A Novel Method for Dynamic Multichannel Access Based on Whittle's Index and KS Sensing Model

Group14: Bo Chen, Wenji Qian, Zizhong Cao

***Abstract*—In the study of cognitive radio, the spectrum access is one of the most important aspect. And every SU(Secondary User) should not only consider how to make its own benefit most, but also pay attention to the influence on other SUs in order to achieve the relatively large benefits of the whole group. Thus game-theory and cooperation problems should be taken into consideration. In this paper, we apply a useful method, the whittle index, for each SU to judge how to make the decision on sensing while considering other SUs and its own benefits. We use the simplified Markovian Chain model while the real differences between sensing and transmitting time are also taken into consideration. And we assume that the sensing process obey the rule of KS(keep sensing) model. Thus, we build such a powerful model and get the proper ratio of one transmitting period to one sensing period by mathematical calculation and deduction, through which we can get the most profit under certain conditions. Then we prove how this algorithm works by giving its upper and lower bound of the benefits and make comparison to the optimal solution. Further illustration of the simulation results are displayed after that.**

***Index Terms*—Markovian Chain, KS model, Whittle Index**

I. Introduction

Today's wireless networks are characterized by a fixed spectrum assignment policy. However, a large portion of the assigned spectrum is used sporadically and geographical variations in the utilization of assigned spectrum ranges from 15% to 85% with a high variance in time. The limited available spectrum and the inefficiency in the spectrum usage necessitate a new communication paradigm to exploit the existing wireless spectrum opportunistically. This new networking paradigm is referred to as NeXt Generation (xG) Networks as well as Dynamic Spectrum Access (DSA) and cognitive radio networks.

A. Cognitive Radio

Cognitive radio techniques provide the capability to use or share the spectrum in an opportunistic manner. Dynamic spectrum access techniques allow the cognitive radio to operate in the best available channel. More specifically, the cognitive radio technology will enable the users to (1) determine which portions of the spectrum is available and detect the presence of licensed users when a user operates in a licensed band (spectrum sensing), (2) select the best available channel (spectrum management), (3) coordinate access to this channel with other users (spectrum sharing), and (4) vacate the channel when a licensed user is detected (spectrum mobility).

B. Hierarchical Access Model

This model adopts a hierarchical access structure with primary and secondary users. The basic idea is to open licensed spectrum to secondary users while limiting the interference perceived by primary users (licensees). Two approaches to spectrum sharing between primary and secondary users have been considered: Spectrum underlay and spectrum overlay. The underlay approach imposes severe constraints on the transmission power of secondary users so that they operate below the noise floor of primary users. By spreading transmitted signals over a wide frequency band (UWB), secondary users can potentially achieve short-range high data rate with extremely low transmission power. Based on a worst-case assumption that primary users transmit all the time, this approach does not rely on detection and exploitation of spectrum white space. Spectrum overlay was first envisioned by Mitola under the term spectrum pooling and then investigated by the DARPA Next Generation (XG) program under the term opportunistic spectrum access. Differing from spectrum underlay, this approach does not necessarily impose severe restrictions on the transmission power of secondary users, but rather on when and where they may transmit. It directly targets at spatial and temporal spectrum white space by allowing secondary users to identify and exploit local and instantaneous spectrum availability in a non-intrusive manner.

C. Restless Multi-armed Bandit Problem

Restless Multi-armed Bandit Processes (RMBP) are generalizations of the classical Multi-armed Bandit Processes (MBP), which have been studied since 1930's. In an MBP, a player, with full knowledge of the current state of each arm, chooses one out of N arms to activate at each time and receives a reward determined by the state of the activated arm. Only the activated arm changes its state according to a Markovian rule while the states of passive arms are frozen. The objective is to maximize the long-run reward over the infinite horizon by choosing which arm to activate at each time.
Whittle generalized MBP to RMBP by allowing multiple $(K \geq 1)$ arms to be activated simultaneously and allowing passive arms to also change states. Either of these two generalizations would render Gittins' index policy suboptimal in general, and finding the optimal solution to a general RMBP has been shown to be PSPACE-hard.

D. The Gilber-Elliot channel model.

Consider the problem of probing N independent Markov chains. Each chain has two states. .good. and .bad.. with different transition probabilities across chains (see Fig. 1). At each time, a player can choose K ($1 \leq K < N$) chains to probe and receives reward determined by the states of the probed chains.

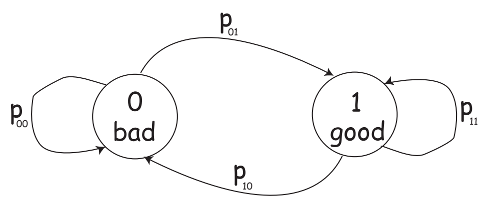

Fig. 1. The Gilber-Elliot channel model

E. Game Theory

Some concepts of game theory date back centuries, but modern game theory began in the mid-20th century. One of its earliest modern making by aggressive superpowers. A more enduring application has been as a powerful array of techniques for modeling economic behavior. The basic unit of game theory is, of course, the game. A game has three basic elements:

- A description of strategic interaction between players
- A set of constraints on the actions the players can take
- A specification of the interests of the players

Games are usually represented in one of two forms: the normal form and the extensive form. The normal form game for two players is represented as a bi-matrix. An extensive form game is depicted as a tree, where each node represents a decision point for one of the players. The normal form is easier to analyze, but the extensive form captures the structure of a real game in time.

F. KS Sensing Model

KS Scheme (Keep-Sensing-if-Busy): After a vacation, the SU(Secondary User) senses the channel. If the channel is idle, the SU transmits a packet and then starts vacation. If the SU senses the channel busy, it keeps sensing until the channel is idle. Then, the SU transmits a packet and starts a random vacation of length V2.

