Is Bayesian Multi-armed Bandit Algorithm Superior?: Proof-of-Concept for Opportunistic Spectrum Access in Decentralized Networks

Sumit J. Darak[1(✉)], Amor Nafkha[2], Christophe Moy[2], and Jacques Palicot[2]

[1] IIIT-Delhi, Okhla Phase III, New Delhi 110020, India
sumit@iiitd.ac.in
[2] SCEE, CentraleSupélec/IETR, CS 47601, 35576 Cesson-Sévigné Cedex, France

Abstract. Poor utilization of an electromagnetic spectrum has led to surge of interest in paradigms such as cognitive radio, unlicensed LTE etc. Such paradigms allow opportunistic spectrum access in the vacant frequency bands of the licensed spectrum. Though various spectrum detectors to check the status of frequency bands (i.e., vacant or occupied) have been studied, the selection of the frequency band from wideband spectrum is a challenging problem especially in the decentralized network. In this paper, a testbed for analyzing the performance of decision making policies (DMPs) for identifying optimum frequency band in the decentralized network is presented. Furthermore, experimental results using real radio signals show that the proposed DMP using Bayesian multi-armed bandit algorithm leads to 7–12 % improvement in an average spectrum utilization over existing DMPs. Added advantages of 6–20 % lower switching cost and 30–46 % fewer collisions make the proposed DMP energy-efficient.

Keywords: Cognitive radio · Multi-armed bandit algorithm · Opportunistic spectrum access · Usrp testbed

1 Introduction

Further boosting the average utilization of an electromagnetic spectrum has led to a surge of interest in paradigms such as cognitive radio (CR), device-to-device (D2D) communications, unlicensed LTE (LTE-U) etc. from the academia as well as industrial partners [1–4]. Such paradigms allow opportunistic spectrum access (OSA) in the vacant licensed bands with the constraint of zero interference to the active licensed users. For example, an underlay inband unlicensed D2D communications allow direct communication between secondary (or unlicensed) users (SUs) over identified vacant licensed subband(s) without the need of base stations or access points for establishing the communication link. Also, CR and LTE-U allow SUs to access vacant TV white and 5 GHz spectrum, respectively. Though most of the existing works focus on the centralized networks, decentralized networks would be an efficient choice over the centralized approach due to

D. Noguet et al. (Eds.): CROWNCOM 2016, LNICST 172, pp. 104–115, 2016.
DOI: 10.1007/978-3-319-40352-6_9

advantages such as ease of implementation, robustness to link or node failures, no communication overhead and lower delay [1,2]. Furthermore, decentralized approach is preferred choice for public safety networks and proximity-aware social networking services.

Decision making policies (DMPs) are desired for OSA in the decentralized networks in order to: (1) Enable SUs to identify optimum vacant subbands, (2) Minimize collisions among SUs, and (3) Keep the subband switching cost (SSC) as low as possible. Here, SSC stands for the total penalty incurred in terms of delay, power, hardware reconfiguration and protocol overhead when SU switches from one frequency subband to another. From the energy efficiency perspective, SSC and number of collisions should be as low as possible. Design of such DMP for decentralized CRNs is a challenging task and one of the objective the work presented in this paper.

In this paper, USRP based testbed for analyzing the performance of DMPs for OSA in decentralized networks is presented. To the best of our knowledge, the proposed testbed is the first proof-of-concept which compares the performance of various DMPs using real radio signals. Furthermore, experimental results using real radio signals show that the proposed DMP using Bayesian multi-armed bandit algorithm leads to 7–12 % improvement in an average spectrum utilization over existing DMPs. Added advantages of 6–20 % lower switching cost and 30–46 % fewer collisions make the proposed DMP energy-efficient.

The paper is organized as follows. The detailed literature review of DMPs for decentralized CRNs is given in Sect. 2. The proposed DMP is presented in Sect. 3 followed by the proposed testbed description in Sect. 4. The experimental results are discussed in Sect. 5. Section 6 concludes the paper.

