Tumbe Group of International Journals
Vol. 6, Issue No. 2, pp.36-47, May-August 2023
ISSN: 2581-8511 https://tumbe.org

Improved Network Energy Consumption and Routing Overhead Through Environment-Aware Rewards Optimized Deep-Q-Learning for Cluster Head Selection in MANET

Haridas S[1], Dr. A. Rama Prasath[2]

[1] Research Scholar, Hindustan Institute of Technology and Science, Chennai, India
Assistant Professor, Dept. of Computer Science, Government First Grade College, Tumkur, Karnataka.
Email: harigoleson@gmail.com

[2]Assistant Professor (Selection Grade), Department of Computer Applications,
Hindustan Institute of Technology and Science, Chennai, India.
E-mail: mraprasath@gmail.com

Abstract

It is obvious that MANETs are dynamic; as a result, network performance declines as the network size grows. Such a problem can be mitigated by implementing clustering. Clustering improves wireless network scalability while decreasing network overhead. In the MANET context, mobile nodes are clustered to reduce processing complexity. A cluster is a group of divided nodes. In a MANET, clustering separates a collection of mobile nodes into virtual logical groupings based on specific criteria. Each cluster comprises a cluster head referred to as CH, cluster members, and cluster gateway; all play distinct functions in the cluster during data transfer in a MANET. Cluster head selection and cluster creation are the two stages of node clustering. Energy state, node degree, distance, trust level, and node mobility are considered when determining the cluster heads' score values. The cluster leader is chosen from among the nodes having the highest score to maintain the maximum cluster size and enhance cluster stability. In the present work, cluster head selection is implemented using Reward optimized DQN-based algorithm. Appropriate rewards are selected by timely environmental surrounding awareness information. The algorithm develops an ideal strategy for CH's selection by continuously learning the network state through forwarding packets and feedback from packets. The RoDQL clustering algorithm simulation ensured that the number of nodes in each cluster was balanced. The cluster head will access each node's connection information, allowing the malicious nodes involved in the Worm Hole attacks to be discovered and destroyed during routing. Better network energy consumption and routing overhead ensured the effectiveness of the proposed algorithm.

Keywords: MANET, Clustering, DQN, Security, Energy.

1 Introduction

It is obvious that MANETs are dynamic; as a result, network performance declines as the network size grows. Such a problem can be mitigated by implementing clustering. Clustering improves wireless network scalability while decreasing network overhead. In the MANET context, mobile nodes are clustered to reduce processing complexity. Clustering offers a robust and resource-efficient network while resolving most of MANET's problems. We considered cluster-based environment which minimizes complexity and overhead in packets and control messages forwarding.

2. Related work

Clustering is a crucial strategy for resolving various MANET issues. It also increases network lifespan and scalability. Furthermore, cluster-based routing improves network administration by reducing the number of nodes in the routing table. Cluster Heads, on the other hand, can handle the extra work burden of communication. As a result, CH energy is drained sooner, and the demise of CHs later divides the network, which shortens the network lifespan. Furthermore, node mobility is the primary cause of connection failure. (Mehrkanoon et al., 2014).

The Stabilized Clustering technique enhances cluster formation stability while simultaneously increasing efficiency. The Moth Flame Optimization technique is used in this approach to determine the CH using the QoS standard. In this technique, the helper CH helps to prevent the CH from deteriorating, allowing it to run successfully (Wang and Qing, 2010).

The QoS-guided Dynamic Scheduling technique identifies the ideal locations for all workloads and schedule clusters. It is a lightweight technique that ensures QoS (Alowish et al., 2020). QoS-oriented scheduling and auto-scaling technique are utilized to schedule jobs in the cluster. This strategy focuses on the critical QoS need. This inevitability predicts if a job will be talented before its target and forecasts appropriate resource formation. However, this method is ineffective. The fuzzy-based clustered network increases complexity since more fuzzy rules increase communication overhead, leading to uncertainty.

