Innovations in Computing and Emerging Technologies

ISSN: 3106-6518 (Print) 3106-6526 (Online) | Issue: 01 2025

Peer-Reviewed Academic Journal

Research Article

Route Optimization for IoT Networks with Reinforcement Learning

Atif Manzoor[1]
Email: raheelahmed860@gmail.com

Muhammad Ishtiaq[2]
Email: ishtiaq.gujjar541@gmail.com

Submitted	Accepted	Published	Article No.
4-Nov-25	28-Nov-25	19-Dec-25	ICET0125-319

Affiliations:
[1] *Federal Government Educational Institute (FGEI), Pakistan*
[2] *Department of Computer Science, University of Lahore, Pakistan*

Abstract

The Internet of Things(IoT) is changing different industries to improve the interactions between devices to facilitate automation, monitoring, and analysis. However, routing within IoT networks is not easily solved due to factors such as the dynamic nature of the networks, congestion, flooding, and resource waste. Efficient routing methods are necessary to guarantee connection availability, longer network duration, and rational resource usage. Reinforcement learning is one of the subfields of machine learning that can potentially address these challenges because it allows for learning in decision-making processes. In this regard, the flooding-controlled adaptive reinforcement learning-based route optimization model (FARLRO) is introduced to mitigate the problem of network flooding as well as make the most of the routing decisions within the networks. The parameters such as residual energy level, available bandwidth, mobility pattern, traffic condition and topological arrangements are incorporated into the state space of the model, and it uses reinforcement learning to adapt the routing decisions. The Q-learning model continuously improves the state variables and optimizes routes to reduce the cases of flooding and enhance the network?s efficiency. It also uses the Bellman equation for assessing future rewards, thus making it a forward-looking method of route optimization. Extensive experiments have shown that the model provides significant improvements in several critical performance metrics, such as a smaller flooding ratio, a lower network congestion index, less frequent broadcast storms, a lower packet drop ratio resulting from flooding, an increased network lifetime, a higher Mobility Aware Packet Delivery Ratio, and higher Resource Utilization Efficiency. As compared to the conventional routing protocols, the proposed model outperforms various state-of-the-art ad adhoc routing schemes. Extensive experiments have been performed to show that the proposed model decreases the flooding ratio, less overhead. Both above parameters are critical to ensure a longer network lifetime compared with the other approaches in high-density and high-mobility environments. Another advantage of the model is its effective stability in solving route optimization problems in IoT networks, which provides a great improvement over traditional routing algorithms.

Keywords: Internet of Things, Io, reinforcement learning, , Machine learning,route optimization

---

1 INTRODUCTION

The Internet of Things (IoT) refers to a network of interconnected devices, such as smartphones, appliances, and sensors, that can communicate with each other and exchange data over the Internet [1]. In traditional IoT networks, several aspects are either human-driven or based on a dynamic template; however, future intelligent networks require Artificial Intelligencem(AI) or machine learning(ML) techniques that can analyze the given environment and make automated decisions. This not only eliminates human dependence but makes the IOT network agile and reliable, from the realm of smart homes and cities to industrial automation and healthcare systems, the integration of IoT technologies has revolutionized various facets of human life [9]. However, with the exponential growth in the number of IoT devices, the optimization of data transmission routes has emerged as a critical challenge confronting IoT networks [10].

Efficient routing of data packets within an IoT network is paramount for ensuring optimal throughput and resource utilization. Conventional routing protocols [11] , while effective in traditional networks, they often fall short in addressing the dynamic and unpredictable nature of IoT environments. Issues such as network congestion, data collisions, and inefficient bandwidth utilization pose significant hurdles to achieving seamless communication within IoT ecosystems [12].

Moreover, flooding-based routing approaches [13], wherein packets are comprehensively broadcast to neighbouring nodes, aggravating network overhead and leading to wasteful utilization of scarce resources. Thus, there is a pressing need for novel routing strategies that can adapt to the unique challenges posed by IoT networks, thereby enhancing their efficiency and scalability.

A typical IoT network architecture [14] comprises three fundamental components: devices and sensors, connectivity infrastructure, and cloud or edge computing resources. IoT devices, equipped with sensors and actuators, serve as the backbone for collecting data from the physical world. Various connectivity options, including Wi-Fi, cellular networks, and Low-Power Wide-Area Networks (LPWAN) [15], facilitate the transmission of data to cloud-based servers or edge computing devices for processing.

IoT network architectural framework supports numerous applications, spanning smart homes, smart cities, industrial automation, healthcare, and environmental monitoring [16]. However, it also presents a host of challenges, including security vulnerabilities, scalability requirements, and efficient data management practices [17]. Addressing these challenges necessitates the development of innovative solutions that leverage emerging technologies, such as reinforcement learning, to unlock the full potential of IoT networks.

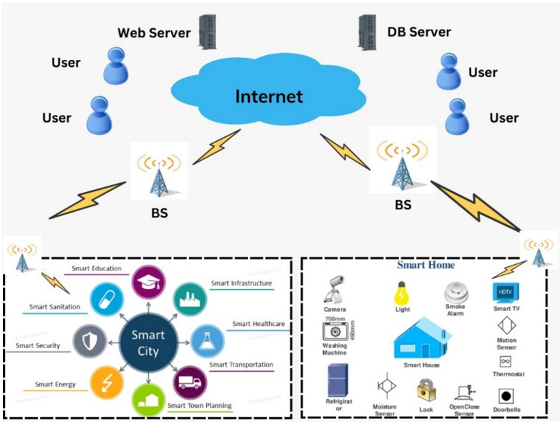

Figure 1: Centralized architecture along with its implications within a LTE network context

1.1 Flooding in network

In flooding [18], when a node wants to transmit a packet to a destination, it broadcasts the packet to all its neighbouring nodes. Each receiving node, in turn, rebroadcasts the packet to its neighbors, and this process continues until the packet reaches the destination or its Time to Live (TTL) value expires. Flooding ensures that the packet reaches its destination regardless of the network topology or the presence of faulty nodes. This makes it particularly suitable for scenarios where the network is highly dynamic and the topology changes frequently.

1.2 Issues and challenges while handling Flooding

This section discusses the various challenges and control mechanisms associated with network flooding, a common issue in networking where packets are broadcast to all neighbouring nodes. Key issues such as network overhead, packet collision, and redundant transmissions are explored, each contributing to network inefficiencies like increased congestion, wasted resources, and delays. To address these problems, the section also examines conventional flooding control mechanisms, including techniques like Time-To-Live (TTL) limits, duplicate packet detection, neighbour table management, and adaptive flooding algorithms. These methods aim to optimize network performance by reducing unnecessary traffic and improving the efficiency of packet delivery.

In highly dense networks or nodes per area, there is a high probability for collisions to occur due to the simultaneous transmission of packets from the nodes [19]. Such collisions cause packet drop and subsequent re transmission, which makes the problem worse, adding to the congestion of the network. As flooding does not depend on any routing information, many nodes may get a copy of the same packet from different neighbour nodes. Such redundancy is not only a waste in terms of available networks but also results in delay and inevitably higher latency and delay in the delivery of packets.

1.3 Conventional Flooding Control Mechanisms

To deal with the effects of flooding, several conventional control measures have been recommended as follows.

One of the most primitive techniques of controlling flooding is the utilization of a TTL value attached to each packet. The TTL values determine how many jumps or relay nodes the packet is allowed to make before it is discarded. Due to TTL, infinite looping and the unnecessary circulation of packets within the network are controlled to reduce network overhead.

Nodes can employ techniques such as sequence numbers or message identifiers that can be used at nodes to knock out those copy instances of the same packet which have already been received to avoid the burden of replay attacks. When applied, duplicate packet detection aids in saving the network cost and reducing congestion as it does not allow the retransmission of packets that have been delivered to the nodes.

There exists the possibility where a node receives multiple copies of the same packet in the network and to deal with this issue; nodes can make use of sequence numbers or message identifiers. By reducing the occurrence of repeated transmissions, duplicate packet detection assists in the preservation of traffic and the reduction of congestion.

Adaptive Predictive flooding algorithms act in a way and flood according to the expectations of the various requirements and conditions of the network. These algorithms may have some heuristics like probabilistic rebroadcasting or geographical forwarding for efficient delivery of packets, except flooding.

1.4 What is Reinforcement Learning

Reinforcement Learning (RL) is a powerful paradigm in the field of artificial intelligence and machine learning, wherein an agent learns to make sequential decisions through interaction with an environment to maximize cumulative rewards [20]. RL is mainly comprised of three components. An agent is an entity that is responsible for making decisions and taking actions within the environment. The external system with which the agent interacts comprises states, actions, and rewards.

In reinforcement learning, it is a plan or chart that an agent uses to determine its actions in view of the observed states with the aim of maximizing future benefits. Thus, the agent figures out the best policies during the different phases of the trial-and-error mechanism of the available actions in the environment. RL algorithms use different approaches such as value iteration, policy iteration and deep learning to estimate the control policies and solve difficult decision-making problems in different fields.

1.5 Role of Reinforcement Learning in flooding control in IoT networks

Reinforcement Learning offers a promising approach to addressing the challenges associated with conventional flooding control techniques in IoT networks. By leveraging the adaptive and learning capabilities of RL agents, IoT networks can dynamically adjust their flooding strategies based on network conditions and performance metrics, thereby optimizing resource utilization and improving communication efficiency [21]. Let's delve into how RL can overcome the limitations of conventional flooding control mechanisms.

1.5.1 Dynamic Adaptation

While static control mechanisms cannot work with a dynamic behavior in the RL agents that can dynamically change their flooding strategies based on the observed states and the feedback received from the environment, even during flooding. This dynamic nature allows IoT networks to change accordingly to the topology, traffic and channel, which in turn enhances the reliability and scalability of the IoT networks.

1.5.2 Optimized Resource Allocation

RL algorithms are capable of learning how best to allocate the scarce network resources through a trade-off between exploration and exploitation. By learning how to decide which nodes to send packets to, the RL agents will limit the number of forwarded packets, and consequently cut on the number of collisions, as well as limit energy consumption, leading to improved throughput and reliability.

