32nd URSI GASS, Montreal, 19–26 August 2017

Bayesian Multi-Armed Bandit Framework for Multi-Band Auction Based Dynamic Spectrum Access in Multi-User Decentralized Networks

Unnam Abhishek(1) and S. J. Darak(2)
(1) Department of Mathematics, Delhi Technological University, New Delhi, India 110042
(2) ECE Department, IIIT-Delhi, India 110020

Abstract

Dynamic spectrum access (DSA) paradigm has shown significant potential to improve the electromagnetic spectrum utilization. For DSA in the multi-user decentralized networks, frequency band characterization and allocation among unlicensed users is a non-trivial task and various decision making policies (DMPs) have been proposed recently. Though auction based DMPs have shown to offer better performance than others, there is still a significant gap between the throughput of the decentralized DMPs and centralized or cooperative DMPs. In this paper, we propose a DMP for multi-user decentralized network comprising of Bayesian multi-armed bandit (MAB) framework for frequency band characterization and multi-band auction mechanism using Distributed Bipartite Matching Algorithm for frequency band allocation. The proposed Bayesian MAB framework makes the frequency band characterization accurate and faster thereby minimizing the selection of sub-optimal bands as well as the frequency band switching. Simulation results validate the superiority of the proposed DMP over existing state-of-the-art DMPs.

1 Introduction

Dynamic spectrum access (DSA) paradigm has generated significant interests in the academia as well as industry due to its potential to improve the utilization of an electromagnetic spectrum [1, 2]. In DSA, unlicensed users (UUs) can access vacant licensed frequency bands provided that they do not interfere with active licensed users. In the dynamic spectrum environment, frequency band quality may not be the same for each UU and hence, decision making, i.e., frequency band characterization and allocation, should be done by each UU independently. However, distinct UUs may transmit on the same band leading to collision. Collision not only leads to loss in throughput but also drainage of battery and hence, the decision making policies (DMPs) which guarantee collision-free transmissions in the decentralized network are desired.

Recently there has been some progress on the design of DMPs for DSA in the decentralized networks [2–8]. In [3], decentralized DMP using online learning algorithm and rank based approach has been proposed. It is extended further in [3–5] to improve the throughput. The main drawback of such DMPs is the large number of collisions thereby leading to loss in throughput as well as battery. In [6], hopping based DMP has been proposed where user can leave or enter the network anytime and it is collision-free. Since all the frequency bands are chosen with equal probability, throughput of the DMP in [6] is lower than other DMPs [3–5]. Recently, DMPs using auction approach have been proposed which are collision-free and hence, offer better throughput than other DMPs at the penalty of small communication cost [7, 8]. Still, there is a significant gap between the throughput of the decentralized DMPs and that of the centralized or cooperative DMPs.

In this paper, we propose a new DMP for the DSA in the multi-user decentralized network comprising of a multi-band auction mechanism using Distributed Bipartite Matching (DBM) algorithm [7, 9] for frequency band allocation and Bayesian multi-armed bandit (MAB) framework for frequency band characterization. The proposed Bayesian MAB framework is based on the Bayesian Upper Confidence Bound (BUCB) algorithm [10] for accurate frequency band characterization and lower upper confidence bounds (LUCB) [11] for identifying the subset of optimal frequency bands thereby reducing the exploration cost (i.e. number of sub-optimal band selection) of BUCB algorithm. Furthermore, proposed auction algorithm may allocate second frequency band to certain UU based on the quality of the first allocated band and the number of active UUs leading to further improvement in the throughput. Simulation results validate the superiority of the proposed DMP over existing state-of-the-art DMPs in terms of total throughput and the number of frequency band switching (FBS).

The paper is organized as follows. The assumed network model is presented in Section 2 followed by the proposed DMP in Section 3. In Section 4, simulation results are discussed. Section 5 concludes the paper.

2 Network Model

Consider the time slotted decentralized network consisting M UUs competing for N frequency bands. The throughput offered by each frequency band, when they are vacant, may not be same for all UUs. When i^{th} UU transmits on the vacant frequency band k, it gets zero throughput if there is a collision with other UU. If no collision, throughput is governed by i.i.d. distribution with mean μ_{ik} and variance, σ_n^2. Let $X_{i,k}(t)$ be the throughput received by user i when it transmits over the frequency band, k, in time slot, t. Then, total loss in throughput, referred to as regret $R(T)$, of the DMP with respect to genie-aided DMP is given by