The routing problem is treated as a Markov decision process, which considers how to route packets in an ideal communication channel (MDP). Unsupervised learning methods, such as reinforcement learning (RL), have also shown useful in learning appropriate MDP rules. Early routing studies relied on the Q-Learning algorithm, the most widely used RL approach (T. Hu and Y. Fei, 2010). On the other hand, the Q-Learning method suffers from slow convergence for larger action spaces. This disadvantage was recently resolved with the introduction of the deep Q-network (DQN) method. (R. Ding et al. 2019) uses DQN for routing in a high-traffic network to minimize network congestion, whereas (A. M. Koushik et al.,2019) create a DQN model to identify the ideal link between nodes. Both, however, need a central unit powerful enough to compute and manage the actions of every node. Deep reinforcement learning emerges as a potential substitute to solve decision-making type problems in this case. DRL, in contrast to traditional reinforcement learning methods, can solve practical problems with large-scale state and action space.

However, no effective scheme exists for intelligently selecting cluster heads in a dynamic network environment. Hence, an efficient clustering strategy is proposed for a large-scale MANET, reducing network energy consumption and increasing the attack detection rate.

Current research proposes an energy-efficient, lifetime-aware, adaptive, and Environment-aware stable Clustering to address the abovementioned challenges. It uses the Rewards optimized Deep-Q-Learning (RoDQL) process to model the dynamic cluster head selection.

3. Problem Statement.

Providing security in the MANET is a big issue because of upcoming factors such as dynamic topology, communication latency, network scalability and high processing security algorithms. Authentication is the main process in MANET security which verifies the credentials provided during the time of registration process. The number of users in the MANET environment is huge and it is not possible to manage the users individually hence clustering of mobile nodes is introduced, this process significantly reduces the complexity of managing the mobile nodes. We considered cluster-based environment which minimizes complexity and overhead in packets and control messages forwarding. Environment aware Clustering is proposed which uses rewards optimized deep-Q-learning (RoDQL) that considers energy status, number of neighbors, mobility and distance. To overcome the challenges faced in routing message packets, the RoDQL model is proposed over clustered MANET environment

4. Comparison Study

This subsection describes evaluation of the proposed blockchain based security model in terms of several QoS metrics. The proposed model is compared with state-of-the-art works. In particular, we considered the following performance metrics as attack detection rate, false positive rate, end-to-end delay, packet delivery ratio, energy consumption, throughput, route overhead ratio, and security strength. Table 1 shows the comparison of existing approaches.

Table.1. Drawbacks of Existing Approaches

Existing work	Contributions	Drawbacks
E2SR [32]	(1). A hash chain dependent certificate authentication (HCCA) is proposed for authentication (2). Then clusters are formed and here dual cluster heads are elected for data transmission. (3). Secure route established between the sources to destination via worst case particle swarm optimization algorithm. (4). Data packets are encrypted before transmission to secured path by means of XOR RC6 encryption with fuzzy logic	•Low level security • High processing time (encryption and decryption) • Suitable for Small number of devices • Higher energy consumption

Multi-Path [33]	(1). Clustered formed using Fuzzy Naïve Bayes algorithm. (2). Secure nodes are selected by hybrid optimization (BSO + WOA). (3). The selection of optimal route is based on the fitness factors as energy, trust, connectivity and throughput.	• Irrelevant optimization algorithms are combined so it is not an optimum solution. • Higher energy consumption •Higher complexity

5.General Model of Deep Q-learning

The Reinforcement Learning (RL) approach offers a paradigm in which a system may learn to achieve a target in control issues built on its experience. Reinforcement learning approaches are required to solve optimum control tasks by interacting with surroundings data. RL aims to maximize an agent's reward by performing a series of behaviours in response to a changing environment.

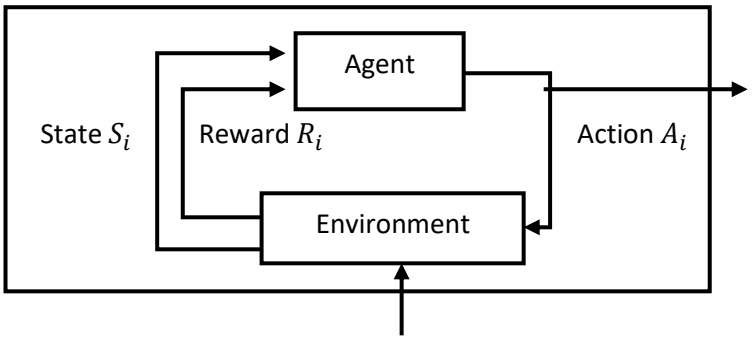

Figure 1: The collaboration between agent and environment in the DQL.

