A New Mechanism of Dynamic Spectrum Access Based on Restless Bandit Allocation Indices

Zhu Jiang
Chongqing key lab of
Mobile Communications
Technology,Chongqing
University of Posts and
Telecommunications
Chongqing, China
zhujiang_cqupt@163.com

Han Chao
Chongqing key lab of
Mobile Communications
Technology,Chongqing
University of Posts and
Telecommunications
Chongqing, China
hancao.cool @163.com

Yang Haolei
Chongqing key lab of
Mobile Communications
Technology,Chongqing
University of Posts and
Telecommunications
Chongqing, China
yanghl0725@163.com

Xiong Jiahao
Chongqing key lab of
Mobile Communications
Technology,Chongqing
University of Posts and
Telecommunications
Chongqing, China
602165121@qq.com

Abstract

Based on the theory of Restless Multi-Armed Bandit model, a novel mechanism of dynamic spectrum access was proposed for the problem that how to coordinate multi-user access multi-channel which were idle. Firstly, take care of the sensing error must be exist in the practical network, the Whittle index policy which can deal with sensing error effectively was derived, in this policy, the users achieved one belief value for every channel based on the historical experience accumulation and chose the channels, which was need to sense and access, by considering the reward of immediate and future based on the belief values. Secondly, this paper used the multi-bid auction algorithm to deal with the collision among secondary users when they select the channels to improve the spectrum utilization. The simulation results demonstrate that, in the same environment, the cognitive users with the proposed mechanism have higher throughput than the mechanism without dealing with sensing error or without multi-bid.

Categories and Subject Descriptors

C.2.1[Computer-Communication Networks]: Network Architecture and Design–Wireless communication, Distributed networks.

General Terms

Design

Keywords

restless multi-armed bandit model, sense error, multi-bid auction,Whittle's index value

1. INTRODUCTION

Now, there are two major problems in the wireless communication network. Firstly, because of the scarcity of spectrum resources, the high-efficiency of spectrum is very important[1]. Secondly, it is very difficult to gain the complete information of network environment. Therefore, it is the key that how to optimize the usage of spectrum resources under limited information of network environment.

In the wireless network, the spectrum access problem based on cognitive radio has become the research hotspot[2-3]. Wherein, RMAB model has adopted in many literatures, because it has the characteristics that dosen't depend on the information of environment to optimize[4]. In [5],the problem of multichannel allocation in single-hop mobile networks with multiple service classes was formulated as RMAB, and the sufficient conditions for the optimality of a myopic type index policy were established. In [6] proposed the spectrum access mechanism based on Whittle's index policy, proved the indexability of Whittle's index policy, and achieved the optimal solution of Whittle's index policy under the Lagrangian relaxation.

The existing spectrum access researches based on RMAB model have the following characteristics: on one hand, they do not consider the impact of the user's current channel access for future access behavior; on the other hand, there are no one corresponding mechanism or rule to deal with the collision generated when multi-user access the same channel. Therefore, we proposed the spectrum access mechanism, considering the impact of the user's current channel access for future access behavior, deducing the Whittle's index policy which can deal with sensing error effectively, and improving the strategy of channels access. Meanwhile, we achieved the spectrum optimization allocation through the multi-bid auction mechanism to deal with the collision of users. Finally, we verified that the proposed mechanism can realize the approximate optimal allocation of spectrum allocation in the network and achieved a better spectrum efficiency.

2. SYSTEM MODEL

2.1. Channel Model

We consider a cognitive network with K users, the users can dynamically use a spectrum which was divided into N channels, the set of users and channels can express as K $\{1,\mathbf{K},K\}$, N $\{1,\mathbf{K},N\}$ respectively. The channels states space can write as $\mathbf{S}$ $\{s_1(t)\mathbf{L}s_N(t)\}\in\{0,1\}^N$. The state of channel n can model as Markov Chain, the transition probability is $\Pr(s_n(t)\rightarrow s_n(t+1))$, and these N channels are modeled as independent[7]. Due to the factor of fading, noise, the difference of users location and so on, different channels have different qualities for the same user, as well, same channel has different qualities for the different users. The channels qualities are time varying for the same user on the same channel, but, the average of channels qualities can be regard as not time-varying. Therefore, if channel n is allocated to user n, then user n can gain the average transfer rate with $g_{n,k}$.