2 Literature Review: Decision Making Policies

Various DMPs have been proposed for decentralized CRNs [3–9]. These DMPs consist of MAB algorithms such as frequentist approach based upper confidence bound (UCB) algorithm and its extensions, $\varepsilon-$greedy, optimization based Kullback-Leibler UCB (KL–UCB) to estimate subband statistics (e.g., probability of being vacant, transition probability from vacant state to occupied state and vice-versa etc.) [5,10]. Such algorithms are based on exploration-exploitation trade-off where they explore all subbands multiple number of times before settling down to optimum subband with respect to desired statistics. These algorithms are asymptotically optimal with logarithmic regret that is the best one can expect when there is no prior information about subband statistics. Such optimality guarantees that the estimated statistics of all subbands, and not just optimum subband, are closed to their actual statistics [5,10].

Another challenging task of the DMP is to orthogonalize SUs to different frequency bands and minimize collisions between them. This is a challenging problem considering SUs do not share any information with each other. In ρ^{rand} DMP with M SUs [4], each SU randomly and independently chooses the rank, $R(k) \in \{1, 2, ..M\}$ in the beginning. In subsequent time slots, underling MAB algorithm calculates the quality index for each subband. Then, the SU with the rank $R(k)$ chooses the subband with the $R(k)^{th}$ best quality index. Another DMP in [6] follows time division fare share approach where the rank of each SU is rotated in circular fashion between 1 to M to allow an equal access to the optimum subbands among all SUs. In [4,6], a new rank is randomly and independently chosen for SU experiencing collision. Though both DMPs are asymptotically optimal with logarithmic regret, SSC of [6] increases linearly with t. The SSC can be minimized using the DMP, ρ^{rand}, in [4], where the rank is changed only when corresponding SU collides with other SUs. The performance of the ρ^{rand} DMP is further improved in [7]. In [7], the range for rank i.e. $1 \leq R(k) < N$, $\forall k$ is made wider to minimize the number of collisions. In [8], variable filtering architecture and its integration with tunable subband access DMP, ρ^{t_rand}, is proposed that takes into account tunable bandwidth requirements of SUs. Still, the average SSC of [7,8] is quite high. Recently, Bayesian MAB algorithms such as Bayes-UCB and Thompson Sampling have become more popular and are proved to be efficient than the other MAB algorithms [9,10]. However, experimental analysis of these MAB algorithms using real radio signals and non-ideal detectors has not been done yet.

3 Proposed Decision Making Policy

The proposed DMP consists of subband statistic estimation using Bayes-UCB algorithm and subband access scheme for orthogonalization of SUs. The proposed DMP is discussed in detail next.

3.1 System Model

Consider the slotted CRN consisting of multiple primary users and M SUs. The wideband spectrum of bandwidth, B, is divided into N uniform subbands of bandwidth, B_{cmin}. Hence, $B_{cmin} = (B/N)$. The status of each subband (i.e., vacant or occupied) is independent of the status of other subbands. For a given subband, vacancy statistic depends on the underlining traffic model which can be either independent and identically distributed (i.i.d.) model or Markovian model. In case of i.i.d. model, the status of $i^{th}, i \in \{1, 2, ...N\}$ subband depends on its $P_{vac}(i)$ distribution and is independent of its status in previous time slots. In case of Markovian model, i^{th} subband switches its state from being vacant to occupied and vice versa according to a discrete Markov process with the probabilities of $p_{vo}(i)$ and $p_{ov}(i)$, respectively. Then, using Markov chain analysis, steady-state probabilities of subband being vacant, denoted as $P_{vac}(i), i \in \{1, 2, ...N\}$, are given by [11]

$$P_{vac}(i) = \frac{p^i_{ov}}{p^i_{ov} + p^i_{vo}}, \ \forall i \tag{1}$$

Basic assumptions made in this paper for SUs in decentralized CRNs are:

1. Infrastructure less decentralized CRN where all SUs employ the same DMP but do not exchange information with other SUs.

2. SU can sense only one subband in each time slot.
3. $P_{vac}(i)$, $p_{vo}(i)$ or $p_{ov}(i), i \in \{1, 2, ...N\}$ are unknown to SUs.
4. All SUs must sense the chosen frequency band at the start of each time slot irrespective of sensing outcomes in the previous time slots.

