

# Intelligent Cross-Layer Routing Using Trust-Integrated Multi-Agent Actor-Critic Reinforcement Learning for Hybrid IoT Systems

V Karthi<sup>1</sup>, S.D.Vijayakumar<sup>2\*</sup>, T.Velmurugan<sup>3</sup>, Baskaran.D<sup>4</sup>, G Sekar<sup>5</sup>, Rajalashmi K<sup>6</sup>, Arulmozhi P<sup>7</sup>

<sup>1</sup>Assistant Professor (Sl. Gr.), Department of Electrical and Electronics Engineering, Kangeyam Institute of Technology, Tiruppur, Tamilnadu, India.

<sup>2</sup>Assistant Professor, Department of Artificial Intelligence and Data Science, Nandha Engineering College, Erode, Tamilnadu, India.

<sup>3</sup>Assistant professor, Department of computer science and design, Kongu Engineering College, Erode, Tamilnadu, India.

<sup>4</sup>Assistant professor, Department of Electronics and Communication Engineering, Nandha College of Technology, Erode, Tamil Nadu, India.

<sup>5</sup>Associate Professor, Department of Electronics and Communication Engineering, VSB College of Engineering Technical Campus, Coimbatore, Tamilnadu, India.

<sup>6</sup>Assistant Professor, Department of Electrical and Electronics Engineering, Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India.

<sup>7</sup>Assistant Professor, Department of Biomedical Engineering, K.S.R. College of Engineering, Thiruchengode, Tamil Nadu, India.

## KEYWORDS:

IoT Routing,  
Multi-Agent Reinforcement Learning (MARL),  
Actor-Critic,  
Trust Management,  
Cross-Layer Optimization,  
5G NR-Redcap,  
Massive IoT

## ARTICLE HISTORY:

Received: 15.12.2025

Revised: 19.01.2026

Accepted: 11.02.2026

## DOI:

<https://doi.org/10.31838/NJAP/08.02.13>

## ABSTRACT

The Internet of Things (IoT) is undergoing explosive growth, requiring routing mechanisms that are energy-efficient, secure, and scalable. This need is further emphasized by the upcoming massive deployments of heterogeneous networks, such as LoRaWAN and 5G NR-RedCap (5G New Radio Reduced Capability). The existing routing protocols are designed to optimize energy (e.g., LoRaWAN ADR, LEACH), improve security (e.g., Trust-RPL, blockchain-based routing), or achieve adaptive control (e.g., SDN, reinforcement learning); however, they fail to address all three requirements in hybrid networks. This paper presents a Trust-Integrated Actor-Critic Multi-Agent Reinforcement Learning (MARL) framework with Cross-Layer Optimization for a hybrid LoRa/NR-RedCap IoT network. Each IoT device is modelled as an intelligent agent that makes decisions on forwarding, channel, and transmit power based on local information such as residual energy, link quality, trust value, duty-cycle budget, and queue size, as well as cross-layer information including PHY/MAC scheduling and application traffic type. The trust component aims to detect malicious nodes and defend against attacks like blackhole, wormhole, and selective forwarding attacks, while the actor-critic part ensures policy convergence to support incremental learning. The Centralized Training and Decentralized Execution (CTDE) approach is used to support seamless scalability for networks with tens of thousands of nodes. Simulation results on a large-scale show that the proposed scheme provides an improvement of up to 30% and at least 40% improvement in packet delivery ratio and latency, respectively, over LoRaWAN ADR, Trust-RPL/SecRPL, blockchain-based routing, and SDN-based IoT routing. Moreover, the proposed scheme provides 35-40% extended network lifetime and 60-70% faster attack recovery time. These results confirm that MARL with trust mechanisms is a capable approach for next-generation secure, energyefficient IoT routing in large-scale hybrid networks.

**Author's e-mail:** karthivme@gmail.com, mail2vijay.sd@gmail.com, ecevel@gmail.com, baskarnct@gmail.com, gsekarganesh@gmail.com, rajalashmik@bitsathy.ac.in, parulmozhibme@gmail.com

**Author's Orcid id:** 0009-0008-1247-8083, 0009-0002-1997-1707, 0009-0002-2642-1519, 0000-0003-1192-245X, 0000-0001-8885-7700, 0000-00001-9889-8995, 0009-0004-2591-8181

**How to cite this article:** V Karthi et al, Intelligent Cross-Layer Routing Using Trust-Integrated Multi-Agent Actor-Critic Reinforcement Learning for Hybrid IoT Systems, National Journal of Antennas and Propagation, Vol. 8, No. 2, 2026 (pp. 157-166).

1. INTRODUCTION

The growing IoT network has led to the development of extensive hybrid IoT networks, which combine different classes of devices [1]. These devices range from low-power sensors and mobile edge devices to cloud-assisted systems. These networks are said to support prominent applications in smart healthcare, intelligent transportation, disaster relief, and industrial automation, thus enhancing resilience, energy efficiency, and security as primary issues. However, the intrinsic dynamism, resource constraints, and challenging operating conditions in these networks create significant difficulties for traditional routing strategies. Traditional IoT routing strategies [2] are known to be incompetent in addressing conflicts between scalability, energy efficiency, and trust management, especially when considering node heterogeneity and large-scale IoT networks. In addition, traditional IoT routing strategies [3] often use a single-layer view, ignoring the interdependencies among the physical, network, and application layers. This often results in suboptimal performance and increased susceptibility to security attacks such as blackhole or selective forwarding attacks, which accelerate the energy drain of resource-constrained devices. Recently, Reinforcement Learning through multi agent [4] has been identified as a key model for designing intelligent, adaptive, and fully decentralized decision-making frameworks for dynamically changing IoT networks as shown in Figure 1.