In RL, an agent chooses actions depending on the current state of a system and the reinforcement learning it gets from the environment. Most RL methods are based on approximating value functions, which are functions of state-action pairs that assess how good it is to do an action in a particular state. A reinforcement learning process with the Markov property is known as a Markov Decision Process-MDP (R.S. Sutton & A.G. Barto,1998), which is critical for understanding the idea of RL. Properties of a certain MDP are represented by a tuple of (A,S,R,P), where A, S, R, and P are the collection of actions, states, rewards, and state transition probabilities. Information probability functions are used to develop by sampling the environment and using experiences to find the best action-value functions q(s; a) for a particular state s.

DQN estimates Q by combining a convolutional neural network (CNN) and Q-Leaning (s; a). Because the CNN can output Q(s; a) of all actions when the current state s is input, it can tackle large-scale RL issues.

Meanwhile, DQ-Learning abandons the Q-table in favour of an experience replay pool to preserve each experience tuple e = (s; a; r; s'). The state is given as input, and the Q-value of all possible actions is given as output experience tuple e = (s; a; r; s'). Its behaviour is influenced

by the reward function, which provides negative or positive reinforcement to the agent once it makes a decision. Its efficiency depends on the careful design of the rewards function. The reward function indicates the current quality of the action decision, which should be developed with the primary goal of intelligent network routing control in mind.

6. Proposed methodology

Environment aware Clustering is proposed which uses rewards optimized deep-Q-learning (RoDQL) that considers energy status, number of neighbors, mobility and distance.

Clustering consists of two processes such as cluster formation and cluster management. In cluster formation, the network splits into different clusters. In each cluster, one node is elected as a CH and others are members of CH. CH are elected using several metrics. The prime motive of clustering is to efficient use of energy resources, maintain and manage routing, and location issue for solving communication and computational complexities. There are two types of cluster maintenance are given follows:

- Inter cluster maintenance – For packet forwarding/routing using more chs
- Intra cluster maintenance – For packet forwarding/routing within a cluster.

CH has the complete responsibility to monitor and manage all the nodes during packet forwarding within a cluster. Clustering is a hierarchical networking scheme that employs flat topology. In this paper, cluster heads are elected by RoDQL algorithm and it is managed by guard nodes.

a) Distance: It is defined by the distance between two nearest nodes. Assume that d_{ij} is the distance between node i and j. It is computed based on its angular position information $angle\ (\theta_2, \theta_1)$ and radius information (r_2, r_1). It is expressed as:

$$d_{ij} = \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 cos(\theta_2 - \theta_1)} \qquad (1)$$

b) Node Mobility: It is defined by the node speed. However node mobility is computed for dynamic network topology and it cause several issues such as link breakage, route failure and degrades network throughput due to increase of mobility. It is expressed as:

$$S_{n_i} = \frac{1}{T}\sum_{t=1}^{T}\sqrt{x(t) - x(t-1)^2 + y(t) - y(t-1)^2 + z(t) - z(t-1)^2} \quad (2)$$

Where S_{n_i} is the nearest node speed, which is calculated for each node in the network with coverage and $x(t) - x(t-1)$,$y(t) - y(t-1)$, & $z(t) - z(t-1)$ are the coordinates of the node at time t and t-1.

c) Residual Energy Level: It is defined by level of energy that nodes consist after certain process at a time scale t. A level of energy per bit/byte consumed for node i at time t. It is expressed as:

$$RE_{n_i} = E_P - E_T \qquad (3)$$

Where RE_{n_i} is the node residual node, E_P is the power consumption of the node in the network, and E_T is the transmission power of a node. Therefore energy consumption of a mobile node is expressed as:

$$C_E(i) = \left[T_p \times \frac{d_s}{d_r} - R_p \times \frac{d_s}{d_r}\right] + i \times L_0 \qquad (4)$$

where $C_E(i)$ is the energy consumption of node i, T_p is the power spend for transmission, d_s is the data size, d_r is the data rate, R_p is the power for receive and L_0 is loss due to overhearing.

