A Thompson Sampling Approach to Channel Exploration-Exploitation Problem in Multihop Cognitive Radio Networks

Viktor Toldov, Laurent Clavier, Valeria Loscrì, Nathalie Mitton

A Thompson Sampling Approach to Channel Exploration-Exploitation Problem in Multihop Cognitive Radio Networks

Viktor Toldov[1,2], Laurent Clavier[2,3], Valeria Loscrí[1], Nathalie Mitton[1]
[1]Inria firstname.lastname@inria.fr, [2]IRCICA USR CNRS 3380, Université Lille 1, IEMN,
[3]Institut Mines-Télécom, Télécom-Lille, firstname.lastname@telecom-lille.fr

***Abstract*—Cognitive radio technology is a promising solution to the exponential growth in bandwidth demand sustained by increasing number of ubiquitous connected devices. The allocated spectrum is opened to the secondary users conditioned on limited interference on the primary owner of the band. A major bottleneck in cognitive radio systems is to find the best available channel quickly from a large accessible set of channels. This work formulates the channel exploration-exploitation dilemma as a multi-arm bandit problem. Existing theoretical solutions to a multi-arm bandit are adapted for cognitive radio and evaluated in an experimental test-bed. It is shown that a Thompson sampling based algorithm efficiently converges to the best channel faster than the existing algorithms and achieves higher asymptotic average throughput. We then propose a multihop extension together with an experimental proof of concept.**

I. Introduction

Electromagnetic spectrum is a limited natural resource. Usable bands of the spectrum have already been assigned to specific applications. However, advances in semiconductor technology have lead to the proliferation of compact embedded wireless devices. Consequently the demand for accessing the spectrum has grown exponentially and does not seem to abate [1]. Cognitive radio is a promising approach to accommodate for this exponential growth in bandwidth demand. In this approach, the already allocated spectrum is opened to secondary users conditioned on limited interference to the primary user of the spectrum. To access the spectrum efficiently, the secondary user has to be able to detect the available opportunities when the primary user is not accessing the spectrum and vacate it. In passive or deterministic spectrum access, the secondary user has a priori knowledge about spectrum utilization by the primary user. However, the prior information about the primary user statistics may not be available or may change over time, *e.g.* some channels may be lightly used in some intervals in parts of a cellular network and could be used to provide connectivity for secondary network users. Hence it is crucial for the secondary network to identify the best available channels as soon as possible. A naive approach is to scan the whole band and identify the available channels. Naive approach imposes large latency and energy cost on the network specially when the number of accessible channels is large. In addition, to the best of our knowledge, advantages of multihop cognitive radio has not been exploiting so far. and not covered by the standard [2]. As a result, finding the best channel with minimum time and energy cost and apply cognitive scheme to multihop communications are still two open problems.

This work is partially supported by CPER NPdC DATA and LIRIMA PREDNET project. We are grateful to A. Maskooki, A. Lazaric and R. Petrolo for their constructive inputs and discussions.

In this work we address the non-trivial channel exploration-exploitation dilemma in cognitive radio in an experimental setting and extend it to multihop communications. The problem is defined as follows. A large number of channels are available for potential usage by secondary network users. The channels are occupied with different rates by primary user which has the exclusive right to use it. The problem is to find and exploit the channel with the best rate as soon as possible. The contributions of this work can be summed up:

- we model the channel selection dilemma as a multi-arm bandit problem and propose a Thompson [3] sampling-based approach;
- we provide an experimentation of our approach and compare its performances to the most efficient and simply implementable learning approaches in terms of throughput both by simulation and experiments.
- we propose a multihop cognitive radio extension together with an experimental proof of concept.

We show that the Thompson sampling formulation finds the best channel significantly faster and the most efficient among the other algorithms in a practical cognitive radio setting. This results in smaller latency in transmission and less energy consumption for channel exploration. This paper is organized as follows: Section II reviews the state of the art in the channel exploration problem in cognitive radio setting with a focus in Section III on the implementation of some existing algorithms. Section IV provides the derivations and implementation details of our Thompson sampling based problem while Section V compares the performances of the algorithms on real-time measurements. Section VI-A presents how to extend our approach to a multihop configuration before Section VII concludes this work.

