10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

APTAPPO+: A Lightweight Trust-Aware Reinforcement Learning Routing Algorithm for IoT Networks

Hitesh Parmar
Assistant Professor
K.S School of Business Management & Information Technology
Gujarat University
Ahmedabad, Gujarat, India
hiteshparmar@gujaratuniversity.ac.in

Dr. Hardik Joshi
Associate Professor
Department of Computer Science
Gujarat University
Ahmedabad, Gujarat, India
hardik.joshi@gujaratuniversity.ac.in

Dr. Kamaljit I. Lakhtaria
Associate Professor
Department of Computer Science
Gujarat University
Ahmedabad, Gujarat, India
kamaljit.lakhtaria@gujaratuniversity.ac.in

Abstract— The proliferation of Internet of Things (IoT) networks has introduced unprecedented challenges in ensuring secure, adaptive, and efficient routing protocols capable of handling dynamic topologies, resource constraints, and malicious entities in IoT networks. This study presents APTAPPO+ (Adaptive Trust-Predictive Advantage Policy Proximal Optimization), an enhanced multi-agent reinforcement learning framework that integrates lightweight trust-anomaly detection for secure IoT routing. The proposed algorithm combines Long Short-Term Memory (LSTM)-based trust prediction models with threshold-based anomaly detection to identify and mitigate malicious node behaviors in real time. A Proximal Policy Optimization (PPO) agent leverages composite trust scores to determine optimal routing paths while maintaining performance optimization and trust-aware decision-making. The Lightweight Trust Anomaly Detection (LTAD) modules filter potentially untrustworthy nodes before routing decisions to ensure security and efficiency. The comprehensive simulation results demonstrate significant improvements over existing approaches, achieving a 95% packet delivery ratio with a 53ms end-to-end delay, representing a 5% improvement in reliability and a 12% reduction in latency compared with the baseline methods. The proposed APTAPPO+ framework offers a scalable and secure solution for next-generation IoT networks, effectively balancing trust management with routing performance optimization.

Index Terms—Internet of Things, Reinforcement Learning, Trust Management, Anomaly Detection, Secure Routing, LSTM, Proximal Policy Optimization, Multi-agent Systems

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

INTRODUCTION

The exponential growth of Internet of Things (IoT) deployments has fundamentally transformed modern communication paradigms, with projections indicating over 75 billion connected devices by 2025 [1]. These networks span di-verse applications including smart cities, industrial automation, healthcare monitoring, and intelligent transportation systems, each demanding reliable, scalable, and secure communication protocols [2], [3]. Recent advances in edge intelligence and digital twin technologies have further expanded IoT capabilities, enabling more sophisticated real-time processing and decision-making [4]. However, the inherent characteristics of IoT networks—including highly dynamic topologies, resource-constrained devices, heterogeneous communication patterns, and open wireless channels—present significant challenges for traditional routing protocols [5].

A. Problem Statement and Motivation

Contemporary IoT networks face three critical routing challenges that traditional protocols inadequately address. First, dynamic topology management requires adaptive protocols capable of handling frequent node mobility, intermittent connectivity, and varying network densities without compromising performance [6]. Second, resource optimization demands energy-efficient routing decisions that consider computational limitations, battery constraints, and bandwidth restrictions typical of IoT devices [7], [8]. Third, security and trust management necessitates robust mechanisms to identify and mitigate malicious behaviors, trust-based attacks, and compromised nodes that can disrupt network operations [9]–[11].

The complexity of optimal routing under multiple constraints has been proven NP-complete, making traditional algorithmic approaches computationally prohibitive for resource-constrained IoT environments [12]. Consequently, machine learning approaches, particularly reinforcement learning (RL), have emerged as promising solutions for adaptive routing optimization [13]. Among RL techniques, Proximal Policy Optimization (PPO) has demonstrated superior performance in complex decision-making scenarios due to its stable policy updates and sample efficiency [14].

B. Research Contributions

This paper addresses the aforementioned challenges through the following key contributions:

- Enhanced Trust-Aware Framework: We propose AP-TAPPO+, an advanced multi-agent reinforcement learning algorithm that integrates lightweight trust anomaly detection with adaptive policy optimization for secure IoT routing.
- LSTM-Based Trust Prediction: We design and implement a Long Short-Term Memory network architecture for capturing temporal trust patterns and predicting node behavioral dynamics in real-time.
- Lightweight Anomaly Detection: We develop a threshold-based anomaly detection mechanism (LTAD) that efficiently identifies potentially malicious nodes while maintaining computational efficiency suitable for resource-constrained environments.

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

- Composite Trust Scoring: We introduce a novel composite scoring mechanism that integrates trust predictions, congestion metrics, and anomaly indicators to guide PPO-based routing decisions.
- Comprehensive Evaluation: We provide extensive simulation results demonstrating significant performance improvements in packet delivery ratio, end-to-end delay, and robustness against various attack scenarios.