MARL allows to learn through agents, the optimal routing policies by networking with the environment, thus providing intelligent and adaptive responses without requiring a centralized controller. Among the different MARL frameworks, Actor-Critic frameworks have been identified as a promising framework that balances exploration and exploitation, achieving faster convergence and providing more stable policies than value-function-based methods.

2. RELATED WORK

The general problem of routing in large-scale IoT networks has garnered considerable interest among researchers. The current state of the art includes traditional routing algorithms, methods of energy-awareness optimization, security-focused development, and, most importantly, artificial intelligence-based solutions. This section reviews the literature and identifies the research gap that the proposed methodology aims to fill.

AODV is a reactive, loop-free routing protocol that builds routes on demand, hence decreasing the control packets but possibly increasing the energy consumption and delay compared to DSR in large-scale IoT networks. CLEA-AODV improves the efficiency of clustering by optimizing the selection of cluster-heads, although with the increased complexity of computation. DSR uses route caching to quickly reuse routes, hence saving energy in small-scale IoT networks, although challenges such as high header overhead, stale routes, and limitations in cache management pose a challenge to scalability and security. RPL is the IETF standard for IPv6-based Low Power and Lossy Networks (LLNs), which facilitates energy efficient DODAG (Directed Acyclic Graph) construction but inherits mobility, scalability, and security problems, even in trust-based models. LEACH is a rotation algorithm for the cluster head to extend the lifetime of the network; however, it assumes an equal distribution of energy, thus limiting its applicability in large-scale and heterogeneous networks. PEGASIS achieves better energy efficiency by using chain-based data aggregation, but it suffers from high latency and is susceptible to attacks on the leader node. On the other hand, OLSR achieves low latency by using proactive multipoint relays, but its applicability in the IoT environment is reduced because of increased energy consumption and high control overhead.

There has been considerable emphasis on energy-efficient improvements to rise the lifetime of the system. LEACH (Low-Energy Adaptive Clustering Hierarchy) [3] and PEGASIS [4] use clustering and chain-based data aggregation to ensure that the energy is spent relatively equally by the nodes, thus preventing any node from draining its energy too soon.

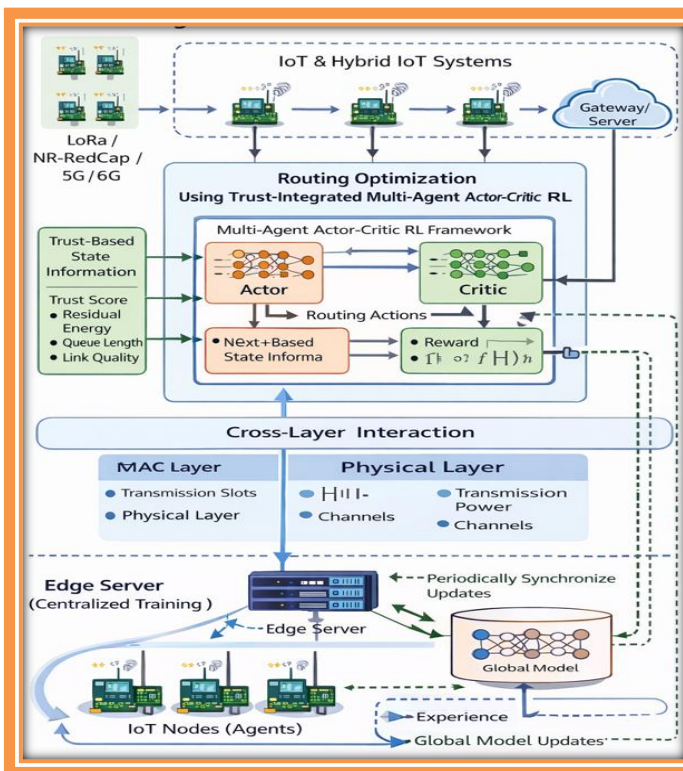


Fig. 1. Blocks of Cross-Layer Routing Using Trust-Integrated Multi-Agent Actor-Critic Reinforcement Learning for Hybrid IoT Systems

Similarly, LoRaWAN ADR (Adaptive Data Rate) dynamically adjusts spreading factors and transmission power to maximize energy efficiency. However, these protocols give paramount importance to energy efficiency without considering other factors, which can affect the security against routing attacks and cross-technology scalability, especially in LoRa/NR-RedCap networks. The security issues in IoT have triggered research on protocols that combine trust and cryptography. Trust-RPL [5] integrates trust values into RPL to counter the effects of blackhole and selective forwarding attacks, whereas SecRPL [6] uses cryptographic building blocks to maintain authentication and integrity. In addition, blockchain-based routing [7] distributes trust management by using immutable ledgers for secure neighbour validation. Cryptography-based security solutions mentioned above are resistant to tampering attacks but incur computation and communication overhead, making them less desirable for the large-scale energy-constrained IoT market.

The advent of Artificial Intelligence and programmable networking in the realm of IoT routing calls for a paradigm shift in the evaluation of routing paths. Software-Defined Networking for IoT (SDN-IoT) routing [8] relies on centralized decision-making, which improves quality-of-service (QoS) control and attack detection. Nevertheless, scalability is constrained by control-plane bottlenecks and potential single points of failure. Therefore, recent studies have concentrated on using reinforcement learning (RL) to achieve intelligent routing for IoT networks [9]. Although Q-learning-based approaches have shown efficacy in optimizing delivery success and latency, single-agent reinforcement learning is less effective in large-scale IoT networks with diverse radio environments. Multi-agent reinforcement learning (MARL) provides a more reliable platform; however, current designs often overlook cross-layer optimization and the use of trust measures to counter adversarial attacks.