II. Related Works

Channel decision process is analyzed based on a channel model as addressed in [4]. The optimal strategy is derived using a partially observable Markov decision process (POMDP) framework which models the channel as a 2^N state Markov chain. Each state is a binary N-tuple indicating the availability of each channel. The model is reduced into N two-state Markov models where the transition probabilities are constant and based on the statistics of the primary user assuming independence between the channels. In [5] and [6], co-existence of a cognitive radio system a IEEE 802.11 primary network is investigated. The channel is modeled by a four state continuous-time Markov model where the states of the Markov chain represent data packet transmission, acknowledgment transmission, short inter-frame spacing time and idle channel. The transition probabilities are obtained from measurement data. The sojourn time is defined as the time that the process spends in each state and estimated by a generalized Pareto distribution as comprehensively studied in [7].

[8] and [9] formulate the channel selection dilemma as a multi-arm bandit in . Algorithms based on upper confidence bound (UCB) algorithm suggested in [10] are applied to identify the channel with the best availability rate. In [8], the performance of UCB based algorithm in terms of regret is investigated for different numbers of accessible channels. Regret is the throughput missed due to not choosing the best channel compared to an oracle who knows and exploits the best channel from the beginning. In [9], dependencies are considered between the channels and a UCB based algorithm is proposed and analyzed in terms of regret. In [11] a cognitive scenario with two competing networks with jamming capability is considered. The decision on jamming or communicating on the channel for each network is modeled in a Bayesian setting for Thompson sampling. In this adversarial scenario, it is shown that the Bayesian framework outperforms UCB based algorithms. Channel decision problem is modeled as a restless multi-arm bandit problem in [12]. The probing time to find an available channel is minimized using the model. The performance of the method is numerically evaluated and is shown to have advantage over previous methods.

Although multi-arm bandit model has been studied extensively in theory, its applicability and efficiency in a cognitive radio framework still needs to be evaluated experimentally before being implemented commercially. The reason is that in practice, the perfect theoretical assumptions are often not valid which can affect the performance of the algorithms. In this work we implemented the theoretically efficient algorithms in a practical setting which closely mimics a cognitive radio context. The details of the simulations and experiments are given in the following sections. Also, to the best of our knowledge, cognitive radio mechanisms have never been studied for multihop communications.

III. Optimal Algorithms for Multi-Arm Bandit

In this section, we formulate channel selection and exploration dilemma in cognitive radio context as a multi-arm bandit problem. Then, we describe the efficient and simple learning algorithms commonly used to solve the multi-arm bandit problems and adapt them into the cognitive radio context.

In a multi-arm bandit problem, an agent tries to obtain as much reward as possible by playing the most rewarding arm among N arms. However, each arm rewards randomly upon being played according to an unknown distribution. Hence, the objective is to minimize exploration to find the most rewarding arm. A policy A is an algorithm that defines the actions of the agent usually based on the previous observations. We assume n_j to be the number of times j^{th} arm has been played after n steps and μ_j to be the expected reward of j^{th} arm. In other words, channel j is found available in average $\mu_j n_j$ times in n_j measurements. μ_j is associated with the statistics of the primary user of the channel. The regret of the policy R_A is defined to indicate how much reward is lost on the average due to the exploration, $R_A = \mu^* n - \sum_{j=1}^{N} \mu_j E(n_j)$ where $\mu^* = \max_{1 \leq j \leq N} \mu_j$ and $E(.)$ indicates the expected value. Minimizing the regret is desirable as it will maximize the average reward.

In [13], authors have derived a logarithmic lower bound for the regret function in a multi-arm bandit problem,

$$R(n) = ln(n) \left[\sum_{i=1}^{N} \frac{p^* - p_j}{D(p^* || p_j)} + o(1) \right], \tag{1}$$

where p_j are the reward density of the arms and p^* is the density of the arm with the maximum average reward (μ^*). D is the Kullback-Leibler divergence between the two densities and $o(1) \to 0$ as $n \to \infty$.

