IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 1 December 2023, accepted 28 January 2024, date of publication 5 February 2024, date of current version 16 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3361945

RESEARCH ARTICLE

Gradient Monitored Reinforcement Learning for Jamming Attack Detection in FANETs

JAIMIN GHELANI, PRAYAGRAJ GHARIA, AND HOSAM EL-OCLA, (Senior Member, IEEE)
Department of Computer Science, Lakehead University, Thunder Bay, ON P7B 5E1, Canada

Corresponding author: Hosam El-Ocla (hosam@lakeheadu.ca)

ABSTRACT Unmanned Aerial Vehicles (UAVs) have several military and civilian applications to perform tasks that do not require a central processing unit or human involvement. There are various vulnerable characteristics, alternatively limitations, in UAV systems such as data loss, signals interference, disabling sensors, misleading weapons, cyber attacks, disrupting services, etc. Jamming attack is one of the cyber threats that likely lead to denial of service that often occurs in wireless communication systems like Flying ad hoc networks (FANETs) and Internet of Drones (IoD). Over years, there are several approaches proposed by researchers to detect jamming attacks such as rule-based jamming attack detection mechanism, Bayesian game-theoretic mechanism, IoD-based protection mechanism, communication channel techniques (channel hopping, spectrum spreading, MIMO-based jamming mitigation, coding, etc), delay tolerant networking technique, and cryptographic algorithms, however, these methods were not suitable for jamming detection in UAV environment. The major challenges are on the delivery efficiency, processing time, accuracy, energy consumption, flight distance, and flight autonomy. In this paper, we introduce a method to detect the jamming attack using Reinforcement Learning-based Gradient Monitored (RLGM) mechanism. RLGM maintains safe regions and reduces gradient variance for intended training and this provides a better accuracy of the learning goal. In addition, RLGM achieves prompt training progress and selects precisely the series of parameters required by the network during the training phase. RLGM produces spontaneous derivation of the essential deep network scale over the training process drawing on automatically unvarying trained weights. Our proposed approach outperforms other reinforcement learning methods such as Federated RL, Deep Q Learning (DQL), in addition to non-machine learning based techniques such as GA-AOMDV.

INDEX TERMS Ad hoc, network, FANET, flying ad hoc network, reinforcement, learning, Q learning, jamming, attack, gradient and detection.

I. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) are becoming increasingly popular due to their ability to execute a variety of difficult jobs in three dimensions [1]. Because of high and continuous mobility of UAVs, there are various security applications where UAVs may be deployed to achieve enhanced efficacy such as border monitoring [2] and relay networks [3]. A flying ad hoc network (FANET) is a decentralized network controlled by UAVs to mitigate the issues that infrastructure-based UAV systems would conduct [4]. UAV nodes send data to each other across wireless links and separately communicate with the base station in their range [5].

Figure 1 [6] shows the architecture of of MANET, VANET, and FANET. The operating factors including adaptability, resilience against topological change, and equipment have different requirements for these different networks [7]. UAVs are widely employed in a various practical applications such as real-time surveillance, search and rescue operations, asset inspection, relay for ad hoc networks, and crop spraying. UAV systems are utilized to fulfill missions over decades. In contrast to MANET (Mobile Ad Hoc Networks) [8] and VANET (Vehicular Ad Hoc Networks) nodes [9], FANET devices are subject to higher levels of mobility and speed inconstancy [10]. FANETs are distributed and self-organized

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang.

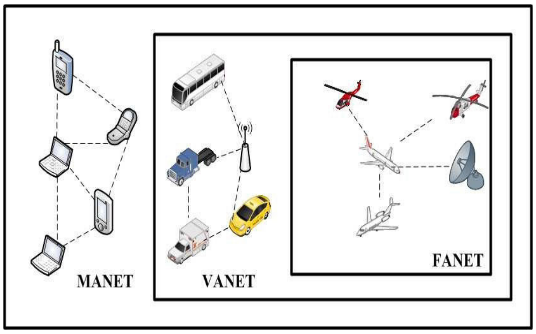


FIGURE 1. Ad Hoc networks.

networks [11]. On the opposite of the nodes in MANETs and VANETs, UAVs move in multi-dimensional and longer transmission range [12]. Also, they must be provided within a limited time range and do not use the common mobility schemes [13]. Due to the uniqueness of FANETs characteristics including mobility in 3D domain [14], sporadic communication [15], topological change [7], and scalable network [16], initiating efficacious routing and establishing FANETs network is a challenge [17]. Alternatively, the base station may be a multiple access edge computing (MEC) server [17]. Data communication in FANET occurs through shared wireless links and as a result, UAV nodes are vulnerable to jamming attacks as shown in Figure 2. Devices in Figure 2 are comprised in Internet of Drones (IoD) which is a layered network control architecture organized fundamentally for regulation the access of UAVs to controlled airspace, and manging navigation services between various locations known as nodes [18]. Jamming attack is one of the cyber threats that likely lead to denial of service that often occurs in wireless communication systems such as FANETs. A jamming attack mainly hinders nodes from interacting and it intervenes with the incoming packets at the receiver end with the lowest transmission power [19]. An attacker can follow several interference models to disrupt the communication between legitimate nodes [20].

Recently, several studies have been conducted to detect jamming attacks [4], [21], [22], [23]. In [4], a rule-based jamming attack detection technique was introduced for UAVs. In [21], an approach using genetic algorithm was proposed in cyber-physical power systems. In [22], a cryptography method was introduced to avoid malicious attacks. In [23], an approach is suggested to detect channel attack utilizing Ordinary Potential Game. However, these mechanisms are improper for jamming attacks detection and data security in FANETs because of the challenges that UAVs encounter. These challenges include [24] 1) Energy constraint particularly with the low node density and hence the communication with the base station is another constraint, 2) High node mobility speed which implies frequent topological change. Therefore, the traditional jamming detection approaches may not be appropriate to respond promptly enough to the unbalanced distributed sensory data, 3) Model-based jamming detection mechanism will be improper assuming that the transition probabilities are initially undetermined. Machine learning (ML) based approach allows an effective learning in a communications network of unbalanced sensory data. In addition, the ML based mechanisms can suite device-level training of mobile nodes particularly in a distributed environment for a jamming attack detection and defense. This in turn would provide more protection for the private data than the centralized model. Also, this local training model would reduce the traffic in the network and consequently lessens the traffic problems such as data congestion and collision that occur often in wireless networks. These challenges were addressed in several studies such as in [5] and [25], however, these mechanisms did not improve the performance sufficiently like the saved energy was marginal. To resist such types of attacks, various methods and spectral retreat mechanisms were proposed [26]. Communication devices fight the blocking node and adaptively senses the risk grade to detect such attacks.

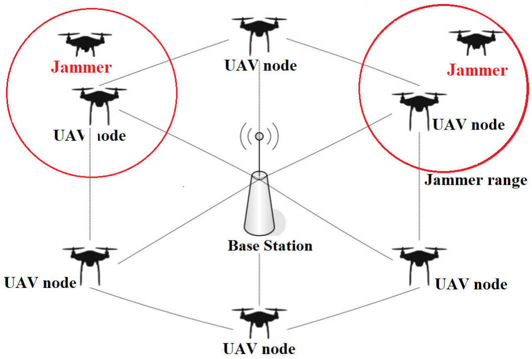


FIGURE 2. Flying Ad Hoc network (FANET).

Reinforcement Learning (RL) has a wide range of applications in recent years and achieved obvious success in different areas such as industrial sector [27], video gaming [28], control problems [29], and multi-agent systems [30]. Over the years, there are advanced studies in deep neural networks (DNN) training and RL algorithms that attracted researchers in these applications [31]. Examples of these advanced methods include Deep Q Networks [32], Deep Deterministic Policy Gradient [33], and Trust Regional Policy Optimization [34]. Training units improved RL mechanism and exploration techniques [35]. RL has two main elements: the agent and the environment that the agent interacts with [36]. The policy that the agent (learner) follows is called the reward and it is needed to take a proper action. RL goes through loops until the ultimate objective is achieved.

UAVs are more vulnerable to jamming attacks than other ad hoc networks (such as MANETs and VANETs) because the covered region in FANET is wider where the communication services provided in FANET are numerous [37]. An anti-jamming mobile mechanism is proposed to enable a node to abandon a frequency or zone that is under jamming attack. Lately, abundant protection mechanisms for UAV communication systems have been proposed. Most of these methods lack the required efficiency to detect and reduce jamming attacks owing to several challenges. Jamming detection demands an excessive amount of energy which is not always available due to the energy limitations in FANET and the low density of UAV devices [38]. As a result of continuous UAV device mobility, sensing data are required often. This data could interfere with the detection mechanism [38].

We propose an efficient RL technique based on the Gradient Monitoring approach [39] (RLGM) to detect jammers in a FANET. RLGM mechanism uses weights as a medium of detection of faulty nodes. Weight parameters of a neural network are decided based on the dynamic development and feedback from the training process itself. The main contributions of our proposed technique are as below:

- Utilize RLGM to detect jamming attacks in FANETs,
- Use RLGM locally where the neural network model is trained on each node with a global model weight update,
- Test our approach through simulation to measure the network performance in comparison with other competitors.

Using RLGM method locally without the need of a support from the global network is quite useful as UAVs have limited FANET communications and also where a global network is sometimes unreachable. This also reduces data traffic communication and sensory data sent over the network and hence this improves the network performance in terms of minimizing traffic jams.

Gradient Monitoring was primarily introduced with the supervised training of deep feed forward neural networks (DNN). Our RLGM mechanism vests for a network scale adjustment through adaptive varying the amount of active training parameters and this expedites the jammers detection and reduces the processing time. The outcome of our RLGM technique is lastly compared to Federated RL and Deep Q-Learning (DQL) techniques.

II. LITERATURE REVIEW

A jamming attack defense strategy was proposed in [24]. The authors introduced a solution based on federated RL which decreases the amount of enroute jammer site hop counts. Federated learning allows sending and receiving of low-level weight to identify the fine-grained properties of the jamming data. However, the end-to-end delay is increased due to relying on learning iterations after each cycle.

