

# Non-Cooperative Learning Based Routing for 6G-IoT Cognitive Radio Network

Tauqeer Safdar Malik<sup>1,\*</sup>, Kaleem Razzaq Malik<sup>1</sup>, Muhammad Sanaullah<sup>2</sup>, Mohd Hilmi Hasan<sup>3</sup> and Norshakirah Aziz<sup>3</sup>

<sup>1</sup>Air University Multan Campus, Department of Computer Science, Multan, 60000, Pakistan

<sup>2</sup>Bahauddin Zakariya University, Department of Computer Science, Multan, 60000, Pakistan

<sup>3</sup>Centre for Research in Data Science, Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Seri Iskandar, 32610, Perak, Malaysia

\*Corresponding Author: Tauqeer Safdar Malik. Email: tauqeer.safdar@aumc.edu.pk

Received: 24 June 2021; Accepted: 05 August 2021

**Abstract:** Cognitive Radio Network (CRN) has turn up to solve the issue of spectrum congestion occurred due to the wide spread usage of wireless applications for 6G based Internet of Things (IoT) network. The Secondary Users (SUs) are allowed to access dynamically the frequency channels owned by the Primary Users (PUs). In this paper, we focus the matter of contention of routing in multi hops setup by the SUs for a known destination in the presence of PUs. The traffic model for routing is generated on the basis of Poison Process of Markov Model. Every SU requires to reduce the end-to-end delay and packet loss of its transmission simultaneously to improve the data rate for the Quality of Service (QoS) of the Secondary Users. The issue of routing is formulated as stochastic learning process of non-cooperative games for the transformation of routing decisions of SUs. We propose a distributed non-cooperated reinforcement learning based solution for solving the issue of dynamic routing that can avert user interferences and channel interferences between the competing Sus in 6G-IoT network. The proposed solution combines and simulate the results to show the effectiveness and working of the proposed solution in decreasing the end-to-end delay, packet loss while meeting the average data rate requirement of QoS for SUs.

**Keywords:** Cognitive radio network; 6G; IoT; quality of service; routing protocols

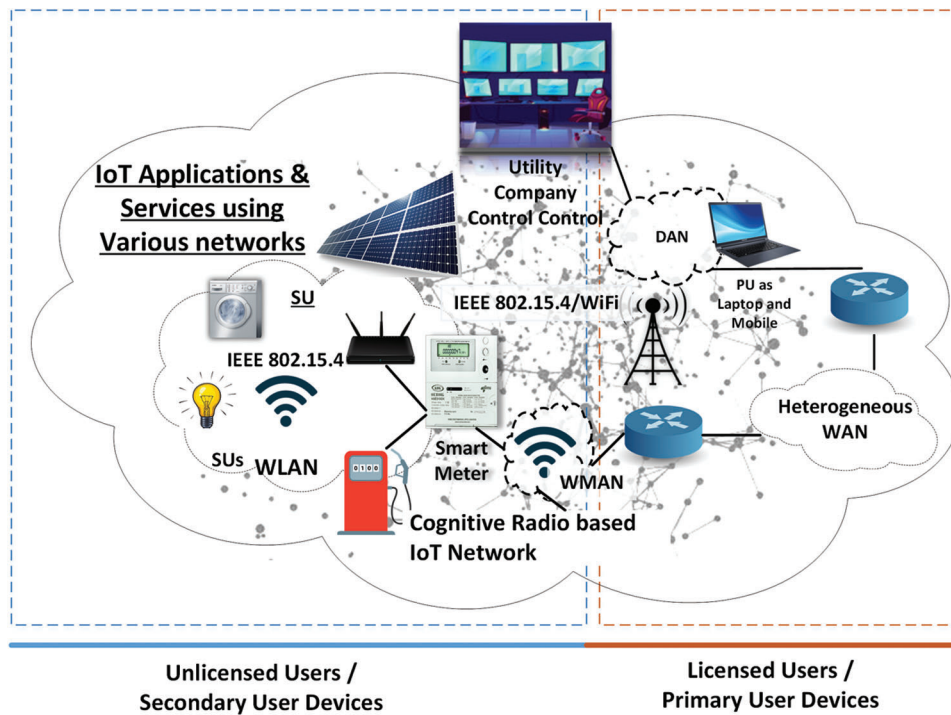
## 1 Introduction

The new archetype of wireless communication, the 6th generation (6G) model with the applicability of artificial intelligence is expected to start in near future. The applicability of 6G demands the stable connectivity of various devices in dynamic environment of Internet of Things (IoT) as shown in Fig. 1. Artificial Intelligence (AI) and Cognitive Radio (CR) will be desegregated into the 6G communication systems for Internet of Things (IoT). All the network management, instruments, resources, services, signal processing, and so on will be incorporated by using AI [1]. The Industry 4.0 is the digital transformation of industrial manufacturing which will be the driving force of such revolution [2]. Fig. 1 represents the 6G based communication architecture scenario toward visualizing the CR enabled 6G-IoT



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

communication systems. There are many important applications and possibilities of 6G-IoT wireless communication as shown in Fig. 1. The marvelous characteristics of 6G will advance the creation of smart cities leading to life quality improvements, environmental monitoring, and automation using CR-based communication technologies [2]. The 6G wireless connectivity will make our society super smart through the use of IoT devices and applications under the heterogeneity of various wide area network (WANs). In addition, many cities in the world will deploy flying taxis based on 6G wireless technology. Hence, this need will create an essential desire to utilize the unutilized spectrum in high processing mode. The right solution for such issues is CR enabled communications which can provide the uninterrupted connectivity for such IoT devices. The differentiation in licensed and unlicensed users' devices can be handled with effectively and without any connectivity issues as shown in Fig. 1. Eventually, the dream of smart homes will become a reality because any device in a remote location will be controlled by using a command given from a smart device. Internet of Things (IoT) is the undoubted combination of smart cities to coordination among a very large number of computing elements and sensors, objects or devices, people, processes, and data using the internet infrastructure [3]. The CR enabled 6G system will provide full IoT support to every type of devices which require any data rate constraints or latency issues. There are various challenges which needs to be addressed beyond 5G such as higher data rate, lower latency, and improved overall Quality of Service (QoS). These devices will require the different communication resources such as spectrum and channel without any interferences in 6G based IoT. Moreover, with the new era of wireless devices and services, the spectrum scarcity problem is solved by the help of Cognitive Radio (CR) [4]. Therefore, today's 6G networks based on CRs encounter issues on the efficient management of this increasing complexity of users and data rate for the increasing demand of QoS for IoT. Cognitive Radio Network (CRN) is likely to transfuse into every range of wireless networks [5].