The user must find the best available channel among N accessible channels as fast as possible considering its time and energy constraints. We assume changes in the statistics of the primary user are slower than the convergence time of the algorithms [11], [14]. To keep track of the primary user statistics, an expiration time can be defined to trigger the search for a new channel together with the number of consecutive unsuccessful channels access. In [10], policies based on upper confidence index are investigated and shown to achieve optimal regret uniformly over time. In this work, we evaluate the performance of learning algorithms from [10] with ϵ_n-greedy and Thompson sampling, an old algorithm which has recently gained interest in research community due to its simplicity and efficiency. Upper confidence bound (UCB) algorithms are based on empirical mean obtained through observations and a term related to the upper confidence interval of the empirical mean. UCB1-based algorithm chooses the channel with the highest upper confidence bound defined as,

$$c_j = \overline{x}_j + \sqrt{\frac{2ln(n)}{n_j}}, \tag{2}$$

where $\overline{x}_j$, n, n_j are the empirical mean of channel j states, total number of channel access and number of times channel j is accessed so far. The values are updated in every iteration based on the observations. The decision criterion is proportional to the empirical average of the obtained rewards and is

known as the exploitation factor. A second term triggers the exploration and is inversely proportional to the square root of the number of times the channel is accessed (n_j). Further details and proofs of optimality are presented in [10].

UCB2 is an improved version of UCB1 presented in [10]. In UCB2 based channel access, the channel access is divided into epochs. At each epoch, the best performing channel is selected according to the selection criterion defined as:

$$c_j = \overline{x}_j + a_j(n, r_j), \tag{3}$$

where $a_j(n, r_j)$ is defined as,

$$a_j(n, n_j) = \sqrt{\frac{(1+\alpha) ln(\frac{en}{\tau(r_j)})}{2\tau(r_j)}} \tag{4}$$

and r_j, the number of epochs channel j is acceded is $\tau(r_j) = \lceil (1+\alpha)^{r_j} \rceil$. In each epoch, the channel is accessed $\tau(r_j + 1) - \tau(r_j)$ times. α is a parameter of the model and should be tweaked based on the setting. Theoretical solution does not define α specifically. The value of α is tuned experimentally based on the context. Our simulation results show that $\alpha = 0.01$ optimizes the performance of the UCB2 algorithm in our experimental setting.

Another well-known and simple policy in bandit problems is ϵ-greedy algorithm. The agent chooses the most rewarding arm based on the previous observations with probability $1-\epsilon$ and a random arm with probability ϵ. As shown in [10], decreasing ϵ proportional to $\frac{1}{n}$ bounds the regret function logarithmically. Through simulation we found the optimum values for the model parameters, $c = 10^{-4}$, $d = 10^{-2}$, $N = 5$, which define ϵ as follows: $\epsilon = \min\{1, \frac{cN}{d^2 n}\}$.

IV. Thompson Sampling Based Algorithm

This section incorporates Thompson sampling into a learning algorithm for channel selection problem. Thompson sampling [3] is best understood in Bayesian context. Assume we observed S_j, the observation vector, after accessing channel j, n_j times. Assuming Bernoulli distribution for each access trial with parameter μ_j, the parametric likelihood function for observation vector S_j is as follows,

$$p_j(S_j|\mu_j) = {\mu_j}^{t_j}(1-\mu_j)^{n-t_j}, \tag{5}$$

where t_j is the number of successful transmissions on j^{th} channel in n trials. Without loss of generality, we use Beta distribution as the prior for the distribution of parameter μ_j. This is because Beta distribution is conjugate prior for the likelihood function in (5) which simplifies the derivations [15]. Using Bayes rule we can write,

$$p_j(\mu_j|S_j) = \frac{p_j(S_j|\mu_j)\frac{\Gamma(\alpha_j+\beta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)}{\mu_j}^{\alpha_j-1}(1-\mu_j)^{\beta_j-1}}{p_j(S_j)}, \tag{6}$$

where,

$$\Gamma(\alpha_j) = \int_0^\infty x^{\alpha_j-1}e^{-x}dx \tag{7}$$

and α_j and β_j are the shape parameters of the Beta distribution; as we assume no prior information on μ_j we initialize $\alpha_j = \beta_j = 1$ which yields uniform distribution in $[0, 1]$. Substituting (5) in (6) yields,

$$p_j(\mu_j|S_j) = \frac{\frac{\Gamma(\alpha_j+\beta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)}}{p_j(S_j)}{\mu_j}^{t_j+\alpha_j-1}(1-\mu_j)^{n_j-t_j+\beta_j-1}. \tag{8}$$