At the beginning of each time slot, DMP of SU chooses the subband for sensing. Let $N_k(t) \in \{1, 2, .., N\}$ be the subband chosen by k^{th} SU in time slot t. The analog-front end and digital front-end of SU filter the chosen subband, down-convert it to baseband and passed it to the spectrum detector. If the subband is vacant, it is assumed that the SU transmits over that subband. When multiple SUs transmit on the same subband i.e., $N_k(t) = N_j(t)$ for any $k \neq j$, collision occurs leading to failed transmission. Otherwise, it is assumed that SU transmits successfully. Let $\Delta_k(t)$ be instantaneous reward of k^{th} SU in time slot t and is given by,

$$\Delta_k(t) = \begin{cases} 1 & \text{No collission} \\ 0 & \text{Collision} \end{cases} \tag{2}$$

Let $r_k(t)$ be the total number of successful transmissions by k^{th} SU and is given by,

$$r_k(t) = r_k(t-1) + \Delta_k(t) \tag{3}$$

Let $S^*(t)$ and $S(t)$ denote the total number of successful transmissions by genie-aided DMP (i.e. the DMP where $P_{vac}(i), \forall i$ are known a priori and central unit allocates distinct subband to each SUs) and decentralized DMP, respectively. Then, total loss in terms of transmission opportunities, $U(t)$, up to time t is given by Eq. 4 and should be as small as possible.

$$U(t) = S^*(t) - S(t) = \sum_{k=1}^{M} \sum_{v=1}^{t} \mathbb{E}[r_k^*(v) - r_k(v)] \tag{4}$$

In addition, SSC and number of collisions, $C(t)$, given by Eqs. 5 and 6, respectively, should be as minimum as possible.

$$SSC(t) = \sum_{k=1}^{M} \sum_{v=2}^{t} \mathbb{E}[\mathbf{1}_{N_k(v) \neq N_k(v-1)}] \tag{5}$$

$$C(t) = \sum_{k=1}^{M} \sum_{v=1}^{t} \Delta_k(v) \tag{6}$$

3.2 Subband Statistic Estimation

The Bayes-UCB algorithm employed in the proposed DMP of k^{th} SU is given in Algorithm 1. Here, H is horizon size, $T_k(i, t)$ indicates the number of times the subband i is chosen by k^{th} SU up to time t, $S_k(i, t)$ indicates the number of times out of $T_k(i, t)$, the subband i is observed as vacant by k^{th} SU and $R(k)$ is the rank of k^{th} SU. The rank calculation is discussed later in Sect. 3.3. Initially, all subbands are sensed once as shown in Steps 1–6 of Algorithm 1. Then, at each time slot, $t > N$, Bayes-UCB algorithm calculates quality index, $q(i,t), \forall i$ for each subband as shown in Step 12 of Algorithm 1. This is done by calculating quantile of order i for a given beta distribution. Based on the values of $q(i,t), \forall i$, SU chooses the subband according to its rank $R(k)$ as shown in Step 14. After getting the status of the chosen subband from spectrum detector, parameters $S_k(N_k(t),t)$ and $T_k(N_k(t),t)$ are updated as shown in Steps 15–18. Note that, $S_k(i,t) \leq T_k(i,t), \forall i, \forall k$.

Algorithm 1: Subband Selection and Statistic Estimation Using Bayes-UCB Algorithm for k^{th} Secondary User

Parameters: $N, R(k), i \in \{1,2,..,N\}, \mathrm{H}$
Input: $\mathrm{S}_k(i,t-1),\ \mathrm{T}_k(i,t-1),\ \forall i$
Output: $N_k(t), \mathrm{S}_k(i,t),\ \mathrm{T}_k(i,t),\ \forall i$

1. **if** $(any(\mathrm{T}_k(:,t-1) == 0))$
2. $\quad N_k(t) = i$ s.t. $\mathrm{T}_k(i,t-1) = 0$
3. $\quad$ **if** $N_k(t)$ is vacant
4. $\quad\quad \mathrm{S}_k(N_k(t),t) = \mathrm{S}_k(N_k(t),t) + 1$
5. $\quad$ **end**
6. $\quad \mathrm{T}_k(N_k(t),t) = \mathrm{T}_k(N_k(t),t) + 1$
7. **else**
8. $\quad$ **for** $t = N + 1$ **to** H **do**
9. $\quad\quad \mathrm{S}_k(:,t) = \mathrm{S}_k(:,t-1)$
10. $\quad\quad \mathrm{T}_k(:,t) = \mathrm{T}_k(:,t-1)$
11. $\quad\quad$ **for** $i = 1$ **to** N **do**
12. $\quad\quad\quad$ **Compute**
$\quad\quad\quad q(i,t) = Q\left\{1 - \frac{1}{t}; Beta[\mathrm{S}_k(i,t) + 1, \mathrm{T}_k(i,t) - \mathrm{S}_k(i,t) + 1]\right\}$
13. $\quad\quad$ **end**
14. $\quad\quad$ Select the band $N_k(t) = i$
$\quad\quad\quad$ s.t. $q(i,t)$ is $R(k)^{th}$ max. value of $q(:,t)$
15. $\quad\quad$ **if** $N_k(t)$ is vacant
16. $\quad\quad\quad \mathrm{S}_k(N_k(t),t) = \mathrm{S}_k(N_k(t),t) + 1$
17. $\quad\quad$ **end**
18. $\quad\quad \mathrm{T}_k(N_k(t),t) = \mathrm{T}_k(N_k(t),t) + 1$
19. $\quad$ **end**
20. **end**