**Figure 1:** Future CR enabled 6G-IoT communication architecture scenario

Although the MAC layer protocols of CRN have been studied for last decade and routing is still the most challenging issue to be solved for their use in new communication paradigms. The routing protocols intents

to provide the route to destination, the channels for the selected route and assurance that the routes do not interfere with other users in 6G-IoT [6]. Therefore, many researchers come up with many routing protocols solutions that suits these aims [7–9].

To utilize these routing protocols in ad-hoc networks environment with Cognitive Radio support for 6G-IoT, Secondary Users (SUs) encounter the challenges such as user interferences, channel interferences, unstable user and lack of required QoS. Therefore, the routing of SUs in multi-hop manner is an important issue to be solved for the improvement of QoS. In some recent works, there are some routing protocols are presented by various researchers [10]. One of them is presented the solution based on the utilization of relaying SUs in case when the source and destination is unable to have a common channel for the communication due to opportunistic spectrum access. In another solution the improvement of QoS at the destination is given when the source encounters the interference from the Primary Users (PUs) [11]. However, there are many other issues still need an attention to design a routing protocol for CR enabled 6G-IoT network. One of the issues is an adaptation of dynamic environment changes during the routing in SUs, when the activities of PU and arrival traffic of other SUs sensed on the channel. To implement the distributed nature of CRN, non-cooperated solutions for routing of SUs must be presented for 6G based IoT. The non-cooperative behavior of a SU must be handled to manage the dynamic routing for multi-hop environment. The existing routing protocols consider a specific interference range in SUs to avoid the user interference. This limitation is useful to avoid the interference with the PUs only and also not fulfill the regulations of Federal Communication Commission (FCC) [12]. To implement such assumptions in routing protocols, lemmatize the possible communication opportunities, depending on limited routes due to the interference range of PUs and to construct the routes where the activities of PUs are very limited. Therefore, in the perspective of non-cooperative routing solutions, the routing decisions must be exchanged between the other SUs which can degrade the network performance and required QoS. Thus, the issue of non-cooperation between the SUs is another main challenge which must be addressed for the routing solutions [13,14].

In this paper, we take a step to answer these questions by applying the non-cooperative reinforcement learning technique to formalize the routing strategy for the CR enabled 6G-IoT network. For this purpose, a new algorithm is proposed by modifying the reconfiguration process in the existing Ad-hoc on-Demand Distance Vector Routing protocol (AODV) routing protocol [15]. This is based on one of the reinforcements learning techniques: namely Q-Learning. The routing scenario is implemented as a non-cooperative game theory model which is very efficient for the stochastic environment model. The dynamic nature of CRN is modelled as a stochastic way to formalize the state-of-art implementation of routing problem.

The rest of the paper is organized as follows: Section 2 presents a brief survey of recent work done in such routing solutions. Section 3 gives an overview of the proposed methodology to implement the reinforcement learning technique for autonomic re-configuration of routing decisions. Section 4 evaluates the proposed algorithm in terms of mathematical modelling according to the implemented scenario of routing strategy. Section 5 shows the obtained results through the simulations and discusses the application of proposed algorithm. Finally, Section 6 concludes the paper and projects the future research directions.

## 2 Related Work

The related work is categorized the literature on the basis of framework of routing solutions and explains very briefly each of these categories with its drawbacks and challenges for routing in the CR enabled future 6G-IoT Network. There are various authors presented [7,16,17] the routing solutions which are based on the extending and adding some new routing parameters of AODV routing protocol, so that they can be useful for

routing in CRN. In this perspective, a routing protocol named SEARCH is developed on the basis of geographic routing by avoiding the region of PU's activity during route formation [7]. The routing solution for multi-hop CRN is proposed in [18] based on the capacity of user to assign the channel for the high mobility of CR for IoT network. A routing scheme for resource-aware is developed in [19] to enhance the routing on the basis of energy usage of a user. In [20] some new metrics added to measure the connectivity and stability of different paths for the distributed environment of CRN and named as Gymkhana. The cognition-based routing strategy is used for the prediction of routing decisions to control the topology and routing issues in [21]. A cross layer approach is used to implement the routing strategy for dynamic CRN and for 5G based IoT and future networks [22]. First time the channel number is also used as routing metrics with other metrics during route formation in as RREQ message broadcast over the network to get the channel information but overloaded as RREP messages [23]. The channel aggregation-based routing protocol is developed with the addition of channel threshold value parameter to enhance the energy efficiency, throughput and end-to-end delay of the network [24]. A routing algorithm to minimize the delay from the source SU to the common destination of SU is presented in [9] but it lacks to handle the activity of PU. To minimize the interference with PU a reactive routing protocol is proposed in [25] for the routing of CR enabled IoT for future networks as beyond 5G. The channel interference is captured using a new routing metric in [26]. The user interference with channel interference is addressed jointly in [27]. All of the above discussed solutions do not come up with the best routing strategy at every stage for the SU in the stochastic nature of CR enabled 6G-IoT network to address the challenges of higher data rate as well as the latency issues. These solutions are not able to adapt the quick changes during routing on the basis of environment changes.

On the other hand, there are some routing solutions implemented on the basis of reinforcement learning algorithms to solve the routing issues encountered in dynamic nature of CR enabled IoT networks [28,29]. Initially, one of the reinforcements learning technique, Q-learning is used for the spectrum aware routing of CR enabled IoT in [30]. This work is based on the assumption that a route which has more available channels can reduce the negotiating time with MAC layer for the channel selection during transmission. Thus, the number of available channels is calculated using the Q function of Q-learning. On the basis of this work, another routing solution is presented in [31] to reduce the interference between SU and PU for the improvement of network performance for IoT based CRN. In [24], the reinforcement learning is used for opportunistic routing solution based on the existing status of a user and the probability of transmission success. This opportunistic routing implementation is lead to introduce the spectrum-aware cluster-based routing in which the reinforcement learning technique is used to choose the stable path without reduction of throughput and end-to-end delay in SU [32]. In all of these routing solutions have not considered the multi-hop concept of CRN in designing and implementing the routing protocol. Hence, the dynamic routing on the basis of game theory is developed for multi-hop CRN in [33]. This is the first time when the time-varying property of CRN is implemented based on the non-cooperative game theory model. In this routing protocol, authors are tried to reduce the overall interference of the PU as well as the average packet delay for a predefined route in multi-hop environment of CR enabled IoT network.

