

# Optimized Reinforcement Learning Based Multipath Transfer Protocol in Wireless Mesh Network

S. Rajeswari<sup>1,\*</sup>, S. A. Arunmozhi<sup>1</sup> and Y. Venkataramani<sup>2</sup>

<sup>1</sup>Department of Electronics and Communication Engineering, Saranathan College of Engineering, Trichy, 620012, Tamil Nadu, India

<sup>2</sup>Dean (R&D), Saranathan College of Engineering, Trichy, 620012, Tamil Nadu, India

\*Corresponding Author: S. Rajeswari. Email: rajeswaris-ece@saranathan.ac.in

Received: 10 December 2021; Accepted: 11 February 2022

**Abstract:** Multiple radios working on different channels are used in Wireless Mesh Networks (WMNs) to improve network performance and reduce Energy Consumption (EC). Effective routing in Backbone WMNs is where each cross-section switch is well-organized with multiple Radio Interfaces (RI), and a subset of hubs is occupied as a gateway to the Internet. Most routing methods decrease the forward overheads by evolving one dimension, e.g., hop count and traffic proportion. With that idea, while considering these dimensions together, the complexity of the routing issue increases drastically. Consequently, an effective EC routing method considers a few performances simultaneously, and the requirement of MRC around the gateways is also considered. In this paper, the proposed Reinforcement Learning (RL) method based routing selection on MPR communication directs the network traffic in WMNs. Here the radio routing path selects the channel depending on the optimized node where optimization is agreed by Particle Swarm Optimization (PSO) technique. This aims to reduce the EC by switching states and utilizing efficient routing with the reduction in traffic demand. Experimental results showed better performance of throughput and EC compared with existing work.

**Keywords:** Wireless mesh network; multipath; reinforcement learning; particle swarm optimization

## 1 Introduction

Wireless Mesh Networks (WMN) initially give low energy utility over a particular region. This shows the WMN with remote access to multi-channel cabling between hubs. It is controlled with radio devices that don't need to be cabled to a wired connection like conventional Wireless Local Area Networks (WLANs). WMN conveys information over long distances, a progression of hops. Central hubs support the sign; however, they pass information from guide towards settling of sending packets dependent on their insight into the network; for example, they perform routing by first determining the network's position [1]. A WMN, for the majority of regions, comprises Mesh Gateway (MG), network switches, users of that network and a network of remote connections among them [2]. A Mesh user is a client node and is, by and large, an end-purpose of a progression of traffic through the network. The Mesh user is fixed by a



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

remote backbone designed by the MPR [3]. The MGs move near which a WMN is associated with a wired network and regularly to the Internet [4]. Accordingly, a network demands initially at a Mesh Client (MC) that would be moved through its related Mesh Route (MR) onto the remote hop, where it takes at least one hop to arrive at a Mesh Grid before arriving at the Internet [5].

The EC routing is an incorporated and static routing discovery, which implies that it won't ever change when a path from source to destination has been chosen [6]. Data gathered by hubs in the network are accessible by any remaining hubs. To find routing paths, a flooding-based path revelation is utilized. In each discovered path from source to destination, choosing and finding loads for all hubs, the hub's weight is the device's energy level [7]. The WMN joins an Institute of Electrical and Electronics Engineers (IEEE) 802.11 radio network with Long Term Evolution (LTE) network. Also, another heterogeneous routing agreement and a routing control dependent on support learning called Cognitive Heterogeneous Routing (CHR) choose the correct transmission novelty based on limits from each network [8]. The heterogeneous network overcomes the issues of sending packets over long ways, key hubs, and traffic in a remote network.

The heterogeneous remote cross-section network attacks the LTE and Wireless Fidelity (Wi-Fi) networks. Furthermore, it expands the general limit of the combined network by using unlicensed recurrence networks instead of obtaining more permissible recurrent networks for LTE [9]. Significant energy-reduced networks utilize the LTE network to adapt high traffic requests. Thus, an additional cost is that one of the other requests have more recurrent clusters [10]. In any case, LTE networks utilize authorized recurrence networks.

The WMN is a standard view created to give comprehensive network inclusion [11]. Consequently, WMNs are achievable to give a significant route network for MANs. In such networks, passages are utilized to give web association with the cross-section network [12]. Be that as it may, the significant problems of utilizing WMNs are their constraints regarding limit, framework implementation, and ensured remote connection quality. The causes of those issues begin with the multi-hop nature of the network. At the point when information packets cross more hops in an enormous WMN, they may either neglect to arrive at their objective through too many network resources. In this paper, the optimized RL of WMS for reducing the traffic demand and EC approaches are designed, and throughput achieves better utilization of the MPR method.

