International Journal of Computer Networks and Applications (IJCNA)
DOI: 10.22247/ijcna/2023/221897 Volume 10, Issue 3, May – June (2023)

RESEARCH ARTICLE

Energy Efficient Multi Hop D2D Communication Using Deep Reinforcement Learning in 5G Networks

Md. Tabrej Khan
Department Faculty of Computer Science, Pacific Academy of Higher Education and Research University, Udaipur (Rajasthan), India.
tabrejmlkhan@gmail.com

Ashish Adholiya
Department Faculty of Computer Science, Pacific Academy of Higher Education and Research University, Udaipur (Rajasthan), India.
asia_1983@rediffmail.com

Received: 16 April 2023 / Revised: 04 June 2023 / Accepted: 09 June 2023 / Published: 30 June 2023

Abstract – One of the most potential 5G technologies for wireless networks is device-to-device (D2D) communication. It promises peer-to-peer consumers high data speeds, ubiquity, and low latency, energy, and spectrum efficiency. These benefits make it possible for D2D communication to be completely realized in a multi-hop communication scenario. However, the energy efficient multi hop routing is more challenging task. Hence, this research deep reinforcement learning based multi hop routing protocol is introduced. In this, the energy consumption is considered by the proposed double deep Q learning technique for identifying the possible paths. Then, the optimal best path is selected by the proposed Gannet Chimp optimization (GCO) algorithm using multi-objective fitness function. The assessment of the proposed method based on various measures like packet delivery ratio, latency, residual energy, throughput and network lifetime accomplished the values of 99.89, 1.63, 0.98, 64 and 99.69 respectively.

Index Terms – 5G Networks, D2D Communication, Energy Efficient Routing, Multi-Hop Path, Deep Q Learning, Optimal Path Selection.

1. INTRODUCTION

Due to the rising demand for multimedia applications and smart phones over the last few years, mobile data traffic has accomplished an accelerated enhancement. Demand for bandwidth is brought on by this sharp rise in data traffic [1]. Although many conventional approaches, such as employing femto and pico cells to reduce cell size and increase throughput, have been suggested to address this issue, their high deployment costs mean that the issue has not yet been resolved. Due to characteristics like resistance, high energy efficiency, high throughput, and cellular traffic offloading to infrastructure failure, device-to-device (D2D) communication is regarded as the best option [2-4]. According to the definition, it is a kind of communication in which mobile nodes interact with one another without the use of a base station or other centralized structure. Instead, they use their own internal communication channels. The two-tier paradigm used by traditional D2D communication relies on cellular architecture to fetch resources. Due to the fact that the devices are closer together, less power is consumed, which improves battery life. Nevertheless, the communication required for this architecture causes interference problems. In order to combat interference, relay nodes are given dedicated resources for lengthy communication, yet these results in resource waste. The utilization of multimedia information is rising, which uses image, audios, videos and several other information having the highest request factor (78%) and driving the majority of the network traffic through the year 2022. Massive congestion on devices and traffic loads, as well as backhaul issues that impair device battery life, can be caused by high demand for videos [5].

With the dramatic rise in demand for faster data rates, D2D communication has attracted a lot of attention from the business world, standards bodies, and academic institutions [6]. Direct wireless communication between two transceivers is made possible via D2D communication without a base station (BS). It is a crucial method for 5G networks since it helps to boost energy efficiency (EE) and use less power while still transferring data without loss [7, 8]. It immediately links the devices while simultaneously enhancing the network's performance, latency, and throughput, which in turn increases the energy efficiency and spectral efficiency of D2D communication. Moreover, it shares frequencies with cellular networks, which degrades network quality because more interference are generated [9]. Cellular networks that support D2D are taken into account, with an emphasis on maximizing energy efficiency. Only single-cell scenarios are supported by the available technologies. Since the user's requirement needs a lot of power, a single-cell situation received more attention

RESEARCH ARTICLE

than multiple bands; therefore, research is concentrated on how to provide the energy efficiency routing with minimal resource utilization through multi-hop routing strategy [10, 11]. To transmit and receive the information, either an uplink or a downlink is taken into account. Energy efficiency optimization is necessary whenever the demand for optimal energy efficiency arises. The answer to the energy efficiency optimization issue is to obtain the greatest EE (energy efficiency) in D2D communication [12].

Several network issues in 5G have been successfully solved using machine learning. One of the best methods of machine learning for controlling strategy is reinforcement learning (RL) [13]. The intelligent resource management challenge in D2D underlay networks has subsequently been addressed in a number of publications using reinforcement learning. Depending on the policy, an agent supporting D2D combination dynamically chooses an acceptable range that was acquired through RL [14-16]. An efficient routing protocol based on Q-learning utilizes the actor-critic (AC) technique. An AC technique is necessary since Q-learning is inadequate to handle continuous valued state and action spaces [17]. Deep Neural Networks (DNNs) are employed in the Deep Q-Network (DQN) algorithm to estimate the value function of Q-learning, relieving the burden of computing and storing Q-values [18, 19]. As a result, certain RL-based distributed resource efficient routing strategies have been put forth to lessen computing complexity [20].

As the world moves closer to the era of digitization that demands for the internet and hence the use of it is increasing dramatically. The internet is connected to billions of physical objects all around the world, which are gathering and exchanging data. The Internet of Things (IoT) makes transportation, environment, industrial automation, smart grid, smart cities, smart homes, as well as healthcare monitoring over the network for transmitting vast amounts of data without human intervention.

Future 5G networks will manage this data and information by offering better connection, larger data rates, ultra-low latency, more energy efficiency, and improved spectrum efficiency. D2D communication is completely realized in a multi-hop communication environment because to these advantages. A multi-hop network might perform worse than a typical mobile system if the improper routing decisions are made without the right processes; hence in order to construct multi-hop D2D communication networks optimally, the routing component should be designed more efficiently.

As a result, this research introduces routing in multi-hop networks. The major goal of the research is to provide the energy efficient cooperative routing protocol. For this a hybrid approach with deep learning based path detection and optimized path selection is proposed. The major contributions of the research are:

- Double Deep Q-Learning: The double deep Q learning is introduced for the detection of paths from source to destination. In the proposed double deep Q learning, the estimation of Q-value and reward are estimated using two various Deep CNN models to avoid the over optimistic issues.
- Gannet Chimp Optimization: The Gannet chimp optimization (GCO) is designed by hybridizing the hunting behavior of the Gannet with the chimp to identify the optimal best path D2D communication.

The organizations of the introduced D2D communication protocol are: Section 2 details the literature review along with the problem statement and the methodology to overcome the issues is detailed in Section 3. The experimental outcome is elaborated in Section 4 and Section 5 presents the conclusion.

2. LITERATURE REVIEW

Some of the prior methods concerning the D2D communication are detailed in this section. An energy efficient D2D communication was devised by [21], in which modified derivative algorithm was introduced for the energy efficient computation overhead. The goal behind the development of the protocol is to minimize the data traffic through optimal resource allocation strategy. The outcome of the experimentation depicts the ideal performance with enhanced efficiency in terms of energy. The failure in considering the significant parameters limits the performance of the model. The energy efficient information sharing using the clustering based criteria was developed by [22]. In this, a weighted algorithm along with the group mobility criteria was designed for the cluster formation and cluster head selection. The direction and speed of the dynamic users were considered for clustering the devices for efficient information sharing. The major challenging aspect of the devised method was the security concerns that create the information leakage. Also, the failure in considering the energy and delay parameters affects the network's scalability.

Machine learning based D2D was designed by [23] using a single relay hop, wherein the recursive learning was utilized for the updating the agents. Besides, the usage of fuzzy logic was incorporated with the recursive learning to choose the best relay node by considering the local knowledge. The performance of the method in terms of power consumption and spectral efficiency depicted the flexibility and effectiveness of the model. However, the failure in considering the bandwidth for performing the cooperative routing limits the efficiency of the model.

Hybrid routing protocol based on reinforcement learning was designed by [24] using the static learning criteria. In this, the channel capacity and traffic intensity were considered for the selecting the best route for information sharing. The analysis

RESEARCH ARTICLE

based on various measures depicts the enhanced quality of service through the constant learning criteria. The failure in considering the processing delay enhances the latency of the information sharing between the devices.

D2D communication using the multi criteria decision making was designed by [25] by considering the factors like contention window, link quality, battery and mobility. In this, multi hop routing was devised for the information sharing with optimal best path. Besides, the energy cost, delay, energy consumption, packet delivery ratio and throughput were considered for the analysis of the performance of the devised protocol and illustrated the superior performance. However, the optimal routing selection criteria failed to consider the bandwidth for traffic-less information flow. An energy efficient routing protocol using the deep reinforcement learning was designed by [19] for communication between two devices. In this, the delay associated with the routing was minimized through the energy consumption based route selection using the deep reinforcement learning technique. The consideration of power consumption and latency in selecting the route enhances the efficiency of the model. The short description of the literature review along with the advantages, disadvantages and methodology utilized were included in Table 1.

Table 1 Short Description of Literature Review

Reference	Methodology	Advantages	Disadvantages
L. Nagapuri et al., [21]	Modified derivative algorithm for energy efficient routing	The optimal D2D user selection enhances the performance of the network in terms of energy efficiency	The scalability of the network is still a challenging task.
N. Khan et al., [22]	Cluster based energy efficient D2D communication	Energy efficient communication is accomplished for dynamic user assignment.	Insecure communication is the challenging aspect.
I. Ioannou et al., [23]	Distributed artificial intelligence framework	Accomplished minimal delay and enhanced spectral efficiency in flexible D2D communication.	Failed to consider the significant features that enhance the energy efficiency.
M. K. Chamran et al., [24]	Reinforcement learning based route selection	Accomplished minimal delay with better quality of service.	The number of routes utilized for communication was minimal.
V. Tilwari et al., [25]	Optimal path selection using multi-criteria based decision making	Acquired better QoS with the optimal route selection.	Failed to consider the bandwidth for traffic-less information flow.
D Han and J. So [19]	Reinforcement learning based route selection	Acquired enhanced power consumption and latency through the best energy efficient route selection algorithm.	Failed to consider the interference between the devices while allocating the resource.