The routing solutions presented in [15–25] and [26–33] do not consider the concept of multi-agent in taking the routing decisions and used a single agent assumption. Also, the concept of non-cooperative game theory is applied only in [28] and [29] for the routing decisions of SUs. The routing technique used in [29] needs to get the routing decisions of all SUs by sharing the information during routing. This information exchange between SUs can increase the routing overhead as well as fail to comply to implement the true nature of non-cooperative game theory for routing solution. The routing solution presented in [31], also needs to flood the messages in case of any breakdown happened to existing channel selection during the transmission due to the activities of PUs. The activity model of PU which is

used in mostly routing solution on the basis of ON/OFF model [30] is also inappropriate for the data traffic [31].

The challenges for effective routing in CR enabled IoT network for future perspective of 6G demand that every SU needs to choose the best path to reduce the end-to-end delay and packet loss so that the QoS requirement of the network met in terms of higher data rates and lower the latency as much as possible. The proposed routing solution based on non-cooperative learning (NCL) should have the capability to reduce the routing overhead in case of channel information exchanging between SUs and non-cooperative nature of SUs. The routing parameters also need to adapt themselves on the basis of environment changes and channel information. Further, the routing solution must be based on the distribution resource management so that the routing decisions must be based only on local spectrum information. The activity of PU also be appropriately modelled and included in designing and evaluating the performance of dynamic routing solution for CR enabled 6G-IoT network.

### 3 Methodology

The learning ability in a radio is desired to enable it to work as cognitive radio [32]. The learning is necessary for receiving specific environmental inputs so that it can predict the future behavior of working in our case routing. The behavior of the system alters overtime so that the performance of the system/network can be improve with the interaction of its environment. Hence, the network can adopt different strategy and respond in different way for the same input overtime depend on its other parameters. The machine learning field emphasis on such kind of issues, properties and performance of environment learning algorithms.

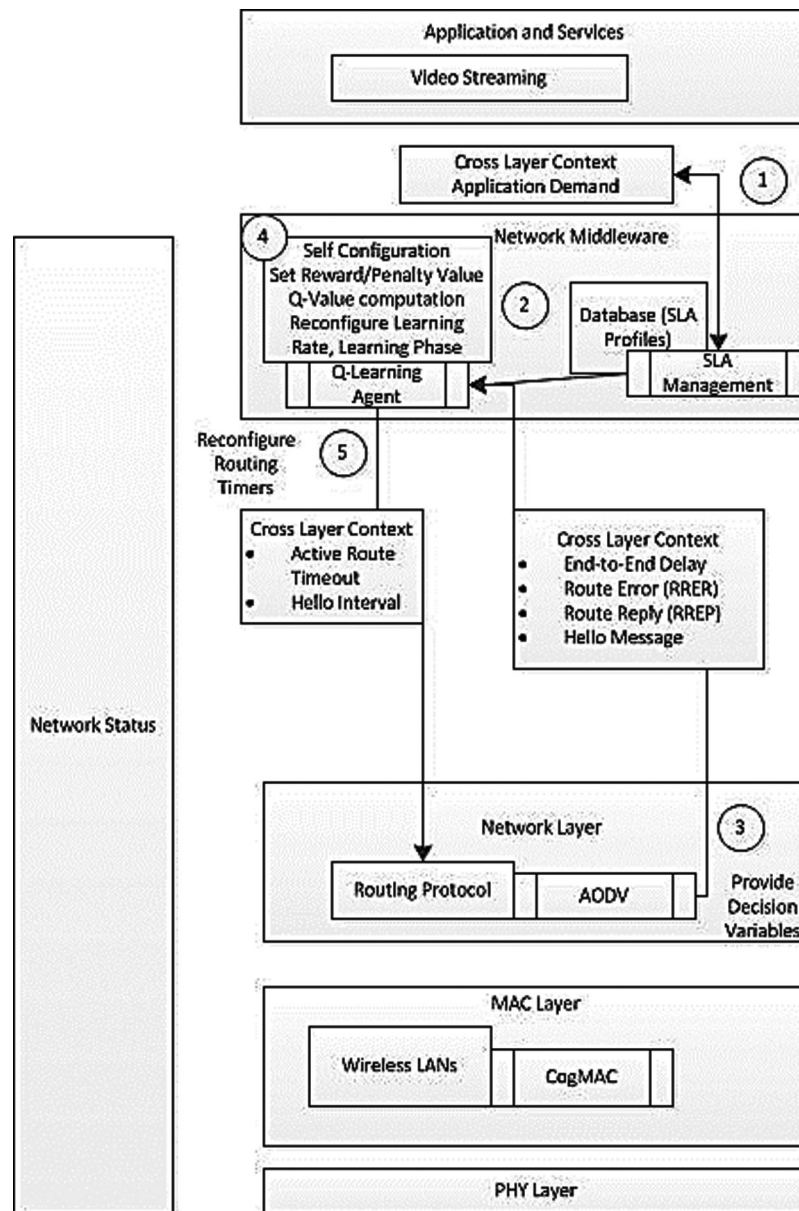
Machine learning techniques can be classified into three broad categories based on the cooperative and non-cooperative models of learning methods. These types are named as: (i) supervised learning, (ii) unsupervised learning, and (iii) reinforcement learning [34]. Briefly, supervised and unsupervised learning techniques are solely based on the cooperative model of future decisions through its environment. The learning is performed through training phase which the learner tries to coordinate to reply correctly according to the inputs it has not seen yet. Unsupervised learning is the other extreme cooperation of participants in which the learner receives no evaluation from the environment at all. The task of learner is to cooperate and classify the inputs in groups, categories, or with reduced set of attributes. A third alternative, is reinforcement learning (RL) in which the learner is getting the appropriate response through non-cooperative and the appropriateness of the response is checked through a reward (or reinforcement). RL, therefore, depends more on exploration from non-cooperative and trial-and-error model between the participants [35].

Routing is the way of making decisions for path selection of a node. This process of decisions for future route is improved through the previous trial and error path. Hence, the applicability of RL technique in case of routing, will be much helpful in CRN. The non-cooperative environment of CRN suits the implementation of RL technique to improve the link quality so that the overall data rate is maximized by minimizing the end-to-end delay and packet loss.

Q-learning is well known RL based learning technique that is a model-free and famous value-iteration learning technique. It is formulated in 1992 by Watkins [36] with limited and non-cooperative computational requirement. It explores and make enable the participating users to learn, plan and act in optimal way in Markovian situation to control the future decisions. Q-learning is based on a model free and hence, does not precisely model the probabilities of reward according to the underlying process. Q-learning initializes by exploring the opportunities for future decisions rather just guessing the value of an action as temporal-difference (TD) learning. The value of an action is compiled over experienced outcomes using an idea known to the system and every user take its decisions in non-cooperative manners. Every player starts its



action selection through incrementing its values of reward that we can say as the Q-values. These improvements incorporate the quality of particular actions at particular states of every user according to the opponents. This Q-value is evaluated through the action-value pair which is calculated and maintained by learning the Q-function. It gives the probability of taking particular action in a given state and using the optimal policy depending on the available choices of every user. The implementation of Q-learning to predict the future routing decisions is done through the cross-layer architecture of network model, so that the availability of path and spectrum is coordinated from MAC layer to the Network layer. This process of cross-layer architecture for our non-cooperative learning based cognitive management framework is having this ability and works in below steps as shown in Fig. 2.