2.2. Protocol Architecture

As Fig 1, in time slot t, assuming every user has a chance to access $|M_k(t)|$ channels. We use the multi-bid auction mechanism to deal with the collision of users. In RMAB model, the users update the belief value of channels according to the history information. The belief values reflect the idle probability of channel. Then, the users will introduce an index that measures how attractive it is to activate a particular arm at its current state, and activate those $|M_k(t)|$ arms with the largest index at each time slot. After all, the users give bid based on the

belief value for the channels which they chose respectively, the seller achieves the optimal channels allocation according to the winner judgment algorithm, then, transmits the outcome to the winners. We define the set of allocation outcome of user k as $C_k(t)$, user k can sense the channels which are in the set, if the sensing outcome is 1, then, user k will access the channel and transmit data. After the transmission, the receiver and transmitter of user k will transmit the ACK or NAK to indicate the success or failure of the communication respectively. In the last of time slot, the users update belief value of channels according to propose algorithm.
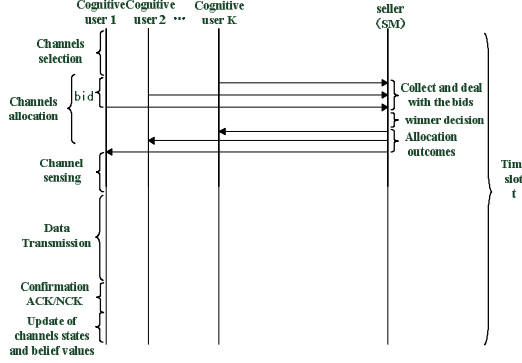

Fig. 1   protocol frame

3. MULTI-USER DYNAMIC SPECTRUM ACCESS MECHANISM BASED ON RMAB MODEL

3.1. Description of RMAB Model and Belief Value Update

It is known from the protocol, at the beginning of time slot t, the users do not sense the channels, so, they don't know the states of channels. But, the users can achieve the conditional probabilities that each channel is in state 1 given all past decisions and observations. Also, the probability is a sufficient statistic for optimal making decision [8] and it is called belief value in this paper. We define the belief vector is $W_k(t)$ $\{w_{1,k}(t), \mathbf{K}, w_{N,k}(t)\}$, $w_{n,k}(t)$ is the probability that user k deems that channel k is idle at the beginning of time slot t. Let $\bar{S}_k(t)$ $\{\bar{s}_{1,k}(t), \mathbf{K}, \bar{s}_{N,k}(t)\}$ denote the sensing outcome of user k for N channels, and if $n \in C_k(t)$, there is $\bar{s}_{n,k} \in \{0,1\}$, otherwise, $\bar{s}_{n,k} = Null$. We consider the sensing errors of users, assuming the false alarm probability and missed detection probability are independent with t and n, the expression of the false alarm probability and missed detection probability are as following, respectively:

$$e_k = \Pr(\bar{s}_{n,k}(t) = 0 \mid s_n(t) = 1)$$
$$d_k = \Pr(\bar{s}_{n,k}(t) = 1 \mid s_n(t) = 0) \tag{1}$$

Since user is the unknown of the real states of channels, user k can verify the accuracy of channel sensing outcome according to response information $a_{n,k}(t) \in \{0(\text{NAK}), 1(\text{ACK})\}$ which was gained after data transmission[8](assumption the response information is completely correct). After considering the false alarm probability and missed detection probability, user k can update the belief value through equation (2) at time slot $t+1$ according to the set of allocated channel $C_k(t)$, the sensing outcome $\bar{S}_k(t)$ and the response information $a_{n,k}(t)$.

$$w_{n,k}(t+1) = \begin{cases} f(F), & n \in C_k(t), \bar{s}_{n,k}(t) = 0 \\ p_n(11), & n \in C_k(t), \bar{s}_{n,k}(t) = a_{n,k}(t) = 1 \\ f(E), & n \in C_k(t), \bar{s}_{n,k}(t) = 1, a_{n,k}(t) = 0 \\ f(w_{n,k}(t)), & n \notin C_k(t) \end{cases} \tag{2}$$

