Learning Algorithm for Reconfigurable Antenna State Selection

Nikhil Gulati, David Gonzalez, and Kapil R. Dandekar

Department of Electrical and Computer Engineering, Drexel University,
Philadelphia, PA 19104
Email: {ng54, dg345, dandekar}@drexel.edu

***Abstract*— In this paper, we propose an online learning algorithm for selecting the state of a reconfigurable antenna. We formulate the antenna state selection as a multi-armed bandit problem and present a selection technique, implemented for a 2×2 MIMO OFDM system employing highly directional metamaterial Reconfigurable Leaky Wave Antennas. We quantify the performance of our selection technique using a software defined radio testbed and present results for a wireless network in a typical indoor environment.**

***Index Terms*— Learning algorithms, bandit problem, reconfigurable antennas, MIMO, OFDM.**

I. Introduction

In recent years, studies have shown that reconfigurable antennas can offer additional performance gains in Multiple Input Multiple Output (MIMO) systems [1], [2], [3], [4], [5], [6]. These reconfigurable antennas are capable of generating multiple uncorrelated channel realizations by changing their electrical and radiation properties and are gradually making their way into commercial wireless systems [7]. The key to effectively utilizing the reconfigurability offered by these antennas is to select a state which provides the highest signal to noise ratio (referred to as optimal state in rest of the paper) among all the available states for a given wireless environment.

Reconfigurable antennas can be employed either at the transmitter or receiver, or at both the ends of the RF chain. This flexibility can create a large search space in order to find an optimal state for communication. Moreover, the effect of node mobility to a different location, changes in physical antenna orientation, and the dynamic nature of the wireless channel can render previously found "optimal" states suboptimal over time. This makes it essential for a wireless system to employ a learning algorithm to find the new optimal states and maintain the highest possible SNR.

In order to be effective, an online learning algorithm for antenna state selection (also referred to interchangeably as selection technique) must overcome certain challenges. We identify such challenges below:

1) Optimal antenna state for each wireless link (between a single transmitter and a receiver location) is unknown *apriori*. Moreover, each wireless link may have a different optimal state. A selection technique should be able to learn and find the optimal state for a given link.
2) For a given wireless link, there might be several states which are near optimal over time based on channel conditions and multipath propagation. A selection technique should provide a policy to balance between exploiting a known successful state and exploring other available states to account for dynamic behavior of the channel.
3) For the purpose of real-time implementation in a practical wireless system, a selection technique must employ simple metrics which can be extracted from the channel without large overhead or extensive feedback data.

Previous work related to state selection is based on estimating channel response of each antenna state which required changing the standard OFDM frame format [1]. Selection techniques using second order channel statistics and average SNR information have also been proposed [8]. Though some of these techniques were successful in showing the benefits of multi-state selection and motivated the need for a selection algorithm, none solved the challenges mentioned above. Previous work in learning for cognitive radios has primarily been focused on link adaptation [9], [10] and channel allocation for dynamic spectrum access [11]. In this paper, we make a case for using learning algorithms for antenna state selection and investigate the feasibility of implementing such algorithms in a practical wireless system.

We propose a solution to the challenges mentioned above by formulating the antenna state selection as a multi-armed bandit problem. Multi-armed bandit problem [12], [13], [14] is a fundamental mathematical framework for learning unknown variables. In its classic form, there are N independent arms with a single player, playing arm $i\,(i = 1, \ldots N)$. Each play of a single arm yields random rewards which are $i.i.d$ with a distribution of unknown mean. The goal is to design a policy to play one arm at each time sequentially to maximize the total expected reward in the long run. Lai and Robbins [12] studied the non-Bayesian formulation and provided a performance measure of an arm selection policy referred to as *regret or cost of learning*. Regret is defined as the difference in the expected reward gained by always selecting the optimal choice and the reward obtained by a given policy. Since

the best arm cannot always be identified in most cases using a finite number of prior observations, the player will always have to keep learning. Due to the continuous learning process, the player will make mistakes which will grow the regret over time. It has been shown in [12] that the minimum rate at which regret grows is of logarithmic order under certain regularity conditions.

