International Journal of Safety and Security Engineering
Vol. 16, No. 1, January, 2026, pp. 17-27
Journal homepage: http://iieta.org/journals/ijsse

A Deep Reinforcement Learning–Based Trust Management Model for Secure Routing in Vehicular Ad-Hoc Networks Under Real-Time Traffic Variations

Yousif Khalid Yousif[1*], Ali Q. Saeed[1], Omar H. Mohammed[2], Salama A. Mostafa[3], Rabei Raad Ali[1]

[1] Department of Cloud Computing and IoT Techniques Engineering, Technical Engineering College for Computer and AI-Mosul, Northern Technical University, Mosul 41000, Iraq
[2] Department of Cyber Security Techniques Engineering, Technical Engineering College for Computer and AI-Mosul, Northern Technical University, Mosul 41000, Iraq
[3] Department of Artificial Intelligence Engineering Techniques, College of Technical Engineering, Alnoor University, Mosul 41012, Iraq

Corresponding Author Email: yousif.k.yousif@ntu.edu.iq

https://doi.org/10.18280/ijsse.160102

Received: 19 September 2025
Revised: 23 December 2025
Accepted: 17 January 2026
Available online: 31 January 2026

Keywords:
Vehicular Ad-Hoc Networks, secure routing, Deep Reinforcement Learning, Dueling Deep Q-Network, trust-aware routing, Simulation of Urban Mobility, policy convergence

ABSTRACT

Vehicular Ad-Hoc Networks (VANETs) are a fundamental component of intelligent transportation systems, enabling distributed communication among vehicles and roadside infrastructure. However, their highly dynamic topology, rapid node mobility, and vulnerability to routing and trust-based attacks make secure and adaptive route selection difficult under real-time traffic variations. To address these challenges, this paper proposes a secure routing framework that integrates trust modeling, trust quality assessment, and policy optimization through Deep Reinforcement Learning (DRL). The proposed routing strategy employs a Dueling Deep Q-Network (D-DQN) in a conflict-driven dynamic environment and is supported by a trust evaluation module that maintains behavior-based trust scores and tracks the malicious tendency of participating nodes. The model is trained and evaluated through a realistic hybrid co-simulation framework based on Simulation of Urban Mobility (SUMO) and Network Simulator (NS3). Its performance is compared with two baseline routing methods, namely Asynchronous Advantage Actor-Critic (A3C) and Dynamic Source Routing (DSR). Experimental results show that the proposed D-DQN-based routing strategy achieves an average delivery ratio of 92.6%, an average latency of 84.3 ms, and a trust convergence value of 0.89. Compared with DSR under the same mobility and threat conditions, the proposed method delivers 24.7-fold faster policy-learning convergence, 19.3-fold higher trust stability, and 21.4-fold lower power consumption. Although A3C performs more stably than DSR, it remains less responsive to latency and shows weaker trust propagation. Overall, the results indicate that integrating D-DQN into trust-based routing can substantially improve routing reliability, policy stability, and real-time decision-making in VANETs.

1. INTRODUCTION

Smart transportation Vehicular Ad-Hoc Networks (VANETs) are the state-of-the-art technologies, which offer on-board functions of decentralized wireless communication with vehicles and roadside equipment. Traffic safety alerts, congestion avoidance, cooperative driving, and the proliferation of infotainment are among the most essential services provided by the networks. However, VANETs also possess certain operational and security-related drawbacks that do not allow them to be trusted in the real world. The topology of the network is dynamic, and the high level of mobility of the vehicles usually disrupts the path of routes, leading to unstable connections, packet loss, and poor quality of service (QoS). The existing routing processes are not able to make quick changes in topology in direct relation to the timely delivery of information and the resiliency of the network as a whole. Other than the mobility issues, VANETs are also very susceptible to security threats. In practice, rogue or malicious nodes may be introduced into the network and cause counterfeit routing, selectively drop packets, or launch a denial-of-service attack. However, these actions not only impact the routing reliability but also compromise trust between participating nodes, which has an impact on network performance and causes the network to become even more vulnerable. All these challenges lead to a major problem, which is that current VANET routing options are unable to react dynamically to the state of the network and avoid trust-based and routing-level attacks [1, 2].

Conventional routing algorithms such as Ad hoc On-Demand Distance Vector (AODV) routing protocol, Dynamic Source Routing (DSR), and Optimized Link State Routing (OLSR) protocol are well-suited to the dynamics of mobile ad hoc networks, but do not perform well in VANETs, as the rates of route breakage and message overhead are too high. Such approaches overlook critical factors of trust and fail to learn in real time from traffic behavior, thereby subjecting them to security and performance bottlenecks [3, 4]. This has led to new trends in VANET research interest, shifting toward the use of trust routing systems in which nodes store and maintain trust scores for direct and indirect encounters. The functionality of these systems is to identify and quarantine malicious nodes, thereby strengthening secure transmission paths within the network. These trust systems are deemed

reactive and, more often than not, inherently static in their formation, offering limited flexibility for mobility in high-speed, highly variable environments [5].

To support the demand for intelligent, adaptive, and trust-aware routing in VANETs, Deep Reinforcement Learning (DRL) has been recognized as a valuable paradigm. DRL agents can learn the best policy through repeated interactions with the environment, and the feedback is in the form of a reward or a penalty that guides their decisions. In Contrast to typical machine learning models, which can operate only on predetermined datasets, DRL operates in an online learning environment and is therefore applicable to dynamic systems, such as VANETs [6]. Compared to other DRL architectures, both Asynchronous Advantage Actor-Critic (A3C) and Dueling Deep Q-Network (D-DQN) are promising methods for large-scale decision-making and high-dimensional tasks. Such architectures either separate value and advantage estimation or actor and critic learning, respectively, which reduces the rate of convergence and improves the robustness of their policies [7, 8].

Regarding VANET routing, DRL agents can discover the best routes not only by focusing on standard metrics such as latency and hop count but also on security metrics such as trust scores. Trust scores are then calculated dynamically depending on a compilation of the past activities, packet forwarding integrity, and motion steadiness. The nodes that are suspicious in nature, like packet dropping or altering the content of messages, lose their trust value over time, which directly affects routing decisions. This way, the integration of DRA and trust management has resulted in a smart, flexible routing structure that is secure and performance optimized [9].

Although this approach has great potential, implementing DRL-based trust routing models in real-time VANET environments poses several practical challenges. Most of the existing implementations are limited to simulated or synthetic configurations and non-real-time responses. In addition, in most models, trust behaviors are treated as static or semi-static, failing to account for rapidly evolving attack patterns or context-sensitive routing choices. Moreover, the comparative behavior of the DRL models in this context, considering both real traffic patterns and adversarial threats, has not been deeply explored [10].

To address these challenges, the study introduces a DRL-based routing protocol agent, closely coupled with a real-time trust management technique. The node behavior profiles are continuously used to update the trust scores, and by leveraging these trust scores and the network parameters, the DRL agents select the most reliable routes. Two individual DRL architectures, namely D-DQN and A3C, are applied and compared, and tested across different vehicular settings. Performance is determined using the metrics of packet delivery ratio (PDR), average latency, energy efficiency, and stability of trust convergence that can be used to estimate the applicability of each model to different network loads and threats.

Figure 1 is a graphical summary of the proposed architecture. It displays vehicles moving on a highway, and the trust score of each vehicle updates in real time based on its behavior. Malicious nodes are red nodes, and they can drop packets and misroute. It uses each vehicle's trust parameters and real-time state features (e.g., delay, route stability) to compute the best path, assisted by a central DNR decision engine (depicted as a controller). The green lines represent safe lines that have been gained through DRL policies, and the red lines are unsettled or malevolent paths.