$$R(T) = T\sum_{i=1}^{M}\mu_{i,k^*} - \mathbb{E}\Bigg[\sum_{t=1}^{T}\sum_{i=1}^{M}[\beta_i(t)X_{i,a_{i,1}(t)}(t) + (1-\beta_i(t))X_{i,a_{i,2}(t)}(t)]\Bigg] \quad (1)$$

$$\beta_i(t) = \begin{cases} 1 & a_{i,1}(t)\text{is vacant} \\ 0 & a_{i,1}(t) \text{ is occupied} \end{cases} \quad (2)$$

where T is horizon, μ_{i,k^*} is the average throughput of the frequency band assigned to the i^{th} UU in genie-aided DMP, $a_{i,1}(t)$ and $a_{i,2}(t)$ are the first and second frequency bands allocated to i^{th} UU, respectively. In Eq. 1, $X_{i,a_{i,2}(t)}(t) < X_{i,a_{i,1}(t)}(t)$ for any $a_{i,1}(t) = a_{i,2}(t)$. In addition to regret, the number of times UU switches the frequency band should be as low as possible. This is because, each FBS incurs penalty in throughput as well as battery due to hardware reconfigurations and communication overheads. The total number of FBS, $S(T)$, is given by,

$$S(T) = \mathbb{E}\Bigg[\sum_{t=1}^{T}\sum_{i=1}^{M}[\beta_i(t)\beta_i(t-1)\mathbf{1}_{\{a_{i,1}(t)=a_{i,1}(t-1)\}} + \beta_i(t)(1-\beta_i(t-1))\mathbf{1}_{\{a_{i,1}(t)=a_{i,2}(t-1)\}} + (1-\beta_i(t))\mathbf{1}_{\{a_{i,1}(t)=a_{i,2}(t)\}}]\Bigg] \quad (3)$$

where $\mathbf{1}_{\{.\}}$ is an indicator function. The aim of the proposed DMP discussed next is to minimize the regret i.e. loss in throughput, $R(T)$ as well as the number of FBS, $S(T)$.

3 Proposed Decision Making Framework

In this section, the proposed decision making framework for frequency band characterization and allocation is presented. The various decision making steps in each time slot of the proposed framework are shown in Algorithm 1. Steps 2-15 deals with the frequency band allocation to UUs and hence, referred to as allocation phase. In the allocation phase, the proposed DMP employs auction algorithm (discussed in sub-section 3.1) for frequency band allocation, BUCB algorithm (discussed in sub-section 3.2) for frequency band characterization and LUCB algorithm (discussed in sub-section 3.3) to identify optimal subset of frequency bands. Execution of allocation phase depends on the value of the counter, η, and t, as shown in Algorithm 1. Please refer to [7] for more details on η selection.

Steps 17-28 of Algorithm 1 are referred to as transmission phase and they are executed in each time slot. In the transmission phase, every UU senses the first allocated frequency band, i.e., $a_{i,1}(t)$. If vacant, corresponding i^{th} UU transmits over it and updates the throughput, $X_{i,a_{i,1}(t)}(t)$. Otherwise, second allocated band, $a_{i,2}(t)$, if any, is sensed. If vacant, i^{th} UU transmits over it and updates the throughput, $X_{i,a_{i,2}(t)}(t)$. Otherwise, UU remains idle until the beginning of the next time slot.

3.1 Proposed Auction Algorithm

For the orthogonal frequency band allocations, i.e. $a_{i,1}(t)$ and $a_{i,2}(t)\forall i$, the proposed modified DBM auction approach is given in Algorithm 2. Auction is executed at certain time slots chosen according to the counter, η, as shown in Algorithm 1.This means that the probability of auction diminishes as more UU settles in their preferred