In this paper, the proposed RL technique based routing selection on MPR directs the traffic in WMNs. Here the radio routing path selects the channel depending on the optimized node where optimization is carried out by PSO technique. This aims to reduce the energy by switching states and utilizing efficient routing with traffic demand reduction. Experimental results showed better performance of throughput and EC when compared with existing work.

This paper is summarized as follows. Section II describes the literature survey related to this work. Section III presents the proposed approach described with RL and PSO approach. Section IV discusses the results of WMNs. Section V concludes the proposed work; finally, there is a description of future work.

## 2 Literature Review

Doraghinejad et al. [13] have presented the RL approach for Multi-Path Transmission Control Protocol (MPTCP). Another MPTCP can improve the throughput by collecting resources of the various paths together. In any case, the demonstration of MPTCP is truly impacted by the network path. With the limited assets and heterogeneous network, the default Round-Robin information scheduling in MPTCP communicates information packets for each path. Thus, it increases the throughput of MPTCP much of the time.

Remote networks are moderately stable networks apart from a periodic failure of new hubs. Q-matrix formation helps to achieve better performance of WMN. Arzani et al. [14] have presented the Q-learning approach with Multi-Path Routing (MPR) in WMS. The physical layer of the network sensing scheme is established with MPR protocol. Here the First-In-First-Out (FIFO) logic is used for reducing the packet drop rate.

Peng et al. [15] have presented the Q-learning based Fuzzy Logic Control (FLC) on mobile ad hoc networks. Unmanned Aerial Vehicle (UAV) showed the network characteristics by controlling Fuzzy Logic (FL) and Q-learning technique. Here, multi-objective/multi-routing is performed on UAVs to perform a better energy utility network. We propose another information scheduling dependent on RL with the recently presented Deep Q-Network system to upgrade the MPTCP information to execute an unbalanced method to focus on this issue. The support learning design gets the data of each way and provides choices of the most suitable method by artificial intelligence.

Ye et al. [16] have presented the RL technique to minimize the channel switching to improve the Cognitive Radio Network (CRN) result. Q-learning technique works in the cross-layer method of the network with layered compositional load, which is executed in the portability administrator to pass the channel data to the network layer. These data emerge at the Medium Access Control (MAC) layer. The channel choice is performed based on RL calculations like No-External Regret Learning, Q-Learning, and Learning Automata. This limits the channel exchanging times and client traffic in the RL based directing convention. Here the CRN tests system is dependent on Network Simulator. The routing convention concludes the better Quality of Service (QoS) of ongoing networks like Cellular and Tele-vision.

Vedantham et al. [17] have proposed Conservative Q-Learning (CQL). They summarized the disadvantages of CQL by stating that lower-bounds of Q function estimated policy value that attains its true value. They have proved that CQL performance is better than offline RL in both continuous and discrete domains. They have shown that learning policies attainment is always 2–5 times higher than the final return, specifically while learning through complex and multi-modal data distributions.

Alicherry et al. [18] have presented the MPTCP by Q-learning algorithm for mobile devices. Mobile phones can use different heterogeneous network methods by MPTCP; Now and then, boosting MPTCP throughput in a remote network is an open allow. The optimised path must be chosen, yet the prevention control factor should be selected. We discover multiple paths and traffic control for various. Therefore, we present the novel MPTCP control increasing the end client throughput, which is less critical than Q-learning's best method in various environments. The results discover a significant impact of exchanging between the various boundaries and changing on throughput improvement with the help of the traffic control system.

Kyasanur et al. [19] have proposed the multi-route/multi-hop WMN model. Remote networks with fixed hub choice are a high-throughput between a source and destination. Their performance allows loads to particular connections dependent on a packet—Expected Transmission Time (ETT) over the connection. The ETT is an element of the failure rate and the transfer speed of the connection. The individual connection loads are fused into a performance called WCETT that openly represents the impedance among connections that utilise a similar channel. The Weighted Cumulative Expected Transmission Time (WCETT) performance is combined into a directing convention called Multi-Radio Link-Quality Source Routing (MRLQSR). Their performance is presented in a remote testbed comprising 23 hubs, each with two 802.11 remotes.

Gao et al. [20] have presented the MRC-WMN design network to limit the damage caused to various MRCs concurrently. Channel task and routing are fundamental problems in MRC designs since both impact traffic propagation over connections and channels. The interdependency between channel tasks and routing towards the proficient travels are performed. In the first place, the key design issues, signifying, and methods are distinguished. Secondly, the existing controls for combined channel tasks and routing are presented, and networks depend on the categories of the channel task.