**Figure 2:** Cross-layer design for Non-Cooperative Learning (NCL) routing of a node

Step 1: The demanded data rate for the specific application is maintained at the network middleware layer through the SLA management module.

Step 2: The transmission requirement in the form of Q-value calculated through the reward and action pair by Q-learning agent implemented in the network middleware layer.

Step 3: Using the RREP and RREQ messages of AODV protocol provides the decision variables, including channel switching, ROUTE ERROR (RERR) at network layer.

Step 4: The action is selected on the basis of reward value obtained through the Q-learning agent to enhance data rate during transmission.

Step 5: The routing parameters are reconfigured using the Q-learning agent decisions of selecting action.

Step 6: This process work in the exploration of loop back to step one so that new observation of environmental actuation and reformulate the routing parameter values.

#### 4 Mathematical Modelling

The licensed primary users (PUs) and unlicensed secondary users (SUs)/cognitive users are assumed as the coexisting of T in the same geographical area. We consider the SUs posing the properties of an ad-hoc network in which users can receive the transmission information from all other participating users and can share the channel information of the network. The C available free channels assumed for the allocation of wireless resource for the SUs transmission purpose. We assumed these channels are freed from the PUs and SUs can access and avail one single free channel at some particular time. There is no central controller available for SUs and every user is taking its own decisions in a distributive and non-cooperative manner. The IEEE 802.11 standard [37] is used to share the channel over a spectrum. This shared channel is accessed by multiple users at the same time as more than one frequency slots available which is free and may use for communication.

The effect of multiple accessed channels is investigated through the parameters such as channel transmission rate ( $R_r$ ) of the network so that the data rate of individual users can be calculated and examined. The random channel assignment to various SUs caused the multiple channels switching events and due to this switching packet loss and end-to-end delay increased. The channel selection information is accessed through the Q-learning agent at the network layer from MAC layer so that the channel transmission rate can be considered at the time of routing. The Q-learning agent is based on RL basis and hence, the channel selection algorithm is implemented at the time of routing for SUs. This channel selection algorithm works in a non-cooperative manner for every SU and provide the better data rate transmission for overall CRN.

We define this channel selection algorithm mathematically in terms of non-cooperative game as  $\{T, \{S_i\} i \in T, \{U_i\} i \in T\}$ , where  $T$  is the set of decision makers as SUs,  $S_i$  is the set of policies/strategies at individual channel until the C available free channels  $\{s_a, s_b, \dots, s_c\}$ , for user  $i$ . Every SU  $i$  uses the learning function  $U_i: S_i \rightarrow R$  to assign the reward ( $R$ ) to the selected strategy  $s_a$  from the available set  $S_i$  according to the opponents:  $S_{-i}$ . The channel is selected for the routing purpose according to the reward of its strategy which is compared to its opponent and helpful in implementing a better strategy. SUs may select a strategy according the reward recorded at any point of the game such that  $S = [s_1, s_2, \dots, s_T]$  so that no users would deviate anymore from the selected strategy. This stage of strategy is known as the Nash equilibrium and this only happens if and only if Eq. (1) exists [13].

$$U_i(S) \geq U_i(s'_a, s_{-a}), \forall i \in T, s'_a \in S_i \quad (1)$$

The learning algorithm performs very well once it is found the strategy which is on Nash Equilibrium point as this is the point where any player can perform much better than its opponents [38]. Hence, the non-cooperative game theory can be very helpful for selecting a free and available channel for transmission of a SU to improve the data rate. The channel transmission rate  $R_{tr}$  of any channel is calculated using the following function through the strategies profile of a user [38].

$$U_i(s_a, s_{-a}) = \beta + \gamma \log\left(\frac{R_{tr}^a}{\lambda} - \delta\right) \quad (2)$$

where  $\log(\cdot)$  stands for natural logarithm function for the distributive environment. Numerical constants  $\beta$ ,  $\gamma$ , and  $\delta$  in the utility function are user defined as it is the stochastic model implementation of learning technique. The packet arrival rate  $\lambda$  of a user  $i \in T$  of the SUs in the non-cooperative game is used. The strategy  $S_a$  is selected using the Q-learning algorithm so that the channel information is also incorporated during the routing parameters selection at network layer. SUs individually maintained a Q table [36] through Q-learning in a distributive manner for its available strategy for the current routing options. The entries of Q-table are modified and updated through the value of reward for any selected strategy and decided to use it for future routing or not [36].

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha[E(U(s_a, s_{-a})) - Q_t(s, a)] \quad (3)$$

where Q entries of Q table is represented as  $Q_{t+1}(s, a)$  and  $Q_t(s, a)$  at time  $(t + 1)$  and  $t$  respectively for selecting action  $a$  from strategy  $s$ . The average reward of a SU is calculated as  $E[U(s_a, s_{-a})]$  through the utility function of strategy according to any its opponents and  $\alpha$  is the learning rate and the value of it is chosen between 0 and 1. In case the Q values of reward for any two strategies are same, then the strategy is selected on the basis of greedy exploration.

The greedy exploration implemented to select a random available channel which has same reward with probability  $\vartheta$  as  $1 - \vartheta$ . The probability of every channel to be selected by any user starting with a very high  $\vartheta$  value and update  $\vartheta$  after each successful transmission of a user [36].

$$\vartheta = \vartheta - \frac{\vartheta}{\text{Update parameter}} \quad (4)$$

From Eq. (4), the probability selection is based on the following Eq. (5) for channel selection process [36].

$$p_{t+1}(s_a) = \begin{cases} 1 - \vartheta + \left(\frac{\vartheta}{\text{update parameter}}\right) \text{ if } Q_t(s, a) \text{ is highest,} \\ \frac{\vartheta}{\text{update parameter}} \text{ otherwise} \end{cases} \quad (5)$$

SUs update channel information through Q-learning so that during the routing availability of free channel must be assured. The SUs relies on its own channel information for Q-learning so that users periodically access the channel and update the channel selection. The channel reward information is passed through the ACK message whenever starting a new transmission. This process is in cooperated by all SUs participating in CRN to update the channel information using the Q table. Every SU makes its self-configuration routing decision based on the local routing information which is represented as the two Q values which estimate the quality of the alternative paths. The reward value is updated on the basis of Q value and the end-to-end (ETE) delay is calculated as [32],



$$Reward = Q(s', a')_{reward} = n \left( \frac{1}{\frac{ETE_t}{ETE_{max}}} \right) \quad (6)$$