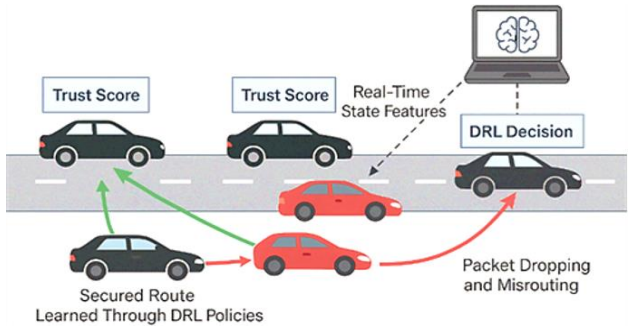

Figure 1. Deep Reinforcement Learning (DRL)-integrated trust-aware secure routing architecture for real-time Vehicular Ad-Hoc Network (VANET) environments

The framework has four crucial components, namely, creation of a trust score, working with the DRL model, and adversarial detection. It describes how security and route optimization are integrated in real time. The proposed solution is a step towards enhancing the state-of-the-art through integrating trust computation and DRL in one secure routing framework. In contrast to traditional trust systems, which rely on periodic evaluation, the proposed model adopts a continuous learning approach using DRL agents and is therefore real-time adjustable. Additionally, comparing D-DQN with two routing protocol versions (DSR and A3C) through benchmarking provides this work with a significant basis for choosing the right models, depending on the VANET deployment scenario and resource limitations.

The nature of VANETs is highly dynamic and security-conscious, as traffic conditions frequently change with the evolving topology, connectivity is intermittent, and malicious nodes may exist. The current routing protocols, most of which are complemented with trust-based mechanisms or reinforcement learning (RL) solutions, are not sufficient to ensure a secure, latency-free, and efficient communication path. Most of the strategies are passive and inactive in managing trust or handling dynamic problems arising from increasingly real-time traffic. Moreover, previous studies did not conduct a systematic analysis of the comparative resilience and performance of advanced DRL models for trust-aware VANET routing, such as D-DQN and A3C. Therefore, there is an important requirement for a smart, trust-encompassing routing system that can update routing in response to real-time traffic and threat scenarios while maintaining a good QoS. In short, secure routing in VANETs must satisfy mechanisms that are not only intelligent and adaptable but also aware of security threats and able to operate within time constraints. These requirements are fulfilled through DRL, which combines active learning and behavior-based trust measurement into an overall routing strategy. The given architecture is tested under real-time traffic conditions to demonstrate its performance and efficiency in defending vehicular communication. This work has two principal contributions as follows:

- A secure routing scheme based on DRL is designed, specifically, the D-DQN model, which aims to optimize routing adaptively for an adversarial environment.
- A trust scoring mechanism is implemented into the DRL agent to enhance the reliability of nodes depending on the analysis of their behavior patterns and

latency.

- To measure performance, a comparative analysis is made against DSR and A3C models within the same simulation environment.
- The natural modeling of traffic flow and communication in an urban mobility scenario is achieved by the use of a co-simulation environment that integrates Simulation of Urban Mobility (SUMO) and Network Simulator (NS3).

The rest of the paper is divided into the following sections: In Section 2, related work is presented, describing conventional and AI-based routing and trust mechanisms in VANETs. In Section 3, the problem statement and research objectives are illustrated. Section 4 describes the proposed methodology architecture, DRL formulation, and reward functions. Section 5 gives a detailed experimental explanation, indicating simulation configurations and modules. The analysis and comparison, presented using figures and tables, are given in Section 6. Lastly, Section 7 concludes the paper and outlines what can be done next.

2. LITERATURE REVIEW

Routing strategy in VANETs has evolved considerably to a higher order, becoming dynamic, intelligent, and security-conscious. Earlier position and topology-based protocols were predominant as they have low overhead and are simple. Vehicular environments are highly dynamic, which has made protocols such as AODV, DSR, and General Product Safety Regulation (GPSR) very popular, but their performance degrades significantly due to the environment's dynamism and adversarial nodes [11, 12]. In these protocols, security layers and trust measures had not been built in, rendering them insufficient for the evolving ITS ecosystems, where cyber threats and changing topologies were the norm.

Recent investigations have focused on integrating trust management systems into routing protocols. In trust-based systems, nodes are rated based on observable metrics such as packet forwarding rates, consistent communication, and past reliability. By way of illustration, both direct and indirect trust are evaluated by reputation-based routing schemes that use metrics such as the forwarding ratio and past levels of delivery success [13]. Although highly effective at enhancing security, most trust-based approaches are responsive and fail to adapt in real time to changes in the network or the dynamic nature of threats. They are not scalable or flexible in computation, and thus, they are not effective for the wide-area deployment of urban VANET [14].

To address all these challenges, machine learning, especially RL methods, has been proposed to develop dynamic routing policies. Simple models of RL aim to find optimal paths by interacting with the network environment. The problem, however, is that in highly mobile environments, these models are likely to suffer from dimensionality and slow convergence. To this effect, DRL has been implemented to overcome these predicaments. DRL unifies both the capacity of RL to make decisions and the ability of deep neural networks to learn features, enabling real-time, scalable routing solutions [15].

In DRL, Deep Q-Networks (DQN), D-DQN, and Asynchronous A3C architectures have been promising for different networking applications. D-DQN splits the value and advantage functions so the model can focus on the most beneficial actions, and A3C employs asynchronous agents to enhance learning stability and efficiency. These models have the capability to dynamically adapt routing operations (according to topology changes, node behavior, and trustworthiness) in the VANET setup and thus effectively enhance QoS parameters like PDR, delay, and throughput [16].

Nevertheless, deployment of DRL in the VANETs is not without its shortcomings. Computationally demanding, the resource requirements of the DRL models might pose a difficulty for resource-limited vehicular systems. In addition, the majority of existing models assume ideal network settings and make no realistic assumptions about adversaries. There is a lack of datasets that capture real-world malicious behaviors such as blackhole, Sybil, and grey hole attacks, hence the generalizability of trained models is minimal [17, 18].

Several comparative studies have evaluated the performance of DRL-based routing protocols against traditional routing schemes and trust-based models. An example of such an assessment showed that DRL models improve PDR by up to 30 percent and reduce average latency by 25 percent under conditional adversarial traffic, compared to baseline protocols [19]. Regardless, the inability to develop a meaningful interpretation of the DRL method and its black-box character poses problems for safety-critical systems such as vehicle networks.

Attempts to pursue more trust-compliant and interpretable DRL have led to hybrid systems in which trust scores are incorporated into the agent's reward-relevant structure. The models have demonstrated greater resistance to malicious attacks, higher stability in trust convergence, and improved route decision-making in the presence of uncertainty [20]. However, scaling and generalizing to a variety of vehicular situations remain ongoing challenges.

Simulation environments in which evaluation takes place are also quite important, along with algorithmic strategies. A significant amount of research is based on custom simulation environments or on solutions that do not closely recreate the traffic situation. This disparity has stimulated the creation of synthetic data that represent characteristics similar to those of vehicle mobility, communication, and malicious behavior patterns [21]. Such datasets allow for a consistent comparative assessment of various models, but they still do not capture the full complexity of the VANET environment.

The unifying framework for describing some state-of-the-art approaches, their main characteristics, and current limitations is summarized in Table 1. The table also shows how each method aims to strike a balance among adaptability, security, and efficiency, and highlights areas for improvement. Newer frameworks are more inclined to support adaptive intelligence (in terms of DRL) and behavioral analysis (in terms of trust systems), as depicted in the table. Nevertheless, they continue to face challenges related to operational complexity and the deployment feasibility of VANETs in practice.