Kyasanur et al. [21] have proposed the MRC frequency hopping on WMN. The network hubs are prepared with different recurrence active RI, each with various receiving wires and the Space-Time Block Coding (STBC) transmission approach. With the STBC approach on the radio, both transient and spatial changes can be utilized on every transmission, and the channel-constricted execution would then be improved. At that point, to lessen the co-channel traffic in the MRC routing task, the measure of co-channel impedances utilizing an error rate based cost work is assessed and allocated an appropriate active example for each radio. Utilizing the STBC actual method together with the MAC-layer channel, the administrators, the mesh network, and the non-vulnerability in contradiction of the channel reduce the co-channel resistance.

Ke et al. [22] have presented the MRC-WMN on minimum RI logic with better channel routing. The multi-path WMN hub is equipped with different RI and multiple channels nearby the node. Relegating reroute to communication links in the network limits network traffic. Since the number of radios on any hub could not be precisely the number of nearby channels, the channel task should accept the required number of channels to the connections on any hub, the maximum number of RI on that hub. Additionally, point-by-point Network Simulator Version 2 (NS-2) recreation considers demonstrating the presentation capability of our channel task measures in 802.11-based MRC-WMN networks.

Zeng et al. [23] have proposed the channel task of MRC-WMN using the position-based method. WMN aims to improve network execution and other networks with the remaining MRC hubs with MPR network design initiated to overcome an unreasonable request of single-radio networks like lack of successful scale to misuse the increasing available network transfer speed. Accordingly, a suitable channel task in MRC cross-section networks can decrease the number of signal flow co-channels and improve the network throughput. In this examination, the key objective is to limit the general problem and expand the network throughput by ensuring network availability.

Balusu et al. [24] have presented the Multi-Radio Channel (MRC) allocation technique in a WMN. Channel assignment was broadly studied in the system of cell networks, yet it was seldom concentrated in the remote, especially fixed networks, particularly in the multi-hop networks. A review of MRC-MPR serving issue in multi-hop remote setups is given in detail. This study the static non-cooperative game and Nash harmony channel allocation methods that are not reasonable for the multi-hop remote networks. In this manner, model the channel portion issue emerges as a hybrid game, including helpful and non-cooperative games. Inside a communication link, it is agreeable, and among links, it is non-cooperative. The min-max cooperation indicates Nash harmony that directs distribution combined in the game, which assumes to increase information of corresponding transmissions.

Remote advances with IEEE 802.11a that utilized non-covering channels. Coudron et al. [25] have presented TCP's routing and link-layer with MPR in an ad hoc wireless network. Ad hoc networks are presently accessible to utilize a single channel. The accessible network limit can be expanded by utilizing multiple channels. Utilization of multiple channels has various computing, albeit quantity of computes and more uncertain than the number of channels. The connection layer is a protocol to deal with multiple channels, and it tends to be represented over existing IEEE 802.11 devices. Direct measurement for MRC-MPR networks is merged into an on-request routing agreement that works over the connection layer protocol.

### **3 Optimized Reinforcement Learning Based Multi-Path Routing Protocol**

Traffic demand reduction on WMN is performed based on the artificial intelligence calculations, and it is determined with the calculated model of optimization approach in WMS-MRC routing. It can improve the drops, get familiar with the approach to accomplish the objective through steady endeavours and discover an ideal approach to challenge the issue. The MPTCP model uses the RL technique with PSO. The RL model is

Deep Q-learning Network (DQN), which combines Q-learning and neural networks [26]. The variance between DQN and Q-learning is that DQN utilizes neural networks to save the memory of the state and activity, so the key refreshing capacity of DQN is more in the light of Q-learning. The center of Q-learning is to discover the Q-value and utilize the Q-worth to enhance the routing approach. The proposed model's results show that RL's proposed information scheduling fundamentally accomplishes a higher throughput than the default computation. The parameter setting is assumed in Tab. 1.