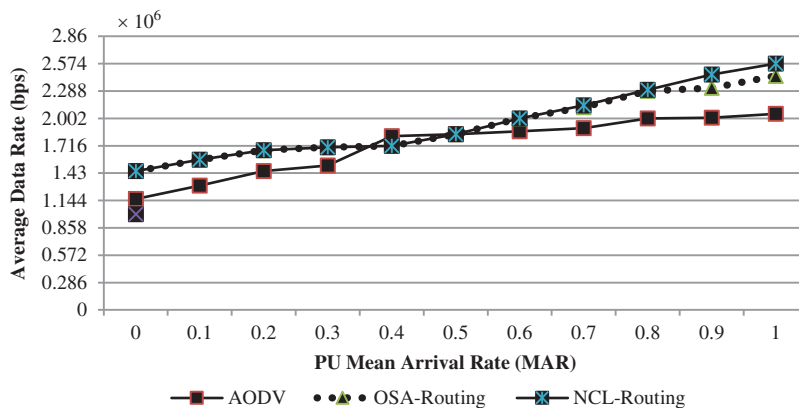
where  $ETE_t$  is the end-to-end delay at current time  $t$  whereas  $ETE_{max}$  is the maximum end-to-end delay observed during the routing for selected route. The  $n$  value shows the effect of normalization constant in case of multiple channel selection for the same route. The end-to-end delay parameter is directly affected data rate for the transmission and hence, the performance of overall network is based on this metrics.

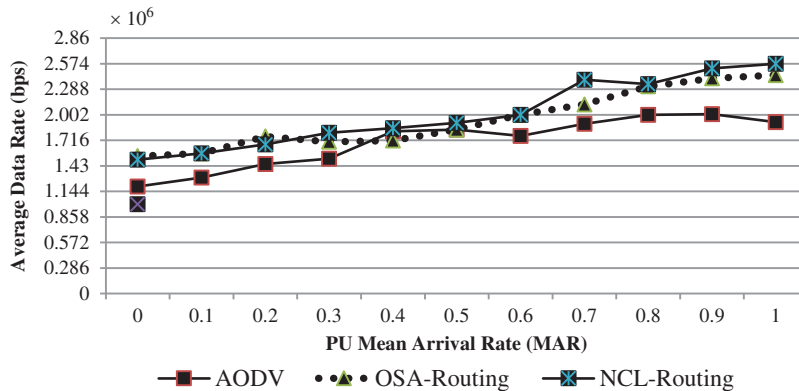
## 5 Simulation Results

Simulations are conducted through Network Simulator-2 (NS-2) versions 2.31 [39], and an additional patch CRCN (Cognitive Radio Cognitive Network) [40] is incorporated for the support of CRN within NS-2. It is an open-source software-based network simulator. CRCN patch offers the variety of functionalities for the CRN environment such as dynamic spectrum resource allocation, power control algorithms at the level of MAC and routing level for the purpose of performance evaluations. The traffic and topology are generated on the basis of realistic stochastic model of CRN. The PHY layer is selected as a reconfigurable multi-radio multi-channel for every SU and the transmission parameters customized such as transmission power, propagation and etc. The CRCN simulator is used for simulation purpose and the simulation area is  $1000 \times 1000 \text{ m}^2$  selected. The simulation area is varied for the different number of nodes during the simulations. The simulation area is varied according to the number of Pus and Sus variations. The impact of simulation area size on the performance of routing is observed as the number of users varies, the proposed routing mechanism performs well in case of users increment due to the maximum availability of routing choices through different users and channels. The packets are generated at the SU by using the Poisson process with a mean arrival rate (MAR) of 0.6 packets/ms in accordance with the Box-Muller transform. The MAR for the activity of the PU indicates the availability of PU on the channel for channel utilization, channel availability information, channel transmission rate and channel transmission time. The PUs are distributed through Poisson Process in a stochastic network environment between the mean arrival rate [MAR] of [0,1] such as low [SD of MAR = 0.0], medium [SD of MAR = 0.4] and high [SD of MAR = 0.8]. The Poisson process is a very important model in queuing theory which can be used when the packets originate from a large population of independent users. The real working of PU on spectrum is not observed due to licensing restrictions. Therefore, The PUs are distributed with the fixed allocation on a spectrum in a stochastic environment of CRN within the mean arrival rate of [0,1]. The mean distribution is checked for the standard error as low [0.0], medium [0.4] and high [0.8] arrival rate of PU using the standard deviation of mean. The performance of proposed model non-cooperative learning (NCL) routing is compared in term of data rate, packet loss and end-to-end delay with very well-known protocol such as AODV routing protocol [7] with Q-learning implementation and OSA (Opportunistic Spectrum Access) Routing [10] protocols. The results are compared with AODV routing protocol which is proposed for the CRN based IoT network and the OSA is indeed the reinforcement learning based latest routing protocol proposed to make devices energy efficient during routing process in CRN. Hence, the proposed routing protocol can also be compared for latest energy constrains in future. The simulation parameters listed in Tab. 1 to build an event driven simulation program for the CR enabled 6G-IoT network. First the average data rate results are evaluated for the routing solution proposed as NCL and shown in Figs. 3–5 for low, medium and high MAR of PUs.

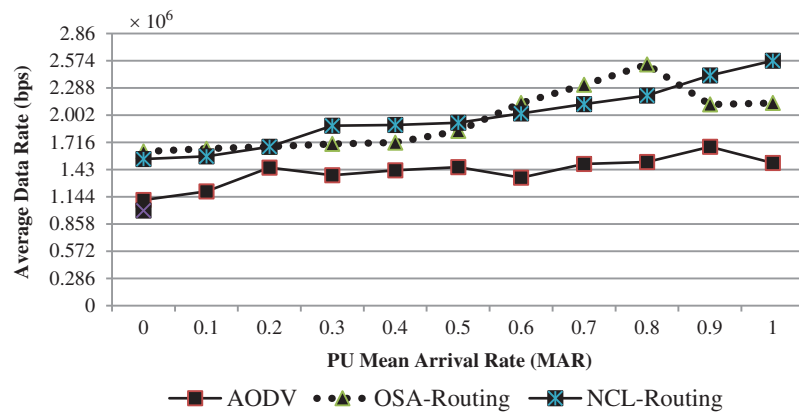
**Table 1:** Parameters of simulations

Parameter	Value
Simulation area	1000 × 1000 m <sup>2</sup>
Secondary users (SUs)	50–100
Primary users (PUs)	1–10
Model for propagation	Two-way round
Spectrum channels	1–6 numbers
Bandwidth of 6 channels 1–6	2–18 MHz
Channel characteristics	0.05 packet/ms mean packet error rate 0.025 sd of packet error rate
Mobility characteristics	Mobility Waypoint = Random Speed = uniform (0–15) m/s Pause time = 0, 60, 120, 180, 240 s
Type of data	Best effort delivery
Payload of packet	8180 bits
Header size	272 bits for MAC
2 <sup>nd</sup> layer protocol for MAC	IEEE 802.11 CogMAC
Header size	127 bits for PHY
Slot time	60 μs
DIFS	128 μs
SIFS	28 μs

