CONCURRENT BANDITS AND COGNITIVE RADIO NETWORKS

MINOR PROJECT II

Submitted by:

PRASOON JAIN (9915103228)
PIYUSH VASHISTHA (9915103265)
PRASHANT MAHESHWARI (9915103266)
PRATAYA AMRIT (9915103267)
SPARSH MAJHE(9915103276)

Under the supervision of:
MR. HIMANSHU AGRAWAL

Department of CSE/IT
Jaypee Institute of Information Technology University, Noida

MARCH 2018

ACKNOWLEDGEMENT

The success of this project required a lot of guidance and assistance from many people and we are extremely fortunate to have got this all along the completion of our project work. Whatever we have done is only due to such guidance and assistance and we could not forget to thank them.

We respect and thank our mentor Mr. Himanshu Agrawal for giving us an opportunity to do the project working fog computing and providing us all support and guidance which made us complete the project on time. We are extremely grateful to him for providing such a nice support and guidance though he had busy schedule managing the academics.

We owe our profound gratitude to our project mentor Mr. Himanshu Agrawal, who took keen interest on our project work and guided us all along by providing all the necessary information for developing a good system.

We would again heartily thank our internal project guide, Mr. Himanshu Agrawal, Department of Computer Science, for his guidance and suggestions during this project work.

We are thankful and fortunate enough to get constant encouragement, support and guidance from all teaching staffs of Department of Computer Science which helped us in successfully completing our project work.

Also, we would like to extend our sincere regards to al the non-teaching staff and Lab Technicians of Department of Computer Science for their timely support.

Signature of Students

PRASOON JAIN (9915103228) ....

PIYUSH VASHISTHA (9915103265) ....

PRASHANT MAHESHWARI (9915103266) ....

PRATAYA AMRIT (9915103267) ....

SPARSH MAJHE (9915103276) ....

TABLE OF CONTENTS

ABSTRACT

We are dealing with the problem of multiple secondary users targeting the arms of a single multi-armed bandit which leads to collision at the arms. The reason to deal with this problem comes from cognitive radio networks. In CRN the secondary user need to coexist without any side communication, cooperation between them. Even the number of user may be unknown and can vary as other user join or leave the network. We are proposing an algorithm that combines an e-greedy learning rule with collision avoidance mechanism. We analyze the expected regret with respect to the system and show that sub-linear regret can be obtained in those scenarios.

LIST OF FIGURES

ABBREVIATIONS

S. No	Abbreviation	Full Form
1.	MEGA	Multi-user e-Greedy collision Avoiding algorithm
2.	CRNs	Cognitive Radio Networks
3.	MAB	Multi-armed bandit

INTRODUCTION

In this paper we deal with the fundamental problem arising in dynamic multi-user communication networks(cognitive radio networks). We tried to design a network of independent users competing over communication channels, represented by arms of stochastic multi armed bandit. Now let us begin by explaining the background, talking about the general model, reviewing the previous work and introducing our contribution.

1.1 Cognitive Radio Networks

A cognitive radio is an intelligent radio that can be programmed and configured dynamically. Radio that automatically detects available channels in wireless spectrum, accordingly changes its transmission or reception parameters to allow more concurrent wireless communication in a given spectrum band. Users in this network are divided into two types namely: primary and secondary. The primary users are licensed users who enjoy precedence over secondary user in term of access to network resources. The secondary users face the challenge of identifying and exploiting the available resources. The characteristics of the primary user vary slowly whereas the characteristics of secondary user tend to be dynamic in nature. The secondary users are unaware of each other, thus there is no reason to assume the existence of any cooperation or communication between them. Also they are not able to know the number of secondary users in the network. Another one of the critical features of cognitive radio networks is their distributive nature, which means that the central control does not exist.

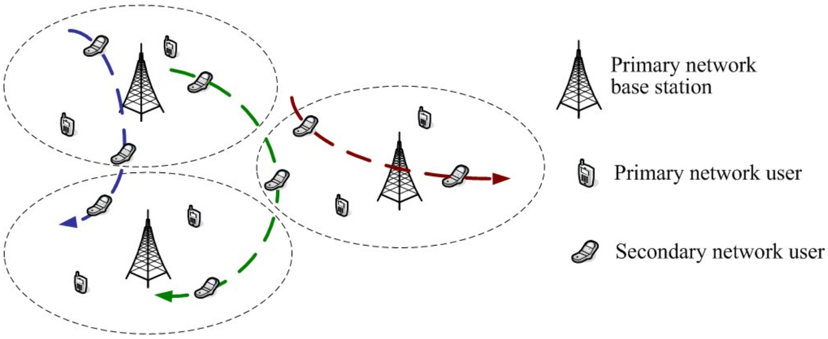


Fig. 1. This fig. shows different users in CRNs.