II. RELATED WORK

Dynamic spectrum access among cognitive radios can be realized by an adaptive, game theoretic learning perspective. Spectrum-agile cognitive radios compete for channels temporarily vacated by licensed primary users in order to satisfy their own demands while minimizing interference. Reference [1] applies an adaptive regret based learning procedure which tracks the set of correlated equilibria of the game, treated as a distributed stochastic approximation. And this illustrates that by adding some degradation factor we can recalculate the value of the decision and make the predicted benefits of whole

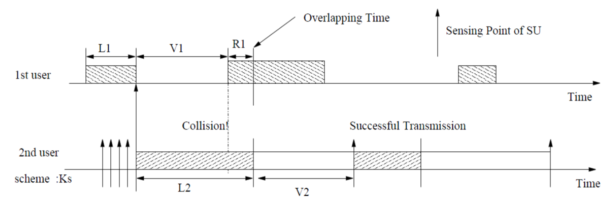

Fig. 2. KS sensing model

group largest by solving the differential equations. By the given degradation index we can get the most proper channel numbers to sense. In Reference [8], it shows the difference of both non-cooperative and cooperative game theory in static and dynamic settings. Careful attention is given to techniques for demonstrating the existence and uniqueness of equilibrium in non-cooperative games. And there are more about the game theory from Reference [16] about the applications of game theory to supply chain analysis and outlines game-theoretic concepts that have potential for future application.

Another aspect of the spectrum access includes the random process analysis. Reference [17] considers a scenario where secondary users can opportunistically access unused spectrum vacated by idle primaries. Supposing the PU's starting using one channel obeying poisson distribution, we can get the max transformation rate under certain limited collision rate by Probability Theory. And Reference [18] develops opportunistic scheduling policies for cognitive radio networks that maximize the throughput utility of the secondary (unlicensed) users subject to maximum collision constraints with the primary (licensed) users. It considers a cognitive network with static primary users and potentially mobile secondary users. The model assumes state whether the channel is idle is a kind of Markov Chain. The paper uses the technique of Lyapunov Optimization to design an online flow control, scheduling and resource allocation algorithm that meets the desired objectives and provides explicit performance guarantees.

In Reference [3],the spectrum access is optimal in that it strikes a balance between two conflicting needs: keeping spectrum assessment overhead low while increasing the likelihood of discovering spectrum opportunities. It study the effect of several network parameters, such as the primary traffic load, the secondary traffic load, and the collaboration level of the sensing method.

Reference [4] deal with multi-armed bandit problem for a gambler is to decide which arm of a K-slot machine to pull to maximize his total reward in a series of trials. It provides a preliminary empirical evaluation of several multi-armed bandit algorithms. It also describes and analyzes a new algorithm, Poker (Price Of Knowledge and Estimated Reward) whose performance compares favorably to that of other existing algorithms in several experiments.

In Reference [9],it considers a class of restless multi-armed bandit problems (RMBP). And it establishes indexability and obtain Whittle's index in closed-form for both discounted and average reward criteria. These results lead to a direct implementation of Whittle's index policy with remarkably low complexity. Furthermore, it has a semi-universal structure that obviates the need to know the Markov transition probabilities. In Reference [12] it provides an method to jointly detect the primary signals over multiple frequency bands rather than over one band at a time. By exploiting the hidden convexity in the seemingly nonconvex problems, optimal solutions can be obtained for multiband joint detection under practical conditions with certain constraints. To address this issue by exploiting the spatial diversity, a cooperative wideband spectrum sensing scheme refereed to as spatial-spectral joint detection is proposed, which is based on a linear combination of the local statistics from multiple spatially distributed cognitive radios. The cooperative sensing problem is also mapped into an optimization problem, for which suboptimal solutions can be obtained through mathematical transformation under conditions of practical interest.

And the most important paper we pay attention to is Reference [20], which make a combination of the Reference [9] and Reference [12]. When arms are stochastically identical, it shows that Whittle's index policy is optimal under certain conditions. Like Reference[9], it need some background knowledge of Markov transition probabilities. The optimality and the semi-universal structure result from the equivalency between Whittle's index policy and the myopic policy established in this work. For non-identical arms, it develops efficient algorithms for computing a performance upper bound given by Lagrangian relaxation.

Other References such as [2], [5], [6], [7], [10], [11], [13], [14], [15], [19] help to clarify the cognitive radio and dynamic spectrum access and offer more relevant background information and knowledge into our research.

III. Problem Statement and Formulation

A. Multi-channel Opportunistic Access

1) the classic Gilbert-Elliot channel model: Consider N independent Gilbert-Elliot channels, each with transmission rate B_i(i=1,$\cdots$,N). The classic model is shows in the [20],The state of channel i–"good"(1) or "bad"(0)– evolves from slot to slot as a Markov chain, as showed in Figure 1.

2) the advanced Gilbert-Elliot channel model: The basic Gilber-Elliot channel model is quite simple and is capable of representing the actual conditions to some extent. However, this basic model assumes that the time period of sensing equals that of transmission, which might not be the best way, comparing the loss of efficiency due to keeping idle with the cost of sensing at a higher frequency. Thus, if we let the cognitive radio sense more frequently in the "bad" case, the total efficiency is likely to improve quite much. Based on this intension, we devide the time slots into smaller pieces and try to incorporate the KS sensing model as showed in Figure 2 into the Gilber-Elliot channel model.

