RESEARCH ARTICLE OPEN ACCESS

Scheduling Multiple Secondary Users in Cognitive Radio: A Deep Reinforcement Learning Approach

P Ravinder Kumar[1], Dr Sandeep V M[2], Dr Subhash S Kulkarni[3]
[1]*(Research Scholar, ECE Department, JNTUH Hyderabad)*
[2]*(Professor and HOD, CSE Department, Jayaprakash Narayan College of Engineering, Mahbubnagar, India)*
[3]*(Principal and Professor, PESIT – Bangalore South Campus, Bengaluru, India)*

ABSTRACT
The Central Unit (CU) schedules multiple Secondary Users (SU) through Reinforcement Learning (RL) using various reward/punishment modes in a single channel Cognitive Radio (CR) system, among which compound reward/punishment mode is making all SUs more obedient in using the channel and achieving higher performance.

***Keywords* -** Cognitive Radio, SU Obedience Level, PU behaviour, SU behaviour, Reinforcement Learning

Date of Submission: 01-01-2023 Date of Acceptance: 11-01-2023

I. INTRODUCTION

Spectrum scarcity serves as a reminder to employ cognitive radio technology (CR). The Secondary User (SU) in cognitive radio technology must use the channel for transmission without interfering with the primary user (PU). This requires sensing which is an overhead. In single SU situation, it senses the channel and uses it without interfering with the PU. However, when multiple SUs competes for a single channel, the simple sensing technique is not sufficient as multiple SUs sense free channel and start using that results in collision. Collision is avoided only when single SU is allowed at a time. Sensing requires additional energy and resources and making every SU to sense in spite of knowing that the only one SU gets the chance to use the channel makes the model highly inefficient. This necessitates to create a Central Unit (CU) for sensing and allowing one SU to use the channel. The SU with the most resources may be assigned to perform as the Central Unit. This relieves all SUs of sensing and allows for more efficient resource utilisation. The CU determines the primary behaviour, its duration, and transition changes. This work has been presented in previous papers. CU determines who can use the channel and who should be assigned to it. This research aims at identifying the best SU among all the SU requests. The rest of the paper is organised as follows: The next section contains background information on CR and Reinforcement Learning. Section 3 addresses how to ascertain SU Scheduling based on their behaviour and achieve high efficiency. Findings and discussion back up for our arguments in section 4, and Section 5 concludes the paper.

II. LITERATURE REVIEW

The Spectrum Cognitive radios have received much attention recently as a proposed solution to the spectrum scarcity problem identified by the Federal Communication Commission (FCC). The problem being that, although there is a shortage of available frequency bands to license out, the current licensed bands are severely underused in both a time and space sense. Mitola proposed that cognitive radios solve this problem by sensing the environment and autonomously adapting to take advantage of the underused spectrum, while staying clear of the incumbent user’s signals [1]. In [2], the authors investigate optimization for the cooperative spectrum sensing with an improved energy detector to minimize the total error rate (sum of the probability of false alarm and miss detection). Follow-up works extend the scenario to imperfect reporting channels and SUs with multiple antennas [3],[4] respectively. Machine Learning (ML) classification algorithms and feature extraction [5], ML-based classifiers include k-nearest neighbour (KNN) [6], Support Vector Machine (SVM), Decision Tree (DT) and Naive Bayesian (NB) [7]. Many Spectrum Modulation Indicator (SMI) methods can be developed by combining different traditional feature extraction techniques are difficult to extract inherent features of different modulations because they are based on statistics [9]. Reinforcement learning [10] is one of the most prominent machine learning research lines that has had a substantial impact on the development of Artificial Intelligence (AI) during the last 20 years [11]. It is a learning process where in an agent is allowed to make decisions on a regular basis, monitor the outcomes, and then automatically alter its strategy to attain the best policy. Deep Reinforcement Learning (DRL) has recently received a lot of attention and success in the field of wireless communications [12]. The DRL trains the learning process using Deep Neural Networks (DNNs) [13], which increases the learning speed and performance of reinforcement learning algorithms. The performance of the secondary user is affected by the primary user's data traffic characteristics, and that when the primary user's arrival rate is high, the mean waiting and average queueing length of the secondary user increase, especially when the combined arrival rate approaches the queue utilisation factor [14]. [15] proposed channel access technique as a knapsack problem in order to maximise the sum of data frame priority under the constraint of restricted transmission time. To address the dilemma in priority scheduling, the priorities of data frames in the queue were dynamically modified based on the wait time at the head of the queues. In [16], for multi-user and multi-channel cognitive radio systems, a channel selection mechanism without negotiation is being studied. To avoid collisions caused by lack of coordination, each user secondary learns how to select channels based on their prior experience. In the scope of Q-learning, multi-agent reinforcement learning (MARL) is used by including the opponent secondary users as part of the environment. In this paper, all SUs must have the natures of all SUs. The RL is required to observe other SUs for each SU. This results in resource wastage. Instead of all SUs, one SU is designated as the Central Unit (CU), and the behaviour of all SUs is examined in this paper. In [17], the Distributed Mutual Exclusion method is used to allocate a single PU, several SUs, and an associated channel. The method's disadvantage is the latency in channel access and the burden for SUs in retaining request information. The SUs' ineffective use of the channel results in low throughput. The resources are needed more for making judgement and keeping a queue with time stamps. The SUs should have the same request queue information (some SUs may not receive the request of other SU). These concerns can be resolved by designating one SU as the CU. This CU will handle any SU requests and assign the channel accordingly. In this research, CU is used to tackle this problem.