The resulting problem is quite challenging: multiple users, coexisting in an environment whose characteristics are initially unknown, acting selfishly in order to achieve an individual performance. We approach this problem from the point of view of a single secondary user, and introduce an algorithm which, when applied by all secondary users in the network, enjoys promising performance.

1.2 Multi-armed bandits

Multi-armed bandits comes in the domain of Machine Learning. The exploration and exploitation in sequential decision problems have been used in context of learning in cognitive radio networks. In classical bandit problem, the users repeatedly choose a single arm (option) from a set of options whose characteristics are initially unknown, receiving a certain reward based on every choice. The user wishes to maximize the acquired reward and in order to do so we must balance the exploration of unknown arms and exploitation of attractive ones.

We adopt the MAB framework in order to solve the challenge of secondary user, choosing between several unknown communication channels. The characteristics of the channels are assumed to be fixed,. The challenges we address in this paper arise from the fact that there are multiple secondary users in the network.

Fig. 2. Shown Machine represents channel as Multi-armed Bandit.

1.3 Multiple users playing MAB

A natural extension of the CRN-MAB framework described above considers multiple users attempting to exploit resources represented by the same bandit. The multi-user setting leads to collisions between users, due to both exploration and exploitation of the attractive arm in terms of reward. This arm will be targeted by all users, once it has been identifieded as such. In real-life communication systems, collisions result in impaired performance. In our model, reward loss is the natural sign or pecrception of collisions. As one might expect, straightforward applications of classical bandit algorithms designed for the single-user case. In the absence of some form of a collision avoidance mechanism, all users attempt to sample the same arm after some time. We therefore face the problem of sharing a resource and learning its characteristics when users cannot communicate.

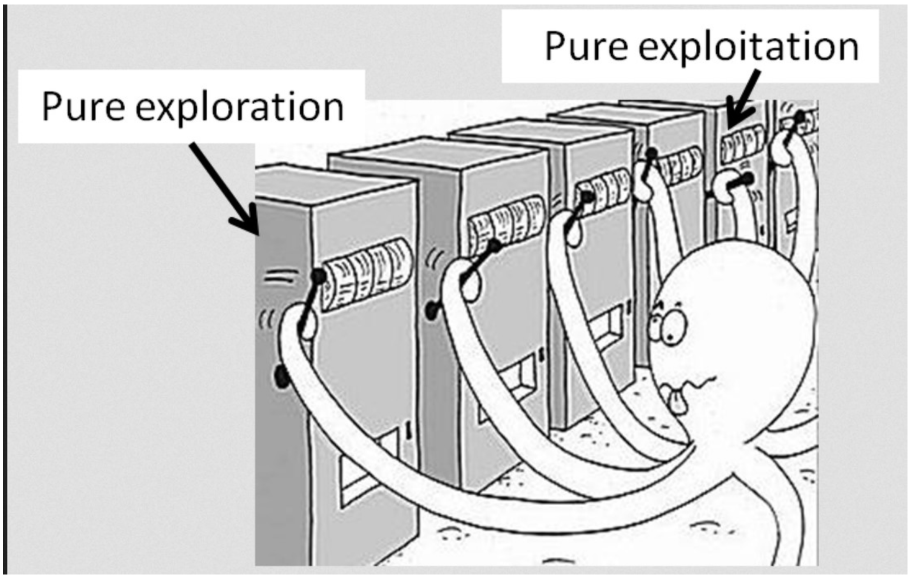


Fig. 3. Showing different users playing MAB's.