Where:

$$E = P(s_n(t) = 1 \mid \bar{s}_{n,k}(t) = 1) = \frac{(1-e_k)w_{n,k}(t)}{(1-e_k)w_{n,k}(t) + d_k(1-w_{n,k}(t))}$$

$$F = P(s_n(t) = 1 \mid \bar{s}_{n,k}(t) = 0) = \frac{e_k w_{n,k}(t)}{e_k w_{n,k}(t) + (1-d_k)(1-w_{n,k}(t))}$$

$$f(w_{n,k}(t)) = w_{n,k}(t)p_n(11) + (1-w_{n,k}(t))p_n(01)$$

The believe value $W_k(t)$ is the idle probability of each channel which was achieved by user k at time slot t. Here, we have modeled the dynamic multi-channel access problem as RMAB model and solved it. The standard pattern of RMAB model can express as: $\left(W(t), \{\Pr(s_n(t) \to s_n(t+1)) : 1 \le n \le N\}, R, b\right)$. The value function denotes the maximum expected reward which was achieved for time slot t under the believe value $W(t)$, which is denoted below:

$$V(t, W_k(t)) = \max_{C_k(t)} \{R_{C_k(t)}(t) +$$
$$b \, \mathrm{E}_{\bar{S}_k(t)}\left(V(t+1, W_k(t+1) \mid (C_k(t), \bar{S}_k(t), a_{n,k}(t)))\right)\} \tag{3}$$

where $R_{C_k(t)}(t) = \sum_{n \in C_k(t)} w_{n,k}(t)(1-e_k)g_{n,k}$ is the immediate reward for the selected channels.

3.2. Whittle's Index Policy with Sensing Error

We have modeled the dynamic spectrum access problem as RMAB model, unfortunately, the optimal solution to a RMAB problem is often intractable: the problem has been shown to be PSPACE-hard[9]. So, we adopt the Whittle's index policy to solve the RMAB problem which was proposed by Whittle in 1988, this policy can gain a index value for each channel based on the belief value, which denotes the attraction of choosing the channel, and then user k chooses the $|M_k(t)|$ channels with the largest index as the set of sense channels at each time slot. In the meantime, we give the subsidy m for not choosing channel to make the selection more comprehensive and accurate, then, gain maximum throughput as possible. What's more, Whittle also proposed the indexability of RMAB model when use the Whittle's index policy [8]. In this paper, we used the two lemmas to achieve the indexability and the expression of Whittle's index.

Lemma 1：An arm is indexable if the passive set $U(m)$ of the corresponding single-armed bandit process with subsidy m monotonically increases from $\varnothing$ to the whole state space $[0,1]$ as m increases from $-\infty$ to ∞. An RMAB $\left(W(t), \{\Pr(s_n(t) \to s_n(t+1)) : 1 \le n \le N\}, R, b\right)$ is indexable if every arm is indexable[9].

Lemma 2：If an arm is indexable, its Whittle index $W(w_{n,k}(t))$ of the state $w_{n,k}(t)$ is the infimum subsidy m so that it is optimal to make the arm passive (not be selected) at $w_{n,k}(t)$. Equivalently, Whittle's index $W(w_{n,k}(t))$ is the

infimum subsidy m that makes the passive and active actions equally rewarding:

$$W\left(w_{n,k}(t)\right) = \inf_{m} \left\{ \begin{array}{l} m : V_{b,m}\left(t, w_{n,k}(t); u_n = 0\right) \\ = V_{b,m}\left(t, w_{n,k}(t); u_n = 1\right) \end{array} \right\} \quad (4)$$

Through a series of analysis and reasoning, the Whittle's index has been achieved under the ideal situation[9].