1.5.3 Learning from Experience

RL agents learn from experience by interacting with the environment and receiving feedback in the form of rewards or penalties. Through iterative exploration and exploitation, RL algorithms can discover optimal flooding strategies that maximize long-term rewards, such as packet delivery ratio, end-to-end delay, and energy efficiency.

1.5.4 Adaptive Policy Optimization

RL agents acquire knowledge through experience gained by interaction with the environment and the feedback by means of rewards or punishment. As a result of exploration and exploitation in RL, RL algorithms can find the best flooding strategies that will enable it to get maximum rewards in the long run, these being PDR, End-to-End delay, and energy efficiency. Reinforcement learning model has a wide range of applications and services, such as augmented reality, virtual reality, the Internet of Things (IoT), and many more [11].

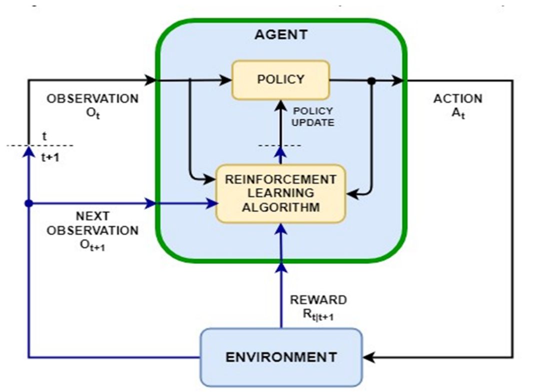

Figure 2: Working principle of reinforcement learning

1.6 The Problem

Enhancing the routes for the data transfer in IoT also remains a problem due to the basic routing protocol that faces the problem of scalability when confronted with dynamic and diverse IoT environments. Challenges like congestion, collision, and resource management become a problem in large-scale, reliable, and real-time networks for IoT, which slows down the extent and take up of these services. Additionally, this kind of route, the flooding-based approaches, complicates the consumption of available network overhead and resource utilization. Thus, there is a pressing need to develop intelligent routing algorithms leveraging Reinforcement Learning (RL) techniques to dynamically optimize routing strategies and mitigate the challenges faced by conventional flooding control mechanisms in IoT networks. This problem raises the several research questions mainly:

1. How can RL intelligently optimize the routing path procedures, such as selection, multipath routing, load balancing and dynamic routing, in resource-scarce IoT networks?
2. What are the specific RL techniques suitable for route optimization, such as deep reinforcement learning for making optimal routing decisions?
3. How is the proposed Flooding-Controlled Adaptive Reinforcement Learning-Based Route Opti-

mization Model compare to traditional routing protocols in terms of performance?

2 LITERATURE REVIEW

The first and foremost objective of the research papers is to gain a broad perspective on the recent work done in the field of IoT network optimization. This includes calling out specific techniques and methods used in the optimization of IoT networks, as well as evaluating the problems of congestion, energy utilization, scalability and QoS. Through the identification of concepts and the establish the relationships between various research papers, this examination is prospective to provide a theoretical foundation for the proposed Flooding Control ML-based route optimization model, and advance the practice and accuracy of IoT networks applied on smart city, industrial applications, healthcare, and environmental condition sensing Internet of Things IoT refers to a complex web of digital connections not only limited to computers, smart phones, home appliances, automobiles etc. But inclusively encompassing machines, gadgets, sensors, industrial products and medical instruments and in fact the list is endless and keeps growing day by day. Such devices are designed to interact and share information independently through the internet, trying to deliver more intelligent solutions and additional functions in several fields. Because IoT systems produce more data than other systems, the efficient solution to transfer these networks? data is critical to their performance, especially regarding the routes connecting them. Optimal management of routes contributes to optimal message delivery, time saving, energy saving and improvement of the performance of IoT networks. This section presents a review of different methodologies and approaches suggested for the purpose of route optimization in IoT networks; further, an analysis of these methodologies and approaches is done, and the strengths and weaknesses of all the approaches mentioned are made clear.

2.1 Flooding-Based Routing

Among all techniques that are used when routing data packets in the network, the first and the simplest is flooding [22]. Researchers have studied several techniques to optimize the flooding for adhoc networks. In this technique, if a node has a data packet that it must send to a particular destination, it broadcasts the packet to all neighbor nodes. It then passes it to the next receiving node, and this process goes on until the packet reaches the destination or till the TTL value is exhausted. It also guarantees that the packet gets to its intended destination, irrespective of the formation of the network at the time of transmission or even if certain nodes are down.

There are several benefits in the use of a flooding-based route. Coupled with the above advantages, the flooding-based routing method provides the following benefits. First, it is very easy to implement since it lacks extensive routing tables and does not require the protocols to hold state information about the network. Further, the method is highly resilient because no specific paths must be maintained when the topology or the rate of communication in the network changes. Most importantly, flooding is also certain to deliver the relevant packet if there is a single path existing from the source node to the destination node.

However, the method also has large drawbacks. The most significant drawback is the amount of traffic it creates in the networks, as many copies of the packet are sent to different neighbouring nodes, thus creating competition for the network resources. Additionally, in high-density nodes, the probability of many nodes transmitting data packets at the same instance causes data packet collisions and loss with the subsequent need for retransmission. Finally, since flooding does not employ any routing information, different neighbours may send similar data to one node, and as a result, there is wastage of resources and time as other nodes receive a similar packet several times.

2.1.1 Limitations Identified

- Slowness in traffic and unnecessary transmissions.
- Greater collisions and congestion in thick networks.
- Unutilized energy and bandwidth Blind broadcasting.

2.1.2 Expected Improvement (Gap)

Current techniques for floods are not smart enough in decision-making to inhibit the rebroadcast of redundant messages. In this work, we present a forwarding mechanism, which is motivated by RL technology, that would re-forward packets selectively in order to minimize redundancy and to ensure reliability.

2.2 Hierarchical Routing

Hierarchical routing [23] organizes the network into clusters, with each cluster having a special cluster head. The key management roles of the cluster head include overseeing communication within the cluster and managing communication between different clusters. This approach offers several benefits, including reduced routing overhead and enhanced scalability by dividing the network into more manageable parts.

The scalability of this routing method is particularly advantageous in large-scale networks. By collecting data at the cluster heads, the method reduces the number of transmissions required, thereby minimizing network congestion. Additionally, hierarchical routing is energy efficient, an essential feature in IoT environments where devices typically have limited power supply capacity. The reduction in the number of transmissions needed not only conserves energy but also contributes to better load balancing, ensuring that the network's resources are used efficiently.

However, there are also some disadvantages to this approach. One significant drawback is the potential for cluster head overload. When cluster heads receive too much traffic, they can become bottlenecks, reducing the overall efficiency of the network. Additionally, the process of forming and maintaining clusters is complex and time-consuming. Decisions, regarding which nodes should be selected as cluster heads, as well as the ongoing management of cluster membership, require careful planning and can be resource-intensive. Furthermore, the reliance on cluster heads introduces a single point of failure; if a cluster head fails, it can cause major communication disruptions within the cluster, posing a serious threat to the network's stability.

2.2.1 Limitations Identified

- Bottlenecks are caused by cluster-head overload.
- Cluster formation and maintenance a complicated processes.
- Single-point failure risk

2.2.2 Expected Improvement (Gap)

Current techniques for floods are not smart enough in decision-making to inhibit the rebroadcast of redundant messages. In this work, we present a forwarding mechanism, which is motivated by RL technology, that would re-forward packets selectively in order to minimize redundancy and to ensure reliability.

2.3 Geographic Routing

This kind of routing arranges the connections according to clusters, and there exists a special cluster head in each cluster [24]. Some of the responsibilities of the cluster head are responsible for communication within the cluster and also communication between two or more clusters. The above approach has several advantages, which include: it reduces the routing overhead since the network is divided into smaller parts to enhance scalability.

It is for this reason that this routing method is more scalable in large network certification. By collecting data at the cluster heads, the method helps to decrease the number of signals transmitted over the network, making the congestion minimal. Furthermore, hierarchical routing is energy-friendly, another important characteristic required in IoT isolates, where often device power supply capabilities are restricted. Such actions save energy for the transmission as well as contribute to load balancing of the necessary network load, which is paramount in determining the network resource usage. But there are also some disadvantages that are associated with this particular way of thinking.

However, there is one major disadvantage, including cluster head overload. When cluster heads are heavily busy, they attract a lot of traffic and thus act as a constraint to the performance of the network. However, the process of cluster formation and maintenance is not very easy and consumes a lot of time. The criteria for selecting nodes to be used as cluster heads and the membership management of the clusters especially involve a lot of planning and may take a lot of resources. Moreover, the use of the cluster head adds up a problem of single point failure; if the cluster is led by a single head and this head fails, then communication within the cluster is significantly affected, and this is a great threat to the network.

2.3.1 Limitations Identified

- Need the proper location of the information.
- Lives in error of localization.
- Continuing to face congestion and redundancy of packets.

2.3.2 Expected Improvement (Gap)

Such procedures fail to control routing processes according to real-time network states. FARLRO implements RL state-feedback to make and change forwarding actions dynamically without making use of entirely geographic information alone.

2.4 Hierarchical Routing

Hierarchical routing [23] organizes the network into clusters, with each cluster having a special cluster head. The key management roles of the cluster head include overseeing communication within the cluster and managing communication between different clusters. This approach offers several benefits, including reduced routing overhead and enhanced scalability by dividing the network into more manageable parts.

The scalability of this routing method is particularly advantageous in large-scale networks. By collecting data at the cluster heads, the method reduces the number of transmissions required, thereby minimizing network congestion. Additionally, hierarchical routing is energy efficient, an essential feature in IoT environments where devices typically have limited power supply capacity. The reduction in the number of transmissions needed not only conserves energy but also contributes to better load balancing, ensuring that the network's resources are used efficiently.