1.4 Related Work

In past few years, considerable effort has been put into finding a solution for the multi-user CRN-MAB problem. One approach, is based on a Time-Division Fair Sharing (TDFS) of the best arms between all users. This policy enjoys good performance guarantees but has two significant drawbacks. First, the number of users is assumed to be fixed and known to all users, and second, the implementation of a TDFS mechanism requires pre-agreement among users to coordinate a time division schedule. Another work that deals with multiuser access to resources, but does not incorporate the MAB setting is, the users reach an orthogonal configuration without pre-agreement or communication, using multiplicative updates of channel sampling probabilities based on collision information. However, this approach does not handle the learning aspect of the problem and disregards differences in the performance of different channels. Thus, it cannot be applied to our problem. Consider a form of the CRN-MAB problem in which channels appear different to different users, and propose an algorithm which enjoys good performance guarantees. However, this approach includes a negotiation phase, based on the Bertsekas auction algorithm, during which the users communicate in order to reach an orthogonal or stable configuration.

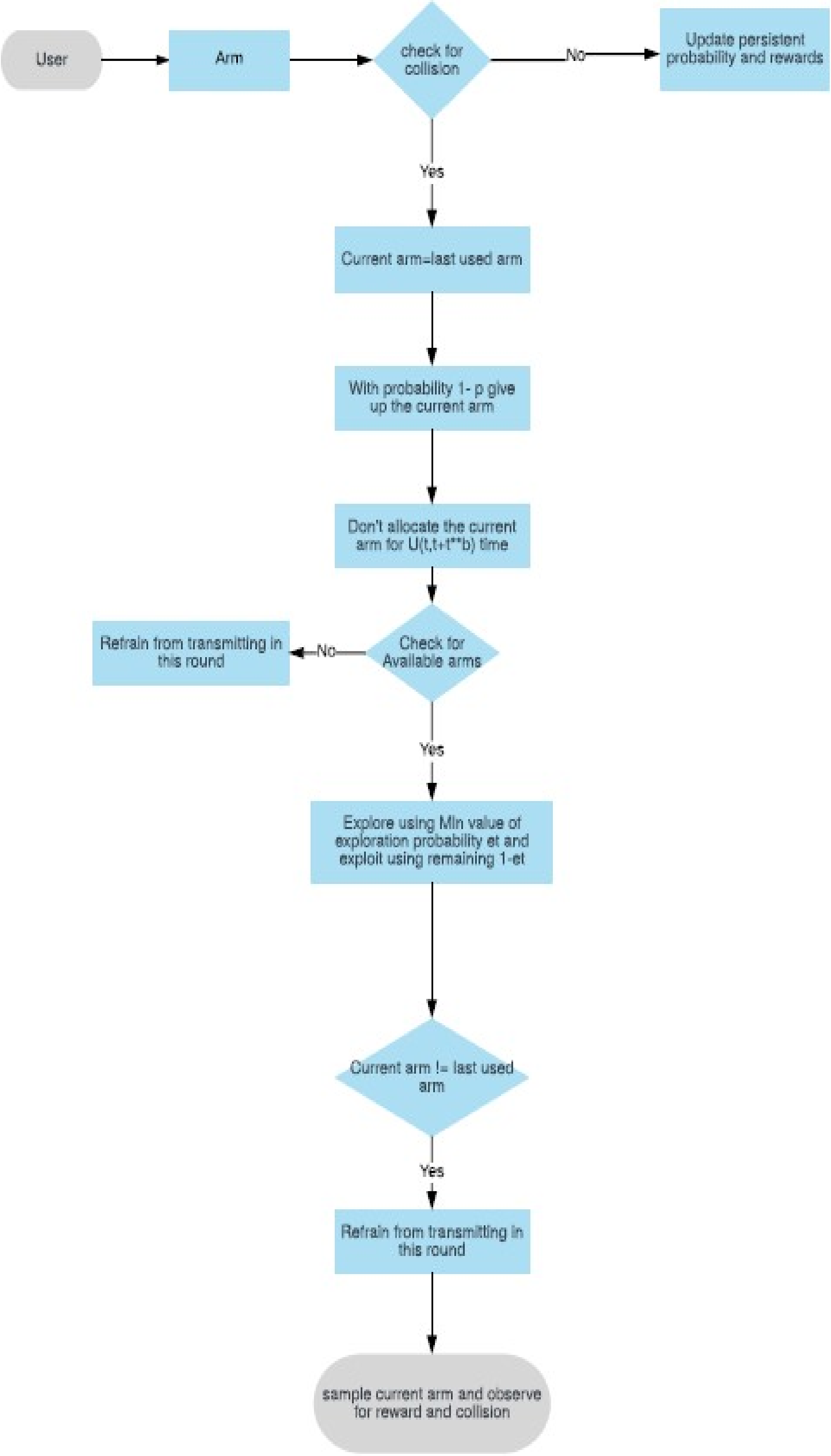
REQUIREMENT ANALYSIS

- **Software**
  - Python 3.5
  - MATLAB 9.3
  - MS Word 2007
  - MS Powerpoint 2007