2.1. Problem Statement

When mobile users randomly moves across one place to another, D2D communication make benefits from the opportunities through an efficient routing strategy. The exchange of information in such chance interactions among individuals is closely tied to physical movement. Services and apps that support D2D visualize highly ad hoc and unexpected movements by taking advantage of user movement. Due to the users' complex requirement, there are difficulties in fully user's requirement. The main focus is on foreseeing the development of communication linkages between D2D users efficiently. The complete D2D network is impacted by mobility, including signal strength, area of operation, and bandwidth demands. In many different application sectors, including the automobile industry, emergency communications, and many other sectors, D2D communication with 5G wireless technologies is extensively used. The vital field of movement research continues to undergo development, despite the existence of a number of

RESEARCH ARTICLE

intriguing investigations on D2D communications that have made significant contributions and greater recognition of D2D communications. For example, the most recent issues in Routing protocol include interference reduction, storage and offload, energy efficiency, delay, and many others; still, the energy efficient routing protocol development is more challenging task. Hence, a novel energy efficient D2D communication protocol is designed using deep learning based path detection and multi-objective function based optimal path selection algorithm. In this, the consideration of residual energy, packet latency, bandwidth, hop count, and degree of connectivity assist to solve the interference, energy efficient communication, and latency issue more effectively.

3. PROPOSED ENERGY EFFICIENT D2D COMMUNICATION FOR 5G NETWORKS

The energy efficient D2D communication between the users of 5G network with multi-hop routing strategy is introduced in this research. Initially, the possible paths for D2D communication with multi-hop is identified using the proposed double deep Q learning. The double deep Q learning utilizes two various Deep CNN for the estimation of the Q-value and reward function to avoid the over optimistic issues. Here, the energy consumption of the node is evaluated by the proposed double deep Q learning method for the acquisition of energy efficient routing. From the detected paths, the optimal bets path is identified by the proposed Gannet Chimp Optimization (GCO) algorithm. The GCO is designed by hybridizing the hunting behavior of the Gannet with the attacking behavior of the chimp in capturing the target. The reason behind the hybridization is to enhance the convergence rate with global best solution. Here, the multi-objective fitness function is considered for the selection of optimal best path. The multi-objective fitness based on residual energy, packet latency, bandwidth, hop count and degree of connectivity are considered for the design of multi-objective fitness function that enhances the efficiency of path selection. The work flow of the proposed energy efficient D2D communication is depicted in Figure 1.

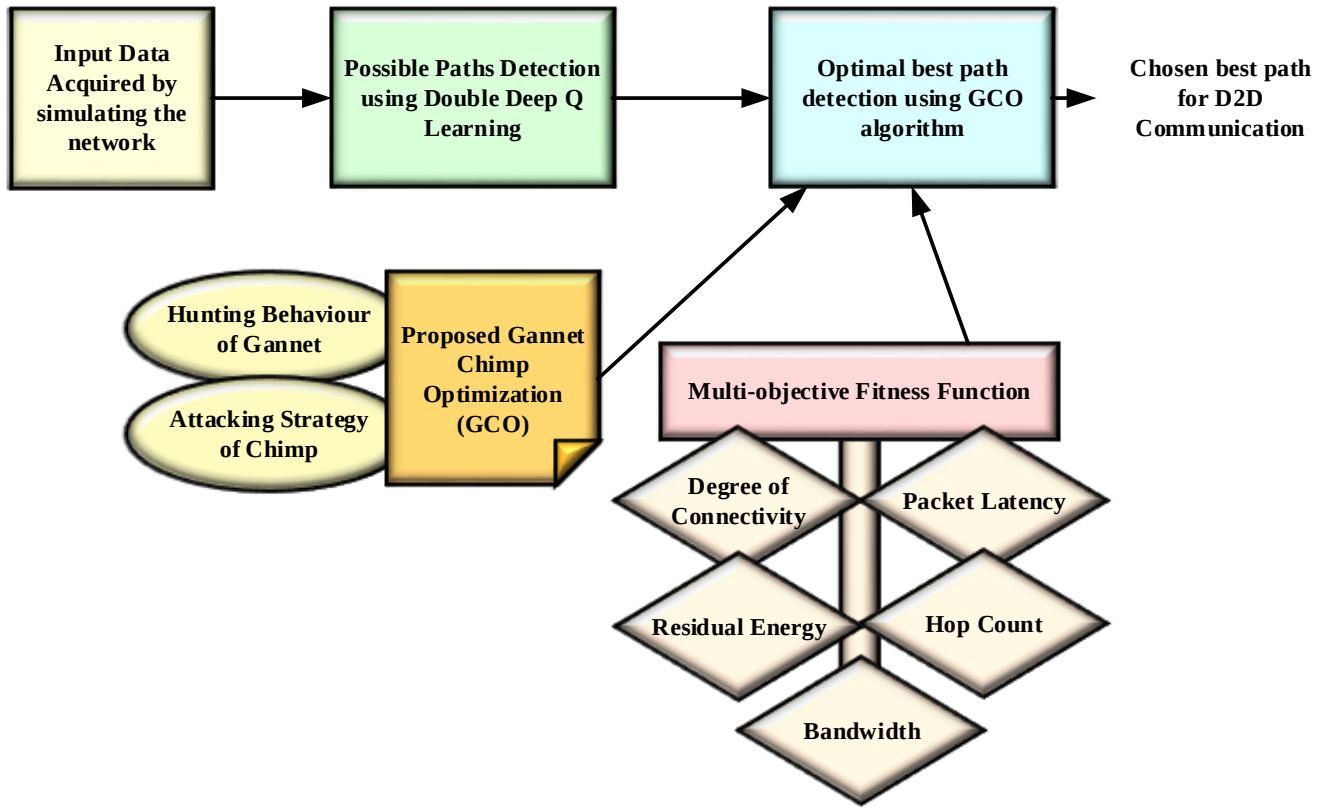


Figure 1 Workflow of Proposed Energy Efficient D2D Communication

3.1. Data Acquisition

The data utilized for the processing of the proposed D2D communication with multi-hop energy efficient routing is acquired by simulating the network. The data acquired is utilized for processing the proposed methodology.

International Journal of Computer Networks and Applications (IJCNA)
DOI: 10.22247/ijcna/2023/221897 Volume 10, Issue 3, May – June (2023)

RESEARCH ARTICLE

3.2. Double Deep Q Learning for Path Detection

The possible paths for D2D communication using the multi-hop routing is identified by the double deep Q learning by considering the energy efficiency. The conventional Q learning uses the Markov decision making strategy for solving the issues in the reinforcement learning; still it is incapable for solving the complex variables. Besides, the curse of dimensionality issues and elevates the computation complexity and limits the convergence speed. These issues are solved using the deep-Q-learning approach, in which the deep neural network (DNN) is utilized for evaluating the reward and Q-value. The DNN is the replacement of the discrete value function of the Q-learning; still, the deep-Q-learning introduces the over optimistic issues due to the usage of single DNN for estimating the reward and Q-value. Thus, the double deep Q learning efficiently solves the over optimistic issue by introducing two separate DNN for estimating the reward and Q-value.

3.2.1. Deep Q-Learning

The conventional Q learning provides the outcome as Q-value by acquiring the inputs state and action. However, the deep Q learning provides various actions as its outcome utilizing the state value. The architecture of the deep-Q-learning and the Q-learning process is depicted in Figure 2 given below.

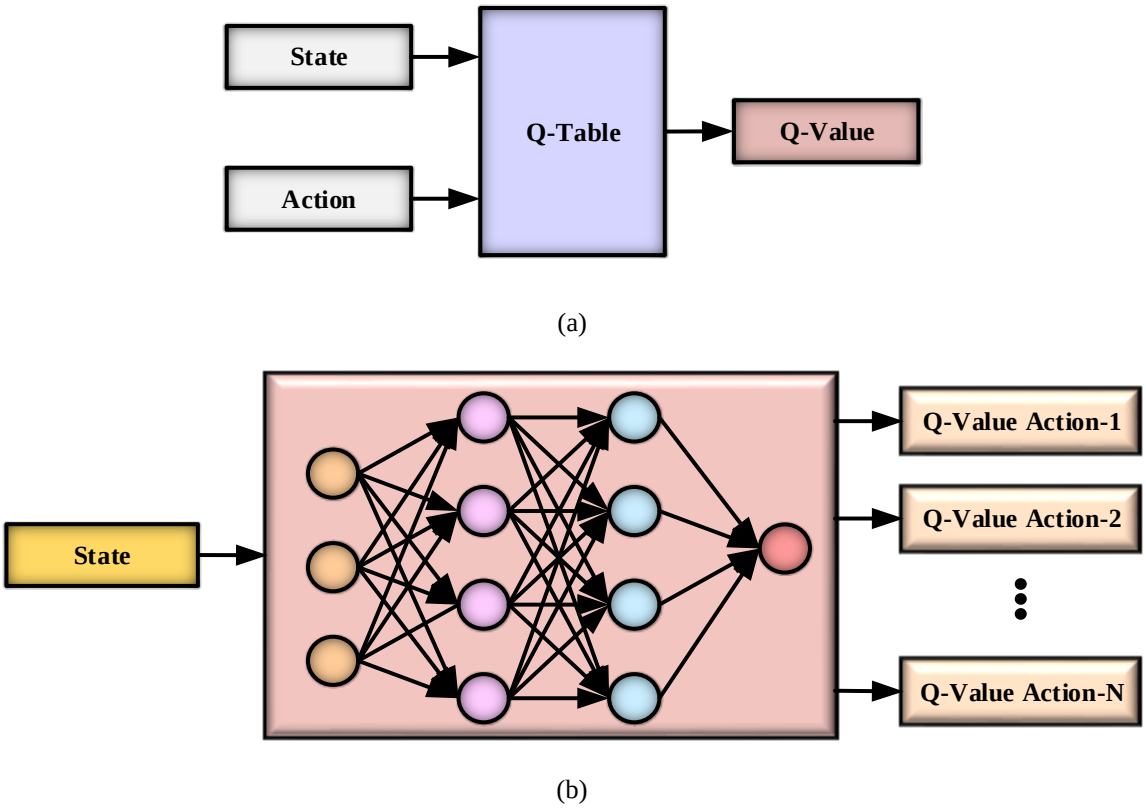


Figure 2 System Model of: (a) Q-Learning and (b) Deep Q Learning

Here, for the state X, the rewards are evaluated as $Y^{f}_{X,X'}$, wherein the action is defined as F. The term β defines the discount factor and $E^{f}_{X,X'}$ refers to the action-state pair probability. The D2D communication among the users of 5G communication is enunciated as states and the multi-hop routing among the nodes to reach the destination is enunciated as action.

