CROSS LAYERING USING REINFORCEMENT LEARNING IN COGNITIVE RADIO-BASED INDUSTRIAL INTERNET OF AD-HOC SENSOR NETWORK

Chetna Singhal and Thanikaiselvan V

School of Electronics Engineering, VIT, Vellore, India

ABSTRACT

The coupling of multiple protocol layers for a Cognitive Radio-based Industrial Internet of Ad-hoc Sensor Network, enables better interaction, coordination, and joint optimization of different protocols in achieving remarkable performance improvements. In this paper, network, and medium access control (MAC) layer functionalities are cross-layered by developing the joint strategy of routing and effective spectrum sensing and Dynamic Channel Selection (DCS) using the Reinforcement Learning (RL) algorithm. In an industrial ad-hoc scenario, the network layer utilizes the sensed spectrum and selected channel by MAC layer for next-hop routing. MAC layer utilizes the lowest known transmission delay of a channel for a single hop as computed by the network layer, which improves the MAC channel selection operation. The applied RL-based technique (Q learning) enables the CR Secondary Users (SUs) to sense, learn, and make the optimal decision on their environment of operations. The proposed RLCLD schemes improve the SU network performance up to 30% as compared to conventional methods.

KEYWORDS

Reinforcement Learning, Cognitive Networks, Cross layer, Industrial Internet, Routing.

1. INTRODUCTION

With the advancement of Industrial Internet of Things (IIoT), ad-hoc sensor networks were found to be useful in a vast area starting from medical to disaster recovery, military scenarios, and smart home security system as they require infrastructure less prompt set-up with the capability of self-organizing/healing. However, their rapid usage in the unlicensed ISM frequency creates congestion and adverse interference effect on network performance. So, the concept of Cognitive Radio based Industrial Internet of Ad-hoc Sensor Network (CR-IIAHSN) emerged [1,2], where sensor nodes can exploit the underutilized licensed frequency band as unlicensed users or secondary users (SUs), when the licensed users or primary users (PUs) are not transmitting. However, there exist varieties of major challenges for CR-IIAHSN like varying industrial network topology, data routing over multiple hops, spectrum availability, and many other parameters. The operating channel characteristic is dependent mostly on PUs activity and largely varies with time.

Spectrum availability is most important for each sensor to operate without harming the communication of PUs. Thereafter, due to the frequent changing of spectrum bands, it is quite troublesome for any sensor to gather network active status information of all the neighboring sensors. Due to dynamic spectrum availability, undesirable communication challenges like end-to-end delay, packet loss, jitter, retransmission, path loss, and interference take place. In such a

scenario, reliable data communication is quite difficult by appropriate route set up through multiple hop nodes and a variety of the spectrum band. Furthermore, mobility of nodes greatly impacts the channel occupancy, received signal strength and interference, and of CR-IIAHSN. The sensor nodes cooperatively sense the spectrum and eventually share it to access the wireless medium. This might lead to large overhead and chances of false spectrum detection. Additionally, frequent spectrum sensing, spectrum handover, and link disconnection cause limited energy of the battery-powered sensors and poor network coverage of the quality of service (QoS) hungry sensor nodes [3].

All such challenges cannot be solved by the traditional layered protocol design like OSI (Open System Interconnection). Hence, this concern demands new paradigm like cross-layer design techniques which can simultaneously address the wide range of communication challenges by considering multiple layers as found in the conventional protocol stack that resides on a cognitive sensor node [4]. This enables a significant performance improvement compared to conventional layering approach, where each layer functions independently with their own protocols. CR-AHSN operates subject to the cognitive cycle, in which radios must observe their operating environment, and thereafter conclude the best ways to adapt with the environment, and act accordingly. This cognitive cycle is repetitive, so, the radio can utilize this opportunity for learning from its earlier taken steps. The fundamental operation relies on the radio's capability to sense, adapt, understand, and learn. For effective real-time processing, the CR-AHSN can be integrated with Artificial Intelligence (AI) and machine-learning (ML) approaches that can achieve an adaptable and robust detection of spectrum hole or vacant spectrum. The key steps in machine learning for such sensor nodes can be described as :(1) Sensing the radio frequency (RF) factors like quality of channel, (2) Monitoring the functional surrounding and inspecting its feedback such as ACK,(3) Learning,(4) Updating the machine learning model by utilizing the decisions and observations to obtain improved accuracy for decision-making, (5) Eventually, settling on problems of managing resources and adapting the communication parameters consequently.

Hence, in this paper, the cross-layer design concept for CR-IIAHSN is combined with AI techniques to achieve better performance in CR networks. The main contribution in this paper is highlighted as follows:

I. Developing a spectrum-aware and reconfigurable routing protocol in CR node Network Layer that will discover suitable network path by avoiding zones occupied by PUs activity, and will adjust with the dynamic characteristics of CR-IIAHSNs, like mobility, and variable link quality

II. The routing method is modelled as a reinforcement learning task, where the cognitive radio source can acquire the most appropriate route towards the destination through a hit-and-miss method of interaction

III. Developing effective spectrum sensing and dynamic channel selection technique for CR nodes which is modelled using RL in MAC layer with the reward mechanism related to RL

IV. Designing a cross-layer approach between the Network and MAC layer of CR nodes based on the sophisticated AI and machine learning technologies, such as Reinforcement Learning (RL), Q learning and Markov Decision Process (MDP)

V. Eventually, deriving a suitable RL model and comparative performance analysis through simulation result for the scenario having CR node coupled with cross layering and without cross layering

The remaining paper is arranged as follows: The several existing methodologies related to the current research area are presented in section 2 from the literature, section 3 highlights the

network model and problem formulation, thereafter the proposed methodology for CR-IIAHSN is depicted in section 4, then in section 5, we brought out few significant outcomes of the proposed cross-layer design, which depicts about how the CR users performed in an CR-IIAHSNs.

2. RESEARCH BACKGROUND

Many researchers collaborated existing CR techniques with AI and machine-learning methods. This makes CR more effective to process real-time information and faster decision making. In [5], the action performance is enhanced using Cooperative spectrum sensing. Abbas et al. [6] presented that Machine learning techniques can be assimilated in the CR network to develop intelligence and knowledge in a wireless system; and brought out relevant challenges, implementation, of AI for CR network nodes. Authors in [7] brought out various application areas of AI-based RL technique for CR networks as: a) Dynamic Channel Selection for CR nodes, b) Channel Sensing and detecting the presence of PU activities, c) Routing for the best route(s) selection for SU nodes, d) Medium Access Control (MAC) protocol to reduce collision of packets and maximize utilization of channels for CR networks, e) Power Control[8] to improve Signal-to-Noise Ratio (SNR) and packet delivery statistics and f) Energy Efficiency method[9] to reduce energy consumption.