Authors of [39] introduced a neural network training process that methodically reduces the gradient variance of the neural network. In this paper, vanilla gradient monitored RL (V-GM) is introduced and this method maintains trust regions and minimizes gradient variance. However, the existence of gradient peaks would obstruct the training process and produces inaccurate gradient data. The selection of the hyperparameter start is another constraint of the momentum-based mechanism of gradient manipulation (M-WGM); this is a complex and unrealistic method as it depletes high amount of energy and accordingly it minimizes the network lifespan.

In [40], the Deep Deterministic Policy Gradient (DDPG) mechanism was proposed which is an efficient RL algorithm. DDPG supports an adaptive critic network that can consider feedback from the actor-network and adjust its loss function according to the policy change rate. However, the incompatibility between neural networks is a challenge. If a function-approximated critic network is incompatible with the actor network, the true gradient cannot be conducted to the actor network.

The authors of [41] introduced a technique to assess the trust values of nodes based on fuzzy logic and proposed an S-OLSR security mechanism using OLSR's multipoint relay (MPR) selection algorithm in FANETs. Simulation results proved the efficiency of the proposed method assuming the existence of black hole attacks. However, S-OLSR assumes that nodes have to choose neighboring nodes which are connected to remote two-hop nodes as MPR nodes, even if these nodes are untrustworthy.

In [42], it was proposed a Qmr routing technique based on Q-learning. Proactive and reactive routing methods are utilized to minimize the transmission delay in UAVs network. However, the energy consumption is ignored and this reduces the efficiency of the network.

The paper [43] suggests a Packet Arrival Prediction (PAP) routing protocol for FANETs, which anticipates each UAV's packet arrival using a Long Short-Term Memory (LSTM) model. It can adaptively prevent packet loss due to buffer overflow. Constrained sorting and routing are presented allowing for collaborative and quick routing decisions. However, it considers a simple and unrealistic mobility model while ignoring the pause times assumption in UAVs that should be predicted.

In [44], it was introduced a novel fully-echoed Q-routing protocol that uses adaptive learning rates. The authors of [45] have proposed a topology-aware resilient routing strategy based on adaptive Q-learning (TARRAQ). This protocol utilizes the rewards to find a stable route and considers the anticipated topological change with low overhead and hence establishes a distinct and distributed routing algorithm. However, these mechanisms require a high energy consumption and frequent update of the Q-table.

In [46], authors have developed a hybrid hierarchical SDN-based mechanism to enhance the network efficiency in terms of reliability and scalability in VANETs. However, with a high flooding rate, the network lifespan is shortened.

Authors in [47] suggested a mechanism to optimize routes across the network utilizing the node's mobility prediction, network connectivity, link permanence, and path existence. This algorithm is complex and, therefore, the latency and energy consumption are quite high.

A hierarchical failure detection mechanism for data communications in VANET is introduced in [48]. This failure detection system adapts to the network configuration while maintaining high-quality of network performance. Links Failure can be avoided through alert messages among vehicles. However, the main shortcoming of this algorithm is the routing overhead which augments the amount of detection messages and network scalability.

In [49] and [50], the concept of authenticated data sharing between legitimate devices is utilized to detect unauthorized nodes. In [49], authors introduced a lightweight distributed mechanism (Lids) to detect and reduce flooding threats in the IoD network. To detect flooding attacks, each drone shares self-counting report with other drones during contacts. In [50], authors presented a quantum-based authenticated communication mechanism for drones in IoD environment. When the size of the network enlarges and the number of drones increases, the amount of data packets exchanged between drones and ground server, augments in the network. As a result, the traffic problems, such as packets collision, frequently occur and this in turn reduces the network performance. In addition, the processing time in [50] is long particularly with large size networks.

In [51], authors used a graph approach to propose a multipath routing framework for SD-UAV networks. This method reduces the outage rate of end-to-end connections in the presence of jammers. However, the end-to-end latency is high. Also, this method is not resilient for the topological change that often happens with UAVs as it uses a graph theory.

In [52], it was introduced a stochastic packet forwarding mechanism to deliver data frames effectively in FANETs deployment. Multipath routing is used to avoid jamming nodes. The algorithm is complex and that would enlarge the latency and lead to fast depletion of the nodes batteries.

A summary of the above discussed methods are summarized in Table 1.

III. THE PROPOSED PROTOCOL

A. PROBLEM STATEMENT

Response time for a request is a crucial factor in the quality of data communications in FANETs. In this regard, in most of the above-described methods, the processing time for jammers detection takes intolerant time. Hence, this depletes a lot of battery energy and this in turn reduces the lifetime of the nodes. As was pointed out earlier, Q-learning is a well-known mechanism for jammers detection in FANETs and it mainly depends on Q-table training. Q-table requires frequent updates for several cycles until all jammers are picked up and this augments the time delay. On the other hand, detection accuracy is a key element to protect the network against such unsafe nodes and therefore security is needed. Also, the network performance in many of the methods described earlier is low in terms of QoS including latency, overhead, and delivery rate.

B. PROPOSED SOLUTION

In this paper, we propose using Reinforcement Learning-based Gradient Monitored (RLGM) mechanism to detect jammer nodes. This method assigns a weight to each node in the FANET network to find jammers which is fast and efficient process. The weights assigning technique used in RLGM is faster than most of the RL methods. Moreover, due to the local training of the nodes' weights, the energy consumption of nodes is low. Additionally, as the training takes place locally, decisions are made for each node when the entire network is unreachable. The main objectives of this paper are below:

- Provide a robust mechanism to detect jammer attacks in FANETs. This method has a fast training progress.
- This mechanism improves the network performance measured through simulation.

In the forthcoming sections, we describe our mechanism to detect jammers in FANETs.

C. PROPOSED METHOD RLGM

Table 2 refers to all the parameters used in our RLGM algorithm. A fully connected FANET network is considered with several jammer nodes trained with mini-batch gradient descent and an arbitrary gradient-based optimizer [39].

We assume that the global weight matrix W_t is for all the nodes in the network. For each node in the FANET network, we start with calculating the decision matrix D_t (i.e. line 7 in Algorithm 1). We calculate the decision matrix D_t using the gradient matrix ∇L_{W_t}, which represents the weights of each node, and the global weight W_t. The weights are locally stored in each node and hence it is easy to fetch the corresponding updated weights. The decision matrix D_t keeps track of every node on the global weight as well as the updated weights.

$$D_t = |\nabla L_{W_t} / W_t| \tag{1}$$

Averaging this decision matrix also gives us the Learning Factor λ that is further used to calculate the masking matrix M_t.

$$\lambda = Average(D_t) \tag{2}$$

To set the elements of gradient matrix ∇L_{W_t} to zero, we calculate the new gradient matrix and define a masking matrix M_t whose values are either one or zero (as calculated in lines 8 and 9 of Algorithm 1). Following that, a typical gradient descent update is used to reflect the weight update.

$$M_t = H(D_t - \lambda\mu) \tag{3}$$

where the learning threshold μ is given by equation (4) with n being the total number of nodes and d_{ij} is the mean of all the elements in the decision matrix D_t

$$\mu = \frac{1}{n} \sum_{ij} d_{ij} \tag{4}$$

$$\nabla \hat{L}_{W_t} = M_t \circ \nabla L_{W_t} \tag{5}$$

where the Hadamard product is indicated by $\circ$ and ∇L_{W_t} is the change in global weight for the mini-batch multiplied by the Masking matrix M_t. Hence and as shown in equation (5),

TABLE 1. Summary of literature review algorithms.

References	Algorithms	ML-based	Strength	Weakness
[4]	rule-base jammers detectiond	x	exhibits a high-level of security	network performance (QoS) is low
[21]	Cumulative Sum	x	can detect jammers effectively	Long processing time
[22]	Hybrid Cryptography	x	provides high security level	energy consumption is not low
[23]	Game theory-based	x	controls channel access	network performance (QoS) is low
[24]	Federated RL	✓	jamming data are identified efficiently	long end-to-end delay
[39]	V-GM/M-WGM	✓	maintains trust regions	gradient data inaccuracy and high energy consumption
[40]	DDPG	✓	supports an adaptive critic network	incompatibility between critic and actor networks
[41]	S-OLSR routing technique	x	efficient to detect blackhole nodes	unreliable routing technique
[42]	Qmr routing technique	✓	latency is low	high energy consumption
[43]	FMCC routing technique	x	can adaptively prevent packet loss	simple and unrealistic mobility model is considered
[44]	Q-routing technique	✓	uses adaptive learning rates	high energy consumption and frequent Q-table update
[45]	TARRAQ routing technique	✓	provides stable route and considers topological change	high energy consumption and frequent Q-table update
[46]	SDN-based routing technique	x	efficient network: reliability and scalability	short lifespan with network flooding
[47]	Obstacle-Aware routing technique	x	optimizes routes and considers node mobility	latency and energy consumption are quite high
[48]	failure detection routing technique	x	adapts to network configuration	large routing overhead
[49]	lightweight distributed	x	detects and reduces flooding threats	Topological change constraint
[50]	quantum-based	x	prevents unauthorized access	long processing time
[51]	multipath routing framework	x	reduces the outage rate of UAV end-to-end connections in the presence of jammers	end-to-end latency is high and is not resilient for the topological change
[52]	packet forwarding	x	provides multipath routing to avoid jamming nodes	energy constraint

TABLE 2. Parameters used in RLGM.

Symbol	Description
λ	Learning factor
μ	Learning threshold
M_t	Masking matrix
W_t	Global weight
H	Heaviside step function
D_t	Decision matrix
η_{start}	Start of GM
η_{repeat}	Mask update frequency

we get the new gradient matrix $\nabla\hat{L}_{W_t}$. As a result, the masking matrix M_t details how much of the updated weight will be derived from the weights that have changed.

