ contains arms with the K highest index values.

$$\hat{\theta}_i(t-1) + \sqrt{\frac{2\ln t}{n_i(t-1)}}.$$

   Step2: with probability $1 - \varepsilon_t$ play the arm with minimum index value in $\mathcal{O}_K$ and with probability ε_t play an arm uniformly at random in $\mathcal{O}_K$.

   Step3: Update $n_i(t)$, $\hat{\theta}_i(t)$ and ε_n.

   EndFor

The above policies are derived and investigated in the scenario where there is no expense when the player switches from one arm to another. However, in CRNs, the handoff cost should be taken into consideration as illustrated in Section 2. Therefore, let

$$H^{\pi}(t) = H\sum_{i=1}^{C}\sum_{j=1}^{M}\mathbb{I}_{i,j}(t)\mathbb{J}_{i,j}(t) \tag{5}$$

be the sum of handoff cost of all users at slot t, where $\mathbb{J}_{i,j}(t)$ is the indicator if user j switches to arm i from other arms. Then, define the handoff regret as

$$HR^{\pi}(\Theta;T) = \mathbb{E}^{\pi}[\sum_{t=2}^{T}H^{\pi}(t)]. \tag{6}$$

We define the total regret as

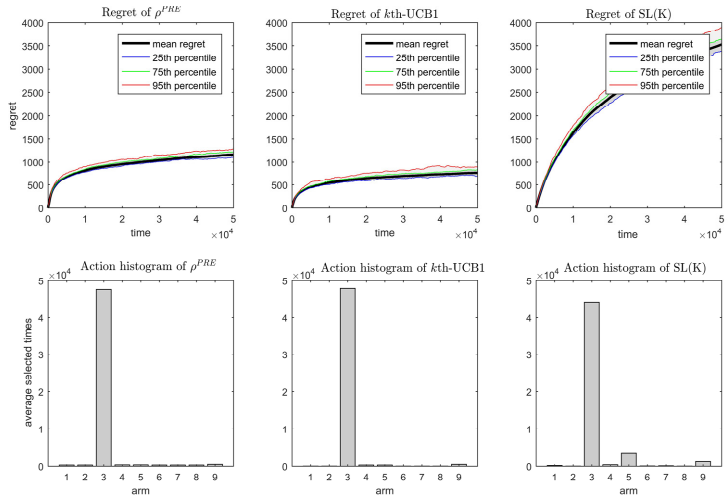
$$R^{\pi}(\Theta;T) = R_{M}^{\pi}(\Theta;T) + HR^{\pi}(\Theta;T) \tag{7}$$

Since the inclusion of the handoff delay H, $HR^{\pi}(\Theta;T)$, and $R_{M}^{\pi}(\Theta;T)$ are correlatively related, it is difficult to make a theoretical analysis of the total regret in a distributed multi-user case, although the authors of [15] considered this problem in a single-user case. In the next section, we examine the above policies by simulation and discuss their performance.

4. Simulation Results and Analysis

In this section, we present simulation results for the scheme proposed in this work. Simulations are done using Matlab, and we assume $C = 9$ channels with channel availabilities B = [0.5, 0.2, 0.8, 0.6, 0.9, .03, 0.4, 0.1, 0.7] and $M = 3$ SUs. The policy parameter configuration is the decay rate $\beta = 400$ for the ϱ^{PRE} policy and $\beta = 50$ for the kth-UCB1 policy, which is an optimal configuration according to the authors of [11], and the SL(K) policy is parameterless. The time scope is $T = 5 \times 10^4$. Every experiment is repeated 50 times.

As the regret of one user already takes the collision into consideration, the total regret of a CRN is simply the sum of all users in that CRN. Therefore, we present the regret and actions of one user in CRN where we take SU with the rank $K = 2$. Figure 3 shows the regrets and actions of these policies under fixed handoff cost $H = 0.1$.