**Table 1:** Parameters used in NS-2

Parameter	Value
Area size	1500 × 1000 m
Node placement	Random selection
Node density	Constant
Packet size	512 Kbits
Communication network	Uni-cast without retransmission
Routing protocol	AODV
Radio Range	3.65E-10
Antenna Model	Omni Directional
Transmission Range	0.2818
Traffic Source	UDP/CBR

The packet moves towards the best bit to fulfill the WMN. The utilized PSO ideally designates the balanced transfer to the connections to decrease the network traffic, which improves the network throughput of the information transmission [27]. PSO is used to attack traffic task problems in MRC-MPR by WMN. The hybrid activity is utilized in the discrete PSO design to deal with the discrete channel task problem. There are many RL related WMN transfer methods in the MRC routing based procedures. Most of these are circulated in nature, with every MR learning logic. Network transmission frequently puts the MGs and MRs at ideal areas to accomplish maximum execution.

Estimation of the expected link load is given by, Eq. (1)

$$C_1 = Q * C_Q \div L_1 \quad (1)$$

where,  $Q \rightarrow$  Number of available channels,

$C_Q \rightarrow$  Capacity per channel and

$L_1 \rightarrow$  Number of Links

MGs are critical in a network where most traffic is predetermined towards the Internet, as in a root network. More hubs are beneficial as they regularly bring about more limited routes for most MCs. Although many other development issues like routing and channel tasks accept a predefined position of these hubs, the presentation of their definitive result relies upon the actual primary design of hubs. Wireless Mesh Protocol (WMP), characterized in IEEE 802.11s, is an essential routing protocol for WMN. WMP protocol is semi-class since it supports two classes of optimal protocols. It depends on Ad-hoc On-demand Distance Vector (AODV) (RFC 3561) and tree-based routing. It depends on the administration protocol peer connect by which each cross-section idea finds and monitors the neighbour's hubs. On the off chance that any of these are associated with a wired communication, it

isn't essential to WMP, which selects the ways of those collected by including all neighbours' network links in an ethernet card.

Data packets are directed along with the packet drop of the target node when compared to the most minimal EC required. The packet drop can be resolved locally without much expansion by contrasting the likely help of neighbouring hubs. The path of the packet is to the neighbour with the best-routing cost. The fields are not restricted to a particular target hub. Various hubs may add to a similar possible field. The network routing involves a significant portion of a network with transfer speed, and the energy of the wireless hub is designed to get better results. Subsequently, with a minor overhead, the solid multipath directing protocol is fundamental for scheduling to limit the cooperation of wireless hubs in a path disclosure period that secures the trustworthiness of data transmission.

### 3.1 Reinforcement Learning Algorithm

The RL method reduces the EC of WMS in routing performances. This characterization of ideal malicious channels to path transmission is assessed. The destination is to decrease the number of inactive links, which reduce the impedance in the network, and consequently, the target work is characterized as a minimization work. Initially, the connectivity graph was designed for MRC, and it was selected as Eq. (2)

$$V_n = u_1, u_2, \dots u_n \quad (2)$$

In this, the node 'n' represents the radios used on the multi-channel 'u' that represents the random vector.

Assessment of expected load on link capacity can be calculated by the Eq. (3)

$$\gamma_1 = \sum_{a,b} P_l(a,b) \div P(a,b) * B(a,b) \quad (3)$$

In the above equation,

The number of the accepted path between a pair of nodes is denoted by  $P(a,b)$

The number of the accepted path between the pass link is represented by  $Pl(a,b)$

The estimated load between the node pair in the traffic profile is given as  $B(a,b)$

The created node is verified with the selected route using an optimization method. Additionally, the link is created to examine the traffic, and it is used for routing when needed. Communication between hubs ensures only just if it falls in the transmission possibility. If the availability is better, the impedance is more because of the signal overlapping. This can be reduced somewhat by utilizing the non-overlapping channels, yet the number of balanced channels can't be kept away. The channel task issue has been comprehensive as a non-linear programming model. Through the investigation utilizing reform, the control accomplishes outstanding throughput and minor packet attacks by ideally allotting the channels to the connections using the proposed method.

Without performing direct gradient computations, it is better to optimize full prediction through random gradient origin for the loss function in the algorithm. Adaptively, the weights are updated by a total time, and the metrics are measured from the behaviour distribution  $\rho$  and the emulator E. We may reach the familiar Q-learning algorithm by replacing expectations into one sample.

