



Reinforcement Learning Based Q-Routing: Performance Evaluation on Cognitive Radio Network Topologies

Anal Paul¹ · Santi Prasad Maity²

Accepted: 7 February 2022 / Published online: 12 March 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The present work explores the scope of cognitive radio networks (CRNs) to support the spectrum demand on future diverse application specific wireless services supported by long-range multi-hop relay-based transmission. The essential requirement is dynamic route establishment where the traditional non-adaptive routing algorithms are not sufficient to support it in dynamic wireless networks. To address this issue, reinforcement learning (RL) based self-adaptable Q-routing in the CRN with different network topologies is explored with an objective to obtain the best performance in the secondary throughput. The traditional Q-routing policy offers the worst-case time-complexity of $\mathcal{O}(ed)$ to find the optimal route. The excessive time complexity for the route establishment reduces the actual secondary data transmission time, this also increases the power consumption during transmission. To reduce the time complexity of Q-routing, the proposed work invokes a strict upper bound on the network topology, which reduces the time complexity to $\mathcal{O}(N^2)$ and increases the secondary network throughput. The further study also reports that the complexity can be reduced to $\mathcal{O}(N)$ if the distributed relays follow a tree topology in the network configuration. A large set of simulation results show that the tree structure-based Q-routing in CR enhances the network throughput by $\sim 6.83\%$, $\sim 7.21\%$ over the other network topologies.

Keywords Cognitive radio networks · Underlay mode · Multi-hop relay network · Reinforcement learning · Q-routing.

✉ Santi Prasad Maity
santipmaity@it.iiests.ac.in

Anal Paul
anal.paul@klh.edu.in

¹ Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, Telangana 500075, India

² Department of Information Technology, Indian Institute of Engineering Science and Technology, Shibpur 711103, India

1 Introduction

The promise and the potential of the Internet of Things (IoT) demands billions of intelligent devices to be connected worldwide to the internet [1]. Each IoT object accumulates the intended information, processes and shares the same over its network to offer the quality of life-enhancing services [2]. According to the forecast, 20 billion IoT-enabled devices to be interconnected that will produce a market value of US\$14.4 trillion across all kinds of industries [3]. Consequently, a huge amount of information to be generated and, more importantly, they need to be processed and transmitted over a wireless medium seamlessly and efficiently [4]. Hence, shortly spectrum regulatory bodies will face a massive demand of allocation in spectrum to support the intelligent IoT infrastructure [5]. Here, it is worth mentioning that the conventional static spectrum allocation policy already initiates the problem of spectrum scarcity to support the extensive growth in the wireless applications [6].

Cognitive radio (CR) is a promising technology to mitigate the spectrum scarcity problem by exploiting the opportunistic access of the licensed spectrum [7–9]. In the CRN (CRN), there are two types of users, namely the primary users (PUs) (i.e., licensed users of the spectrum) and the secondary users (SUs) (i.e., unlicensed users or cognitive users) [10, 11]. In the underlay CRN mode, both PUs and SUs simultaneously transmit their respective data over the same spectrum in a power controlled manner so that SUs must maintain a limit of interference power level to guarantee the high quality-of-service (QoS) of PU network [12]. The underlay mode of communication is recently explored and investigated in the CR enabled IoT framework to improve the network throughput [13, 14]. The underlay paradigm allows IoT devices to be operated as a secondary transmitter [13]. The massive deployment of IoT devices also needs an expansion its network coverage [14]. However, the long-range direct transmission from the secondary IoT transmitter (SU_T) to the secondary receiver (SU_D) requires a high transmission power which increases the interference power level to the PU [15]. It is worth mentioning that geographically distributed low-powered battery sources drive IoT sensors [16, 17]. Sometimes the presence of multiple cooperative SUs, over a particular area, forms a multi-hop relay assisted CRN [18–21]. Relay-based network minimizes the transmission power consumption through the optimal relay selection (i.e., routing), which significantly reduces the interference power level, extends the coverage area, and enhances the network lifetime [16, 18, 19, 22–30]. In the energy-constrained CRN, several reports explore the different aspects of relay power allocation and their optimal routing strategies to enhance the performance [18, 31–34].

1.1 Routing in Cognitive Radio Networks

The work reported in [18] proposed an optimal power allocation scheme followed by an optimal relay selection strategy that minimized the actual transmission power consumption and enhanced the secondary network lifetime. The work [18] applied the conventional Bellman-Ford (BFD) and Dijkstra's routing algorithms to obtain the best route for secondary data transmission. It is seen that routing becomes more challenging in the dynamic channel states due to PU's uncertain activities, variations in channel gain and availability, and mobility in the network [22]. The effectiveness of the traditional routing algorithms [18, 35] in the above-stated scenario is found to be inefficient, as the prior route set up policy is non-adaptable with the channel dynamicity [36]. Several self-adaptable routing

strategies are explored in the CRN to achieve the best performance during dynamic channel states [31, 36–41].

The specific searching policies and self-adaptability associated with the dynamic routing algorithm increase the computational time complexity. The algorithm proposed in [40] requires the worst-case time complexity of $\mathcal{O}(N^4)$. In contrast, the genetic algorithm-based adaptive routing strategy [39] increases the runtime complexity to $\mathcal{O}(N^6)$ in multi-hop CRN, where N denotes the number of relay nodes. It is reported that the BFD algorithm has the worst-case time complexity of $\mathcal{O}(N^3)$ [35], but the BFD algorithm suffers from its static route set up policy. The authors in [41] developed two different routing strategies to minimize the average transmission delay in the CR framework. However, the reported work did not investigate the time complexity of the proposed routing algorithms, which is an essential issue in the CRN [41]. The authors in [42] modified the low-energy adaptive clustering hierarchy (LEACH) protocol to ECO-LEACH for the routing in wireless sensor networks. The clustering process associated with LEACH is time-consuming and increases the optimal path setup time. The LEACH routing technique is also investigated in the CRNs from the aspects of energy conservation and security [43]. Reinforcement learning (RL) based Q-routing in CRN is explored recently due to its less time complexity, presence of tuneable control parameters, ease of implementation, fast convergence procedure, and self-learning capability in the dynamic CRN environment [36, 37]. Saleem et al. applied the typical Q-routing due to its self-adaptive nature to maximize the CRN throughput [36]. The Q-routing algorithm requires the run time complexity of $\mathcal{O}(ed)$ [44] to find the shortest paths in the deterministic network topology, the symbols ' e ', and ' d ', denote the number of links, and the maximum depth of the optimal path.