However, there are also some disadvantages to this approach. One significant drawback is the potential for cluster head overload. When cluster heads receive too much traffic, they can become bottlenecks, reducing the overall efficiency of the network. Additionally, the process of forming and maintaining clusters is complex and time-consuming. Decisions, regarding which nodes should be selected as cluster heads, as well as the ongoing management of cluster membership, require careful planning and can be resource-intensive. Furthermore, the reliance on cluster heads introduces a single point of failure; if a cluster head fails, it can cause major communication disruptions within the cluster, posing a serious threat to the network's stability.

2.5 Machine Learning-Based Routing

In routing, the use of ML adapts previous knowledge and data to predict good routes for vehicles or data packets [27]. One method that has a good result is supervised learning which can help predict the state of a network and adapt routing to that state.

The use of ML [21] for routing comes with the following benefits: One of them is the predictive nature of the system. The result of the analysis of historical data is used by the ML algorithms to provide a prediction of problems that may be expected, such as network congestion, and modify routing decisions to avoid these problems. It further optimizes the execution and dependability of the network to improve the management of the routing procedures. Further, it is quite flexible; the algorithms can also be adjusted to work using different performance measures, which in turn means that the Basic ML Algorithms can be retrained to work using different routing objectives or scenarios.

Still, there are certain disadvantages of utilizing ML in routing as well. That is the reason one of the biggest problems can be considered as the dependence on the data. Another challenge is that the training process of ML algorithms necessitates accurate quantities of data, most of which are difficult to get or could be voluminous, requiring even large storage and computational intensities. This dependence on data as a medium might prove disadvantageous, especially where access to this input is restricted. Also, the traffic of training and feeding the ML models, and particularly deep learning-based models, might be quite high. This is a big drawback, especially for IoT lower-tier devices that are rarely endowed with high computational power. The final concern is that of generalization; the current learning algorithms train the ML models using certain network conditions, and therefore, when applied to networks with different or even new conditions, it may cause the algorithm to make suboptimal routing decisions.

2.5.1 Limitations Identified

- Distributed on high dependency on labelled datasets.
- Poor flexibility to invisible network conditions.
- Training the inappropriate overhead of the resource-constrained IoT nodes.

2.5.2 Expected Improvement (Gap)

ML models do not possess the dynamic performance of learning and controlling flooding in real-time. RL is an adaptive version of ML that substitutes of static mechanisms with continuous environmental learning.

2.6 Genetic Algorithm-Based Routing

Genetic Algorithms (GAs) are based upon the natural selection process and in search of the best solution; a population of options evolves over several steps. Specifically, in the routing context, the GAs can be used in finding out the best path through which data can be transmitted.Another benefit one should mention is the fact that GAs are capable of global optimization. They are suitable for solving problems where the goal is to find global optima, especially in large and complex search space, such as in large and heterogeneous networks, where they are suitable for finding the best routing path. Also, the GAs is very flexible; they are able to adapt themselves to new conditions in the network and develop new routing strategies that are better fitted for the current situation. The third advantage that GA is the high level of robustness. Here, GAs can get stuck in local optima and this contributes to the productivity of the algorithm in finding the desired routing solutions.

The main disadvantage of this technique implies the necessity of significant computational loads. Although the process of creating and comparing the multiple runs of the possible solutions is beneficial, it is computationally costly, especially for those environments with limited computational power. Besides, the problem-solving process of GAs may take a longer time before arriving at the near-optimal solution. It is not uncommon to require multiple generations in finding an effective solution and this eats up time, which can be resource wasting especially in real time contexts which require speed in delivery. In addition, GAs are known to be dependent on various factors whose values have to be appropriately set in order to yield good results: the population size, mutation rate, and crossover rate. These parameters may take higher values and may also be hard to find, and they must be set well before to maximize the efficiency of the algorithm.

2.6.1 Limitations Identified

i Poor performance of the code. ii Slow convergence iii Sensitive to the tuning of parameters.

2.6.2 Expected Improvement (Gap)

In the case of IoT networks, GA-based routing is unable to respond to the fast topology changes. FARLRO can react in real time to changes in states by updating policy in response to changes.

2.7 Other Advanced Routing Techniques

One of the meta-heuristic algorithms is Ant Colony Optimization (ACO), which is inspired by the ant?s foraging behavior. Described below is applied in the discovery of the best solutions to pass through a network by replicating how ants drop pheromones on the best paths. This approach is based on the organization of ant?s behavior since they work in networks.

The fact that ACO is distributed is one of the main benefits of this approach. This characteristic indeed proves ACO as appropriate for the modern IoT networks, where the devices are not necessarily in the same network. Also, ACO is flexible; it can quickly adapt to changes in the network architecture, for instance, inclusion or exclusion of nodes. This adaptability enables the algorithm to perform well in dynamic conditions since the connectivity is done on the go. Moreover, ACO is considerably accurate, flexible and capable of managing large and dynamically growing networks, because of the distributed computing methodology.

However, the adoption of ACO also has some disadvantages as follows however, one of its shortcomings is the longevity to converge, that is, how long it takes before it stabilizes itself. ACO can be computationally expensive to arrive at the best solution whereby more time is consumed, especially where the network base is large. While this had made their integration a slow process, this can be a disadvantage when quick outcomes are needed. Additionally, there might be an issue of making ACO optimization complex since some elements in ACO are quite complex.

2.8 Swarm Intelligence-Based Routing

Swarm Intelligence (SI) is therefore defined as a method of solving a problem that incorporates services that imitate behaviours of natural systems, for instance bees, bird or fish. These algorithms solve complicated optimization issues by having many quite simple partners all work together in a coordinated manner obeying some local regulations. Another benefit of SI is that it is more resistant to such attacks as compared to other authentication methods. This is true because SI algorithms can easily adapt when one of the nodes has failed or when there is a change in the communication graph. Further, SI algorithms are inherently scalable in that the new solutions possessed are tabled along with the archive of previous solutions. They are also partially decentralized which enables them to administer large networks without necessarily having to implement a central command system. In addition, it can also be considered that SI algorithms have quite a high level of flexibility because they can be used to solve various optimization tasks within IoT networks.

2.8.1 Limitations Identified

- Huge computational burden.
- Delays in the large networks.
- Parameters Design and tuning are complex.

2.8.2 Expected Improvement (Gap)

The methods fail to tackle the issue of flooding redundancy directly or make real-time decisions. Reward-based optimization in FARLRO can reduce the time required to converge and minimize superfluous transmissions.

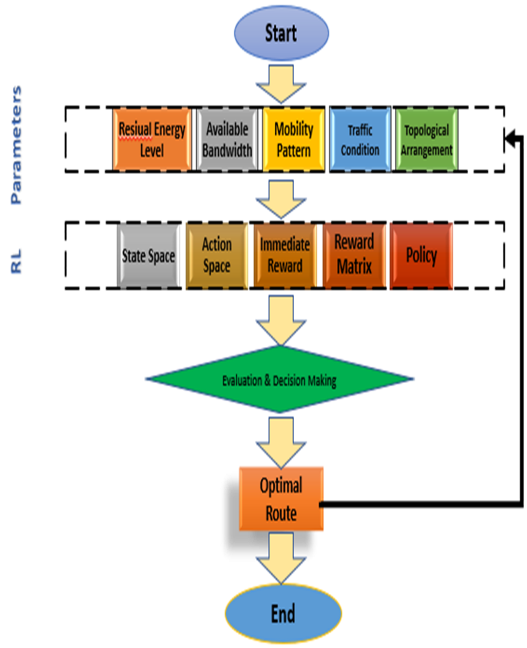

Figure 3: Proposed framework

2.8.3 Summary of Literature Gaps

In all the techniques that have been reviewed, the consistent gaps are:

- Deficiency of adaptive management of the intensity of flooding.
- Redundant packet transmissions are high.

- High dynamism or density leads to poor performance.
- Poor scalability and poor overhead.
- Lack of learning based forwarding policies.

These concerns in the proposed model are addressed by:

- A new system: RL-based adaptive rebroadcasting.
- Achieving resource efficiency.
- Minimizing the number of collisions and congestion.
- Improving the ratio of packet delivery and energy efficiency.

3 FLOODING-CONTROLLED ADAPTIVE REINFORCEMENT LEARNING-BASED ROUTE OPTIMIZATION

The main goal of FARLRO is to build an effective model that prevents the occurrence of floods while optimizing the route in network communication. Network congestion and, hence, flooding of the network is a major issue in dynamic networks. To this end, FARLRO incorporates the key parameters of the network and applies refined reinforcement learning methods. FARLRO considers some important network parameters, including Residual Energy Level (REL), Available Bandwidth (ABW), Mobility Pattern (MP), Traffic Condition (TC), and Topological Arrangements (TPA). These parameters provide a complete status of the network from which the model can make the most appropriate choices.

Therefore, FARLRO, which uses reinforcement learning to define the reward matrix, evaluates the short-term rewards of potential routing actions. This reward matrix is so well designed that the pros and cons of energy consumption, bandwidth, mobility, traffic and network architecture are balanced to the maximum. The model also applies the Bellman equation to update the Q-values which are the expected future discounted rewards of each state-action pair. Thus, FARLRO identifies the most suitable route from the source to the destination and thus facilitates the establishment of proper communication. As it has been mentioned previously, FARLRO uses the Bellman equation, which enables it to consider future rewards, thus making the routing more future-oriented. Therefore, FARLRO is a relatively new technique in the domain of route optimization and does not permit the network to be overloaded, thus enhancing the stability of the network.

3.1 Parameters and their importance in FARLRO

- Residual Energy Level (REL): REL is used to avoid frequent use of nodes that have low energy, to increase the network?s lifetime. Add REL to the state space to track the energy levels of nodes and add the REL to the reward function to encourage the selection of energy-efficient paths.
- Available Bandwidth (ABW): ABW is aimed here to avoid congestion and proper utilization of the bandwidth in the network so that there is no congestion. Add ABW to the state list to see the current bandwidth usage and incorporate it into the reward function to consider routes with enough bandwidth.
- Mobility Pattern (MP): The purpose is to adapt to the dynamic nature of mobile nodes, ensuring stable and reliable routes. Include MP in the state space to capture the movement patterns of nodes and in the reward, function to favor stable routes that are less likely to break.
- Traffic Condition (TC): The mission of TC is to manage network traffic load and avoid congestion by balancing the traffic across the network. Include TC in the state space to monitor current traffic conditions and in the reward function to incentivize load-balanced routing.
- Topological Arrangements (TPA): TA here is to understand the network topology for optimal route selection, avoiding unnecessary packet flooding. Include TPA in the state space to represent the network topology and in the reward function to select efficient and direct routes.