II. Problem Formulation and Algorithm

Our work is influenced by the work done in [14] where arms have non-negative rewards that are $i.i.d$ over time with an arbitrary un-parametrized distribution. We consider the set up where there is a single transmitter and M wireless receiver nodes and both the transmitter and the receivers employ the reconfigurable antennas. The transmitter has a fixed antenna state and the receivers can select from N available antenna states. This reduces the problem to selecting an antenna state only at the receiver end where each receiver can select state i independently. The decision is made at every packet reception n to select the state to be used for the next reception. If a receiver node selects a state i and assuming the transmission is successful, a random reward is achieved which we denote as $R_i(n)$. Without loss of generality, we normalize $R_i(n) \in [0,1]$. When a receiver selects a state i, the value of $R_i(n)$ is only observed by that receiver and the decision is made only based on locally observed history.

We base our selection technique on the deterministic policy UCB1 given in [14]. To implement this policy, each receiver stores the average of all the reward values observed for state i up to the current packet n denoted as $\bar{R}_i(n)$ and the number of times state i has been played, $n_i(n)$. The UCB1 policy is shown below as Algorithm I.

Algorithm 1 UCB1 Policy, Auer [14]

// Initialization
$n_i, \bar{R}_i \leftarrow 0$
Play each arm at least once and update $n_i, \bar{R}_i$ accordingly.
// Main Loop
while 1 **do**
  Play arm i that maximizes $\bar{R}_i + \sqrt{\frac{2ln(n)}{n_i}}$
  Update n_i, $\bar{R}_i$ for arm i
end while

We also implemented the ϵ-GREEDY policy which is a randomized policy and the UCB1-Tuned policy [14] which has been shown to work better for practical purposes. In the ϵ-GREEDY policy, the arm with current highest average is selected with probability $1-\epsilon$ and a random arm is selected with probability ϵ. UCB1-Tuned is a fine tuned version of UCB1 policy which accounts for the variance measured independently across arms. In this policy, the upper confidence bound of UCB1 policy is replaced by

$$\sqrt{\frac{ln(n)}{n_i}\min\left\{\frac{1}{4}, V_i(n_i)\right\}} \tag{1}$$

where V_i is defined as

$$V_i(s) \equiv \left(\frac{1}{s}\sum R_{i,s}^2\right) - \bar{R}_{i,s}^2 + \sqrt{\frac{2ln(t)}{s}} \tag{2}$$

when arm i has been played s times during the first t plays.

III. Reconfigurable Leaky Wave Antennas

The Reconfigurable Leaky Wave Antenna (RLWA) is a two port antenna array designed to electronically steer two highly directional independent beams over a wide angular range. Initially proposed by the authors in [15], the prototype shown in Fig. 1 is a composite right/left-handed leaky wave antenna composed of 25 cascaded metamaterial unit cells [3]. Moreover, the application of various combinations of bias voltages "S" and "SH" controls the beam direction allowing for symmetrical steering of the two radiation beams at the two ports over a 140° range.

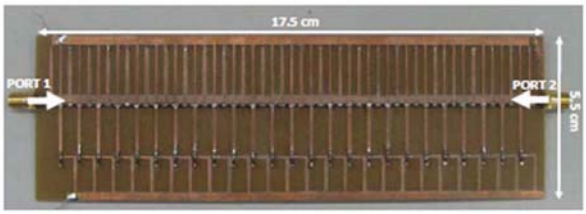


Fig. 1. Two port reconfigurable leaky wave antenna [15]

In order to characterize the effect of beam direction on the efficacy of a wireless system with RLWAs deployed at both ends of a link, a subset of states was selected to allow the beam to steer over a range of 140° in the elevation plane. Fig. 2 shows the measured radiation patterns for the selected states and their corresponding bias voltages.