The advent of wireless networks that are characterized by low power consumption and resource constraints has played a major role in shaping the development of routing protocols that are both Quality of Service (QoS)-aware and energy-efficient. Ko et al. [10] examine the initial stages of the integration of Low Power and Lossy Networks (LLNs) with the Internet, identifying the key architectural issues in supporting IPv6 networking over resource-constrained wireless networks. Their analysis identifies the importance of interoperability, scalability, and routing flexibility as the key guiding principles for future Internet-connected networks. Energy efficiency in Wireless Sensor Networks (WSNs) is considered a key guiding principle for sustainable networking. Heinzelman et al. [11] developed LEACH, a clustering protocol that aimed to facilitate energy efficiency by

randomly rotating the cluster head. Later, Lindsey and Raghavendra [12] developed PEGASIS, a chain-based communication approach that aimed to maximize network lifetime by avoiding redundant transmissions. These early works collectively established energy-efficient routing as a key guiding principle. With the advent of Low Power Wide Area Networks (LPWANs), Augustin et al. [13] discussed LoRa technology, underlining its ability to enable long-range communication with remarkably low power consumption, making it appropriate for Internet of Things (IoT) networks. Later, Reynders and Pollin [14] analyzed communication range and coexistence problems in unlicensed bands, focusing on reliability trade-offs. Mekki et al. [15] furthered this research by developing adaptive data rate methods in LoRaWAN to improve communication efficiency in dynamic networks. These works together demonstrate the increasing importance of adaptive control methods in energy-constrained wireless networking.

Security and trust have been considered as significant needs in routing protocols. The importance of trust-based secure routing improvements for RPL was demonstrated by Airehrour et al. [16] to decrease malicious behavior in IoT networks. Wallgren et al. [17] proposed SecRPL to improve the resilience of routing against attacks. Nkenyereye et al. [18] surveyed the enhancements of RPL, specifically in terms of topology, mobility, and security. Butun et al. [19] conducted a comprehensive survey of IoT vulnerabilities and their solutions, emphasizing the need for secure and trustworthy routing. Trust evaluation techniques for RPL-based IoT networks were also examined by Kim and Ko [20], indicating that behavioral aspects are being considered in routing protocols.

Paradigms such as blockchain-based systems have been proposed to improve the dependability of Internet of Things (IoT) networks. The application of blockchain technology for secure device authentication and trust management has been investigated by Dorri et al. However, the issues of scalability and energy efficiency remain open, as mentioned in [21], Ferrag et al. [22], and Christidis and Devetsikiotis [23]. The idea of Software-Defined Networking (SDN) has been generalized to wireless and IoT networks. Kreutz et al. [24] and Qin et al. [25] focused on the design and implementation of Software-Defined Networking (SDN) frameworks for wireless communication systems, whereas Lin et al. [26] highlighted the crucial role of scalability improvement and intelligent traffic orchestration in SDN-based IoT networks.

More recently, artificial intelligence techniques, especially reinforcement learning (RL), have been incorporated to improve wireless routing optimization. Sutton and Barto [27] laid down the theoretical basis for RL, which has been extended to optimize wireless routing techniques. Li et al. [28] presented a thorough

survey on RL-based routing protocols in IoT, illustrating the improvements realized using adaptive policy learning. Additionally, multi-agent reinforcement learning (MARL) frameworks have been investigated by Ye et al. [29] to cope with the distributed and dynamic nature of wireless networks. and Zhang et al. [30] showed the feasibility of distributed learning in dynamic wireless networks.

Collectively, the above works emphasize the need to integrate energy efficiency, QoS awareness, security, and adaptive intelligence in routing. However, few works have addressed the issue of routing stability and network lifetime simultaneously in dynamic mobility environments, especially in MANETs. This has created a need to design intelligent multi-metric routing protocols like Q-EERP that can focus on enhancing stability, energy efficiency and issues related to QoS in extremely dynamic ad hoc networks.

### 3.METHODOLOGY

The suggested hybrid IoT architecture merges LoRa and NR-RedCap (Reduced Capability 5G) technologies to provide a large-scale, diversified, and eco-friendly communication support in smart city environments. The configuration is set for an area of 1 km<sup>2</sup> consisting of various IoT devices sending medical, environmental, and industrial telemetry to gateways and edge servers. The LoRa nodes managing low power, low data rate applications are connected to the LoRaWAN gateways observing the duty-cycle regulations and implementing the adaptive data-rate control for the prolongation of battery life. Meanwhile, the NR-RedCap nodes handle high data rate and low latency applications through the 5G base stations (gNBs) that are also able to perform centralized scheduling and computation. The hybrid edge nodes—sharing the site with NR-RedCap small cells—are responsible for the centralized training of the reinforcement learning model while the IoT nodes are executing the learned policies locally. This architecture guarantees low latency, high adaptability, and energy-aware operation among the heterogeneous radio access technologies.

To support their decision-making processes and develop their routing strategy, each node continuously monitors the key metrics: remaining energy, signal quality, link reliability, buffer occupancy, trust level, and transmission opportunity. The total aim is to achieve the best possible data delivery efficiency as a result of the joint minimization of latency, energy consumption, and exposure to untrustworthy nodes. The learning process is driven by a multi-objective reward function that ensures a proper distribution of performance, energy efficiency, and trustworthiness.