- **Hardware**
  - A computer system
  - RAM-4GB (minimum)
  - Operating System – Windows 10

- **Functional Requirements**
  - Downloading and installing MATLAB 9.3.
  - Downloading and installing Python 3.6.

DETAILED DESIGN


User
Arm
check for collision
No
Update persistent probability and rewards
Yes
Current arm=last used arm
With probability 1- p give up the current arm
Don't allocate the current arm for U(t,t+t**b) time
Check for Available arms
No
Refrain from transmitting in this round
Yes
Explore using Min value of exploration probability et and exploit using remaining 1-et
Current arm != last used arm
Yes
Refrain from transmitting in this round
sample current arm and observe for reward and collision

IMPLEMENTATION

Framework:

The framework consists of two components: environment and user. The environment is a communication system that consists of K channels with initially unknown reward characteristics.
We model these channels as the arms of a stochastic Multi-Armed Bandit (MAB). We denote the expected values of the reward distributions by μ = $(\mu 1, \mu 2 \ldots\ldots \mu K)$, and assume that channel characteristics are fixed. Rewards are assumed to be bounded in the interval [0,1].
The users are selfish agents and they have no means of communicating with each other and they are not subject to any form of central control. We assume they have no knowledge of the number of users. We assume the number of users is fixed in our first section, equal to N and dynamic in the second. In both cases we assume K >=N. The situations in which K < N are supposed to be over-crowded situations. Two users or more attempting to sample the same arm at the same time will encounter a collision, resulting in a zero reward for all of them in that round. A user sampling an arm k alone at a certain time t receives a reward r (t).
The expected regret whose formula in case of single user is:

$$|E[R(t)] = \mu k * t + \sum_{\tau=1}^{t} |E[r(\tau)]$$

Where, $\mu k * = maxk \in \{1 \ldots. K\}\mu k$
In the multi-user scenario not all users can be allowed to select the optimal arm. Therefore, the number of users defines a set of optimal arms, namely the N best arms, which we denote by K*. Thus, the appropriate expected regret definition is:

$$|E[R(t)] = t \sum_{k \in K*} \mu k - \sum_{n=1}^{N} \sum_{\tau=1}^{t} |E[rn(\tau)]$$

Where, $rn(\tau)$ is the reward user n acquired at time .
Our algorithm is based on several principles which are:

- We are assuming an arm that experiences collision as an attractive arm in terms of expected reward, we want one of our users to continue sampling it.
- As we know each users need to learn the characteristics of all the arms, we would like to know that none of the arm is sampled by single user exclusively.
- To avoid frequent collisions on optimal arms, we need users to back off or skip the arms on which they have experienced collisions.
- To avoid interfering with on-going transmissions in steady state, we would like to prevent exploring users from throwing off exploiting users.

In our work, we are implementing mega algorithm to avoid collision between secondary user and to minimize the loss of reward (regret). MEGA (Multi-user e-Greedy collision Avoidance) algorithm combines an e-greedy rule and Aloha as a collision avoidance mechanism. In this algorithm learning is achieved by balancing between two terms, the exploration and exploitation through a time dependent exploration probability. In Collision avoidance mechanism, the user sampling an arm has a persistence probability that controls their determination once a collision occurs.

In this algorithm, users get randomly allocated arm but there could be many users present at any time. so, if collision detected between the users then the value of persistence probability remains fixed until collision ends. A collision event ends even if one user has given up and stop sampling the arm under dispute. If there is no collision, then value of persistence probability and reward will update.

Now, user checks for available arm from the set of uniformly available arms. If no arms are available then refrain from sampling the arm in this round. i.e., user will go to the next round. If the arms are available then user will calculate the minimum value of exploration probability and explore for the search of any other optimal arm. After this, user will exploit with the remaining exploration probability for the current optimal arm.