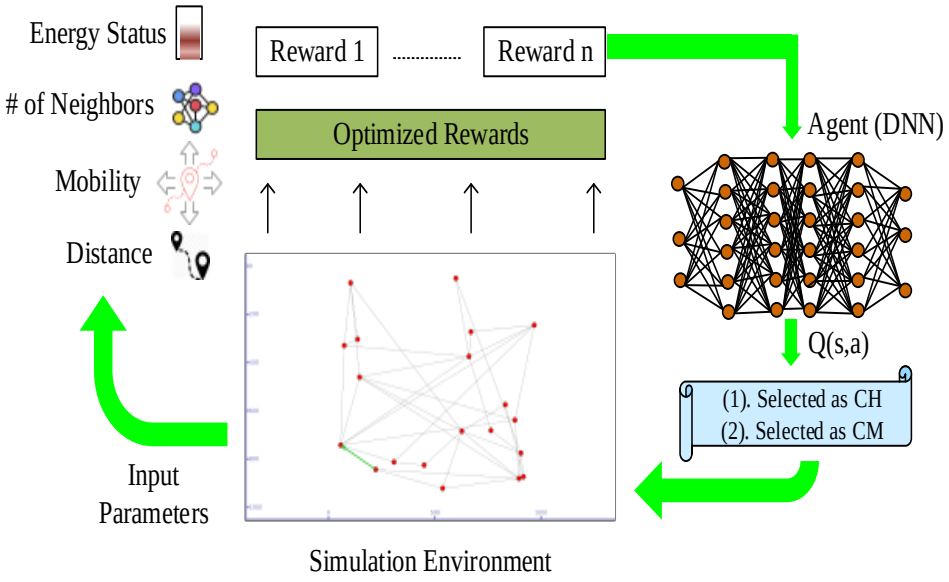

Fig.2.Flow of RoDQL algorithm

In this work, number of neighbors is known as node relative degree D_p which is computed by,

$$D_p = \left|d_p - \sqrt{N}\right| \qquad (5)$$

where d_p is the node density which is computed by:

$$d_p = \begin{cases} 1 & 0 < D_{pq} < R \\ 0 & otherwise \end{cases} \qquad (6)$$

RoDQL algorithm follows three principles such as (1) utilize deep neural network for representing the policy, value functions and model, (2). Optimize the policy, end-to-end model and value functions, and (3). It uses stochastic gradient descent. Fig 2 depicts the RoDQL algorithm.

For each node in the network $N_i\ i = \{1,2, \ldots n\}$, the MDP model consists of following elements:

$$N_i = (1,2, \ldots n) = \begin{Bmatrix} States\ (S_i), Actions\ (A_i), Transition\ Model\ (T_i), \\ and\ a\ set\ of\ Reward\ functions\ (R_i) \end{Bmatrix} \qquad (7)$$

In a given time scale t, the state $S_{i,k} \in S_i$ of N_i is the residual energy of node RE_{n_i}, trust values Tr_{n_i}, distance d_{n_i}, transmission distance Td_{n_i}, delay D_{n_i}. The description of the reinforcement learning algorithm is follows:

a) States $\boldsymbol{S_i}$: For each node N_i, states of nodes computed and change by the node RE_{n_i}, Tr_{n_i}, d_{n_i}, Td_{n_i}, and D_{n_i}. In this model S_i denotes the available set of state transitions in the environment. This element results any of the node as next hop (1st relay node R1) for packets transmission from the source node.

b) Actions $\boldsymbol{A_i}$: This denotes the set of agents action or behavior in a given time period t. It may possible to change from current state to the next state. All set of actions A_i are self-possessed by all the nodes energy value that each node can choose the next node. Thus the finite set of actions is follows:

$$A_i = \left\{RE_{ni,k} | RE_{ni,k}\{0, \delta_k, 2\delta_k, \dots, RE_{max,k}\}\right\} \quad (8)$$

Where δ_k is the step size. For any node the possible action covers: (1) Choose one of the nodes from the set of possible nodes (2). Data packets are terminated and never route the packets.

c) Transition Model $\boldsymbol{T_i}$: This model is depends on the action and states transition. It defines the state transition probabilities from state $S_{i,k}$ to $S_{i+1,k}$ and the state transition probability function is defined in below:

$$T_i = S_i \times A_i \times S_i \rightarrow [0,1] \quad (9)$$

From the result of the action $a_t \in A_i$. The selection probability of a particular forwarder node is a basis of neighbouring node routing score.

d) Reward functions $\boldsymbol{R_i}$: It is also known as reinforcement function, which purpose is to compute the immediate action a_t. It represents the state transition from one state to another state. It is computed as:

$$\rho_i = S_i \times A_i \times S_i \rightarrow R \quad (10)$$

In this stage, routing policy π_k maximize throughput of each node by reward functions. Routing policy is mapped from the given $S_{i,k}$ to $\rho_{i,k}$ that should be elected and it is written by:

$$\rho_{i,k} = \pi_k(S_{i,k}) \quad (11)$$

π_k is determined using action value function such that $Q_k^{\pi}(S_{i,k}, \rho_{i,k})$. It is an exact reward function computed starting from state $S_{i,k}$, and $\rho_{i,k}$.