Still we assume that the channel condition remains the same as the Gilbert-Elliot channel, and the changes only affect the sensing mode. Assume that n represents the ratio of a single series of transmission period to one sensing period. And assume that before the transmission period n, the SU will not sense the channel even if it has finished sending the data. Based on the assumptions above, we put forward an advanced model as showed in the Figure 5.

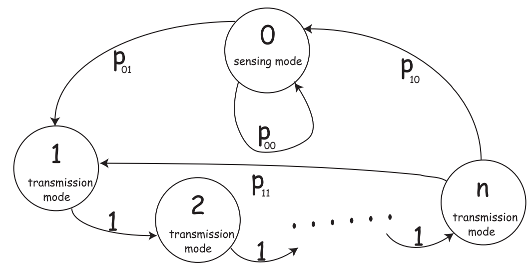

Fig. 3. the advanced Gilbert-Elliot model

At the beginning of slot t, if the state S_i(t) of the sensed cannel is 1, the SU transmits and collects B_i units of reward in this channel and the transmission period last for n_i(t) time slot, which is no more than n. Otherwise, the user collects no reward in this channel.

Our objective is to maximize the expected long-run reward by designing a sensing policy that which channels are selected to sense in each slot, and prove that based on the advanced model a suitable number n will improve the performance of the channel significantly.

B. Basic Analyze of The Advanced Gilbert-Elliot Channel

Obviously if n equals 1, the advanced model will have the same performance as the basic model do and since the channel's performance will remain the same as it only related to the other users while has no relationship with the sensing period. Then the transfer matrix of the Markov chain will change according to different n. Assume that the when n=1, the transfer time from one state to another in the Markov chain is T, the transfer matrix $\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$ can be the index of the channel. Then when $n > 1$, which means transfer time changes from T to $\frac{T}{n}$, Transfer Matrix will change to $\begin{pmatrix} p_{00}^{(n)} & p_{01}^{(n)} \\ p_{10}^{(n)} & p_{11}^{(n)} \end{pmatrix}$. Obviously the two matrices have the following relationship:

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} p_{00}^{(n)} & p_{01}^{(n)} \\ p_{10}^{(n)} & p_{11}^{(n)} \end{pmatrix}^n \quad (1)$$

We can see that

$$\begin{cases} p_{00} + p_{11} - 1 = (p_{00}^{(n)} + p_{11}^{(n)} - 1)^n, \\ \frac{1-p_{00}}{1-p_{11}} = \frac{1-p_{00}^{(n)}}{1-p_{11}^{(n)}}. \end{cases} \tag{2}$$

In condition that $p_{11} > p_{01}$,

$$\begin{cases} p_{11}^{(n)} = \frac{p_{01}+(1-p_{11})\sqrt[n]{p_{11}-p_{01}}}{1+p_{01}-p_{11}}, \\ p_{01}^{(n)} = \frac{p_{01}(1-\sqrt[n]{p_{11}-p_{01}})}{1+p_{01}-p_{11}}. \end{cases} \tag{3}$$

The channel states are not directly predictable before the sensing action is made. The user can, however, partially infer the channel states from its decision and observation history. Assume $\omega(t)$ is the conditional probability that the channel is "good". Refered to as the belief vector or information state, the belief state in the time slot t+1 can be obtained recursively as follows:

$$\omega(t+1) = \begin{cases} p_{01}^{(n)}, & S(t) = 0 \\ p_{11}, & S(t) = n \\ T(\omega(t)), & not - sensed \end{cases} \tag{4}$$

where

$$T(\omega(t)) \triangleq \omega(t)p_{11}^{(n)} + (1-\omega(t))p_{01}^{(n)} \tag{5}$$

denotes the operator for the one-step belief update for unobserved channels.

If no information on the initial system state is available, the initial belief vector can be set to the stationary distribution ω_0 of the underlying Markov chain:

$$\omega_0 = \frac{p_{01}^{(n)}}{p_{10}^{(n)} + p_{01}^{(n)}} = \frac{p_{01}}{p_{01} + p_{10}} \tag{6}$$

We suppose to work on the model and to find the long-run reward based on the benefit from the data transmitted and the penalty from the sensing cost.

IV. How to get the subsidy m

We now present the formal definition of indexability and Whittle's index. We first consider the discounted reward criterion. Their definitions under the average reward criterion can be similarly obtained. Denoted by $V_{\beta,m}(\omega)$, the value function represents the maximum expected total discounted reward that can be accrued from a single-armed bandit process with subsidy m when the initial belief state is ω. Considering the two possible actions in the first slot,we have

$$V_{\beta,m}(\omega) = max\{V_{\beta,m}(\omega; u=0), V_{\beta,m}(\omega; u=1)\} \tag{7}$$

where $V_{\beta,m}(\omega; u)$ denotes the expected total discounted reward obtained by taking action u in the first slot followed by the optimal policy in future slots. Consider $V_{\beta,m}(\omega; u = 0)$. It is given by the sum of the subsidy m obtained in the first slot under the passive action and the total discounted future reward $\beta * V_{\beta,m}(T(\omega))$ which is determined by the updated belief state $T(\omega)$. $V_{\beta,m}(\omega)$ can be similarly obtained, and we arrive at the following dynamic programming.

$$V_{\beta,m}(\omega)(\omega; u=0) = m + \beta \times V_{\beta,m}(T(\omega)) \tag{8}$$

$$V_{\beta,m}(\omega)(\omega; u=1) = \omega + \beta(\omega \times V_{\beta,m}(p_{11}) + (1-\omega) \times V_{\beta,m}(p_{01})) \tag{9}$$

The optimal action $u_m^*(\omega)$ for belief state ω under subsidy m is given by

$$u_m^*(\omega) = \begin{cases} 1, & if\ V_{\beta,m}(\omega; u=1) > V_{\beta,m}(\omega; u=0) \\ 0, & otherwise \end{cases} \tag{10}$$

The passive set P(m) under subsidy m is given by

$$P(m) = \begin{cases} \{\omega : u_m^*(\omega) = 0\} \\ \{\omega : V_{\beta,m}(\omega; u=0) \geq V_{\beta,m}(\omega; u=1)\} \end{cases} \tag{11}$$

An arm is indexable if the passive set P(m) of the corresponding single-armed bandit process with subsidy m monotonically increases from ϕ to the whole state space [0; 1] as m increases from -∞ to +∞. An RMBP is indexable if every arm is indexable. Under the indexability condition, Whittle's index is defined as follows.