Introducting $C = \frac{\frac{\Gamma(\alpha_j+\beta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)}}{p_j(S_j)}$ can re-write (8) as:

$$p_j(\mu_j|S_j) = C{\mu_j}^{\alpha_j+t_j-1}(1-\mu_j)^{\beta_j+t_j-1} \tag{9}$$

Using $\int p_j(\mu_j|S_j)d\mu_j = 1$ and $\int x^{\alpha_j-1}(1-x)^{\beta_j-1}dx = \frac{\Gamma(\alpha_j)\Gamma(\beta_j)}{\Gamma(\alpha_j+\beta_j)}$, we obtain,

$$p_j(\mu_j|S_j) = \frac{\Gamma(\alpha_j+\beta_j+2t_j)}{\Gamma(\alpha_j+t_j)\Gamma(\beta_j+t_j)}{\mu_j}^{\alpha_j+t_j-1}(1-\mu_j)^{\beta_j+t_j-1}, \tag{10}$$

which is the beta distribution with parameters $\alpha_j + t_j$ and $\beta_j + t_j$,

$$p_j(\mu_j|S_j) = beta(\alpha_j + t_j, \beta_j + t_j). \tag{11}$$

Based on this computing, each node chooses a channel and tries to transmit on it as described in Algorithm 1.

Algorithm 1 Thompson Sampling

Parameters: j: channel index n: total number of channel accesses t_j: number of successful transmissions so far

$\alpha_j = \beta_j = 1$
$t_j = n = 0$
while True **do**
  for all j **do**
    sample $r_j \sim beta(\alpha_j + t_j, \beta_j + t_j)$
  end for
  $m = \arg\max\{\overline{r}_j\}, n++$
  if channel m is free **then**
    TRANSMIT(), $t_m += 1$
  end if
end while

V. Performance Evaluation

We evaluate the performance of Thompson sampling based algorithm and compare it with the existing methods, UCB1, UCB2 and ϵ_n-greedy in a cognitive radio setting. The channel observations are modeled and created as Bernoulli trials with parameter μ (μ is the probability that the channel is available). We created $N = 10k$ samples for 10 channels with μ values randomly distributed in $(0, 1)$. The sampling rate is assumed 1500 samples per second to match the measurements discussed in the following. Algorithms are implemented in MATLAB and numerical analysis are performed to find the optimum values for the parameters of the algorithms. The average throughput of algorithm A ($\bar{T}_A$) is obtained as the ratio of the cumulative number of transmitted bits ($N_b(A)$) to the time passed since the beginning of the transmission t: $\bar{T}_A = \frac{N_b(A)}{t}$

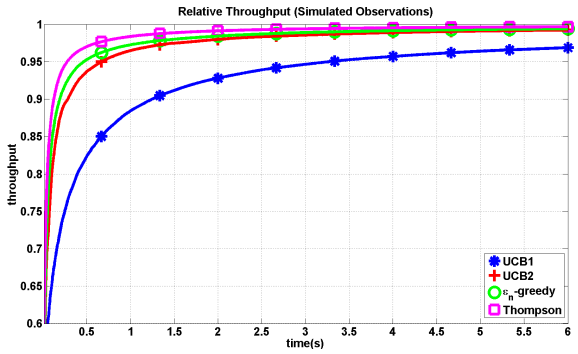


Fig. 1: Relative throughput comparison on 10 accessible channels ($\mu = 0.80$).

The relative throughput of A was obtained as the ratio of the average throughput of A to the average throughput of an oracle agent ($\bar{T}_O$) that always operates on the best channel,

$$\overline{RT}_A = \frac{\bar{T}_A}{\bar{T}_O}. \tag{12}$$

Note that the relative throughput of the oracle algorithm is always one. The simulations are run for 100 iterations and averaged to yield the relative throughput. Fig. 1 shows the relative throughput for different algorithms vs. time. In the exploration phase, the throughput of all algorithms is significantly lower than the oracle agent. However, all algorithms eventually find the best channel and start exploiting it almost all the time and hence reach a throughput close to the oracle. Yet, Thompson sampling based algorithm converges to the best channel in fewer steps than the other policies. In addition, it achieves higher average throughput as it spends less time on exploring the channels and converges to the best channel faster. This results in smaller latency in transmission and lower energy consumption for channel exploration.