1.2 Cross-Layer Routing Algorithm

Recently, the use of a cross-layer design strategy to increase CRNs' performance receives a lot of attention [45–49]. Several issues such as channel diversity, dynamic route discovery with spectrum selection, user mobility, and so on, are considered during routing protocol using a cross-layer architecture in CRNs. The idle channel for each relay is identified in the route establishment method via cross-layer design among the physical layer, MAC layer, and network layer, taking into account of opportunistic spectrum access strategy [45]. The multi-radio link quality source routing (MR-LQSR) protocol works with the layers as mentioned earlier to determine the best channel assignment and route selection strategy in cross-layer design [48]. In cross-layer routing design, every layer shares different information among others, e.g., the collaboration of the physical and MAC layers can provide crucial information regarding spectrum sensing to the routing algorithm, which helps to make routing decisions. It is an excellent paradigm for balancing guidance and resource performance across layers which improve the data rate with high spectrum efficiency and reduces the data conflict error [34]. The mobility among the SUs or PUs influences the routing protocol to identify the idle channel, with the lack of information SU faces problem while transmitting or relaying the data. As a result, SU must be capable of accommodating the dynamic change in spectrum usage by PU. Therefore, the dynamic routing strategy offers an important role for the cross-layer design with a multi-channel and multi-path concept. Each relay can select the best channel for relaying the data based on the available local information. For designing the cross-layer routing protocol, the Q-learning algorithm is found to be suitable and explored in [33, 50]. In the present work, the Q-routing algorithm

is exploited intensively on different network topologies to find the best network layout, increasing CRNs throughput and energy efficiency.

1.3 Scope and Contributions

The literature review indicates a good deal of works on routing in CRN, however, none of those works considers the issue of run time complexity in the process of secondary data transmissions. All the routing algorithms consume some time (Δt) during the optimal path set up. This Δt is dependent on the respective routing algorithm and the network topology. Consideration of the Δt from the available data transmission time (T) influences the required relay power and effects in actual network throughput. The present work considers the RL-based dynamic Q-routing due to its low time complexity over the other dynamic algorithms for the secondary data transmission.

The proposed work implies some unique properties on the network topology to reduce the worst-case time complexity of Q-routing from $\mathcal{O}(ed)$ to $\mathcal{O}(N^2)$. The work further investigates and reduces the routing complexity to $\mathcal{O}(N)$ if the network topology follows a tree-like structure [31]. The present work maximizes the secondary network throughput in the underlay CRN mode under the constraints of a certain level of PU interference and a power budget for the secondary transmission. A closed-form expression is derived for the exact power allocation to the relay to maximize the network throughput. The optimal path setup time is considered and deducted from the actual data transmission time to get more accurate power allocation to the relays. The multi-hop multi-path network is reformed to a tree-like topology. Several constraints are imposed on the network to minimize the run time complexity of the Q-routing algorithm to obtain better CRN performance.

The brief contributions of the work are as follows:

- A simple tree structure formation algorithm is proposed, capable of converting any grid-like multi-hop network to a tree topology.
- The proposed tree structure topology for the multi-hop CRN significantly reduces the run time complexity of the Q-routing algorithm to $\mathcal{O}(N)$ which increases the network throughput and reduces the relay power consumption over the other network topologies.

The rest of the paper is organized as follows: the proposed system model is presented in Sect. 2, while the problem formulation and the proposed solution are stated in Sect. 3. Numerical results and analysis are demonstrated in Sect. 4 to highlight and compare the efficiency of the tree topology-based network. Finally, the concluding remarks are made in Sect. 5.

2 System Model

The proposed system model is depicted in Fig. 1. It contains PU and SU networks operated on the same spectrum. The PU network consists of a PU transmitter (PU_T) and a PU receiver (PU_R), while the SU network includes a SU transmitter (SU_T) and a SU receiver (SU_D), and N number of decode-and-forward (DF) relays. The relays are denoted as R_1, R_2, \dots, R_{N+1} , where R_1 indicates the SU_T and R_N represents the SU_D .

Figure 2 represents the proposed time frame structure of duration T operated in an underlay CRN mode. A suitable routing algorithm finds an optimal path in Δt time in the

Fig. 1 System model for CRN

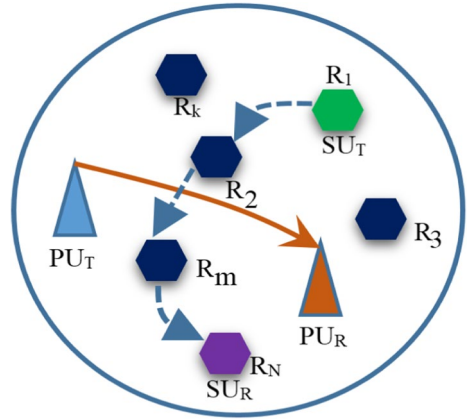
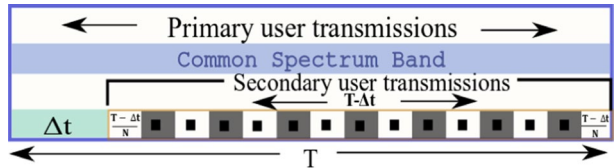


Fig. 2 Proposed timeframe structure



multi-hop CRN for relaying the data from $SU_T - SU_D$. This Δt is deducted from the actual data transmission time. The remaining time slot $T - \Delta t$ is equally partitioned into the N number of TDMA slots and is accessed by the N number of different relays.

2.1 Signal Modelling

The wireless channels between $PU_T - PU_R$ link and the other links among the relays are modelled as Rayleigh flat fading channels. The signal transmitted from SU_T is received by the relay R_m and it forwards the received signal to R_k . The power P_m is consumed by the R_m during this relaying process over the instantaneous channel Φ_m . The average noise power over the link $R_m - R_k$ is denoted by η_m . The average SNR between $R_m - R_k$ is represented by $\bar{\gamma}_{m,k}$ and is expressed as

$$\bar{\gamma}_{m,k} = \frac{(P_{m,\Phi_m})d_{\Phi_m}^{-\theta_{\Phi_m}}}{\eta_m} \tag{1}$$

The symbols d_{Φ_m} and θ_{Φ_m} denote the distance and the path loss exponent of the $R_m - R_k$ link, respectively. Due to the licensed user of the spectrum, PU does not impose any constraints on its own transmission while SUs maintain an interference power level (I_{th}) to PU_R . The interference constraint on SU transmission is expressed as

$$\sum_{m=1}^N P_{m,\Phi_m} d_{\Phi_m}^{-\theta_{\Phi_m}} \leq I_{th} \tag{2}$$

In the energy-constrained CRN, SU must minimize the power consumption during relaying to extend the CRN lifetime. Hence, a constraint on the total power consumption (P_T) in the SU network is imposed.

$$\sum_{m=1}^N P_{m,\Phi_m} \leq P_T \tag{3}$$

The secondary network throughput between $R_m - R_k$ is represented by \mathcal{T}_{out_m} and is expressed as

$$\mathcal{T}_{out_m} = \left(\frac{T - \Delta t}{NT} \right) \log_2 \left(1 + \frac{(P_{m,\Phi_m}) d_{\Phi_m}^{-\theta_{\Phi_m}}}{\eta_m} \right) \tag{4}$$

3 Problem Formulation and Proposed Solution

This section aims to obtain the optimal values of P_{m,Φ_m} for maximizing the overall secondary network throughput under the constraints of PU interference and secondary transmission power budget.