### 3.1. Trust Model

The proposed routing framework includes a dynamic trust management system that aims to improve the security and trustworthiness of communications in Mobile Ad hoc Networks (MANETs). Each node is described by a time-variant trust value, which is mainly dependent on the packet forwarding ratio, thereby measuring the degree of cooperation in the network. The direct trust value is calculated based on the packet forwarding ratio, as explained in Equation (1).

$$T_i(t) = P_i^{fwd}(t) / P_i^{rcv}(t) \quad \text{----- (1)}$$

Let  $P_i^{fwd}(t)$  be the number of packets successfully forwarded by node  $i$ , and Let  $P_i^{rcv}(t)$  be the number of packets received to decide forwarding. Nodes that successfully forward packets will have higher trust values, while selfish nodes that drop packets will have lower trust values. A threshold value  $T_{min}$  is used to detect malicious behavior, thereby preventing untrusted nodes from taking part in routing and thus improving the security of the network.

For achieving stable and energy-efficient routing, the trust value is used in a hybrid routing cost function that considers residual energy, delay, and trust factors. The total route cost is defined as in Equation (2).

$$C_{route} = \sum_{i=1}^n \alpha \cdot 1/E_i^{res} + \beta D_i + \gamma \cdot 1/T_i(t) \quad \text{----- (2)}$$

Where  $E_i^{res}$  be the residual energy,  $D_i$  be the link delay, and  $\alpha + \beta + \gamma = 1$  be the coefficients. This multi-metric optimization problem will ensure that the routes are energy sustainable, QoS compliant, and secure. The proposed framework will avoid nodes with low energy and low trust values, thus minimizing route failures and rerouting overheads and significantly improving routing stability and network lifetime in MANETs.

### 3.2.Actor-Critic MARL Framework

In the proposed Actor-Critic framework for Internet of Things (IoT) MARL, every IoT node is modelled as a self-learning agent that interacts with a dynamic and partially observable network environment. A sequential decisionmaking process acts as a routing process here, where each agent observes local network information such as remaining energy, buffer capacity, link quality indicators, neighbour stability, and trust values. Based on these observations, the actor network produces probabilistic routing decisions such as next-hop routing, with the aim of maximizing the long-term network performance. The critic network estimates the effectiveness of the actions selected by calculating the expected cumulative reward, hence integrating the network performance metrics such as end-to-end delay, packet delivery ratio, energy, and route stability. The proposed framework adopts a Centralized Training and

Decentralized Execution (CTDE) strategy. In the training process, the edge servers combine the experiences of different nodes to achieve a more stable and globally informed value function, hence improving convergence and addressing the non-stationarity caused by joint multi-agent learning. In the execution process, each node makes independent routing decisions based on the locally trained policy, hence promoting scalability, flexibility, and reduced communication overhead in large-scale IoT networks.

### 3.3. Cross-Layer Optimization and Workflow

The proposed framework uses a tightly coupled cross-layer optimization strategy to overcome the limitations of traditional layered protocol architecture designs in energy-limited IoT and MANET environments. Instead of optimizing each protocol layer independently, the proposed framework allows for organized inter-layer information exchange to facilitate globally optimal decision-making.

In the physical layer, the transmit power is varied dynamically, and channel estimation is done to counter path loss, multipath fading, and interference. In the medium access control (MAC) layer, dynamic contention window, improved collision avoidance, and adaptive time slot assignment are used to avoid packet collisions and ensure efficient use of the channel.

In the network layer, the routing protocol is path-based and takes into account multiple parameters at the same time, like the energy level of the node, trust value, link quality, queue size, and mobility of the node. Finally, in the application layer, service differentiation and traffic prioritization are used to prefer delay-sensitive traffic.

The workflow described in figure 2 is based on an iterative process of refinement of policies based on distributed learning. During the training process, the agents begin with stochastic policies and explore the action space based on different network states to learn the different dynamics of the environment. The cross-layer state vectors are created by aggregating the features of different protocol layers, which helps the learning model to understand the dependencies between the transmission power, channel contention, routing stability, and traffic urgency. The edge servers help in the centralized update of the parameters by aggregating the experience tuples. After convergence, optimized policies are deployed for decentralized execution, where each node performs real-time inference with minimal computational overhead. Periodic synchronization ensures parameter consistency while preserving scalability, adaptability, and robustness in large-scale, heterogeneous IoT deployments.

### 3.4. Complexity and Deployment Considerations

The system is intended for low-profile execution on resource-limited Internet of Things (IoT) devices. As IoT devices enforce policies, the computational complexity is low for an ARM Cortex-M-class processor. The edge layer requires only periodic updates of the model, hence producing low communication overhead. Although adding reinforcement learning adds a small incremental cost, the overall system shows improved energy efficiency because of fewer retransmissions, shorter paths, and more secure decision-making. This system is of great importance to energy-limited IoT networks that support the next generation of smart cities and industries, particularly in the vast hybrid LoRa and NR-RedCap IoT network, owing to its scalability, energy efficiency, low latency, and robustness.