3.2.2. Reward and Q Value Evaluation

The D2D communication among the user depends on the behaviour of the agent in the deep-Q-Learning through the reward estimated from the state. Here, for the energy efficient D2D communication among the users considers the energy consumption for providing the energy efficient communication. Let us consider the user m_c, who is considered as source node and the receiver node is defined as m_b. The evaluation of the reward function is defined in equation (1).

RESEARCH ARTICLE

$$Y_{m_c,m_b}^{n_c} = -p - \alpha_1[l(m_c)+l(m_b)] + \alpha_2[n(m_c)+n(m_b)] \quad (1)$$

Where, the action-state pair is defined as (m, f_s) and α_1 *and* α_2 refers to the weighting parameter. The reward function is defined as $Y_{m_c,m_b}^{n_c}$; then the cost function enunciated as the punishment factor is defined as p. The reward function estimated in equation (2) is utilized for the successful communication between the nodes, if the communication gets dropped; then, the reward is estimated in equation (2).

$$Y_{m_c,m_b}^{n_c} = -p \times \eta - \gamma_1 l(m_c) + \gamma_2 n(m_c) \quad (2)$$

Where, η refers the drop case of communication and the energy evaluation for the communication is defined as $l(m_c)$ and is formulated in equation (3).

$$l(m_c) = 1 - \frac{E_{resi}(m_c)}{E_{ini}(m_c)} \quad (3)$$

Where, the initial energy varies from $[0,1]$ and is referred as E_{ini}, then, the residual energy is represented as E_{resi}. The normalized form of energy is indicated as $l(m_c)$ that plays a crucial role in communication between the nodes. Because, for the energy efficient routing protocol, E_{resi} is highly essential. The communication between the nodes takes place when the E_{resi} value becomes higher for the avoidance of communication dropping. Then, the reward function formulation for the group is enunciated in equation (4).

$$n(m_c) = \frac{2}{\pi}\arctan\left(E_{resi}(m_c) - \bar{E}(m_c)\right) \quad (4)$$

Where, the term $\bar{E}$ defines the residual energy of a group in average. Then, the final reward function is enunciated in equation (5).

$$\text{Re}\,ward = E_X \times Y_{m_c,m_b}^{f_c} + E(1-E_X) \times Y_{m_c,m_b}^{f_c} \quad (5)$$

Estimation of Q-Value: For the acquisition of the highest reward value, the Q-value is evaluated to make the required action. The Q-value is enunciated in equation (6).

$$Q-V(X, f) = \text{Re}\,ward + \beta\left[Q-V(X, f) + Max_{f'}(Q-V(X',:))\right] \quad (6)$$

Where, estimation of the Q-value is defined as $Q-V$ and is highly helpful in choosing the energy efficient node for D2D communication.

3.2.3. Double Deep Q Learning based on Deep CNN

The traditional double deep Q learning utilizes the DNN for estimating the Q-value and reward function. In the proposed methodology, the deep convolutional Neural Network (Deep CNN) is utilized for estimating the Q-value and reward function. The detailed description is given below.

3.2.3.1. Architecture of Deep CNN

The complex features are learned by the deep learning models to enhance the generalization capability that provides the outcome more efficient through various layers. Recent years, the deep learning methods are widely utilized for solving various application domains concerning the computer vision related tasks like classification, prediction, recognition and various other tasks due to the promising outcome of the deep learning models. The convolutional neural network (CNN), recurrent neural network (RNN), and deep belief networks (DBN) are the some of the examples for the deep learning methods. Besides, the automatic feature extraction criterion of the deep learning methods diminishes the need for external feature extraction technique. Thus, in the proposed path detection model, the deep CNN (Deep CNN) is introduced for the estimation of Q-value and reward function. The architecture of the Deep CNN is depicted in Figure 3.

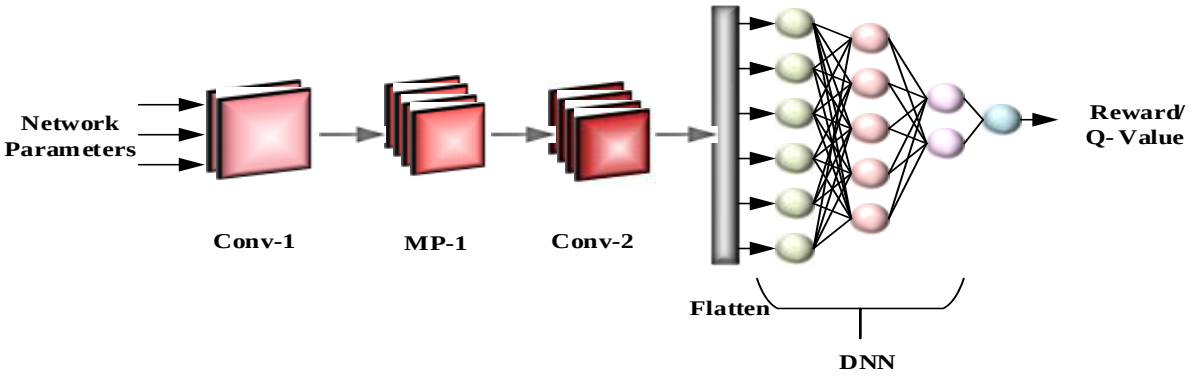


Figure 3 Architecture of Deep CNN

RESEARCH ARTICLE

The layer-wise details of the Deep CNN for estimating the Reward or the Q-value is detailed here.

Conv Layer: The input parameters of the network for energy efficient routing is acquired by the Conv layer-1 and it convolves the input with the kernel function for obtaining the feature mapping. The formulation for the conv layer outcome is defined in equation (7).

$$R-Q_v = \sum X^w * Y^w + Q^w \quad (7)$$

Where, the outcome of the conv layer is defined as $R-Q_v$. The input feature is referred as X^w and the weight is represented as Y^w. The bias value is notated as Q^w, wherein the output map corresponding to the w^{th} feature is indicated as w.

Max-Pooling Layer: The most informative attributes are extracted in the pooling operation for the minimization of the redundant features, which in turn reduces the complexity concerning the computation overhead. In the proposed method, the max-pooling operation is utilized for extracting the significant attributes. An example for the max-pooling operation is illustrated in Figure 4.

Figure 4 Max-Pooling Operation

Flatten Layer: The transformation of the features into the single dimension is employed in the flatten layer.

Fully Connected Layer: The Q-value and the reward estimation is provided at the outcome of the fully connected layer, wherein the softmax activation is utilized. The estimation of the outcome is defined as,

$$R-Q_{v\,out} = \frac{e^{z_m}}{\sum_{n=1}^{i} e^{z_n}} \quad (8)$$

Where, the softmax function is indicated as $R-Q_{v\,out}$, the element corresponding to the input attribute is indicated as z_m, and i refers to the outcome.

3.3. Optimal Path Detection Using the Proposed Gannet Chimp Optimization Algorithm

The possible path detected by the proposed Double Deep Q Learning consists of various paths. From the all detected paths, the optimal best path for the D2D communication is chosen by the proposed GCO algorithm. The optimal best path is chosen by the GCO based on the Multi-objective fitness function with the factors like energy consumption, packet latency, bandwidth, hop count, and trust factor.

3.3.1. Multi-objective Fitness Function

The factors considered for the estimation of the optimal best path using the proposed GCO algorithm for estimating the multi-objective fitness function are energy consumption, packet latency, bandwidth, hop count, and trust factor. The detailed description is given below.

3.3.1.1 Residual Energy

The residual energy is crucial factor for the energy efficient D2D communication between the users. The node with higher residual energy is considered for communication between the users, because the node with higher energy has enough energy for communication without any interruption due to the lack of energy. The estimation of the residual energy is formulated in equation (9).

$$RE = E_c - \left(E_{txn} + E_{rxn}\right) \quad (9)$$

Where, the energy utilized for sender is indicated as E_{txn}, the energy utilized by the receiver is indicated as E_{rxn}, the residual energy is defined as RE, and the present remaining energy of the node is indicated as E_c. The node with higher RE is preferred for D2D communication.

3.3.1.2. Packet Latency

The latency is defined as the time taken by the network for D2D communication. The estimation of the packet latency is defined in equation (10).

$$PL = a\frac{P+Q(N)}{d} \quad (10)$$

where, the packet latency is notated as PL, the count of bits in the packet is notated as a, the number of packet is represented as N, the capacity of the link is indicated as d, the size of data is indicated as Q and the bit size of header is notated as P.

3.3.1.3. Bandwidth

The larger amount of bandwidth is essential for the uninterruptable communication between the D2D users. However, the limited amount of bandwidth the resource must be utilized in minimal amount for the efficient routing. Thus, the minimal amount of bandwidth needs to be considered for the efficient information routing, which is indicated as F_{BW}. The minimal bandwidth is utilized through communication the devices using the energy efficient node sensing.

RESEARCH ARTICLE

3.3.1.4. Hop Count

The proposed D2D communication routing protocol uses multi-hop path for user communication, wherein the path with large number of hop consumes more energy. Thus, the path with minimal hop is considered for the minimization of energy consumption. The hop count is defined using the variable F_{HC}.

3.3.1.5. Degree of Connectivity

The estimation of degree of connectivity is essential for identifying the capability of the node to handle the number of devices within the specified time t. The connectivity is defined as DC_i and the neighbour node is indicated as NN_i. Then, the expression for calculating the degree of connectivity is formulated in equation (11).

$$DC_i = \frac{NN_i}{D_{i,j} \le R_T} \tag{11}$$

Where, the transmission range is represented as R_T, the distance between the nodes is indicated as $D_{i,j}$.Thus, the multi-objective fitness function is formulated in equation (12).

$$MO_{fitness} = Max(RE, DC_i)Min(PL, F_{HC}, F_{BW}) \tag{12}$$

Here, the multi-objective fitness function is indicated as $MO_{fitness}$. The fitness function is normalized within the range of $[0,1]$ for making the computation simpler.