Bayes-UCB algorithm is asymptotically optimal which means that

$$\limsup_{t\to\infty} \frac{\mathbb{E}[T_k(i,t)]}{\ln t} \geq \frac{1}{KL(P_{vac}(i), P_{vac}(i^*))}, \quad \forall i, \forall k \tag{7}$$

where KL stands for the Kullback-Leibler divergence. Equation 7 indicates that Bayes-UCB algorithm achieves optimal balance between exploration of all subbands and exploitation of optimum subband. Also, it guarantees that learned subbands statistics are closed to their actual values. Another advantage of Bayes-UCB algorithm is that SSC is low through empirical observations. This means that Bayes-UCB algorithm chooses the same band consecutively more number of times than other MAB algorithms like UCB. This feature might be advantageous for accurate estimation of transition probabilities, i.e., P_{vo} and P_{ov}. However, usefulness of transition probabilities for DMP is out of scope of this paper. Next, subband access scheme is presented.

3.3 Subband Access in Multi-user Decentralized Network

For any DMP, the upper bound on $U(t)$ in Eq. 4, is given by [4],

$$U(t) \leq P_{vac}(1^*)\left\{ \sum_{k=1}^{M} \sum_{i \in Z\_worst} \mathbb{E}[T_k(i,t)] + \mathbb{E}[C(t)] \right\} \tag{8}$$

where $P_{vac}(1^*)$ is the highest vacancy statistics among N subbands, Z_worst is set of all subbands excluding first M subbands when arranged according to decreasing values of P_{vac} and $C(t)$ are the number of collisions when SU chooses any of the first M subbands. For lower $U(t)$, subband ordering at each SU should be accurate and rank selection must be orthogonal. Hence, for OSA in a multi-user decentralized network, accurately estimation of subband statistics alone is not sufficient. In addition, DMP needs to avoid collisions, $C(t)$, among active SUs.

In the proposed DMP, modified randomization based subband access scheme, proposed in [8], has been used. The proposed subband access scheme in [8] is based on ρ_{rand} [4]. The novelty of the proposed scheme is that the upper limit of rank is tunable based on subband statistics compared fixed value of rank (equal to M) in ρ_{rand} [4]. This leads to fewer collisions and hence, higher number of transmission opportunities and lower SSC compared to ρ_{rand} [4].

4 Proposed USRP Testbed

The proposed USRP testbed is shown in Fig. 1 and is an significant extension of the testbed in [13]. It consists of two units: (1) Left hand side unit is primary user traffic generator, and (2) Right hand side unit acts as DMP of secondary user(s). Both the units are discussed in detail next.

4.1 Primary User Traffic Generator

The chosen design environment for the primary user traffic generation is GNU Radio Companion (GRC) and the hardware platform is made of a USRP from Ettus Research. The main reason for choosing GRC is the precise control on each parameter of the transmission chain compared to other environments. The proposed primary user traffic generator is shown in Fig. 2. In the beginning, number of frequency bands, traffic model (i.i.d. or Markovian) and corresponding subband statistics are taken from the user using the block named Traffic Model in