If an arm is indexable, its Whittle's index $W(\omega)$ of the state ω is the infimum subsidy m such that it is optimal to make the arm passive at ω. Equivalently, Whittle's index W(ω) is the infimum subsidy m that makes the passive and active actions equally rewarding.

$$\begin{aligned} W(\omega) &= \inf_m \{m : u_m^*(\omega) = 0\} \\ &= \inf_m \{m : V_{\beta,m}(\omega; u=0) = V_{\beta,m}(\omega; u=1)\}. \end{aligned} \tag{12}$$

The optimality of Whittle's Index Policy under a Relaxed Constraint: Whittle's index policy is the optimal solution to a Lagrangian relaxation of RMBP. Specifically, the number of activated arms can vary over time provided that its discounted average over the infinite horizon equals to K. Let K(t) denote the number of arms activated in slot t. The relaxed constraint is given by

$$E_\pi[(1-\beta)\sum_{n=1}^{\infty} \beta^{t-1} K(t)] = K. \tag{13}$$

Let $\bar{V}_\beta(\Omega(1))$ denote the maximum expected total discounted reward that can be obtained under this relaxed constraint when the initial belief vector is $\Omega(1)$. Based on the Lagrangian multiplier theorem, we have

$$\bar{V}_\beta(\Omega(1)) = \inf_m \{\sum_{n=1}^{N} V_{\beta,m}^{(i)}(\omega_i(1)) - m\frac{N-K}{1-\beta}\} \tag{14}$$

The above equation reveals the role of the subsidy m as the Lagrangian multiplier and the optimality of Whittle's index policy for RMBP under the relaxed constraint given in (13). Specifically, under the relaxed constraint, Whittle's index policy is implemented by activating, in each slot, those arms whose current states have a Whittle's index greater than a constant m^*. This constant m^* is the Lagrangian multiplier that makes the relaxed constraint given in (13) satisfied, or equivalently, the Lagrangian multiplier that achieves the infimum in (14). It is not difficult to see that Whittle's index policy implemented by comparing to a constant m^* is the optimal policy.

Now we provide the whittle index's equation.

First, we define an important quantity $L(\omega, \omega\prime)$. Referred to as the crossing time, $L(\omega, \omega\prime)$ is the minimum aomount of time required for a passive arm to transit across $\omega\prime$ starting from ω.

$$L(\omega, \omega\prime) \triangleq min\{k : T^k(\omega) > \omega\prime\}. \tag{15}$$

For a positively correlated arm, we have the formulation of the crossing time in equation (16).

Based on it the whittle index can be formulate in equation (17).

V. General Analyze Based on the whittle index

We can get the properties of the Whittle's Index that $W_\beta(\omega)$ is a monotonically increasing function of ω(the lemma has been proved in the reference [9]), such that when the Index is settled as m^* the strategy of one channel will be settled as well.

Suppose that the initial value ω is initialized as ω_0. Since that we focus on the positively correlated channels. The belief of the unobserved channels update as showed in fig.4,fig.5.

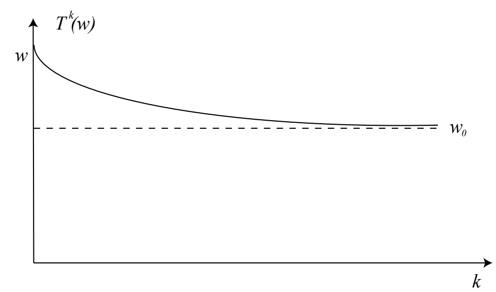

Fig. 4. Belief update of an unobserved arm(1)

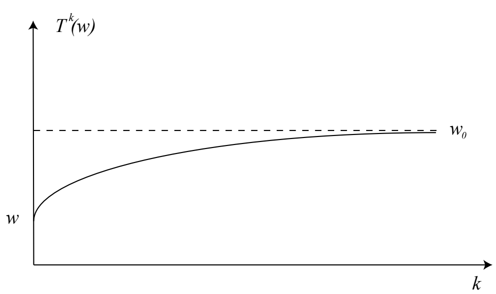

Fig. 5. Belief update of an unobserved arm(2)

Then to certain channel, when $W_\beta(\omega_0) \leqslant m^*$ the channel will never be sensed. IN that And when $W_\beta(p_{01}^{(n)}) > m^*$. the channel will be always sensed if considered. Then we focus on the left area: at time 0, $W_\beta(\omega) = W_\beta(\omega_0) > m^*$, the channel will be sensed.

CASE1:If the channel is available, then after the transmission of the data, ω will be p_{11} , the whittle index definitely will be higher than m^*, then the channel will be sensed again;

CASE2:If the channel is not available, then at next period, the belief ω will fall down to $p_{01}^{(n)}$, which is lower than m^*, the channel will not be sensed, but after constant steps—$L(p_{01}^n, m^*)$,it will be sensed again.