Now, there is another condition in which, if two arms, i.e., the current arm and the previous arm are not same then value of current persistence probability will be same as initial persistence probability. After all the procedures, we will sample the arm and check for the reward and collision. Now, we come to know that when all users apply MEGA algorithm, then the expected regret grow at a sub-linear rate means that expected regret will be minimum. Hence, MEGA is no regret algorithm, and much efficient that the other proposed algorithm.

The parameters in below pseudo code are:

p : Persistence Probability of sampling arm

p_0 : Initial Persistence Probability

t : time slot in which user sample the arm

η (t) : collision indicator (could be 0 or 1)

a(t) : current arm selected to be sampled at time t

U(1,…..,K) : Uniformly distributed set of arms

t_{nk}: time in which user will hold the arm

d: $\mu_{kN}-1-\mu_{kN}$

(difference between rewards of best arm and second best arm (fixed value=0.05))

$c = 0.1$, $p_0 = 0.6$, $\alpha = 0.5$, $\beta = 0.8$

Pseudo Code

```
Init: $p=p_0$, t=1, ɳ(0)=0, a(0)~U({1,.....,K}), $t_{nk}=1$
    assign : a(1,....K)=0
    //this shows that arms are available
    while(t) :
        if(ɳ (t-1)==1) then :
            a(t)=a(t-1)
            $p=p_0$
        else
            $p=p.\alpha+(1-\alpha)$
            update $\mu_{\alpha\ (t-1)}$
        end if
    //Now check for Available arms...
        A={k : $t_{nk}$<=t}
        if(A==0) then :
            Refrain from transmitting in this round
        assign : a(1,...K)=1
        //this shows no arm is available for this slot
        end if
        $\varepsilon_t = \min\{1, \frac{cK2}{d2(K-1)t}\}$
        explore using et :
            a(t)~U(A)
        exploit : with $1-\varepsilon_t$
            a(t)~a(t-1)
        if(a(t)!=a(t-1)) :
            $p=p_0$
        end if
        sample arm a(t) and observe for r(t), n(t)
    end loop
```

Limitations

- In MEGA, whenever there is collision all users but one have ‘given up’, again when the loop starts there remains some refrained users which are not sampled till the end. Thus these refrained users decreases the total expected reward.

- The collisions between the users may not be reduced to absolute zero, there may be partial collisions between the users.

- Different users may receive different rewards.

- It may be possible that the all the users may not receive the same amount of rewards, as in our algorithm the users are not able to communicate with each other.

- Another limitation is that the explicit collision is not always available in practice meanings of this is there might be some partial collisions.

CONCLUSION

Our proposed algorithm, a combination of an e-greedy policy with an availability detection mechanism, which exhibits good experimental results for both fixed and dynamic numbers of users in the network. We conclude these results with a theoretical analysis guaranteeing sub-linear regret. One another challenge that came in light i.e. different users has different expected rewards while sampling the same arm. Our proposed algorithm does not involve the communication between users, so they are not able to share which channel give the maximum expected reward therefore resulting in fewer number of collisions .

GANTT CHART

TASK	PERSON RESPONSIBLE	FROM DATE	TO DATE
Research Papers	All Members	26 JANUARY 2018	-
Synopsis	All Members	9 FEBURARY 2018	15 FEBURARY 2018
Pseudo Code	All Members	10 MARCH 2018	15 MARCH 2018
SRS/Report	All Members	16 MARCH 2018	21 MARCH 2018

REFERENCES

- O. Avner, S. Mannor Concurrent bandits and cognitive radio networks. In 29th International Conference on Machine Learning, 22 April 2012.

- O. Avner, S. Mannor, and O. Shamir exploration and exploitation in multi-armed bandits. In 29th International Conference on Machine Learning, December 2012.

- D. Kalathil, N. Nayyar, and R. Jain Referential learning for multi-player multi-armed bandits. In 51st IEEE Conference on Decision and Control, pages 3960{3965, 2012.

- Decision-Theoretic Distributed Channel Selection for Opportunistic Spectrum Access: Strategies, Challenges and Solutions, Yuhua Xu, Student Member, IEEE, Alagan Anpalagan, Senior Member, IEEE, Qihui Wu, Senior Member, IEEE, Liang Shen, Zhan Gao, and Jinglong Wang, Senior Member, IEEE