III. MODEL

To When an SU requests a channel in the CR system, CU checks the channel status and provides the information to the SU. The information contains the behaviour of the PU at that time and how long this behaviour is expected to continue. When SU is permitted to use a channel, SU must synchronise itself with informed PU behaviour and its duration. This results in minimal interference, minimal transmission loss, and maximum throughput. When multiple SUs request for the channel but only one can be allowed, then it is advisable to allocate the SU that guarantee minimum interference to PU and maximum throughput. Allotted SU must adhere to this behaviour for best performance. Any deviation reduces the CR performance. The quality of adherence of SU to the PU behaviour is measure of its obedience. In attempting to improve its throughput the SU may deviate from adhering to the PU behaviour. This can be classified in three. categorises i. Synchronisation error: Minor deviation in synchronisation of SU with PU. This creates interference to PU but small in quantity. ii. Overtime usage: SU continues to work even after the allotted time. The time allocated usually refers to particular behaviour of PU and after that period PU behaviour will change this is unaware to SU hence more interference to PU. iii. Unauthorized usage: SU using the channel without authorisation. SU requests indicates that some other SU with higher obedience is allotted to the channel. Usage of channel in this case adds interference to both PU and SU officially permitted to use the channel. To deter the SU taking these deviations some sort of reward/punishment technique is to be implemented. If SU synchronises exactly with PU then it must be rewarded and punished if it deviates. So, the deviation of first type creates smaller interference to PU and hence attracts the smaller punishment. The overtime deviation creates more interference to PU and hence attracts higher punishment. The third case of deviation the unauthorised SU is disturbing both PU and authorised SU hence highest punishment. The obedience level of SU is the measure of its past performance in CR system. It can be calculated as cumulative function of reward/punishment through its past actions. The CU uses the obedience levels of SU competing for the channel any time and allots the channel to the best SU. To start, all SUs behaviours are unknown and hence their obedience. Knowing the behaviour is a runtime process which calls for Reinforcement Learning (RL). Thompson method of RL ranks all the SUs permanently and hence it is a static approach. This does not change the situation even if the obedience levels of SUs are modified with time. It is quite possible that disobedient SU, due to multiple rejections, may want to improve its obedience level by showing better behaviour. So the learning process must be dynamic. The Upper Confidence Bound (UCB) method qualifies for this approach. Initially, In UCB method, all SUs have the same level of obedience, and the behaviour is observed whenever the secondary is assigned the channel. SU is rewarded or punished depending on how well the information provide by CU is followed. The obedience level is updated accordingly. The experiment is carried out with a single PU and multiple SUs competing for the channel. The CU allocates the channel to the SU with highest obedience level. The obedience level of SUs are continuously updated through UCB on real time basis. The SUs with lower obedience levels will never get opportunity to use the channel. With no competition a disobedient SU may get the channel providing a channel opportunity to improve its behaviour. and uplift its obedience level. This provides opportunity for misbehaving allowing other SUs get opportunities in future. On the other hand, if the SU with the highest obedience level starts misbehaving then its obedience level falls and other SU may get opportunity in future. Each SU keeps track of its usage opportunities and always striving to improve its behaviour.

IV. RESULT AND DISCUSSION

The experiment is conducted to study the effect of various types of reward and punishments given to the SU for its obedience.