3.3.2. Gannet Chimp Optimization

The novel Gannet Chimp Optimization is designed by combining the hunting behaviour of the Gannet with the attacking behaviour of the chimp in capturing the target. The goal of the hybridization is to accomplish the global best solution with balanced randomization and local search capability. The balanced optimization assures the best solution for solving optimization issues without trapping at local optimal solution.

Motivation behind the Proposed Gannet Chimp Optimization

The Gannet [26] is a carnivorous bird that hunts the target (fish, amphibians, crustaceans, and so on) along the sea shore and lakes. They live in flocks with powerful eyes, slender necks and stubby. The enhanced vision of the bird helps to capture the target very easily by accurately identifying from a very large distance. Thus, the target that falls within the vision of the carnivorous bird never has the chance of escaping. Besides, the V-shaped and U-shaped diving behaviour of the bird assures the better encircling of the target. High capturability behaviour of the bird by ignoring the water resistance helps to capture the target very easily. Here, for enhancing the capturability of the bird, the attacking strategy of the chimp is integrated for obtaining the solution with fast convergence rate. The chimp [27] is a great ape African species belongs to the family Hominoid. Four various types of chimp are considered for solving the optimization issues, which are attacker, chaser, barrier and driver. Here, each chimp has their own role in capturing the target. The attacking strategy of the chimp is devised through the combined behaviour of all the four categories of chimp. Thus, the local search capability of the Gannet is enhanced by hybridizing the attacking strategy of the chimp to obtain the global best solution in solving the optimization issue. In the proposed methodology, the global best solution is utilized for identifying the energy efficient best path for D2D communication among the users through the multi-hop routing strategy.

3.3.2.1. Mathematical Modeling

The initialization of the proposed Gannet chimp optimization (GCO) algorithm is the first step, wherein the candidate solutions (Gannets) and the prey (target solution) are located randomly in the feature space. The solutions accomplished by each candidate in the feature space are utilized for solving the optimization issues. The candidate initialization in the feature space is formulated in equation (13).

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,y} & \cdots & a_{1,V-1} & a_{1,V} \\ a_{2,1} & \cdots & a_{2,y} & \cdots & a_{2,V-1} & a_{2,V} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & a_{x,y} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{U-1,1} & \cdots & a_{U-1,y} & \cdots & a_{U-1,V-1} & a_{U-1,V} \\ a_{U,1} & \cdots & a_{U,y} & \cdots & a_{U,V-1} & a_{U,V} \end{bmatrix} \tag{13}$$

Where, the x^{th} candidate's location in the feature space is defined as a_x. The solution accomplished by the x^{th} candidate in the feature space with the dimension y is expressed in equation (14).

$$a_{x,y} = g_1 \times (Q_y - S_y) + S, \qquad x = 1,2,...U, \quad y = 1,2,...V \tag{14}$$

Where, V refers to the solution's dimension and the population of the candidate solution is defined as U. The random number is notated as g_1 with the value of $[0,1]$. The boundary of the solution is between Q_y and S_y, in which the upper limit is mentioned as Q_y. The solution obtained by the candidates is stored in the memory D, wherein all the

RESEARCH ARTICLE

candidates have the space for the storage. The solutions are upgraded in D by checking the past solution. If the present solution is better then, D is upgraded; else, it is maintained in the past solution.

(i) Exploring Solution: In the randomization phase, the candidate's searches for the solution by exploring the feature space by performing various diving techniques. The expressions for the diving tactics are defined in equation (15-16).

$$M = 2 * \cos(2 * \pi * g_2) * t \tag{15}$$

$$N = 2 * B(2 * \pi * g_3) * t \tag{16}$$

Where, M refers to the diving of Gannet in U-shape and N refers to the diving of Gannet in V-shape. The expression for the iteration in the randomization phase is defined in equation (17).

$$t = 1 - \frac{\tau}{\tau_{\max}} \tag{17}$$

Where, t is the iterations utilized by GCO in the randomization phase, in which the maximal value is defined as $\tau_{\max}$. g_2 and g_3 are the random numbers with range $[0,1]$. The diving angle is defined as $B(.)$ and is formulated in equation (18).

$$B(a) = \begin{cases} -\frac{1}{\pi} * a + 1, & a \in (0, \pi) \\ \frac{1}{\pi} * a - 1 & a \in (\pi, 2\pi) \end{cases} \tag{18}$$

After making various dives randomly for exploring more space in the feature area, the corresponding locations are updated in D. The possibility of both the V and U dives are assigned equally and is defined as f. The update of memory (D) is stated in equation (19-23).

$$D_x(t+1) = \begin{cases} A_x(t) + u_1 + u_2 \ , f \geq 0.5 \\ A_x(t) + v_1 + v_2 \ , f < 0.5 \end{cases} \tag{19}$$

Where,

$$u_2 = K * (A_x(t) - A_e(t)) \tag{20}$$

$$v_2 = L * (A_x(t) - A_c(t)) \tag{21}$$

$$K = (2 * g_4 - 1) * M \tag{22}$$

$$L = (2 * g_5 - 1) * N \tag{23}$$

Where, randomly chosen solution is referred as $A_e(t)$, the solutions evaluated at the present iteration $A_c(t)$ is averaged based on equation (24).

$$A_c(t) = \frac{1}{U} \sum_{x=1}^{U} A_x(t) \tag{24}$$

The range of u_1 is $[-M, M]$ and the range of v_1 is $[-N, N]$.

(ii) Exploiting Solution: In the randomization search, the candidates identify a solution globally, which is further exploited deeply in the local search for acquiring the solution. Here, the candidate uses its capturability behaviour for capturing the target. It is defined as in equation (25).

$$C = \frac{1}{G * t_2} \tag{25}$$

Where, the iterations utilized in the local search is indicated as t_2 and is formulated as in equation (26).

$$t_2 = 1 + \frac{\tau}{\tau_{\max}} \tag{26}$$

Let G be the energy of the candidate to capture the target, which depends on the mass and velocity. It is defined as in equation (27).

$$G = \frac{H * s^2}{P} \tag{27}$$

where, G is the mass with 2.5kg, s is the velocity with the value 1.5m/s and P is a variable that is estimated as in equation (28).

$$P = 0.2 + (2 - 0.2) * g_6 \tag{28}$$

Where, the random variable is indicated as g_6 and has the value of $[0,1]$. Then, the solution update is stated as in equation (29).

$$D_m(t+1)_{Gannet} = \begin{cases} t * \gamma * (A_x(t) - A_{better}(t)) + A_x(t), & C \geq d \\ A_{better}(t) - (A_x(t) - A_{better}(t)) * R * t & C < d \end{cases} \tag{29}$$

Where, the best candidate is indicated as $A_{better}(t)$ and the factors γ *and* R are estimated as in equation (30-31).

RESEARCH ARTICLE

$$\gamma = C * \left| A_x(t) - A_{better}(t) \right| \quad (30)$$

$$R = Levy(V) \quad (31)$$

Here, the levy flight behavior is utilized by the candidate to capture the solution and is indicated as R and is expressed as in equation (32-33).

$$Levy(V) = 0.01 \times \frac{\alpha \times \beta}{|v|^{1/\mu}} \quad (32)$$

Where,

$$\beta = \left(\frac{\Gamma(1+\mu) \times \sin\left(\frac{\pi\mu}{2}\right)}{\Gamma\left(\frac{1+\mu}{2}\right) \times \mu \times 2^{\left(\frac{\mu-1}{2}\right)}} \right)^{1/\mu} \quad (33)$$

The values of the random variables $\gamma \; and \; \beta$ lie in the range of $[0,1]$ and the predefined constant μ has the value of 1.5. Here, in the Gannet optimization, the sudden turning of the cunning fish escapes from the Gannet and hence the capturing of solution is not possible and hence the Gannet searches for another fish. Thus, in order to minimize the capability of fish escaping, the attacking behaviour of the chimp is incorporated in the proposed GCO algorithm.

The solution is updated by the chimp is devised based on all the four types of chimps. Its solution update is expressed as in equation (34).

$$D_x(t+1) = \frac{D_A + D_B + D_D + D_C}{4} \quad (34)$$

where, the solution update is indicated as $D_x(t+1)$, the solution obtained by the attacker is notated as D_A, the solution acquired by the barrier is notated as D_B, the solution updated by the driver is represented as D_D, the solution acquired by the carrier is indicated as D_C. Here, the position updated by the individual chimp is expressed as in equation (35-38).

$$D_A = D_1 - k_1(q_A) \quad (35)$$

$$D_B = D_2 - k_2(q_B) \quad (36)$$

$$D_C = D_3 - k_3(q_C) \quad (37)$$

$$D_D = D_4 - k_4(q_D) \quad (38)$$

Where, q_A refers to the distance between the target and the attacker chimp, q_B refers to the distance between the target and the barrier chimp, q_C refers to the distance between the target and the carrier chimp, and q_D refers to the distance between the target and the driver chimp. The coefficient $k_1, k_2, k_3, and \; k_4$ ranges between $[0,1]$ forces the candidates to capture the target. $D_1, D_2, D_3, and \; D_4$ refers to the best solutions acquired by the attacker, barrier, carrier and driver. Then, the hybridized solution updating using the proposed GCO is formulated as in equation (39-40).

$$D_x(t+1) = 0.5 D_x(t+1)_{Gannet} + 0.5 D_x(t+1)_{Chimp} \quad (39)$$

$$W_m(T+1) = \begin{cases} 0.5\left[t * \gamma * \left(A_x(t) - A_{better}(t)\right) + A_x(t)\right] + \\ 0.5\left[\dfrac{D_A + D_B + D_D + D_C}{4}\right], & C \geq d \\ 0.5\left[A_{better}(t) - \left(A_x(t) - A_{better}(t)\right) * R * t\right] + \\ 0.5\left[\dfrac{D_A + D_B + D_D + D_C}{4}\right], & C < d \end{cases} \quad (40)$$

(iii) Feasibility Estimation: For the solutions updated in the previous stage the feasibility is evaluated through the multi-objective fitness function defined in equation (12).

(iv) Stopping Criteria: The attainment of $\tau_{\max}$ or the optimal best solution stop the iteration of the algorithm. The pseudo-code for the proposed GCO algorithm is depicted in Algorithm 1.