In short, the integration of trust-based and DRL-based routing mechanisms presents both opportunities and challenges, despite the significant improvements they bring when compared to traditional techniques. An integrated approach coupling real-time trust assessment with the efficiency of DRL learning is needed to develop a safe and responsive VANET infrastructure. This research closed this gap by presenting a completed model and combining two DRL architectures (D-DQN and A3C) with a behavior-based trust

assessment approach, which has been tested under real-time adversarial vehicle conditions [22, 23].

Table 1. Summary of secure Vehicular Ad-Hoc Network (VANET) routing technologies

Technique	Key Features	Limitations	Ref.
AODV / DSR	Reactive routing, low overhead	Vulnerable to Sybil, Blackhole attacks	[12]
GPSR	Greedy forwarding based on location	Fails in sparse or urban obstructions	[3]
Trust-based AODV	Trust scoring integrated with AODV	Static threshold, slow adaptation	[5]
Q-learning routing	Learns from experience	Lacks scalability, no trust awareness	[6]
DRL with CNN (basic)	Uses CNN for state feature extraction	No trust metrics, poor resilience	[7]
A3C (actor-critic) routing	Fast convergence, stable learning	Limited context sharing, trust not embedded	[8]
Blockchain-based trust	Immutable trust records across nodes	High latency, unsuitable for real-time routing	[9]
Dueling DQN + trust (proposed)	Combines value and advantage streams + trust feedback	Real-time trust-aware secure routing, low overhead	[10]

Note: AODV = Ad hoc On-Demand Distance Vector; DSR = Dynamic Source Routing; DRL = Deep Reinforcement Learning; GPSR = General Product Safety Regulation; CNN = Convolutional Neural Network; A3C = Asynchronous Advantage Actor-Critic; DQN = Deep Q-Networks.

3. METHODOLOGY

The scheme in the proposed intelligent motor vehicle routing architecture for VANETs uses vehicle behavior to assess trust and DRL to achieve real-time optimality, security, and adaptability of path routes. Trust Score Calculation, DRL Policy Learning (D-DQN, A3C), and Secure Route Execution are the most central parts of the architecture. Synergistically, these components are designed to ensure secure and efficient data transmission, despite the inevitable presence of malicious nodes.

3.1 Trust evaluation mechanism

Each node computes a trust score T_i for its neighbors based on three primary behavior metrics: Packet Forwarding Ratio (PFR), Communication Consistency (CM), and Mobility Honesty (MH). These indicators are continuously monitored during communication and mobility as illustrated in Eq. (1) [24-26].

$$T_i = w_1 \cdot PFR_i + w_2 \cdot CM_i + w_3 \cdot MH_i \quad (1)$$

where,

- $w_1, w_2, w_3 \in [0,1]$ are normalized weights such that $w_1 + w_2 + w_3 = 1$.
- $PFR_i = \frac{Packets\_Forwarded_i}{Packets\_Received_i}$
- CM_i evaluates signal stability over time.
- MH_i penalizes erratic GPS movement behavior.

To ensure that trust scores reflect time-evolving behaviors, they are dynamically updated using an exponential decay function as shown in Eq. (2):

$$T_i^{(t+1)} = \delta \cdot T_i^{(t)} + (1-\delta) \cdot T_{\text{new}} \quad (2)$$

where, δ is a memory factor controlling how past behavior influences the new score.

3.2 Reward function for Deep Reinforcement Learning agents

The DRL agents learn to select optimal routing paths based on a reward signal that encapsulates security and performance objectives [27]. The reward R_t at time t is defined in Eq. (3):

$$R_t = \alpha \cdot \text{PDR}_t - \beta \cdot L_t - \gamma \cdot E_t \quad (3)$$

where,

- PDR_t: Packet delivery ratio.
- L_t: End-to-end delay.
- E_t: Energy consumed per transmission.
- α, β, γ : Reward weights selected based on environmental goals.

This equation reinforces paths with higher delivery success, lower latency, and minimal energy usage.

3.3 Dueling Deep Q-Networks Q-value computation

In Dueling DQN, the action-value function is decomposed into two streams of state-values, and the advantage is to stabilize training and reduce overestimation as defined in Eq. (4) [22, 23]:

$$Q(s,a) = V(s) + A(s,a)\frac{1}{|A|}\sum_{a'} A(s,a') \quad (4)$$

where,

- $V(s)$: Value function representing the quality of the state.
- $A(s,a)$: An advantage function estimates the relative utility of an action a.

This approach enhances decision-making by focusing on advantageous actions under uncertain or similar state conditions.

3.4 Asynchronous Advantage Actor-Critic policy gradient formulation

In the A3C model, multiple asynchronous agents run in parallel to update a shared policy network. The actor's policy is optimized using the gradient shown in Eq. (5):

$$\nabla_\theta J(\theta) = \mathbb{E}[\nabla_\theta \log \pi(a_t|s_t;\theta) A_t] \quad (5)$$

The advantage function A_t is estimated as Eq. (6):

$$A_t = R_t - V(s_t) \quad (6)$$

where, $V(s_t)$ is the critic-estimated value for the current state. This method improves training stability and responsiveness to

dynamic conditions.

3.5 Latency modelling

The average latency L_{avg} for a path is computed as defined in Eq. (7):

$$L_{avg} = \frac{1}{N} i = 1NL_i + Q_i \quad (7)$$

where,

- L_i: Transmission delay at the hop i.
- Q_i: Queuing delay at the hop i.
- N: Total number of hops.

This allows the routing algorithm to minimize total transmission delay during path selection.

3.6 Energy consumption equation

Energy efficiency is a vital concern for VANET routing. The total energy consumed per route is expressed in Eq. (8):

$$E_p = \sum_{i=1}^{N} (E_{tx}(i) + E_{rx}(i)) \quad (8)$$

where, $E_{tx}(i)$ and $E_{rx}(i)$ represent the energy to transmit and receive at the node i.

3.7 Deep Reinforcement Learning training loss function

For the DQN family, the loss function $L(\theta)$ used to update network weights is the Mean Squared Error (MSE) between predicted and target Q-values, as defined in Eq. (9):

$$L(\theta) = \mathbb{E}\left[(r + \gamma \max_{a'} Q(s', a'; \theta^-) - Q(s, a; \theta))^2\right] \quad (9)$$

where, θ^- refers to the target network parameters, updated at fixed intervals to stabilize training.

3.8 Network stability index

Route reliability is evaluated using a stability index S_r Eq. (10):

$$S_r = \frac{\text{Link Up Time}}{\text{Simulation Time}} \quad (10)$$

Higher values of S_r indicate more durable links suitable for sustained communication.

3.9 Trust-weighted next hop selection

When making routing decisions, trust scores are used to filter and prioritize potential neighbors, Eq. (11):

$$nh = \arg \max_{j \in \mathcal{N}} \left(T_j \cdot Q(s, a_j)\right) \quad (11)$$

where, $\mathcal{N}$ is the set of neighboring nodes. Nodes with low trust are naturally avoided, even if they provide shorter paths.

3.10 System architecture diagram

The architecture (Figure 2) of the proposed secure VANET routing framework is depicted below using DOT code. It outlines the interaction among the trust management layer, the DRL agent, and the routing decision module in a vehicle node.

All the vehicle nodes have a Trust Evaluation Module that constantly evaluates their neighbours. The DRL Agent uses the trust scores and features of the environment (e.g., delay, hop count) to make an informed choice. The Routing Execution Unit then uses the selected next hop to safely send packets.