The optimal policy π_k^* is the policy whose value function is greater than or equal to any other policy for all states. The final action value for the optimal policy π_k^* is also known as Q_k^* and

$Q_k^*(S_{i,k}, \rho_{i,k})$ is an optimal action for the selection of large probability score at every hop that increases reward function at all the destination

If any attack patterns found by guard node, then it will immediately isolate the particular malicious node and inform this message throughout the network.

Algorithm1: Reward optimized Deep Q-learning Algorithm **Input:** $tuple\ of\ (A, S, R, P)$ **Output:** π(s) Cluster head	
1	Set buffer capacity to N
2	Set the action value function Q and random weight θ;
3	Copy original model parameters to build the network Q , $\theta^- = \theta$;
4	for 1: M do
5	The initial state is s_1
6	for 1: T do
7	Choose an action a_t with the possibility of ε, or select the existing best with the possibility of $1-\varepsilon$;
8	$a_t = max_a Q^*(\phi(s_t), a; \theta)$
9	Complete action a_t , increment state s_{t+1} and return r_{t+1} ;
10	Store $\{\phi_t, a_t, r_t, \phi_{t+1}\}$ in D to ;
11	Arbitrarily sample $\{\phi_j, a_j, r_j, \phi_{j+1}\}$from replay buffer D ;
12	Calculate $y = \begin{cases} r_{j+1} \text{ for terminal } \phi_{j+1} \\ r_{j+1} + \gamma\, max_a\, Q(\phi_{j+1}, a; \theta^-) \text{ for non\_terminal } \phi_{j+1} \end{cases}$
13	Use the Stochastic gradient descent function to solve $(y_i - Q(\phi_j, a_j; \theta))^2$;
14	For every C step, Update parameters $\theta^- \leftarrow \theta$
15	end for
16	end for

7. Result Discussion

Cluster head selection is implemented using Reward optimized DQN-based algorithm. Appropriate rewards are selected by timely environmental surrounding awareness information. The algorithm develops an ideal strategy for CH's selection by continuously learning the network state through forwarding packets and feedback from packets. As the guard node is deployed in the network, which performs node verification, the cluster head will access each node's connection information, allowing the malicious nodes involved in the Worm Hole attacks to be discovered and destroyed during routing.

7.1 Impact of Energy Consumption

Energy consumption is a QoS based metric that determines the difference between the initial energy of node and then residual energy after the energy consumption for packet transmission or any other operations implementation in the network. However, energy consumed for several processes as packet transmission, route request, reply message reception, waiting to sleep after packet acknowledgement. Fig shows the energy consumption for the number of malicious nodes. From the graphical analysis, it is observed that the proposed blockchain model consumes lesser energy compared to E2SR and multi-path model. In particular, the proposed work has obtained 5J for 2 malicious nodes.

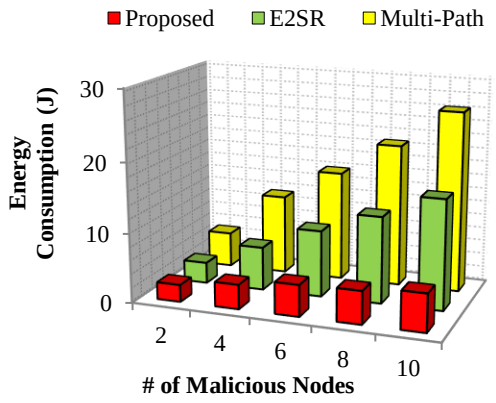

Fig.3.Energy Consumption vs. # of Malicious Nodes

Fig 3 shows the performance of energy consumption in terms of simulation time. In our proposed work, timer is used to listen the node's current state namely, sleep, listen or active. According to the network density, mobility of node, energy consumption is affected over a time. The proposed work learns environment and adaptively changes the rewards for deep reinforcement learning (DRL) algorithm.