---

**Deep Q-Learning Algorithm**


---

- Step 1. Sequence  $S_1$ ,
- Step 2. Preprocessed sequenced  $\phi_1$ ,
- Step 3. Replay memory  $M$ ,
- Step 4. Capacity  $C$
- Step 5. Q function action value with random weights
- Step 6. Representation of sequence  $S_1 = \{u_1\}$  and preprocessed sequence  $\phi_1 = \phi(S_1)$
- Step 7. For  $i = 1:1:N$
- Step 8.  $S_1 = \{u_1\}$
- Step 9.  $\phi_1 = \phi(S_1)$
- Step 10. For  $T = 1:1:T$
- Step 11. Choose an action randomly, with a probability  $p$
- Step 12. Else
- Step 13. Choose  $a_t = \text{Max}_a Q^*(\phi(S_t), a; \theta)$
- Step 14. Perform an execution function: action  $a_t$  in emulator and observe the result in terms of metric reward 'rt' and image  $x_t + 1$
- Step 1. Set the updation as  $S_t + 1 = S_t, a_t, u_t + 1$
- Step 2. Preprocess updation  $\phi_t + 1 = \phi(S_t + 1)$
- Step 3. Save the changes in variables ( $\phi_t, a_t, r_t, \phi_t + 1$ ) in  $M$
- Step 4. Randomly sample batch of transitions ( $\phi_j, a_j, r_j, \phi_j + 1$ ) from  $M$
- Step 5. Set  $v_j = \{r_j\}$
- Step 6. For terminal  $\phi_j + 1$
- Step 7.  $r_j + \gamma$
- Step 8.  $\text{Max}_a Q(\phi_j + 1, a_0; \theta)$
- Step 9. For non-terminal  $\phi_j + 1$ ; random gradient origin performed over the step on  $(v_j - Q(\phi_j, a_j; \theta))^2$
- Step 10. End
- Step 11. End
- 

**3.2 Discrete Particle Swarm Optimization Algorithm**

The Discrete PSO is used on the effective MPR of WMS. This condition is utilized to reduce the cyclic process of the packets in the detection interval. There are distinctive speed models with the point of giving assortment in PSO. The behaviour of the packets will, in general, search locally instead of investigating the entire network. This behaviour is functional to our nearby routing table to transfer channels locally. Since this model is somewhat slower, to cause it to combine closer, we have acquainted the latency weight with speed. The capacity of MRC networks depends on the approach of single channels allocated to different RI to build the network with negligible effects. The EC of forwarding packets node is assumed by Eq. (4)

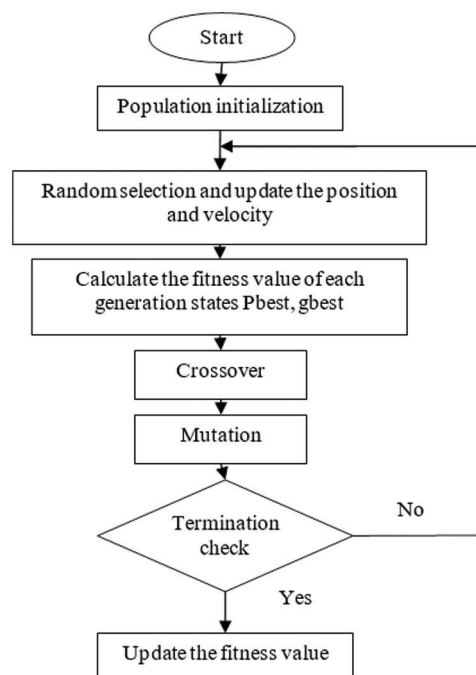
$$E_t(n, m) = E_{el}(N) + E_a(n, m) \quad (4)$$

where 'n' is the number of bits and 'm' is the distance between nodes.  $E_{el}$  and  $E_a$  are energy dissipated per bit to forward and receive packets, respectively. The transmission cost is specified by Eq. (5)

$$Tr = \frac{1}{a \times b} \quad (5)$$

where 'a' and 'b' are the probabilities value of forwarding packets to decide transmission cost, the MRC network is utilized to address fewer fuzzy attacks from certain misbehaves. The channel's task is to comply with the signal flow request that allots channels to the connections of the hub in the region corresponding to the signal flow of that hub.

Process flow of PSO in routing analysis is given in Fig. 1. PSO is utilized to take care of the multi-dimensional issue, and the balance time reserved is significantly less and the QoS to be tuned is trivial, and every bit gives a possible solution for the issue.



**Figure 1:** Flow of PSO in WMS routing

A network of random packets is formulated in PSO initially. The optimal value for this problem is searched by updating the generations and iterations. The optimum value must be considered to maximize accuracy and minimize time, error and cost. Hence, the two best values are assigned to each bit in echo. The best solution (fitness) is the first solution using this method. The name of the variable is set as  $p_{best}$ . The fitness value is stored with another "best" value: the PSO paths. This best value is named as global best and called  $g_{best}$ . Higher probability, typically 0.8–0.95, is the cross overvalue. A new solution is obtained by mutation, which turns on some digits in a string. Hence, on utilizing crossover, we may generate new solutions with a higher probability, which will generate some new solutions by transformation effect.