Moving ahead, authors in [10], brought out a survey on multiple AI methodologies like game theory, reinforcement learning, Markov decision model, fuzzy logic, neural networks, artificial bee colony algorithm, Bayesian, support vector, entropy, and multi-agent systems. They further investigated by comparative analysis of various learning techniques related to the CR network. Further, AI and ML -based methods [11] were introduced into routing protocol design of cognitive radio nodes. Here, multiple artificial intelligence techniques like machine learning, game theory, optimization, control theory, economics are integrated to present a useful solution for decision-making as required for the CR network layer. They also brought out that if the learned knowledge from the present and previous scenarios are consolidated it can further improve the CR performance.

Reinforcement learning (RL) is an unsupervised AI method, where the system learns directly or indirectly from reinforcements (positive reinforcement are known as rewards and negative reinforcement as punishments) based on actions selected. RL allows CR nodes to learn from their past state which empowers a CR node to observe, learn, and take suitable strategy on selecting necessary action to maximize overall network performance [7]. For example, using RL, every SU observes the spectrum, understands its present transmission factors, and determines required steps whenever a PU turns up. In this process, when such SUs interferes with PUs, penalty values are enforced. This in turn, improves the SU learning technique and the overall spectrum usage.

RL-based cross-layer and decision-making [12] methodology to select minimum-cost route between a set of source and destination nodes, is investigated to improve the efficacy of CR node networks. The simulation shows that different attributes of RL like exploitation vs. exploration, reward function, and rate of RL learning boosts SUs' network performance. Such an approach for SU nodes significantly brought down overall network delay and loss of packets and further enhanced throughput without impacting PUs' operations. Authors further introduced a new routing scheme as Cognitive Radio Qrouting (CRQ-routing) which focuses on minimizing the interference by SUs' to PUs throughout the network. Further, the authors [13] showed an example of Q-learning for CR-IIAHSNs by incorporating RL techniques to jointly allocate power and spectrum for CR nodes. They utilised RL techniques to solve complex issues like interference-management in CR-IIAHSNs, where every CR node learns the most appropriate combination of various factors (for example, spectrum and power). Simulation showed improved performance of the RL-based method as compared to the non-learning scheme under both static

and dynamic spectrum scenarios. Authors in [14] presented a cross-layer routing technique in CR system through deep learning. Here, cross-layering is carried out by collecting various attributes between network and physical layer to improve the Quality of Service (QoS) in multimedia application.

3. CROSS LAYERING IN CR-IIAHSN USING REINFORCEMENT LEARNING

3.1. Network Model

In the present research, the CR-AHSN model assumes the presence of both licensed PUs and unlicensed SUs all over the network. We also assumed that no base station is available to control the network and communication is completely ad-hoc i.e. every SU is responsible for an entire range of CR-related operations like channel selection strategy, spectrum sensing, data transmission, etc. All the PUs are authorised to transmit/receive using the licensed frequency spectrum at any random time and in the absence of PU (PU OFF state), SUs can use such spectrum. It is also considered that if a Primary User is active at present and its transmission band matches partly or fully over a Secondary User's channel, then the SU will experience interference from the Primary User and will not be permitted to continue communication anymore in this duration. It is considered that every SU will have two transceiver radios: one is for transmitting data and the other one for exchanging control messages among neighbouring SUs. Furthermore, our assumption considers that each CR user performs perfect sensing to accurately deduce the existence of the PU as and when they come under PU's transmission range.
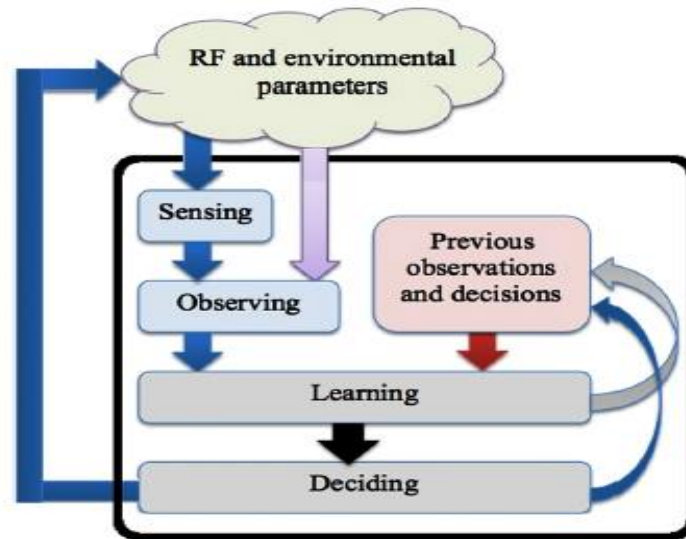

Figure 1. Learning States in CR-IIAHSN

Each CR user is capable of deciding about its operational spectrum, channel and transmission path independently of the other users in the neighbourhood. The CR users can continuously monitor the chosen spectrum each time slot and can switch from one channel to another as per the operational scenario (Figure 1). This problem statement can be formulated mathematically as follows using Q functions. Let's assume at decision epochs t (t ϵ T = {1, 2, . . .}), the learning awareness acquired by CR node agent i for a specific state-action pair is expressed by [15]

$$Q_{t+1}^i(s_t^i, a_t^i) \leftarrow (1 - \alpha)Q_t^i(s_t^i, a_t^i) + \alpha[r_{t+1}^i(s_{t+1}^i) + \gamma max Q_t^i(s_{t+1}^i, a)] \qquad (1)$$

where, $s_t^i \in S$ as state, $a_t^i \in A$ as action, $r_{t+1}^i(s_{t+1}^i) \in R$ is the reward received at time instant t+1 for action occurred at time t and known as a delayed reward. $0 \leq \gamma \leq 1$ is the discount factor. $0 \leq \alpha \leq 1$ is the learning rate.

The higher learning rate indicates that the agent is more dependent on the delayed reward. At time instant t+1, the state of the CR node agent changes from s_t^i to a new state s_{t+1}^i as a result of an action a_t^i and the CR node obtains a delayed reward $r_{t+1}^i(s_{t+1}^i)$ for which the value is updated using (1). Henceforth, for all remaining time instant, i.e. t+1, t+2…., the CR node should take optimal action concerning the states. As this procedure evolves with time and the CR agent continues receiving the sequence of rewards, the key goal should be to find out an optimal policy for a longer duration through maximizing the value functions as following

$$V^\pi(s_t^i) = \max(Q_t^i(s_t^i, a)) \tag{2}$$

The CR agent would aim to select an action or derive the optimum policy as:

$$\pi_i(s_t^i) = \arg\max(Q_t^i(s_t^i, a)) \tag{3}$$

Hence, the main goal is to attain this optimum policy.

3.2. Problem Formulation

Based on the above system model and operational characteristics of CR nodes, we further took into consideration the following challenges in the purview of CR ad-hoc sensor network.