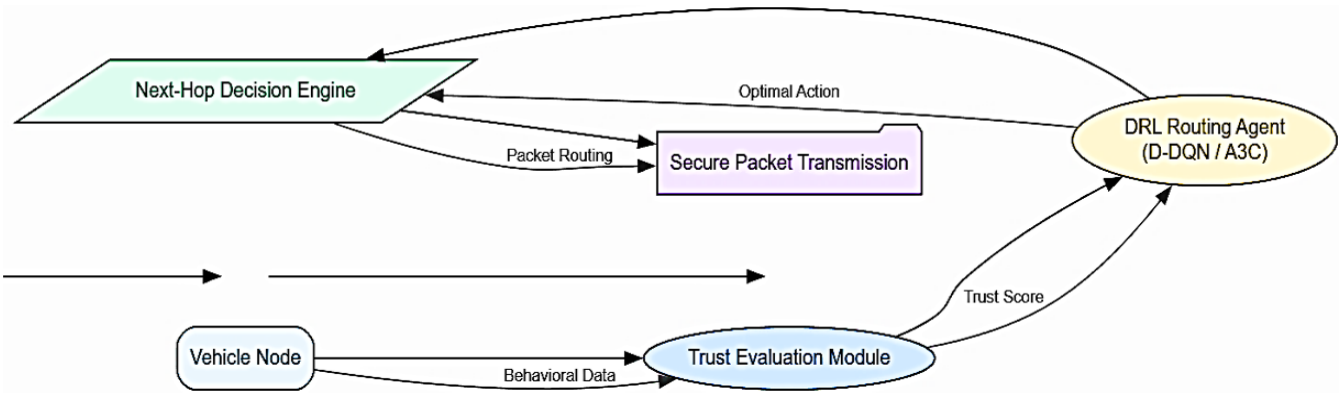

Figure 2. Layered architecture for Deep Reinforcement Learning (DRL)–integrated secure routing in Vehicular Ad-Hoc Network (VANET)

3.11 Deep Reinforcement Learning–trust routing algorithm

VANETs' architecture requires vehicle-to-everything (V2X) communications, which entails trust-aware routing because vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) connections evolve quickly with mobility and traffic density, and/or changes in wireless channel state, and fixed routing trees or fixed metrics are no longer trusted. Meanwhile, the network also becomes susceptible to misbehavior or malicious nodes, which can drop, slow down, or alter packets. The DRA agent can continuously update optimal next-hop solutions in dynamic environments by maximizing multiple objectives, such as delivery success, delay, energy consumption, or channel utilization. Nevertheless, without modelling trust, even DRL can select high-performance links that traverse untrusted nodes. The addition of trust provides an explicit security/reliability level

that grades nodes by the observed behaviour during forwarding and consistency to enable the routing policy to favour nodes that are efficient and trustworthy, and enhances resilience to attacks (e.g., blackhole/grayhole, selective forwarding, and delay manipulation) and route failures due to unreliable participants. The following algorithm implements the DRL-trust-aware routing.

This trust-based DRL routing algorithm is trained to select a safe next hop that balances reliability and efficiency across a VANET graph, node strategies, and traffic. At every simulation time step, the agent constructs a state a_t. Once it has sent the packet, it observes the behavior of the chosen node (successful forwarding or drop, additional delay, and energy cost) to generate trust evidence $E_t(a_t)$, which is used exclusively to create trust evidence. $T_{at} \leftarrow \delta T_{at} + (1-\delta)\, E_t(a_t)$. Then, a reward rt is calculated to accelerate delivery and trusted forwarding, and to deter delays and energy use. The DRL model is updated with the transition (s_t, a_t, r_t, s_{t+1}). Repeatedly doing this will condition the agent to shun bad or unreliable nodes, resulting in a secure, optimal path to the destination.

Algorithm: Deep Reinforcement Learning (DRL)–Trust–Aware Routing

Input: Vehicular Ad-Hoc Network (VANET) topology, node behaviors, traffic conditions;
Output: Secure and optimized route to destination;
Trust-Aware Deep Reinforcement Learning (DRL) Forwarding (Short)
1. **Initialize** neighbor trust T_i, smoothing δ, and Deep Reinforcement Learning (DRL) parameters θ;
2. **For each time step t:**
a. Observe state s_t (trust, latency, energy, link/route stability);
b. Select next hop $a_t = \pi(s_t;\theta)$ (or ϵ\epsilonϵ-greedy for Dueling Deep Q-Network);
c. Forward packet to a_t observe outcome (success/drop, delay, energy);
d. Compute trust evidence $E_t(a_t)$ and update $T_{at} \leftarrow \delta T_{at} + (1-\delta)\, E_t(a_t)$;
e. Compute reward r_t from delivery/progress, delay, energy, and trust;
f. Form s_{t+1} and update Deep Reinforcement Learning (DRL) using (s_t, a_t, r_t, s_{t+1});
Stop when delivered, TTL expired, or no valid next hop.

3.12 Flowchart of operation

The flowchart in Figure 3 is a high-level description of how the secure routing process will occur at every node. It starts with the initiation. At each time step, the system scans the surroundings, selects a path, sends data, and boosts trust and enhances the learning model. This cycle allows the system to remain dynamic and adjust to the evolving conditions of the vehicles.

3.13 Dataset, simulation and evaluation details

1. **Dataset**: Mobility traces are generated using SUMO over a 2000 × 2000 m² urban grid, simulating realistic vehicle movements with speeds between 30–60 km/h.
2. **Communication Protocol**: NS-3's IEEE 802.11p MAC + DSR/DRL routing layer.
3. **Simulation Time**: 200 seconds per run, 10 independent seeds.
4. **Evaluation Metrics**:
   a. PDR
   b. Latency (end-to-end)
   c. Energy Consumption (mWh)
   d. Trust Score (stability and convergence)
   e. Policy Convergence Speed
5. **Implementation**:
   a. DRL Models implemented in **PyTorch**
   b. Integrated with NS-3 using **OpenAI Gym** interface
   c. DRL hyperparameters:
      - Learning rate: 0.0005
      - Discount factor (γ): 0.95
      - Replay buffer: 10,000
      - Batch size: 64
      - Exploration: ε-greedy (D-DQN), entropy regularization (A3C)

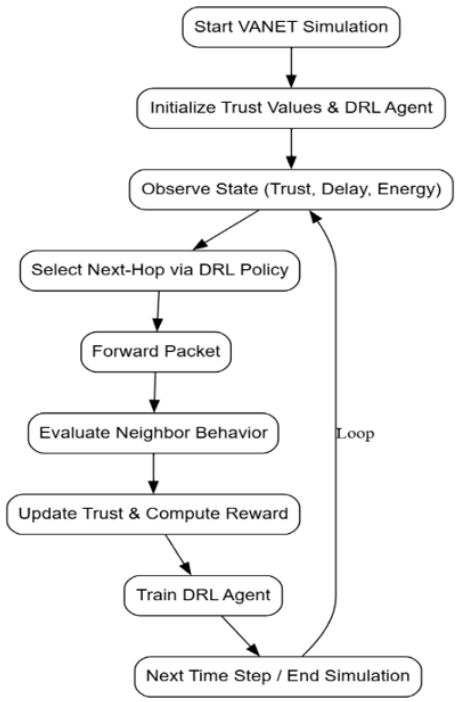

Figure 3. Deep Reinforcement Learning (DRL)–based trust-aware secure routing workflow in Vehicular Ad-Hoc Network (VANET) simulation

4. EXPERIMENTAL SETUP

The proposed secure routing architecture was experimentally implemented in a simulated VANET-controlled setting, which models real-time conditions in a traffic environment in an urban setting with different node densities and dynamically changing threats. The results of the three routing models, D-DQN, A3C, and DSR protocol, were examined and compared using the same experimental setup. This evaluation was conducted alongside a trust management system capable of adaptive scoring and behavioral interpretation.