Arrived solutions are verified with their value so that the process may be terminated. Suppose the process has been stopped, then the fitness value is updated at once—the weight factor for optimal path and the remaining energy of a node shows better performance. A route with a better link quality is selected for forwarding data from source to destination. This paper utilises RL with PSO as a modern device for presenting an ideal channel task method for the multi-route WMN. Besides minimising a



network attack, the actual point of this work is to limit the network throughput. The correctness value is determined by Eq. (6)

$$Fit(k) = f(k) + p(k) \quad (6)$$

where  $f(k)$  is the global search updation and  $p(k)$  denotes the destination function of PSO. Here the destination function is determined based on the transmission cost function. The optimal MPR-MRC task introduced a RL method. Sequence update of crossover purpose is selected with crossover operator and separate route. The transformation is the process to flip the individual bits to the next group of PSO. The PSO is a cutting-edge procedure to discover the solution and search range. The PSO selection, transformation, and crossover process are the population-based meta-heuristic efficient routing. PSO can undoubtedly fall into a nearby, optimal procedure; concurrently, the Genetic Algorithm (GA) isn't reasonable for dynamic information because of the unique hidden network with RL. A base route to utilise PSO is to look for better connection quality hubs in the MPR phase. PSO efficiency is an issue by iteratively improving the proportion of value plan. The proposed approach is executed in two multiple links to determine the virtual channel and eliminate traffic signal requirements for the developed real channel.

#### 4 Results and Discussion

The proposed model of WMS performed with the routing approach to reduce the EC computations are performed with NS-2. About 100 nodes have been taken by implementing traffic demand reduction with optimized switching FL using RL and PSO. We have improved the performance, and network throughput of communication with the model of DQN blended with Q-learning and neural networks. We discovered the approach that increases transmission throughput in long-distance direction by utilising the RL model. An optimized routing process for MPR-WSN is proposed with RL and PSO. WMN with hubs will fill in a route that upgrades some restricted utility by the position of the energy-efficient hub.

The network is brought to an effective action with a particular routing method for every hub by providing the need of the improvement issue. The routing data would have been forwarded through the best-discovered routing path, which can control the network towards the optimized routing as the AQN with PSO learn the characteristic to improve the motivation to formulate the model. The various conditions of the model logically accept the excellent method to select the best path and improve the network throughput, as shown in Fig. 2.

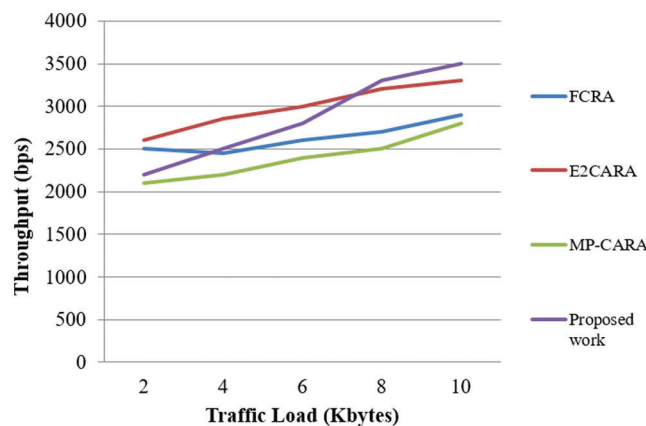
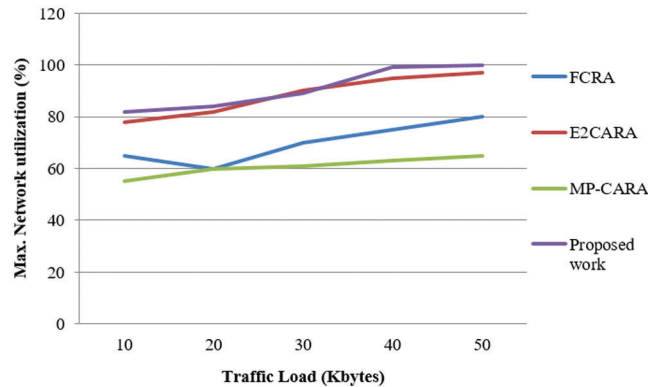