Now we give out the expected benefit in discounted reward in the three cases.

1) $W_\beta(\omega_0) \leqslant m^*$**:** No sense no transmission, then the gain will be 0, since the cost of sensing is 0 too, the final benefit is 0;
2) $W_\beta(p_{01}^{(n)}) > m^*$**:** The channel will be always sensed. Assume that E_0 represents the expected transmission gain if at this moment the channel is sensed occupied by the primary user, while E_1 represents the expected transmission gain if at this moment the transmission of last packet is just finished. Obviously we will have the following equations:

$$\begin{cases} E_0 = \sqrt[n]{\beta} p_{01}^{(n)}(B + \beta E_1) + \sqrt[n]{\beta} p_{00}^{(n)} E_0 \\ E_1 = \sqrt[n]{\beta} p_{11}^{(n)}(B + \beta E_1) + \sqrt[n]{\beta} p_{10}^{(n)} E_0 \end{cases} \tag{20}$$

So that we will get the following expression:

$$\begin{cases} E_0 = \frac{\sqrt[n]{\beta} p_{01}^n}{(1 - \sqrt[n]{\beta} p_{00}^n)(1 - C_3\beta)} B \\ E_1 = \frac{C_3}{1 - C_3\beta} B \end{cases} \tag{21}$$

Where,

$$C_3 = \frac{\sqrt[n]{\beta} p_{11}^n - \sqrt[n]{\beta^2}(p_{11}^n p_{00}^n + p_{10}^n p_{01}^n)}{1 - \sqrt[n]{\beta} p_{00}^n} \tag{22}$$

Then the expected total gain will be:

$$E = \omega_0(B + \beta E_1) + (1 - \omega_0)E_0 \tag{23}$$

We can see that E will increase with n, and when $n \to \infty, C_3 \to 1$, so that

$$\begin{cases} E_1 \to \frac{B}{1-\beta} \\ E_0 \to \frac{\beta p_{01} B}{(p_{10}p_{11} - p_{10}p_{01} + p_{01})(1-\beta)} \end{cases} \tag{24}$$

In that case, E will converge to a constant. The conclusion is meaningful for that the infinite represents that the channel will start to transmit data every time when the channel was not occupied.

Now given a real parameter α, which represents the source consumed one time sensed Compared to the data transmission. Then assume that S_0 represents the expected cost if at this moment the channel is sensed occupied by the primary user, while S_1 represents the expected cost if at this moment the transmission of last packet is just finished. Then

$$\begin{cases} S_0 = \sqrt[n]{\beta} p_{01}^{(n)} \beta S_1 + \sqrt[n]{\beta} p_{00}^{(n)} S_0 + \alpha \sqrt[n]{\beta} \\ S_1 = \sqrt[n]{\beta} p_{11}^{(n)} \beta S_1 + \sqrt[n]{\beta} p_{10}^{(n)} S_0 + \alpha \sqrt[n]{\beta} \end{cases} \tag{25}$$

$$L(\omega,\omega\prime)=\begin{cases}0, & if\ \omega>\omega\prime\\ \lfloor\log_{p_{11}^{(n)}-p_{01}^{(n)}}\frac{p_{01}^{(n)}-\omega\prime(1-p_{11}^{(n)}+p_{01}^{(n)})}{p_{01}^{(n)}-\omega(1-p_{11}^{(n)}+p_{01}^{(n)})}\rfloor, & if\ \omega\leq\omega\prime<\omega_0\\ \infty, & if\ \omega\leq\omega\prime\,and\,\omega\prime\geq\omega_0\end{cases} \tag{16}$$

$$W_{\beta,n}(\omega)=\begin{cases}\omega B_{(i)}, & if\ \omega\leq p_{01}^{(n)}or\omega\geq p_{11}^{(n)}\\ \frac{\omega}{1-\sqrt[n]{\beta}p_{01}^{(n)}+\sqrt[n]{\beta}\omega}B_{(i)}, & if\ \omega_0\leq\omega<p_{11}^{(n)}\\ \frac{\omega-\sqrt[n]{\beta}T^1(\omega)+C_2(1-\sqrt[n]{\beta})(\sqrt[n]{\beta}(1-\sqrt[n]{\beta}p_{11}^{(n)})-\sqrt[n]{\beta}(\omega-\sqrt[n]{\beta}T^1(\omega)))}{1-\sqrt[n]{\beta}p_{11}^{(n)}-C_1(\sqrt[n]{\beta}(1-\sqrt[n]{\beta}p_{11}^{(n)})-\sqrt[n]{\beta}(\omega-\sqrt[n]{\beta}T^1(\omega)))}, & if\ p_{01}^{(n)}<\omega<\omega_0\end{cases} \tag{17}$$

where

$$C_1=\frac{(1-\sqrt[n]{\beta}p_{11}^{(n)})(1-\sqrt[n]{\beta}^{L(p_{01}^{(n)},\omega)})}{(1-\sqrt[n]{\beta}p_{11}^{(n)}(1-\sqrt[n]{\beta}^{L(p_{01}^{(n)},\omega)+1})+(1-\sqrt[n]{\beta})\sqrt[n]{\beta}^{L(p_{01}^{(n)},\omega)+1}T^{L(p_{01}^{(n)},\omega)}(p_{01}^{(n)})} \tag{18}$$

$$C_2=\frac{\sqrt[n]{\beta}^{L(p_{01}^{(n)},\omega)}T^{L(p_{01}^{(n)},\omega)}(p_{01}^{(n)})}{(1-\sqrt[n]{\beta}p_{11}^{(n)}(1-\sqrt[n]{\beta}^{L(p_{01}^{(n)},\omega)+1})+(1-\sqrt[n]{\beta})\sqrt[n]{\beta}^{L(p_{01}^{(n)},\omega)+1}T^{L(p_{01}^{(n)},\omega)}(p_{01}^{(n)})} \tag{19}$$