Theorem 1: Whittle's index with sensing error is shown as follow:

While $p_n(11) < p_n(01)$

$$W\left(w_{n,k}(t)\right) = \begin{cases} \dfrac{w_{n,k}(t)\,g_{n,k}\,(1-b\,I)(1-e_k)}{1-b\,I(1-e_k)} & w_{n,k}(t) \le p_n(11) \\[2ex] \dfrac{\left(w_{n,k}(t)+b\left(p_n(01)-w_{n,k}(t)\right)\right)(1-e_k)}{1+b\left(p_n(01)-w_{n,k}(t)\right)(1-e_k)} g_{n,k} \\ \qquad\qquad f_n\left(p_n(11)\right) \le w_{n,k}(t) < p_n(01) \\[2ex] \dfrac{\left(w_{n,k}(t)+b\left(p_n(01)-w_{n,k}(t)\right)\right)(1-e_k)\,g_{n,k}}{1+b\,(1-b)\left(p_n(01)-w_{n,k}(t)\right)(1-e_k)-H} \\ \qquad\qquad w_n(0) \le w_{n,k}(t) < f_n\left(p_n(11)\right) \\[2ex] \dfrac{w_{n,k}(t)(1-e_k)}{1-b\left(w_{n,k}(t)-p_n(11)\right)(1-e_k)} g_{n,k} \\ \qquad\qquad p_n(11) < w_{n,k}(t) < w_n(0) \\[1.5ex] w_{n,k}(t)\,g_{n,k}\,(1-e_k) \qquad w_{n,k}(t) \ge p_n(01) \end{cases} \quad (6)$$

where:

$$H = \left(b\left(1-e_k\right)\right)^2 p_n(11)\,I$$

While $p_n(11) > p_n(01)$

$$W\left(w_{n,k}(t)\right) = \begin{cases} \dfrac{g_{n,k}\,w_{n,k}(t)\left((1-e_k)(1-b\,p_n(01))-G\right)}{1-b\,(1-e_k)\,I} \\ \qquad\qquad w_{n,k}(t) \le p_n(01) \\[2ex] \dfrac{w_{n,k}(t)(1-e_k)\,g_{n,k}}{1-b\left(p_n(11)-w_{n,k}(t)\right)(1-e_k)} \\ \qquad\qquad w_n(0) \le w_{n,k}(t) < p_n(11) \\[2ex] \dfrac{\left(P\left(1-b^{L+2}\right)+J\right)g_{n,k}}{(1-G)(1-b)+J+b\,P\left(b-b^{L+2}\right)} \\ \qquad\qquad p_n(01) < w_{n,k}(t) < w_n(0) \\[1.5ex] g_{n,k}\,w_{n,k}(t)(1-e_k) \qquad w_{n,k}(t) \ge p_n(11) \end{cases} \quad (5)$$

where,

$$f_n^z\left(w_{n,k}(t)\right) = \frac{p_n(01)-(I)^z\left(p_n(01)-(1-I)w_{n,k}(t)\right)}{1-I}$$

$$L = \left\lceil \frac{\left|\lg\left(\dfrac{p_n(01)-w_{n,k}(t)(1-I)}{p_n(01)(I)}\right)\right|}{\lg(I)} \right\rceil + 1$$

$$I = p_n(11) - p_n(01)$$

$$J = b^{L+2}(1-b)\,f_n^{L+1}\left(p_n(01)\right)(1-e_k)$$

$$P = \left(w_{n,k}(t) - b\,f\left(w_{n,k}(t)\right)\right)(1-e_k)$$

$$G = b\,p_n(11)(1-e_k)$$

3.3. Multi-Bid Auction Mechanism

From the protocol frame, there would generate collisions when many users choose the same channel. So, we use the multi-bid auction mechanism to deal with the collisions, and achieve the reasonable utilization of spectrum resources. We regard the environment of the users as the marketplace, the goods in the marketplace are the two-dimensional time-frequency channel resources. The process of multi-bid auction is shown above, each auction round corresponding to a time slot, and the key steps include of bid, winner decision, Levy a tax. the concrete process is shown below:

a) Bid

Define $\boldsymbol{B}_k(t)$ as the vector of bid of user k at time slot t, also, $\boldsymbol{B}_{-k}(t)$ is the vector of bid which stands for excepting for user k at time slot t. $b_{n,k}(t) \in \boldsymbol{B}_k(t)$ denotes the bids which was paid by user k for channel n:

$$\boldsymbol{B}_k(t) \quad \left\{ b_{n,k}(t) \mid n \in \boldsymbol{N} \right\} \quad (7)$$