To make our scenario more realistic, we used real-time measurements of the channel to compare the performance of different algorithms. We used TelosB node (IEEE 802.15.4) as the secondary user of a Wi-Fi channel which is considered as the primary user channel. As in cognitive radio scenario, we suppose that the TelosB is able to receive the messages from all available channels simultaneously. Although this is not exactly a cognitive radio setting as both standards have the same priority in using the ISM band, it can closely mimic a realistic cognitive radio setting. In addition, our experiment setup can be considered as a case of dynamic spectrum sharing in heterogeneous networks. The IEEE 802.11g Wi-Fi access point (AP) as well as the primary user client equipment, a laptop with a Wi-Fi interface card, were 4.5 m away on the line of sight. The Wi-Fi network used channel 6 of Wi-Fi. The AP was wired (Ethernet) to the PC running the Apache web-server providing the access to a web-page and a big data file (about 1 GB). The measurements were performed by the TelosB node which was situated near the Wi-Fi AP at the distance of 30 cm. The TelosB node was driven by Contiki operating system running the code to read the RSSI of channel 17 of IEEE 802.15.4 (the center frequency 2435 MHz) that overlaps with channel 6 of IEEE 802.11 (the center frequency 2437 MHz)

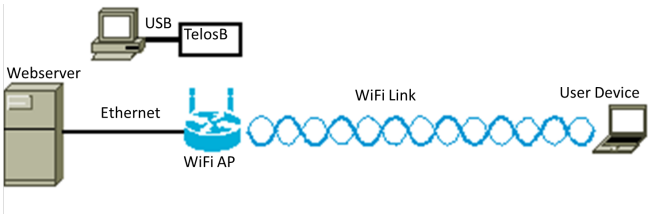
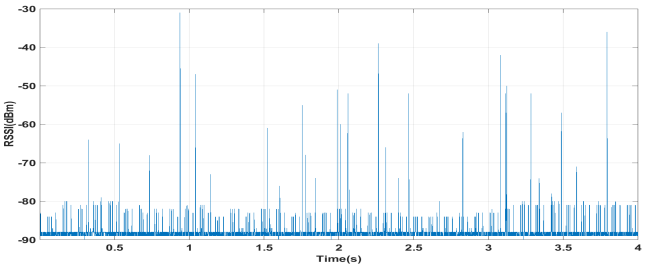
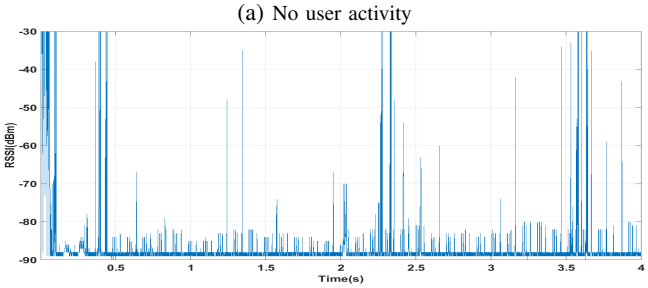


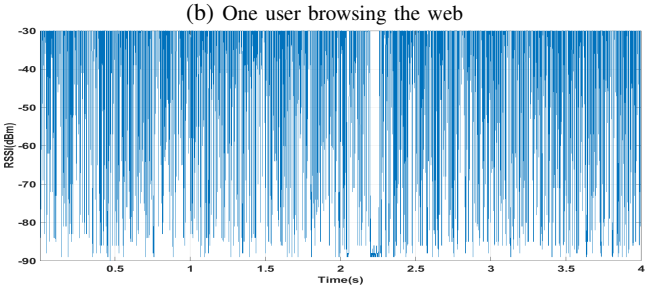
Fig. 2: Schematic of experiment setup.



