Cross-Entropy Optimized Cognitive Radio Policies

Boris Oklander, Moshe Sidi

Cross-Entropy Optimized Cognitive Radio Policies[1]

Boris Oklander and Moshe Sidi

Department of Electrical Engineering, Technion – Israel Institute of Technology,
32000 Haifa, Israel
oklander@tx.technion.ac.il, moshe@ee.technion.ac.il

Abstract. In this paper we consider cognitive processes and their impact on the performance of cognitive radio networks (CRN). We model the cognition cycle, during which cognitive radio (CR) sequentially senses and estimates the environment state, makes decisions in order to optimize certain objectives and then acts. Model-based analysis of CRN is used to solve control and decision making tasks, which actually gives the radio its "cognitive" ability. Particularly, we design an efficient strategy for accessing the vacant spectrum bands and managing the transmission-sampling trade-off. In order to cope with the high complexity of this problem the policy search uses the stochastic optimization method of cross-entropy. The developed model represents CRN ability to intelligently react to the network's state changes and gives a good understanding of the cross-entropy optimized policies.

Keywords: cognitive radio networks; dynamic spectrum access; state estimation; queueing analysis; cross-entropy

1 Introduction

CRN are expected to cope with a wide spectrum of challenges arising in the face of the growing demand for wireless access in voice, video, multi-media and other high rate data applications. Although researchers and standardization bodies agree that CR should sense the environment and autonomously adapt to changing operating conditions, there are different views concerning the levels of cognitive functionality [1]. This functionality of CRN can be represented by the cognition cycle [2].

Cognition cycle is the main control process that enables CR to stay aware of its communication environment and to adapt to its changing conditions. There are different views of what phases the cognition cycle consists [2],[3], but basically all the versions share the observation, orientation, decision and action (OODA) phases. During the observation phase, CR continuously senses the environment in order to collect the input information for the cognition cycle. During the orientation phase, CR uses the gathered information to estimate the current network state. Next, CR enters the decision-making phase, in which it applies some policy to decide on the course of action. Finally, CR completes the cognition cycle by carrying out the chosen actions.

[1] This research has been partially supported by the CorNet consortium funded by the chief scientist in the Israeli Ministry of Industry, Trade and Labor.

It is not common to find studies that directly address the interdependent processes composing the cognition cycle. The main reason for this is the difficulty to design analytically tractable models for systems characterized by cognitive behavior. Different studies have addressed CRNs capability of opportunistic spectrum access [3],[4], in which spectrum bands licensed to primary users (PU) are shared with the cognitive users called secondary users (SU). It is well known that a significant part of the allocated spectrum is vastly underutilized and the CRN goal in this scheme is to improve spectrum utilization while avoiding interference with the PUs [5]. In [6] state of the art protocols for medium access in cognitive radio networks are overviewed. The authors point out that the existing works do not fully integrate both the spectrum sensing and access in one framework which is required in order to maintain the capability of adaptation to the environment changes [7]. The authors of [8] derive a queueing framework to study the performance of CRN. Although this model allows an analytic study of CRN performance, it lacks the modeling of the cognition cycle. A basic version of cognition cycle model is given in our previous work [9].

This paper presents three significant contributions to the problem of modeling the CRN. Firstly, we enhance the model of [9] by introducing CRN with penalty for interfering PU. The penalty provides an incentive for CRN to enhance its perception level in order to avoid interference with PU, which is an essential requirement in any realistic scenario. Next, we introduce a decision-making process, which is responsible both for selecting the channels to be accessed and for managing the sampling-transmission tradeoff [10]-[12]. The third contribution of this paper is the introduction of cross-entropy optimized policy for controlling the CRN. The task of policy optimization is rather hard due to the high complexity of the model. To overcome this problem we use the method of stochastic optimization of cross-entropy, which is an efficient tool at hand for the task of policy optimization [13]-[15]. The resulting policies reflect the intelligent behavior induced by the described above cognition cycle.

In sections 2, we present our model of cognition cycle. Then in section 3, we use cross-entropy method to optimize the control policy and we evaluate its performance. Section 4 summarizes.