---

Algorithm 1 Proposed Decision Making Policy for i^{th} UU

1: **Initialization** Set counter $\eta = 1$, $\bar{N}_i$=N, $\bar{k}_i \in \{1,2,..,\bar{N}_i\}$, B_i = $\{1,2,...,N\}$.
2: **while** (t <= T) **do**
3:    **if** ($t = p\cdot T^*$ for some $p = 0,1,2....$) **then**
4:       $\bar{N}_i$=min($2M$, N)
5:       Obtain the indexes of $\bar{N}_i$ number of optimal frequency bands using LUCB algorithm and store their indices in B_i, i.e. $|B_i| = \bar{N}_i$ $\forall i$.
6:    **if** ($\eta = 2^p$ for some $p = 0,1,2....$) **then**
7:       Using BUCB algorithm, obtain quality factor of $\bar{N}_i$ frequency bands whose indexes are stored in B_i.
8:       Participate in DBM Auction given in Algorithm 2 to obtain allocated frequency bands, $a_{i,1}(t)$and $a_{i,2}(t)$.
9:       **if** ($a_{i,1}(t) \neq a_{i,1}(t-1)$) **then**
10:          Send an INTERRUPT to all UUs about frequency band change.
11:          Reset $\eta = 1$
12:       **if** INTERRUPT Received **then**
13:          Reset $\eta = 1$
14:    **else**
15:       $a_{i,1}(t) = a_{i,1}(t-1)$, $a_{i,2}(t) = a_{i,2}(t-1)$
16:    Sense the frequency band $a_{i,1}(t)$
17:    $T_{i,a_{i,1}(t)}(t) = 1$
18:    **if** $a_{i,1}(t)$ is vacant **then**
19:       Transmit and observe the instantaneous throughput, ξ_1. Update $X_{i,a_{i,1}(t)}(t) = \xi_1$
20:    **else**
21:       Sense the frequency band $a_{i,2}(t)$
22:       Update $X_{i,a_{i,1}(t)}(t) = 0$ and $T_{i,a_{i,2}(t)}(t) = 1$
23:       **if** $a_{i,2}(t)$ is vacant **then**
24:          Transmit and observe the instantaneous throughput, ξ_2. Update $X_{i,a_{i,2}(t)}(t) = \xi_2$
25:       **else**
26:          $X_{i,a_{i,2}(t)}(t) = 0$.
27:    Increment counter $\eta = \eta + 1$, $t = t + 1$

Algorithm 2 Proposed DBM Auction Algorithm for UU i

1: **Initialization** Initialize prices of all bands to zero, i.e., $p_{\bar{k}} = 0 \; \forall k$
2: UU identifies the preferred frequency band from unallocated bands using BUCB quality index and calculates its bid. For the preferred band index, $\bar{k}$, bid is $\chi_{i,\bar{k}} = \max_{\bar{k}}(\mu_{i,\bar{k}} - p_{\bar{k}}) - \text{second} \max_{\bar{k}}(\mu_{i,\bar{k}} - p_{\bar{k}}) + \frac{\varepsilon}{M} - p_{a_{i,1}(t)}$
3: After receiving bids from all UUs, UU with the highest bid wins the corresponding band. Set $p_{\bar{k}} = \mu_{i,\bar{k}}$ where UU i is the winner of the band $\bar{k}$
4: Follow steps 2-4 till the UU gets the band.
5: **if** $M \leq N$ **then**
6: Follow steps 2-4 until UU gets second band or all bands are allocated.

frequency band. In the beginning, each UU sends the bid for its most preferred frequency band to all other UUs as shown in step 2 of Algorithm 2. The UUs with highest bid get their respective preferred bands. Remaining UUs again follow the same process iteratively but every time they send the revised bid for their preferred frequency band chosen from the unallocated bands. Fist stage auction completes when each UU gets the band which is denoted as $a_{i,1}(t) \forall i$. If $M < N$, then the proposed auction algorithm allocates the remaining frequency bands as second choice frequency bands, referred to as $a_{i,2}(t)$, as shown in steps 6-7 of the Algorithm 2. However, the priority is given to the UU which has the most sub-optimal band in the first allocation. This is accomplished by adding the term $p_{a_{i,1}(t)}$ in bid calculation which is zero in the first band allocation and gives lower preference to UUs with optimal first band.

3.2 Frequency Band Characterization

In the auction algorithm, each UU needs to send its bid (or price) for the preferred frequency band which has not been allocated to any UU yet. In order to determine the preferred band, UU needs to characterize the frequency bands. In the proposed DMP, we have formulated the frequency band characterization problem into MAB framework. Here, each arm is analogous to frequency band and the task is identify the best arm, i.e. band. Selection of bands in each time slot is done by MAB algorithm which needs to balance between the exploration of N bands and exploitation of optimal bands. MAB algorithms include frequentist approach based Upper Confidence Bound (UCB) algorithm [12], Bayes-UCB (BUCB) [10] and Thompson Sampling (TS) algorithms [4, 10]. Though these algorithms are asymptotically optimal, it has been recently proved that BUCB and TS algorithms offer better performance and have lower computational complexity than others [10]. Empirically, we observed that the BUCB algorithm offers slightly better performance than TS algorithm and more importantly, it leads to fewer number of the FBS. These advantages make the BUCB algorithm a preferred choice for the proposed DMP.