Initialize the $\tau_{\max}$, U and V

Locate the population (candidate) of Gannet in the search space

Create the memory matrix D

Estimate the fitness for all the updated solutions

While

If $f \geq 0.5$

Update the solution using equation (18) based on first condition

Else

RESEARCH ARTICLE

Update the solution using equation (18) based on second condition

End if

If $d \geq 0.2$

Update the solution using equation (40) based on first condition

Else

Update the solution using equation (40) based on second condition

End if

Recheck the feasibility of the solution

Replace the memory matrix D with best solution

End while

$t = t + 1$

end

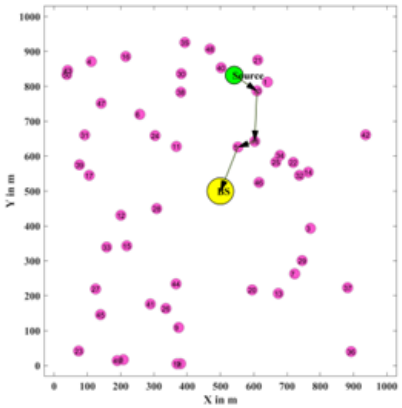
Algorithm 1 Pseudo-Code for Proposed GCO Algorithm

Thus, using the GCO algorithm, the optimal best path with multi-hop energy efficient routing is chosen for D2D communication between the users in the 5G networks.

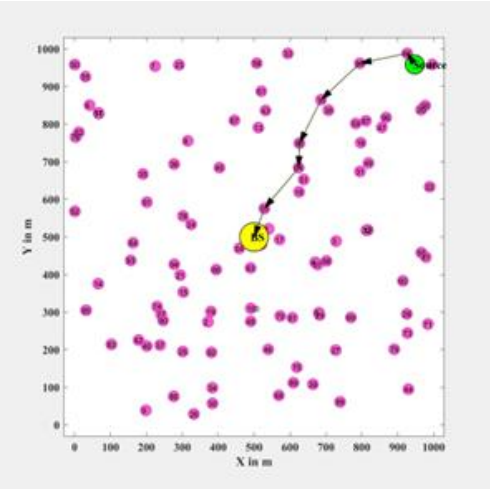
4. RESULTS AND DISCUSSION

The proposed energy efficient multi hop routing protocol is implemented in MATLAB using Windows 10 OS, and 8GB RAM PC. The experimental outcome is measured through various assessment measures to depict the excellence of the devised model. For this, the conventional energy efficient D2D routing protocols like MBLCR [25], Modified Derivative Algorithm [21], 5G-EECC [22], and DRL [24] are utilized for comparison with the proposed method.

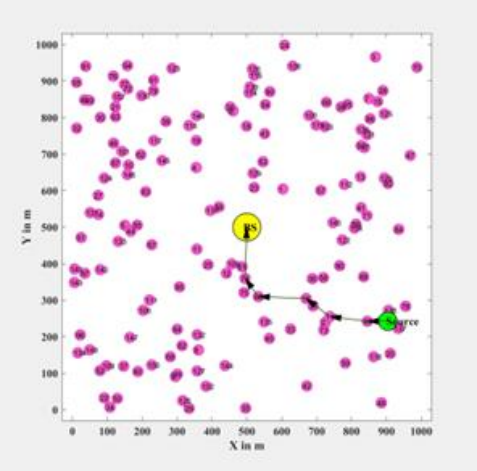
4.1. Simulation Outcome



(a)



(b)



(c)

Figure 5 Simulation Outcome of the Proposed Routing Protocol Based on (a) 50 Nodes, (b) 100 Nodes and (c) 150 Nodes

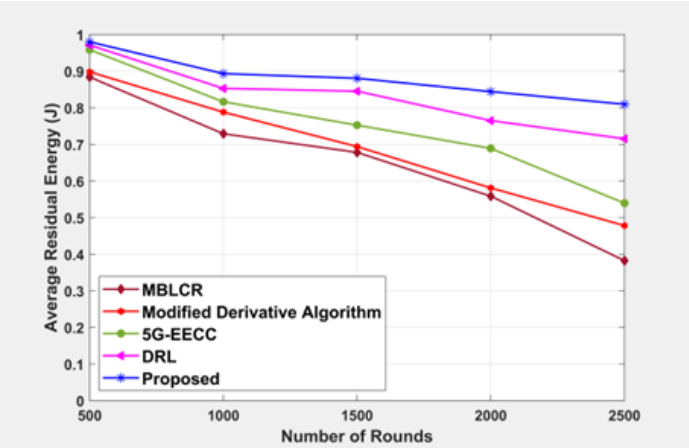
The simulation outcome of the proposed D2D communication between the users by varying the rounds is portrayed in Figure 5. Here, the communication between the users in the 5G network is devised through multi hop path by considering multi-objective fitness function. Besides, the deep learning based path detection and optimal path selection criteria enhance the energy efficiency of the proposed routing protocol.

4.2. Analysis based on Average Residual Energy

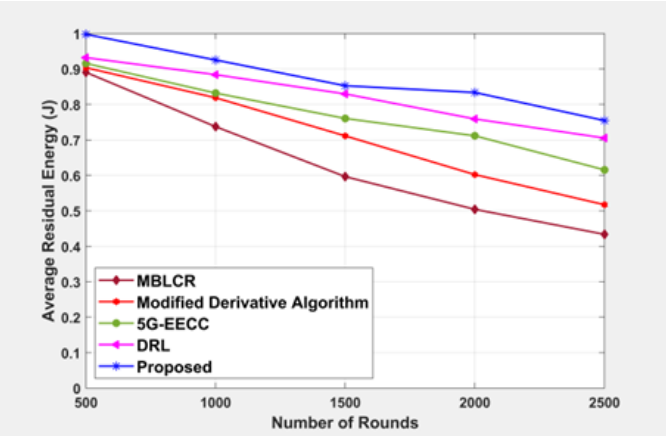
The estimation of the energy depleted by the node during the communication between the users in the network is obtained through the average residual energy estimation. The average

RESEARCH ARTICLE

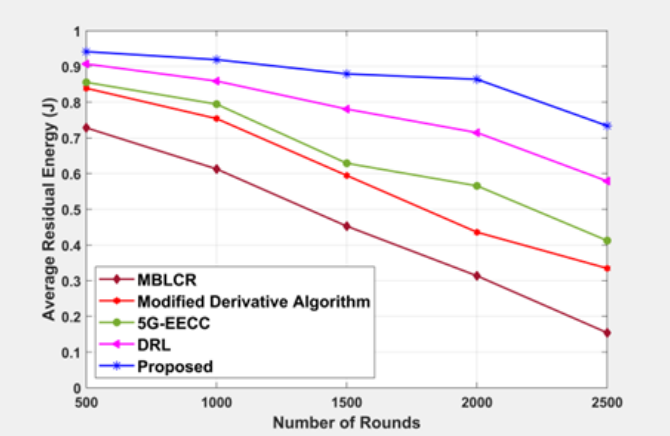
residual energy of the proposed energy efficient routing protocol along with the comparative methods is depicted in Figure 6. The average residual energy with 50 nodes is depicted in Figure 6(a), 100 nodes in Figure 6(b) and 150 nodes in Figure 6(c). The average residual energy of the proposed method with 500 rounds is 0.98, which is 9.81%, 8.37%, 2.18%, and 0.90% improved outcome than the MBLCR, Modified derivative algorithm, 5G-EECC, and DRL methods. For 2500 round, the average residual energy of the proposed method is 0.81, which is 52.80%, 40.99%, 33.41%, and 11.68% improved outcome than the MBLCR, Modified derivative algorithm, 5G-EECC, and DRL methods. Thus, the analysis by varying the number of rounds depicts that the increase in rounds depletes the residual energy due to the increase in energy consumption with larger number of rounds. However, the residual energy of the proposed method is better than the traditional methods due to the energy efficient routing protocol design. The consideration of energy consumption in detecting the possible paths using the double deep Q learning and the optimal path selection helps to accomplish the maximal residual energy compared to the conventional methods. The more detailed analysis is depicted in Table 2.



(a)



(b)



(c)

Figure 6 Average Residual Energy for (a) 50 Nodes, (b) 100 Nodes and (c) 150 Nodes

Table 2 Analysis based on Average Residual Energy

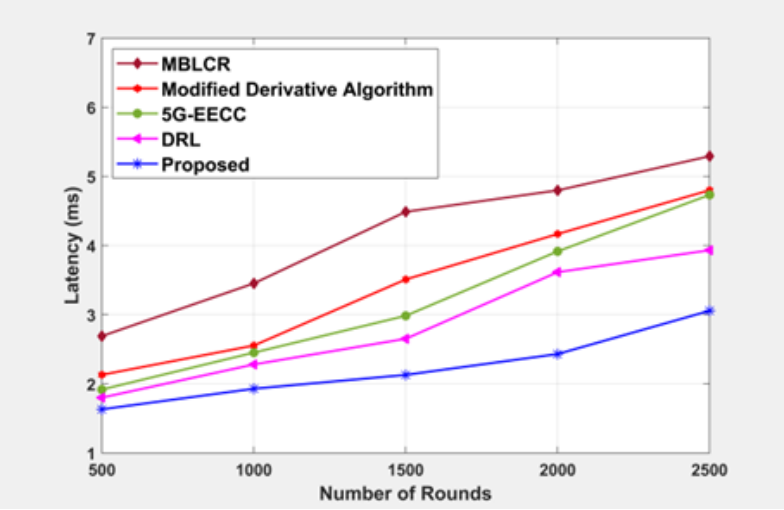
Methods/ Rounds	MBLCR	Modified Derivative Algorithm	5G-EECC	DRL	Proposed
Using 50 nodes					
500	0.88	0.90	0.96	0.97	0.98
1000	0.73	0.79	0.82	0.85	0.89

RESEARCH ARTICLE

1500	0.68	0.69	0.75	0.85	0.88
2000	0.56	0.58	0.69	0.76	0.84
2500	0.38	0.48	0.54	0.72	0.81
Using 100 nodes					
500	0.89	0.90	0.92	0.93	1.00
1000	0.74	0.82	0.83	0.88	0.93
1500	0.60	0.71	0.76	0.83	0.85
2000	0.50	0.60	0.71	0.76	0.83
2500	0.43	0.52	0.62	0.71	0.75
Using 150 nodes					
500	0.73	0.84	0.86	0.91	0.94
1000	0.61	0.75	0.79	0.86	0.92
1500	0.45	0.59	0.63	0.78	0.88
2000	0.31	0.44	0.57	0.71	0.86
2500	0.15	0.33	0.41	0.58	0.73