3.1 Problem Formulation

The objective function \mathcal{F} is defined as

$$\begin{aligned} \mathcal{F}(P_{m,\Phi_m}) &= \max_{P_{m,\Phi_m}} \sum_{m=1}^N \mathcal{T}_{out_m} \\ \text{s.t. } i) \quad &\sum_{m=1}^N P_{m,\Phi_m} d_{\Phi_m}^{-\theta_{\Phi_m}} \leq I_{th}, \quad ii) \quad \sum_{m=1}^N P_{m,\Phi_m} \leq P_T \end{aligned} \tag{5}$$

3.2 Proposed Solution

Lagrange multiplier method and Karush–Kuhn–Tucker (KKT) conditions are applied to solve the proposed optimization problem. The Lagrangian function $\mathcal{L}(P_{m,\Phi_m})$ for the stated problem is represented as

$$\mathcal{L}(P_{m,\Phi_m}) = \sum_{m=1}^N \mathcal{T}_{out_m} + \mu_1 \left(P_T - \sum_{m=1}^N P_{m,\Phi_m} \right) + \mu_2 \left(I_{th} - \sum_{m=1}^N P_{m,\Phi_m} d_{\Phi_m}^{-\theta_{\Phi_m}} \right) \tag{6}$$

where the symbol μ_1 and μ_2 are the Lagrange multiplier. Now, $\frac{\delta \mathcal{L}}{\delta P_{m,\Phi_m}}, \frac{\delta \mathcal{L}}{\delta \mu_1}, \frac{\delta \mathcal{L}}{\delta \mu_2}$ are calculated as

$$\frac{\delta \mathcal{L}}{\delta P_{m, \Phi_m}} = \frac{T - \Delta t}{NT} \frac{d_{\Phi_m}^{-\theta_{\Phi_m}}}{\eta_m + P_{m, \Phi_m} d_{\Phi_m}^{-\theta_{\Phi_m}}} \tag{7}$$

$$\frac{\delta \mathcal{L}}{\delta \mu_1} = P_T - \sum_{m=1}^N P_{m, \Phi_m} \tag{8}$$

$$\frac{\delta \mathcal{L}}{\delta \mu_2} = I_{th} - \sum_{m=1}^N P_{m, \Phi_m} d_{\Phi_m}^{-\theta_{\Phi_m}} \tag{9}$$

There are two solutions (10) and (11) obtained for P_{m, Φ_m} , where $C = \sum_{m=1}^N d_{\Phi_m}^{-\theta_{\Phi_m}} \eta_m - \eta_m d_{\Phi_m}^{-\theta_{\Phi_m}}$, and $\mathcal{Y} = \eta_m d_{\Phi_m}^{-\theta_{\Phi_m}} - \sum_{m=1}^N d_{\Phi_m}^{-\theta_{\Phi_m}} \eta_m$. The solution (10) can be used, provided that it satisfies the constraint in (2). On the other hand, solution (11) can be used, provided that it maintains the constraint (3).

$$P_{m, \Phi_m} = \frac{C - \sqrt{\mathcal{Y}^2 - 4(d_{\Phi_m}^{-\theta_{\Phi_m}})^3 \eta_m P_T}}{2N(d_{\Phi_m}^{-\theta_{\Phi_m}})^2} \tag{10}$$

$$P_{m, \Phi_m} = \frac{C - \sqrt{\mathcal{Y}^2 - (2d_{\Phi_m}^{-\theta_{\Phi_m}})^2 \eta_m I_{th}}}{2N(d_{\Phi_m}^{-\theta_{\Phi_m}})^2} \tag{11}$$

3.3 RL Based Q-routing

Q-Routing is an RL-based self-adaptive algorithm. To obtain the optimal path from source to destination, each relay node in the network participates in the path establishment procedure. Relay node R_m searches the next suitable hop R_k for forwarding the data using the available local information at R_m . The Q-function $Q(s, a)$ in the Q-routing algorithm is defined as

$$Q_{t+1}^m(s_t^m, a_t^m) \leftarrow (1 - \alpha)Q_t^m(s_t^m, a_t^m) + \alpha \left(\alpha_{t+1}^m(s_t^m, a_t^m) + \beta \max_{k \in a} Q_t^{a_t^m}(s_t^m, k) \right) \tag{12}$$

The symbols ‘ s ’ and ‘ a ’ denote the set of all relay nodes and the set of all existing links (see Fig. 3a) in the network, respectively. In the above equation, the state transition function $Q_{t+1}^m(s_t^m, a_t^m)$ signifies that at the t^{th} instant the relay $s_t^m = R_m$ forwards the data to the next relay ‘ R_k ’ by selecting the link a_t^m , where $a_t^m \in a$. The reward function is denoted by $\alpha(s_t^m, a_t^m)$. The reward function for the proposed work is defined in Eq. (4). The symbols ‘ α ’ ($0 \leq \alpha \leq 1$) and ‘ β ’ ($0 \leq \beta \leq 1$) represent the learning rate and the discount factor in the reward function, respectively. Here, $\alpha = 0$ indicates that the agent does not learn anything, while $\alpha = 1$ signifies that the agent adapts only the most recent information. The optimal value of $\alpha = 1$ is possible in fully deterministic environments, but in stochastic problems, the α value decreases to 0 for converging the solution. However, if $\beta = 0$, then the system considers only the current reward and behaves like a greedy algorithm, leading to the

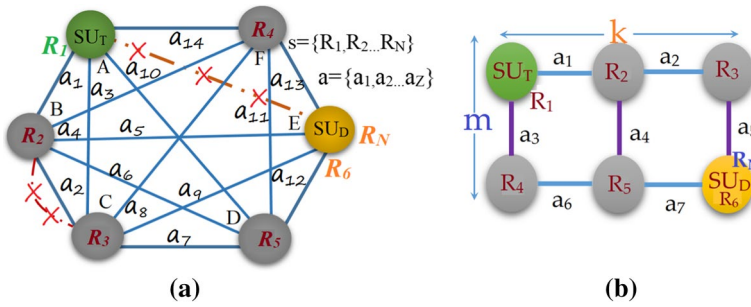


Fig. 3 a Strongly (fully) connected network b Reformed to grid network

optimal local solution. For $\beta = 1$, the system is always looking for long-term high reward, and the future reward cannot be estimated accurately, which reduces the system’s overall performance.