2 Cognition Cycle Model

We start here with modeling the environment's dynamics, the cognition cycle and the CRN transmission process. Then, we close the loop by unifying these models under the entire system framework. The resulting system model makes it possible to analyze the cognition cycle and to evaluate the performance of the CRN.

2.1 Environment Model

We consider a general scenario of wireless communication system which consists of M channels. Every channel alternates between transmitting and idle states. The ON (OFF) period of a channel corresponds to the time interval T_{ON} (T_{OFF}) during which PU transmits (is idle). We assume that T_{ON} and T_{OFF} intervals are exponentially distributed with parameters α and β, respectively. Since the channels are statistically independent, the number of channels available for SU S_t ($S_t \in \{0,1,\ldots,M\}$) at time t is a birth-death process with birth-rate $(M-m)\alpha$ and death-rate $m\beta$ when $S_t=m$, $m \in \{0,1,\ldots,M\}$.

2.2 Perception Model

Next, we model the perception process, which aggregates the observation and the orientation phases of OODA. CRN generates the estimation $\hat{S}_t$ of the environment state S_t through sensing. We assume that the time it takes to update the estimation $\hat{S}_t$ is exponentially distributed. CRN adaptively tunes the sampling rate δ according to its current estimate $\hat{S}_t$, we denote this by $\delta_{\hat{S}}$. For example, CRN could increase the sample rate $\delta_{\hat{S}}$ in order to keep track of rapidly changing network states characterized by high throughput potential, while decreasing it for slowly changing states. The compound process $\{S_t,\hat{S}_t\}$ describes the mutual evolvement of both the environment and the estimation and is actually a continuous time Markov chain (CTMC) (see Fig. 1).

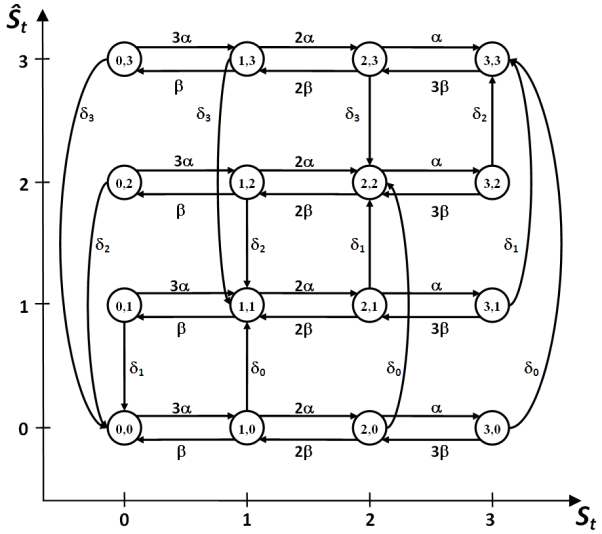

Fig. 1. CTMC of the $\{S_t,\hat{S}_t\}$ process for M=3. The horizontal transitions describe the changes of environment state S_t. The vertical transitions describe the updates of the estimator $\hat{S}_t$.

2.3 Decision Making

The decision-making phase of the cognition cycle employs some policy P for both transmission-sampling tradeoff management and for channels allocation. The transmission rate of SU over a single channel is μ [bit/sec]. We introduce the tradeoff parameter θ ($0\leq\theta\leq1$) which divides the available bandwidth between the transmissions and sampling, where the portion θ of the channel is assigned for transmission and the remaining part $(1-\theta)$ is assigned for sampling. The effective transmission rate over a single channel is therefore $\theta\mu$ [bit/sec] and the resulting update rate of the estimations is $(1-\theta)\mu B$ [1/sec]. The constant $1/B$ [bit] is the number of bits required for updating the estimation $\hat{S}_t$ and it is subject to the physical layer issues. The other responsibility of the policy P is the channel allocation C_t, which is the number of channels over which CR tries to transmit at time t.