I.   The routing protocol in the CR node network layer should discover a route by avoiding areas occupied by PUs activity, and to adjust with the dynamic characteristics of CR-IIAHSNs, namely, spectrum availability, mobility, and variable link quality
II. The spectrum consciousness and dynamic switchover are major criteria for CR-IIAHSN routing protocol[16]. The routing method can be modelled as a function of RL, where the CR source node can find out the superior route towards the destination by a trial-and-error method of interaction
III. Every SU or CR node should decide an appropriate frequency channel to transmit, with the goal of achieving maximum aggregate performance with enhanced QoS arrangement and minimizing the interference on PUs
IV. Different channels in the CR network will have various properties, namely, transmission range, Bit Error Rate (BER), data rate, and varying PU activity over time. So, effective spectrum sensing and dynamic channel selection for CR nodes can be modelled using RL in the MAC layer with the reward mechanism associated with RL
V.   To cater to the requirements of points I and II, a cross-layer approach needs to be designed among the Network and MAC layer of CR nodes based on the futuristic AI and ML approaches, such as RL, Q learning[17]and Markov Decision Process (MDP)
VI. Eventually, a suitable RL model must be derived and comparative performance analysis to be carried out through simulation result for the scenario having CR node coupled with cross layering [18] and without cross layering

4. PROPOSED METHODOLOGY

Reinforcement learning (RL) based concept (Q learning) has been utilized in CR-IIAHSN, such that the SUs is able to properly sense, understand, and determine optimal actions for their own

surrounding functional environment. As an example, a SU tries to sense its surrounding spectrum to discover white spaces, finds out the most suitable channel to transmit data, and selects actions like transmitting required data through the best possible channel along with the most efficient route. Network and MAC layer functionalities are cross-layered by developing the joint strategy of routing and effective spectrum sensing and Dynamic Channel Selection (DCS) using the RL algorithm. The Network Layer utilizes the sensed spectrum and selected channel by the MAC layer for next-hop routing. On the contrary, the MAC layer utilizes the lowest known transmission delay of a channel for a single hop as computed by the Network layer, which improves the MAC Channel Selection [19] operation. Rewards for RL, are formulated based on the information accumulated at each node from the successful transmission, collision due to SU/PU transmission, link errors, and Acknowledgment messages carrying performance metrics (delay and path-stability), etc.

4.1. Markov Decision Process (MDP)

The RL model through MDP is formulated as follows [14] :

States (St): It is the environmental variable that an agent will be sensing very often. In this case, SUs will play the role of CR agents.

Action (Ac): The agent adopts a set of possible actions to improve its performance based on interaction with the environment. This is known as an action space.

Transition Probabilities (Tp): The transition probabilities describe the probability with which an agent will settle down at any specific state. Considering the present running state and a particular action chosen by the agent, this is denoted by

$$Tp\,(st\,,\,ac,\,st\,') = PrT\,(st_{d+1} = st\,'\,|ac_d = ac,\,st_d = st) \tag{4}$$

Where, PrT denotes the probability of state transition at time instant d for the specific state st_d and selected action ac_d.

Rewards (R): Reward signifies the outcome of an action taken by the agent in terms of success or failure. After the agent performs an action, the reward is calculated by monitoring the changes in the environmental states.

Discount factor (γ): The discount factor calculates the return over long-term.

Initially, the agent will be at state $st_0 \in St$. With the progress of every time instant d, the agent selects a specific action from the set of possible actions known as action space (Ac). Then, the system migrates to the subsequent state according to the probability Tp and the agent gets an instant reward R (Figure 2). Here, agent aims to achieve the maximum promotional summation of rewards through a longer time period. The phenomena of agent's selection of action with the variation in the surrounding, is indicated as the policy *(Pol)* adopted by agent. The agent communicates with the surrounding environment and selects particular action according to the policy.
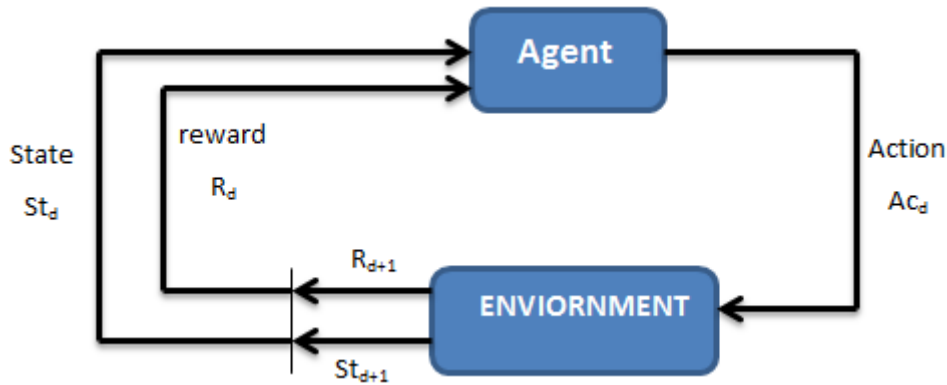
Figure 2. Agent State-Action diagram in CR-IIAHSN

Furthermore, the policy *Pol* establishes the relation between the actions and relevant states, for every step d. Here, the agent aims to discover the optimal policy, as designated by various reward models. The reinforcement learning approach can be used to establish the best possible policy for the MDP. The associated equations are as following:

St: set of states and Ac: set of actions

Tp: $St \times Ac \times St \times \{0,1,...,H\} \rightarrow [0,1]$, $Tp_d(st,ac,st') = P(st_{d+1} = st' | st_d = st, ac_d = ac)$

R: $St \times Ac \times St \times \{0, 1,...,H\}$, $R_d(st,ac,st') =$ reward for $(st_{d+1} = st', st_d = st, ac_d = ac)$

H: denotes the horizon upon which the agent takes action

Goal: Figure out (*Pol*) : $St \times \{0, 1, ...,H\} \rightarrow$ Ac that attains the maximum anticipated sum of rewards, i.e., *Pol**= arg max E [$\sum_{d=0}^{H}$ $R_d(St_d, Ac_d, St_{d+1}) | Pol]$

 Pol

4.2. Q Learning

Q-learning, a wide spread RL method that does not follow any model, is applied while implementing cross-layer CR Networks. Q-learning aims to assess the optimal action-state function Q(st,ac),without enforcing any environmental model. If the agent is in state st at present, performs an action ac, and acquires the reward r and eventually transits into the following state st' , then Q-learning renews the Q function in the following manner [15] :

$$Q(st,ac) \rightarrow Q(st, ac) + \alpha [r + \gamma \max_{ac'} Q(st',ac') - Q(st,ac)] \qquad (5)$$

Where $0 \leq \gamma \leq 1$ refer to discount factor, $0 \leq \alpha \leq 1$ is a parameter denotes the agent learning rate. For each state st, the agent chooses ac that maximizes the Q(st, ac) values. The Q-learning algorithm functions by considering the transitional reward from the environment that makes progressive optimization of the SU transmission parameters. A table (Q table) is used by Q-learning to store the state-action pair values. The Q learning steps are briefly shown in Figure 3 and thereafter the algorithm steps are shown in the next subsection.