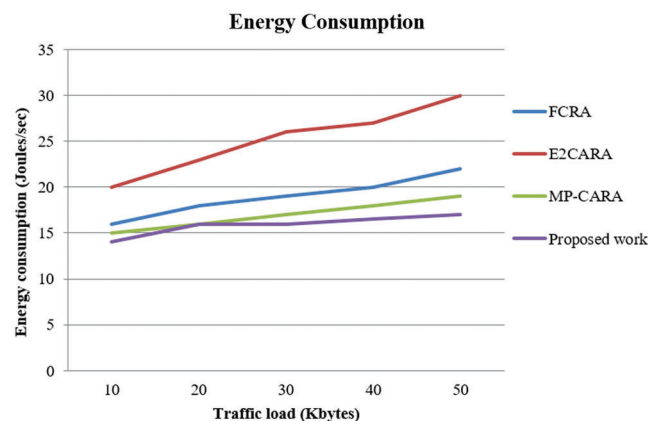
Figure 2: Throughput

Network utilization is the ratio of current bandwidth usage to the hub's maximum allocated bandwidth, which can handle traffic load, as shown in Fig. 3. The network can utilize the EC-related data collected from hubs during WMN action. These data incorporate the EC levels of each hub, the EC of the whole network, and the potential energy use of choosing a specific path naturally.



**Figure 3:** Maximum network utilization

Being a communication network, the EC of the network proposes that the hubs can make routing selections reliant on the EC at their perception. Comparison of proposed and existing works results in less EC, as shown in Fig. 4. The route is selected based on PSO-fitness value up-gradation—selection of best route results with more EC than the existing methods. The shortest path is searched as the best route, which minimizes the energy and delay and maximizes the network throughput. The proposed method improves all three metrics throughputs, EC, and maximum network utilization by comparing. The proposed work was achieved with 100% network utilization, 50% EC, and better network throughput under heavy traffic loads.



**Figure 4:** Energy consumption

## 5 Conclusion and Future Work

The proposed WMNs obtained better performances on MRC routing of the MPR transmission period, and the results proved it. Here the MRC-WMN is designed to reduce the traffic demand and improve the switching approach for EC. The RL with a discrete PSO approach is used. Here, the Discrete PSO is used on the routing path selection of MRC transmission, giving better network utilization results.

Overall, the RL with PSO model evaluation determines better performances than other existing works based on traffic load.

In future, the work may be extended with hybrid techniques on path selection and mapping and analysis with more resulting parameters.

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

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

## References

- [1] B. S. Roh, M. H. Han, J. H. Ham and K. I. Kim, "Q-LBR: Q-learning based load balancing routing for UAV-assisted VANET," *Sensors*, vol. 20, no. 19, p. 5685, 2020.
- [2] S. Rajeswari, S. A. Arunmozhi and Y. Venkataramani, "Q-Learning algorithm with network coding in multi-path transfer protocol for wireless mesh network," *International Journal of Recent Technology and Engineering*, vol. 9, no. 3, pp. 153–157, 2020.
- [3] Q. Yang, S. J. Jang and S. J. Yoo, "Q-learning-based fuzzy logic for multi-objective routing algorithm in flying ad hoc networks," *Wireless Personal Communications*, vol. 113, no. 1, pp. 115–138, 2020.
- [4] T. S. Malik and M. H. Hasan, "Reinforcement learning-based routing protocol to minimize channel switching and interference for cognitive radio networks," *Complexity*, vol. 2020, pp. 1–24, 2020.
- [5] S. Kosunalp, Y. Chu, P. D. Mitchell, D. Grace and T. Clarke, "Use of Q-learning approaches for practical medium access control in wireless sensor networks," *Engineering Applications of Artificial Intelligence*, vol. 55, no. 8, pp. 146–154, 2016.
- [6] E. F. G. M. Beig, P. Daneshjoo, S. Rezaei, A. A. Movassagh, R. Karimi *et al.*, "MPTCP throughput enhancement by Q-learning for mobile devices," in *IEEE 20th Int. Conf. on High-Performance Computing and Communications; IEEE 16th Int. Conf. on Smart City; IEEE 4th Int. Conf. on Data Science and Systems (HPCC/Smart City/DSS)*, Exeter, United Kingdom, pp. 1171–1176, 2018.
- [7] D. Yong, P. Kanthakumar and X. Li, "Hybrid multi-channel multi-radio wireless mesh networks," in *Int. Workshop on Quality of Service*, Charleston, USA, 1–5, 2009.
- [8] R. Draves, J. Padhye and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Proc. of the 10th Ann. Int. Conf. on Mobile Computing and Networking-MobiCom '04*, Philadelphia, PA, USA, pp. 114–128, 2004.
- [9] J. Luo, X. Su and B. Liu, "A reinforcement learning approach for multipath TCP data scheduling," in *IEEE 9th Ann. Computing and Communication Workshop and Conf. (CCWC)*, Las Vegas, NV, USA, pp. 276–280, 2019.
- [10] O. M. Zakaria, A. H. A. Hashim, W. H. Hassan, O. O. Khalifa, M. Azram *et al.*, "Joint channel assignment and routing in multi-radio multi-channel wireless mesh networks: Design considerations and approaches," *Journal of Computer Networks and Communications*, vol. 2016, no. 2769685, pp. 1–24, 2016.
- [11] K. Davis and L. Qilian, "Improving performance of multi-radio frequency hopping wireless mesh networks," in *Int. Conf. on Wireless Algorithms, Systems and Applications*, Chengdu, China, pp. 134–145, 2011.
- [12] A. P. Subramanian, H. Gupta, S. R. Das and J. Cao, "Minimum interference channel assignment in multi-radio wireless mesh networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 12, pp. 1459–1473, 2008.
- [13] M. Doraghinejad, H. Nezamabadi-pour and A. Mahani, "Channel assignment in multi-radio wireless mesh networks using an improved gravitational search algorithm," *Journal of Network and Computer Applications*, vol. 38, no. 4, pp. 163–171, 2014.
- [14] B. Arzani, A. Gurney, S. Cheng, R. Guerin and B. T. Loo, "Impact of path characteristics and scheduling policies on MPTCP performance," in *28th Int. Conf. on Advanced Information Networking and Applications Workshops*, Victoria, BC, Canada, pp. 743–748, 2014.
- [15] Q. Peng, A. Walid, J. Hwang and S. H. Low, "Multipath TCP: Analysis design and implementation," *IEEE/ACM Transactions on Networking*, vol. 24, no. 1, pp. 596–609, 2016.