We consider state-dependent policies, meaning that the decisions are made based on the estimation of the network state $\hat{S}_t$ and the internal buffer state X_t. The internal buffer state X_t is the number of SU packets waiting for transmission at time t. For the sake of simplicity, in the following modeling we assume that CRN makes decisions based on a greedy policy P_G:

$$C_t=P_G(\hat{S}_t,X_t)=\begin{cases}\hat{S}_t & X_t>0\\ 0 & X_t=0\end{cases} \tag{1}$$

The greedy policy aims to increase the throughput by scheduling transmissions over all the channels that are estimated as unoccupied by PU. This assumption of greedy policy is removed later in section 3 when we optimize the control policies.

2.4 Transmission Process

The arrivals generated by SU are modeled as a Poisson process with rate λ [bit/sec] and service time exponentially distributed with rate μ_t [bit/sec], which changes with time, dependent on a few factors. These factors include the number of accessed channels C_t, the proportion of the bandwidth allocated for transmission θ, the environment state S_t and the penalty for interfering with PU. The combination of these factors results in:

$$\mu_t=\begin{cases}\theta C_t\mu & C_t\le S_t\\ 0 & C_t>S_t\end{cases} \tag{2}$$

It can be seen from (1) that when for the greedy policy P_G, we may substitute $\hat{S}_t$ for C_t since transmissions occur only for $X_t>0$. From (2), our model introduces penalty for CRN when it accesses channels that are in use of PU ($C_t>S_t$). This means that when CRN tries to access channels erroneously estimated as vacant, the transmissions fail. This type of service models CRN, giving the highest priority to PU transmissions.

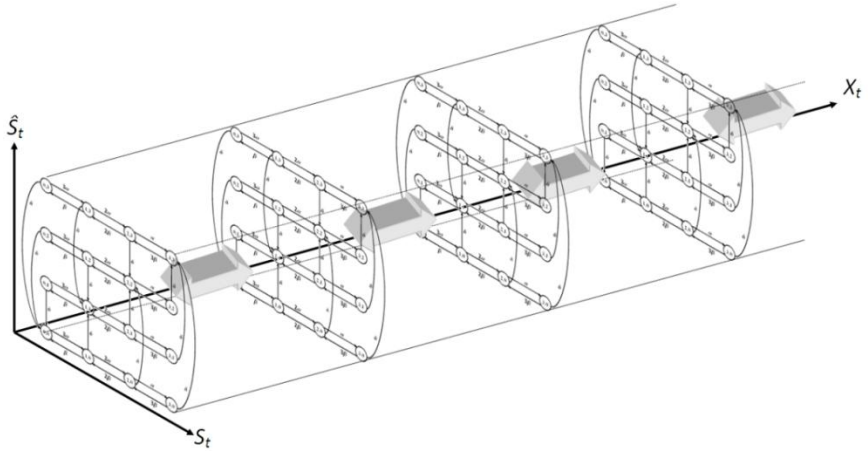

Fig. 2. Illustration of the CTMC of the CRN model. The transitions in the $(S_t,\hat{S}_t)$ plane are identical to those in Fig. 1. The transitions between the levels of the process (along the X_t axis) are omitted here in sake of keeping the clarity.

2.5 System Process

Finally, we aggregate the environment dynamics, the cognition cycle and the transmission process into a unified system model. We define $\{X_t,S_t,\hat{S}_t\}$ to be the process of the entire system for which at time t there are X_t ($X_t\in\{0,1,2,\ldots\}$) queued packets of the SU. This process forms a three dimensional CTMC illustrated in Fig. 2, which is homogeneous, irreducible and stationary. The exact structure of transitions within the CTMC is constructed in the same manner as in [9],[13]. Next, matrix-geometric approach is applied to calculate the steady state probabilities of CTMC. Then, the average number of queued packets of SU can be calculated and by using the Little's law one can obtain the waiting time W of the SU.

3 CRN Policy Optimization

In previous sections we modeled the cognition cycle, in which the decisions were made based on some arbitrarily chosen greedy policy P_G. In this section, we aim to improve the performance of the cognition cycle and the CRN by optimizing the decision making process, i.e., by optimizing the policy.