Firstly, the Q-table has to be built with state-action pair. This table will have n number of actions as columns and m rows as states. The table parameters will be set to zero initially. Now, an action

(ac) in the state (st) will be chosen by the agent based on the Q-Table. After taking an action, the outcome will be observed and a reward will be measured accordingly. In this way, the Q-Table gets updated and the value function Q is boosted to maximum state. The combination of choosing an action and performing the action will continue for multiple iterations until the learning is completed and a good Q table is formed to choose the optimal policy.
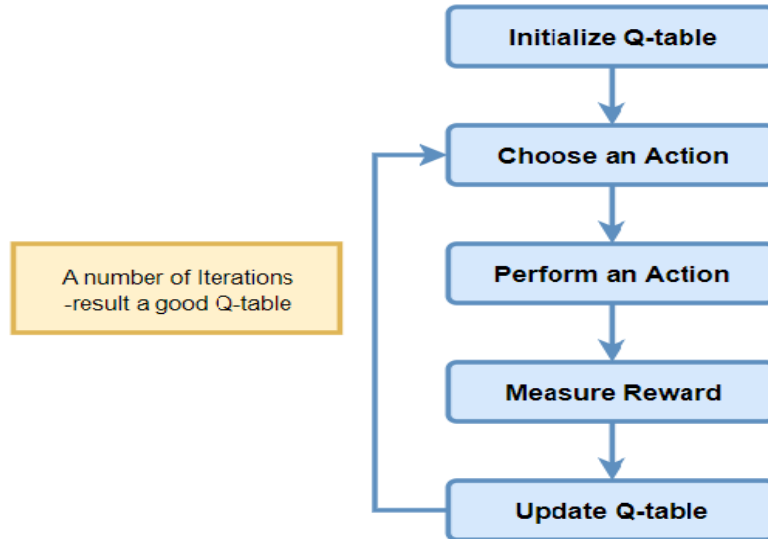

Figure 3. Q Learning Steps in CR-IIAHSN

4.2.1. Q Learning Algorithm

(1) Assign the gamma (γ) parameter value and reward from the environment in R.
(2) Set Q as all zeros
(3) Pick an initial state randomly
(4) Assign initial state = current state
(5) Choose one actions (amid all set of probable schedules) in present state
(6) Use chosen action in the direction of subsequent state
(7) Acquire highest Q cost for subsequent value through probable potential actions
(8) Calculate: Q(value, deed) = R(value, deed) + Gamma* Max[Q(subsequent value, entire deeds)]
(9) Repeat above steps until current state = goal state

4.3. Routing based on Q Learning

Each CR node decides its routing choice depending on the local routing intelligence, which consists of a table having Q values to assess the quality of the substitute routes. All such Q values are revised every instant when the node transmits a packet to one of its peers or neighbours. In such fashion, as the node dispatches packets, its Q-values progressively accumulate fresh global information of the network. The CR nodes collect information they receive back from their neighbouring CR nodes when they transmit packets to them (known as forward exploration for source node) and the information added to the packets as they receive from their neighbouring nodes (known as backward exploration for the destination). This method is known as Dual RL-based Q learning technique that has been adopted in the current research for routing. Our model amends two Q-functions for the subsequent and preceding states concurrently, to speed up the process of learning. The dual RL technique has been incorporated to solve the routing issues in

CR-IIAHSNs, such that every time a packet is dispatched over a wireless channel, the Q(st, ac) are renewed at both of the transmitter and receiver end (Figure 4).
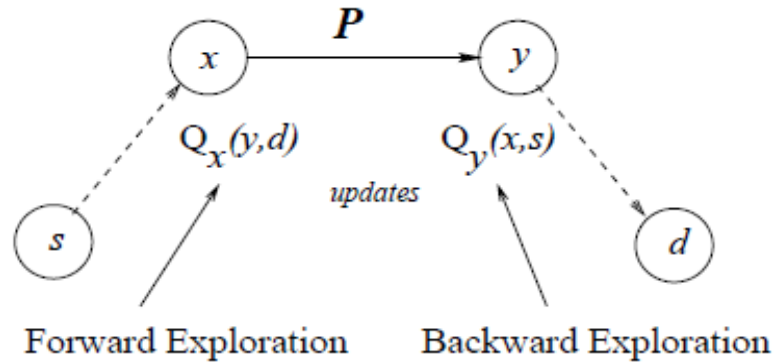

Figure 4. Dual RL Q Learning in CR-IIAHSN

At the time of onward investigation, CN x (at sender side) apprises its Qx (y,d) cost related to leftover route of packet P through CN y for destination d. While retrograde search, the recipient CN y (at receiver side) apprises its Qy (x,s) cost about its travelled route P through node x towards node s. By finding out associations with minimum delay that indicates the time interval needed for successful transport of secondary users' packet to a CN located at next-layer. Here, secondary users' interference with primary users is minimized through reinforcement learning and secondary users' overall system performance can be improved. The routing protocol has been designed to use three types of packets: packets carrying network control information like route discovery, then packets to carry data traffic , and packets that carry ACK feedback information as required by the reward function of RL. This reward function has been designed as a combination of high stability, low delay route, offered bandwidth, packet drop ratio, link reliability and energy consumption [22] criteria to find out routes that can be active for longer duration without hampering the overall delay. We utilised the acknowledgment-based Q-routing technique to decide routing strategies that can adapt network changes in real time, allowing nodes to learn effective   routing policies.

A Q-value can be represented as Q (state1, action1) and describes the anticipated rewards of choosing action1 at state1. Here, every state signifies a probable destination network node d. Hence, after the elapse of sometime instants (RL stages), the Q-values can illustrate the network precisely. This indicates that at every state, the maximum Q-value represents actions (choosing suitable neighbour node) which consists of the best options. Every node x develops its perspective about various network states and their corresponding Q-values aimed at all state-action pair in its Q-table as shown in Table 1.

Table 1. Q Table for node x

	Neighbor1	Neighbor2	Neighbor3
Destination1	$Q^x(d1,n1)$	$Q^x(d1,n2)$	......
Destination2	$Q^x(d2,n1)$	$Q^x(d2,n2)$	......
Destination3	$Q^x(d3,n1)$	$Q^x(d3,n2)$	......