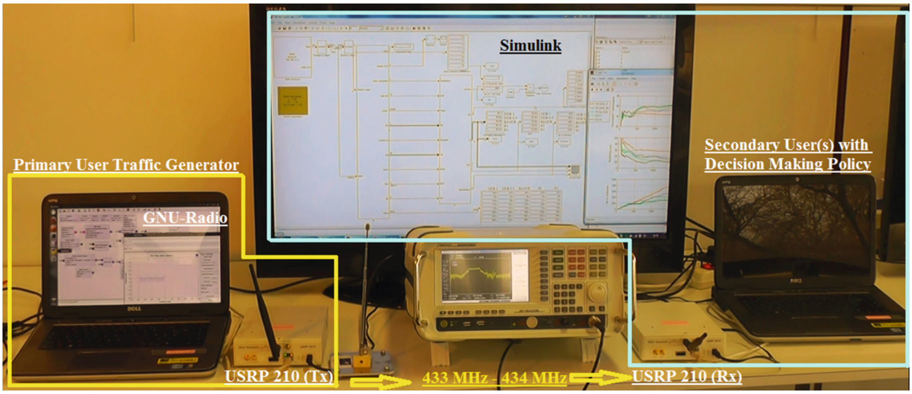

Fig. 1. Proposed USRP based testbed for analyzing the performance of DMPs using real radio signals and non-ideal spectrum detectors.

Fig. 2. The transmission bandwidth, which is restricted by bandwidth of analog front-end of USRP, is divided into N subbands. In each time slot, masking vector of size N is generated by Traffic Model block based on given subband statistics. This masking vector can have 1 or 0 values where 1 and 0 indicate that corresponding band is occupied and vacant, respectively. Next step is mapping data to be transmitted on sub-carriers of occupied bands. The data modulation used is a differential QPSK modulation with Gray encoding. This is followed by sub-carrier mapping using OFDM and transmission via USRP. In the proposed tested, number of sub-carriers, center frequency and transmission bandwidth are 256, 433.5 MHz and 1 MHz, respectively. For demonstration purpose, each time slot duration is one second so that it can be followed by human eye. However, it can be reduced to the order of milliseconds and will have no direct effect on the performance of DMP.

4.2 Secondary User with Decision Making Policy

The chosen design environment for the SU terminal is Matlab/Simulink and USRP from Ettus Research. USRP is tuned to receive signal of bandwidth 1 MHz centered at 433.5 MHz. The received signal is then down-sampled, digitized and passed to the DMP implemented using Simulink. The DMP uses MAB algorithm to select single subband in each time slot. The chosen subband is sensed using energy detector. Note that energy detector is not ideal and sensing error may occur [12]. If subband is sensed as vacant, it is assumed that SU transmits over the chosen subband. If multiple SUs choose the same subband, it is assumed that both users suffer collision and fail to transmit any data. In case of multiple SUs, each user is independently implemented in Simulink with their respective DMP. In existing works, sensing is assumed as perfect which is not true in real radio conditions. Thus, proposed testbed with non-ideal spectrum detectors will enable to study performance of DMPs in the presence of sensing errors. Note

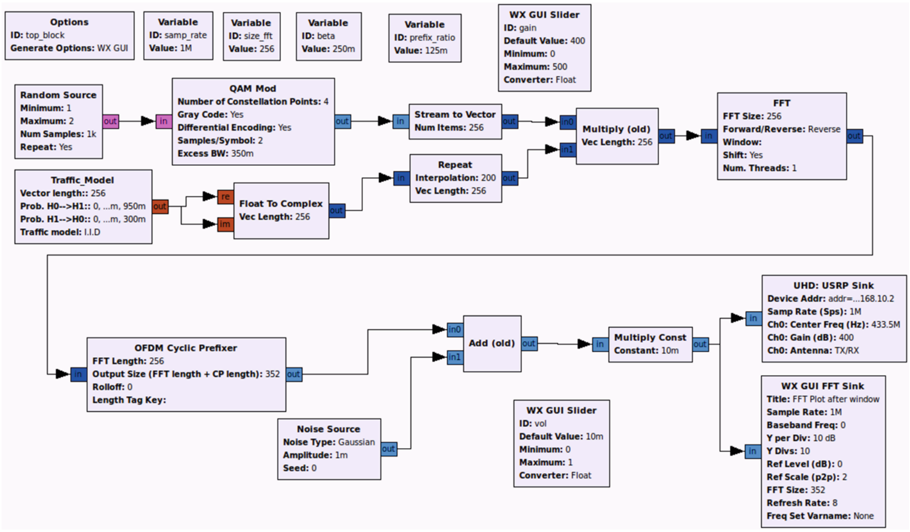

Fig. 2. Proposed primary user traffic generator.

that performance comparisons of various detectors and their effect on DMPs is not discussed here due to brevity of the paper.