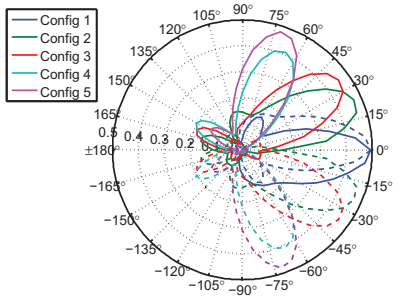
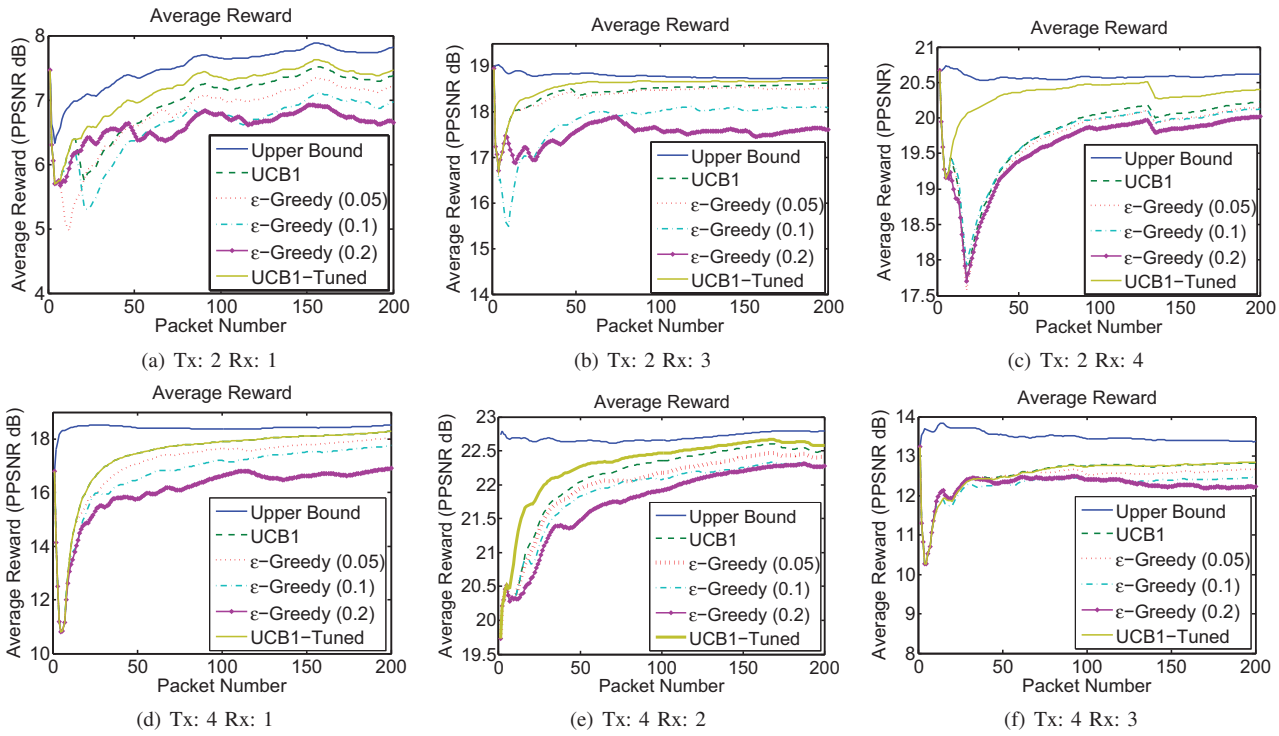


Fig. 2. Measured radiation patterns for port 1 (Gain $\approx -3dB$)



(a) Tx: 2 Rx: 1 (b) Tx: 2 Rx: 3 (c) Tx: 2 Rx: 4

(d) Tx: 4 Rx: 1 (e) Tx: 4 Rx: 2 (f) Tx: 4 Rx: 3

Fig. 3. Average reward for each algorithm for the designated links

IV. Experimental Setup and Results

In our experiments we make use of the Wireless Open Access Research Platform (WARP), an FPGA-based software defined radio testbed and WARPLab, the software development environment used to control WARP nodes from MATLAB [16]. Four WARP nodes were distributed throughout the fifth floor of the Drexel University Bossone Research Center as shown in Fig. 4. By using WARPLab, each of the nodes were centrally controlled for the synchronization of the transmission and reception process and to provide control over the antenna state selected at each of the nodes. Although, the nodes were controlled centrally for data collection purposes, the learning algorithm was decentralized. Specifically, no information during the learning process was shared with the transmitter.