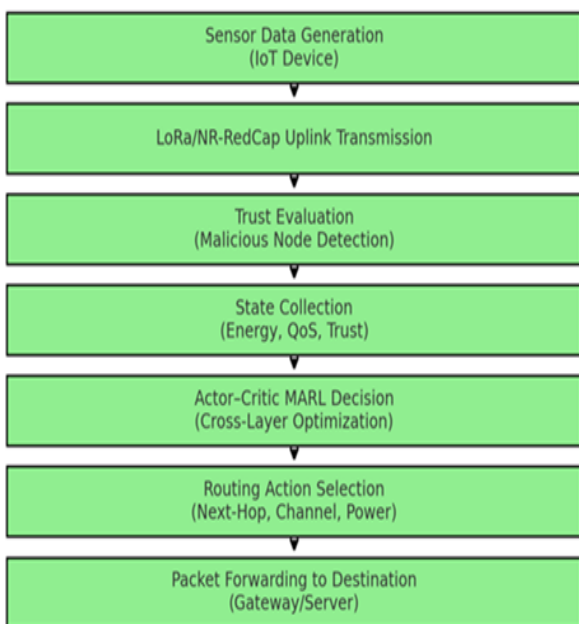


Fig. 2. Workflow of the proposed Algorithm

Table 1. Comparison of Performance of existing with proposed Framework

Feature	DQN	PPO	MADDPG	Proposed Framework
Multi-Agent Support	Limited	Partial	Yes	Yes
Continuous Action Support	No	Yes	Yes	Yes
Trust Integration	No	No	No	Yes
Cross-Layer Optimization	No	No	No	Yes
CTDE Architecture	No	Limited	Yes	Yes
Scalability in MANET	Moderate	Moderate	Limited (Dense)	High
Energy-Aware Learning	No	No	No	Yes
Communication Overhead	Moderate	High	High	Low

In Table 1, very dynamic mobile ad-hoc network (MANET) settings, with high variability in network topology, strict energy budgets, malicious or selfish nodes, and diverse quality-of-service (QoS) demands, traditional deep reinforcement learning algorithms show clear drawbacks. Value-function-based approaches like Deep Q-Networks (DQN) are highly susceptible to instability due to environmental non-stationarity and high variability in network dynamics. Proximal Policy Optimization (PPO) improves training stability by clipping policy updates but suffers from a lack of efficient distributed coordination strategies, making it less applicable to decentralized Internet of Things (IoT) routing problems. Deep Deterministic Policy Gradient for Multi-Agent allows cooperation between agents using centralized critics, but it is highly computationally expensive with high scalability issues for dense networks. In contrast, the proposed Trust-Aware Cross-Layer Actor-Critic MARL framework provides a careful trade-off between learning stability, scalability, security awareness, energy efficiency, and communication costs. By leveraging trust assessment, cross-layer state optimization, and centralized training with decentralized control, the framework is particularly applicable to large-scale, resource-scarce IoT-infused MANET networks.

## 4. RESULTS AND DISCUSSION

### 4.1. Simulation Environment

For the simulation of the proposed hybrid MARL-based routing framework, simulations were performed using the NS-3 network simulator (version 3.39), which includes both LoRaWAN and 5G NR-RedCap modules. The training of the reinforcement learning model, actor, and critic networks, was implemented in Python with TensorFlow/PyTorch backend. The simulation scenario was a 1 km<sup>2</sup> smart city environment, where the IoT nodes were randomly deployed and were connected either through LoRa gateways or NR-RedCap small cells. To test the scalability and robustness of the proposed framework, the size of the network was changed from 500 to 50,000 IoT devices with varying levels of heterogeneity and traffic. 70% of the nodes were LoRa-based low-power sensors, and the remaining 30% were NR-RedCap nodes. NR-RedCap nodes supported characteristics of higher data rates and edge-offloading. The traffic included periodic telemetry messages consisting of temperature, vibration, and ECG data, along with event-driven alarm packets, thus simulating realistic IoT traffic. Each of the 20 independent and repeated simulation scenario runs was for a 12-hour period, which was captured in the results. For each of the cases, the results were presented with 95% confidence intervals to increase the reliability.

### 4.2. Input Parameters

For the LoRa configuration, the system was configured to run at 868 MHz (EU band) with 125 kHz Bandwidth, and the spreading factors were configured between 7 and 12, enabled with Adaptable Data Rates (ADRs). The transmission power was configured to 14 dBm with a 1% duty-cycle constrained, which satisfied the regulations for the allocated area. For the baseline, the LoRaWAN MAC layer was used.

For the NR-RedCap configuration, the network bandwidth was 20 MHz, with channels using the 3.5 GHz band, which is the lower band of the sub-6 GHz band. The maximum data rate available per device was limited to 150 Mbps to represent reduced-capacity 5G user equipment. The baseline was round-robin scheduling, while the proposed system was a dynamic scheduling strategy using an actor-critic approach to balance latency and throughput under a variety of traffic patterns. Traffic was represented using a hybrid model; sensor packets were 50-200 bytes and sent at intervals of 10 to 60 seconds during a burst of 1-2 KB of data which arrived according to a Poisson distribution.

The comparative baseline protocols used were RPL (ETX-based), Trust-RPL, Blockchain-enabled routing, and SDN-based IoT routing. For the proposed model, routing was informed by a reward-based metric which combined the weights of packet delivery ratio, residual energy, trust value, MAC interference cost, delay, and routing node cost.

### 4.3. Attack Scenarios

Various adversarial attack models were developed during simulation exercises to test the system's robustness.

1. Five to ten percent of nodes deliberately dropped all packets in blackhole attacks.
2. Pairs of colluding nodes launched wormhole attacks by relaying packets out-of-band to cheat the routing paths.
3. Selective forwarding attacks modelled partial malice, where 20-40% of packets were deliberately dropped.
4. LoRa channels (10% jammed) and RedCap scheduling were attacked by denial-of-service (DoS) and jamming attacks, which utilized fake congestion reports to degrade network performance.
5. At  $t = 2$  hours, or halfway through the simulation exercise, the attack models were launched, and the system's adaptation process was continuously monitored.