4.3 Synchronization

The synchronization between transmitter and receiver is an important aspect of slotted decentralized network infrastructure considered in this paper. For demonstration purposes, synchronization has been achieved by switching first band from occupied to vacant states or vice-versa in each time slot. This enables SUs to detect the transitions between OFDM symbols as well as to synchronize the energy detection phase on an entire OFDM symbol of the primary traffic. In a real OSA scenario, SU should be able to synchronize with PU network via synchronization signals or pilot carriers. Note that the synchronization band in the proposed approach is not wasted because DMP does not restrict it as synchronization band but consider it as possible option for data transmission.

5 Experimental Results and Analysis

In this section, experimental comparison of various DMPs on the proposed test-bed, discussed in Sect. 4, is presented. For $N = 8$ and $B = 1$ MHz, we have $B_{cmin} = 125$ KHz. In case of i.i.d. rewards, two different P_{vac} distributions, denoted as case 1[1] and case 2[2] are considered. Similarly, P_{vo} and P_{ov} distributions

[1] P_{vac} :- [.50 .10 .20 .30 .40 .60 .80 .90].
[2] P_{vac} :- [.50 .05 .95 .10 .80 .60 .40 .75].

for Markovian rewards are given by case 3[3]. Each numerical result reported hereafter is the average of values obtained over 15 independent experiments on USRP testbed and each experiment consider a time horizon of 1000 iterations i.e. 1000 time slots for each SU and one time slot corresponds to one second. It is assumed that all SUs employ the same DMP but do not exchange any information with others.

For Case 1, Fig. 3a and b show total number of successful transmissions, $S(t)$, in percentage for various DMPs w.r.t. genie-aided DMP when $M = 2$ and $M = 4$, respectively. For simplicity, policies UCB($\alpha = 2$)+ρ_{rand} [4], UCB($\alpha = 0.5$)+ρ_{rand} [4] and KLUCB+ρ_{rand} are hereafter referred to as E1, E2 and E3, respectively. Here, α is an exploration factor of UCB algorithm and it should be between 0.5 to 2. It can be observed that the proposed DMP offers higher transmission opportunities compared to existing DMPs. Since the probability of collision among SUs increases with M, $S(t)$ in % is lower when $M = 4$ compared to $M = 2$. On the other hand, average spectrum utilization for $M = 4$ is higher than the same when $M = 2$. For instance, average spectrum utilization due to licensed users was only 47 % for Case 1. Due to OSA with $M = 2$, average spectrum utilization can be increased to 58 %, 63 %, 66 % and 70 % using E1, E2, E3 and proposed DMPs, respectively. In case of $M = 4$, average spectrum utilization can be increased to 70 %, 76 %, 78 % and 81 % using E1, E2, E3 and proposed DMP, respectively.

Figures 4 and 5 show total number of successful transmissions, $S(t)$, in percentage for various DMPs w.r.t. genie-aided DMP in Case 2 and Case 3, respectively. Improvements in the average spectrum utilization, similar to Case 1, can also be seen for Case 2 and 3. To conclude, proposed DMP leads to higher avearge spectrum utilization and hence, superior compared to existing DMPs for various traffic distributions.

As discussed in Sect. 1, SSC should be as minimum as possible for making SU terminals energy efficient. In Fig. 6 (a), the SSC of different DMPs are compared

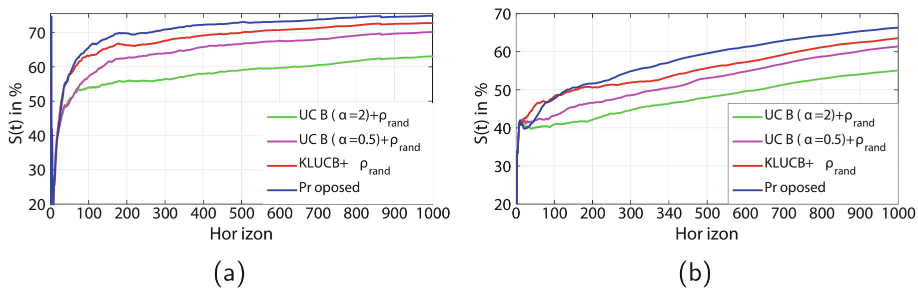

Fig. 3. Comparisons of average $S(t)$ in % of different DMPs with respect to the genie-aided DMP in Case 1 for (a) $M = 2$ and (b) $M = 4$.