**Figure 3:** Average data rate for NCL, AODV and OSA at MAR of PUs = 0.0



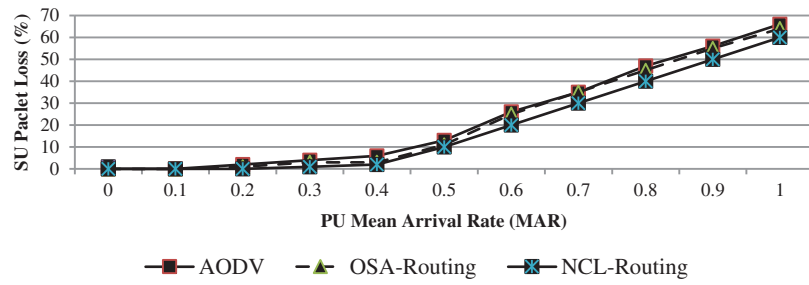
**Figure 4:** Average data rate for NCL, AODV and OSA at MAR of PUs = 0.4



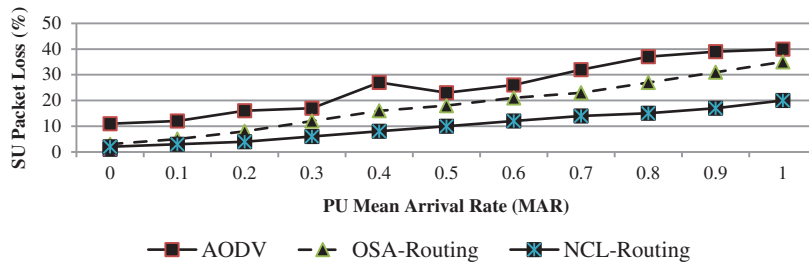
**Figure 5:** Average data rate for NCL, AODV and OSA at MAR of PUs = 0.8

It can notice that the proposed NCL routing has achieved higher data rate as compared to the other existing routing solutions of CR enabled protocols during the transmission. In the start, the data rate is not significantly improved and higher as compared to other protocols due to unavailability of PUs as the MAR of PUs is 0.0 and therefore, the other protocols easily manage to compensate this. However, OSA routing is giving the same results but once the MAR is increasing due to the increment in number of primary users, its data rate is decreasing as shown in Fig. 5 when the MAR of PUs is high at 0.8. The proposed NCL routing is using the reinforcement learning technique on routing layer rather than on MAC layer for the free and available channel selection due to this as the number of users and channels are increasing in the transmission, it performs much better.

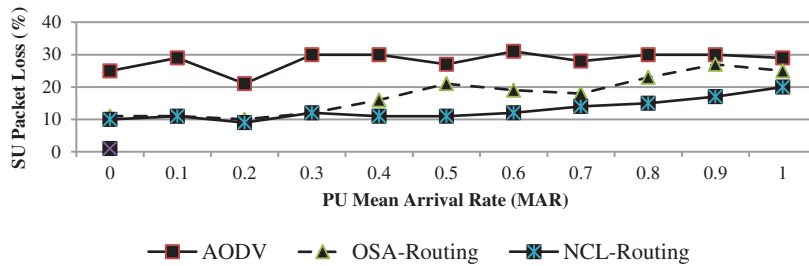
The AODV routing protocol with Q-learning implementation on MAC layer also does not achieve such data rate due to the channel management is done on MAC layer and routing layer is not managed accordingly. The average data rate is observed low when the numbers of SUs are high for AODV and OSA routing protocol while the proposed routing NCL reaching almost 90% increment in delivered packets. This enables it to improve the average data rate by 69% compared to the AODV routing protocol and by almost 39% compared to the OSA routing protocol. This behavior is also confirmed through packet loss and end-to-end delay results as shown in Figs. 6–8 and Figs. 9–11 respectively.



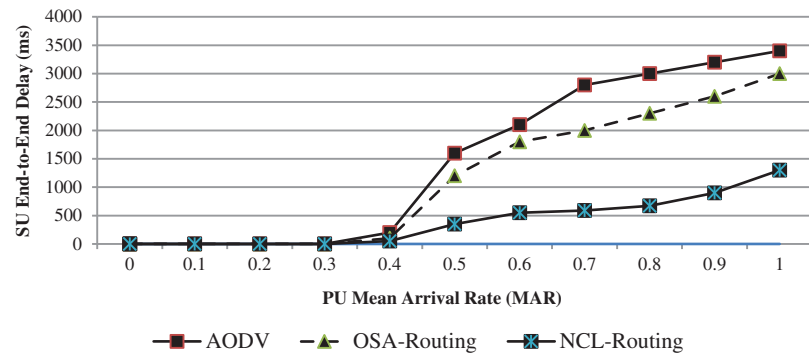
**Figure 6:** Packet loss for NCL, AODV and OSA at MAR of PUs = 0.0



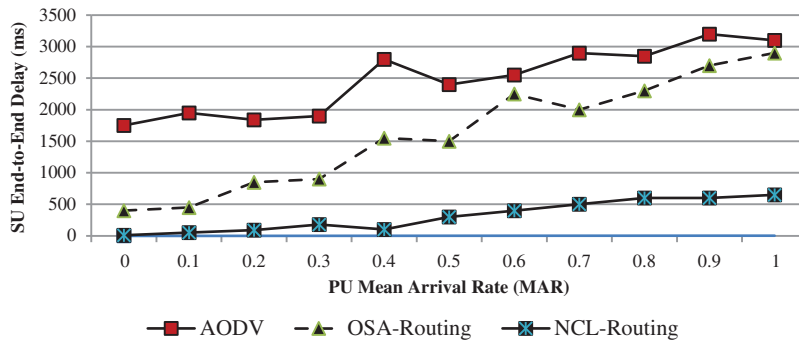
**Figure 7:** Packet loss for NCL, AODV and OSA at MAR of PUs = 0.4