Under the assumption of the effect that each parameter has on the forward and backward propagation, the learning process is triggered. Therefore, we can calculate a function that accepts the weights W_t, their corresponding gradients ∇L_{W_t} from the backward pass, a learning threshold $\mu(W_t, \nabla L_{W_t})$, and a learning parameter as inputs is used to create the masking matrix, M_t. As the quantity of learning is involved, we use the absolute values. $D_t(W_t, \nabla L_{W_t})$ and $\mu(W_t, \nabla L_{W_t})$ are denoted as just D_t and μ, respectively, for simplicity. As shown in equation (3), H is known as the Heaviside Step Function which is used to deactivate the gradients that do not reach the required amount of learning.

So going further every node is assigned a new weight $W_{t+1}$and this weight helps in finding the jammer node. Our proposed algorithm is shown in the next section.

$$W_f = W_t + \rho\nabla\hat{L}_{W_t} \tag{6}$$

Example: We initialize the global weight W_t with a unity array. So, we can calculate the decision matrix D_t using the gradient matrix L_{W_t} and get:

$$D_t = \begin{bmatrix} 6.7 & -3.4 & 8.9 & \ldots & -2.6 \\ 2.8 & 4.2 & 9.1 & \ldots & 6.3 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 5.6 & -1.6 & 9.3 & \ldots & 3.8 \end{bmatrix}$$

where each element in the matrix represent the weight of the nodes. Next, we calculate the average of the decision matrix to get the learning factor λ and we can say we get a value of 0.5. Accordingly, we can calculate the masking matrix M_t and update the gradient matrix $\hat{L}_{W_t}$ to obtain the final weights where we get a matrix of weights with its respective nodes as assigned in the decision matrix D_t.

$$W_f = \begin{bmatrix} 0.33 & 0.5 & 0.36 & \ldots & 0 \\ 0.44 & 0.65 & \mathbf{1.5} & \ldots & 0.56 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \mathbf{1.03} & 0.23 & 0.76 & \ldots & 0.22 \end{bmatrix}$$

As we see in the final weights matrix, we have 2 jammer nodes that we can mark in the network and later eliminate from the whole network.

As we see, Figure 3 [24] represents the network configuration and how the weights are assigned to every node in the network. The 2 nodes on the top of the network with weights 1.5 and 1.03 are marked as jammer nodes and those nodes will be eventually eliminated from the network.

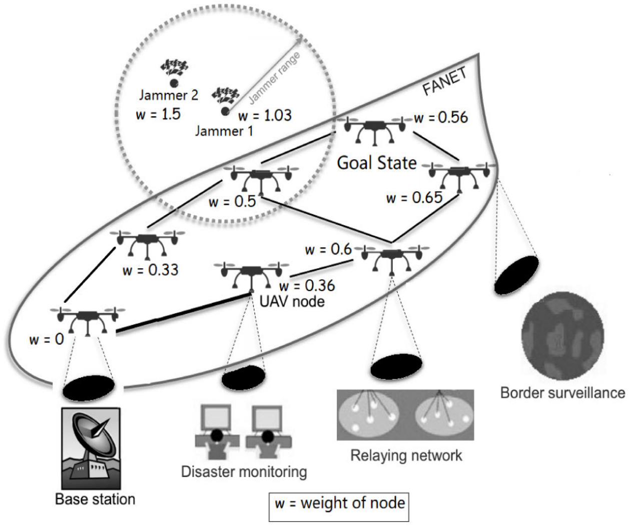


FIGURE 3. Network configuration.

IV. METHODOLOGY

Algorithm 1 shows how the weights are calculated for each node. The weight that is achieved as the output is finally assigned to the respective node and hence we can check if the node is a jammer node or a non jammer node. So for understanding how the training process using RLGM works, we consider the following procedure and the algorithm:

Algorithm 1 RLGM Jammer Detection

```
Input ($\nabla L_{wt}$, $\rho$, $\lambda$, $\eta$, $\eta_{start}$, $\eta_{repeat}$)
Initialize: Masking Matrix ($M_t$), Global Weight ($W_t$)
Sequence:
  if $\eta >= \eta_{start}$
  every $\eta_{repeat}$
  For each node $n$
  Decision matrix $D_t = |\nabla L_{W_t}/W_t|$
  Learning Factor $\lambda = Average(|\nabla L_{W_t}/W_t|)$
  Masking matrix $M_t = H(D_t - \lambda\mu)$
  New Gradients : $\nabla\hat{L}_{W_t} = \nabla L_{W_t} \circ M_t$
Output $W_f = W_t + \rho\nabla\hat{L}_{W_t}$
  if $W_f <= 1$
      Non-Jammer node is detected
  else
      Jammer node is detected
  End If
```

- **Initialization and Required Parameters:** We consider a fully connected FANET network with multiple jammer nodes introduced within the network. We first split the data to create a local environment for the procedure. Each node including the jammer nodes is initialized with a specific value of gradient (i.e. Global Weight) which eventually is responsible to create a gradient matrix ∇L_{W_t}.
- **Decision Matrix D_t:** Next, we move forward with creating the Decision Matrix D_t. As shown in the Algorithm 1 D_t can be calculated using the Gradient Matrix ∇L_{W_t} and Weight W_t. A global weight is initially assigned to each node in the network and in turn, we use it as the initial weight W_t. Once we have the decision matrix D_t we then move forward to the Learning Factor λ.
- **Learning Factorλ and Learning Threshold μ:** RLGM is the derivation of appropriate circumstances for the activation and deactivation of the gradients, ∇L_{W_t}, flow which includes determining the learning threshold μ and factor λ, and time for freezing and unfreezing gradients based on the actual state of learning. It is clear that maintaining a fixed integer value for the learning threshold λ across all gradients is inappropriate since the distribution of learning represented by the gradients may vary across layers and time. Additionally, it is not simple to select a single constant learning value for a variety of learning activities. Thus, by using functions like the mean or percentile of the values of the decision matrix D_t, the learning threshold is made modifiable. This offers a way to guarantee that a specific amount of the gradients will always be permitted.
- **Masking Matrix M_t and Heaviside Step Function *H*:** The gradients that are under the learning condition are rendered inactive according to the definition of the Heaviside function, H when we calculate the Masking Matrix M_t. Finally, we get the new gradient values when we apply the masking matrix M_t using the Hadamard product to the gradient matrix.
- **The Final Weight:** The new weights W_f for each node in the network are calculated as shown in the output in Algorithm 1. If the new weight W_f is 0 then the node is considered to be a non-jammer node. If the new weight W_f is 1 then we say that the jammer is detected. In short, our proposed RLGM is shown in Algorithm 1.

We went through various simulations to carry out the most efficient and fastest approach to present as our approach. We used various typologies and different evaluations to check which of them were the most efficient. We have presented 2 different cases as our final results in this paper that we found to be the best approaches.

Figure 4 shows the flowchart of our workflow:

- **Creation of the network:** We started with creating a network, it consists of UAV Nodes, 1-Base Station, and 1- Multi-access edge server. Every node in the network is treated as a FANET node irrespective of the jammer node. We implement AOMDV to return possible routes.
- **Global weight and Training** The global weight is then initialized for all the client nodes and is based on the local model. The global weight is responsible to assign initial gradients to every node so that further procedures can be carried out. For training the model we consider the whole network and the new weights are trained on each and every node locally. Since the training takes place locally, it is easy for the nodes to make necessary decisions.

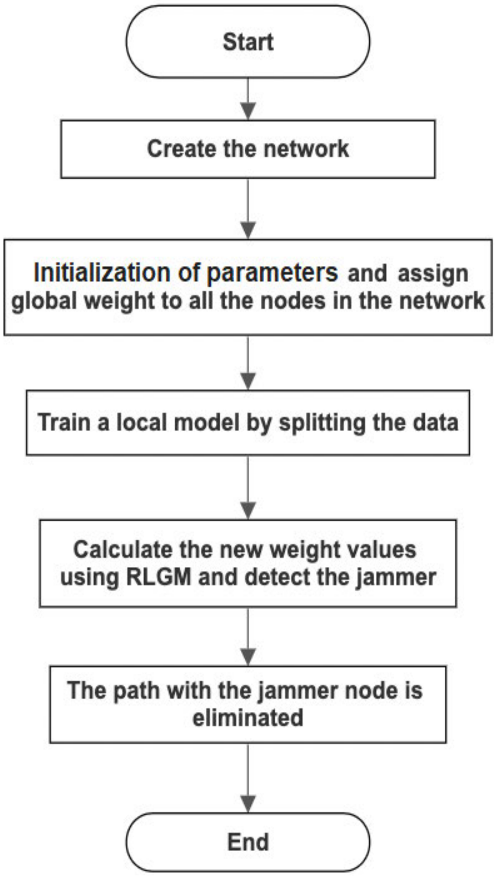

Start
Create the network
Initialization of parameters and assign global weight to all the nodes in the network
Train a local model by splitting the data
Calculate the new weight values using RLGM and detect the jammer
The path with the jammer node is eliminated
End

FIGURE 4. Flowchart for jammer detection using RLGM method in FANET.

At every point when the algorithm is running, the decision matrix D_t is updated with the new weights so we can determine the weights of every node in the network. The final weight W_f is calculated to find the jammer node.

- **Detect jammers using RLGM:** Based on the global weight W_t, new weights W_f are calculated using the RLGM approach and based on the calculated weights we can finally detect the jammer nodes in the network. The path with the jammer node can hence be eliminated from the AOMDV routes.

V. PERFORMANCE EVALUATION

A. EXPERIMENTAL SETTINGS

A 64-bit, Intel i7 CPU @ 2.50-GHz processor and 8.00-GB RAM simulation environment was used to train the detection model. We started with installing ubuntu 14.04 on the system for setting up the simulation using ns3.26 [14] assuming parameters in Table 2. We designed a FANET topology and introduced a jammer node with constant jamming signals that caused interference in the communication of the UAV nodes that were nearby as proposed in the Adaptive Federated RL [7]. Network parameters are set to these values in Table 2. unless otherwise noted.

For the dataset, we set a binary class problem in which the two classes were identified as jammer and non-jammer classes using 50% of the training dataset and 50% of the validation dataset. We pre-processed the testing dataset to have two sets of instances: one set with 80% non-jammer and 20% jammer instances, and the other with 80% jammer and 20% non-jammer instances. The average parameter of the imbalanced dataset was then obtained by averaging the performance parameters from these two sets.