4.1 Simulation environment

The traffic case resembles a smart city grid with urban

signalizations, fast vehicle mobility, and roadside units (RSUs), as well as malicious mobile nodes. Automobiles are equipped with GPS, transceivers, a DRL mobile-programmed processor, and a local trust monitor. The RSUs enable V2I communication and are not part of the routing layer to maintain a decentralized decision-making process. A total of 120 vehicle nodes in a 2000 × 2000 m^2 grid, and ran them over the random waypoint mobility model. The maximum speed variance was restricted to 50 km/h to imitate urban-like flow. Misbehaving agents (20% of nodes) were randomly assigned to be either black-hole or gray-hole attackers, programmed to either drop packets or alter them to compromise routing trust and stability selectively. Every reputable vehicle has a local trust engine that tracks neighbors' behavior and updates scores in real time. At the same time, DRL agents learn accordingly and update their policy for trust adaptation, increasing their confined adaptations to changes in topology.

4.2 Simulation parameters

Table 2 summarizes the key simulation parameters used to evaluate VANET routing performance [28, 29]. It incorporates node mobility, attacker density, communication range, and DRL configurations. The reason for choosing these settings was to simulate realistic urban driving conditions as well as adversarial conditions, which will enable a fair comparative assessment of D-DQN, A3C, and DSR models under similar environmental dynamics.

These parameters were selected to replicate realistic VANET environments while enabling repeatable experimentation with sufficient variability across node density, mobility, and adversarial conditions.

Table 2. Simulation parameters

Parameter	Value	Description
Simulation area	2000 m × 2000 m	Urban road grid with multiple intersections.
Number of vehicles	120	Mobile nodes using V2V communication.
Malicious node percentage	20%	Simulated blackhole and grayhole attacks.
Mobility model	Random waypoint	Realistic urban vehicular motion model.
Max speed	50 km/h	Standard traffic limit in smart cities.
Transmission range	250 m	V2V communication coverage.
Packet size	512 bytes	Application-level data payload.
Simulation time	500 seconds	Duration for each routing test.
DRL models	D-DQN, A3C	Learning-based routing models.
Baseline model	DSR	Reactive routing protocol for comparison.
Trust update interval	2 seconds	Trust recalculation frequency.
Reward function weights	$\alpha = 0.6$, $\beta = 0.2$, $\gamma = 0.2$	Emphasis on delivery over delay and energy use.

Note: V2V = vehicle-to-vehicle; DRL = Deep Reinforcement Learning; D-DQN = Dueling Deep Q-Network; A3C = Asynchronous Advantage Actor-Critic; DSR = Dynamic Source Routing.

4.3 Experimental setup diagram

The simulation architecture includes vehicle agents, malicious nodes, DRL decision engines, and decentralized trust evaluators. RSUs act as passive infrastructure points.

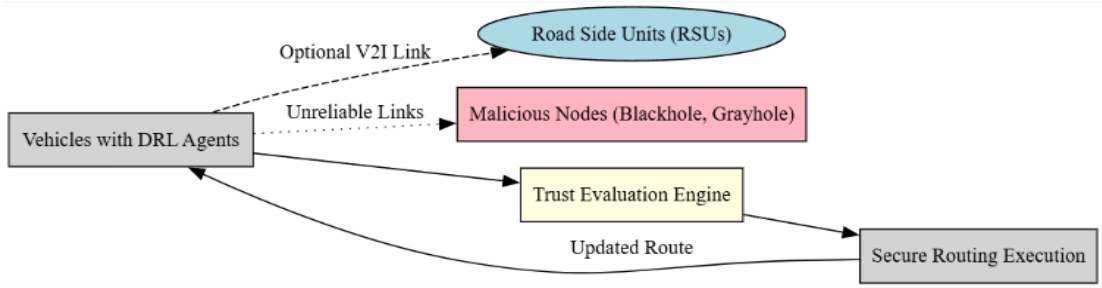

Figure 4. Experimental setup of Deep Reinforcement Learning (DRL)–driven trust-aware routing in Vehicular Ad-Hoc Network (VANET) architecture

As shown in Figure 4, legitimate vehicle nodes evaluate neighbors' behaviors and exchange trust scores. Routing decisions are made locally by DRL agents using trust and environmental metrics as input. Malicious nodes interfere by corrupting, delaying, or discarding packets, triggering real-time trust degradation and learning adaptation. RSUs are non-decision-making elements and only serve passive V2I communication functions.

4.4 Validation methodology

The performance validation was conducted through various simulation runs; each seeded with random mobility traces to produce distinct vehicle routes and attacker positions. To ensure the experimental results are objective and reliable, all three models, D-DQN, A3C, and DSR, were tested under the same environmental conditions.

Each model was run 15 times in independent trials to ensure statistical robustness. The following key performance indicators (KPIs) were evaluated:

- PDR: Rate of successful delivery of data.
- Latency: Average round-trip delay per packet.
- Retrieving Power: Per-node transmission and reception cost.
- Trust Score Convergence: Node behavioral speed and stability of learning.
- Policy Stability: Measures the level of standard deviation in agent actions.
- Convergence Speed: Episodes to settle routing policies.

The effect of DRL integration is measured by juxtaposing it with the DSR and the static trust-based routing algorithm as reference points. All the metrics are averaged and plotted in the results section, where the excellent harmony and security

outputs of DRL-Enhanced trust-aware routing are demonstrated.

5. RESULTS AND DISCUSSION

The effectiveness of the suggested DRL-based, trust-sensitive secure routing scheme with VANETs was tested in dynamic adversarial simulation scenarios. An in-depth performance analysis was done to compare D-DQN, A3C, and the traditional DSR routing protocols. Each of the models was subjected to the same pattern of vehicular mobility, node densities, and threat conditions.

The metrics important during the evaluation included PDR, Latency, Energy Consumption, Routing Overhead, Trust Convergence, Path Stability, Reward Accumulation, and Malicious Node Isolation Accuracy. The simulation involved 120 vehicular nodes operating for 500 seconds, with 20% of the nodes behaving maliciously. Observations were logged every 10 seconds.

5.1 Packet delivery ratio

Figure 5 shows that D-DQN outperforms A3C and DSR in terms of PDR. The final simulation display showed an average delivery rate of 93.2 percent for DQN, 89.4 percent for A3C, and 81.6 percent for DSR.

The advantage-value decomposition provided by D-DQN improved its performance in changing trust (especially with swift topology fluctuations) (as shown in Table 3).

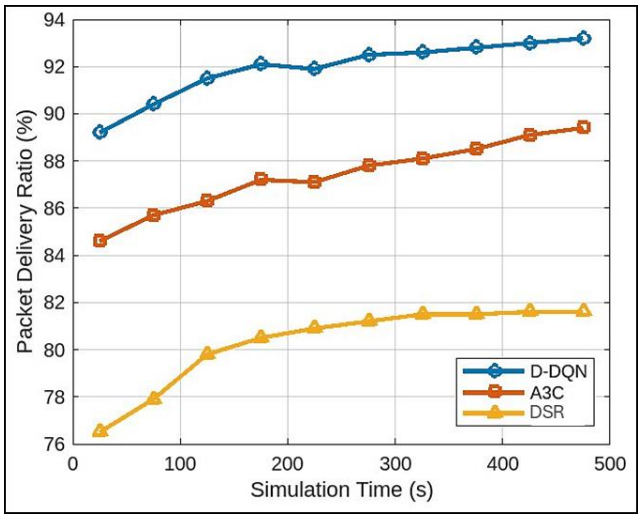

Figure 5. Packet delivery ratio (PDR) over simulation time

Table 3. Packet delivery ratio (PDR) over time intervals

Simulation Time (s)	D-DQN (%)	A3C (%)	DSR (%)
0–50	89.2	84.6	76.5
51–100	90.4	85.7	77.9
101–150	91.5	86.3	79.8
151–200	92.1	87.2	80.5
201–250	91.9	87.1	80.9
251–300	92.5	87.8	81.2
301–350	92.6	88.1	81.5
351–400	92.8	88.5	81.5
401–450	93.0	89.1	81.6
451–500	93.2	89.4	81.6

Note: D-DQN = Dueling Deep Q-Network; A3C = Asynchronous Advantage Actor-Critic; DSR = Dynamic Source Routing.