4.4. Dynamic Channel Selection (DCS) and Spectrum Sensing based on Q Learning

Q-learning is further explored to dynamically access the spectrum by introducing cognitive self-learning for unlicensed SUs. The key parts of the proposed algorithm consist of two modules: channel selection and channel access [20,21]. The algorithm selects the channel with highest Q value for CR node's transmission. The sequences of events are shown in Figure 5. At the starting of each slot, SU will first analyse all the free channels through spectrum sensing and estimate whether the appointed channel's status is BUSY for N quantity of slots. If entire amount of BUSY slots is not crossing number N, the SU believes that no PU is present on the present channel and spectrum access can be permitted else it switches the channel (considering the severe interference due to PU presence in the previous channel) and selects a channel with highest Q cost and update the Q cost of previous BUSY channel. Each SU determines its Q value as per the success/failure observed and then forecasts reward by using Q value. Such Q function is defined by $Q (st, ac)$ that shows the reward obtained by SU at state st having action ac.

$$Q_u(st,ac_j)[\ d+1]= Q_u(st,ac_j)[d] + \alpha\ [r\ (u,j)\ [d] + \gamma V(st)[d+1]- Q_u(st,\ ac_j)\ [d]] \qquad (6)$$

Where $Q_u(st,ac_j)$ specifies the Q function arrived when user u takes up the relevant action (choosing channel j). r(u, j) specifies reward attribute after SU u chooses channel j, V(st)[d+1] refer to assessment cost of subsequent cost function that is represented as following, where Ac is set of actions :

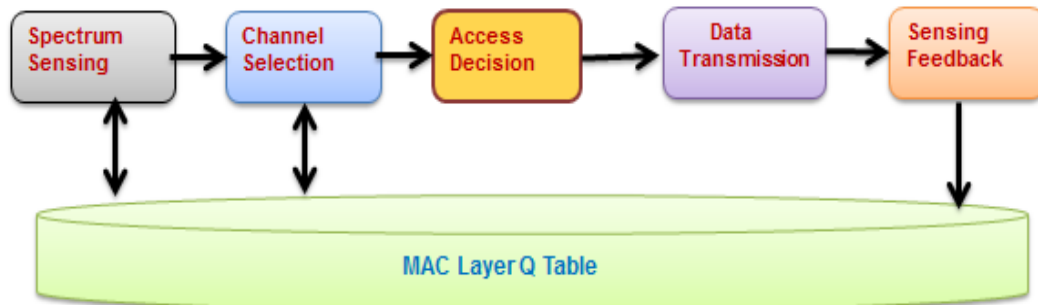$$V(st)\ [d+1] =\ maximum(\ Q\ (st\ [d+1],\ b)\ ) \qquad (7)$$
$$b\ \epsilon\ Ac$$

Figure 5. DCS and Sensing Mechanism in CR-IIAHSN

Once the SU decides to access the present channel via self-learning, it starts the transmission. During transmission, users acknowledge a successful transmission by ACK feedback information. A SU keeps on checking this feedback for transmission status. Based on the sensing feedback various rewards are assigned for SU and Q function is updated accordingly. The algorithm Steps are as follows:

1) Initialise SU's Q cost and relevant constraints
2) Select a channel arbitrarily for each SU such that SU can readily access this channel
3) Analyze if the total number of BUSY slots at present channel ch surpasses the threshold N. If yes, Step 4, else Step 5
4) Select the state for SU through highest Q cost and whose BUSY slot number is less than the given threshold N
5) Perform channel access of ch for SU

6) Investigate present user's channel position
7) Adjust Q value

4.4.1. Reward Assignment for DCS and Spectrum Sensing

The reward assignment mechanism for SU is presented in Figure 6 followed by reward values for various scenarios.

I. SU-PU interference: If PU is already present in the current channel and SU selects it for transmission, then an expensive penalty value of −11 is allocated.

II. SU-SU interference: If a packet experiences a clash with one more simultaneous SU transmission, then a penalization value of −3 is enforced.
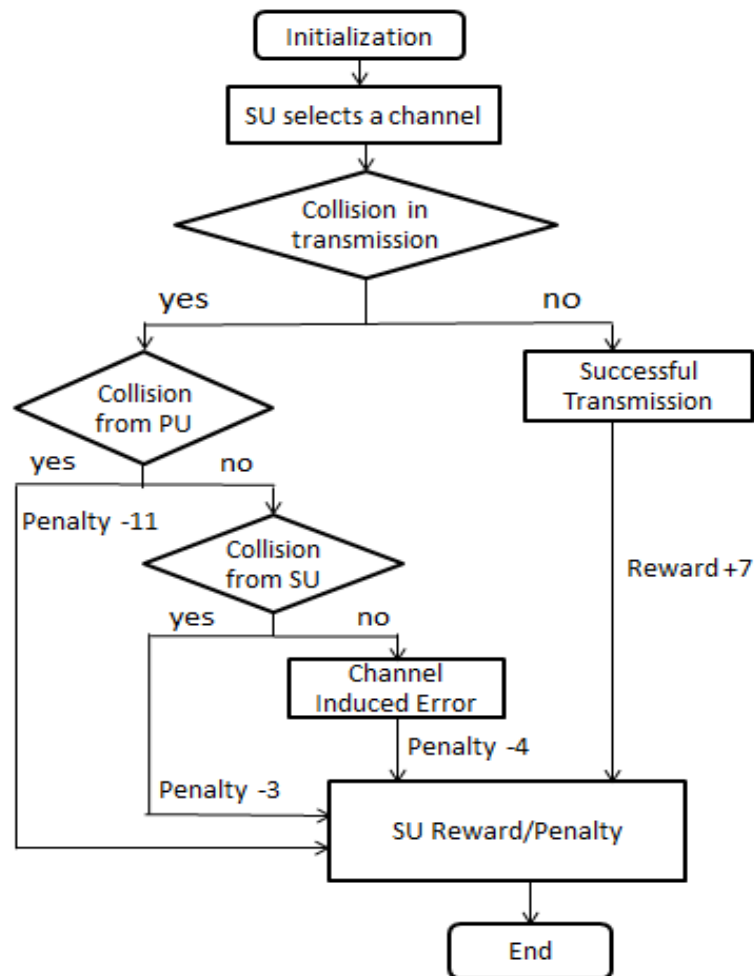
Figure 6. SU Reward Assignment in CR-IIAHSN

III. Channel Stimulated Errors: The intrinsic unpredictability in the wireless medium affects in fading that is caused by multipath reflections. Such an event is assigned a penalty of −4.

IV. Successful Transmission: If any of the earlier conditions are not found in the provided SU transmission slot and the packet is transmitted successfully between sender and receiver, then a reward of +7 is allocated.