Exploration and Exploitation of the network topology are the two main critical functions in the Q-routing. In the exploration process, each relay node searches the alternative paths to obtain a better routing path over the current path. The information is gathered during the exploration phase and is used in future decisions, known as exploitation. Thus, the current number of links in the network topology is directly proportional to the full exploration’s minimal time requirement. The practical design of the network topology reduces the number of links. It decreases the number of exploration steps, resulting in a less time complexity solution for obtaining the optimal routing path.

3.4 Q-Routing Complexity for Strongly Connected Network

The proposed work explores the zero-initialized Q-routing strategy to support the dynamicity of the CRN. The terminology ‘zero initialized’ refers to the reward values of all the existing links initialized to zero. For any network topology, the worst-case time complexity of Q-routing is $\mathcal{O}(ed)$, where the symbols ‘d’, and ‘e’ represent the maximum depth and the total number of links in the network, respectively. The maximum depth (d) of the routing path from SU_T to SU_D is defined by the number of relays associated in the path and calculated as $d = \max_{R_m, R_k \in s} d(SU_T, SU_D)$.

Figure 3a depicts a strongly connected network topology with a consideration that there is no direct link between $SU_T - SU_D$. The topology does not contain any duplicate links, i.e. if $R_m \leftrightarrow R_k$, then $R_k \leftrightarrow R_m$. Thus, the possible number of links in the network (Fig. 3a) is $|\frac{N(N-1)}{2} - 1|$, and $\approx \frac{N(N-1)}{2}$. If the network has no duplicate links then Q-routing terminates after at most $\mathcal{O}(ed^2)$ steps. The maximum depth from the source to the destination nodes would be $d = N - 1$. Hence, the worst case time complexity of Q-routing algorithm becomes $\mathcal{O}\left(\frac{N(N-1)}{2}(N - 1)\right), \equiv \mathcal{O}\left(\frac{N^3 - 2N^2 + N}{2}\right)$, for very large N value the upper bound will be $\approx \mathcal{O}(N^3)$.

3.5 Reduced Time Complexity to $\mathcal{O}(N^2)$ of Q-Routing

The strongly connected network is reformed to a grid-world network form, as shown in Fig. 3b. In the grid-world topology, a node can communicate with a certain maximum

number of adjacent nodes. Thus a constant individual upper link bound is imposed on the network and denoted by the symbol ‘ \mathcal{B} ’. From the Fig. 3b, the maximum \mathcal{B} value is 2. The linear boundary constraints on the network implies $e \leq \mathcal{B}N$. Hence, in the worst-case $\mathcal{O}(ed)$ becomes $\mathcal{O}(\mathcal{B}N(N - 1))$. Here, \mathcal{B} is a constant term and the complexity is $\equiv \mathcal{O}(\mathcal{B}N^2 - \mathcal{B}N)$. For the large number of N , it can be written as $\mathcal{O}(N^2)$.

3.6 Reduced Time Complexity to $\mathcal{O}(N^{1.5})$

The rectangular grids are shown in Figs. 3b and 4a. The size of each grid is $m \times k$. If the each grid has N number of relays then $N = m \times k$. The number of maximum links in the grid is $e = 4mk - 2m - 2k$ and the possible depth in the worst case is $d = m + k - 2$. Hence, the required time complexity of $\mathcal{O}(ed)$ can be written as $\mathcal{O}\{(4mk - 2m - 2k)(m + k - 2)\}$, $\equiv \mathcal{O}(4m^2k - 2m^2 - 12mk + 4mk^2 + 4m + 4k)$, which is $\equiv \mathcal{O}(m^2k + mk^2)$. Now, if the grid is in square form then $m = k$, which implies $\mathcal{O}(m^3 + m^3)$, $\equiv \mathcal{O}(m^3)$. However, $N = m \times k$ is modified as $N = m^2$, then $m = \sqrt{N}$. Finally, the time complexity is obtained as $\mathcal{O}(\sqrt{N^3})$, $\equiv \mathcal{O}(N^{1.5})$.

3.7 Complexity Reduction to $\mathcal{O}(N)$

The worst-case complexity is further reduced. A constraint of linear total upper link bounds is imposed on the network topology. It implies that the total number of links linearly increases with the number of relays in the network. From Fig. 4a, it is observed that the constant individual upper bound is four (i.e., $\mathcal{B} = 4$). The grid-world network of Fig. 4a is converted to a tree structure (see Fig. 4b) without reducing any existing links. In the tree-topology, the depths of all the relays in a particular level are the same. It means that d remains constant. Thus, in the constant upper bound and the liner total upper bound, the $\mathcal{O}(ed)$ is $\equiv \mathcal{O}(e)$. Here, $e = 4mk - 2m - 2k$, $\equiv \mathcal{O}(4mk - 2m - 2k)$, it is written as $\mathcal{O}(mk)$, which implies the worst-case time complexity will be $\mathcal{O}(N)$.

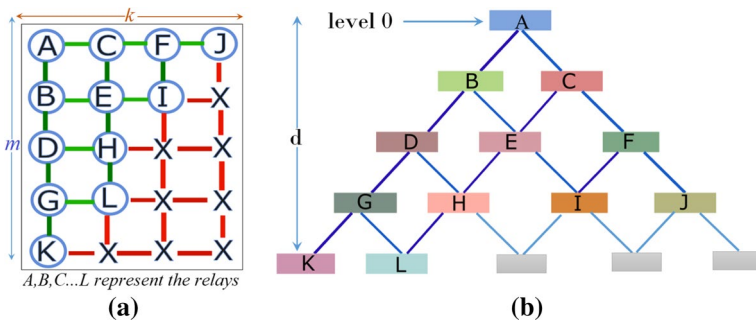


Fig. 4 a Grid-world network b Converted to tree-topology

Table 1 Q-routing execution time on different topologies

Number of relays	Fully connected	Rectangular grid	Square grid	Tree topology
12	12.405 ms	10.515 ms	9.117 ms	7.371 ms
25	24.612 ms	21.339 ms	18.213 ms	16.297 ms
50	36.237 ms	29.811 ms	25.671 ms	24.332 ms