A. Without Rewards or Punishment

In this case, the past behaviour of SUs is not considered while allocating the channel. The allocation may be first in first out or random. If the channel is assigned without knowing their behaviour, there is a chance that they will deviate from the rules, causing more interference to the PU. The plot for this case in figure Fig. 2 shows no improvement in the interference to the primary. As a result, some sort of incentive or punishment is required to get them to follow the rules. From figure Fig. 3 it is evident that the efficiency with this scheme at its worst.

B. Simple Reward, Simple Punishment

In this case, the behaviour of SU is binary, either obedient or disobedient. Here, all three types of disobedience are equally punished. To discourage this, the deviations must find different punishments. This type of reward/punishment model may encourage to make larger deviation instead of smaller one. In this case, (figure Fig. 2) it is observed that the performance is improving with time. But this improvement is slower. Larger disturbance is found as deviation types are not differentiated while punishing. An improvement is observed in the efficiency (figure Fig. 3).

C. Simple Reward, Simple Variable Punishment

Here the SU is punished according to its level of deviation, this makes the obedience level of the SU to drop in accordance with level of deviation. This deters, CU allotting the channel to highly misbehaving SUs, making more and more chances for obedient SUs to be allotted. This further improves CR efficiency. This faster and smoother improvement is clearly visible in figure Fig. 2 and figure Fig. 3

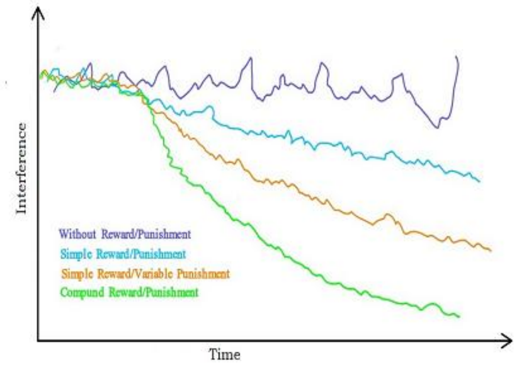

Fig.2. Interference measure for various rewarding schemes

D. Compounding Reward, Compounding Punishment

To encourage sincere SUs the reward may be compounded for showing successive obedience. Similarly regular misbehaving SUs be compound punished. This will speed-up the process of CR system to reach stable and efficient state. Figure Fig. 2 justifies that the compound effect of reward(punishment) will highly encourage (discourage) the obedient (disobedient) SUs, forcing each SU to be more and more obedient with time. It is also evident from the figure.2 that earlier stability is reached more smoothly and higher efficiency (figure.3) is confirmed than other schemes.

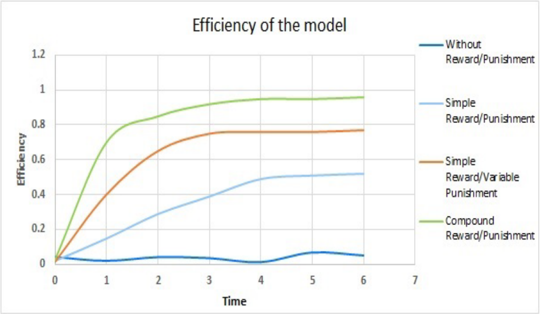

Fig. 3. Efficiency for various rewarding schemes

V. CONCLUSION

This paper presented, a Reinforcement Learning model that used to schedule the multiple SUs, encouraging higher obedient SUs in obtaining the channel and the lower obedience SU are discouraged at the same time. This forces each SU to improve its obedience level. In the long run, all SUs are found to increase their obedience hence best performance in terms of efficiency, throughput and interference. It is also observed that the compound reward/punishment attains faster steady state. Hence, CR system with multiple SUs and single channel are advised to have one CU and use compound reward/punishment mode of operation. However, CU is overburdened limiting its own performance. Knowing CU in advance makes/forces CU to be available all the time and hence must be a static SU. The research can be extended to accommodate multiple PUs and SUs. Our work may also be extended to dynamic CU and some mechanism to transfer the duties of the CU from one SU to other SU. The burden of CU may further be reduced through usage of multiple CUs.

Acknowledgements

The paper concepts, experimentation, validation, formal analysis, paper research, and original draft preparation of writing were developed by the first author. The project was supervised and managed by the second and third author.

REFERENCES

[1]. J. Mitola, G.Q. Maguire, Cognitive radio: Making software radios more personal, IEEE Personal Communications 6 (4) (1999) 13–18. doi: 10.1109/98.788210.