5. RESULT ANALYSIS

To estimate the significance of the industrial internet of cognitive radio scenario, an area of 1300 square meters with 6 channels shared among 50 SUs and 5 PUs is considered. An Omnidirectional antenna is assumed with 2-ray ground propagation model. Here, we analysed network and MAC layer performance metrics. The routing performance is compared against the traditional Ad-hoc On-demand Distance Vector (AODV) [23] routing. Figure 7and Figure 8 depicts that SUs' performance metrics like aggregated packet delivery ratio is improved by 12-15% [24] and end-to-end network delay is reduced by 8-12% [25] in our design as compared to AODV based routing. This is achieved by applying RLCLD routing technique, which is a part of our cross-layer design framework. As displayed in Figure 8, the network behaviour with the proposed protocol is influenced by the initial channel state. Hence, the AODV protocol display an improved packet delivery ratio for a pause time of 125 sec. Channel state information with reinforcement learning based technique enables the CR secondary users to sense unoccupied spectrum and thereby improves the packet delivery ratio of the system. This also reflects in end-to-end delay of the network and hence a cross-over is displayed in the performance curves.
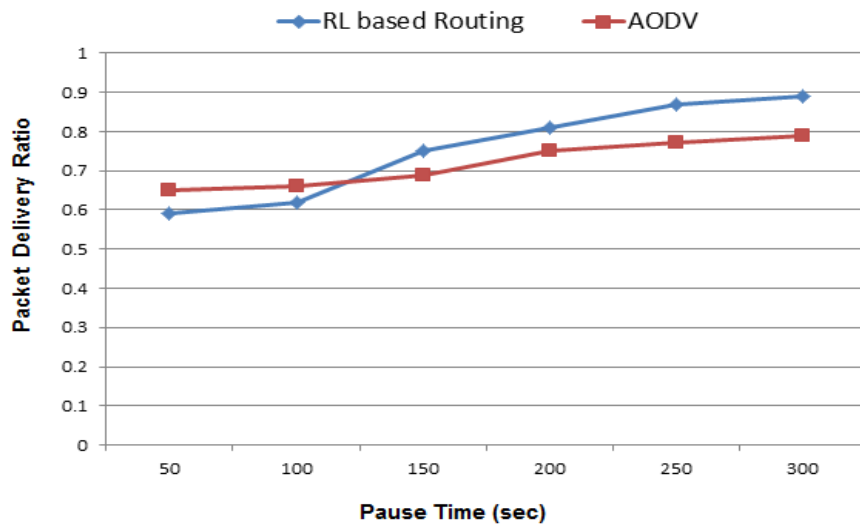

Figure 7. Performance Analysis of Packet Delivery Ratio
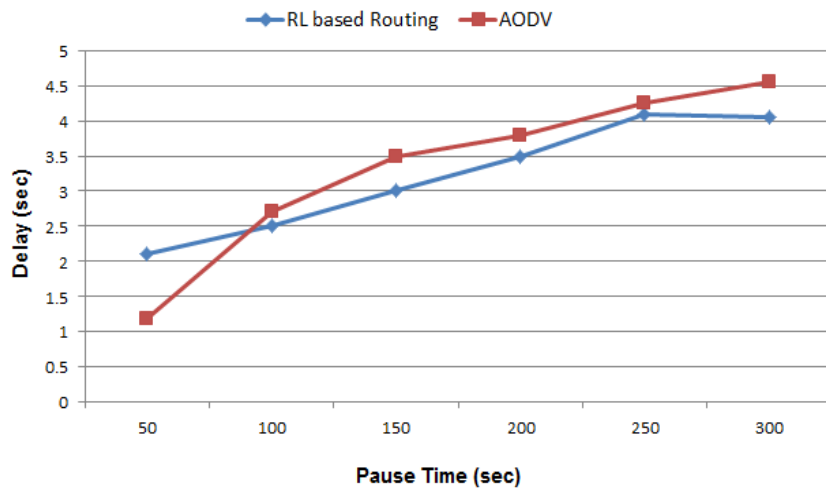

Figure 8. Performance Analysis of end-to-end network Delay
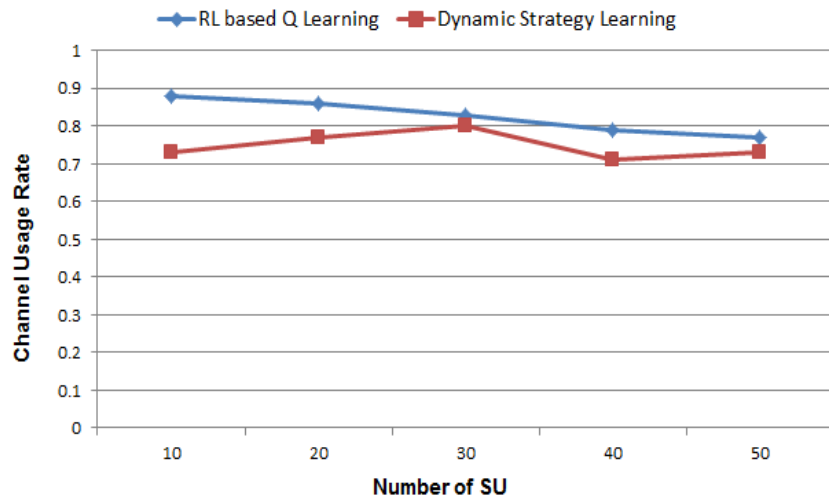
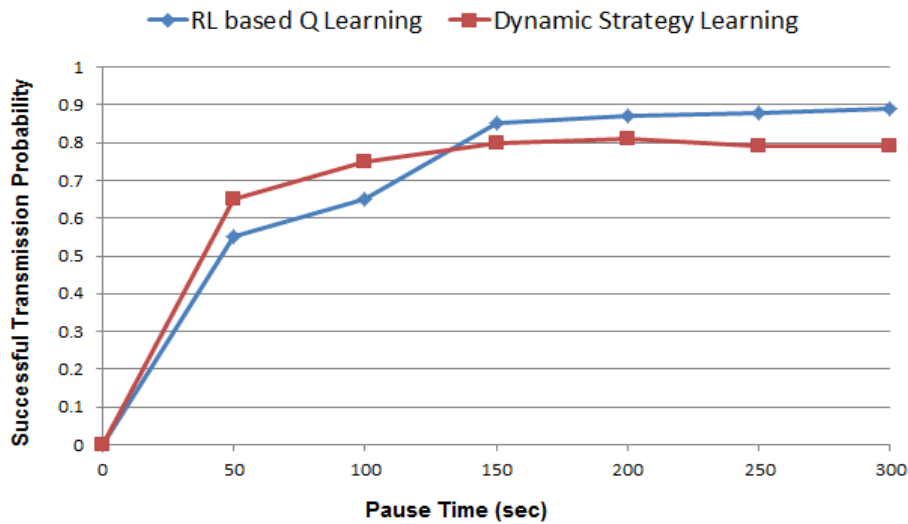Figure 9. Comparison of Channel Usage Rate