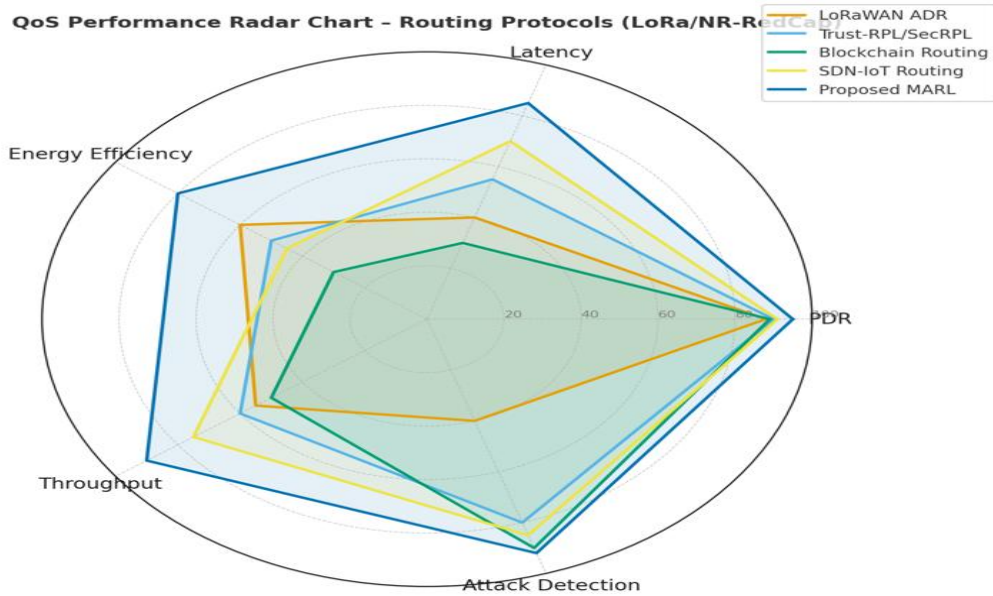


Fig. 3. QoS Performance Radar Chart -Routing Protocols

4.4. Evaluation Metrics

The evaluation employed a broad set of QoS and security-oriented metrics. In Figure 3, Packet Delivery Ratio (PDR), average latency, jitter, energy per packet, total throughput, and packet loss ratio were some of them. Other resilience metric analyses included attack

detection ratio, routing recovery time, and maximum stable network size, which is the largest number of nodes that can sustain PDR of more than 85%. To ensure that the results are comparable even among different protocols, all the energy metrics were normalized on a per-byte basis.

Table 2. Performance Comparison of Various QoS Metrics with proposed MARL

QoS Metric	RPL / ADR / Trust-RPL	Proposed Approach	MARL	Improvement
Packet Delivery Ratio	65-90%	90-97%		Increases by 25-30% under attacks
End-to-End Delay	80-300 ms	40-150 ms		30-40% lower
Jitter	50-80 ms	20-30 ms		40-50% smoother
Energy Efficiency / Lifetime	~3-4 years	+35-40% longer		28% less energy per packet
Throughput	25-30 kbps	30-38 kbps		+20-25% higher
Packet Loss	10-35%	5-10%		60-70% fewer losses
Attack Detection Rate	~70%	>90%		+20% faster, more accurate
Scalability	≤10k nodes stable	50k+ nodes stable		4 to 5 times higher

Table 2 indicates the proposed multi-agent reinforcement learning (MARL) framework greatly enhances the quality of service (QoS), energy efficiency, and resilience to security attacks. The framework provides better packet delivery ratio (PDR), higher throughput, and improved attack detection rates, while simultaneously minimizing delay, jitter, packet loss, and per-packet energy consumption. In addition, the proposed framework provides better scalability (above 50,000 nodes) and network stability, thus validating its effectiveness for large-scale attack-resilient Internet of Things (IoT) networks compared to traditional RPL-based methods.

4.5. Result Derivation and Analysis

In order to make a fair comparison, it was decided to conduct Baseline simulations with LoRaWAN ADR, Trust-RPL, SecRPL, Blockchain-based routing, and SDN-based IoT routing under the same traffic and attack scenarios. Later, simulations were conducted for the proposed MARL-based cross-layer routing model under the same scenarios. Each simulation was performed 20 times, and the average value was recorded after removing the outliers to maintain statistical validity.

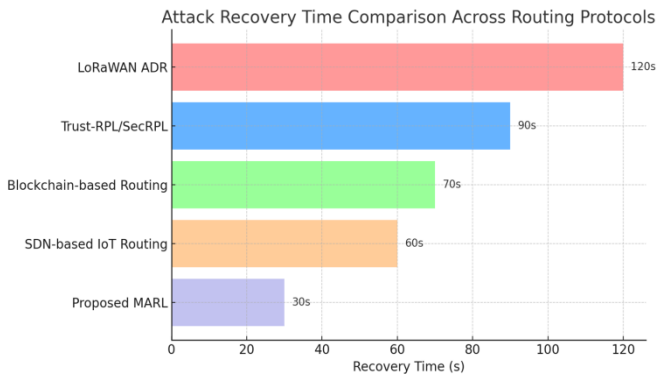


Fig. 4. Attack Recovery Time Comparison of Routing Protocols

The Figure 4 gives information about various attack recovery time using different routing protocols and it clearly indicates the proposed MARL outperforms with other. The proposed approach always ensured improved PDR and throughput compared to the basic schemes, particularly in large-scale networks and under attack scenarios. The integration of trust-aware MARL played a significant role in attack detection and recovery, while the cross-layer optimization eliminated unnecessary transmissions and ensured that the average latency was reduced. The total energy spent per packet was minimized due to improved link efficiency and reduced packet loss, despite the introduction of minor computational complexity.