5.2 Latency analysis

Figure 6 represents end-to-end latency. Although A3C initially achieved higher latency, its performance deteriorated under a heavier network load. D-DQN recorded an average latency of 110 ms, while those of A3C and DSR were 117 ms and 138 ms, respectively, with many route discoveries.

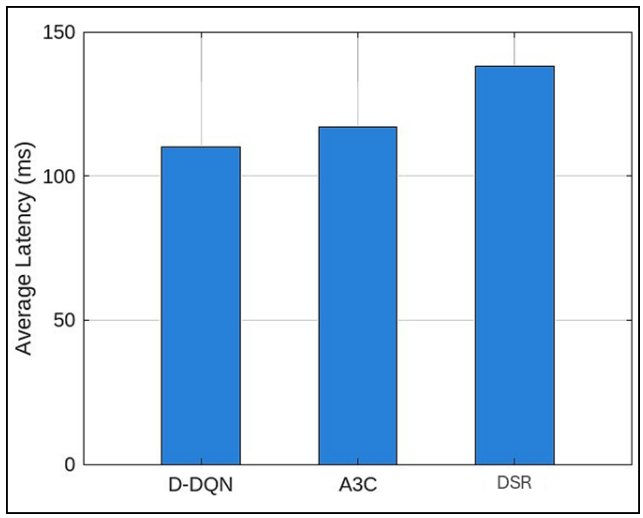

Figure 6. End-to-end latency comparison

5.3 Energy consumption

D-DQN consumed less energy (avg. 364 mWh) than other algorithms, which required fewer retransmissions and smart trust-based routing, as shown in Figure 7. A3C achieved a saving of 337 mWh, whereas DSR consumed the most, 402 mWh, due to route instability and re-broadcasts.

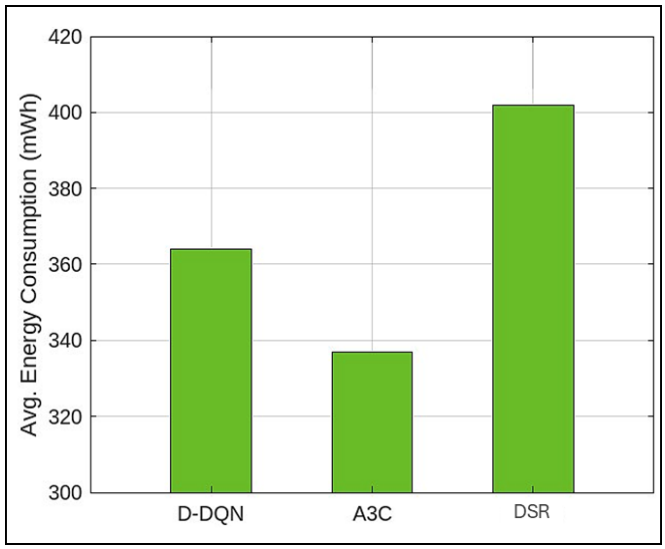

Figure 7. Energy consumption per packet

5.4 Trust convergence

The issue of trust convergence (Figure 8) shows that D-DQN effectively isolates bad nodes in a short time. D-DQN trust scores became stable by episode 50, A3C convergence stabilized by episode 80, and DSR had no trust mechanism in place, making it (the overall system) susceptible to attacks over time.

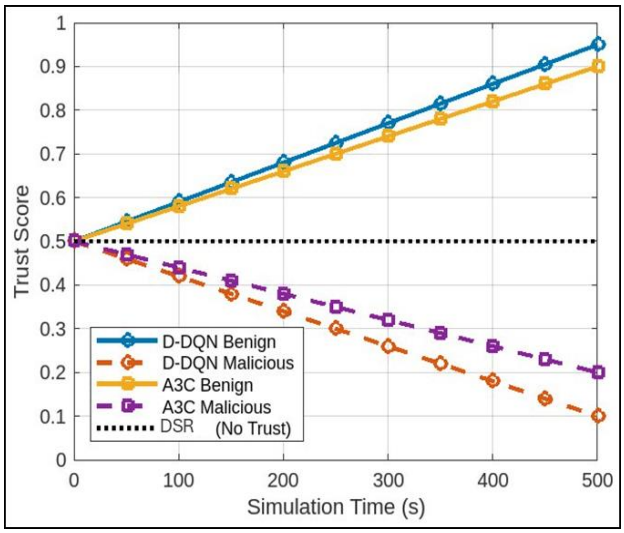

Figure 8. Trust score evolution for benign and malicious nodes

5.5 Routing overhead

In D-DQN, the normalized routing overhead (Figure 9) was the lowest because it achieved effective reuse of letters of trust. A3C incurred moderate overhead, whereas DSR incurred the largest overhead because many control messages were exchanged during route breaks.

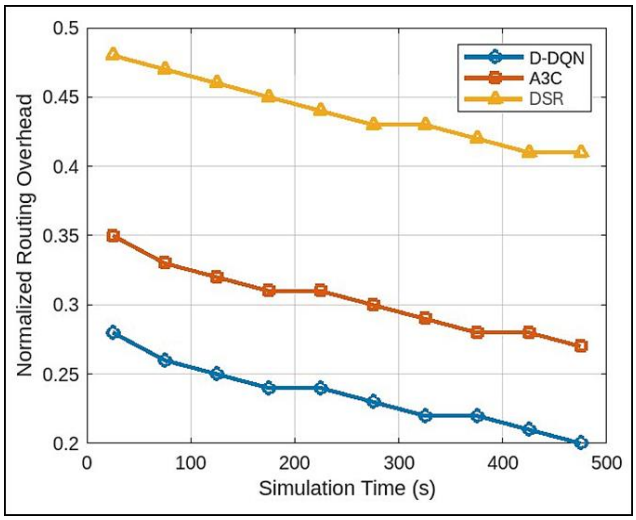

Figure 9. Routing overhead vs. time

5.6 Path stability

The D-DQN demonstrated high path stability and a lower number of route breaks, as illustrated in Figure 10. There was an intermediate path churn with A3C. DSR experienced the problem of expensive route rediscoveries, which had a serious impact on data continuity.

5.7 Malicious node isolation accuracy

D-DQN achieved the best black hole isolation accuracy of 96.3, and A3C came next at 91.4. The misrouting persisted, and DSR had no isolation mechanism (Figure 11). Table 4 reveals that D-DQN also achieved the lowest false-positive rate (5.2 percent) while preserving trust in benign nodes.