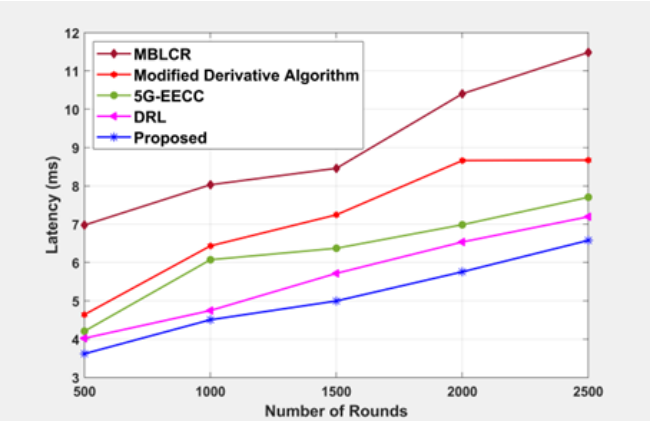
4.3. Analysis based on Latency

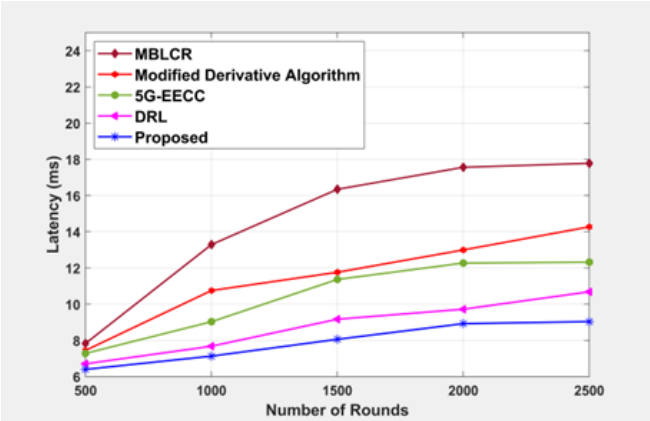
The time taken for sharing the communication request between the sender and the destination measures the latency that helps to measure the communication delay in the network. The analysis of the D2D routing protocol based on the latency assessment is portrayed in Figure 7. For example, the analysis with 1000 rounds and 100 nodes, the proposed method evaluated the latency of 4.51, which is 43.84%, 29.90%, 25.74%, and 4.99% improved outcome than the MBLCR, Modified derivative algorithm, 5G-EECC, and DRL methods. Here, the analysis indicates the elevated outcome of the proposed routing protocol by acquiring minimal latency compared to the conventional methods. In the proposed GCO algorithm for the optimal path selection, a multi-objective fitness is considered, wherein the packet latency is considered as one of the significant parameter for selecting the optimal best path. Thus, the optimal path selection criterion using the proposed GCO algorithm reduces the D2D communication latency and hence enhances the efficiency of the routing protocol. The detailed description is presented in Table 3.



(a)

RESEARCH ARTICLE





(b) (c)

Figure 7 Latency Analysis for (a) 50 Nodes, (b) 100 Nodes and (c) 150 Nodes

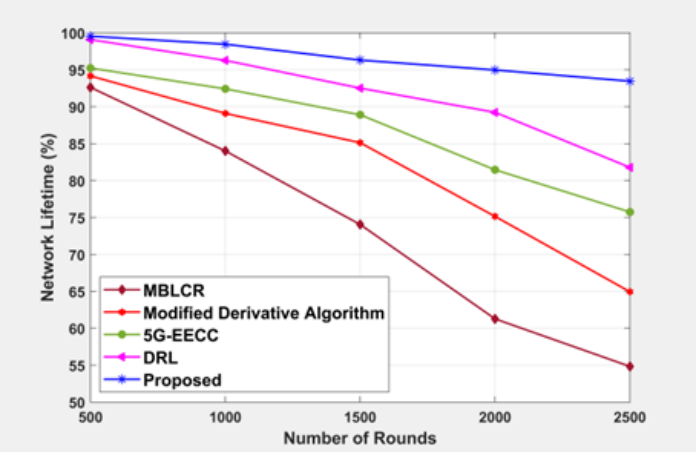
Table 3 Analysis based on Latency

Methods/ Rounds	MBLCR	Modified Derivative Algorithm	5G-EECC	DRL	Proposed
Using 50 nodes					
500	2.69	2.13	1.92	1.80	1.63
1000	3.45	2.55	2.45	2.28	1.93
1500	4.49	3.51	2.98	2.65	2.13
2000	4.80	4.17	3.92	3.62	2.43
2500	5.29	4.80	4.73	3.93	3.06
Using 100 nodes					
500	6.98	4.64	4.21	4.02	3.62
1000	8.03	6.43	6.07	4.75	4.51
1500	8.46	7.25	6.37	5.72	5.00
2000	10.40	8.66	6.99	6.54	5.76
2500	11.48	8.67	7.70	7.19	6.58
Using 150 nodes					
500	7.83	7.44	7.28	6.70	6.40
1000	13.29	10.75	9.03	7.67	7.13
1500	16.34	11.76	11.37	9.17	8.06
2000	17.56	12.99	12.27	9.72	8.92
2500	17.78	14.27	12.32	10.68	9.04

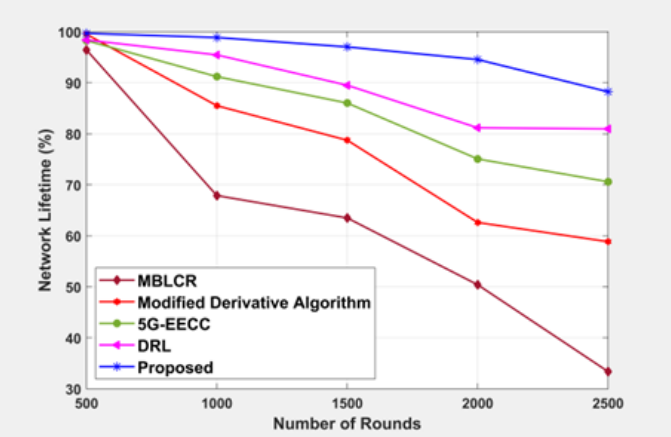
RESEARCH ARTICLE

4.4. Analysis based on Network Lifetime

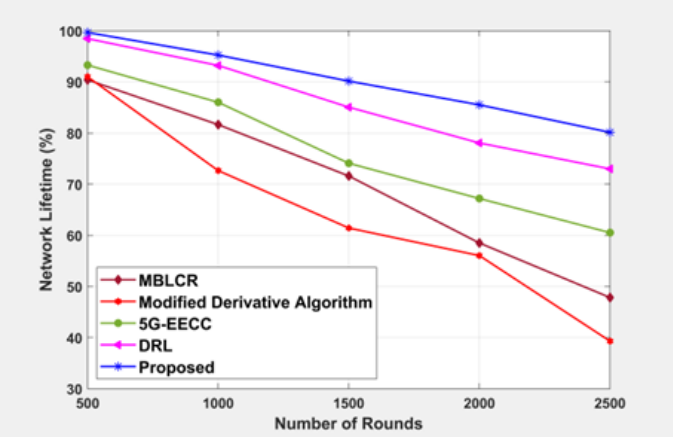
The time period of the network until the first node drops out of energy is considered as the network lifetime. The network lifetime based analysis is depicted in Figure 8 and its detailed analysis is presented in Table 4. For example, using 150 nodes and 2000 communication rounds, the proposed method acquired 85.52% of network lifetime, which is 31.60%, 34.48%, 21.43%, and 8.72% improved outcome than the MBLCR, Modified derivative algorithm, 5G-EECC, and DRL conventional methods. The network lifetime depends on the energy of the nodes. Once the energy of the network gets depleted and it runs out of energy, the network becomes dead and hence the lifetime gets completed. The proposed multi hop routing protocol considers energy consumption as a significant criteria for D2D communication between the users, wherein the possible multi hop paths for D2D communication utilizes energy consumption for identifying the next hop using the double deep Q learning. Here, the consideration of energy efficient node for communication between the users enhances the residual energy of the node. The network with minimal energy consumption further enhances the lifetime of the network. Thus, the energy efficient multi hop routing protocol design enhances the lifetime of the network.



(a)



(b)



(c)

Figure 8 Network Lifetime Analysis for (a) 50 Nodes, (b) 100 Nodes and (c) 150 Nodes

Table 4 Analysis based on Network Lifetime

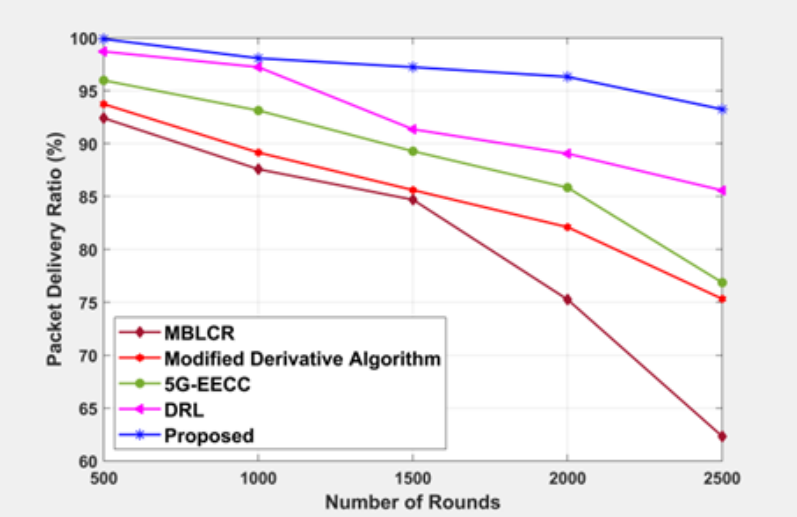
Methods/ Rounds	MBLCR	Modified Derivative Algorithm	5G-EECC	DRL	Proposed
Using 50 nodes					
500	92.62	94.17	95.24	99.09	99.55