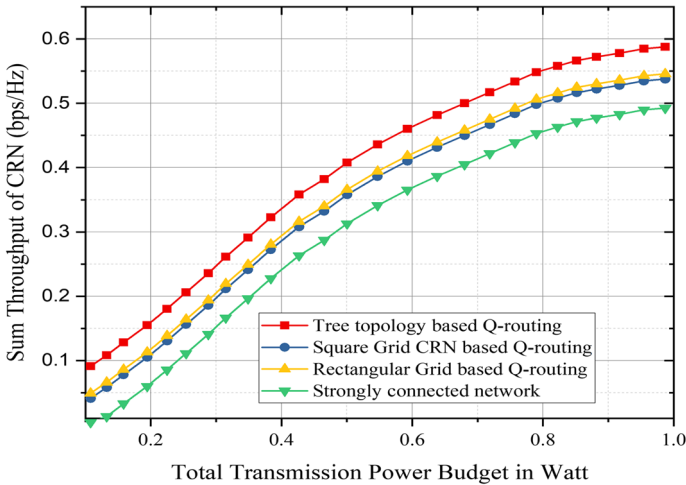


Fig. 5 Sum throughput of CRN vs maximum power budget

4 Numerical Analysis and Discussions

This section explains the simulation results for the performance analysis of Q-routing in different network topologies. Monte Carlo simulations of 10,000 times are performed to average the outcomes of each experiment to consider the wireless channels’ variability. Simulations are performed on 64-bit Matlab-2017, 2.13GHz Intel i5 processor with 8GB RAM through self-coding. Numerical values of the required system parameters in the simulation environment are considered as follows: $T = 100$ ms, $\max d_{\phi_m} = 10$ m, $\min d_{\phi_m} = 2.5$ m, $\min \theta_{\phi_m} = 2$, $\max \theta_{\phi_m} = 4$, $P_T = 1W$, $I_{th} = 1.25 W$, $\eta_m = 0.10 W$, $\alpha = 0.49$, $\beta = 0.58$ and $N = 36$.

The execution time of Q-routing on the different network topologies are shown in Table 1. It is clearly observed that tree structure based network minimizes the path establishment time over all other network topologies for the same number of relays in the network.

Figure 5 illustrates the simulation results for the sum throughput vs. maximum power allocation in the secondary network. It is observed from the figure that tree topology-based Q-routing outperforms all other network architectures. The throughput for the CRN is increased with the power level for all the network topologies. Here, ~ 38.86% increase on throughput is observed for tree topology-based network, while the total

power budget is increased from $0.2W$ to $0.35W$. When the power is increased from $0.85W$ - $1W$, the throughput is improved by only $\sim 6.27\%$. It is noted that the rate of throughput improvement is higher in the first case compared to the latter one. This happens due to the restriction of interference level at SU network. The tree topology-based Q-routing offers higher throughput at $P_T = 1W$ by $\sim 6.83\%$, $\sim 7.21\%$, and $\sim 18.11\%$ over the square grid, rectangular grid, and strongly connected network topologies, respectively. This is justified as Q-routing in tree topology requires less time to obtain its optimal path due to its less time complexity than all other cases. The less time needed in the path set up increases the data transmission period, which improves the throughput of the CRN.

Figure 6 demonstrates the performance comparison on the sum throughput vs. the interference power level. In this case, tree topology-based Q -routing outperforms all the other network topologies. It is observed from Fig. 6 that the tree network improves the throughput by $\sim 6.89\%$, and $\sim 52.89\%$ at $I_{th} = 0.25 W$ over the square and the rectangular grid based topologies, respectively. The self-adaptive Q-routing in all the different network topologies performs better than the typical BFD routing algorithm. The Q-routing in the strongly connected network, which is similar to BFD routing in terms of the worst-case time complexity of $\mathcal{O}(N^3)$, improves the throughput of the CRN by $\sim 37.21\%$ at $I_{th} = 0.75 W$ compared to the BFD algorithm.

Figure 7 demonstrates the performance analysis on the network throughput with the variation in the discount factor (β) in Q-routing. The throughput is increased in all the cases when β value is increased from low to high. But suddenly, after a particular value of β , the throughput is decreased. This is due to $\beta \rightarrow 0$, the system considers only the current reward and behaves like a greedy, leading to the optimal local solution. And $\beta \rightarrow 1$, the system always is looking for a long-term high reward, and the future reward can't be estimated accurately, which reduces the system's overall performance. It is noted from the figure that for different network topologies, there exists a unique β value for which the throughput is maximized. The optimal value of the β is dependent on the network topology and its size. For the tree-based network, the optimal β value is 0.62, while for the square grid-based network, the optimal β value is 0.56.

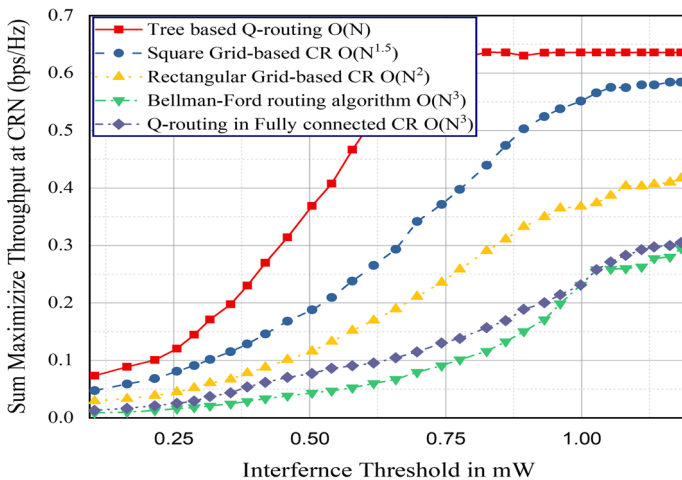


Fig. 6 Sum throughput of CRN vs interference threshold

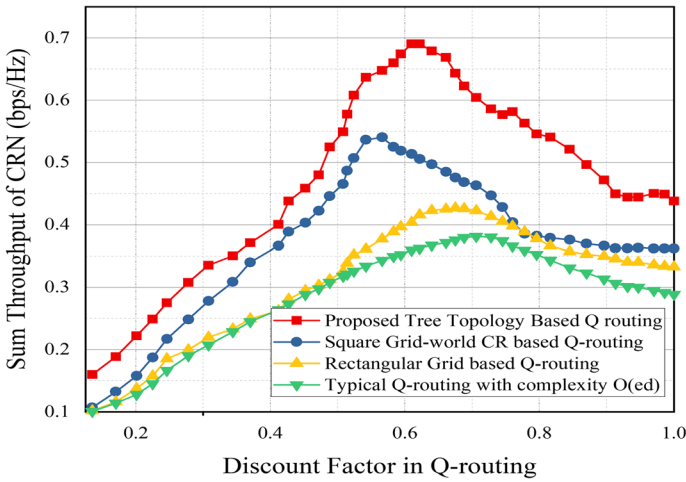


Fig. 7 Sum throughput of CRN vs discount factor β in Q-routing

Figure 8 illustrates the influence of the learning rate factor (α) in the required number of steps for convergence of Q-routing in the different network topologies. It is observed in Fig. 8 that initially, for the low value of α , the required number of steps are high for obtaining the minimal path. The initial increment in α value reduces the number of steps up to a limit. However, after a particular value of α , the required number of steps again increase. The value of $\alpha \rightarrow 0$ indicates that the agent does not learn anything, while $\alpha \rightarrow 1$ signifies that the agent adapts only the most recent information. The value $\alpha \rightarrow 1$ leads to the fact that network topology seems to be deterministic. It is observed that the tree structure-based CRN converges faster than the other network topologies. The faster convergence rate reduces the path selection time, which enhances the network throughput. The tree topology-based Q routing obtains its minimal route at $\alpha = 0.69$, and the route selection is faster