3.1 Problem formulation

In our framework, a policy P governs the decision-making phase of the cognition cycle. This policy is responsible for managing the sampling-transmission tradeoff by tuning the parameter θ_t, and for allocation of the number of channels, C_t. The values of C_t and θ_t are determined dependently on the network's state estimation $\hat{S}_t$, current CRN buffer state X_t, entire system model and its parameters, which we denote by Ω:

$$(C_t,\theta_t)=P(\hat{S}_t,X_t;\Omega) \tag{3}$$

We aim to optimize CRN performance by minimizing the average waiting time W of SU. In the previous section, we calculated W by applying the matrix geometric analysis to the 3-D CTMC and the Little's law. The 3-D CTMC structure embeds the policy P as follows: the levels transitions (X_t) are affected by the service rate μ_t (eq. 2) and the state transitions $\{S_t,\hat{S}_t\}$within the level are affected by the estimation update rate δ_t given by $\delta_t=(1-\theta_t)\mu B$. Therefore, given the system structure and its parameters Ω, we regard the average waiting time W of SU as a function of the policy P, $W=W(P;\Omega)$. The resulting optimization problem is given by:

$$P^*=\arg\min_{P\in\Pi} W(P;\Omega) \tag{4}$$

where Π is the set of all the feasible policies, i.e., policies which for valid inputs $\hat{S}_t\in\{0,\ldots,M\}$ and $X_t\in\{0,1,2,\ldots\}$ decide on valid values for $\theta\in[0,1]$ and $C_t\in\{0,1,2,\ldots,M\}$. Our optimization problem (4) is complicated. First, it can be shown that the problem is not convex, and the gradient-based techniques are not applicable since it is difficult to obtain a gradient for W. Next, the set Π consists of policies comprising both continuous (θ) and discrete (C_t) action spaces, which requires special approach for optimization. Additionally, the problem exhibits a high computational complexity, due to the rapidly growing (with M) set Π.

We solve this problem by applying the cross-entropy (CE) method of stochastic optimization. CE method is a state-of-the-art method for solving combinatorial and multi-extremal optimization problems. In the following subsection, we review briefly the CE method and demonstrate its application for our optimization problem. The readers interested in further details are referred to [13].

3.2 Cross-Entropy based Stochastic Optimization

The main idea behind the CE method is to define for the original optimization problem an associated stochastic problem (ASP) and then to solve efficiently the ASP by an adaptive scheme. The described below procedure sequentially generates random solutions which converge stochastically to the optimal or near-optimal one.

We define a stochastic policy $P((C_t,\theta_t)|\sigma(\hat{S}_t,X_t))$ as the ASP for (4). $P((C_t,\theta_t)|\sigma(\hat{S}_t,X_t))$ is the probability of choosing action (C_t,θ_t) when CRN's state is $(\hat{S}_t,X_t)$ according to the parameter $\sigma(\hat{S}_t,X_t)$. In the following we use shorthand notation of σ for $\sigma(\hat{S}_t,X_t)$. For the defined ASP, the CE method iteratively draws sample policies $P^{(k)}$ (k=1,2,…,K) from the defined above probability and calculates the average waiting time $W(P^{(k)};\Omega)$ for each sample. Then, N (N<K) best samples graded by their related average waiting time, are used to update the parameters σ, in order to produce better samples in the next iteration. The algorithm stops when the score of the worst selected sample no longer improves significantly.

3.3 Cross-Entropy Optimized Policies

We present here policies obtained from CE optimization and examine them in order to get insights concerning the optimal decision-making process in CRN. As in the previous sections we are interested to reveal the impact of the cognition cycle and the dynamics of the environment on the optimal policy. We set the parameters of the environment (Ω): the number of PU channels is M=6, and the transmission rate over every channel is μ=1, the constant B is set to unity, the parameters responsible for the environment dynamics are set to α=β=k – as before we will check the performance for different values of k={0.001, 1,1000}, the arrival rate of CRN traffic is λ=4.