RESEARCH ARTICLE

1000	84.02	89.10	92.42	96.28	98.47
1500	74.05	85.13	88.91	92.52	96.31
2000	61.28	75.15	81.46	89.24	94.98
2500	54.84	64.94	75.75	81.78	93.46
Using 100 nodes					
500	96.44	99.53	98.27	98.40	99.69
1000	67.88	85.50	91.21	95.46	98.87
1500	63.49	78.73	86.03	89.51	97.03
2000	50.40	62.61	75.07	81.17	94.55
2500	33.38	58.85	70.59	80.95	88.24
Using 150 nodes					
500	90.41	91.02	93.30	98.47	99.68
1000	81.66	72.66	86.02	93.23	95.25
1500	71.60	61.43	74.10	85.06	90.16
2000	58.50	56.03	67.19	78.06	85.52
2500	47.83	39.31	60.52	73.00	80.14

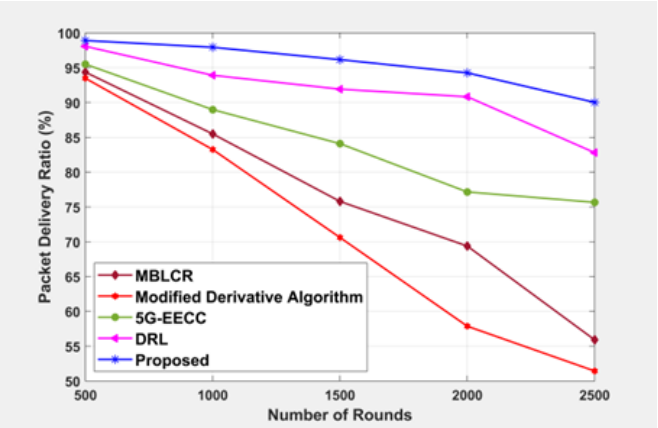
4.5. Analysis based on Packet Delivery Ratio

The ratio that estimates the number of data packets delivered to the destination to the total data shared by the sender measures the packet delivery ratio. Figure 9 depicts the packet delivery ratio by varying number of nodes in the network. In this analysis, the packet delivery ratio estimated by the proposed method is 85.38 with 150 nodes and 2500 rounds of communication that is 56.79%, 49.34%, 34.57%, and 9.62% improved outcome than the MBLCR, Modified derivative algorithm, 5G-EECC, and DRL conventional methods. The higher packet delivery rate of the proposed method depicts the efficiency of the routing protocol with minimal packet loss. The packet loss is minimized by routing the packet through the energy efficient node. The data shared through the node with higher energy minimizes the chance of information loss due to the high capability of the node in terms of lifetime. Thus, the proposed method accomplished enhanced packet delivery ratio compared to the conventional methods. The detailed analysis is presented in Table 5.

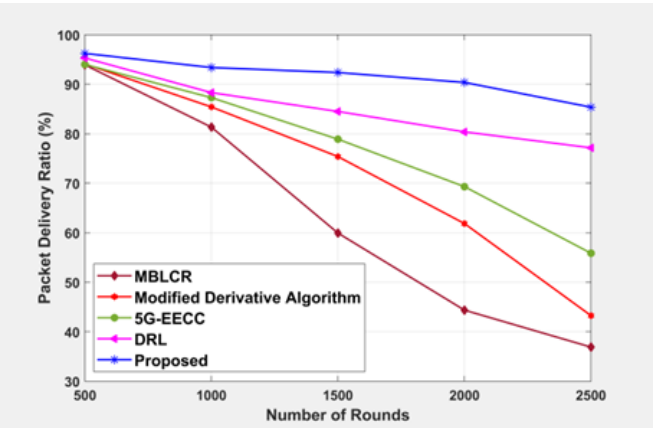


(a)

RESEARCH ARTICLE



(b)



(c)

Figure 9 Packet Delivery Ratio Analysis for (a) 50 Nodes, (b) 100 Nodes and (c) 150 Nodes

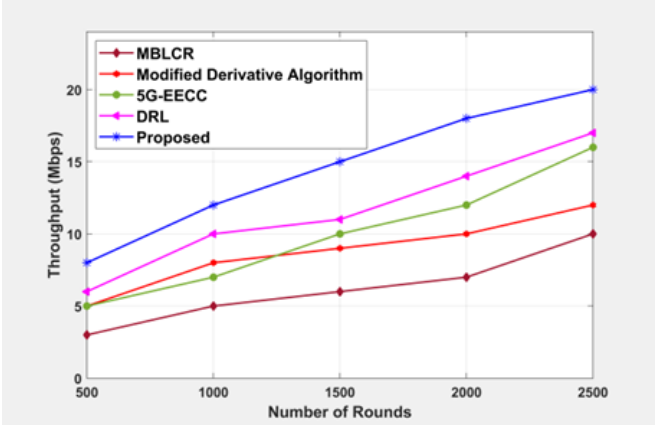
Table 5 Analysis based on Packet Delivery Ratio

Methods/ Rounds	MBLCR	Modified Derivative Algorithm	5G-EECC	DRL	Proposed
Using 50 nodes					
500	92.41	93.74	95.97	98.70	99.89
1000	87.58	89.16	93.13	97.23	98.07
1500	84.69	85.61	89.28	91.35	97.22
2000	75.25	82.10	85.84	89.05	96.31
2500	62.32	75.32	76.86	85.57	93.24
Using 100 nodes					
500	94.37	93.49	95.51	98.08	98.91
1000	85.50	83.25	88.98	93.92	97.94
1500	75.80	70.61	84.09	91.92	96.16
2000	69.38	57.87	77.16	90.83	94.26
2500	55.91	51.44	75.67	82.81	90.03
Using 150 nodes					
500	93.92	94.11	94.01	95.32	96.25
1000	81.34	85.42	87.30	88.31	93.38
1500	59.94	75.39	78.90	84.51	92.38
2000	44.37	61.86	69.32	80.39	90.37
2500	36.90	43.25	55.86	77.17	85.38

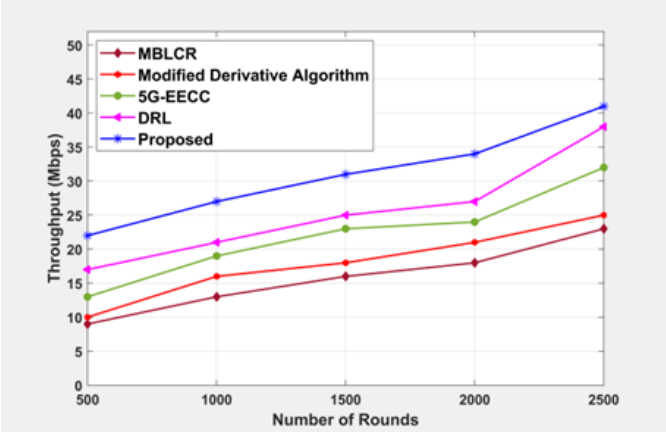
RESEARCH ARTICLE

4.6. Analysis based on Throughput

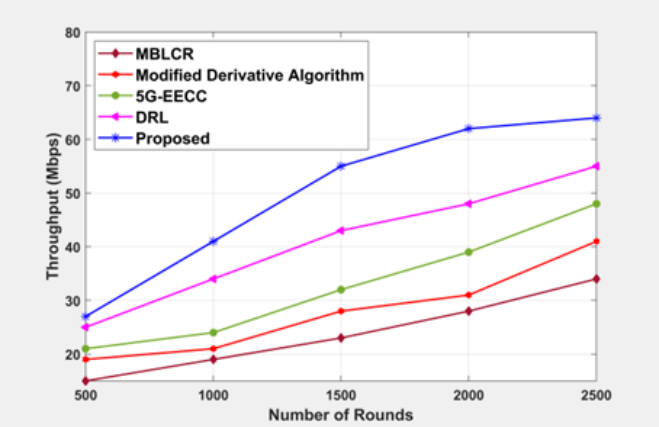
The data packet received by the destination node within the specified time is measured through the throughput analysis, which is portrayed in Figure 10. The throughput estimated by the proposed method with 1000 rounds and 100 nodes is 27, which is 51.85%, 40.74%, 29.63%, and 22.22% improved outcome than the MBLCR, Modified derivative algorithm, 5G-EECC, and DRL conventional methods. The proposed optimal route selection technique chooses the best path with minimal hop count, which helps to communicate the sender faster and hence more communication is possible within the specified time. The larger amount of communication without information loss enhances the throughput of the network. The detailed analysis is depicted in Table 6.



(a)



(b)



(c)

Figure 10 Throughput Analysis for (a) 50 Nodes, (b) 100 Nodes and (c) 150 Nodes

Table 6 Analysis based on Throughput

Methods/ Rounds	MBLCR	Modified Derivative Algorithm	5G-EECC	DRL	Proposed
Using 50 nodes					
500	3	5	5	6	8
1000	5	8	7	10	12
1500	6	9	10	11	15
2000	7	10	12	14	18

RESEARCH ARTICLE

2500	10	12	16	17	20
Using 100 nodes					
500	9	10	13	17	22
1000	13	16	19	21	27
1500	16	18	23	25	31
2000	18	21	24	27	34
2500	23	25	32	38	41
Using 150 nodes					
500	15	19	21	25	27
1000	19	21	24	34	41
1500	23	28	32	43	55
2000	28	31	39	48	62
2500	34	41	48	55	64

4.7. Comparative Discussion

The proposed energy efficient routing protocol with multi hop for D2D communication between the users in the 5G network accomplished enhanced performance while assessing the performance based on various measures like packet delivery ratio, latency, residual energy, throughput and network lifetime. The proposed method utilizes the multi-hop possible path detection using the proposed double deep Q learning technique. Here, the consideration of energy consumption between the nodes for selecting the next hop node identifies the best energy efficient node for communication. Besides, the consideration of Deep CNN for estimating the Q-value and reward function enhances the detection accuracy of finding the possible paths by solving the over optimistic issues. Also, the proposed GCO algorithm utilizes the multi-objective fitness function for finding the optimal best path for communication among the identified paths. Thus, the consideration of the combined behavior of the double deep Q learning along with the GCO algorithm helps to identify the optimal best energy efficient path for D2D communication and is depicted based on various assessment measures.