TABLE 3. Simulation parameters.

Parameter Type	Parameter Value
Environment	ns3
Operating System	Ubuntu 14.04
Parameters	Throughput, End to End Delay, Packet Delivery Ratio, Energy Consumption, Routing Overhead, Network lifetime
Number of nodes	100
Mobility Speed	10m/s
MAC Protocol	IEEE 802.11n
Bandwidth	20 MHz
Initial energy of UAV node	1000 J
Energy threshold	100 J
Power of jammer node	100 dBm
Mobility Model	3-D UAV ad-hoc Gauss-Markov mobility model
Routing Protocol	Reinforcement Learning using Gradient Monitoring (RLGM)
Comparison Protocols	Federated RL, Deep Q Learning (DQL), GA-AOMDV.

There are no initial weights assigned to the nodes as per RLGM and thus each node is still vulnerable [8]. For the successful completion of our method, the nodes should have assigned weights between 0 and 1, considering 0 as non-jammer node and 1 as a jammer node as we followed the RLGM algorithm [8].

B. FEATURES

The optimal set of features are utilized during the training process. Those features are for parameters that distinguish between legitimate data packets and the malicious ones. We consider the applied features below:

- Data type: These feature are for those related to transmission bytes such as IP addresses of the source and destination, transmission protocol, time to live value (TTL).
- Performance: This is for the nodes performance such as sending rate in terms of amount of packets/second, pause time, packets duration, packet delivery ratio (PDR).
- Existence: This is for the position of nodes within the range of the network.

A FANET is usually a private network where the jammer would have a global IP address that should not have access to this local network of UAV devices. TTL for legitimate packets in these private networks should not be higher than 30. For the malicious node, TTL would have an abnormal values higher than 30. Jammers tend to send excessive amount of data packets with a limited packet size and duration to overwhelm nodes. PDR is low when links are under jamming attacks, and vice versa. Therefore, through the average packets interval, with other features, the learning process can converge preciously.

For features selection, we use Information Gain technique where the amount of mutual information gained from a combination of monitored variables [53]. With respect to ML, Information Gain method is beneficial for selecting various remarkable features based on theories that measure the significance of information concluded from a specific feature. The global weight W_t stored at each node represents the average of features scores.

C. BACKGROUND

In this section, we compare RLGM with three new competitors including Federated RL [5], [24], Deep Q-Learning (DQL) [54], [55], and energy-efficient multi-path algorithm called genetic algorithm (GA)-based Ad Hoc On-Demand Multipath Distance Vector routing protocol (GA-AOMDV) [56]. Federated learning allows sending and receiving of low-level weight to identify the fine-grained properties of the jamming data. However, the end-to-end delay is increased due to relying on learning iterations after each cycle.

With DQL method, the optimal communication path has been learned through deep Q-networks. In DQL, the action is selected randomly that occasionally maximizes the total predicted reward linked with a state. This method is approximate and the time consumed during the learning is long due to the slow convergence process.

With GA-AOMDV, the route is selected based on the minimum energy consumption at the nodes using the GA. The optimum array of routes are sorted based on the GA algorithm scores during the route discovery phase. These best routes should detect and avoid jamming attack nodes. If there is a threat of jamming attack at a node that occurs within the data path during the data transfer phase, data packets will not arrive at the destination. In this case, the sender can know about this delivery failure through the ICMP message sent by the router which is unable to forward the data packets. Alternatively, the sender may know through the absence of the acknowledgment (ACK) packet during a time-out. In these cases, the sender does not need to go through the route discovery phase as it will select an alternative route as GA-AOMDV is a multipath routing method.

D. BENCHMARKS

We test RLGM to examine the performance of our proposed RLGM mechanism through two approaches. In the first approach, we measure the quality of service parameters such as energy consumption, throughput, end-to-end delay, etc. In the second approach, we ran a simulation for a training experiment with various topology. We measure the accuracy, energy, and routing time.

To test our algorithm, we compare RLGM with two types of mechanisms: one is based on ML (Federated RL and DQL) and the other mechanism is based on non-ML technique (GA-AOMDV).

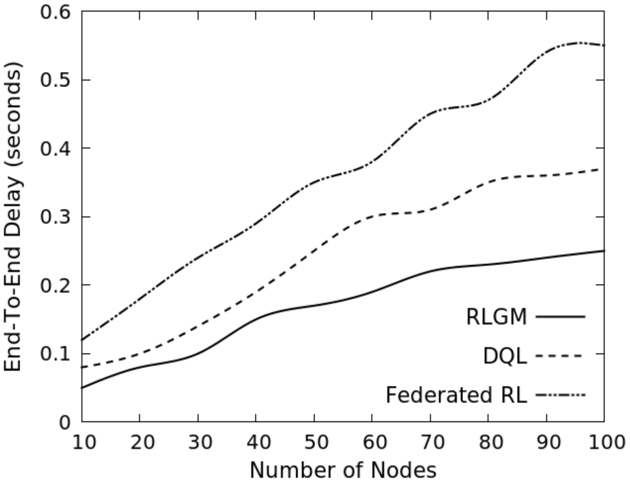


FIGURE 5. Comparison of Federated RL, DQL and proposed method RLGM for end-to-end delay with number of nodes.

E. ML-BASED TECHNIQUES

In this section, we compare RLGM with methods based on ML approach.

1) END-TO-END DELAY

Figure 5 Shows the effect of average end-to-end delay as a function of network density. At low-density networks, RLGM has the lowest delay. Our algorithm decreases the need for the route discovery mechanism through utilizing jammer's detection information. As the number of UAVs rises, so does the delay of RLGM, although our protocol has a substantially lower E2E delay when compared to DQl and Federated RL. RLGM selects the route using AOMDV for those routes avoiding passing through jammer attackers links.

Because of the perimeter forwarding technology utilized, Federated RL has the longest latency as shown in Table 3.

The average E2E latency is shown in Figure 6 as a function of UAVs speed. RLGM selects the route with the minimum delay and hence reduces the packet delivery latency. At high node speeds, RLGM has less delay. This is due to the RLGM mechanism which considers the node velocity and location and accordingly avoids unstable and jammed links. In this case, data retransmissions are minimized and therefore the E2E is lowered obviously.

2) ENERGY CONSUMPTION

Figure 7 shows the energy usage of the comparative routing algorithms for various UAV densities. Gradient Monitoring has less energy-intensive than Federated RL and DQL. RLGM picks the non-jammer nodes forwarder and only those nodes use less energy since it is deemed an energy-consumer of nodes in a weight reward function. Furthermore, RLGM takes into account a node's residual energy, and only nodes with an energy level greater than the threshold level can engage in communication. Table 4 shows the remarkable energy saving of RLGM compared to other methods. In this

TABLE 4. End-to-end delay in Figure 5.

No. of nodes	Federated RL	DQL	RLGM
10	0.12	0.08	0.05
20	0.18	0.1	0.08
30	0.24	0.14	0.09
40	0.29	0.19	0.12
50	0.35	0.25	0.15
60	0.38	0.3	0.2
70	0.45	0.31	0.22
80	0.47	0.35	0.22
90	0.54	0.36	0.24
100	0.55	0.37	0.25
Sum	3.57	2.45	1.68
Saving %	52.94	31.24	

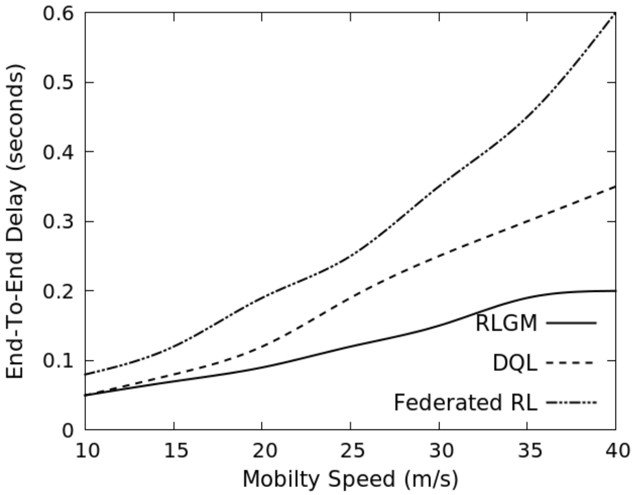


FIGURE 6. Comparison of Federated RL, DQL, and proposed method RLGM for mobility with end-to-end delay.

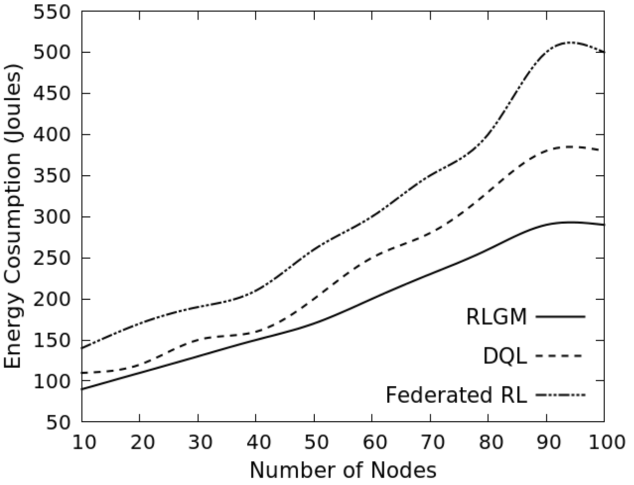


FIGURE 7. Comparison of Federated RL, DQL and proposed method RLGM for energy consumption with number of nodes.

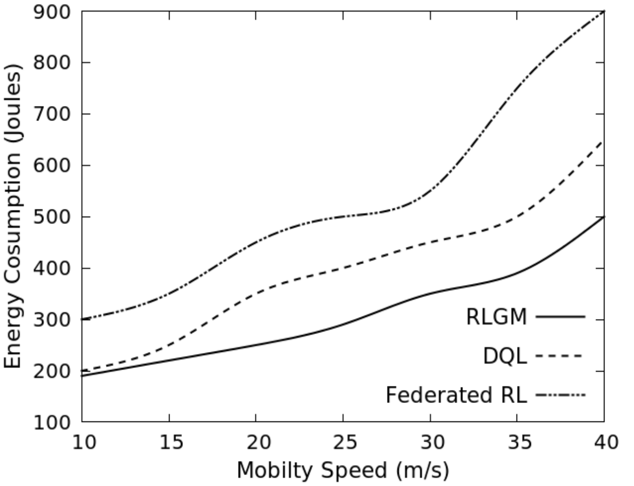


FIGURE 8. Comparison of Federated RL, DQL, and proposed method RLGM for mobility with energy consumption.

regard, RLGM has the minimum energy consumption with nodes mobility as shown in Figure 8.