Figure 10. Successful Transmission Probability

On a similar line, we further compared our technique with the Dynamic Strategy Learning approach (DSL) [26]. It is observed that, even with the varying number of SUs, our RL based approach achieves a higher channel usage rate up to 20% and 6-12% higher successful transmission probability[13]than DSL as shown in Figure 9and Figure 10. As different form the performance of AODV routing, the DSL approach display reduced channel usage rate with increase in the number of secondary users. The proposed RL based Q learning display stable channel usage rate with increase in the number of secondary users with dynamic channel selection strategy. However, the successful transmission probability with RL based Q learning display saturation curves with increased pause time as the primary users may reuse the channel that is selected by the secondary users. However, with an increase in pause time the RL based Q learning display stable performance.
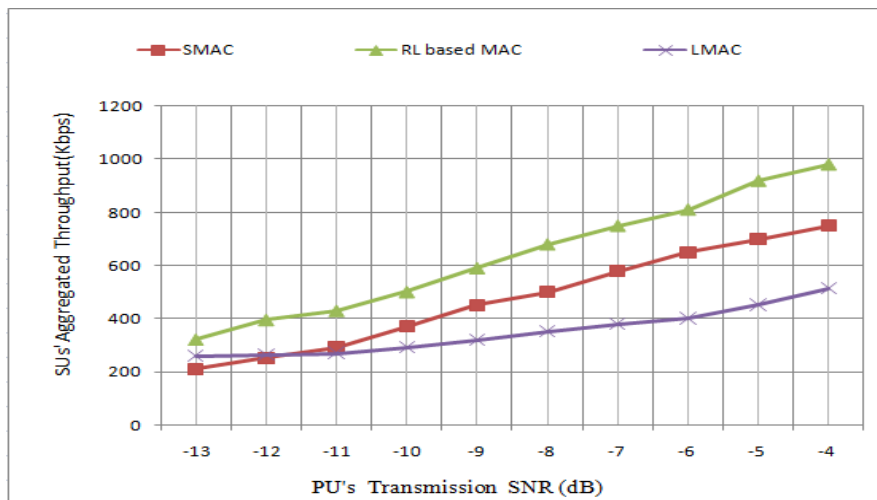
Figure 11. Secondary Users (SU) Aggregated throughput for the transmitted SNR
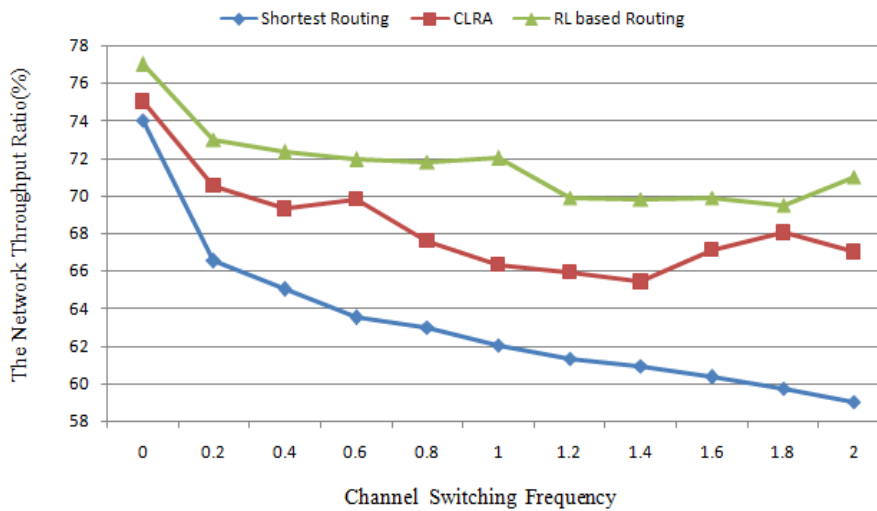

Figure 12. Performance Analysis of the Network Throughput with respect to Channel switching Frequency

We further investigated the effect of transmission SNR from PU on the SU's overall throughput and observed the performance. We compared our technique with two existing cross-layer MAC protocols which are large-scale backoff based MAC (LMAC) and small-scale backoff based MAC (SMAC) [27] and presented the comparative analysis of three protocols as depicted in Figure 11. It is observed that the secondary user's overall throughput improves when there is a surge of primary users signal to noise ratio as perceived by secondary users. Such phenomena are fairly contradicting with the conventional wireless concept, as in general, throughput should decrease as interference upsurges. Nevertheless, in these circumstances, whenever, primary users signal-to-noise ratio rises, it turns out handy for secondary users to sense activity of primary users, which eventually improves secondary users throughput. Here, the RL-based MAC outperforms the LMAC (up to 35%) and SMAC (up to 25%) method [3] in terms of aggregated throughput. Another experimental analysis as revealed in Figure 12, shows the impact on network throughput with channel switching frequency. It is observed for secondary users channel switching frequency, both shortest route-based routing and Cross-Layer Routing Algorithm

(CLRA) [28] exhibit a rapidly declining network throughput ratio as compared to our proposed RL-based routing.

6. CONCLUSION

This paper has discussed a design framework based on cross-layer techniques, that considered various functionalities of medium access control and network layers to enhance the operational cost of cognitive radio in industrial internet of ad-hoc sensor networks. This sets up a fair idea related to multiple potential cross-layering combinations for a cognitive radio node. The reinforcement learning-based Q learning algorithm for routing and dynamic channel access displays an improved performance in comparison to the shortest path in cognitive radio enabled on-demand routing and the dynamic strategy learning approach of spectrum access.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES

[1] Feng Li, & Kwok-Yan, (2018) "Q-Learning-Based Dynamic Spectrum Access in Cognitive Industrial Internet of Things", Mobile Networks and Applications, vol. 23, pp1636–1644.

[2] Md Sipon Miah, Michael Schukat and Enda Barrett.,(2021)"A throughput analysis of an energy efficient spectrum sensing scheme for the cognitive radio based Internet of things", J Wireless Com Network (2021) 2021:201. https://doi.org/10.1186/s13638-021-02075-2

[3] Muteba, K. F., Djouani, K., & Olwal, T. O., (2020) "Deep Reinforcement Learning Based Resource Allocation for Narrowband Cognitive Radio-IoT Systems", Procedia Computer Science, vol. 175, pp315-324.

[4] Singhal, C., & Rajesh, A., (2020) "Review on cross layer design for cognitive adhoc and sensor network", IET Communications, vol. 14, no. 6, pp897-909.

[5] Sheetal Naikwadi , and Akanksha Thokal .,(2022) "Cooperative Spectrum Sensing for Cognitive Radio", International Journal of Progressive Research and Engineering, Vol.3,No.04,April 2022

[6] Abbas, N., Nasser, Y., & Ahmad, K. E., (2015) "Recent advances on artificial intelligence and learning techniques in cognitive radio networks", EURASIP Journal on Wireless Communications and Networking, vol. 2015, no. 1, pp1-20.