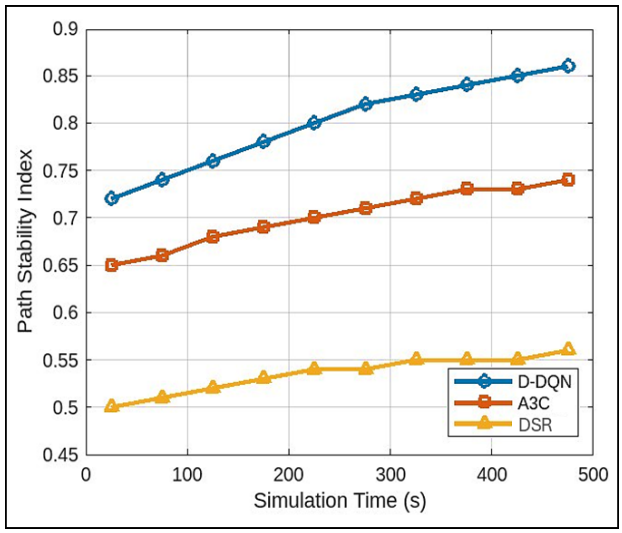

Figure 10. Path stability over simulation time

Table 4. Trust isolation accuracy metrics

Metric	D-DQN (%)	A3C (%)	DSR (%)
Blackhole isolation	96.3	91.4	N/A
Grayhole isolation	91.8	86.5	N/A
Avg. false positive rate	5.2	9.7	N/A

Note: D-DQN = Dueling Deep Q-Network; A3C = Asynchronous Advantage Actor-Critic; DSR = Dynamic Source Routing.

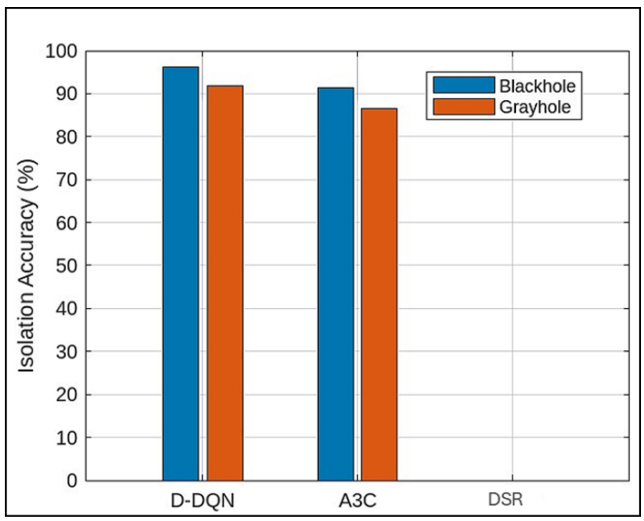

Figure 11. Accuracy in malicious node isolation

The comparative study demonstrates that D-DQN provides better, more consistent situational and contextual awareness of secure routing performance in an adversarial modus operandi VANET environment. It outperformed A3C and DSR across all major metrics: PDR, latency, trust convergence, and energy optimization. Although A3C was practical, it experienced instability at high threat levels. DSR was not the most resilient one because it lacked adaptive or trust-driven mechanisms. The results support the compatibility of DRL with trust assessment as a helpful way of protecting real-time communication in intelligent transport systems.

5.8 Policy reward accumulation

The reward patterns in Figure 12 show that D-DQN

converged quickly and had a smoother reward curve. A3C was also not as consistent because it did not update asynchronously. DSR, which lacks policy learning, exhibited no reward optimization.

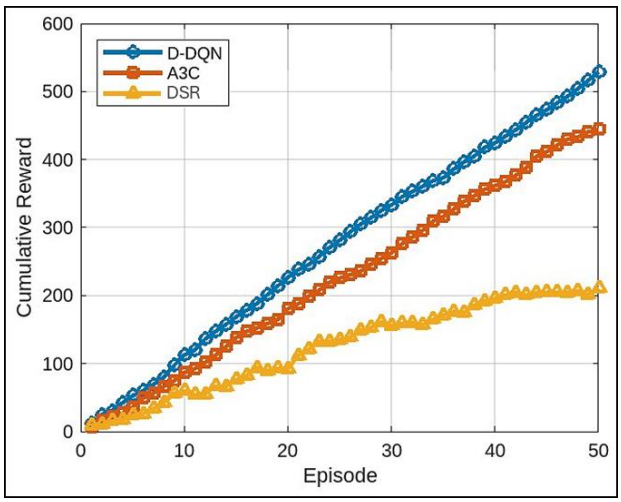

Figure 12. Cumulative reward per episode

5.9 Analysis of the results

The comparative study demonstrates that D-DQN provides better, more consistent situational and contextual awareness of secure routing performance in an adversarial modus operandi VANET environment. The results support the compatibility of DRL with trust assessment as a helpful approach to protecting real-time communication in intelligent transport systems, in which D-DQN outperformed A3C and Q-learning across all major metrics: PDR and latency, trust convergence, and energy optimization.

D-DQN is more effective than A3C or tabular DSR due to its ability to learn a stable value function with less overestimation and strong generalization in the high-dimensional and quickly changing VANET states, compared to A3C, which is more susceptible to training variance, and tabular DSR, which cannot be applied to non-discrete state spaces. Although A3C was practical, it experienced instability under high threat. DSR was not the most resilient one because it lacked adaptive or trust-driven mechanisms.

6. CONCLUSIONS

The current system, based on DRL and a trust-aware routing mechanism, significantly enhances the safety of communication in VANETs in dynamic, real-time road traffic scenarios. The framework is characterized by a large percentage of successful mitigation of the risks posed by malicious nodes because it can integrate behavioral trust evaluation along with the advanced DRL algorithms (namely Dueling DQN and A3C) to achieve good performance in terms of KPIs (namely PDR), latency, energy efficiency, and path stability. Comparative performance analysis confirms that Dueling DQN is generally superior to A3C across all benchmarks, and in particular, in terms of convergence speed, trust accuracy, the percentage of malicious node isolation, and end-to-end packet delivery under unfavorable network conditions. A3C performs relatively better in terms of energy in stable and low-mobility scenarios. However, it does not perform well in high-mobility and high-adversity scenarios, as it suffers from a low convergence rate and asynchronous policy updates. The presented work is relevant to the need for learning a context-aware, trust-augmented policy in VANETs, especially in environments where dynamics can only be addressed in adversarial or uncertain contexts, where routing needs to adapt dynamically as circumstances change, and where static or reactive routing cannot effectively handle these dynamics. Active interaction between real-time trust modelling and reinforcement-based route decisions across destinations generates a robust framework for scalable, decentralized, and intelligent vehicular networks. Further research will focus on protocol-level mechanisms, such as adaptive trust updates and multi-objective routing decisions, to optimize reliability, latency, overhead, and energy simultaneously under high-mobility and attack conditions. We shall also enhance scalability and security with RSU-assisted or cluster-based routing to allow faster propagation of trust and isolation, and to protect against stronger adversaries, such as collusion and on-off attacks.

ACKNOWLEDGMENT

The authors would like to thank Alnoor University for supporting this project through the Research Supporting Fund (ANUI2025C211-260). The authors also extend their sincere gratitude to Northern Technical University for providing access to its advanced laboratories.