With the same process as the transmit gain, we have the expression:

$$\begin{cases}s_1=\frac{\alpha\sqrt[n]{\beta}}{1-\sqrt[n]{\beta}p_{11}^n\beta-\sqrt[n]{\beta}p_{10}^nC_4}\\ s_0=C_4S_1\end{cases} \tag{26}$$

where

$$C_4=\frac{1+\sqrt[n]{\beta}(p_{01}^n-p_{11}^n)\beta}{1+\sqrt[n]{\beta}(p_{10}^n-p_{00}^n)} \tag{27}$$

This time when $n\rightarrow\infty, S_0\rightarrow\infty$,so that with the increase of n, the cost will diverge to infinite. We know that the general benefit will be the gap between the two parameter.

3) $W_\beta(p_{01}^{(n)})\leqslant m^*<W_\beta(\omega_0)$**:** It is the most complicated case, but with the fundamental analysis above and the transform function of Whittle Index, we will find that the result are nearly the same.

First of all, we should use the Whittle's index to get the critical ω, which stands for $W(\omega)=m^*$(as discussed before the inverse function is a single-valued function), then we can get the crossing time t_n with the equation (16).

Then we can get the function of E_0,E_1,S_0,S_1 follow the equation:

$$\begin{cases}E_0=\beta^{\frac{t_n}{n}}T^{t_n}(p_{01}^{(n)})(B+\beta E_1)+\beta^{\frac{t_n}{n}}(1-T^{t_n}(p_{01}^{(n)}))E_0\\ E_1=\sqrt[n]{\beta}p_{11}^{(n)}(B+\beta E_1)+\sqrt[n]{\beta}p_{10}^{(n)}E_0\\ S_0=\beta^{\frac{t_n}{n}}T^{t_n}(p_{01}^{(n)})\beta S_1+\beta^{\frac{t_n}{n}}(1-T^{t_n}(p_{01}^{(n)}))S_0+\alpha\sqrt[n]{\beta}\\ S_1=\sqrt[n]{\beta}p_{11}^{(n)}\beta S_1+\sqrt[n]{\beta}p_{10}^{(n)}S_0+\alpha\sqrt[n]{\beta}\end{cases} \tag{28}$$

The only difference between these equations and which in the former case are some parameters which will not feeble the basic properties discussed before. So the Gain and the cost will still increase with n, and when $n\rightarrow\infty$, the expected gain will converge and the expected cost will diverge.

The properties of the function discussed above is very important, for it strongly support our work for that the optimal choice n is a finite constant.

VI. Analyze for Stochastically Identical Channels

Since the channels are stochastically identical channels, they share the same transfer rate, then according to the monotonicity of the whittle index, we know that the optimal policy is the myopic policy. In this case our policy is much more simpler.

Assume that there are M i.i.d channels waiting to be sensed, and we only have k sensing machines.

VII. Simulation Results and Analysis

In the simulation process, we use the Matlab to simulate the process of channel sensing and data transmitting. Here n means the ratio of the transmission period verse the sensing period. We suppose that the transmission time is fixed according to the transmission protocol which protects the Qos of PUs. A larger n means ahigher sensing frequency of the SU(seconcdary user). And the sensing cost is positively related to the sensing count during the simulation. We set the total process 1000 transmission steps. Then we get the sensing cost-n relationship and throughput-n relationship picture and some other relevant pictures.

Thus we get the following pictures as results:

Through Pic.6, we can see that the overall throughput of a secondary user is booming when n is relative small. As n becomes larger, the throughput gradually get close to a fixed limitation which means it can not increase any more. This does make sense since the channel capacity and use rate has a limitation.

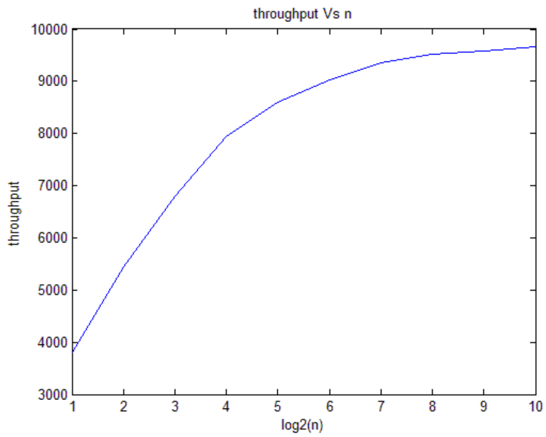

Fig. 6. Throughput-N

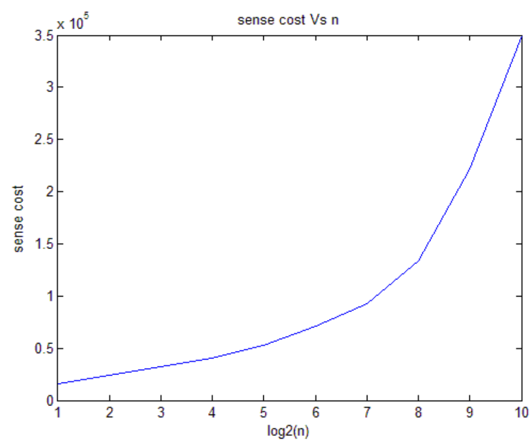

Fig. 7. Sensing cost-N

Through Pic.7, we find out that the overall sensing cost is increasing rapidly when n gets larger. What's more, the larger n is, the higher the increasing rate of sensing cost is. It is reasonable because when n increases, it is more difficult to meet with a "good" state.