In the proposed DMP, frequency bands are characterized based on their quality (i.e. throughput) using BUCB index as shown in Eq. 4 [10].

$$G_{i,\bar{k}}(t) = Q\left\{ 1 - \frac{1}{t}; Beta\left[\sum_{v=1}^{t} X_{i,\bar{k}}(v) + 1, \sum_{v=1}^{t} T_{i,\bar{k}}(v) - \sum_{v=1}^{t} X_{i,\bar{k}}(v) + 1 \right] \right\} \forall \bar{k} \quad (4)$$

where $Q(x)$ is the probability that any normal random variable gets a value larger than x standard deviations above the mean and $Beta$ represents the complete beta function, i.e., Euler integral of the first kind. In order to choose the preferred frequency band in step 2 of Algorithm 2, UU selects the frequency band having the maximum value of the BUCB quality index in Eq. 4 among the unallocated bands.

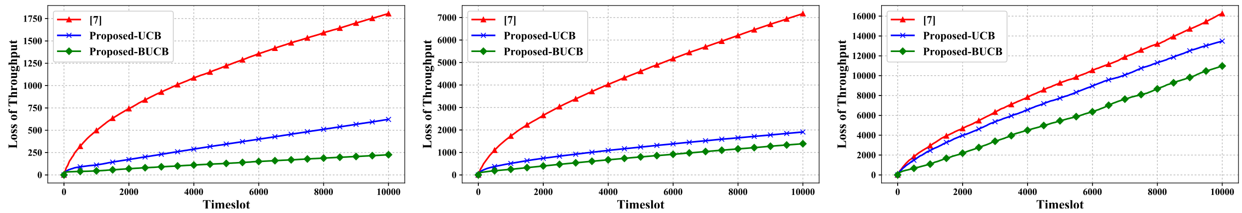
3.3 Frequency Band Partitioning: LUCB

The task of the frequency band characterization using BUCB algorithm becomes challenging as the number of frequency bands, N, increases. This is because of exploration-exploitation trade-off which requires BUCB algorithm to select all frequency bands sufficient number of times to guarantee accurate characterization. In order to further limit the selection of sub-optimal bands without compromising on the characterization accuracy of optimal bands, LUCB [11] algorithm is used to identify $\bar{N}_i$ bands for all UU. This is achieved by comparing the lower and upper confidence bounds on the throughput offered by each band. Please refer to [11] for more details. The LUCB is invoked at regular intervals after initial exploration period as shown in Steps 3-5 of Algorithm 1.

4 Results

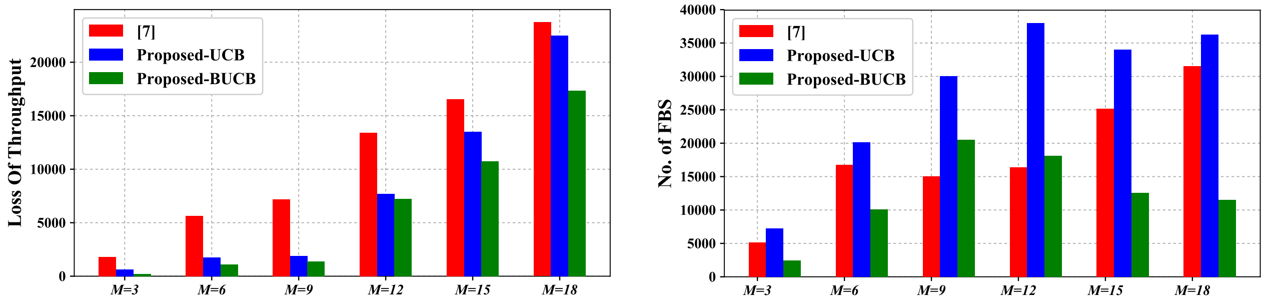
In this section, the performance of the proposed DMP is compared with the existing state-of-the-art DMP in [7] in terms of throughput and the number of FBS. Since the DMP in [7] has shown to be superior than DMP in [8], the comparison with the latter is not done here for clarity of the figures. Simulations consider 20 distinct frequency bands and the number of UUs are varied from 3 to 18. Each numerical result reported hereafter is the average of the values obtained over 50 independent experiments and simulations consider a time horizon of 10000 iterations. The statistics, $\mu_{i,k}$, are randomly generated at the start of each experiment and remains the same throughout but may change from one experiment to another. Also, $\mu_{i,k}$ may not be same as $\mu_{j,k}$ for UUs, i and j. The value of T^* in Algorithm 1 is 1000.

In Fig. 1a, Fig. 1b and Fig. 1c, the regret, i.e. loss in throughput, at various stages of the horizon is shown for $M = \{3, 9, 18\}$, respectively. We consider two variations of the proposed DMP: 1) Proposed DMP designed with UCB algorithm and, 2) Proposed DMP designed using BUCB (+LUCB) algorithms. Note that DMP in [7] employs UCB



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

Figure 1. Total loss in throughput at various stages of horizon for (a) M=3, (b) M=6, and (c) M=18.