Use $\boldsymbol{V}_k(t)$ denote the vector of value of user k at time slot t, $v_{n,k}(t) \in \boldsymbol{V}_k(t)$ is the value which was gained by user k from channel n at time slot t, $v_{n,k}(t)$ depends on the belief value $w_{n,k}(t)$, the reward in history, average transfer rate $g_{n,k}$:

$$\boldsymbol{V}_k(t) \quad \left\{ v_{n,k}(t) = g_{n,k}\,w_{n,k}(t) \mid n \in \boldsymbol{N} \right\} \quad (8)$$

b) Winner Decision

After the t round bid, the channel i was allocated to user which was determined by the winner decision. The basic standard of winner decision is to maximize the auctioneer's revenue, that is:

$$\max\left(\sum_{n=1}^{N} \sum_{k=1}^{K} b_{n,k}(t)\,a_{n,k}(t) \right)$$
$$\text{subject to: } \sum_{n=1}^{N} a_{n,k}(t) = 1, \quad a_{n,k}(t) \in \{0,1\} \quad (9)$$

c) Strategy of Levy a Tax

Because users are rational and selfish, so there are some users obtaining the usage of some certain channels by cheating ways. For example, user k will obtain the usage of channel n by improving the bid of channel n, i.e. $b_{n,k}(t) > v_{n,k}(t)$. So, we use levy strategy to deal with the problem of lying about the bid. The levy strategy is defined as:

$$h_k\left(\boldsymbol{B}_{-k}(t), \boldsymbol{B}_k(t)\right) \quad \sum_{n=1}^{N} \sum_{i=1, i \ne k}^{K} b_{n,i}(t)\,a_{n,i}^*(t) -$$
$$\max_{\boldsymbol{A}(t)}\left(\sum_{n=1}^{N} \sum_{i=1, i \ne k}^{K} b_{n,i}(t)\,\aleph_{n,i}(t) \right) \quad (10)$$

The above equation denotes the tax which was levied by the seller, while the bid vector of user k is $\boldsymbol{B}_k(t)$ and other users' are $\boldsymbol{B}_{-k}(t)$. The first term is the sum of real utilities that other users make under the allocation vector $\mathbf{A}^*(t)$ in presence of user k. The second summation, however is the best aggregate utility that others would have made under the new allocation vector $\aleph(t)$ which had not user k been present at all. Obviously, $h_k\left(\boldsymbol{B}_{-k}(t), \boldsymbol{B}_k(t)\right)$ is non-positive. The reward $s_k\left(\boldsymbol{B}_{-k}(t), \boldsymbol{B}_k(t)\right)$ of user k at time slot t can be calculated according to the above-mentioned levy scheme:

$$s_k\big(\boldsymbol{B}_{-k}(t),\boldsymbol{B}_k(t)\big)\ \sum_{n=1}^{N}v_{n,k}(t)a_{n,k}^{*}(t)+h_k\big(\boldsymbol{B}_{-k}(t),\boldsymbol{B}_k(t)\big)\quad (11)$$

4. SIMULATION ANALYSIS

In this paper, we mainly simulate in two different miss detection probabilit $\big(d=0.05;d=0.12\big)$ environments(for without loss of generality, we set $d_1=\boldsymbol{K}=d_K=d$), we set $\big|\boldsymbol{M}_k(t)\big|=4$ in the following simulation. The labels in the simulation diagrams will be explained as follow: 'W + error + bid' denoting the mechanism that the Whittle's index policy which consider the sensing error and the multi-bid auction. After all, we define the enumerative algorithm to gain the approximate system traffic limit as the basis of mechanism performance judgment:
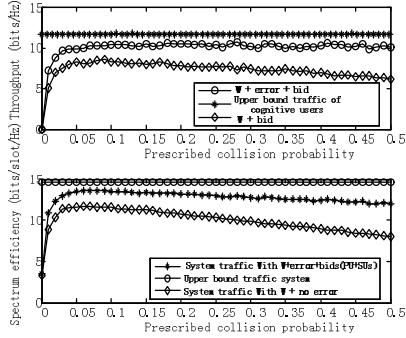