[3] P_{vo} :- [.50 .05 .10 .20 .30 .40 .50 .60]
P_{ov} :- [.50 .80 .70 .60 .50 .40 .30 .20].

for subband distributions in Case 1, Case 2 and Case 3. It can be observed that SSC of the proposed DMP is low. Numerically, average SSC of the proposed DMP is 20 %, 6 % and 2 % lower than that of E1, E2 and E3, respectively. In additions to SSC, the number of collisions of DMP should be as minimum as possible. This is because, collision leads to waste of the energy required for transmission of corresponding data and it may be higher than the energy required for subband switching. In Fig. 6 (b), the number of collisions suffered by all SUs are compared for subband distributions in Case 1, Case 2 and Case 3. Numericaly, SUs employing proposed DMP suffers 46 %, 30 % and 13 % less number of collisions than SUs employing policies E1, E2 and E3, respectively. Thus, lower SSC and collisions of the proposed DMP makes it energy efficient and suitable for resource constrained battery operated SU terminals.

In terms of computational complexity, KLUCB is based on optimization approach and is the most computationally complex. The complexity of Bayes-UCB is slightly lower than that of UCB [10]. Based on experimental results, we argue that proposed DMP using Bayesian MAB algorithm for OSA in multi-user decentralized network is not only superior but also energy efficient.

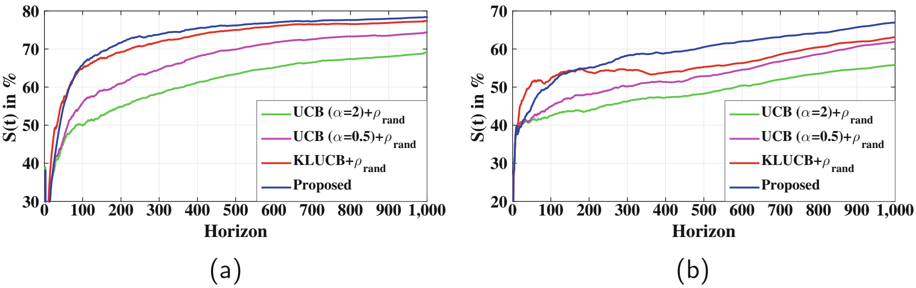

Fig. 4. Comparisons of average $S(t)$ in % of different DMPs with respect to the genie-aided DMP in Case 2 for (a) $M = 2$ and (b) $M = 4$.

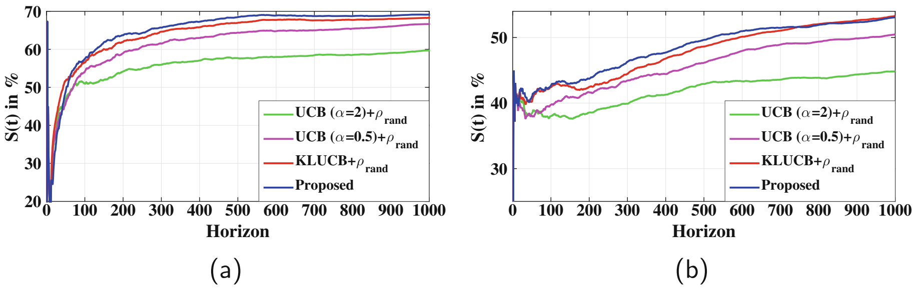

Fig. 5. Comparisons of average $S(t)$ in % of different DMPs with respect to the genie-aided DMP in Case 3 for (a) $M = 2$ and (b) $M = 4$.

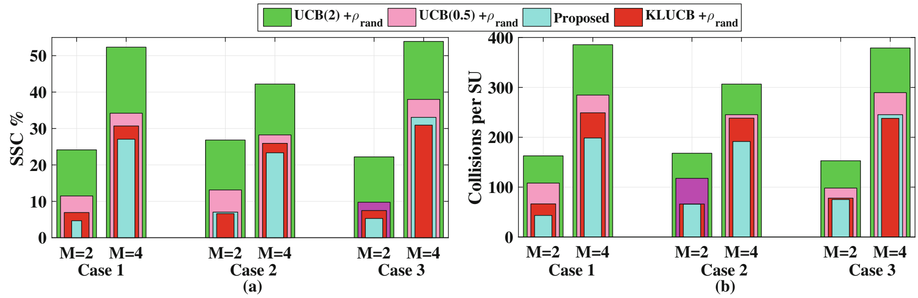

Fig. 6. (a) Comparison of SSC for various P_{vac} distributions, (b) Comparison of number of collisions for various P_{vac} distributions.