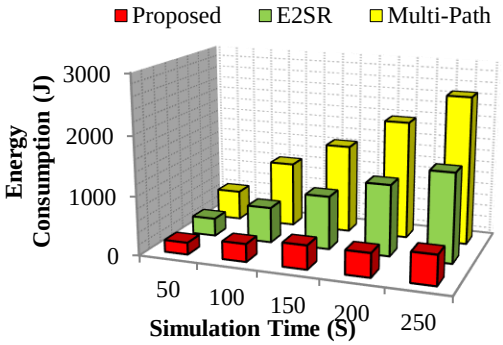

Fig.4.Energy Consumption vs. Simulation Time

7.2 Impact of Throughput

In MANET, throughput is defined as sum of data forwarded from the sender to the receiver node. On the other hand, it is defined as the complete data transmission through

communication link to the receiver node. We compute the throughput with respect to the malicious nodes count, which is depicted in fig.5.

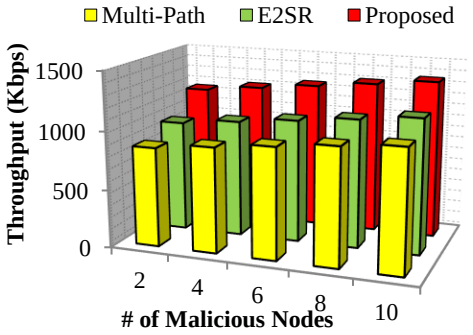

Fig.5.Throughput (Kbps) vs. # of Malicious Nodes

7.3 Impact of Routing Overhead Ratio

This metric is defined as the ratio between sums of packets generated for route selection to the sum of packets transmitted. However, routing overhead refers to the amount of routing packets forward in route discovery and maintenance.

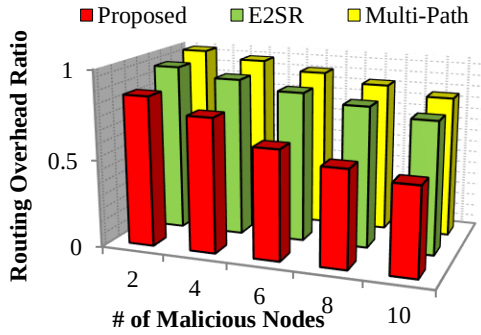

Fig.6.Routing Overhead Ratio vs. # of Malicious Nodes

These control messages forwarding introduce the routing overhead. Active route is determined by control messages to the neighbor nodes. Under low mobility environment, routing overhead is less, but highly dynamic networks produce frequent control messages forwarding. Another reason behind a high routing overhead is size of packet header transmitted through a link. Fig shows the performance of routing overhead ratio for number of malicious nodes. From the result, it is observed that the proposed model has obtained small routing overhead due to monitoring of link stability. Further, the selected route transfers packets in a reliable way. Network topology is controlled by CH, which reduces the sum of routing packets transmitted. For instance, when number of malicious node is 2, then the routing overhead by proposed model is 0.85, and the previous works are 0.95 and 0.99 for E2SR and multi-path, respectively.

The RoDQL clustering algorithm simulation ensured that the number of nodes in each cluster was balanced. Better network energy consumption and routing overhead ensured the

effectiveness of the proposed algorithm. When this algorithm is compared with other traditional approaches, it is clear that the RoDQL algorithm maintains higher residual energy and the network operation cycle will be longer.

7. References

[1] Kavitha, G. (2018). Qos improvement based security enhancement for link activity monitoring service in mobile ad hoc network. Cluster Computing.

[2] Janani, V. S., & Manikandan, M. S. K. (2017). Enhanced Security Using Cluster Based Certificate Management and ECC-CRT Key Agreement Schemes in Mobile Ad hoc Networks. Wireless Personal Communications, 97(4), 6131–6150.

[3] Krishnan, R.S., Julie, E.G., Robinson, Y.H., Kumar, R., Son, L., Tuan, T.A., & Long, H.V. (2020). Modified zone based intrusion detection system for security enhancement in mobile ad hoc networks. Wireless Networks, 26, 1275-1289.