(a) **(b)**

Figure 2. (a) Total loss in throughput and (b) Total number of FBS at the end of Horizon for different values of M.

for frequency band characterization. The superior performance of UCB based proposed DMP validates the superiority of the proposed auction algorithm for frequency band allocation over DMP in [7] while the superior performance of BUCB based proposed DMP over other DMPs validates the superiority of proposed MAB framework consisting of BUCB and LUCB algorithms for accurate frequency band characterization. In Fig. 2a and Fig. 2b, the regret and FBS for different values of M are shown. It can be observed that proposed DMP offers superior performance for all M.

5 Conclusions and Future Works

In this paper, we propose a decision making policy (DMP) for dynamic spectrum access in multi-user decentralized network. The proposed DMP consists of Bayesian multi-armed bandit framework for frequency band characterization and multi-band auction mechanism using Distributed Bipartite Matching Algorithm for frequency band allocation. Simulation results validate the superiority of the proposed DMP over existing state-of-the-art DMPs in terms of total throughput and the number of frequency band switchings. Future works includes in-depth regret analysis and extension of the proposed DMP for full duplex radios.

Acknowledgment

This work is supported by the DST Inspire Faculty Fellowship granted by the Department of Science and Technology, Govt. of India.

References

[1] X. Hong, J. Wang, C-X. Wang and J. Shi, "Cognitive Radio in 5G: A Perspective on Energy-Spectral Efficiency Trade-off," in *IEEE Communication Magazine*, vol. 52, no. 7, July 2014.

[2] Y. Xu, A. Anpalagan, Q. Wu, L. Shen, Z. Gao and J. Wang, "Decision-Theoretic Distributed Channel Selection for Opportunistic Spectrum Access: Strategies, Challenges and Solutions," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 4, pp. 1689–1713, Nov. 2013.

[3] A. Anandkumar, N. Michael, A. Tang and A. Swami, "Distributed Algorithms for Learning and Cognitive Medium Access With Logarithmic Regret," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 731–745, April 2011.

[4] S. J. Darak, S. Dhabu, C. Moy, H. Zhang, J. Palicot and A. P. Vinod , "Decentralized Spectrum Learning and Access for Heterogeneous Cognitive Radio Networks," *Elsevier Digital Signal Processing*, vol. 37, pp. 13–23, Feb. 2015.

[5] S. J. Darak, H. Zhang, J. Palicot and C. Moy, "An Efficient Policy for D2D Communications and Energy Harvesting in Cognitive Radios: Go Bayesian!," in 23th *European Signal Processing Conference (EUSIPCO)*, pp. 1236–1240, Nice, France, Aug. 2015.

[6] G. Zhang, A. Huang, H. Shan, J. Wang, T. Q. S. Quek and Y. D. Yao, "Design and Analysis of Distributed Hopping-Based Channel Access in Multi-Channel Cognitive Radio Systems with Delay Constraints," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 11, pp. 2026–2038, Nov. 2014.

[7] D. Kalathil, N. Nayyar and R. Jain, "Decentralized Learning for Multiplayer Multiarmed Bandits," *IEEE Transactions of Information Theory*, vol. 60, no. 4, pp. 2331-2345, April 2014.

[8] M. Zandi and A. Grami, "Dynamic Spectrum Access via Channel-Aware Heterogeneous Multi-Channel Auction With Distributed Learning" *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 5913 - 5926, Nov. 2015.

[9] D. P. Bertsekas, "The Auction Algorithm: A distributed Relaxation Method for the Assignment Problem," *Ann Oper Res*, vol. 14, no. 1, pp. 105-123, 1988.

[10] E. Kaufmann, O. Cappé and A. Garivier, "On Bayesian Upper Confidence Bounds for Bandit Problems," in 15th *Int. Conf. on Artificial Intelligence and Statistics*, pp. 592–600, Canary Islands, Apr. 2012.

[11] S. Kalyanakrishnan, A. Tewari, P. Auer and P. Stone, "PAC Subset Selection in Stochastic Multi-armed Bandits," *Proceedings of the* 29th *International Conference on Machine Learning (ICML)*, pp. 655–662, Jun. 2012.

[12] P. Auer, N. Cesa–Bianchi and P. Fischer, "Finite–time Analysis of the Multiarmed Bandit Problem," *Machine Learning*, vol. 47, no. 2, pp. 235–256, 2002.