The performance of the RLWA was evaluated in a 2×2 MIMO system with spatial multiplexing as the transmission technique [15]. For baseline measurements, each designated WARP node transmitter broadcasted packets modulated using BPSK. For each packet transmission, the receiver nodes stored channel estimates and measured the post-processing signal-to-noise ratio (PPSNR) by evaluating the error vector magnitude of the received symbol constellations. Furthermore, the antenna states for each receiver node were switched after each packet until all 5 possible antenna states between the transmitter and receivers were tested. This process was repeated until 200 realizations were achieved for all state combinations and

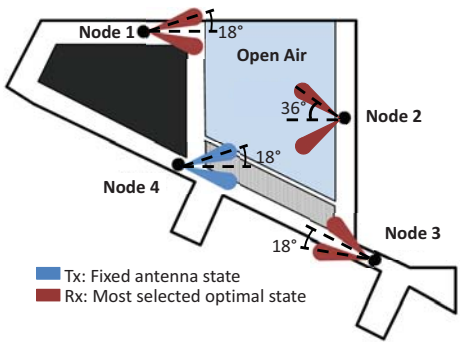


Fig. 4. Node positions on the 5th floor of the Drexel University Bossone Research Center.

for each node acting as a transmitter. The beam directions in Fig. 4 corresponds to the optimal state selected most often at each of the receivers when node 4 was transmitting. The algorithm described in Section II is an online algorithm, but note that we collected the channel realizations corresponding to each state and evaluated the algorithm in post-processing. This is essential in order to benchmark the performance of different policies under the same channel conditions and make sure that channel conditions do not bias the performance results.

We present the results for three multi-armed bandit policies (UCB1, UCB1-Tuned, ϵ-GREEDY) in Fig. III verifying the empirical performance of our selection technique.

Each sub-figure represents the average reward achieved by all three policies for a given wireless link over 200 packets. We define the upper bound as the reward obtained by a genie which always selected the optimal state with perfect channel knowledge of all antenna states at each trial. For most of the links (3(a)-3(c), 3(e)), we found UCB1-Tuned outperformed the other policies. UCB1-Tuned has been found to work better for practical purposes [14] since it is not sensitive to the variance of the states. Also, ϵ-GREEDY did not perform well because ϵ-GREEDY explores uniformly over all states and can select sub-optimal states more often, thereby reducing the average reward. It is evident from the figure that among three instances of ϵ-GREEDY policy, the instance with highest ϵ performed the worst. However, were we to consider mobile users in our experiment, it is possible that ϵ-GREEDY policy will adapt better to substantial variations in channel condition.

Also, we show in Table I the percentage of time the optimal state was selected by each policy. For four links ((3(b)-3(e)), UCB1-Tuned selected the optimal state more than 95% of the time. For the links where all the policies had lower success rate, we attribute that to the fact that, even though some states had higher instantaneous rewards, those states did not consistently generate highest rewards and were not selected.

TABLE I
PERCENTAGE OF TIME EACH POLICY SUCCESSFULLY SELECTED THE OPTIMAL STATE

Tx	Rx	UCB1	$\epsilon(0.05)$	$\epsilon(0.1)$	$\epsilon(0.2)$	UCB1-Tuned
2	1	82	79	78	69	83
2	3	95.5	95	90	79.5	97.5
2	4	94.5	87.5	87.5	79.5	95.5
4	1	98	93.5	87	86.5	98
4	2	88.5	89.5	83.5	81	95
4	3	67	64	63	60	67.5

V. CONCLUSION AND FUTURE WORK

We have proposed a learning algorithm for antenna state selection and have shown that wireless systems employing reconfigurable antennas can benefit from such technique. We have shown empirically that the multi-armed bandit problem is a useful online learning framework for antenna state selection in a practical wireless system. For a network of four nodes employing reconfigurable antennas equipped with five states, the learning algorithm improves the received PPSNR and thereby improving the achievable throughput of the system. In the future work, we would like to consider a system in which multiple antenna states can also be selected at the transmitter and evaluate the feedback requirements and overhead. In addition, we would like to evaluate the algorithm's performance in more diverse indoor as well as outdoor environments. Another area of future work can involve performance evaluation of the learning algorithm with mobile nodes.