- [16] F. Ye, S. Roy and Z. Niu, "Flow oriented channel assignment for multi-radio wireless mesh networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, no. 1, pp. 930414, 2010.
- [17] R. Vedantham, S. Kakumanu, S. Lakshmanan and R. Sivakumar, "Component-based channel assignment in single radio, multi-channel ad hoc networks," in *Proc. of the 12th Ann. Int. Conf. on Mobile Computing and Networking-MobiCom'06*, Los Angeles, CA, USA, pp. 378, 2006.
- [18] M. Alicherry, R. Bhatia and L. E. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 11, pp. 1960–1971, 2006.
- [19] P. Kyasanur, S. Jungmin, C. Chereddi and N. H. Vaidya, "Multichannel mesh networks: challenges and protocols," *IEEE Wireless Communications*, vol. 13, no. 2, pp. 30–36, 2006.
- [20] L. Gao, X. Wang and Y. Xu, "Multiradio channel allocation in multi-hop wireless networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 11, pp. 1454–1468, 2009.
- [21] P. Kyasanur and N. H. Vaidya, "Routing and link-layer protocols for multi-channel multi-interface ad hoc wireless networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 10, no. 1, pp. 31–43, 2006.
- [22] F. Ke, M. Huang, Z. Liu, Q. Liu and Y. Cao, "Multi-attribute aware multipath data scheduling strategy for efficient MPTCP-based data delivery," in *22nd Asia-Pacific Conf. on Communications*, Yogyakarta, Indonesia, pp. 248–253, 2016.
- [23] G. Zeng, B. Wang, Y. Ding, L. Xiao and M. Mutka, "Efficient multicast algorithms for multi-channel wireless mesh networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 1, pp. 86–99, 2010.
- [24] N. Balusu, S. Pabboju and G. Narsimha, "An intelligent channel assignment approach for minimum interference in wireless mesh networks using learning automata and genetic algorithms," *Wireless Personal Communications*, vol. 106, no. 3, pp. 1293–1307, 2019.
- [25] M. Coudron and S. Secci, "An implementation of multipath TCP in NS3," *Computer Networks*, vol. 116, no. 5, pp. 1–11, 2017.
- [26] P. Bahl, A. Adya, J. Padhye and A. Wolman, "Reconsidering wireless systems with multiple radios," *ACM SIGCOMM Computer Communication Review*, vol. 34, no. 5, pp. 39–46, 2004.
- [27] A. Kumar, A. Zhou, G. Tucker and S. Levine, "Conservative Q-learning for offline reinforcement learning," *Neural Information Processing Systems*, vol. 33, pp. 1179–1191, 2020.