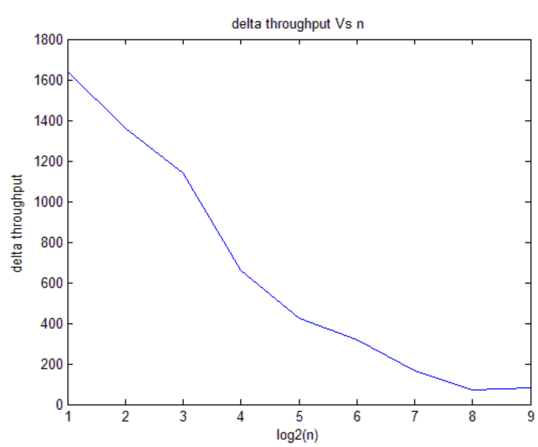

Fig. 8. Delta throughput-N

Through Pic.8, we discover that the changing rate of the throughput based on n, namely the increment of throughput when n increases by 1. And we find that the changing rate is monotonously decreasing. This can be predicted by Pic.6 since the Pic.8 is just like the differentiate of Pic.6 on N.

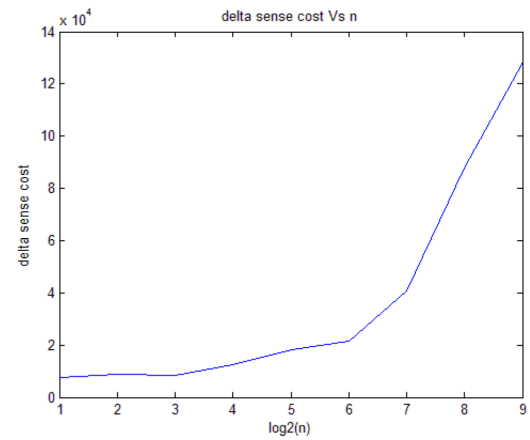

Fig. 9. Delta sensing cost-N

Through Pic.9, we know the changing rate of the sensing cost which means the sensing increment of cost when n pluses 1, is increasing. And it accelerates when n continuously gets larger. It is related to Pic.7 and can be seen as the differentiate of Pic.7 on n.

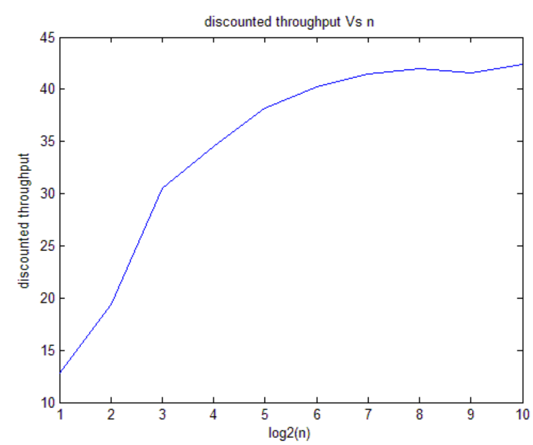

Fig. 10. Discounted throughput-N

Through Pic.10, we discuss about the discounted throughput under the discount reward criterion, which means the far future benefits are less valuable. And we can get such result that Pic.10 is similar to Pic.6, which shows that the discount reward criterion and the average reward criterion has the similar results. Yet by comparing their changing rate, Pic.7 and Pic.11, we can see they have some differences. There are some fluctuations in Pic.11 and the curve is not monotonously decreasing. This may have something to do with the randomness of the "good" state. Since the discounted criterion pays more attention on the first several steps' gain,

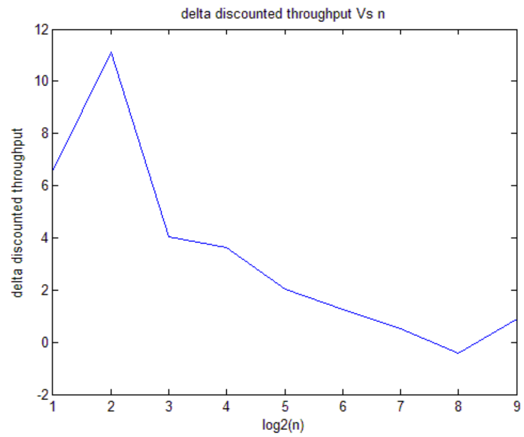

Fig. 11. Pic.6 Delta discounted throughput-N

which may vary greatly. While the curve in Pic.7 is using the average standard, which can undermine the fluctuation by calculate the average values.

By all the pictures above, we can see that throughput is increasing to a fixed limitation with the growth of n. Meanwhile, the sensing cost is accelerating. The total benefit of SU can be written as

$$totalbenefit = overallthroughput - totalsensingcost \tag{29}$$

Thus we can find out that there must be a optimal n, which can make the total benefit most. And we also realize that this benefit can be derived by minusing the curve in Pic.6 with the cuvre in Pic.7. Consequently, we see that the result curve have a peak and its corresponding n value. Another way to get the optimal n is to put Pic.8 and Pic.9 in one picture. And the curves must have a intersection point. And the point's corresponding n is the value we want. It is quite understandable because when each step n pluses 1, the increment of throughput decrease and the sensing cost increases. When the two variance get the same, the increment of total benefit will no longer increase, which means we get the benefit peak right now.

By using the methods mentioned above, we can get the optimal n to get most benefits considering both the throughput and the sensing cost.