(**a**) (**b**) (**c**)

Figure 3. The regrets and action histogram of ϱ^{PRE} (**a**), kth-UCB1 (**b**), and SL(K) (**c**) when $H = 0.1$.

From Figure 3, we can see that the three policies all achieve the logarithm regret and the actions converge to the third arm, which is the arm with the second-best channel availability. The regret of policy *k*th-UCB1 (Figure 3b) is smaller than that of policy ϱ^{PRE} (Figure 3a). This is caused by two aspects. Firstly, the decay rate β in *k*th-UCB1 can be smaller than ϱ^{PRE} and does not make the policy diverge. Secondly, the *k*th-UCB1 policy can distinguish the order-optimal arm from other arms more precisely than the ϱ^{PRE} policy by comparing the action histogram of Figure 3a,b, in which the number of arm 9 selected by the *k*th-UCB1 policy is smaller than that of policy ϱ^{PRE}. The regret of policy SL(K) in Figure 3c is largest, which means it has the largest leading constant in the logarithmic order.

We also investigated the regret of the three policies with varying handoff cost H as shown in Figure 4. As handoff cost is meaningless when the value is larger than 0.5, an H between 0 and 0.5 is chosen. From Figure 4, we see that the regret increases as H increases. Moreover, the growth rates of the three policies are all small when $H < 0.3$, which shows that these policies perform well. However, they become considerably large when $H > 0.3$. Intuitively, this is because a large H causes the arms to become indistinguishable.

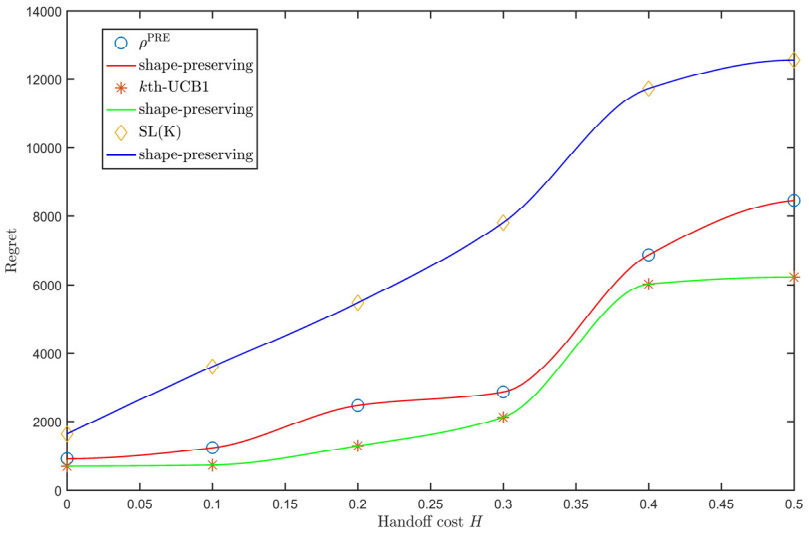


Figure 4. The regret with handoff cost (H).

5. Conclusions

In this work, we studied the blind spectrum selection in a decentralized cognitive radio networks in the MAB framework considering the handoff delay as a fixed handoff cost. We formulate this problem and investigate three representative policies and prove the uniform goodness of these policies for our scenario. Through simulation, we further show that, despite the inclusion of the fixed handoff cost, ϱ^{PRE}, SL(K), and *k*th-UCB1 achieve the same asymptotic performance as they do without handoff cost. Through comparison of these three policies, we found that the *k*th-UCB1 policy has better overall performance.

Acknowledgments: This work was supported by the International S & T cooperation Program of China under grand No. 2013DFA12460 and partially supported by the Zhejiang Provincial Natural Science Foundation of China under Grant No. LY16F010015 and by a grant from Public Projects of Wenzhou Science & Technology Bureau (Grant No. G20150020 & No. G20160007).