[7] Yau, K. L. A., Poh, G. S., Chien, S. F., & Al-Rawi, H. A., (2014) "Application of reinforcement learning in cognitive radio networks: Models and algorithms", The Scientific World Journal, vol. 2014, no. 209810, pp1-23.

[8] X. Li, J, Fang, W. Cheng, & H. Duan, (2018) "Intelligent power control for spectrum sharing in cognitive radios: A deep reinforcement learning approach", IEEE Access, vol. 6, pp. 25463–25 473.

[9] M. C. Hlophe . & B. T. Maharaj, (2019) "Spectrum Occupancy Reconstruction in Distributed Cognitive Radio Networks Using Deep Learning", IEEE Access, vol. 7, no. 2, pp14294 - 14307.

[10] Ganesh Babu, R., & Amudha, V., (2019) "A survey on artificial intelligence techniques in cognitive radio networks", In Emerging technologies in data mining and information security, Springer, Singapore, pp99-110.

[11] Junaid Q, & Jones M, (2016) "Artificial intelligence based cognitive routing for cognitive radio networks", Artificial Intelligence Review, vol. 45, no. 1, pp1-25.

[12] Al-Rawi, H. A., Yau, K. L. A., Mohamad, H., Ramli, N., & Hashim, W., (2014) "Reinforcement learning for routing in cognitive radio ad hoc networks", The Scientific World Journal, vol. 2014, no. 960584, pp1-22.

[13] Felice, M. D., Chowdhury, K. R., Wu, C., Bononi, L., & Meleis, W., (2010). "Learning-based spectrum selection in cognitive radio ad hoc networks", In International Conference on Wired/Wireless Internet Communications, pp133-145.

[14] Chitnavis, S., & Kwasinski, A. (2019) "Cross layer routing in cognitive radio networks using deep reinforcement learning", In 2019 IEEE wireless communications and networking conference (WCNC), pp1-6.

[15] Wu, C., Chowdhury, K., Di Felice, M., & Meleis, W., (2010) "Spectrum management of cognitive radio using multi-agent reinforcement learning", In Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Industry track, pp1705-1712.

[16] Amulya, S., (2021), "Survey on Improving QoS of Cognitive Sensor Networks using Spectrum Availability based routing techniques", Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no. 11, pp2206-2225.

[17] Nasir, Y. S., & Guo, D., (2019) "Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks", IEEE Journal on Selected Areas in Communications, vol. 37, no. 10, pp2239-2250.

[18] Vij,Sahil. (2019).,"Cross-layer design in cognitive radio networks issues and possible solutions", 10.13140/RG.2.2.15041.71520.

[19] Jang, S. J., Han, C. H., Lee, K. E., & Yoo, S. J., (2019) "Reinforcement learning-based dynamic band and channel selection in cognitive radio ad-hoc networks", EURASIP Journal on Wireless Communications and Networking, vol. 2019, no. 1, pp1-25.

[20] Hema Kumar Y, Sanjib Kumar D, & Nityananda S, (2017) "A Cross-Layer Based Location-Aware Forwarding Using Distributed TDMA MAC for Ad-Hoc Cognitive Radio Networks", Wireless Pers Commun, vol. 95, pp4517–4534.

[21] Deepti S, & Rama Murthy G, (2015) "Cognitive cross-layer multipath probabilistic routing for cognitive networks", Wireless Netw, vol. 21, pp1181–1192.

[22] Du, Y., Xu, Y., Xue, L., Wang, L., & Zhang, F., (2019) "An energy-efficient cross-layer routing protocol for cognitive radio networks using apprenticeship deep reinforcement learning", Energies, vol. 12, no. 14, pp2829.

[23] Singh, J., & Mahajan, R., (2013) "Performance analysis of AODV and OLSR using OPNET", Int. J. Comput. Trends Technol, vol. 5, no. 3, pp114-117.

[24] Do, V. Q., & Koo, I., (2018) "Learning frameworks for cooperative spectrum sensing and energy-efficient data protection in cognitive radio networks", Applied Sciences, vol. 8, no.5, pp1-24.

[25] Safdar Malik, T., & Hasan, M. H., (2020) "Reinforcement learning-based routing protocol to minimize channel switching and interference for cognitive radio networks", Complexity, vol. 2020, no. 8257168, pp1-24.

[26] H.-P. Shiang , & M. Van Der Schaar, (2008) "Queuing-based dynamic channel selection for heterogeneous multimedia applications over cognitive radio networks", IEEE Transactions on Multimedia, vol. 10, no. 5, pp896–909.

[27] Shaojie Z, & Abdelhakim S.H, (2016) "Cross-layer aware joint design of sensing and frame durations in cognitive radio networks", IET Commun., vol. 10, no. 9, pp1111–1120.

[28] Honggui, W. (2015) "Cognitive radio system cross-layer routing algorithm research", International Journal of Future Generation Communication and Networking, vol. 8, no. 1, pp237-246.

AUTHORS

Ms. Chetna Singhal received her BTech and MTech degrees in Electronics and Communication Engineering from Kurukshetra University, India in 2008, and 2010 respectively and currently pursuing her Ph.D. from the School of Electronics Engineering, VIT University, Vellore. She has worked with global leaders like Aricent Technologies and Nokia Siemens Network. She also served as a Lecturer at NCIT College Panipat and an Assistant Professor at SKIT College, Bangalore. She is a recipient of 'Sammana Patra' for outstanding performance in education by Aggarwal Yuva Club, India, and qualified UGC NET all India examination. She secured various top positions at her graduate and PG level in Kurukshetra University. She has published papers in international conference proceedings and journals. Her current research interest is in the areas of wireless sensor networks, cross layer design, and cognitive radio technology.

Dr. Thanikaiselvan V received his Ph.D. degree in the field of Information security form the Vellore Institute of Technology (VIT), Vellore, Tamilnadu in the year 2014. He received MTech in Advanced Communication Systems from SASTRA University, Thanjavur, in 2006 and B.E in Electronics and Communication Engineering from Bharathidasan University, Trichy, in 2002. He joined Vellore Institute of Technology in 2006 as a Lecturer and currently working as an Associate Professor in the Department of Communication Engineering under the School of Electronics Engineering, Vellore Institute of Technology, Vellore. His teaching and research interest includes Digital communication, Wireless communication, Digital signal and Image Processing, Wireless Sensor Networks and Information Security. He has published more than 75 research articles. Currently, he is guiding 4 Ph.D. candidates in the areas of Information Security and Digital Image Processing. He is an active IEEE member and reviewer for Elsevier, Wiley and springer journals.