3.2 Implementation of Reinforcement learning

It is implemented the above-mentioned parameters to find the optimum route based on the mentioned parameters by using reinforcement learning (RL). RL generally consists of four phases, which are discussed below.

3.2.1 State space (S)

State space S is a representation of all the relevant information about the network at a given time. For each node iii in the network, the state can be represented as a vector that includes the Residual Energy Level (REL), Available Bandwidth (ABW), Mobility Pattern (MP), Traffic Condition (TC), and Topological Arrangements (TPA). For a network with N nodes, where ?N? is the network density, the state space for the node can be mathematically expressed in equation

1 below,

$$S_i = \{\mathrm{REL}_i, \mathrm{ABW}_i, \mathrm{MP}_i, \mathrm{TC}_i, \mathrm{TPA}_i\} \tag{1}$$

Combining the state of all nodes, the overall state space S for the entire network can be represented in equation 2 below,

$$S = \{s_1, s_2, s_3, \ldots, s_N\} \tag{2}$$

The state space for any node i in the network may, therefore, be described in terms of several parameters defining its running condition and immediate surroundings. First, the residual energy level has the meaning of the residual battery or energy resources of node i, this parameter reflects its capability to perform work and interact with other nodes. Second, the available bandwidth at node i means a part of the network capacity or data rate which can be utilized for communication and can affect the efficiency of the node. The mobility pattern is usually captured in terms of movement vector and defines a nodes activity in the network, affecting connectivity and coherence of the paths. The traffic condition at node i in the example describes the current traffic load, or the current state in terms of the amount of data that is flowing through at any one time and hence affecting latency and overall system performance. Last are neighbours or near nodes which signify the immediate nodes with which node i can share information and are used in routing algorithms, topology and other collaborations for forwarding of data. All together they create a comprehensive state space of node ?I? within the network, which enables adaptive decision making where constraints of the node matter for instance in routing and resources management.

3.2.2 Action Space (A)

Action space A defines the set of all possible actions that can be taken by the RL model to find the optimal route. For each node iii, the actions can include selecting the next hop for packet forwarding, adjusting transmission power, changing route discovery frequency, and applying congestion control mechanisms. Mathematically, the action space for node iii can be expressed in equation 3 below.

$$a_i = NH_i, PTx_i, RF_i, CC_i \tag{3}$$

In above Eq(3) NH, PTx, RF and CC repersents the next hop, transmit power, route frequency and congestion control for the ith node respectively. Looking at the network, which includes all nodes, and then the overall action space for optimizing the network design and to ensure delivery of packets efficiently can be defined as a set of actions that each node can take. Such action includes that is deciding the next jump node for packets, where each node analyses its neighbour nodes and selects the best node according to the parameters like available energy, available bandwidth and link quality etc. Further, nodes have the capability of setting their signal transmission power to manage the area and quality of their nodes? connections, which is essential in managing the direction of energy consumption while at the same time ensuring good signal quality. Another key action involves the frequency of route discovery where, through the broadcasting process, nodes decide how often to perform route discovery activities given the dynamism in the network, mobility and route stability. High frequency update could be useful particularly in conditions where the surrounding environment changes often, although, in this case, there could be some overhead introduced to the system, whereas low update frequency could lead to the case when the route is already out-dated in conditions that are rapidly changing. Lastly, the congestion control mechanisms are an important action that nodes must perform to control traffic; to prevent some areas within the network from leading to congestion of data traffic. This entails controlling the speed at which the packets are sent or directing the traffic to other channels that have few congestion. These are the constituent actions that comprise the total recognizable action space by which the network can respond to fluctuating conditions in real-time on its accord. This can be mathematically represented in the following equation 4.

$$A = \{a_1, a_2, a_3, \ldots, a_N\} \tag{4}$$

To explain the state and action space let's see an example, for node 1, based on the given parameters are REL= 80 Joules, ABW= 10 Mbps, MP= 1.2 m/s, TC= 15 packets/s, TPA= 2, 3, 4 The state for node 1 can be expressed as given in equation 5,

$$S_1 = 80, 10, 1.2, 15, (2, 3, 4) \tag{5}$$

Some things that node 1 might do is select one of the neighbouring nodes, like 2, 3 or 4, as the Next Hop. It could also involve change in the Transmission Power which ranges from High, Medium or Low. Furthermore, node 1 can also select the frequency of Routing Discovery which may get high frequency or low frequency. At the end, it could apply Congestion Control actions such as rate limiting or traffic shaping to curb congestion.

3.2.3 Reward Matrix

Combining these states and actions, the RL model can learn to optimize routing by considering the entire state and action space. The Q-learning algorithm can be used to update the Q-values for state-action pairs, guiding the model towards optimal routing decisions that avoid flooding and enhance overall network performance. Design a reward function that balances the objectives of avoiding flooding and optimizing routing, The Reward matrix calculation follows the relationship as described in Eq. (6) and (7).

$$R(s,a) = \alpha \cdot \mathrm{REL} + \beta \cdot \mathrm{ABW} - \gamma \cdot \mathrm{MP} - \delta \cdot \mathrm{TC} - \varepsilon \cdot \mathrm{HC} \quad (6)$$

And for a possible node reward can be obtained,

$$\begin{aligned} Q((3,1)) &= 87.0 + \max(Q((1,\cdot))) \\ &= 87.0 + 87.0 = 174.0 \\ Q((3,2)) &= 84.2 + \max(Q((2,\cdot))) \\ &= 84.2 + 170.2 = 254.4 \\ Q((3,6)) &= 78.3 + \max(Q((6,\cdot))) \\ &= 78.3 \\ Q((3,7)) &= 65.5 + \max(Q((7,\cdot))) \\ &= 65.5 \end{aligned} \quad (1)$$

In this model, weights identify the importance and frequency of various attributes incorporated into this model. To decide between these options for simplicity and to better demonstrate this process, as given these weights values of 1.

Here S stands for the existing state, while ?a means an action that has been made, such as choosing the next hop. The reward is influenced by several factors: The reward is influenced by several factors: In communication networks, there are several ways that are used to assess and remunerate for the services that providers give to the users. Residual Energy Level (REL) quantifies the power left in nodes of the network, and the node with having higher REL is perceived as a more dependable and renewable node. Higher REL values are incentivised, thus increasing the network stability, and high node energy levels increase the network?s longevity. Available bandwidth (ABW) is the measurement of bandwidth facilities available for data transmission. According to the work done, a higher value of ABW means more availability of bandwidth, which is beneficial for the networks. Hence, nodes that have higher ABW should be rewarded more to encourage optimum utilization of the effective bandwidth. In contrast, Mobility Pattern (MP) is a measure of the node mobility standard and depicts the motions of the nodes within the network. Mobility affects the reward of nodes based on changes in topological structures; hence, network stability would reduce for node mobility to reflect instabilities. Traffic Congestion or TC is defined as the measure of the amount of traffic on a network and since traffic congestion leads to delays in the flow of traffic then its effect is to slow the throughput of traffic within a network. Accordingly, nodes, which encountered more traffic congestion should be rewarded with a lesser reward. Finally, we have Hop Count, which indicates the number of intermediate nodes to which the packet passes. Table 1 shows the nodes and deployment parameter values.

Bellman Equation for Route Optimization

Table 1: Node and Deployment Parameter Values

Node	REL (Joules)	ABW (Mbps)	vMP (Speed; m/s)	TC (Packets/s)	TPA (Neighbor Nodes)
1	80	10	v1.2	15	2, 3, 4
2	75	8	v0.8	20	1, 3, 5
3	60	12	v1	18	1, 2, 6, 7
4	70	7	v1.5	25	1, 8
5	85	9	v0.5	10	2, 6
6	65	11	v1.3	22	3, 5, 7, 9
7	55	6	v0.9	12	3, 6, 10
8	90	14	v1.1	8	4, 9
9	50	13	v0.7	30	6, 8, 10
10	60	5	v1.4	28	7, 9

Hops, if too many, bring about more time delay and hence routes with high hop values should be discouraged by lower rewards. Scripts 2 and 3 can then be used to assign the weight based on the above parameters and create a value function that enables the decision maker to arrive at the best action based on the state of the network importance of is the fact that the parameters and weights will enable the computation of network performance and reliability, hence improving the same Bellman equations are used to find the optimum route as per given conditionsasuse the immediate reward matrix to iteratively calculate the maximum expected reward for each node. Bellman equation for this context now becomes, Bellman Equation implementation

$$Q(s,a) = R(s,a) + \gamma \max_{a'} Q(s',a') \quad (8)$$

The Q-value of taking an action ?a? in each state ?s? denotes the potential sum of rewards that can be earned if that particular action is taken and subsequently the best policy is to be followed. The Immediate reward, which is represented by the function refers to the reward gained once an action ?a? has been taken in state ?s?. (that takes a value between 0 and 1) defines the discount factor that is the amount of credit given to the subsequent rewards, where a value close to 1 means that the future rewards are as valuable as the immediate rewards. For simplicity and clarity, the value given to is taken as 1 that is the future rewards are not at all discounted. The term is the maximum expectation of future reward from the next state when playing the best action in that state. However, during calculations, this term is left blank because the actual future rewards depend upon many factors (such as residual energy, available bandwidth, mobility factor, traffic etc.) all of which are included in the parameter matrix. Thus, by using the value known as Q, an action can be chosen that will help receive the best sum of the reward while operating in an unknown environment and/or choosing in networks, a route.