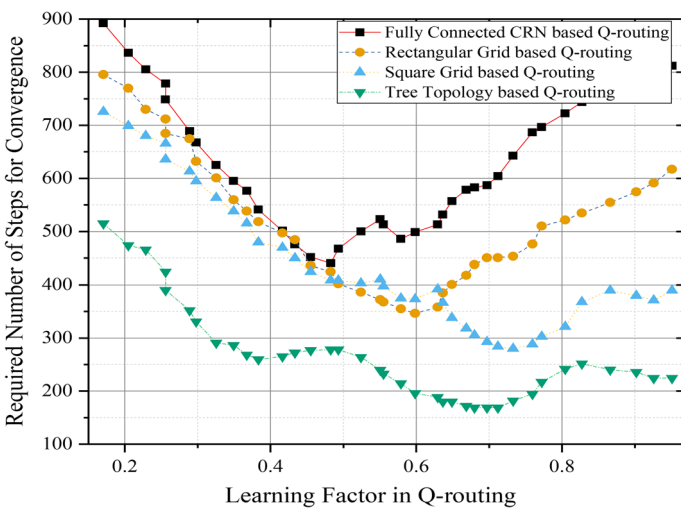


Fig. 8 Required steps for convergence vs learning factor α in Q-routing

by 1.86, 2.33, and 3.1 times compared to the square grid, rectangular grid, and fully connected network, respectively.

Figure 9 represents a 3D plot that shows the change in sum throughput (bps/Hz) of the SU network for the fixed value of α while β value varies. The plot is drawn only for the tree topology-based network. For the fixed value of $\alpha = 0.6$, initially the throughput is improved from 0.22 to 0.88 while the β value is increased from 0.12 to 0.57. The maximum throughput is achieved at $\beta = 0.57$, and after that, for the same α value, the throughput is decreased. For each kind of network topology, there exists an optimal α and β values for which the system offers the maximum throughput.

Figure 10 compares the performance improvement in tree topology-based Q-routing over the other existing algorithms in the literature [18, 36, 42]. Figure 10 plots the sum throughput of the CRN vs. the number of relays in the CRN. It is observed that the proposed tree topology based Q-routing outperforms the SU throughput by $\sim 7.14\%$, $\sim 11.11\%$, and $\sim 19.93\%$ while compared to [36, 42], and [18] when the number of relay nodes is fixed to 35. The Q-routing technique proposed in [36] follows a random network architecture, which increases the run time complexity and plunges down the network throughput compared to the tree topology-based Q-routing. On the other hand, the ECO-LEACH routing strategy initially adopts the clustering technique, and this reduces the secondary data transmission time.

5 Conclusions

The proposed work explores the Q-routing algorithm in the CRN. It is observed that the typical Q-routing enhances the network throughput by $\sim 37.21\%$ over the popular BFD algorithm in a similar framework. The work investigates and explores the different network

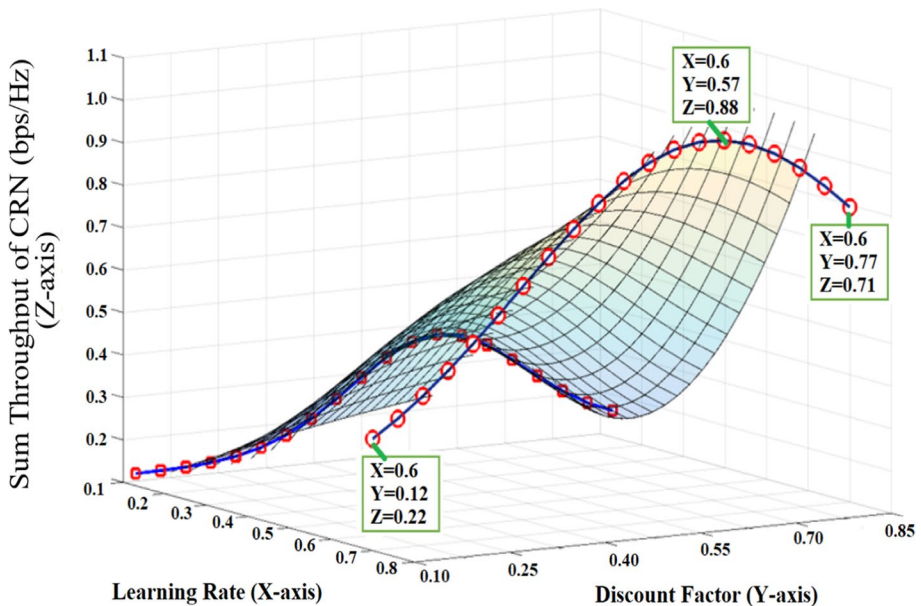


Fig. 9 Sum throughput (z-axis) vs α (x-axis) vs β (y-axis)

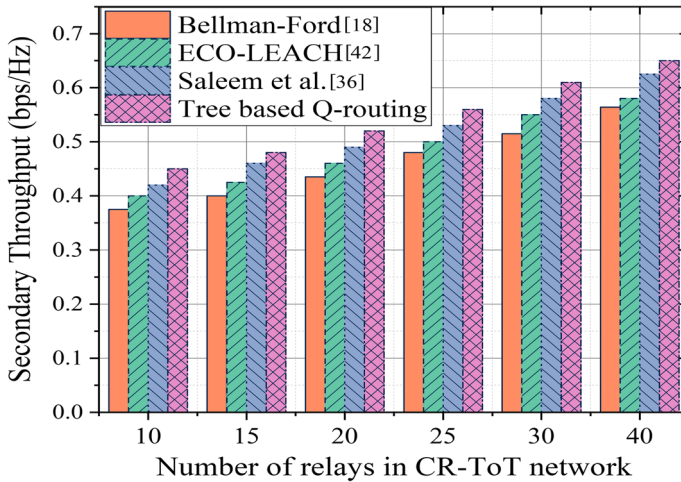


Fig. 10 Sum throughput vs number of relays in CRNs

topologies to reduce the run time complexities of the Q-routing. It is found that the worst-case time complexity of Q routing can be decreased to $\mathcal{O}(N)$ from $\mathcal{O}(ed)$ if the network follows some particular constraints. Tree topology-based Q-routing provides enhanced throughput by $\sim 6.83\%$, $\sim 7.21\%$, and $\sim 18.11\%$ over the square grid, rectangular grid, and strongly connected CRN topologies, respectively.