(a) No user activity



(b) One user browsing the web



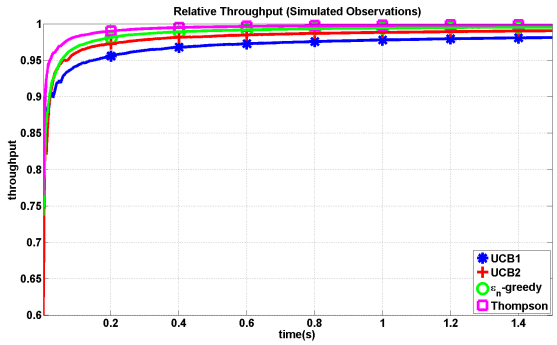
(c) One user downloading a large file

Fig. 3: Spectrum measurements

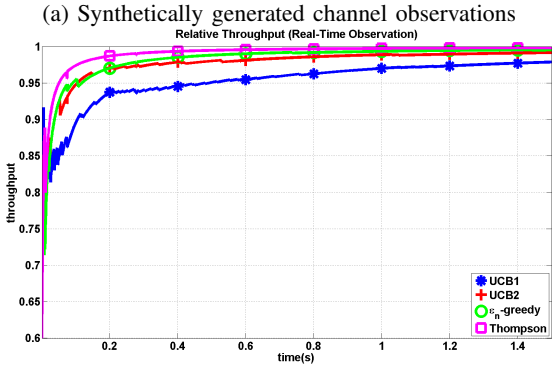
used by the AP. The TelosB node performed measurements at about 1500 samples per second and sends them to another PC via an USB link as depicted in Fig. 2.

Fig. 3 shows the RSSI values received by the secondary user when only one primary user is connected to the network. The average availability rates for the setting depicted in are $\mu = 0.99, \mu = 0.92, \mu = 0.12$ resp. Fig. 3(a) shows the RSSI when the user is idle and most of the traffic comes from the control frames and beacons of the base station. As expected, the channel is available most of the time. In Fig. 3(b), the user is browsing Internet pages frequently while Fig. 3(c) shows the RSSI values when they download a large file (Approx. 1 GB). Average Availability rate of the channel (μ) is obtained by dividing the number of idle samples to the total number of observations where idle channel was obtained by applying a threshold of $-44dBm$ to the RSSI measurements.

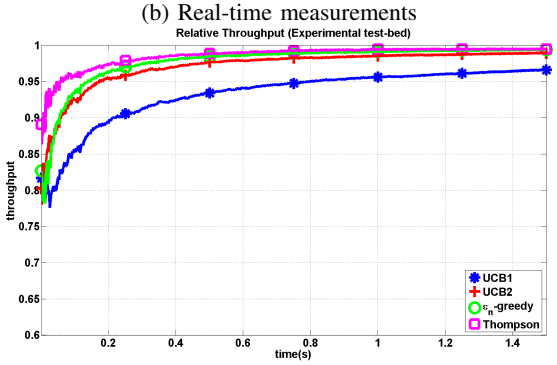
We performed simulation for 3 channels with average availability rate equal to the measured channels above (Fig. 4(a)).



(a) Synthetically generated channel observations



(b) Real-time measurements



(c) Real-time test-bed

Fig. 4: Relative throughput on 3 channels.

In addition, we used the real-time channel observations instead of synthetically generated data to compare the performance of the learning algorithms (Fig. 4(b)). Real-time data results are similar to the ones got through synthetic channel. The real-time measurements show that Thompson sampling algorithm reaches 99% of the oracle throughput after 0.26s (390 samples) approx. 57% faster than the next best algorithm (ϵ_n-greedy) which achieves the same throughput in 0.60s (900 samples).

In the final evaluation step, we implemented the algorithms in TelosB nodes to compare their real-time performances and confirm their relatively low computational complexity through execution on a simple 16-bit microcontroller. Our setup (Fig. 5) features used 3 pairs of laptops occupying 3 orthogonal Wi-Fi (IEEE 802.11g) channels, 1, 6 and 11 overlapping with standard 802.15.4 channels, 12, 17 and 22 resp. The traffic was generated using "Distributed Internet Traffic Generator" [16] in single flow mode with average packet size on the Internet of 500 bytes [17]. Two TelosB

Fig. 5: Experimental setup: 3 pairs of laptops occupy 3 orthogonal channels of Wi-Fi (1, 6, 11)

nodes use Contiki operating system with a learning algorithm while the third node (oracle) is fixed on the best channel. Beta distribution samples used in Thompson sampling algorithm are generated with "GEN_SEQUENCE" library[1].