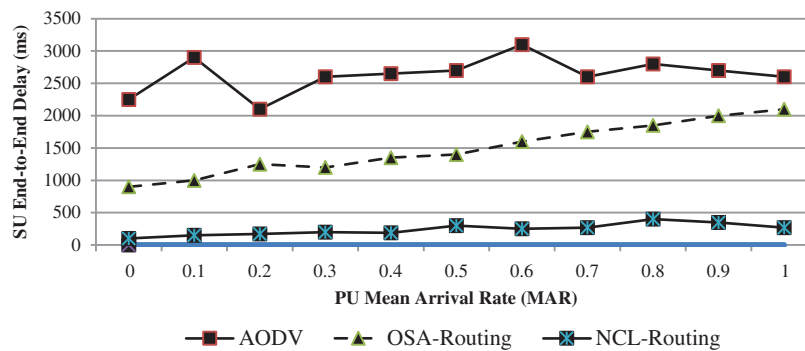
**Figure 8:** Packet loss for NCL, AODV and OSA at MAR of PUs = 0.8



**Figure 9:** End-to-end delay for NCL, AODV and OSA at MAR of PUs = 0.0



**Figure 10:** End-to-end delay for NCL, AODV and OSA at MAR of PUs = 0.4



**Figure 11:** End-to-end delay for NCL, AODV and OSA at MAR of PUs = 0.8

The number of packet loss for a SU is shown in Figs. 6–8 at different MAR and NCL routing performs much better than other two routing protocols. The packet loss decreases almost 56% compared to the AODV routing protocol and 25% compared to the OSA routing protocol when the MAR of the PU = 0.8 (user/ms). This is happened due to the less user and channel switching due to the prior channel information updated for every user at network layer. Finally, the end-to-end delay is minimized as compare to the AODV and OSA routing protocols due to the learning capability of NCL routing as shown in Figs. 9–11 for low, medium and high MAR of PUs. The EED consisted of different types of delays and when the MAR for the activity of PU is low than the EED becomes higher due to the less routing choices available to the SUs. This can be observed from Fig. 11 as the availability of PU is higher when the MAR = 0.8, the channel and user pairs have greater differences in the activity of the PU. Hence, the EED of the SU is lower when the MAR of the PU is 0.8 compared to when the MAR = 0.4 and NCL routing achieves a lower EED for the SU of up to 89% when the MAR = 0.8 (user/ms) in comparison to that of other routing protocols. This happens due to the increment in sd’s MAR of PUs which creates more routes and channel choices.

The proposed routing NCL is very suitable for low latency and higher data rate applications for the upcoming generation networks such as 6G based IoT which also posse the capability of cognitive radio to utilize and manage the dynamic spectrum access. The proposed routing is recommended in large and having a greater number of users’ environment as observed from the results that as the MAR of PUs is increasing, it is performing much better. This is due to the reinforcement learning mechanism and stability of network as with the passage of time previous routing decisions are more and hence, not much exploration learning will take place. Initially the exploration learning is much high due to this the proposed routing is not able to perform significantly as compare to other.



## 6 Conclusion

Routing is at the core of the vision of any network and very demanding and challengeable in wireless networks especially for higher data rate and lower latency rate. Previously, a lot of focus on general Machine learning based techniques explored for optimizing the cognitive radio parameters of PHY and MAC layers in wireless networks beyond 5G, but less focus on utilizing the learning techniques at the network layer especially for the routing solution. We have presented and simulated a machine learning technique and showed through results that it can be very productive and can lead to improve beyond 5G network performance in terms of average data rate, packet loss and end-to-end delay. This proposed technique for autonomously reconfigured network systems of 6G based IoT and is implemented through a CR enabled network architecture approach. We also presented mathematical modeling of the technique and simulations using NS-2 which shows the performance evaluation. There are many other non-cooperative learning techniques available to be tested for routing purpose in CR enabled 6G-IoT networks for its data rate and latency requirements.

**Funding Statement:** The authors received no specific funding for this study.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] S. J. Nawaz, S. K. Sharma, S. Wyne, M. N. Patwary and M. Asaduzzaman, "Quantum machine learning for 6G communication networks: State-of-the-art and vision for the future," *IEEE Access*, vol. 7, pp. 46317–46350, 2019.
- [2] 6G Enabling Technologies, 6G mm Wave Networking Group, 2021. [Online]. Available: <http://mmwave.dei.unipd.it/research/6g/>.
- [3] N. Chen and M. Okada, "Towards 6G internet of things and the convergence with RoF system," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 8719–8733, 2020.
- [4] A. M. Alberti, D. Mazzer, M. M. Bontempo, L. H. de Oliveira and R. R. Righi, "Cognitive radio in the context of internet of things using a novel future internet architecture called NovaGenesis," *Computers & Electrical Engineering*, vol. 57, no. 2, pp. 147–161, 2017.
- [5] D. S. Gurjar, H. H. Nguyen and H. D. Tuan, "Wireless information and power transfer for IoT applications in overlay cognitive radio networks," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3257–3270, 2018.
- [6] H. B. Salameh, S. Otoum, M. Aloqaily, R. Derbas, I. Al Ridhawi *et al.*, "Intelligent jamming-aware routing in multi-hop IoT-based opportunistic cognitive radio networks," *Ad Hoc Networks*, vol. 98, no. 3, pp. 102035, 2020.
- [7] S. Anamalamudi, A. R. Sangi, M. Alkathairi and A. M. Ahmed, "AODV routing protocol for cognitive radio access-based internet of things (IoT)," *Future Generation Computer Systems*, vol. 83, no. 1, pp. 228–238, 2018.
- [8] K. Katzis and H. Ahmadi, "Challenges implementing internet of things (IoT) using cognitive radio capabilities in 5G mobile networks," in *Internet of Things (IoT) in 5G Mobile Technologies*. Berlin: Springer, vol. 8, pp. 55–76, 2016.
- [9] H. B. Salameh, R. Derbas, M. Aloqaily and A. Boukerche, "Secure routing in multi-hop IoT-based cognitive radio networks under jamming attacks," in *Proc. of MSWIM '19*, FL, Miami Beach, USA, pp. 323–327, Nov. 2019.
- [10] R. A. A. Diab, A. Abdrabou and N. Bastaki, "An efficient routing protocol for cognitive radio networks of energy-limited devices," *Telecommunication Systems*, vol. 73, no. 4, pp. 577–594, 2020.
- [11] R. A. Diab, N. Bastaki and A. Abdrabou, "A survey on routing protocols for delay and energy-constrained cognitive radio networks," *IEEE Access*, vol. 8, pp. 198779–198800, 2020.
- [12] S. Kalambe, P. Lohiya and P. Malathi, "Performance evolution of energy detection spectrum sensing technique used in cognitive radio," in *Proc. of ICSPCT'14*, Ajmer, India, pp. 786–790, July 2014.
- [13] J. Felegyhazi, P. Hubaux and L. Buttyan, "Nash equilibria of packet forwarding strategies in wireless ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 5, no. 5, pp. 463–476, 2006.