6 Conclusion and Future Works

A USRP based testbed for experimentally analyzing the performance of decision making policies (DMPs) for opportunistic spectrum access (OSA) in the decentralized cognitive radio networks has been proposed. To the best of our knowledge, the proposed testbed is the first proof-of-concept which compares the performance of various DMPs using real radio signals. Furthermore, experimental results showed that the proposed DMP designed using Bayesian multi-armed bandit (MAB) algorithm offers superior performance over existing policies in terms of average spectrum utilization, subband switching cost as well as number of collisions. Thus, we argue that Bayesian MAB algorithm based DMPs are superior for OSA in the decentralized networks. Future work involves study of the effect of various spectrum detectors on the performance of DMPs and realization of actual data transmission on the chosen frequency band.

Acknowledgments. The authors would like to thank Department of Science and Technology (DST), Government of India for INSPIRE fellowship in support of this work.

References

1. Asadi, A., Wang, Q., Mancuso, V.: A survey on device-to-device communication in cellular networks. IEEE Commun. Surv. Tutorials **16**(4), 1801–1819 (2014)
2. Phunchongharn, P., Hossain, E., Kim, D.I.: Resource allocation for device-to-device communications underlaying LTE-advanced networks. IEEE Wirel. Commun. **20**(4), 91–100 (2013)
3. Xu, Y., Anpalagan, A., Wu, Q., Shen, L., Gao, Z., Wang, J.: Decision-theoretic distributed channel selection for opportunistic spectrum access: strategies, challenges and solutions. IEEE Commun. Surv. Tutorials **15**(4), 1689–1713 (2013)
4. Anandkumar, A., Michael, N., Tang, A., Swami, A.: Distributed algorithms for learning and cognitive medium access with logarithmic regret. IEEE J. Sel. Areas Commun. **29**(4), 731–745 (2011)

5. Auer, P., Cesa-Bianchi, N., Fischer, P.: Finite-time analysis of the multiarmed bandit problem. Mach. Learn. **47**(2), 235–256 (2002)
6. Liu, K., Zhao, Q.: Distributed learning in multi-armed bandit with multiple players. IEEE Trans. Signal Process. **58**(11), 5667–5681 (2010)
7. Zandi, M., Dong, M.: Learning-stage based decentralized adaptive access policy for dynamic spectrum access. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5323–5327, Vancouver, Canada, May 2013
8. Darak, S.J., Zhang, H., Palicot, J., Moy, C.: Efficient decentralized dynamic spectrum learning and access scheme for multi-standard multi-user cognitive radio networks. In: 11th International Symposium on Wireless Communication Systems (IEEE ISWCS 2014), pp. 271–275, Barcelona, Spain, August 2014
9. Darak, S.J., Zhang, H., Palicot, J., Moy, C.: An efficient policy for D2D communications and energy harvesting in cognitive radios: go bayesian! In: 23rd European Signal Processing Conference (EUSIPCO), pp. 1236–1240, Nice, France, August 2015
10. Kaufmann, E., Cappé, O., Garivier, A.: On bayesian upper confidence bounds for bandit problems. In: 15th International Conference on Artificial Intelligence and Statistics, pp. 592–600, La Palma, Canary Islands, April 2012
11. Hoang, D.T., Niyato, D., Wang, P., Kim, D.I.: Opportunistic channel access and RF energy harvesting in cognitive radio networks. IEEE Trans. Sel. Areas Commun. **32**(11), 1–14 (2014)
12. Bahamou, S., Nafkha, A.: Noise uncertainty analysis of energy detector: bounded and unbounded approximation relationship. In: 21st European Signal Processing Conference (EUSIPCO), pp. 1–4, Marrakech, Morocco, September 2013
13. Robert, C., Moy, C., Wang, C.X.: Reinforcement learning approaches and evaluation criteria for opportunistic spectrum access. In: IEEE International Conference on Communications (ICC), pp. 1508–1513, Sydney, Australia, June 2014