3.2.4 Policy:

The policy in FARLRO dictates which node to forward packets to, based on the current network conditions represented by the parameters Residual Energy Level (REL), Available Bandwidth (ABW), Mobility Pattern (MP), Traffic Condition (TC), and Topological Arrangements (TPA). In FARLRO, the policy is the cornerstone that guides the route optimization process. It determines the best routing actions based on current network conditions and learned experiences, ensuring efficient, reliable, and flood-free network communication. By continuously adapting to the network environment and leveraging the Bellman equation, the policy helps FARLRO achieve its goal of optimal route selection while avoiding the pitfalls of network flooding.

3.3 FARLRO Real-World Testing

Here is the parameters table which is gained by implementing the model in MATLAB. To assess the nodes? performance in the communication networks and efficiency of the networks, several parameters are used. Residual Energy Level (REL) has been used to denote the residual energy of each node in Joules, and it ranges from 50 Joules to 90 Joules. Nodes with higher REL values are those which have more energy available and are more reliable and sustainable. The nodes that have higher residual energy levels are given preference so that energy consumption for communication is minimized, and thereby the life span of the networks is maximized. Available Bandwidth (ABW) which is in units of Mbps shows the actual availability of bandwidth at different node ranges from 5 to 14 Mbps. ABW is typically higher in nodes that provide better data transmission and better network performance; therefore, it should be rewarded so that bandwidth usage in the network is optimised. Mobility Pattern (MP) is quantified by the speed of each node in meters per second (m/s) and its values lie between 0.5 to 1.5 m/s. Hence, while higher speeds translate to higher mobility, these are often associated with instabilities in the network; as such nodes that have higher mobility are likely to be awarded lower rewards. Traffic Congestion (TC) is defined as the packets per second that are passed through a certain node, and this is in the scale of 8 to 30 packets per second. More traffic congestion implies more delays on the networks and hence decreased efficiency; nodes that have greater congestion should be awarded less. Last, Topological Arrangement (TPA) is the list of the neighbouring nodes each node is directly connected to, which portrays the structure of the network. The position of these adjoining nodes determines the level of communication as well as routing performance. When these metrics are incorporated in a value function, network management can make good decisions relating to energy usage, bandwidth provisioning, mobility, traffic patterns, and topology for networks to improve their performance.

3.4 Reward Matrix on given data

Combining these states and actions, the RL model can learn to optimize routing by considering the entire state and action space. The Q-learning algorithm can be used to update the Q-values for state-action pairs, guiding the model towards optimal routing decisions that avoid flooding and enhance overall network performance.

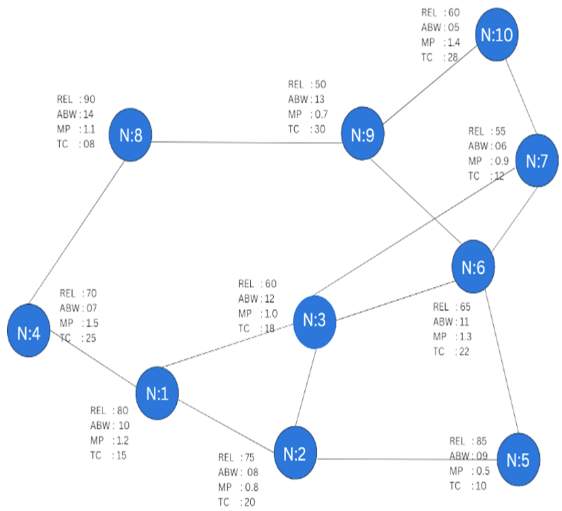

Figure 4: Node deployment and parameter values

3.5 Bellman equation implementation on given data

The detailed calculations for bellman equation are as below; **From node 1:**

$$\begin{aligned} Q((1,2)) &= 83.2 + \max(Q((2,\cdot))) = 83.2 \\ Q((1,3)) &= 87.0 + \max(Q((3,\cdot))) = 87.0 \\ Q((1,4)) &= 55.5 + \max(Q((4,\cdot))) = 55.5 \end{aligned} \tag{9}$$

From node 2:

$$\begin{aligned} Q((2,1)) &= 83.2 + \max(Q((1,\cdot))) \\ &= 83.2 + 87.0 = 170.2 \\ Q((2,3)) &= 84.2 + \max(Q((3,\cdot))) \\ &= 84.2 \\ Q((2,5)) &= 86.5 + \max(Q((5,\cdot))) \\ &= 86.5 \end{aligned} \tag{10}$$

From node 3:

$$\begin{aligned} Q((3,7)) &= 59.5 + \max(Q((7,\cdot))) \\ &= 59.5 + 409.6 = 469.1 \\ Q((3,6)) &= 51.5 + \max(Q((9,\cdot))) \\ &= 51.5 + 475.1 = 526.6 \end{aligned} \tag{11}$$

Table 2: Values Reward of Each Node to its Neighboring Nods

Nodes	N1	N2	N3	N4	N5	N6	N7	N8	N9
Node 1	0	83.2	87.0	55.5	–	–	–	–	–
Node 2	83.2	0	84.2	–	86.5	–	–	–	–
Node 3	87.0	84.2	0	–	–	78.3	65.5	–	–
Node 4	55.5	–	–	0	–	–	–	94.9	–
Node 5	–	86.5	–	–	0	84.2	–	–	–
Node 6	–	–	78.3	–	84.2	0	68.7	–	61.3
Node 7	–	–	65.5	–	–	68.7	0	–	59.5
Node 8	–	–	–	94.9	–	–	–	0	65.5
Node 9	–	–	–	–	–	61.3	65.5	65.5	0
Node 10	–	–	–	–	–	–	59.5	–	51.5

From node 4:

$$\begin{aligned} Q((4,1)) &= 55.5 + \max(Q((1,\cdot))) \\ &= 55.5 + 87.0 = 142.5 \\ Q((4,8)) &= 94.9 + \max(Q((8,\cdot))) \\ &= 94.9 + 0 = 94.9 \end{aligned} \tag{12}$$

From node 5:

$$\begin{aligned} Q((5,2)) &= 86.5 + \max(Q((2,\cdot))) \\ &= 86.5 + 170.2 = 256.7 \\ Q((5,6)) &= 84.2 + \max(Q((6,\cdot))) \\ &= 84.2 + 0 = 84.2 \end{aligned} \tag{13}$$

From node 6:

$$\begin{aligned} Q((6,3)) &= 78.3 + \max(Q((3,\cdot))) \\ &= 78.3 + 254.4 = 332.7 \\ Q((6,5)) &= 84.2 + \max(Q((5,\cdot))) \\ &= 84.2 + 256.7 = 340.9 \\ Q((6,7)) &= 68.7 + \max(Q((7,\cdot))) \\ &= 68.7 + 0 = 68.7 \\ Q((6,9)) &= 61.3 + \max(Q((9,\cdot))) \\ &= 61.3 + 0 = 61.3 \end{aligned} \tag{14}$$

From node 7:

$$\begin{aligned} Q((7,3)) &= 65.5 + \max(Q((3,\cdot))) \\ &= 65.5 + 254.4 = 319.9 \\ Q((7,6)) &= 68.7 + \max(Q((6,\cdot))) \\ &= 68.7 + 340.9 = 409.6 \\ Q((7,10)) &= 59.5 + \max(Q((10,\cdot))) \\ &= 59.5 + 0 = 59.5 \end{aligned} \tag{15}$$

From node 8:

$$\begin{aligned} Q((8,4)) &= 94.9 + \max(Q((4,\cdot))) \\ &= 94.9 + 142.5 = 237.4 \\ Q((8,9)) &= 65.5 + \max(Q((9,\cdot))) \\ &= 65.5 + 0 = 65.5 \end{aligned} \tag{16}$$

From node 9:

$$\begin{aligned} Q((9,6)) &= 61.3 + \max(Q((6,\cdot))) \\ &= 61.3 + 340.9 = 402.2 \\ Q((9,7)) &= 65.5 + \max(Q((7,\cdot))) \\ &= 65.5 + 409.6 = 475.1 \\ Q((9,8)) &= 65.5 + \max(Q((8,\cdot))) \\ &= 65.5 + 237.4 = 302.9 \\ Q((9,10)) &= 51.5 + \max(Q((10,\cdot))) \\ &= 51.5 + 0 = 51.5 \end{aligned} \tag{17}$$

From node 10:

$$\begin{aligned} Q((10,7)) &= 59.5 + \max(Q((7,\cdot))) \\ &= 59.5 + 409.6 = 469.1 \\ Q((10,9)) &= 51.5 + \max(Q((9,\cdot))) \\ &= 51.5 + 475.1 = 526.6 \end{aligned} \tag{18}$$

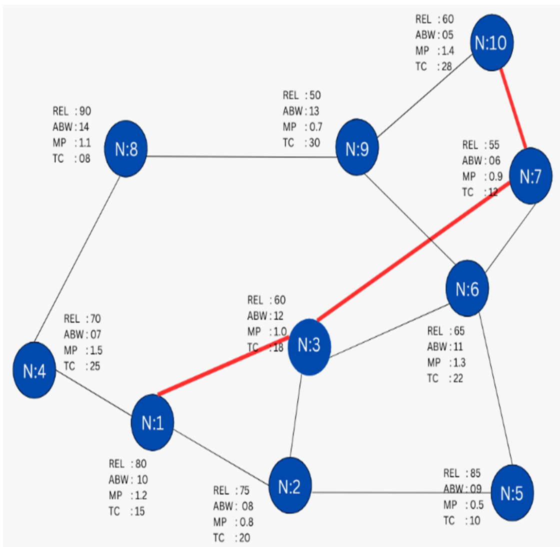

Figure 5: The optimum path from source to destination

Table 3: Values Reward Matrix of Each Nodes to its Neighboring Nodes

Source	Destination	IR	Q_{max}	Q_v
1	2	83.2	170.2	253.4
1	3	87.0	254.4	341.4
1	4	55.5	142.5	198.0
2	1	83.2	87.0	170.2
2	3	84.2	254.4	338.6
2	5	86.5	256.7	343.2
3	1	87.0	87.0	174.0
3	2	84.2	170.2	254.4
3	6	78.3	340.9	419.2
3	7	65.5	409.6	475.1
4	1	55.5	87.0	142.5
4	8	94.9	237.4	332.3
5	2	86.5	170.2	256.7
5	6	84.2	340.9	425.1
6	3	78.3	254.4	332.7
6	5	84.2	256.7	340.9
6	7	68.7	409.6	478.3
6	9	61.3	475.1	536.4
7	3	65.5	254.4	319.9
7	6	68.7	340.9	409.6
7	10	59.5	526.6	586.1
8	4	94.9	142.5	237.4
8	9	65.5	475.1	540.6
9	6	61.3	340.9	402.2
9	7	65.5	409.6	475.1
9	8	65.5	237.4	302.9
9	10	51.5	526.6	578.1
10	7	59.5	409.6	469.1
10	9	51.5	475.1	526.6