To monitor the availability rate of the channel we programmed one TelosB node as monitor which just sampled the channel. The availability rate obtained as the average number of samples the channel is detected available. The RSSI sensitivity of TelosB node was set to $-40dbm$. This relatively high threshold was set to suppress the RSSI received from other networks present in the building. With this sensitivity, the monitor node registers approximately 90% availability rate for the channels. The availability rate of channel 6 drops to approximately 40% when the traffic generator is activated at $2000pkt/sec$ and packet size of 500 bytes. The availability rate of the channel 1 drops to approx. 60% when the traffic generator occupies the channel with $500pkt/sec$. Channel 11 is left without traffic although the server and client were connected. The monitor shows approx. 90% availability on the channel. Note that the channel occupancy rate is affected by our traffic, other networks traffic and noise. However, it was roughly constant during the experiment at the given rates.

We programmed two TelosB nodes; one as an oracle which always operated on the channel with the best availability rate. The other node was programmed with the implementation of a learning algorithm to find and use the best channel. In our results, we considered an available channel as a successful transmission. In reality, the packet transmission can be disrupted in the middle of the transmission and cause the transmission to fail. However, the collision will affect the throughput results of all algorithms including the oracle the same way. Hence the comparison results would not be affected.

In each set of experiments, we performed 3 experiments where occupancy rate of the channels were inverted. The relative throughput of each algorithm in each experiment is divided by the oracle performance of the best channel and then averaged over all the experiments for each algorithm. The results are shown in Fig. 4(c). As seen in the figure, similar results are obtained in empirical evaluation of the algorithms where Thompson based method achieves the best performance

[1]https://compbio.soe.ucsc.edu/gen_sequence

followed by ϵ_n-greedy and UCB2 algorithms which performed similarly in this context.

VI. Multi-hop extension

A. Multihop approach

Main challenges that appear in multihop is to have sender and receiver nodes use the same channels at the same time. Our approach proposes that a sender node applies our Thomson-based scheme (Algo.1) to select the channel to use to broadcast its messages. To increase its chances, it will not only select the best channel but the N best channels. Note that the higher N value, the more chance for a sender node to use the channel selected by the receiver node but the more energy consumed.

Receiving nodes apply our scheme to select the best channel to listen based on the same approach. Once a message is received, it is processed and if it needs to be retransmitted, the receiver node switches to the sending mode.

B. Experimental proof of concept

1) Set up: We implemented our multihop approach in the framework of the EWSN competition [18]. The simulation set up was as follows. Experimentation was run on TelosB / Tmote Sky nodes with either a SMA or PCB antenna. A source node sends a message every time the luminosity sensed changes. The network was composed of 15 wireless nodes distributed in a $150m^2$ building with a large proportion of concrete and metal. Some additional nodes have been randomly deployed to disturb the network as illustrated on Fig.6. Source and destination nodes were between 3 and 5 hops away from each others. Each node forwards over $N = 3$ different channels.

2) Results and discussions: We run the above algorithm for 5min within the tough conditions of the competition. At the end of the experimentation, the sink has received more than 77% of packets sent by the source with paths composed of between 4 and 6 hops, depending of the channels chosen by each node. We have witnessed that forwarding nodes dynamically learn in an efficient way the best channel to listen since when the environment becomes to disturbed due to interference nodes, they change channel after 2μs. This reliability could be enhanced by increasing N but to the detriment of the energy cost. This first experimentation was useful to settle the feasibility and applicability of our approach even on constrained nodes. A deeper investigation regarding the energy consumption and latency and trade-off between them will allow a better qualification of our proposed technique. In the future, we intend to exploit these preliminary feedbacks to improve the design of our multihop approach to allow a better connectivity and thus improve the reliability and latency by keeping a low energy consumption.

VII. Conclusion

This work addresses the channel exploration-exploitation dilemma in a cognitive radio context where secondary network has no prior information about the primary user channel utilization statistics. The problem is formulated as a multi-arm bandit problem and addressed in a Thompson sampling framework. The performances of the most efficient mathematical methods in multi-arm bandit formulation are compared both numerically and by implementation. Results show that Thompson sampling formulation converges to the best channel in fewer steps than the other policies in a cognitive radio setting. In addition, we provide first steps and pave the way for further investigations to extend such cognitive approaches to multihop communications.