[2]. A. Bhowmick, K. Yadav, S.D. Roy, S. Kundu, Throughput of an Energy Harvesting Cognitive Radio Network Based on Prediction of Primary User, IEEE Transactions on Vehicular Technology 66 (9) (2017) 8119–8128. doi: 10.1109/TVT.2017.2690675.

[3]. K. Zhang, Y. Mao, S. Leng, H. Bogucka, S. Gjessing, Y. Zhang, Cooperation for optimal demand response in cognitive radio enabled smart grid, 2016 IEEE International Conference on Communications, ICC 2016 (2016) 0–5doi: 10.1109/ICC.2016.7511433.

[4]. Z.Wang, S. Salous, Time series arima model of spectrum occupancy for cognitive radio, IET Seminar Digest 2008 (12338) (2008) 2–5. doi: 10.1049/ic:20080405.

[5]. S. Zhang, T. Wu, M. Pan, C. Zhang, Y. Yu, A-SARSA: A Predictive Container Auto-Scaling Algorithm Based on Reinforcement Learning, in: Proceedings - 2020 IEEE 13th International Conference on Web Services, ICWS 2020, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 489–497. doi: 10.1109/ICWS49710.2020.00072.

[6]. T.J. O'Shea, J. Corgan, T.C. Clancy, Convolutional radio modulation recognition networks, Communications in Computer and Information Science 629 (2016) 213–226.

arXiv:1602.04105, doi: 10.1007/978-3-319-44188-716.

[7]. C.S. Park, J.H. Choi, S.P. Nah, W. Jang, D.Y. Kim, Automatic modulation recognition of digital signals using wavelet features and SVM, International Conference on Advanced Communication Technology, ICACT 1 (1) (2008) 387–390. doi: 10.1109/ICACT.2008.4493784.

[8]. Z. Huang, C. Li, Q. Lv, R. Su, K. Zhou, Automatic Recognition of Communication Signal Modulation Based on the Multiple-Parallel Complex Convolutional Neural Network, Wireless Communications and Mobile Computing 2021. doi: 10.1155/2021/5006248.

[9]. H.A. Shah, I. Koo, Reliable machine learning based spectrum sensing in cognitive radio networks, Wireless Communications and Mobile Computing 2018. doi: 10.1155/2018/5906097.

[10]. X. Tan, L. Zhou, H. Wang, Y. Sun, H. Zhao, B.C. Seet, J. Wei, V.C.M. Leung, Cooperative Multi-Agent Reinforcement Learning Based Distributed Dynamic Spectrum Access in Cognitive Radio Networks (2021) 1–28arXiv:2106.09274, doi: 10.1109/jiot.2022.3168296. URL: http://arxiv.org/abs/2106.09274.

[11]. F. Muteba, K. Djouani, T. Olwal, A comparative survey study on LPWA IoT technologies: Design, considerations, challenges and solutions, Procedia Computer Science 155 (2019) 636–641. doi: 10.1016/j.procs.2019.08.090. URL: https://doi.org/10.1016/j.procs. 2019.08.090.

[12]. G. Yunxin, Z. Yue, M. Hong, Modulation Recognition of Digital Signals Based on Deep Belief Network, in: IOP Conference Series: Materials Science and Engineering, vol. 563, 2019. doi: 10.1088/1757-899X/563/5/052009.

[13]. M. Ul Hassan, M.H. Rehmani, M. Rehan, J. Chen, Differential Privacy in Cognitive Radio Networks: A Comprehensive Survey, Cognitive Computation 14 (2) (2022) 475–510. arXiv:2111.02011, doi: 10.1007/s12559-021-09969-9.

[14]. I. Suliman, J. Lehtomaki, Queueing analysis of opportunistic access in cognitive radios, in: 2009 Second International Workshop on Cognitive Radio and Advanced Spectrum Management, 2009, pp. 153–157. doi: 10.1109/COGART.2009.5167252.

[15]. B. Choi, H. Lim, H. Kang, B.J. Jeong, Dynamic priority scheduling for heterogeneous cognitive radio networks, in: 2012 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2012, pp. 62–64. doi: 10.1109/SECON.2012.6275837.

[16]. H. Li, Multi-agent q-learning of channel selection in multi-user cognitive radio systems: A two by two case, in: 2009 IEEE International Conference on Systems, Man and Cybernetics, 2009, pp. 1893–1898. doi: 10.1109/ICSMC.2009.5346172.

[17]. M. Hosen, M. Mridha, M. Hamza, Secondary user channel selection in cognitive radio network using adaptive method, 2018, pp. 1–6. doi: 10.1109/I2CT.2018.8529521.