[4] Harold Robinson, Y., & Golden Julie, E. (2019). MTPKM: Multipart Trust Based Public Key Management Technique to Reduce Security Vulnerability in Mobile Ad-Hoc Networks. Wireless Personal Communications.

[5] Janani, V.S., & Manikandan, M.S. (2020). Hexagonal Clustered Trust Based Distributed Group Key Agreement Scheme in Mobile Ad Hoc Networks. Wireless Personal Communications, 1-20.

[6] Kavitha, T., & Muthaiah, R. (2018). A light weight FFT based enciphering system for extending the lifetime of mobile ad hoc networks. Cluster Computing.

[7] Desai, A. M., & Jhaveri, R. H. (2018). Secure routing in mobile Ad hoc networks: a predictive approach. International Journal of Information Technology.

[8] Ponguwala, Maitreyi & Rao, Sreenivasa. (2019). E2-SR: A Novel Energy Efficient Secure Routing Scheme to Protect MANET from Adversaries in Internet of Things. IET Communications. 13.

[9] Veeraiah, N., & Krishna, B. T. (2020). An approach for optimal-secure multi-path routing and intrusion detection in MANET. Evolutionary Intelligence.

[10] Ochola, E.O., Mejaele, L., Eloff, M.M., & Poll, J. (2017). Manet Reactive Routing Protocols Node Mobility Variation Effect in Analysing the Impact of Black Hole Attack.

[11] Khan, B. U. I., Anwar, F., Olanrewaju, R. F., Pampori, B. R., & Mir, R. N. (2020). A Game Theory-based strategic approach to ensure reliable data transmission with optimized network operations in Futuristic Mobile Adhoc Networks. IEEE Access, 1–1.

[12] Kanagasundaram, H., & A, K. (2018). EIMO-ESOLSR: Energy Efficient and Security-Based Model for OLSR Routing Protocol in Mobile Ad-Hoc Network. IET Communications.

[13] Mehrkanoon, S., Alzate, C., Mall, R., Langone, R., & Suykens, J. A. (2014): Multiclass semisupervised learning based upon kernel spectral clustering. IEEE transactions on neural networks and learning systems, 26(4), pp.720-733.

[14] Feng, Y., Teng, G. F., Wang, A. X., & Yao, Y. M. (2007, September). Chaotic inertia weight in particle swarm optimization. In Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on(pp. 475-475). IEEE.

[15] Alowish, M., Shiraishi, Y., Takano, Y., Mohri, M., & Morii, M. (2020): Stabilized Clustering Enabled V2V Communication in an NDNSDVN Environment for Content Retrieval. IEEE Access, 8, pp.135138-135151.

[16] Haridas,S & Rama Prasath,A (2020). Enhancement of Network Lifetime in MANET: Improved Particle Swarm Optimization for Delayless and Secured Geographic Routing. Jour of Adv Research in Dynamical & Control Systems, Vol. 12, 07-Special Issue, 2020

[17] Haridas.S, A.Rama Prasath (2021), Bi-Fitness Swarm Optimizer: Blockchain Assisted Secure Swarm Intelligence Routing Protocol for MANET, Indian Journal of Computer Science and Engineering (IJCSE), ISSN:0976-Vol. 12 No. 5 Sep-Oct 2021. 5166 DOI: 0.21817/ indjcse/2021/ v12i5/211205158.

[18] Alowish, M., Shiraishi, Y., Takano, Y., Mohri, M., & Morii, M. (2020): Stabilized Clustering Enabled V2V Communication in an NDNSDVN Environment for Content Retrieval. IEEE Access, 8, pp.135138-135151.

[19] Ding, R., Xu, Y., Gao, F., Shen, X., & Wu, W. Deep reinforcement earning for router selection in a network with heavy traffic, IEEE Access, vol. 7, pp. 37 109–37 120, Apr. 2019.

[20] Koushik, A. M., Hu, F., & Kumar, S. Deep Q -learning-based node positioning for throughput-optimal communications in dynamic UAV swarm network, IEEE TCCN, vol. 5, no. 3, pp. 554–566, Sept. 2019.

[21] Hu, T., & Fei, Y. Qelar: A machine-learning-based adaptive routing protocol for energy-efficient and lifetime-extended underwater sensor networks, IEEE Trans. Mobile Comput., vol. 9, no. 6, pp. 796–809, Jun. 2010.