3) PACKET DELIVERY RATIO

The influence of varying node densities in a FANET on the PDR is investigated in Figure 9. Simulation results reveal that our proposed RLGM routing protocol outperforms Federated RL and DQL routing protocols. Initially, the network encounters frequent disconnection due to the low density of UAV nodes resulting in poor PDR.

RLGM, on the other hand, delivers a high performance gain even at low UAV densities since it uses Weight Rewards information to forecast the availability of an appropriate forwarding node. As the number of UAV nodes grows, so does network connectivity, and therefore the PDR increases. RLGM has a higher PDR than DQl due to the use of Weight data rather than the Q-table.

The PDR as a function of UAV node mobility is seen in Figure 10. RLGM outperforms Federated RL and DQL. RLGM considers the dynamic mobility of nodes and this is based on two different types of speeds: essential mobility and a mobility assumed by Gradient vectors. In this regard, RLGM updates its mobility dynamically to successfully deliver the packet. This adaptive mobility would reduce the packet drop rate and accordingly enhance the PDR.

4) THROUGHPUT

Figure 11 shows throughput and it is a metric that measures the real transmission capacity in a channel. RLGM outperforms other protocols particularly when the number of nodes increases. This is because the amount of safer alternative routes augments when jammer attackers exist in the network. Throughput gain is shown in Table 6.

When the mobility speed increases, the nodes can deliver the data quicker and this ameliorates the throughput as shown in Figure 12. RLGM selects the most stable routes and hence improves the throughput compared to other protocols.

TABLE 5. Energy consumption in Figure 7.

No. of nodes	Federated RL	DQL	RLGM
10	140	110	90
20	170	120	105
30	190	150	120
40	210	160	139
50	260	200	160
60	300	250	190
70	350	280	220
80	400	330	250
90	500	380	298
100	500	380	293
Sum	3020	2360	1865
Saving %	38.24	20.97	

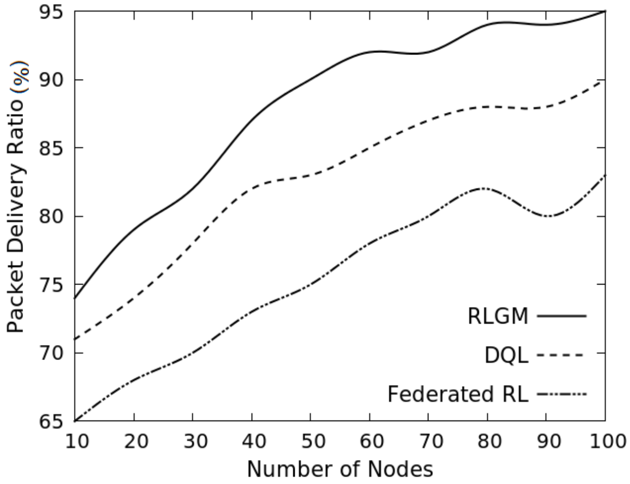


FIGURE 9. Comparison of Federated RL, DQL and proposed method RLGM for packet delivery ratio with number of nodes.

TABLE 6. Packet delivery ratio in Figure 9.

No. of nodes	Federated RL	DQL	RLGM
10	65	71	74
20	68	74	79
30	70	78	82
40	73	82	87
50	75	83	90
60	78	85	92
70	80	87	91
80	82	88	94
90	80	88	94
100	83	89	95
Sum	754	825	879
Gain %	16.57	6.54	

5) NETWORK LIFETIME

Figure 13 depicts the network lifespan as the number of nodes grows. The suggested RLGM ensures that energy consumption in the network is balanced. As a result, the network lifespan is greatly increased. Furthermore, RLGM may locate an appropriate end-to-end channel with minimal latency for data packet forwarding, lowering the number of re-transmissions and maximizing energy use.

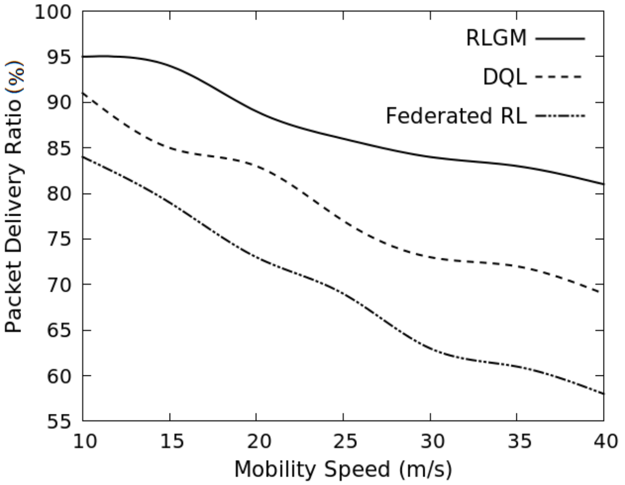


FIGURE 10. Comparison of Federated RL, DQL, and proposed method RLGM for mobility with packet delivery ratio.

Figure 14 shows the network lifetime for different UAV velocities. RLGM is a weight based monitoring routing protocol where we utilize the Gradient monitoring scheme to maximize the network lifetime. In RLGM, the weight function is considered the residual energy of UAV nodes and energy distribution among the neighboring nodes. Selecting the high weights vector will maximize the network lifetime compared to Federated RL and DQL.

F. NON-ML TECHNIQUE

In this section, we compare RLGM with GA-AOMDV as a non-ML based technique. GA-AOMDV is a multipath method where the route discovery process is not needed often whenever there is a jamming threat as alternative routes are likely available. However, because of the time required for the arrival of the ICMP message or the time-out that the sender has to wait to receive the ACK message, E2E enlarges. This E2E augment will occur particularly when the number of nodes increases as shown in Figure 15 where the packets have to go through a longer route.

More energy will be wasted when jamming attackers are involved in the selected routes. In this case, data packets should be retransmitted and hence more energy is consumed at the nodes of the data path as shown in Figure 16. Accordingly, the PDR and the throughput will reduce as shown in Figures 17 and 18.

VI. EXPERIMENTAL RESULTS

As previously stated, in our case, the UAV agent receives a weight 1 if it finds a jammer and a weight of 0 if it detects a non-jammer node. We created 2 different cases of the topology to consider all the possibilities that can occur. To analyze the Reinforcement Learning Gradient Monitoring

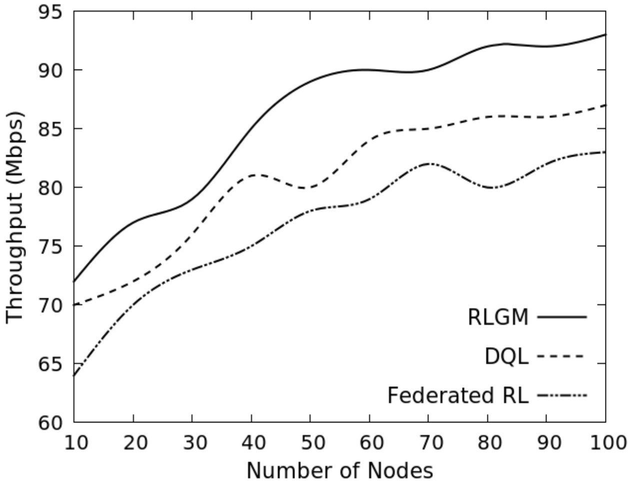


FIGURE 11. Comparison of Federated RL, DQL, and proposed method RLGM for throughput with number of nodes.

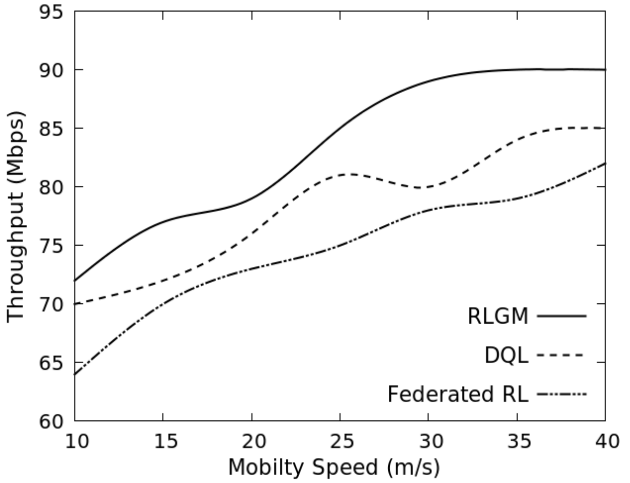


FIGURE 12. Comparison of Federated RL, DQL, and proposed method RLGM for mobility with throughput.

TABLE 7. Throughput in Figure 11.

No. of nodes	Federated RL	DQL	RLGM
10	64	70	72
20	70	72	77
30	73	76	79
40	75	81	85
50	78	80	89
60	79	84	90
70	82	85	90
80	80	86	92
90	82	87	92
100	83	87	93
Sum	766	807	859
Gain %	12.14	6.44	

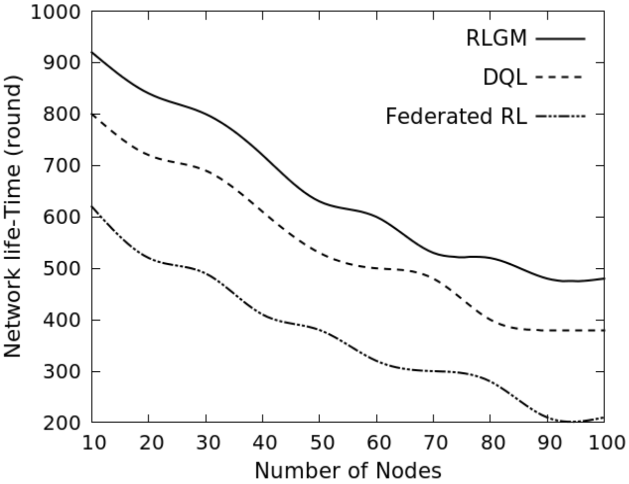


FIGURE 13. Comparison of Federated RL, DQL, and proposed method RLGM for network lifetime with number of nodes.