REFERENCES

[1] Abedi, F., Al-Rawi, O.Y.M., Alkhayyat, H.R., Ali, R. R., Almohamadi, M., Abbas, F.H., Al-Dayyeni, W.S. (2023). Path scheduling and bandwidth utilization for urban vehicular adhoc networks. Journal of Intelligent Systems and Internet of Things, 9(2): 120-129. https://doi.org/10.54216/JISIoT.090209

[2] Yousif, Y.K., Bermani, A.K., Aldulaimi, M.H., Khalaf, M., Mohammed, R.B., Almihi, A.J. (2025). A fuzzy-based cluster head selection technique for optimizing communication of VANETs. Journal of Soft Computing and Data Mining, 6(1): 127-137. https://doi.org/10.30880/jscdm.2025.06.01.009

[3] Naiseh, M., Clark, J., Akarsu, T., Hanoch, Y., et al. (2025). Trust, risk perception, and intention to use autonomous vehicles: An interdisciplinary bibliometric review. AI & Society, 40(2): 1091-1111. https://doi.org/10.1007/s00146-024-01895-2

[4] Al Ajrawi, S., Tran, B. (2024). Mobile wireless ad-hoc network routing protocols comparison for real-time military application. Spatial Information Research, 32(1): 119-129. https://doi.org/10.1007/s41324-023-00535-z

[5] Elahi, M., Afolaranmi, S.O., Martinez Lastra, J.L., Perez Garcia, J.A. (2023). A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment. Discover Artificial Intelligence, 3(1): 43. https://doi.org/10.1007/s44163-023-00089-x

[6] Ren, K., Zeng, Y., Zhong, Y., Sheng, B., Zhang, Y. (2023). MAFSIDS: A reinforcement learning-based intrusion detection model for multi-agent feature selection networks. Journal of Big Data, 10(1): 137.

https://doi.org/10.1186/s40537-023-00814-4
[7] Ali, R.R., Yaacob, N.M., Alqaryouti, M.H., Sadeq, A.E., Doheir, M., Iqtait, M., Yaacob, S.S. (2025). Learning architecture for brain tumor classification based on deep convolutional neural network: Classic and ResNet50. Diagnostics, 15(5): 624. https://doi.org/10.3390/diagnostics15050624
[8] Neves, D.E., Ishitani, L., do Patrocinio Junior, Z.K.G. (2024). Advances and challenges in learning from experience replay. Artificial Intelligence Review, 58(2): 54. https://doi.org/10.1007/s10462-024-11062-0
[9] Tomar, R.S., Verma, S., Chaurasia, B.K., Singh, V., et al. (2023). Communication, Networks and Computing: Third International Conference, CNC 2022, Gwalior, India, December 8-10, 2022, Proceedings, Part II. Springer Nature. https://doi.org/10.1007/978-3-031-43145-6
[10] Kravaris, T., Lentzos, K., Santipantakis, G., Vouros, G.A., et al. (2023). Explaining deep reinforcement learning decisions in complex multiagent settings: Towards enabling automation in air traffic flow management. Applied Intelligence, 53(4): 4063-4098. https://doi.org/10.1007/s10489-022-03605-1
[11] Abujassar, R.S. (2025). An innovative algorithm for multipath routing and energy efficiency in IoT across varied network topology densities. International Journal of Networked and Distributed Computing, 13(1): 14. https://doi.org/10.1007/s44227-024-00041-0
[12] Yousif, Y.K., Hasan, S.J., Mahmood, I.Y., Saeed, A.Q., Tanash, M., AbuAin, T., Abuain, W.A. (2025). Homomorphic encryption-enabled federated transfer learning for privacy-preserving of Internet of Vehicles. Journal of Soft Computing and Data Mining, 6(3): 416-428.https://doi.org/10.30880/jscdm.2025.06.03.026
[13] Mahamune, A.A., Chandane, M.M. (2024). Trust-based co-operative routing for secure communication in mobile ad hoc networks. Digital Communications and Networks, 10(4):1079-1087. https://doi.org/10.1016/j.dcan.2023.01.005
[14] Iftikhar, A., Qureshi, K.N., Shiraz, M., Albahli, S. (2023). Security, trust and privacy risks, responses, and solutions for high-speed smart cities networks: A systematic literature review. Journal of King Saud University-Computer and Information Sciences, 35(9): 101788. https://doi.org/10.1016/j.jksuci.2023.101788
[15] Wang, Z., Goudarzi, M., Buyya, R. (2025). ReinFog: A deep reinforcement learning empowered framework for resource management in edge and cloud computing environments. Journal of Network and Computer Applications, 242: 104250. https://doi.org/10.1016/j.jnca.2025.104250
[16] Shah, P., Kasbe, T. (2021). A review on specification evaluation of broadcasting routing protocols in VANET. Computer Science Review, 41: 100418. https://doi.org/10.1016/j.cosrev.2021.100418
[17] Alzainalabdin, I., Alzedawi, F. (2025). Self-supervised learning for speech recognition: A comprehensive review. Al-Noor Journal for Information Technology and Cybersecurity, 2(1): 19-22. https://doi.org/10.69513/jncs.v1.i1.a3
[18] Kakulla, S., Malladi, S. (2022). Sybil attack detection in vanet using machine learning approach. Ingenierie des Systemes d'Information, 27(4): 605-611. https://doi.org/10.18280/isi.270410
[19] Ch, M.S.S., Yamarthi, N.R. (2025). Design of an iterative method leveraging deep Q-networks for intrusion detection system operations. IEEE Access, 13: 48720-48745. https://doi.org/10.1109/ACCESS.2025.3551718
[20] Chen, J.M., Li, T.T., Panneerselvam, J. (2018). TMEC: A trust management based on evidence combination on attack-resistant and collaborative internet of vehicles. IEEE Access, 7: 148913-148922. https://doi.org/10.1109/ACCESS.2018.2876153
[21] Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., et al. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71: 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642
[22] Wu, Q., Han, J., Yan, Y., Kuo, Y.H., Shen, Z.J.M. (2025). Reinforcement learning for healthcare operations management: Methodological framework, recent developments, and future research directions. Health Care Management Science, 28(2): 298-333. https://doi.org/10.1007/s10729-025-09699-6
[23] Sivayazi, K., Mannayee, G. (2024). Modeling and simulation of a double DQN algorithm for dynamic obstacle avoidance in autonomous vehicle navigation. E-Prime-Advances in Electrical Engineering, Electronics and Energy, 8: 100581. https://doi.org/10.1016/j.prime.2024.100581
[24] Cho, J.H., Swami, A., Chen, R. (2010). A survey on trust management for mobile ad hoc networks. IEEE Communications Surveys & Tutorials, 13(4): 562-583. https://doi.org/10.1109/SURV.2011.092110.00088
[25] Buchegger, S., Le Boudec, J.Y. (2002). Performance analysis of the CONFIDANT protocol. In Proceedings of the 3rd ACM International Symposium on Mobile Ad Hoc Networking & Computing, pp. 226-236. https://doi.org/10.1145/513800.513828
[26] Jøsang, A., Ismail, R., Boyd, C. (2007). A survey of trust and reputation systems for online service provision. Decision Support Systems, 43(2): 618-644. https://doi.org/10.1016/j.dss.2005.05.019
[27] Ye, H., Li, G.Y., Juang, B.H.F. (2019). Deep reinforcement learning based resource allocation for V2V communications. IEEE Transactions on Vehicular Technology, 68(4): 3163-3173. https://doi.org/10.1109/TVT.2019.2897134
[28] Abdulsattar, N.F., Alkhayyat, A.H., Abbas, F.H., Abosinnee, A.S., Ibrahim, R.K., Ali, R.R. (2023, April). Improved chicken swarm optimization with zone-based epidemic routing for vehicular networks. In International Conference on Frontiers of Intelligent Computing: Theory and Applications, pp. 405-418. https://doi.org/10.1007/978-981-99-6702-5_34
[29] Abbas, F.H., Al-Rawi, O.Y.M., Ali, R.R., Mahmood, S.N., Alkhayyat, A., Hassan, M.H. (2025). A hybrid bio-inspired optimization based enhanced cluster head selection to improve communication in vehicular ad-hoc networks. In Tech Fusion in Business and Society, pp. 475-485. https://doi.org/10.1007/978-3-031-84628-1_40