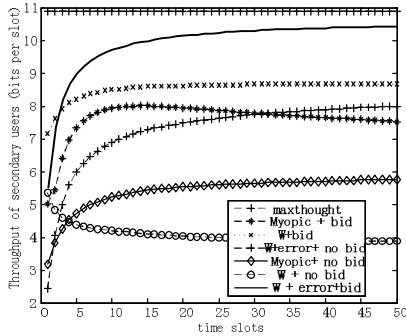**Fig 2：the impact of missed detection probabilitys**
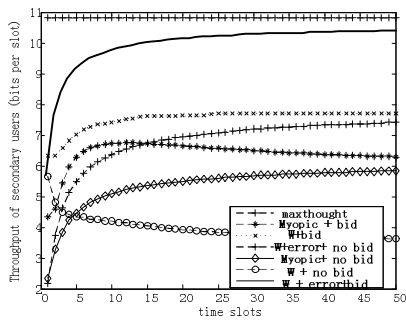

Fig 3: $K=7,N=15,d=0.05$,the comparison of algorithms

Fig 4: $K=7,N=15,d=0.12$,the comparison of algorithms.

In Fig.2 $\big(K=7,N=15\big)$, we can see that the mechanism which have processed the sensing error making the throughput of users not changing under different miss detection probability, but for the mechanism with not dealing with sensing errors, the throughput are declining. Also, the total system throughputs are all declining with the increasing of miss detection probability in both mechanisms.

In Fig.3 and Fig.4 we used different miss detection probability, In Fig.3 we set $d=0.05$, Fig.4 $d=0.12$, we can see that the Whittle's index policy which have processed the sensing error can make the user evolve the idle probability of channel in this slot accurately, and improve the throughput of his own. In the meantime, there is decline for the performance of access mechanism with no sensing error with the increase of miss detection probability

5. SUMMARY

In this paper, we have considered the problem that the users didn't know the states and the usage of the channels, when they inter to the new network area. We proposed one adaptive channel access mechanism based on RMAB model which is under the situation that multi-user can access multi-channel simultaneously. In addition, we combine the Whittle's index policy and multi-bid auction under the RMAB model and achieve the optimal channel bandwidth reward (throughput) for the users. The simulation results show that, the proposed mechanism can make the users gain higher channel bandwidth than other access mechanism, and when the channel model was modeled as a continuous switching process model, the users can still learn the state transfer situation accurately according to the historical experience, and improve the utilization rate of the channel effectively.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multi-channel opportunistic access: Structure, optimality, and performance,"[J] IEEE Trans. Wireless Commun., vol. 7, no. 12, pp. 5431–5440, Dec.2008.

[2] S. H. Ahmad, M. Liu, T. Javadi, Q. Zhao, and B. Krishnamachari, "Optimality of myopic sensing in multi-channel opportunistic access ,"IEEE Trans. Inf. Theory, vol. 55, no. 9, pp. 4040–4050, Sep. 2009.

[3] K. Liu, Q. Zhao, and B. Krishnamachari, "Dynamic multichannel access with imperfect channel state detection," IEEE Trans. Signal Process. , vol. 58, no. 5, pp. 2795–2808, May 2010.

[4] K. Liu, Q. Zhao, and B. Krishnamachari, "Dynamic multichannel access with imperfect channel state detection," [J] IEEE Trans. Signal Process. , vol. 58, no. 5, pp. 2795–2808, May 2010.

[5] C. Lott and D. Teneketzis, "On the optimality of an index rule in multichannel allocation for single-hop mobile networks with multiple service classes," Probab. Eng. Inf. Sci., vol. 14, pp. 259–297, 2000

[6] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in Ad hoc networks: A POMDP framework," [J] IEEE J. Sel. Areas Commun., vol. 25, no. 3, pp. 589–600, Apr. 2007.

[7] C. H. Papadimitriou and J. N. Tsitsiklis, "The Complexity of Optimal Queueing Network Control," in Mathematics of Operations Research [R] ,Vol. 24, No. 2, May 1999, pp. 293-305.

[8] P. Whittle, "Restless bandits: Activity allocation in a changing world", in Journal of Applied Probability, [J] Volume 25, 1988.

[9] Ny J L, Dahleh M, Feron E. Multi-UAV dynamic routing with partial observations using restless bandit allocation indices[C]//American Control Conference, 2008. IEEE, 2008: 4220-4225.