Funding No funding is available for this work.

Availability of data and material Not shareable currently.

Declarations

Conflicts of interest Not applicable.

Code availability Not shareable currently.

References

- Hassan, A. M., & Awad, A. I. (2018). Urban transition in the era of the internet of things: Social implications and privacy challenges. *IEEE Access*, 6, 36428–36440.
- Ma, Y., Wu, Y., Ge, J., & Li, J. (2018). An architecture for accountable anonymous access in the internet-of-things network. *IEEE Access*, 6, 14451–14461.
- Reeves, D. (2018). How to create a smart city: Future-proofed cities that foster growth and innovation. *IEEE Electrification Magazine*, 6, 34–41.
- Salameh, H. A. B., Almajali, S., Ayyash, M., & Elgala, H. (2018). Spectrum assignment in cognitive radio networks for internet-of-things delay-sensitive applications under jamming attacks. *IEEE Internet of Things Journal*, 5, 1904–1913.
- Zaheer, K., Othman, M., Rehmani, M. H., & Perumal, T. (2018). A survey of decision-theoretic models for cognitive internet of things (CIoT). *IEEE Access*, 6, 22489–22512.
- Han, R., Gao, Y., Wu, C., & Lu, D. (2018). An effective multi-objective optimization algorithm for spectrum allocations in the cognitive-radio-based internet of things. *IEEE Access*, 6, 12858–12867.

7. Kakalou, I., Psannis, K. E., Krawiec, P., & Badea, R. (2017). Cognitive radio network and network service chaining toward 5G: Challenges and requirements. *IEEE Communications Magazine*, 55, 145–151.
8. Katta, S., & Prasad, M. S. G. (2020). Performance analysis of cognitive radio under spectrum sharing using CTMC queuing model. *Journal of Critical Reviews*, 7(5), 555–558.
9. Paul, A., Banerjee, A., & Maity, S. P. (2019). Throughput maximisation in cognitive radio networks with residual bandwidth. *IET Communications*, 13(10), 1327–1335.
10. Lee, S., Duong, T. Q., da Costa, D. B., Ha, D. B., & Nguyen, S. Q. (2018). Underlay cognitive radio networks with cooperative non-orthogonal multiple access. *IET Communications*, 12(3), 359–366.
11. Ganesh, D., Kumar, T. P., & Kumar, M. S. (2021). Optimised levenshtein centroid cross-layer defence for multi-hop cognitive radio networks. *IET Communications*, 15(2), 245–256.
12. Zhang, X., Zhang, X., Han, L., & Xing, R. (2018). Utilization-oriented spectrum allocation in an underlay cognitive radio network. *IEEE Access*, 6, 12905–12912.
13. Yan, Z., Chen, S., Zhang, X., & Liu, H. L. (2018) Outage performance analysis of wireless energy harvesting relay-assisted random underlay cognitive networks. *IEEE Internet of Things Journal*, pp. 1.
14. Tian, R., Wang, Z., & Tan, X. (2018) A new leakage-based precoding scheme in iot oriented cognitive MIMO-OFDM systems. *IEEE Access*, pp. 1.
15. Khakzad, H., Taherpour, A., Shakeri, R., & Khattab, T. (2017). Dynamic interference-limited relay sharing in cognitive radio networks by using hierarchical modulation. *IET Communications*, 11(12), 1903–1912.
16. Huang, X., Yu, R., Kang, J., Xia, Z., & Zhang, Y. (2018). Software defined networking for energy harvesting internet of things. *IEEE Internet of Things Journal*, 5, 1389–1399.
17. NavnathDattatraya, K., & Rao, K. R. (2019). Maximising network lifetime and energy efficiency of wireless sensor network using group search ant lion with levy flight. *IET Communications*, 14(6), 914–922.
18. Banerjee, A., Paul, A., & Maity, S. P. (2018). Joint power allocation and route selection for outage minimization in multihop cognitive radio networks with energy harvesting. *IEEE Transactions on Cognitive Communications and Networking*, 4, 82–92.
19. Huang, M., Liu, Y., Zhang, N., Xiong, N. N., Liu, A., Zeng, Z., & Song, H. (2018). A services routing based caching scheme for cloud assisted crns. *IEEE Access*, 6, 15787–15805.
20. Sonti, S. R., & Prasad, M. S. G. (2019). Enhanced fuzzy c-means clustering based cooperative spectrum sensing combined with multi-objective resource allocation approach for delay-aware crns. *IET Communications*, 14(4), 619–626.
21. Parida, R. K., Mishra, R. K., Sahoo, N. K., Muduli, A., Panda, D. C., & Mishra, R. K. (2020). A hybrid multi-port antenna system for cognitive radio. *Progress in Electromagnetics Research*, 106, 1–16.
22. Tang, F., & Li, J. (2017). Joint rate adaptation, channel assignment and routing to maximize social welfare in multi-hop cognitive radio networks. *IEEE Transactions on Wireless Communications*, 16, 2097–2110.
23. Paul, A., & Maity, S. P. (2018). On outage minimization in cognitive radio networks through routing and power control. *Wireless Personal Communications*, 98, 251–269.
24. Anusha, M., & Vemuru, S. (2020). An effective mac protocol for multi-radio multi-channel environment of cognitive radio wireless mesh network (CRWMN). In: *Proceedings of the First International Conference on Sustainable Technologies for Computational Intelligence*, pp. 21–35, Springer, New York.
25. Ch, S., Ramesh, K., et al. (2020). Fuzzy guided integrative factors-based spectrum decision-making in cognitive radio networks. *International Journal of Intelligent Unmanned Systems*.
26. Paul, A., & Maity, S. P. (2016). Kernel fuzzy c-means clustering on energy detection based cooperative spectrum sensing. *Digital Communications and Networks*, 2(4), 196–205.
27. Paul, A., Kunarapu, P., Banerjee, A., & Maity, S. P. (2019). Spectrum sensing in cognitive vehicular networks for uniform mobility model. *IET Communications*, 13(19), 3127–3134.
28. Anumandla, K. K., Sabat, S. L., Peesapati, R., AV, P., Dabbakuti, J. K., & Rout, R. Optimal spectrum and power allocation using evolutionary algorithms for cognitive radio networks. *Internet Technology Letters*, p. e207.
29. Nayak, D. K., Muduli, A., Hussain, M. T., Mirza, A. A., Gummadipudi, J. R., & Kumar, N. S. (2020). Channel allocation in cognitive radio networks using energy detection technique. *Materials Today: Proceedings*, 33, 934–938.
30. Reddy, S. S., & Prasad, M. S. G. (2021). Improved whale optimization algorithm and convolutional neural network based cooperative spectrum sensing in cognitive radio networks. *Information Security Journal: A Global Perspective*, 30(3), 160–172.