Legends:
IR - Immediate Reward
Q_{max} - Max Future Reward
Q_v - Q-value $(R + Q_{max})$

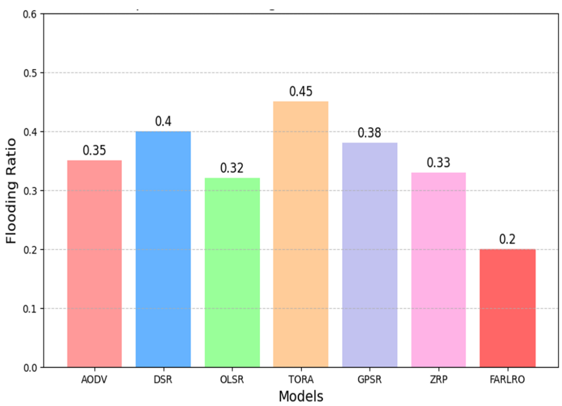

Figure 6: Comparison of flooding rate or all models

3.5.1 Optimum path calculation

For the optimum path from Node 1 to Node 10 based on the Q-values, we begin by looking at the highest Q-values at each step at the end of the sequence. First, from Node I, the maximum Q – value directs to Node 3. After that, starting from Node 3, the link with the higher Q-value leads to Node 7. Last but not least, from Node 7; we come across the highest Q value, which leads us to Node 10. Thus, the optimum path from Node 1 to Node 10, according to the Q-values, follows the sequence: The path is: Node 1 ? Node 3 ? Node 7 ? Node 10. This path is selected because it has the largest total Q-value; this implies that a given path is the best or the most optimal path based on the rewards or penalties given.

4 Results and analysis

To apply the FARLRO model, the following fundamental network parameters are to be defined in MATLAB: Residual Energy Level (REL), Available Bandwidth (ABW), Mobility Pattern (MP), Traffic Condition (TC), and Topological Arrangements (TPA). These parameters constitute the state space, which is the current state of the network. Subsequently, a set of actions can be defined that include next hop, transmission power, and frequency of route discovery and congestion control. Apply the Q-learning algorithm, which requires the creation of a Q-table where Q-values are stored, these being the expected future rewards of state-action pairs. The reward function should include objectives such as reducing energy consumption and congestion. In the training process in a simulated network environment, update the Q-values using the Bellman equation. After the training of the model, export the Q-values and the corresponding parameters of the network in a dataset. In Python, load this dataset with the help of libraries such as Pandas and then apply path optimization logic to determine the best path with the help of Q-values. Python can also be used for further analysis and visualization of network performance, thus, we can be sure about the effectiveness and stability of the model. Such an approach offers a clear plan of how to apply and enhance the FARLRO model systematically, using MATLAB for model development and Python for optimization and analysis.

In this part of this research, FARLRO model used the following performance evaluation metrics that have been chosen carefully. Flooding Ratio, Network Congestion Index, Broadcast Storm Frequency, Packet Drop Due to Flooding, and Impact on Network Lifetime, Mobility-Aware Packet Delivery Ratio (MA-PDR) and Resource Utilization Efficiency (RUE) are the metrics that are used to measure the performance of the FARLRO model in mitigating flooding and managing routing in mobile ad hoc networks. All of them offer different insights into the performance of the model for controlling the network resources, avoiding network congestion, and improving the overall network utilization. To compare the effectiveness of the FARLRO model, it is compared with other seven routing protocols such as Ad hoc On-Demand Distance Vector (AODV) Routing [2], Dynamic Source Routing (DSR)[4], Optimized Link State Routing (OLSR)[8], Temporally Ordered Routing Algorithm (TORA)[7], Greedy Perimeter Stateless Routing (GPSR)[5], Zone Routing Protocol (ZRP)[4] and Epidemic Flooding. These protocols were selected because of their applicability in similar types of network settings and because they employ different techniques for dealing with network congestion and flooding. In this way, we intend to show that FARLRO provides a better performance concerning the protocols described above, thus resulting in a more efficient and stable network in terms of flooding, packet loss, and resource management. Analyzing the results of this study will offer essential information regarding the benefits of employing FARLRO in networks that are vulnerable to floods and mobility.

4.1 Flooding Ratio

The Flooding Ratio is defined as the ratio of the number of unnecessary or redundant broadcast packets to the total number of packets that have been transmitted in the network. The bar graph given below shows the Flooding Ratio of different network routing models such as AODV, DSR, OLSR, TORA, GPSR, ZRP and FARLRO. As has been illustrated in the graph above, the proposed FARLRO model has the lowest Flooding Ratio compared to the other models. 20. This is in contrast with the other models, where the Flooding Ratios vary between 0 32 0. 45, which shows higher redundant packet transmission and probable congestion in the network. The reason why FARLRO has better performance than the other approaches is that FARLRO has an adaptive learning mechanism that regulates flooding through the intelligent control of routing decisions with the real-time conditions of the network. This helps to reduce the number of broadcasts that take place and, hence, more effective utilization of the available networks. In contrast, the existing models, such as AODV and TORA, are based on relatively more static or reactive approaches that may cause a lot of flooding, especially in large or dynamic networks. The graph effectively illustrates how FARLRO?s techniques to solve the problem of route optimization are not only a solution to the problem of flooding but also have a greater advantage in improving the efficiency of the entire network than other solutions to solve the problem of traffic congestion in the network.

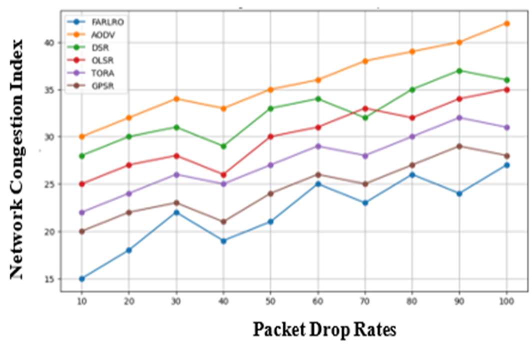

Figure 7: Congestion index with changing PDR

4.2 Network Congestion Index

The four graphs (Fig. 8-11 given below show the Network Congestion Index considering the network capacity, the packet drop rates, the buffer overflow levels and the queue lengths in the different routing models, namely FARLRO, AODV, DSR, OLSR, TORA and GPSR. In all the cases, FARLRO has higher values, and this shows that it has better performance than the other models in terms of congestion management. In the first graph, the congestion index of FARLRO is relatively low when the network capacity increases from 10 to 100 nodes, where FARLRO is 10 at the beginning and 26 at the end, while AODV is 50, which shows that FARLRO can effectively use the network resources.

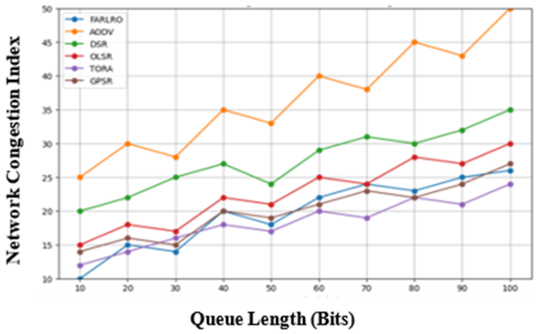

Figure 8: Congestion index with changing Queue lengths

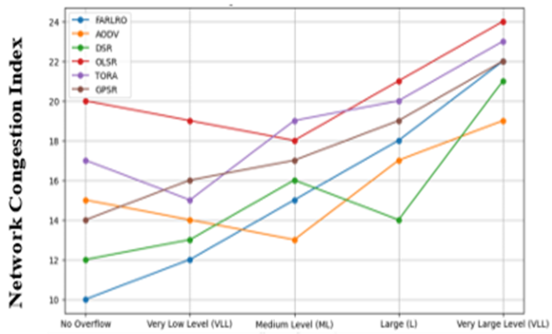

Figure 9: Congestion index with changing BOL

The second graph compares congestion with packet drop rates, and here too FARLRO is depicted to be quite robust with an index of 25 to 35 when the drop rates are high, compared to TORA, which has an index of 45 at the same drop rates.

The third graph, which is on the buffer overflow levels, indicates that FARLRO manages to keep the congestion index considerably lower at all the levels of no overflow, small overflow, large overflow and very large overflow, starting from level 10 and only rising to level 22, while DSR and OLSR risesteeply to level 24. FARLRO is 10 at the beginning and 26 at the end, while AODV is 50, which shows that FARLRO can effectively use the network resources.

Last but not least, in the graph of queue lengths, FARLRO is almost static with a congestion index varying in the range of 10 to 26 while other models such as GPSR are highly volatile with values peaking up to 50. These results show that FARLRO has better congestion control mechanisms than the other models tested in terms of load management, packet loss rate and resource utilization.