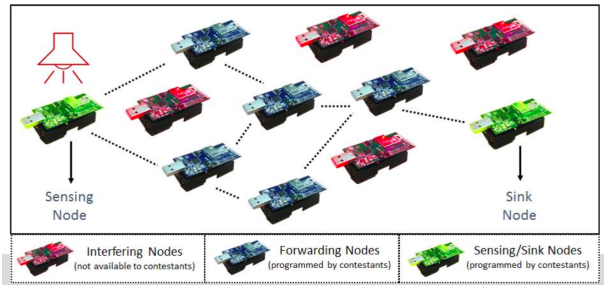


Fig. 6: Deployment scenario

References

[1] J. Lunden, V. Koivunen, and H. V. Poor, "Spectrum exploration and exploitation for cognitive radio: Recent advances," *Signal Processing Magazine, IEEE*, vol. 32, no. 3, pp. 123–140, 2015.

[2] S. 802.22, "IEEE standard for wireless regional area networks," 2014.

[3] O. Chapelle and L. Li, "An empirical evaluation of thompson sampling," in *Advances in neural information processing systems*, 2011.

[4] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive mac for opportunistic spectrum access in ad hoc networks: A pomdp framework," *JSAC*, vol. 25, no. 3, 2007.

[5] S. Geirhofer, L. Tong, and B. M. Sadler, "A measurement-based model for dynamic spectrum access in wlan channels," in *Military Com. Conf. (MILCOM)*, 2006.

[6] ——, "Cognitive medium access: constraining interference based on experimental models," *JSACn*, vol. 26, no. 1, 2008.

[7] M. López-Benıtez and F. Casadevall, "Time-dimension models of spectrum usage for the analysis, design and simulation of cognitive radio networks," *IEEE Trans. Veh. Technol*, vol. 62, no. 5, 2013.

[8] W. Jouini, D. Ernst, C. Moy, and J. Palicot, "Multi-armed bandit based policies for cognitive radio's decision making issues," in *3rd international conference on Signals, Circuits and Systems (SCS)*, 2009.

[9] Y. Gai, B. Krishnamachari, and R. Jain, "Learning multiuser channel allocations in cognitive radio networks: A combinatorial multi-armed bandit formulation," in *DySPAN*, 2010.

[10] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2, 2002.

[11] Y. Gwon, S. Dastangoo, and H. Kung, "Optimizing media access strategy for competing cognitive radio networks," in *Global Com. Conf. (GLOBECOM)*, 2013.

[12] K. Wang, L. Chen, Q. Liu, W. Wang, and F. Li, "One step beyond myopic probing policy: A heuristic lookahead policy for multi-channel opportunistic access," *Wireless Com., IEEE Trans.*, vol. 14, no. 2, 2015.

[13] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in applied mathematics*, vol. 6, no. 1, pp. 4–22, 1985.

[14] L. Lai, H. El Gamal, H. Jiang, and H. V. Poor, "Cognitive medium access: Exploration, exploitation, and competition," *Mobile Computing, IEEE Transactions on*, vol. 10, no. 2, pp. 239–253, 2011.

[15] P. Diaconis, D. Ylvisaker *et al.*, "Conjugate priors for exponential families," *The Annals of statistics*, vol. 7, no. 2, pp. 269–281, 1979.

[16] A. Botta, A. Dainotti, and A. Pescapè, "A tool for the generation of realistic network workload for emerging networking scenarios," *Computer Networks*, vol. 56, no. 15, pp. 3531–3547, 2012.

[17] K. Thompson, G. J. Miller, and R. Wilder, "Wide-area internet traffic patterns and characteristics," *Network, iEEE*, vol. 11, no. 6, 1997.

[18] A. Maskooki, V. Toldov, L. Clavier, V. Loscrí, and N. Mitton, "Competition: Channel Exploration/Exploitation Based on a Thompson Sampling Approach in a Radio Cognitive Environment," in *Int. C. on Embedded Wireless Systems and Networks (EWSN)*, Graz , Austria, 2016.