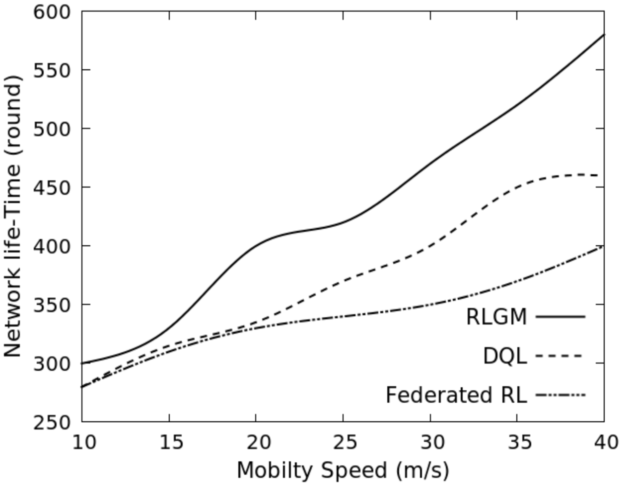


FIGURE 14. Comparison of Federated RL, DQL, and proposed method RLGM for mobility with network lifetime.

(RLGM) defense strategy, we examine a topology of 25 discrete communication cells (as shown in Figures 19 and 20) in which a UAV travels from a source point to a target point outlined in 3-D geometry. In Figures 19 and 20, N-checked cells are for those legitimate nodes while the J-stamped cells are those representing a jammer. The spatial coordinates of the communication cells are represented by the x and y axes. We assume 3 jammers are distributed in the network of the 25 cells. The initial point is set at cell 0, while the destination point is set at cell 24. Two cases are presented below:

Case 1: As shown in Figure 19, the distance between the start point and the target point is shorter. The UAV must go through 5 routers from the source to the destination. At cells 6, 12, and 18, three jammers are placed in the UAV's path.

Case 2: We consider a longer distance separating the source and destination points as shown in Figure 20. The UAV must go through 8 routers from the source to the end point. Three jammers are positioned in the path of the UAVs at cells 3, 4, and 9.

The comparison between the proposed RLGM, DQL, and Federated RL for jamming detection during 10

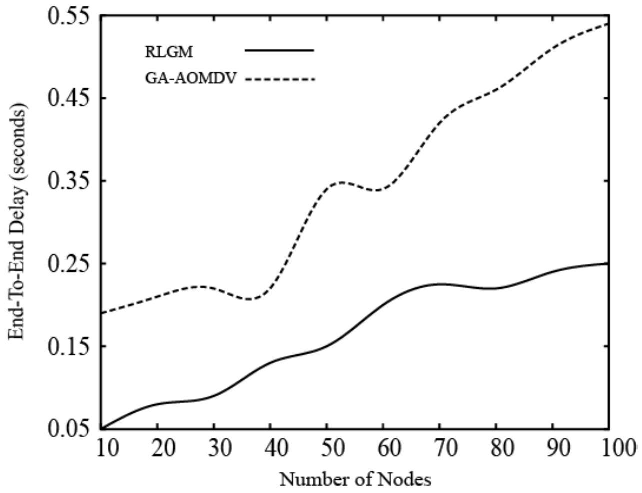


FIGURE 15. Comparison of proposed method RLGM and GA-AOMDV for number of nodes with end-to-end delay.

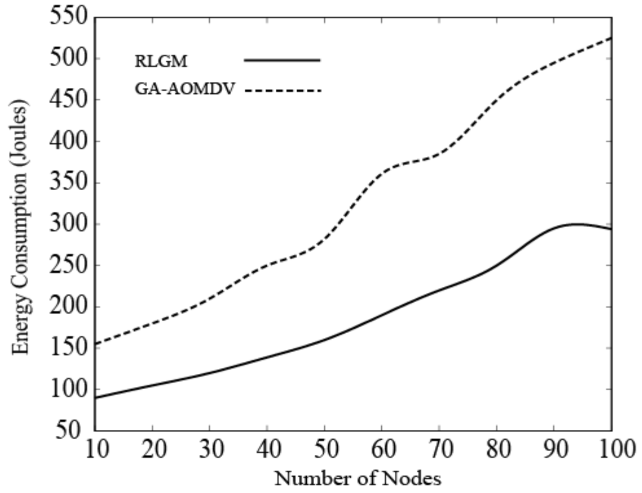


FIGURE 16. Comparison of proposed method RLGM and GA-AOMDV for the number of nodes with energy.

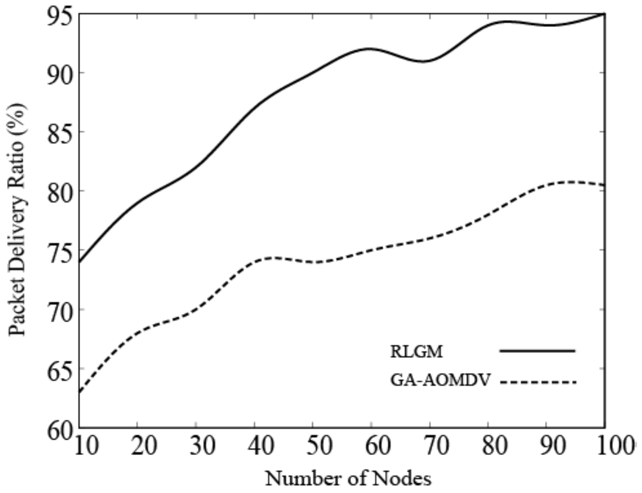


FIGURE 17. Comparison of proposed method RLGM and GA-AOMDV for the number of nodes with packet delivery ratio.

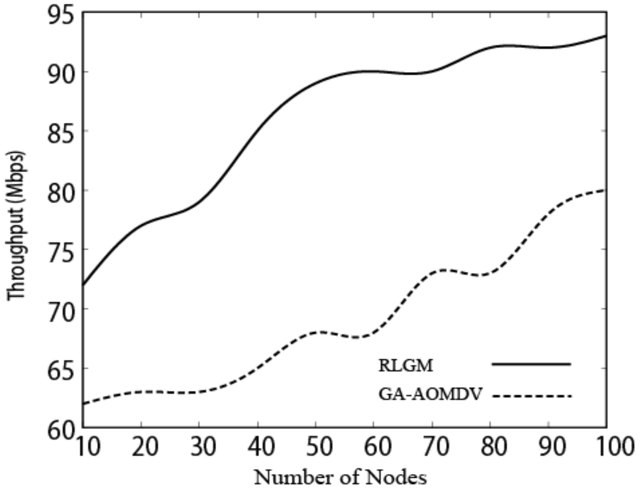


FIGURE 18. Comparison of proposed method RLGM and GA-AOMDV for the number of nodes with throughput.

communication rounds after 25 epochs each averaged over 10 samples is shown in Figure 21. Following round 1, RLGM, DQL, and Federated RL displayed average accuracy of 35%, 33%, and 30%, respectively, in the ns-3-simulated FANET dataset. However, after round 10, the average accuracy of RLGM grew to 80%, whereas the average accuracy of the DQL and Federated RL models increased to 60% and 50%, respectively.

For local jamming attack detection throughout 10 communication rounds, Figure 22 compares the average running times of the RLGM, DQL, and Federated RL model. Throughout the 10 communication cycles, the RLGM model, DQL, and the Federated RL model had average running times of 7.5, 8.0, and 8.1 seconds, respectively. While the suggested RLGM model's performance in detecting jamming attacks (with an average accuracy of 80%) is substantially greater than that of the DQL and Federated RL model (with an average accuracy of 60% and 50%), their average running times are nearly identical. This demonstrates unequivocally that the RLGM model can achieve much-improved performance without adding any more running time. Figure 23 compares the average energy consumption of node between Federated RL, DQL, and the proposed RLGM model. Here, our proposed RLGM consumes less energy than DQL and Federated RL.

VII. PROTOCOL COMPLEXITY

Through the training, peaks in the gradients take place irregularly interrupting the training process and this is a possible drawback of the RLGM mechanism. Precise selection of the hyperparameter η_{start} is another limitation of the RLGM. This can be likely avoided by benefiting from the input of the reward collecting over the training process.

There are several constraints that characterize FANETs compared to MANETs such as node density is low, nodes mobility speed is high and hence the network has a

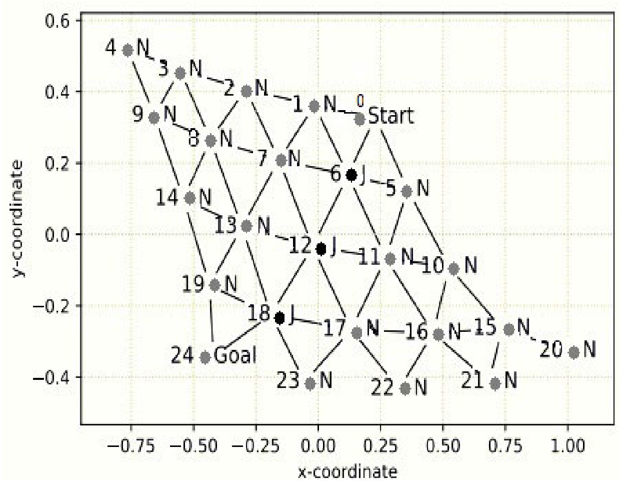


FIGURE 19. Experimental scenario of network topology consisting of 25 communication cells: Case 1 topology.