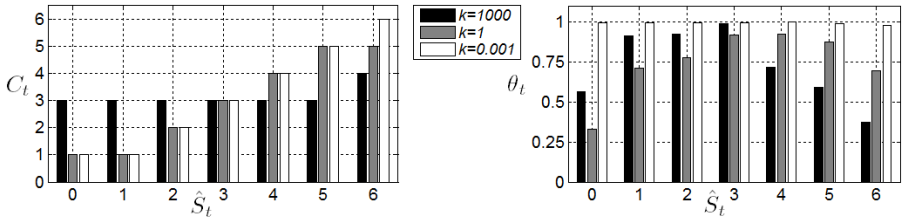

Fig. 3. The CE optimized policy for parameters M=6,μ=1,λ=4, α=β=k={1000,1,0.001}.

In our ASP the policy chooses action (C_t,θ_t) when CRN is in state $(\hat{S}_t,X_t)$. We assume that, C_t is a discrete random variable that takes integer values {0,1,…,M}, while the tradeoff parameter θ_t is normally distributed according to a truncated normal distribution in the range [0,1]. Recall, that our policy is state dependent. We distinguish between the cases X_t=0 and X_t>0. Obviously, for X_t=0 CRN has no packets to transmit

and in this case it is reasonable to allocate the bandwidth resources to the sensing process ($\theta_t = 1$). The CE algorithm optimizes the policy for $X_t>0$.

The resulting CE optimized policies are presented in Fig. 3. For the case k=1000, CRN fails to keep track of the rapidly changing network state. This can be seen through the channel allocation C=(3,3,3,3,3,3,4), which is insensitive to the estimation $\hat{S}_t$, and the number of accessed channels is approximately the average number of unoccupied channels E[S_t]. Nevertheless, the tradeoff parameter θ=(0.57,0.91,0.93,0.99,0.72,0.59,0.37) shows that CRN tries to avoid collisions with PU; a simple analysis of the CTMC (in Fig. 2) shows that for $\alpha=\beta$, S_t resides only a small portion of time in the states 0 and M while it spends more time in the inner states. This fact is reflected in the low values of θ when $\hat{S}_t$ is 0 or M. In order to better react to the fast network changes, CRN accelerates the sampling rate $\delta=(1-\theta)\mu B$ in these states.

For the case k=1, the resulting policy is more sensitive to the estimation of the environment state $\hat{S}_t$, and the number of accessed channels C=(1,1,2,3,4,5,5) is approximately $\hat{S}_t$ except for the rapidly switching states 0 and M. As in the previous case, the tradeoff parameter θ=(0.33,0.71,0.78,0.92,0.92,0.87,0.69), allocates more bandwidth for transmissions when $\hat{S}_t$ indicates that the network state is a persistent one. When the environment changes occur in a significantly slower manner compared to the rate of the perception process k=0.001, the tradeoff parameter θ=(0.99,0.99,0.99,0.99,0.99,0.98,0.98) takes very high values independently of the estimation $\hat{S}_t$. The allocation of the channels C=(1,1,2,3,4,5,6) is equal to $\hat{S}_t$ even for the rapidly switching state M.

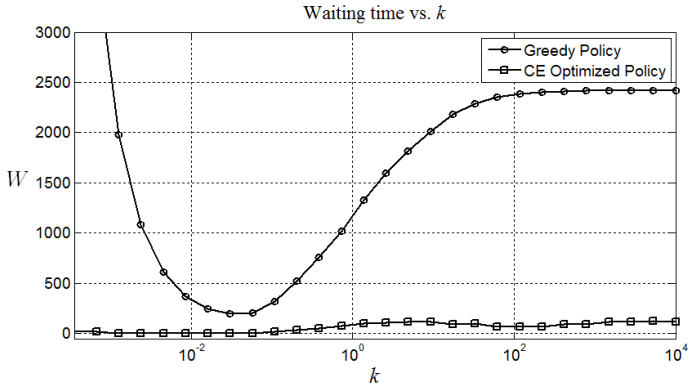

Fig. 4. CRN waiting time under greedy and CE optimized policies for parameters M=4,B=1,μ=1,λ=0.5, $\alpha=\beta=k\in[10^{-3},10^{4}]$.