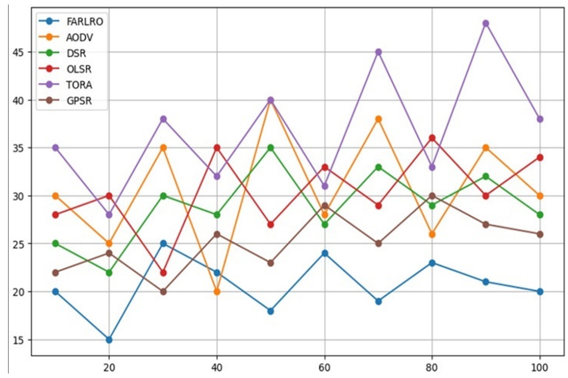

Figure 10: Congestion index with changing network density

4.3 Broadcast Storm Frequency

This measures the ability of your model to prevent broadcast storms which are severe network congestion events due to flooding. The graph above shows the Broadcast Storm Frequency of the proposed FARLRO model, AODV, DSR, OLSR, TORA, GPSR and ZRP, when the number of nodes in the network is increased from 10 to 100.

The FARLRO model is always shown to perform better than the other models, in terms of having a lower Broadcast Storm Frequency, especially as the size of the network increases. For example, with 50 nodes FARLRO has a frequency of 10 while AODV, DSR and OLSR have a higher frequency of 24, 29 and 35 respectively. This trend is maintained as the node count goes up; FARLRO has a minimum frequency of 6 at 100 nodes while ZRP and GPSR have 46 and 41 respectively. The benefit of the FARLRO model is that it uses reinforcement learning which can adapt the routing decision depending on the current network conditions thus minimizing the number of broadcast messages and avoiding congestion. However, models such as AODV and DSR that depend on route discovery mechanisms are likely to produce more broadcast storms as network density increases hence frequent. Due to their proactive and hybrid natures, OLSR and ZRP have higher storm frequencies in a large network, although the overhead is a problem. FARLRO's capability of reducing such incidents not only improves efficiency in the network but also guarantees more stable connections as the network expands, making it a great tool for managing floods and routing in a dynamic network.

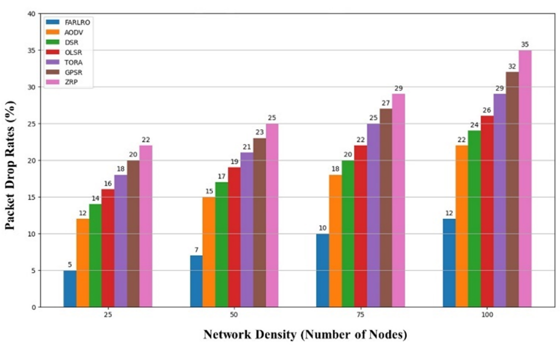

Figure 11: Comparison of packet drop ratio during the flooding period

4.4 Packet Drop Ratio Due to Flooding

Packet drops ratio due to flooding is the ratio of the packets that are dropped because of flooding rather than other reasons, such as link failure or congestion. This is a metric of packet drops due to flooding only, and this gives a clear picture of the impact of flooding on the packets and how well the model can mitigate it.

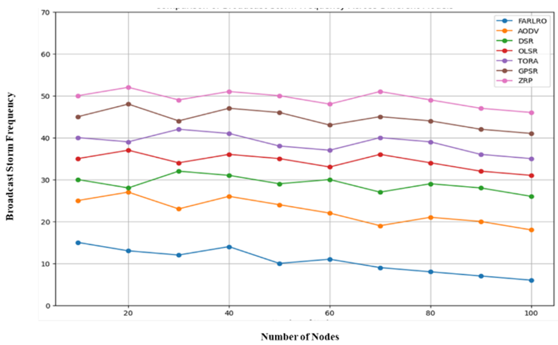

Figure 12: Broadcast storm frequency for all models

4.5 Impact of flooding on network lifetime

This metric measures how the control of flooding enhances the lifetime of the network by avoiding the wastage of energy in unnecessary transmission of packets. Reducing flooding should mean that more energy will be saved across the network, especially for nodes that would have been overloaded with processing and forwarding of unnecessary packets. This in turn, increases the network?s life cycle time. The figure above shows the network lifetime for the FARLRO (Proposed Model) and other routing models, such as AODV, DSR, OLSR, TORA, GPSR, and ZRP, at different densities of the network (25, 50, 75, and 100). The y-axis is the time to the first node death in the network, which is an essential measure of the network?s sustainability and productivity. The FARLRO model is always superior to other models in terms of network lifetime at all network densities. For example, when the network density is 100 nodes, the FARLRO model preserves the network lifetime of 100 units, while ZRP is only 50 units, and GPSR – 55 units. This is true for lower densities as well; FARLRO has achieved 120 units at 25 nodes, while ZRP has only 70 units. FARLRO outperforms other algorithms because of the use of reinforcement learning to determine the best routing path based on the node energy, thus avoiding the early exhaustion of the nodes. The Fig. (14) shows the results of various models.

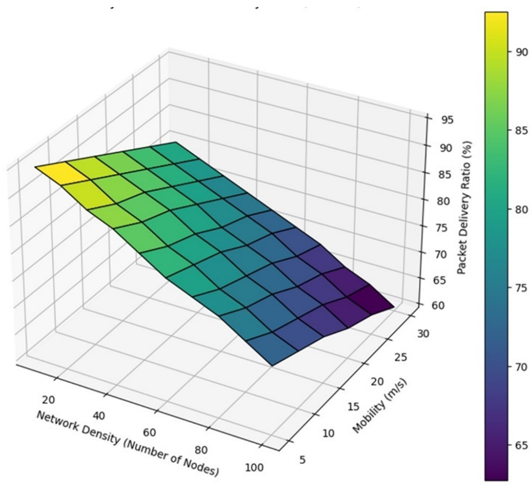

Figure 13: Comparison of mobility award PDR for all models

However, models like ZRP and GPSR have the drawback of faster node depletion because of poor utilization of energy resources and high overhead due to continuous routing updates and flooding. AODV and DSR are better than ZRP and GPSR, but still have a shorter network lifetime than FARLRO, with 75 and 70 units in 100 nodes respectively. This graph clearly shows how FARLRO can prolong the life of the network through the optimization of energy consumption in a densely built network.

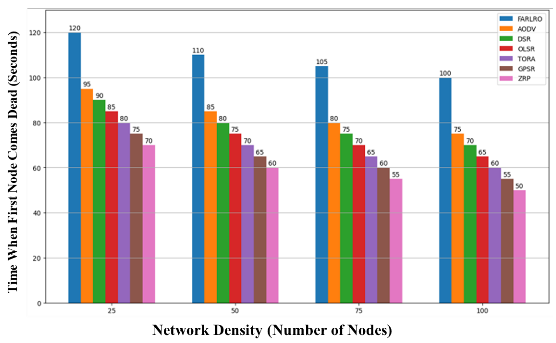

Figure 14: Comparison of impact of network lifetime during the flood period

4.6 Mobility-Aware Packet Delivery Ratio (MA-PDR)

Mobility of nodes directly affects the Packet Delivery Ratio due to its increased frequency of route disturbance and spending more time on the route discovery and maintenance process. It measures the number of packets that have been delivered, considering the mobility of the nodes and how the protocol can respond to it. MA-PDR is used in the analysis of the model in scenarios where the position of nodes is constantly changing. When the MA-PDR is higher, it implies that the proposed FARLRO model can address route disruptions due to mobility and, hence, results in better delivery rates and reduced flooding. The graph shown below shows the comparison of MA-PDR in terms of network density and node mobility speed to understand the performance of different routing models. The x-axis is for network density from 10 to 100 nodes and the y-axis is for node mobility from 5 m/s to 30 m/s. The color intensity or the height of the bar in the graph represents the MA-PDR and the scale goes up to 1 where the bars show better performance. The FARLRO model always shows better performance in terms of packet delivery ratio while the network density and mobility are high. For instance, at 100 nodes density and 30 m/s mobility FARLRO has an MA-PDR of 72% while other models decrease to 60% or below at the same conditions. This robustness is due to the use of the adaptive learning-based technique in FARLRO that can address the dynamic factor inherent in a high mobility environment and high node density by making optimal routing decisions in real time. However, conventional protocols such as AODV, DSR, and OLSR fail to achieve high packet delivery ratios as mobility rises because they are unable to react efficiently to frequent changes in topology. Consequently, these models have a significant reduction in the MA-PDR, especially in the conditions of high node mobility and density, which results in a higher level of packet loss and lower network stability. The graph depicts FARLRO in maintaining the network performance during unfavorable circumstances and is more efficient in packet delivery than other routing protocols.

4.7 Resource Utilization Efficiency (RUE)

RUE measures the degree of utilization of the network resources, for instance, bandwidth or energy, during the routing process with the view of assessing the effects of resource utilization on the general performance of the network. The management of resources should be done effectively to avoid congestion of the network and to increase the lifespan of the network. This metric evaluates how the FARLRO model deals with resources to eliminate floods and maintain network continuity. This means that the higher the RUE, the better the model in leveraging available resources to minimize the occurrence of floods. The RUE results are shown in Fig. (15).

The RUE results are shown in Fig. (15). The graph presented is a 3D surface plot which shows the Resource Utilization Efficiency (RUE) with network density and node mobility as two parameters; the former is in terms of nodes, while the latter is in meters per second. The z-axis is the percentage of RUE, which shows how effectively the network resources are being used depending on the situation. The above graph shows that as the density of the network increases from 10 to 100 nodes and node mobility from 5 m/s to 30 m/s, the RUE generally the decreases. This trend is an indication of the fact that network protocols are facing more difficulty in managing resources efficiently in more complicated and dynamic situations. The FARLRO model is also distinguished by the fact that it keeps higher levels of RUE in all types of organizations. For example, at the network density of 50 nodes and the mobility of 15 m/s, RUE for FARLRO is about 75% while other models are less efficient and can go as low as 65% under the same circumstances. This is because FARLRO has an adaptive reinforcement learning method of managing the network resources that allows it to allocate the resources optimally based on the current network conditions. However, in the case of increased node mobility and overall node density, conventional models such as AODV and OLSR are not very efficient and result in poor utilization of resources. These models are based on static or reactive approaches that cannot efficiently handle the changes in the topology and the increased density of the networks. Therefore, their RUE reduces at a steeper rate than the competitors, suggesting poor network utilization. The graph also emphasizes the effectiveness of FARLRO in controlling resources in a network environment that is dynamic and dense hence making it the best option.