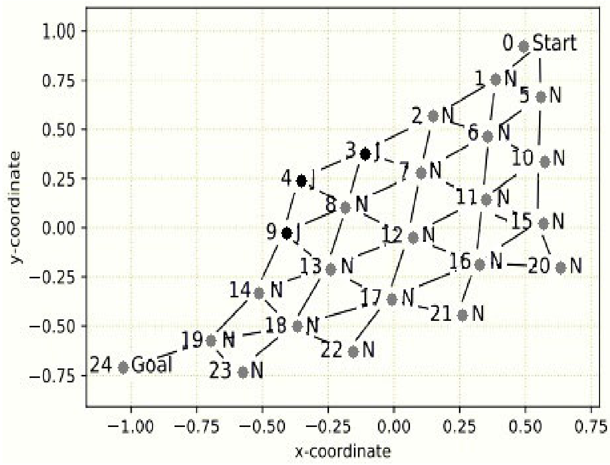


FIGURE 20. Experimental scenario of network topology consisting of 25 communication cells: Case 2 topology.

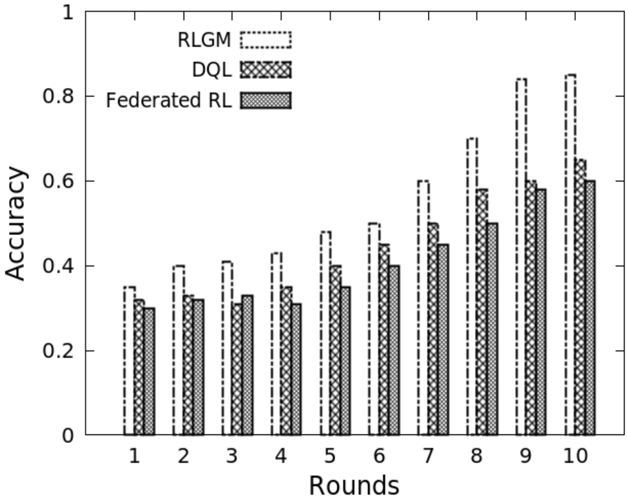


FIGURE 21. Comparison between Federated RL and RLGM for jamming detection in terms of average accuracy.

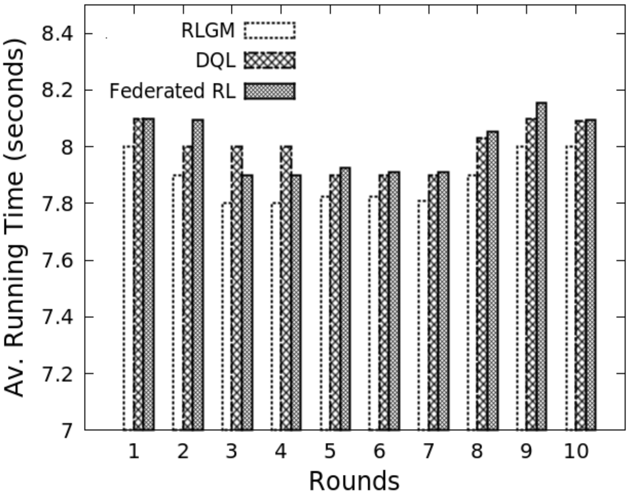


FIGURE 22. Comparison of the average local running time between Federated RL and RLGM.

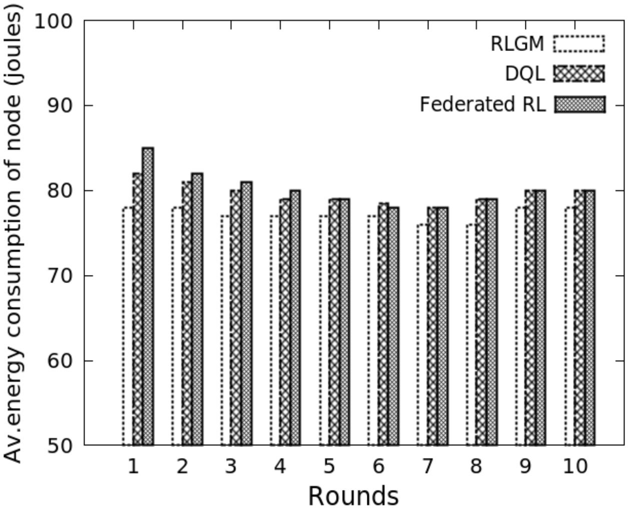


FIGURE 23. Comparison of the average energy consumption of node between Federated RL and RLGM.

frequent topological change, and the communications with a centralized unit is limited [42]. In this case, RLGM is suitable to work with FANETs where it does not require intensive communications with the base station as the weights calculation is local. This in turn reduces the data traffic with the global network and hence maximizes the network performance. High nodes mobility does not affect the performance as it was proved in our results. RLGM does not degrade the network performance with its scalability, which is one of the characteristics of FANETs, as was shown in our results. RLGM can though work with a MANET given that its nature relaxes all the above limitations. As a result, RLGM can be applied in other environments as it works with a distributed system.

Jamming threats target mostly networks at the physical layer but it can be as well at the cross-layer attacks. To prevent or reduce the ability of jammers to intercept the data communication, some security precautions can be considered such as spread spectrum modulations, coding, channel hopping, etc. Also, legitimate data packets should use sophisticated encryption techniques to prevent the attacker from decoding the data configuration. Data rate tuning is another security solution to reduce the packet size transmitted to the jammed nodes and hence send data at a lower rate [57], and this in turn minimizes the possibility of data jamming. In this paper, we introduced RLGM to detect jamming blocks and as a result select the routes that avoid jammed nodes.

VIII. CONCLUSION

This paper addresses an efficient and faster approach based on reinforcement learning (RL) to detect jammers in Flying Ad Hoc Networks (FANETs). Our proposed mechanism utilizes the Gradient Monitoring approach (RLGM) which is robust and faster than most RL techniques. Here, we introduced the concept of weights, also known as gradients, to be used in jammers detection. After a faster detection of the jammer in the network, it is easy to eliminate those attacker nodes to select a safe route. Our algorithm outperforms recent methods such as Federated RL, DQL, and GA-AOMDV in terms of QoS metrics. RLGM has a high accuracy and hence the selected route is quite secure and this in turn would maximize the performance of the QoS parameters such as throughput and latency.

REFERENCES

[1] M. Hessel, H. Soyer, L. Espeholt, W. Czarnecki, S. Schmitt, and H. Van Hasselt, "Multi-task deep reinforcement learning with PopArt," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 3796–3803, doi: 10.1609/aaai.v33i01.33013796.

[2] L. Xiao, D. Jiang, D. Xu, H. Zhu, Y. Zhang, and H. V. Poor, "Two-dimensional antijamming mobile communication based on reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9499–9512, Oct. 2018.

[3] L. Xiao, X. Lu, D. Xu, Y. Tang, L. Wang, and W. Zhuang, "UAV relay in VANETs against smart jamming with reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4087–4097, May 2018.

[4] H. Sedjelmaci, S. M. Senouci, and N. Ansari, "A hierarchical detection and response system to enhance security against lethal cyber-attacks in UAV networks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 9, pp. 1594–1606, Sep. 2018.

[5] N. I. Mowla, N. H. Tran, I. Doh, and K. Chae, "Federated learning-based cognitive detection of jamming attack in flying ad-hoc network," *IEEE Access*, vol. 8, pp. 4338–4350, 2020.

[6] I. Sumra, P. Sellappan, A. Abdullah, and A. Ali, "Security issues and challenges in MANET-VANET-FANET: A survey," *EAI Endorsed Trans. Energy Web*, vol. 5, no. 17, Apr. 2018, Art. no. 155884.

[7] Y. Shi, Y. E. Sagduyu, T. Erpek, K. Davaslioglu, Z. Lu, and J. H. Li, "Adversarial deep learning for cognitive radio security: Jamming attack and defense strategies," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[8] A. Guillen-Perez and M.-D. Cano, "Flying ad hoc networks: A new domain for network communications," *Sensors*, vol. 18, no. 10, p. 3571, Oct. 2018.

[9] NS-3. *Wireless Jamming Model*. Accessed: 2022. [Online]. Available: https://www.nsnam.org

[10] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konečný, S. Mazzocchi, H. Brendan McMahan, T. Van Overveldt, D. Petrou, D. Ramage, and J. Roselander, "Towards federated learning at scale: System design," 2019, *arXiv:1902.01046*.

[11] T. Nishio and R. Yonetani, "Client selection for federated learning with heterogeneous resources in mobile edge," in *Proc. IEEE ICC*, May 2019, pp. 1–7.

[12] N. I. Mowla, I. Doh, and K. Chae, "On-device AI-based cognitive detection of bio-modality spoofing in medical cyber physical system," *IEEE Access*, vol. 7, pp. 2126–2137, 2019.

[13] H. H. Zhuo, W. Feng, Q. Xu, Q. Yang, and Y. Lin, "Federated reinforcement learning," 2019, *arXiv:1901.08277*.

[14] B. Liu, L. Wang, and M. Liu, "Lifelong federated reinforcement learning: A learning architecture for navigation in cloud robotic systems," 2019, *arXiv:1901.06455*.

[15] N. H. Tran, W. Bao, A. Zomaya, M. N. H. Nguyen, and C. S. Hong, "Federated learning over wireless networks: Optimization model design and analysis," in *Proc. IEEE Conf. Comput. Commun.*, Apr. 2019, pp. 1387–1395.

[16] G. Noubir, "On connectivity in ad hoc networks under jamming using directional antennas and mobility," in *Proc. IFIP WWIC*. Berlin, Germany: Springer, 2020, pp. 186–200.

[17] S. Bhunia, P. A. Regis, and S. Sengupta, "Distributed adaptive beam nulling to survive against jamming in 3D UAV mesh networks," *Comput. Netw.*, vol. 137, pp. 83–97, Jun. 2018.

[18] M. Gharibi, R. Boutaba, and S. L. Waslander, "Internet of Drones," *IEEE Access*, vol. 4, pp. 1148–1162, 2016.

[19] A. Chriki, H. Touati, H. Snoussi, and F. Kamoun, "FANET: Communication, mobility models and security issues," *Comput. Netw.*, vol. 163, Nov. 2019, Art. no. 106877.