In Fig. 4, we compare the performance of the CRN under greedy and CE optimized policies. Under the greedy policy, the waiting time W decreases when the network transitions accelerate (k<0.1). This happens since CRN tracks well the channels and efficiently utilizes the vacant ones. For k>0.1, W grows since CRN fails to track the fluctuating state of the network. When comparing the two policies, it can be seen that the waiting time, under CE optimized policy, does better in orders of magnitude for the entire range of network dynamics ($k\in[10^{-3},10^{4}]$).

4 Summary

In this paper, a three-dimensional CTMC process has been introduced to model the operation of CRN where PU form a birth-death process and SU can queue. The analytical framework combines the environment dynamics, perception and decision making components of the cognition cycle and the spectrum access processes. The cognition cycle is treated as an integral part of the system's overall behavior, and we optimize policies controlling simultaneously the interdependent perception and transmission processes. In this way, the resources are allocated according to the needs of the overall task. The CE optimized policies demonstrate adaptive behavior in which the resources are intelligently allocated to the perception and the transmission processes in a task-relevant manner.

5 References

[1] Zhao, Y., Mao, S., Neel, J., Reed, J.: Performance Evaluation of Cognitive Radios: Metrics, Utility Functions and Methodologies. Proceedings of the IEEE vol 97, Issue 4, (2009)

[2] Mitola, J., Maguire, G.Q.: Cognitive radio: Making software radios more personal. IEEE Pers. Commun., vol. 6, no. 4, pp. 13–18, (1999)

[3] Haykin, S.: Cognitive radio: brain-empowered wireless communications. IEEE J. on Selected Areas in Communications, vol. 23 pp. 201-220, (2005)

[4] Akyildiz, I. F., Lee W.-Y., Vuran, M. C., Mohanty, S.: A survey on spectrum management in cognitive radio networks. IEEE Communications Magazine, vol. 46, no. 4, pp. 40–48, (2008)

[5] Akyildiz, I. F., Lee, W.-Y., Vuran, M.C., Mohanty S.: NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey. Computer Networks, vol.50 pp.2127-2159, (2006)

[6] Cormio, C., Chowdhury, K. R.: A survey on MAC protocols for cognitive radio networks. Ad Hoc Networks, vol. 7, no. 7, pp. 1315-1329, (2009)

[7] Maldonado, D., Le, B., Hugine, A., Rondeau, T.W., Bostian, C.W.: Cognitive radio applications to dynamic spectrum allocation: a discussion and an illustrative example. DySPAN, (2005)

[8] Rashid, M.M., Hossain, M.J., Hossain, E., Bhargava, V.K.: Opportunistic spectrum scheduling for multiuser cognitive radio: a queueing analysis. Trans. Wireless. Comm. 8, pp. 5259-5269, (2009)

[9] Oklander, B., Sidi, M.: Modeling and Analysis of System Dynamics and State Estimation in Cognitive Radio Networks. PIMRC'10 (CogCloud Workshop), Istanbul, Turkey, (2010)

[10] Hoang, A. T., Liang, Y.C.: Adaptive Scheduling of Spectrum Sensing Periods in Cognitive Radio Networks. In Proc. IEEE GlobeCom, Washington D.C., USA, (2007)

[11] Liang, Y.-C., Zeng, Y., Peh, E., Hoang, A.T.: Sensing-throughput tradeoff for cognitive radio networks. In Proc. IEEE Int. Conf. Commun.(ICC), pp. 5330-5335, (2006)

[12] Ghasemi, A., Sousa, E.S.: Optimization of spectrum sensing for opportunistic spectrum access in cognitive radio networks. In Proc. 4th IEEE CCNC, pp.1022-1026, (2007)

[13] Oklander, B., Sidi, M.:Cross-Entropy Optimized Cognitive Radio Policies. CCIT Technical report #780, Technion, (2011)

[14] Rubinstein, R.Y., Kroese, D.P.: The Cross Entropy Method. A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning. ser. Information Science and Statistics, Springer, (2004)

[15] Mannor, S., Rubinstein, R.Y., Gat, Y.: The cross-entropy method for fast policy search. In Proceedings 20th International Conference on Machine Learning (ICML-03), Washington, US, pp. 21–24, (2003)