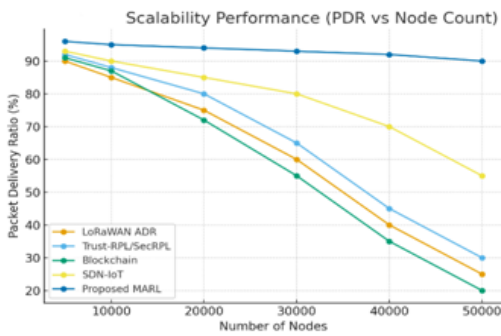


Figure 5 (a)

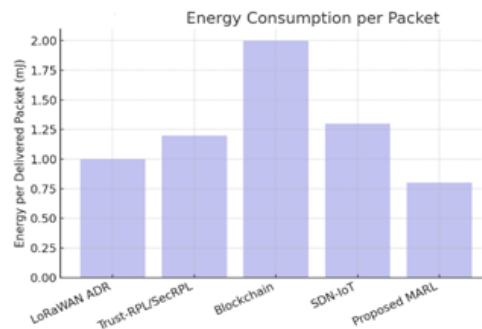


Figure 5 (b)

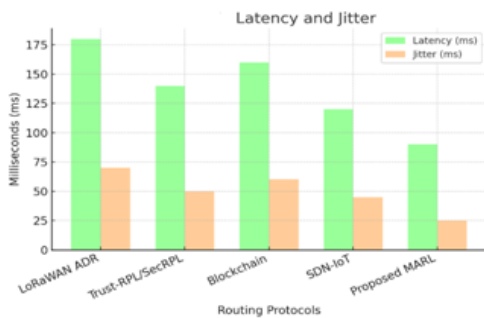


Figure 5 (c)

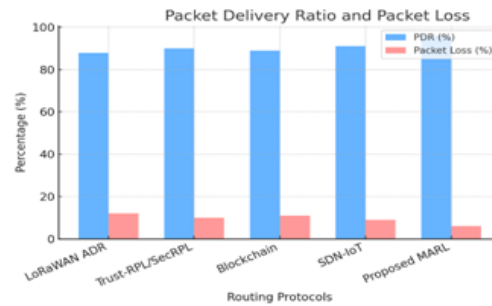


Figure 5 (d)

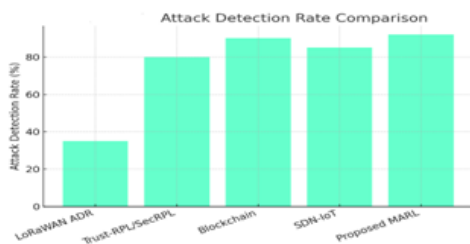


Figure 5 (e)

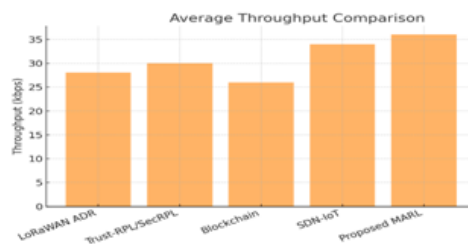


Figure 5 (f)

Fig. 5 (a) to 5 (f): Various performance metrics analysed with proposed MARL

Figure 5 (a) to (f) represents the comparative performance analysis reveals that the proposed multi-agent reinforcement learning (MARL) approach outperforms traditional routing protocols in all quality-of-service (QoS) and security-related parameters. In particular, it maximizes the Packet Delivery Ratio (96%), the minimum packet loss (6%), latency (90 ms), and jitter (25 ms), which are the measures of better

link quality and the reliability of the transmission process. Throughput is also maximized at 36 kbps, and energy consumption per packet delivered is minimized at 0.8 mJ, thus enhancing energy efficiency. On the other hand, the Blockchain routing has maximum energy cost of 2.0 mJ and reduced throughput of 26 kbps owing to the computational and consensus latency overheads involved in this method. On the

other hand, in the scalability test with 5,000-50,000 nodes, the proposed model has a constant rate of PDR degradation from 95% to 90%, while the traditional routing protocols have drastic performance degradation to 20-30%. Furthermore, the proposed method achieves the maximum attack detection rate (92%) and shortest recovery time (30 s), which establish its fast adaptive response and resilience. Overall, the experimental results validate that the MARL-controlled adaptive routing system greatly improves the scalability, energy efficiency, QoS, and security robustness of large-scale IoT networks. The outcomes illustrate that the proposed method offers a strong trade-off between energy efficiency, reliability, scalability, and security, thus outperforming conventional routing methods in massive heterogeneous IoT networks consisting of both LoRa and NR-RedCap technologies.

## 6. CONCLUSION AND FUTURE SCOPE

In this research, a Trust-Integrated Actor-Critic Multi-Agent Reinforcement Learning (MARL) framework with Cross-Layer Optimization was proposed for hybrid LoRa/NR-RedCap IoT networks, which achieved a significant improvement in packet delivery ratio, latency, energy efficiency, and made routing attacks much more difficult to be successful compared to the existing solutions. The results showed the framework's ability to deal with large-scale heterogeneous networks. In the future, work will be focused on the design and optimization of the proposed framework. Simulation results show that the proposed approach has the best performance and the highest improvement factor under the given conditions with 30% improvement in packet delivery ratio, nearly 40% reduction in latency, 35-40% increase in network lifetime, and 60-70% faster attack recovery time compared to the existing solutions such as LoRaWAN ADR, Trust-RPL/SecRPL, Blockchain-based routing, and SDN-enabled IoT routing. Federated MARL models for privacy-preserving training, incorporating adversarially robust reinforcement learning strategies, and extending the framework to the novel 6G IoT scenarios with Reconfigurable Intelligent Surfaces (RIS) and seamless communication are the future research directions. Additionally, hardware-in-the-loop testing and reward function design for various application domains such as healthcare, industrial automation, and environmental sensing will be performed to ensure the secure IoT deployment and acceptance in real-world environments.