- [14] M. Huang, A. Liu, N. N. Xiong, T. Wang and A. V. Vasilakos, "An effective service-oriented networking management architecture for 5G-enabled internet of things," *Computer Networks*, vol. 173, no. 2, pp. 107208, 2020.
- [15] A. Bagwari and G. S. Tomar, "Cooperative spectrum sensing in multiple energy detectors based cognitive radio networks using adaptive double-threshold scheme," *International Journal of Electronics*, vol. 101, no. 11, pp. 1–13, 2014.
- [16] A. Carie, M. Li, B. Marapelli, P. Reddy and H. Dino, "Cognitive radio assisted WSN with interference aware AODV routing protocol," *Ambient Intelligence and Humanized Computing*, vol. 10, no. 10, pp. 4033–4042, 2019.
- [17] A. Guirguis, M. Karmoose, K. Habak, M. El-Nainay and M. Yousef, "Cooperation-based multi-hop routing protocol for cognitive radio networks," *Network and Computer Applications*, vol. 110, no. 1, pp. 27–42, 2018.
- [18] A. Bagwari and G. S. Tomar, "Enriched the spectrum sensing performance of estimated SNR based detector in cognitive radio networks," *International Journal of Hybrid Information Technology*, vol. 8, no. 9, pp. 143–156, 2015.
- [19] A. Bagwari, J. Kanti, G. Tomar and A. Samara, "A Robust detector using SNR with adaptive threshold scheme in cognitive radio networks," *International Journal of Signal Processing, Image Processing, and Pattern Recognition*, vol. 9, no. 5, pp. 173–186, 2016.
- [20] A. Bagwari, J. Kanti and G. Tomar, "Novel spectrum detector for IEEE 802.22 wireless regional area network," *International Journal of System Dynamics Applications*, vol. 59, pp. 4443–4452, 2015.
- [21] S. Khadim, F. Riaz, S. Jabbar, S. Khalid and M. Aloqailay, "A non-cooperative rear-end collision avoidance scheme for non-connected and heterogeneous environment," *Computer Communications*, vol. 150, no. 6, pp. 828–840, 2020.
- [22] M. Compare, P. Baraldi and E. Zio, "Challenges to IoT-enabled predictive maintenance for industry 4.0," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4585–4597, 2019.
- [23] J. Li, B. N. Silva and M. Diyan, "A clustering-based routing algorithm in IoT aware wireless mesh networks," *Sustainable Cities and Society*, vol. 40, no. 1, pp. 657–666, 2018.
- [24] A. Ali, L. Feng, A. K. Bashir, S. El-Sappagh and S. H. Ahmed, "Quality of service provisioning for heterogeneous services in cognitive radio-enabled internet of things," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 328–342, 2018.
- [25] S. Jabbar, R. Iram, A. A. Minhas, I. Shafi and S. Khalid, "Intelligent optimization of wireless sensor networks through bio-inspired computing: Survey and future directions," *International Journal of Distributed Sensor Networks*, vol. 9, no. 2, pp. 421084, 2013.
- [26] G. Kaur, P. Chanak and M. Bhattacharya, "Energy efficient intelligent routing scheme for IoT-enabled WSNs," *IEEE Internet of Things*, vol. 8, no. 14, pp. 1, 2021.
- [27] K. Habak, M. Abdelatif, H. Hagrass, K. Rizc and M. Youssef, "A location-aided routing protocol for cognitive radio networks," in *Proc. of ICNC'13*, San Diego, CA, USA, pp. 729–733, May 2013.
- [28] F. A. Awin, Y. M. Alginahi, E. Abdel-Raheem and K. Tepe, "Technical issues on cognitive radio-based Internet of Things systems: A survey," *IEEE Access*, vol. 7, pp. 97887–97908, 2019.
- [29] N. Javaid, A. Sher, H. Nasir and N. Guizani, "Intelligence in IoT-based 5G networks: Opportunities and challenges," *IEEE Communications Magazine*, vol. 56, no. 10, pp. 94–100, 2018.
- [30] R. Ali, Y. A. Qadri, Y. Bin Zikria, T. Umer, B. Kim *et al.*, "Q-learning-enabled channel access in next-generation dense wireless networks for IoT-based eHealth systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–12, 2019.
- [31] S. Vimal, M. Khari, R. G. Crespo, L. Kalaivani, N. Dey *et al.*, "Energy enhancement using multiobjective ant colony optimization with double Q learning algorithm for IoT based cognitive radio networks," *Computer Communications*, vol. 154, no. 4, pp. 481–490, 2020.
- [32] H. A. A. Al-Rawi, K. A. Yau, H. Mohamad, N. Ramli and W. Hashim, "A reinforcement learning-based routing scheme for cognitive radio ad hoc networks," in *Proc. of WMNC'14*, Vilamoura, Portugal, pp. 1–8, May 2014.
- [33] Y. Zhou, F. Zhou, Y. Wu, R. Q. Hu and Y. Wang, "Subcarrier assignment schemes based on Q-learning in wideband cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 1168–1172, 2019.

- [34] M. Bkassiny, Y. Li and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1136–1159, 2012.
- [35] M. Wiering and M. Van Otterlo, "Reinforcement learning," *Adaptation, Learning, and Optimization*, vol. 12, pp. 441–467, 2012.
- [36] C. J. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3–4, pp. 279–292, 1992.
- [37] G. R. Hiertz, D. Denteneer, L. Stibor, Y. Zang, X. P. Costa *et al.*, "The IEEE 802.11 universe," *IEEE Communications Magazine*, vol. 48, no. 1, pp. 62–70, 2010.
- [38] R. Rajabioun, E. Atashpaz-Gargari and C. Lucas, "Colonial competitive algorithm as a tool for Nash equilibrium point achievement," in *Proc. of ICCSA'08*, Berlin, pp. 680–695, 2008.
- [39] Network Simulator -2 (NS-2), 2021. [Online]. Available: <http://www.isi.edu/nsnam/ns/ns-build.html>.
- [40] Cognitive Radio Cognitive Network (CRCN) Simulator, 2021. [Online]. Available: [http://faculty.uml.edu/tricia\\_chigan/Research/CRCN\\_Simulator.htm](http://faculty.uml.edu/tricia_chigan/Research/CRCN_Simulator.htm).