ACKNOWLEDGMENT

The authors wish to acknowledge Dr. Bhaskar Krishnamachari, Kevin Wanuga, Prathaban Mookiah and Alex Lackpour for their valuable suggestions and feedback. This material is based upon work supported by the National Science Foundation under Grant No. 0916480.

REFERENCES

[1] A. Grau, H. Jafarkhani, and F. De Flaviis, "A reconfigurable multiple-input multiple-output communication system," *Wireless Communications, IEEE Transactions on*, vol. 7, no. 5, pp. 1719–1733, 2008.

[2] H. Pan, G. Huff, T. Roach, Y. Palaskas, S. Pellerano, P. Seddighrad, V. Nair, D. Choudhury, B. Bangerter, and J. Bernhard, "Increasing channel capacity on MIMO system employing adaptive pattern/polarization reconfigurable antenna," in *Antennas and Propagation Society International Symposium, 2007 IEEE*. IEEE, 2007, pp. 481–484.

[3] D. Piazza, M. D'Amico, and K. Dandekar, "Performance improvement of a wideband MIMO system by using two-port RLWA," *Antennas and Wireless Propagation Letters, IEEE*, vol. 8, pp. 830–834, 2009.

[4] D. Piazza, N. Kirsch, A. Forenza, R. Heath, and K. Dandekar, "Design and evaluation of a reconfigurable antenna array for MIMO systems," *Antennas and Propagation, IEEE Transactions on*, vol. 56, no. 3, pp. 869–881, 2008.

[5] A. Forenza and R. Heath Jr, "Benefit of pattern diversity via two-element array of circular patch antennas in indoor clustered MIMO channels," *Communications, IEEE Transactions on*, vol. 54, no. 5, pp. 943–954, 2006.

[6] J. Boerman and J. Bernhard, "Performance study of pattern reconfigurable antennas in MIMO communication systems," *Antennas and Propagation, IEEE Transactions on*, vol. 56, no. 1, pp. 231–236, 2008.

[7] Y. Jung, "Dual-band reconfigurable antenna for base-station applications," *Electronics Letters*, vol. 46, no. 3, pp. 195 –196, 4 2010.

[8] D. Piazza, J. Kountouriotis, M. D'Amico, and K. Dandekar, "A technique for antenna configuration selection for reconfigurable circular patch arrays," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 3, pp. 1456–1467, 2009.

[9] R. Daniels, C. Caramanis, and R. Heath, "Adaptation in convolutionally coded MIMO-OFDM wireless systems through supervised learning and SNR ordering," *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 1, pp. 114–126, 2010.

[10] H. Volos and R. Buehrer, "Cognitive engine design for link adaptation: an application to multi-antenna systems," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 9, pp. 2902–2913, 2010.

[11] Y. Gai, B. Krishnamachari, and R. Jain, "Learning multiuser channel allocations in cognitive radio networks: a combinatorial multi-armed bandit formulation," in *New Frontiers in Dynamic Spectrum, 2010 IEEE Symposium on*. IEEE, 2010, pp. 1–9.

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

[13] V. Anantharam, P. Varaiya, and J. Walrand, "Asymptotically efficient allocation rules for the multiarmed bandit problem with multiple plays-part I: I.I.D rewards," *Automatic Control, IEEE Transactions on*, vol. 32, no. 11, pp. 968–976, 1987.

[14] 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.

[15] D. Piazza, D. Michele, and K. Dandekar, "Two port reconfigurable CRLH leaky wave antenna with improved impedance matching and beam tuning," in *Antennas and Propagation, 2009. EuCAP 2009. 3rd European Conference on*. IEEE, 2009, pp. 2046–2049.

[16] Rice University WARP project. [Online]. Available: http://warp.rice.edu