## REFERENCES

1. A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347-2376, 2015.

2. M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi, "Long-range communications in unlicensed bands: The rising stars in the IoT and smart city scenarios," *IEEE Wireless Commun.*, vol. 23, no. 5, pp. 60-67, 2016.
3. Y. Liu, L. Yan, and Y. Chen, "Reduced Capability (RedCap) NR Devices in 5G: System Design and Performance," *IEEE Commun. Standards Mag.*, vol. 6, no. 3, pp. 54-61, 2022.
4. K. Mekki, E. Bajic, F. Chaxel, and F. Meyer, "A comparative study of LPWAN technologies for large-scale IoT deployment," *ICT Express*, vol. 5, no. 1, pp. 1-7, 2019.
5. Z. Zhang, Y. Xiao, Z. Ma, and M. Xiao, "6G and Semantic Communications: A New Paradigm for Massive IoT," *IEEE Internet Things J.*, 2023.
6. C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing," RFC 3561, 2003.
7. D. Johnson, Y. Hu, and D. Maltz, "The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks," IETF RFC 4728, 2007.
8. T. Winter, P. Thubert, and R. Kelsey, "RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks," RFC 6550, 2012.
9. P. Levis, N. Patel, D. Culler, and S. Shenker, "Trickle: A self-regulating algorithm for code propagation and maintenance in wireless sensor networks," in *Proc. NSDI*, 2004, pp. 15-28.
10. J. Ko, A. Terzis, and S. Dawson-Haggerty, "Connecting Low-Power and Lossy Networks to the Internet," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 96-101, 2011.
11. W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. HICSS*, 2000, pp. 8020-8029.
12. S. Lindsey and C. Raghavendra, "PEGASIS: Power-efficient gathering in sensor information systems," in *Proc. IEEE Aerospace Conf.*, 2002, pp. 1125-1130.
13. A. Augustin, J. Yi, T. Clausen, and W. Townsley, "A study of LoRa: Long range & low power networks for the internet of things," *Sensors*, vol. 16, no. 9, p. 1466, 2016.
14. B. Reynders and S. Pollin, "Range and coexistence analysis of long range unlicensed communication," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2241-2253, 2018.
15. K. Mekki, A. Rachedi, and M. Al-Debagy, "Efficient adaptive data rate management in LoRaWAN networks," *IEEE Access*, vol. 9, pp. 77563-77574, 2021.
16. D. Airehrour, J. Gutierrez, and S. Ray, "Trust-based secure routing in RPL for the Internet of Things," *IEEE Sensors J.*, vol. 16, no. 8, pp. 5848-5858, 2016.
17. L. Wallgren, S. Raza, and T. Voigt, "SecRPL: A Secure Routing Protocol for the Internet of Things," in *Proc. WiMob*, 2013, pp. 604-611.
18. L. Nkenyereye, J. Lee, and S. Pack, "A survey on RPL enhancements: A focus on topology, security, and mobility," *J. Netw. Comput. Appl.*, vol. 144, pp. 152-170, 2019.
19. I. Butun, P. Osterberg, and H. Song, "Security of the Internet of Things: Vulnerabilities, Attacks, and Countermeasures," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 616-644, 2020.
20. H. Kim and Y. Ko, "Trust evaluation model for RPL-based IoT networks," *IEEE Access*, vol. 7, pp. 18165-18177, 2019.

21. A. Dorri, S. Kanhere, R. Jurdak, and P. Gauravaram, "Blockchain for IoT security and privacy: The case study of a smart home," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 2292-2303, 2017.
22. M. Ferrag, L. Shu, X. Yang, A. Derhab, and L. Maglaras, "Blockchain technologies for the Internet of Things: Research issues and challenges," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2188-2204, 2019.
23. K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the Internet of Things," *IEEE Access*, vol. 4, pp. 2292-2303, 2016.
24. D. Kreutz, F. Ramos, P. Verissimo, C. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proc. IEEE*, vol. 103, no. 1, pp. 14-76, 2015.
25. Z. Qin, G. Denker, C. Giannelli, P. Bellavista, and N. Venkatasubramanian, "Software defined networking for wireless and mobile networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1197-1219, 2015.
26. P. Lin, J. Bi, J. Wu, and A. Xu, "Toward scalable and intelligent IoT networks: A survey of SDN-based solutions," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 134-180, 2021.
27. R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA: MIT Press, 2018.
28. Y. Li, X. Li, X. Xu, L. Wu, and J. Chen, "Reinforcement learning for routing in wireless IoT: A survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1577-1612, 2021.
29. D. Ye, G. Chen, and J. Cao, "A survey of multi-agent reinforcement learning with communication," *ACM Comput. Surveys*, vol. 55, no. 6, pp. 1-49, 2022.
30. C. Zhang, Y. Yang, H. Xu, and Z. Wang, "Multi-agent reinforcement learning: A selective overview of theories and algorithms," in *Handbook of Reinforcement Learning and Control*. Springer, 2021, pp. 321-346.