VIII. Conclusion and Future Work

In this period, we provide the idea of subsidy m, which mean the offset of not sensing. And by considering the effect of m, we can judge whether the channel should be sensed or not more scientifically and rationally. We find two criterions to judge the decision's future benefits. Here we have future average criterion and future discounted criterion. Based on this two criterions, we use the tool of Whittle Index to make mathematic deductions to decide the value of m. Then we achieve the formation of future benefits, which helps to make the sensing decision. Then we find the method of using Whittle Index is equal to myopic under certain conditions. Then it is easy for us to make simulation of our theory by using myopic method instead. And through simulation results, we get more interesting discoveries which have been discussed in 'Simulation Results and Analysis'.

In the next period, we will finish our proofs of the theory by giving out detail justifications. Now, we only find proper proofs for the discounted criterion, yet the average criterion still need further consideration. And we will also simplify the equations and get more simple and clear formations. Meanwhile, we find out that Whittle Index is quite complex and we will try to find an easier way to get the subsidy m.

References

[1] Michael Maskery, Vikram Krishnamurthy and Qing Zhao *"Decentralized Dynamic Spectrum Access for Cognitive Radios: Cooperative Design of a Non-cooperative Game,"* New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN

[2] Qing Zhao and Brian M. Sadler], *"A Survey of Dynamic Spectrum Access,"* IEEE Signal Processing Magazine, 2007

[3] Bechir Hamdaoui, *"Adaptive Spectrum Assessment for Opportunistic Access in Cognitive Radio Networks,"* IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, VOL. 8, NO. 2, FEBRUARY 2009

[4] Joann'es Vermorel and Mehryar Mohri, *"Multi-Armed Bandit Algorithms and Empirical Evaluation,"* Courant Institute of Mathematical Sciences, 719 Broadway, New York, NY 10003, USA

[5] Ian F. Akyildiz *, Won-Yeol Lee, Kaushik R. Chowdhury, *"CRAHNs: Cognitive radio ad hoc networks,"* Volume 7, Issue 5, July 2009, Pages 810-836

[6] Zhu Ji and K. J. Ray Liu, University of Maryland, *"Dynamic Spectrum Sharing: A Game Theoretical Overview,"* Z Ji, KJR Liu - IEEE Communications Magazine, 2007 - cspl.umd.edu

[7] P. S. Ansell, K. D. Glazebrooky, J. Nina-Mora and M. O keeffe, *"Whittles index policy for a multi-class queueing system with convex holding costs,"* PS Ansell, KD Glazebrook, J Nino-Mora, MO'Keeffe - Mathematical Methods of Operations Research (ZOR), 2003 - Springer

[8] Gerard P. Cachon, *"GAME THEORY IN SUPPLY CHAIN ANALYSIS,"* GP Cachon, S Netessine - INTERNATIONAL SERIES IN OPERATIONS RESEARCH AND MANAGEMENT, 2004

[9] Keqin Liu, Qing Zhao, *"Indexability of Restless Bandit Problems and Optimality of Whittle's Index for Dynamic Multichannel Access"* J Nino-Mora - Advances in Applied Probability, 2001

[10] Yunxia Chen, Student Member, IEEE, Qing Zhao, Member, IEEE, and Ananthram Swami, Fellow, IEEE, *"Joint Design and Separation Principle for Opportunistic Spectrum Access in the Presence of Sensing Errors",* IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 54, NO. 5, MAY 2008

[11] Ian F. Akyildiz, Won-Yeol Lee, Mehmet C. Vuran *, Shantidev Mohanty, *"NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey"* Computer Networks 50 (2006) 2127C2159

[12] Zhi Quan, Student Member, IEEE, Shuguang Cui, Member, IEEE, Ali H. Sayed, Fellow, IEEE, and H. Vincent Poor, Fellow, IEEE, *"Optimal Multiband Joint Detection for Spectrum Sensing in Cognitive Radio Networks"* IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 57, NO. 3, MARCH 2009

[13] D. Bertsimas and J. Nino-Mora, *"Restless Bandits, Linear Programming Relaxations and a Primal-Dual Heuristic"* D Bertsimas, J Ni?o-Mora - Operations Research, 2000

[14] Qing Zhao and Brian M. Sadler, *"A Survey of Dynamic Spectrum Access"*Q Zhao, BM Sadler - IEEE Signal Processing Magazine, 2007

[15] Qing Zhao, Lang Tong, Ananthram Swami, and Yunxia Chen, *"Decentralized Cognitive MAC for Opportunistic Spectrum Access in Ad Hoc Networks: A POMDP Framework"* IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 25, NO. 3, APRIL 2007

[16] Gerard P. Cachon, *"GAME THEORY IN SUPPLY CHAIN ANALYSIS"* GP Cachon, S Netessine - INTERNATIONAL SERIES IN OPERATIONS RESEARCH AND MANAGEMENT , 2004

[17] Senhua Huang, Xin Liu, and Zhi Ding, *"Opportunistic Spectrum Access in Cognitive Radio Networks"*

[18] Rahul Urgaonkar, Student Member, IEEE, Michael J. Neely, Member, IEEE, *"Opportunistic Scheduling with Reliability Guarantees in Cognitive Radio Networks"* R Urgaonkar, MJ Neely - ieeexplore.ieee.org

[19] Qing Zhao, Bhaskar Krishnamachari, Keqin Liu, *"On Myopic Sensing for Multi-Channel Opportunistic Access: Structure, Optimality, and Performance"* IEEE Transactions on Wireless Communications, 2007
[20] Keqin Liu, Qing Zhao, *"Indexability of Restless Bandit Problems and Optimality of Whittle's Index for Dynamic Multichannel Access"* IEEE Transactions on Information Theory , 2008