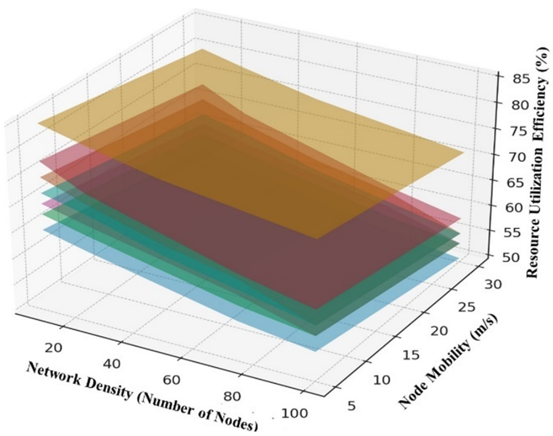

Figure 15: RUE with changing network density and mobility levels

5 Conclusion

In conclusion, the findings of this study show that the developed FARLRO model can help to overcome the major issues related to routing in IoT networks. Some of the network parameters that are incorporated in FARLRO include Residual Energy Level (REL), Available Bandwidth (ABW), Mobility Pattern (MP), Traffic Condition (TC), and Topological Arrangements (TPA). The use of reinforcement learning, especially the Q-learning algorithm and Bellman equation, allows FARLRO to make dynamic routing decisions to avoid flooding and, at the same time, make the best use of the available resources. FARLRO was compared with several other routing protocols, such as AODV, DSR, OLSR, TORA, GPSR and ZRP, under different network scenarios. The results indicate that FARLRO consistently outperforms these models in key performance metrics: it has better flooding ratios, network congestion index, broadcast storm frequency, packet drop ratio due to flooding, network lifetime, mobility-aware packet delivery ratio, and resource utilization efficiency. It is therefore evident that the usage of adaptive and intelligent routing mechanisms such as FARLRO can greatly improve the stability, efficiency and durability of IoT networks. The ability to sustain high performance in conditions of high density and dynamism of the network makes FARLRO a highly suitable solution for optimizing the route in IoT systems. The study is useful in the understanding of reinforcement learning in networking and provides a direction for the future development of IoT routing technologies.

References

[1] Varga, A., Lin, M., S?rensen, J., & Adebayo, Y. (2025). Reinforcement Learning for Energy-Efficient Routing in IoT-Enabled Transport Systems.

[2] Saqhib, M. N., & Lakshmikanth, S. (2025, May). Enhancing AODV routing with machine learning for intelligent IoT path selection. In 2025 IEEE 5th International Conference on Smart Information Systems and Technologies (SIST) (pp. 1-6). IEEE.

[3] Rahmani, A. M., Haider, A., Ali, S., Mohammadi, M., Mehranzadeh, A., Khoshvaght, P., & Hosseinzadeh, M. (2025). A routing approach based on combination of gray wolf clustering and fuzzy clustering and using multi-criteria decision making approaches for WSN-IoT. Computers and Electrical Engineering, 122, 109946.

[4] Debnath, S., & Arif, W. (2025). A hybrid zone-based routing protocol based on ZRP and DSR for emergency applications. International Journal of Communication Networks and Distributed Systems, 31(1), 1-18.

[5] Chen, H., Luo, F., Zhou, J., & Dong, Y. (2024). Multi-Objective Optimized GPSR Intelligent Routing Protocol for UAV Clusters. Mathematics, 12(17), 2672.

[6] Gupta, M., Pandey, S. K., Pareek, P. K., Shukla, P. K., Aggarwal, P. K., & Reddy, P. V. (2025). Efficient Routing and Lifetime Prolongation in IoT founded Wireless Sensor Network Performance with Bee Colony-Inspired Lifetime Enhancement. Journal of Intelligent Systems & Internet of Things, 14(1).

[7] Wei, W., Wu, H., He, Y., & Li, Q. (2024). A multi-objective optimized OLSR routing protocol. Plos one, 19(4), e0301842. Vignesh, G., Vengatesan, R., & Balasubramanian, S. (2024, December). A Review of Temporally Ordered Routing Algorithm Protocol's Challenges and Future Directions. In 2024 9th International Conference on Communication and Electronics Systems (ICCES) (pp. 733-740). IEEE.

[8] Vignesh, G., Vengatesan, R., & Balasubramanian, S. (2024, December). A Review of Temporally Ordered Routing Algorithm Protocol's Challenges and Future Directions. In 2024 9th International Conference on Communication and Electronics Systems (ICCES) (pp. 733-740). IEEE.

[9] B. Tasci, "A Survey: Internet of Things (IoTs) Technologies, Embedded Systems and Sensors," in Communication Technologies and Security Challenges in IoT: Present and Future: Springer, 2024, pp. 173-19

[10] A. Husen, M. H Chaudary, and F Ahmad. 2022. A Survey on Requirements of Future Intelligent Networks: Solutions and Future Research Directions. ACM Comput. Surv. 55, 4, Article 73 (April 2023), 61 pages. https://doi.org/10.1145/3524106

[11] W. a. Kassab, K. A. J. J. o. N. Darabkh, and C. Applications, "A?Z survey of Internet of Things: Architectures, protocols, applications, recent advances, future directions and recommendations," vol. 163, p. 102663, 2020.

[12] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): a vision, architectural elements, and future directions," Futur Gener Comput Syst, vol. 29, 2013// 2013.

[13] R. K. Poluru, S. J. J. o. E. S. Naseera, and T. Review, "A Literature Review on Routing Strategy in the Internet of Things," vol. 10, no. 5, 2017.

[14] A. Dhumane, R. Prasad, and J. Prasad, "Routing issues in internet of things: a survey," in Proceedings of the international multiconference of engineers and computer scientists, 2016, vol. 1, pp. 16-18.

[15] A. M. Hanashi, A. Siddique, I. Awan, M. J. S. M. P. Woodward, and Theory, "Performance evaluation of dynamic probabilistic broadcasting for flooding in mobile ad hoc networks," vol. 17, no. 2, pp. 364-375, 2009.

[16] C. J. W. P. C. Sobin, "A survey on architecture, protocols and challenges in IoT," vol. 112, no. 3, pp. 1383-1429, 2020.

[17] B. S. Chaudhari, M. Zennaro, and S. J. F. I. Borkar, "LPWAN technologies: Emerging application characteristics, requirements, and design considerations," vol. 12, no. 3, p. 46, 2020.

[18] A. Khanna and S. Kaur, "Internet of Things (IoT), Applications and Challenges: A Comprehensive Review," Wireless Personal Communications, vol. 114, no. 2, pp. 1687-1762, 2020/09/01 2020.

[19] Y. B. Zikria, R. Ali, M. K. Afzal, and S. W. Kim, "Next-Generation Internet of Things (IoT): Opportunities, Challenges, and Solutions," vol. 21, no. 4, p. 1174, 2021.

[20] [20] W. Wang, X. Liu, Y. Yao, Z. Chi, Y. Pan, and T. Zhu, "Coexistent Routing and Flooding Using WiFi Packets in Heterogeneous IoT Network," IEEE/ACM Transactions on Networking, vol. 29, no. 6, pp. 2807-2819, 2021.

[21] N. Jiang, Y. Deng, A. Nallanathan, X. Kang, and T. Q. J. I. T. o. W. C. Quek, "Analyzing random access collisions in massive IoT networks," vol. 17, no. 10, pp. 6853-6870, 2018.

[22] M. Naeem, S. T. H. Rizvi, and A. Coronato, "A Gentle Introduction to Reinforcement Learning and its Application in Different Fields," IEEE Access, vol. 8, pp. 209320-209344, 2020.

[23] Y. Akbari and S. Tabatabaei, "A New Method to Find a High Reliable Route in IoT by Using Reinforcement Learning and Fuzzy Logic," Wireless Personal Communications, vol. 112, no. 2, pp. 967-983, 2020/05/01 2020.

[24] R. Farivar, M. Fazeli, and S. G. Miremadi, "Directed flooding: a fault-tolerant routing protocol for wireless sensor networks," in 2005 Systems Communications (ICW'05, ICHSN'05, ICMCS'05, SENET'05), 2005, pp. 395-399.

[25] A. Husen, A. Soahil, M. Hijji, M.-H. Chaudary and F. Ahmed, Optimizing Energy Conservation in V2X Communications for 5G Networks, Computers, Materials \& Continua 2022 Vol. 71 Issue 2 Pages 3479–3495 doi:10.32604/cmc.2022.023840, http://www.techscience.com/cmc/v71n2/45878

[26] L. Chan, K. Gomez Chavez, H. Rudolph, and A. Hourani, "Hierarchical routing protocols for wireless sensor network: a compressive survey," Wireless Networks, vol. 26, no. 5, pp. 3291-3314, 2020/07/01 2020.

[27] A. R. Hameed, S. u. Islam, M. Raza, and H. A. Khattak, "Towards Energy and Performance-aware Geographic Routing for IoT-enabled Sensor Networks," Computers & Electrical Engineering, vol. 85, p. 106643, 2020/07/01/ 2020.

[28] A. Alharin, T. N. Doan, and M. Sartipi, "Reinforcement Learning Interpretation Methods: A Survey," IEEE Access, vol. 8, pp. 171058-171077, 2020.

[29] R. A. Nazib and S. Moh, "Reinforcement Learning-Based Routing Protocols for Vehicular Ad Hoc Networks: A Comparative Survey," IEEE Access, vol. 9, pp. 27552-27587, 2021.

[30] R. Amin, E. Rojas, A. Aqdus, S. Ramzan, D. Casillas-Perez, and J. M. Arco, "A Survey on Machine Learning Techniques for Routing Optimization in SDN," IEEE Access, vol. 9, pp. 104582-104611, 2021.

[31] [A. Telikani, A. Tahmassebi, W. Banzhaf, and A. H. J. A. C. S. Gandomi, "Evolutionary machine learning: A survey," vol. 54, no. 8, pp. 1-35, 2021.