Author Contributions: Huibei Zhou and Yongqun Chen conceived and designed the experiments; Li Zhu and Huaqing Mao analyzed the data; Yongqun Chen and Ruoshan Kong wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Akyildiz, I.F.; Lee, W.-Y.; Vuran, M.C.; Mohanty, S. NeXt Generation/Dynamic Spectrum Access/Cognitive Radio Wireless Networks: A Survey. *Comput. Netw.* **2006**, *50*, 2127–2159. [CrossRef]
2. Salahdine, F.; Kaabouch, N.; el Ghazi, H. A survey on compressive sensing techniques for cognitive radio networks. *Phys. Commun.* **2016**, *20*, 61–73. [CrossRef]
3. Arjoune, Y.; Kaabouch, N.; el Ghazi, H.; Tamtaoui, A. Compressive Sensing: Performance Comparison of Sparse Recovery Algorithms. In Proceedings of the 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 9–11 January 2017; pp. 1–7.
4. Sun, H.; Nallanathan, A.; Wang, C.; Chen, Y. Wideband Spectrum Sensing for Cognitive Radio Networks: A Survey. *IEEE Wirel. Commun.* **2013**, *20*, 74–81.
5. Bagheri, S.; Scaglione, A. The Restless Multi-Armed Bandit Formulation of the Cognitive Compressive Sensing Problem. *IEEE Trans. Signal Process.* **2015**, *63*, 1183–1198. [CrossRef]
6. Wu, L.; Wang, W.; Zhang, Z. A POMDP-Based Optimal Spectrum Sensing and Access Scheme for Cognitive Radio Networks with Hardware Limitation. In Proceedings of the 2012 IEEE Wireless Communications and Networking Conference (WCNC), Paris, France, 1–4 April 2012; pp. 1281–1286.
7. Liu, K.; Zhao, Q. A Restless Bandit Formulation of Opportunistic Access: Indexablity and Index Policy. In Proceedings of the 5th IEEE Annual Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks Workshops, SECON Workshops '08, San Francisco, CA, USA, 16–20 June 2008; pp. 1–5.
8. Anandkumar, A.; Michael, N.; Tang, A. Opportunistic Spectrum Access with Multiple Users: Learning under Competition. In Proceedings of the 2010 IEEE INFOCOM Conference, San Diago, CA, USA, 15–19 March 2010; pp. 1–9.
9. Gai, Y.; Krishnamachari, B. Distributed Stochastic Online Learning Policies for Opportunistic Spectrum Access. *IEEE Trans. Signal Process.* **2014**, *62*, 6184–6193. [CrossRef]
10. Auer, P.; Cesa-Bianchi, N.; Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* **2002**, *47*, 235–256. [CrossRef]
11. Chen, Y.; Zhou, H.; Kong, R.; Huang, J.; Chen, B. QoS-Based Blind Spectrum Selection with Multi-armed Bandit Problem in Cognitive Radio Networks. *Wirel. Pers. Commun.* **2016**, *89*, 663–685. [CrossRef]
12. Anandkumar, A.; Michael, N.; Member, S.; Tang, A.K.; Swami, A. Distributed Algorithms for Learning and Cognitive Medium Access with Logarithmic Regret. *IEEE J. Sel. Areas Commun.* **2011**, *29*, 731–745. [CrossRef]
13. Liu, K.; Zhao, Q. Decentralized Multi-Armed Bandit with Multiple Distributed Players. In Proceedings of the 2010 Information Theory and Applications Workshop (ITA), La Jolla, CA, USA, 31 January–5 February 2010; pp. 1–10.
14. Kalathil, D.; Nayyar, N.; Jain, R. Decentralized learning for multiplayer multiarmed bandits. *IEEE Trans. Inf. Theory* **2014**, *60*, 2331–2345. [CrossRef]
15. Agrawal, R.; Hedge, M.V.; Teneketzis, D. Asymptotically efficient adaptive allocation rules for the multiarmed bandit problem with switching cost. *IEEE Trans. Autom. Control* **1988**, *33*, 899–906. [CrossRef]