[20] M. Rothmann and M. Porrmann, "A survey of domain-specific architectures for reinforcement learning," *IEEE Access*, vol. 10, pp. 13753–13767, 2022, doi: 10.1109/ACCESS.2022.3146518.

[21] K.-D. Lu and Z.-G. Wu, "Genetic algorithm-based cumulative sum method for jamming attack detection of cyber-physical power systems," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022.

[22] A. Kardi and R. Zagrouba, "Hybrid cryptography algorithm for secure data communication in WSNs: DECRSA," in *Proc. Congr. Intell. Syst.*, vol. 1334, H. Sharma, M. Saraswat, A. Yadav, J. H. Kim, and J. C. Bansal, Eds. Singapore: Springer, 2020, pp. 643–657.

[23] Y. Yang, W. Wang, R. Xu, G. Srivastava, M. Alazab, T. R. Gadekallu, and C. Su, "AoI optimization for UAV-aided MEC networks under channel access attacks: A game theoretic viewpoint," in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 1–6.

[24] N. I. Mowla, N. H. Tran, I. Doh, and K. Chae, "AFRL: Adaptive federated reinforcement learning for intelligent jamming defense in FANET," *J. Commun. Netw.*, vol. 22, no. 3, pp. 244–258, Jun. 2020.

[25] W. Wang, Z. Lv, X. Lu, Y. Zhang, and L. Xiao, "Distributed reinforcement learning based framework for energy-efficient UAV relay against jamming," *Intell. Converged Netw.*, vol. 2, no. 2, pp. 150–162, Jun. 2021.

[26] C. Sun, W. Liu, and L. Dong, "Reinforcement learning with task decomposition for cooperative multiagent systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 2054–2065, May 2021.

[27] K. Cobbe, O. Klimov, C. Hesse, T. Kim, and J. Schulman, "Quantifying generalization in reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 1282–1289.

[28] E. Conti, V. Madhavan, F. P. Such, J. Lehman, K. Stanley, and J. Clune, "Improving exploration in evolution strategies for deep reinforcement learning via a population of novelty-seeking agents," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 5027–5038.

[29] G. S. Chadha, E. Meydani, and A. Schwung, "Regularizing neural networks with gradient monitoring," in *Proc. INNS Big Data Deep Learn. Conf.*, 2019, pp. 196–205.

[30] A. N. Gomez, I. Zhang, S. R. Kamalakara, D. Madaan, K. Swersky, Y. Gal, and G. E. Hinton, "Learning sparse networks using targeted dropout," 2019, *arXiv:1905.13678*.

[31] M. Belkin, D. Hsu, S. Ma, and S. Mandal, "Reconciling modern machine-learning practice and the classical bias–variance trade-off," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 32, pp. 15849–15854, Aug. 2019.

[32] A. Brutzkus and A. Globerson, "Why do larger models generalize better? A theoretical perspective via the XOR problem," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 822–830.

[33] J. Zhang, T. He, S. Sra, and A. Jadbabaie, "Why gradient clipping accelerates training: A theoretical justification for adaptivity," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–21.

[34] N. Lee, T. Ajanthan, and P. Torr, "SNIP: Single-shot network pruning based on connection sensitivity," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–15.

[35] N. Lee, T. Ajanthan, S. Gould, and P. H. S. Torr, "A signal propagation perspective for pruning neural networks at initialization," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–16.

[36] D. Blalock, J. J. G. Ortiz, J. Frankle, and J. Guttag, "What is the state of neural network pruning?" *Proc. Mach. Learn. Syst.*, vol. 2, pp. 129–146, Mar. 2020.

[37] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, ''A tutorial on UAVs for wireless networks: Applications, challenges, and open problems,'' *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2334–2360, 3rd Quart., 2019.

[38] A. Tahir, J. Boling, M. H. Haghbayan. H. T. Toivonen, and J. Plosila, ''Swarms of unmanned aerial vehicles—A survey,'' *J. Ind. Inf. Integr.*, vol. 16, Dec. 2019, Art. no. 100106.

[39] M. S. A. Hameed, G. S. Chadha, A. Schwung, and S. X. Ding, ''Gradient monitored reinforcement learning,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 8, pp. 4106–4119, Aug. 2023, doi: 10.1109/TNNLS.2021.3119853.

[40] D. Wang and M. Hu, ''Deep deterministic policy gradient with compatible critic network,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 8, pp. 4332–4344, Aug. 2023, doi: 10.1109/TNNLS.2021.3117790.

[41] S. Ren, D. Li, Q. Hu, Y. Liu, and J. Liu, ''An improved security OLSR protocol against black hole attack based on FANET,'' in *Proc. 13th Asian Control Conf. (ASCC)*, 2022, pp. 383–388, doi: 10.23919/ASCC56756.2022.9828257.

[42] J. Liu, Q. Wang, C. He, K. Jaffres-Runser, Y. Xu, Z. Li, and Y. Xu, ''QMR: Q-learning based multi-objective optimization routing protocol for flying ad hoc networks,'' *Comput. Commun.*, vol. 150, pp. 304–316, Jan. 2020.

[43] B. Mahalakshmi and D. S. R. Kumari, ''An adaptive routing in flying ad-hoc networks using FMCC protocol,'' *Int. J. Recent Technol. Eng. (IJRTE)*, vol. 8, no. 5, pp. 2473–2480, Jan. 2020, doi: 10.35940/ijrte.E5782.018520.

[44] A. Rovira-Sugranes, F. Afghah, J. Qu, and A. Razi, ''Fully-echoed Q-routing with simulated annealing inference for flying adhoc networks,'' *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 3, pp. 2223–2234, Jul. 2021, doi: 10.1109/TNSE.2021.3085514.

[45] Y. Cui, Q. Zhang, Z. Feng, Z. Wei, C. Shi, and H. Yang, ''Topology-aware resilient routing protocol for FANETs: An adaptive Q-learning approach,'' *IEEE Internet Things J.*, vol. 9, no. 19, pp. 18632–18649, Oct. 2022, doi: 10.1109/JIOT.2022.3162849.

[46] M. Kumar and R. S. Raw, ''A novel routing protocol for hierarchical software defined vehicular adhoc network,'' in *Proc. 9th Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2022, pp. 771–775, doi: 10.23919/INDIACom54597.2022.9763267.

[47] P. K. Pattnaik, B. K. Panda, and M. Sain, ''Design of novel mobility and obstacle-aware algorithm for optimal MANET routing,'' *IEEE Access*, vol. 9, pp. 110648–110657, 2021, doi: 10.1109/ACCESS.2021.3101850.

[48] J. Liu, F. Ding, and D. Zhang, ''A hierarchical failure detector based on architecture in VANETs,'' *IEEE Access*, vol. 7, pp. 152813–152820, 2019, doi: 10.1109/ACCESS.2019.2948599.

[49] C. Pu and P. Zhu, ''Defending against flooding attacks in the Internet of Drones environment,'' in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2021, pp. 1–6.

[50] H. Abulkasim, B. Goncalves, A. Mashatan, and S. Ghose, ''Authenticated secure quantum-based communication scheme in Internet-of-Drones deployment,'' *IEEE Access*, vol. 10, pp. 94963–94972, 2022.

[51] G. Secinti, P. B. Darian, B. Canberk, and K. R. Chowdhury, ''Resilient end-to-end connectivity for software defined unmanned aerial vehicular networks,'' in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Montreal, QC, Canada, Oct. 2017, pp. 1–5, doi: 10.1109/PIMRC.2017.8292772.

[52] C. Pu, I. Ahmed, E. Allen, and K.-K.-R. Choo, ''A stochastic packet forwarding algorithm in flying ad hoc networks: Design, analysis, and evaluation,'' *IEEE Access*, vol. 9, pp. 162614–162632, 2021.

[53] A. Azhari, A. W. Muhammad, and C. F. M. Foozy, ''Machine learning-based distributed denial of service attack detection on intrusion detection system regarding to feature selection,'' *Int. J. Artif. Intell. Res.*, vol. 4, no. 1, pp. 1–8, Feb. 2020.

[54] N. Gao, Z. Qin, X. Jing, Q. Ni, and S. Jin, ''Anti-intelligent UAV jamming strategy via deep Q-networks,'' *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 569–581, Jan. 2020.

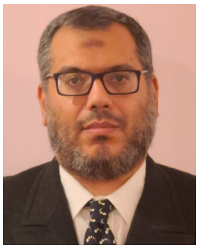
[55] X. Chen, M. W. Ulmer, and B. W. Thomas, ''Deep Q-learning for same-day delivery with vehicles and drones,'' *Eur. J. Oper. Res.*, vol. 298, no. 3, pp. 939–952, May 2022.

[56] J. Patel and H. El-Ocla, ''Energy efficient routing protocol in sensor networks using genetic algorithm,'' *Sensors*, vol. 21, no. 21, p. 7060, Oct. 2021, doi: 10.3390/s21217060.

[57] K. Grover, A. Lim, and Q. Yang, ''Jamming and anti-jamming techniques in wireless networks: A survey,'' *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 17, no. 4, pp. 197–215, Dec. 2014.

JAIMIN GHELANI received the Bachelor of Engineering degree in information technology from Gujarat Technological University, India, in 2019. He is currently pursuing the M.Sc. degree with Lakehead University, Canada. His research interests include wireless networks, routing, and network security.

PRAYAGRAJ GHARIA received the Bachelor of Engineering degree in information technology from Gujarat Technological University, India, in 2019. He is currently pursuing the M.Sc. degree with Lakehead University, Canada. His research interests include wireless networks, routing, and network security.



HOSAM EL-OCLA (Senior Member, IEEE) received the M.Sc. degree from the Department of Electrical Engineering, Cairo University, in 1996, and the Ph.D. degree from Kyushu University, in 2001. He joined the Graduate School of Information Science and Electrical Engineering, Kyushu University, as a Research Student, in 1997. He joined Lakehead University as an Assistant Professor, in 2001, where he has been an Associate Professor, since 2007. He has more than 100 publications in international journals and conferences. His current research interests include networks performance and security in wireless sensor and mobile networks, neural networks, and the Internet of Things. He is a referee and an editor of several journals and conferences.