5. CONCLUSION

An energy efficient multi hop routing protocol is introduced in this research for D2D communication between the 5G network users. Here, a deep reinforcement learning technique named double deep Q learning is proposed for the identification of multi hop paths for D2D communication. In this, the Deep CNN is introduced for the estimation of the Q-value and reward function of the double deep Q learning for enhancing the path detection accuracy and to solve the issue concerning the over optimization. Also, a hybrid optimization named GCO is introduced by hybridizing the hunting behavior of the Gannet with the chimp to obtain the global best solution in choosing the optimal best path. The balanced exploration and exploitation capability of the proposed GCO algorithm with multi-objective fitness function chooses the best path for D2D communication. The assessment of the proposed method based on various measures like packet delivery ratio, latency, residual energy, throughput and network lifetime accomplished the values of 99.89, 1.63, 0.98, 64 and 99.69 respectively. In the future, a novel architecture will be designed based on fuzzy concept for the reduction of computational complexity.

REFERENCES

[1] Z. Li, C. Guo, and Y. Xuan, “A Multi-Agent Deep Reinforcement Learning Based Spectrum Allocation Framework for D2D Communications,” in 2019 IEEE Global Communications Conference (GLOBECOM), Dec. 2019, pp. 1–6. doi: 10.1109/GLOBECOM38437.2019.9013763.

[2] V. Sridhar and S. E. Roslin, “Energy Efficient Device to Device Data Transmission Based on Deep Artificial Learning in 6G Networks,” Int. J. Comput. Networks Appl., vol. 9, no. 5, pp. 568–577, 2022, doi: 10.22247/ijcna/2022/215917.

[3] M. Alnakhli, S. Anand, and R. Chandramouli, “Joint Spectrum and Energy Efficiency in Device to Device Communication Enabled Wireless Networks,” IEEE Trans. Cogn. Commun. Netw., vol. 3, no. 2, pp. 217–225, Jun. 2017, doi: 10.1109/TCCN.2017.2689015.

[4] M. Waqas et al., “A Comprehensive Survey on Mobility-Aware D2D Communications: Principles, Practice and Challenges,” IEEE Commun. Surv. Tutorials, vol. 22, no. 3, pp. 1863–1886, 2020, doi: 10.1109/COMST.2019.2923708.

[5] 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, doi: 10.1109/ACCESS.2020.3035325.

[6] L. Li, L. Chang, and F. Song, “A Smart Collaborative Routing Protocol for QoE Enhancement in Multi-Hop Wireless Networks,” IEEE Access, vol. 8, pp. 100963–100973, 2020, doi: 10.1109/ACCESS.2020.2997350.

RESEARCH ARTICLE

[7] X. Zhou, M. Sun, G. Y. Li, and B. H. Fred Juang, "Intelligent wireless communications enabled by cognitive radio and machine learning," China Commun., vol. 15, no. 12, pp. 16–48, 2018.
[8] K. M. Thilina, Kae Won Choi, N. Saquib, and E. Hossain, "Machine Learning Techniques for Cooperative Spectrum Sensing in Cognitive Radio Networks," IEEE J. Sel. Areas Commun., vol. 31, no. 11, pp. 2209–2221, Nov. 2013, doi: 10.1109/JSAC.2013.131120.
[9] R. Joon and P. Tomar, "Energy Aware Q-learning AODV (EAQ-AODV) routing for cognitive radio sensor networks," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 9, pp. 6989–7000, Oct. 2022, doi: 10.1016/j.jksuci.2022.03.021.
[10] J. Ramkumar and R. Vadivel, "Improved Wolf prey inspired protocol for routing in cognitive radio Ad Hoc networks," Int. J. Comput. Networks Appl., vol. 7, no. 5, pp. 126–136, 2020, doi: 10.22247/ijcna/2020/202977.
[11] M. C. Hlophe and B. T. Maharaj, "QoS provisioning and energy saving scheme for distributed cognitive radio networks using deep learning," J. Commun. Networks, vol. 22, no. 3, pp. 185–204, Jun. 2020, doi: 10.1109/JCN.2020.000013.
[12] H. B. Salameh, S. Mahasneh, A. Musa, R. Halloush, and Y. Jararweh, "Effective peer-to-peer routing in heterogeneous half-duplex and full-duplex multi-hop cognitive radio networks," Peer-to-Peer Netw. Appl., vol. 14, no. 5, pp. 3225–3234, Sep. 2021, doi: 10.1007/s12083-021-01183-6.
[13] Y. Zhi, J. Tian, X. Deng, J. Qiao, and D. Lu, "Deep reinforcement learning-based resource allocation for D2D communications in heterogeneous cellular networks," Digit. Commun. Networks, vol. 8, no. 5, pp. 834–842, Oct. 2022, doi: 10.1016/j.dcan.2021.09.013.
[14] S. Yu and J. W. Lee, "Deep Reinforcement Learning Based Resource Allocation for D2D Communications Underlay Cellular Networks," Sensors, vol. 22, no. 23, p. 9459, Dec. 2022, doi: 10.3390/s22239459.
[15] X. Li, G. Chen, G. Wu, Z. Sun, and G. Chen, "Research on Multi-Agent D2D Communication Resource Allocation Algorithm Based on A2C," Electronics, vol. 12, no. 2, p. 360, Jan. 2023, doi: 10.3390/electronics12020360.
[16] S. H. A. Kazmi, F. Qamar, R. Hassan, and K. Nisar, "Routing-Based Interference Mitigation in SDN Enabled Beyond 5G Communication Networks: A Comprehensive Survey," IEEE Access, vol. 11, pp. 4023–4041, 2023, doi: 10.1109/ACCESS.2023.3235366.
[17] J. Zhang, W. Gao, G. Chuai, and Z. Zhou, "An Energy-Effective and QoS-Guaranteed Transmission Scheme in UAV-Assisted Heterogeneous Network," Drones, vol. 7, no. 2, p. 141, Feb. 2023, doi: 10.3390/drones7020141.
[18] X. Li, G. Chen, G. Wu, Z. Sun, and G. Chen, "D2D Communication Network Interference Coordination Scheme Based on Improved Stackelberg," Sustainability, vol. 15, no. 2, p. 961, Jan. 2023, doi: 10.3390/su15020961.
[19] D. Han and J. So, "Energy-Efficient Resource Allocation Based on Deep Q-Network in V2V Communications," Sensors, vol. 23, no. 3, p. 1295, Jan. 2023, doi: 10.3390/s23031295.
[20] P. Tam, R. Corrado, C. Eang, and S. Kim, "Applicability of Deep Reinforcement Learning for Efficient Federated Learning in Massive IoT Communications," Appl. Sci., vol. 13, no. 5, p. 3083, Feb. 2023, doi: 10.3390/app13053083.
[21] L. Nagapuri et al., "Energy Efficient Underlaid D2D Communication for 5G Applications," Electronics, vol. 11, no. 16, p. 2587, Aug. 2022, doi: 10.3390/electronics11162587.
[22] N. Khan, I. A. Khan, J. U. Arshed, M. Afzal, M. M. Ahmed, and M. Arif, "5G-EECC: Energy-Efficient Collaboration-Based Content Sharing Strategy in Device-to-Device Communication," Secur. Commun. Networks, vol. 2022, pp. 1–13, Jan. 2022, doi: 10.1155/2022/1354238.
[23] I. Ioannou, C. Christophorou, V. Vassiliou, and A. Pitsillides, "A novel Distributed AI framework with ML for D2D communication in 5G/6G networks," Comput. Networks, vol. 211, p. 108987, Jul. 2022, doi: 10.1016/j.comnet.2022.108987.
[24] M. K. Chamran, K.-L. A. Yau, M. H. Ling, and Y.-W. Chong, "A Hybrid Route Selection Scheme for 5G Network Scenarios: An Experimental Approach," Sensors, vol. 22, no. 16, p. 6021, Aug. 2022, doi: 10.3390/s22166021.
[25] V. Tilwari, T. Song, and S. Pack, "An Improved Routing Approach for Enhancing QoS Performance for D2D Communication in B5G Networks," Electronics, vol. 11, no. 24, p. 4118, Dec. 2022, doi: 10.3390/electronics11244118.
[26] J.-S. Pan, L.-G. Zhang, R.-B. Wang, V. Snášel, and S.-C. Chu, "Gannet optimization algorithm : A new metaheuristic algorithm for solving engineering optimization problems," Math. Comput. Simul., vol. 202, pp. 343–373, Dec. 2022, doi: 10.1016/j.matcom.2022.06.007.
[27] M. Khishe and M. R. Mosavi, "Chimp optimization algorithm," Expert Syst. Appl., vol. 149, p. 113338, Jul. 2020, doi: 10.1016/j.eswa.2020.113338.

Authors

Md. Tabrej Khan was born in Bokaro, India, in 1981. He received the M.Sc. degree in computer science from Jamia Hamdard University, Delhi, India, in 2008.In 2008, he started his career as a Software Developer at Software Company, Delhi. He is currently pursuing PhD from Faculty of Computer science Pacific Academy of Higher Education and Research University Udaipur (Rajasthan), India. He is also an excellent teacher and a talented researcher with over seven years of teaching and research experience in 5G, Deep Learning, Machine Learning, and image processing. He has produced many publications in the journal of international repute and presented articles at international conferences. His current research interests include 5G, deep learning, medical informatics, and machine learning. He is also a member of the International Association of Engineers (IAENG) and a member of the following societies: the IAENG Society of Bioinformatics, the IAENG Society of Computer Science, and the IAENG Society of Data Mining.

Dr. Ashish Adholiya is right now working as Assistant Professor at Pacific Academy of Higher Education and Research University Udaipur (Rajasthan), India. He pursued his Ph.D. in the area of Database Flexibility from JRN Rajasthan Vidyapeeth (Deemed to be University, NAAC –A Grade). He has a total experience of 13 years in academics and 3 years of IT companies. He has authored 29 research articles for international journals and 23 research articles for national journals with impact factor. He also has been contributed two chapters in ISBN books of international repute. He has been conducting Management Development Program to various organizations like IOC. He is managing editor of two national journals published by Pacific University, Udaipur since last 3 years. He has been the editor of three books published by the Pacific University, Udaipur.

RESEARCH ARTICLE

How to cite this article:

Md. Tabrej Khan, Ashish Adholiya, "Energy Efficient Multi Hop D2D Communication Using Deep Reinforcement Learning in 5G Networks", International Journal of Computer Networks and Applications (IJCNA), 10(3), PP: 401-421, 2023, DOI: 10.22247/ijcna/2023/221897.