31. Paul, A., & Maity, S. P. (2020). Outage analysis in cognitive radio networks with energy harvesting and q-routing. *IEEE Transactions on Vehicular Technology*, 69(6), 6755–6765.
32. Paul, A., & Maity, S. P. (2017). Optimal cluster power for joint spectrum sensing and secondary data transmission in cognitive radio networks. In: *Proceedings of the IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, pp. 1–6, IEEE.
33. Du, Y., Xue, L., Xu, Y., & Liu, Z. (2019). An apprenticeship learning scheme based on expert demonstrations for cross-layer routing design in cognitive radio networks. *AEU-International Journal of Electronics and Communications*, 107, 221–230.
34. Lavanya, S., & Bhagyaveni, M. A. (2017). Design of sop based cross-layered opportunistic routing protocol for cr ad-hoc networks. *Wireless Personal Communications*, 96(4), 6543–6556.
35. Saad, M. (2014). Joint optimal routing and power allocation for spectral efficiency in multihop wireless networks. *IEEE Transactions on Wireless Communications*, 13, 2530–2539.
36. Saleem, Y., Yau, K. L. A., Mohamad, H., Ramli, N., Rehmani, M. H., & Ni, Q. (2017). Clustering and reinforcement-learning-based routing for cognitive radio networks. *IEEE Wireless Communications*, 24, 146–151.
37. Syed, A. R., Yau, K. L. A., Qadir, J., Mohamad, H., Ramli, N., & Keoh, S. L. (2016). Route selection for multi-hop cognitive radio networks using reinforcement learning: An experimental study. *IEEE Access*, 4, 6304–6324.
38. Wang, J., Yue, H., Hai, L., & Fang, Y. (2017). Spectrum-aware anypath routing in multi-hop cognitive radio networks. *IEEE Transactions on Mobile Computing*, 16, 1176–1187.
39. Mansourkiaie, F., Ismail, L. S., Elfouly, T. M., & Ahmed, M. H. (2017). Maximizing lifetime in wireless sensor network for structural health monitoring with and without energy harvesting. *IEEE Access*, 5, 2383–2395.
40. Babae, R., & Beaulieu, N. C. (2011). Power-optimized routing with bandwidth guarantee in multihop relaying networks. In: *Proceedings of the IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE.
41. Zhang, L., Cai, Z., Li, P., Wang, L., & Wang, X. (2017). Spectrum-availability based routing for cognitive sensor networks. *IEEE Access*, 5, 4448–4457.
42. Bahbahani, M. S., & Alsusa, E. (2018). A cooperative clustering protocol with duty cycling for energy harvesting enabled wireless sensor networks. *IEEE Transactions on Wireless Communications*, 17, 101–111.
43. Vivekanand, C. V., & Bagan, K. B. (2020). Secure distance based improved leach routing to prevent puea in cognitive radio network. *Wireless Personal Communications*, 113(4), 1823–1837.
44. Koenig, S., & Simmons, R. G. (1992). Complexity analysis of real-time reinforcement learning applied to finding shortest paths in deterministic domains. Technical report
45. Salih, Q. M., Rahman, M. A., Al-Turjman, F., & Azmi, Z. R. M. (2020). Smart routing management framework exploiting dynamic data resources of cross-layer design and machine learning approaches for mobile cognitive radio networks: A survey. *IEEE Access*, 8, 67835–67867.
46. Elangovan, K., & Subashini, S. (2018). Particle bee optimized convolution neural network for managing security using cross-layer design in cognitive radio network. *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–9.
47. Gawas, M. A., & Govekar, S. (2021). State-of-art and open issues of cross-layer design and qos routing in internet of vehicles. *Wireless Personal Communications*, 116(3), 2261–2297.
48. Singhal, C., & Rajesh, A. (2020). Review on cross-layer design for cognitive ad-hoc and sensor network. *IET Communications*, 14(6), 897–909.
49. Awang, A., Husain, K., Kamel, N., & Aissa, S. (2017). Routing in vehicular ad-hoc networks: a survey on single-and cross-layer design techniques, and perspectives. *IEEE Access*, 5, 9497–9517.
50. Chen, J., Ping, S., Jia, J., Deng, Y., Dohler, M., & Aghvami, H. (2016). Cross-layer optimization for spectrum aggregation-based cognitive radio ad-hoc networks. In: *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6.



Anal Paul is currently working as Associate Professor in the Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, Telangana, India, 500075. He received his Ph. D degree from Indian Institute of Engineering Science and Technology, Shibpur, India in January, 2021. His research interests include cooperative spectrum sensing, wireless routing, and energy harvesting.



Santi P. Maity received his B.E. degree in Electronics and Communication Engineering from National Institute of Technology Durgapur and M. Tech. in Microwaves from the University of Burdwan, India in 1993 and 1997, respectively. He received his Ph.D. degree in Engineering from Bengal Engineering and Science University, Shibpur, (BESU Shibpur), at present know as the Indian Institute of Engineering Science and Technology, Shibpur, India in 2008. He received a couple of post-doctoral positions in several institutes, like University of Vigo, Spain, Nanyang Technological University, Singapore, and Universite Paris-Sud 11. He did his Postdoctoral work from January 2009 to July 2009 and February 2011 to July 2011 at the Laboratoire des Signaux et Systems (CNRS-Supelec-Universite Paris-Sud 11) in France. He has been working as Professor in the Dept. of Information Technology, IEST, Shibpur, India since September 2012. He served as Head, Department of Information Technology, BESU Shibpur from January 2012 to January 2014, head of PDSIT, IEST Shibpur from July, 2018 to January, 2019, January, 2020 to July, 2020 and at present

head of School of VLSI at IEST Shibpur since January, 2021. His research interests include CDMA, MC-CDMA, Cognitive Radio Network; Compressed Sensing Image reconstruction and Transmission, Digital Image Watermarking and Secret Sharing. His recent research works focus on Machine Learning in image reconstruction and CRN. He has published about 300 research articles in several Journals, Conference Proceedings and book chapters. He has guided (as single and joint supervisors) so far 19 Ph. D students on the above mentioned research areas and few are also in working.