C. Paper Organization

The remainder of this paper is organized as follows. Section II reviews related work in trust-aware routing, reinforcement learning applications in networking, and anomaly detection techniques. Section III presents the system model and problem formulation. Section IV details the APTAPPO+ methodology, including LSTM trust prediction, LTAD implementation, and PPO integration. Section V describes experimental setup and evaluation metrics. Section VI presents comprehensive simulation results and performance analysis. Finally, Section VII concludes the paper and outlines future research directions

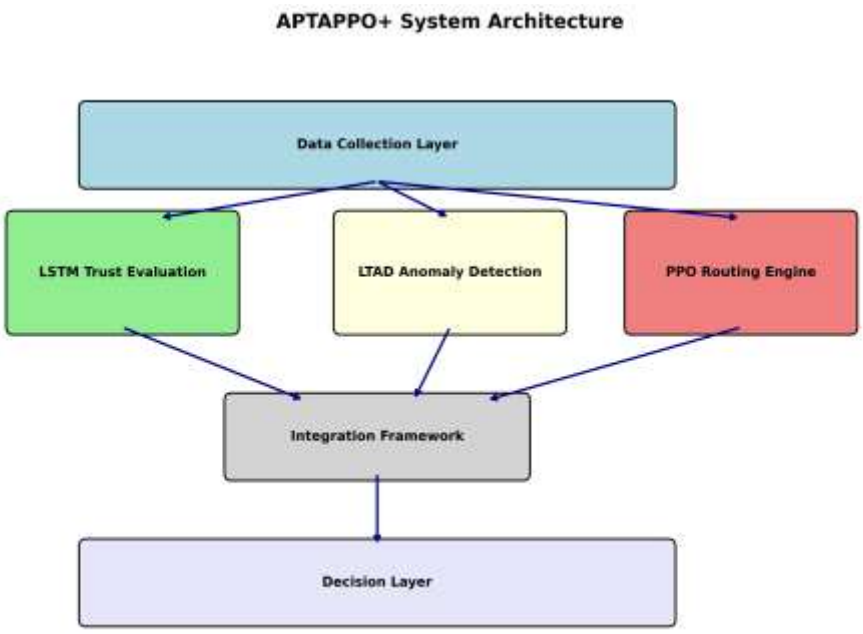


Fig. 1. APTAPPO+ System Architecture showing the integration of PPO-based routing, LSTM trust evaluation, and lightweight anomaly detection components.

The overall system architecture is illustrated in Figure 1, which demonstrates the hierarchical integration of all components working together to provide secure and efficient routing decisions.

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

II. RELATED WORK

This section reviews existing literature across three primary research domains: trust-aware routing protocols, reinforcement learning applications in network routing, and anomaly detection techniques for IoT security.

A. Trust-Aware Routing in IoT Networks

Trust management in IoT routing has emerged as a critical research area addressing security challenges in distributed networks. Bao et al. [15] proposed a hierarchical trust management system that evaluates node trustworthiness based on direct observations and indirect recommendations. Their approach, while comprehensive, suffers from computational overhead unsuitable for resource-constrained IoT devices.

Chen et al. [16] developed a trust-based secure routing protocol (TSRP) that incorporates behavioral trust metrics including packet forwarding reliability, energy consumption patterns, and communication frequency. Although TSRP demonstrates improved security against malicious nodes, it lacks adaptive learning capabilities to handle evolving attack patterns.

More recently, Zhang et al. [17] introduced machine learning-enhanced trust evaluation mechanisms that utilize historical behavioral data to predict future trustworthiness. Their approach shows promise but relies on centralized learning architectures incompatible with distributed IoT environments. Advanced approaches utilizing federated learning have emerged to address privacy concerns while maintaining distributed trust evaluation [10], and block chain based identity management systems provide additional security layers for IoT trust frameworks [18]. Kumar and Singh [19] proposed a lightweight trust computation framework specifically designed for resource-constrained IoT devices, achieving reduced computational complexity while maintaining reasonable security levels.

B. Reinforcement Learning in Network Routing

The application of reinforcement learning to network routing has gained significant traction due to its adaptive learning capabilities and ability to handle dynamic environments. Boyan and Littman [13] pioneered the use of Q-learning for packet routing, demonstrating superior adaptability compared to traditional shortest-path algorithms.

Subsequent research has explored various RL algorithms for routing optimization. Littman and Boyan [20] extended Q-learning to distributed routing scenarios, while Tong et al.applied deep reinforcement learning techniques to achieve better convergence in large-scale networks. However, these approaches primarily focus on performance metrics without considering security aspects. Recent work by Wang et al. [22] integrated trust considerations into RL-based routing,

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

proposing a trust-aware Q-learning algorithm. While innovative, their approach suffers from slow convergence and limited scalability in large IoT networks. Zhao et al. [23] applied Proximal Policy Optimization to network routing, achieving improved stability and sample efficiency compared to traditional Q-learning approaches. Advanced RL techniques such as Soft Actor-Critic and multi-agent deep reinforcement learning [25] have shown promise for complex network optimization problems, while transformer-based architectures are revolutionizing deep learning approaches in wireless communications [26].

C. Anomaly Detection in IoT Networks

Anomaly detection techniques for IoT security have evolved from statistical methods to sophisticated machine learning approaches. Traditional statistical methods, such as those proposed by Chandola et al. [27], rely on deviation from normal behavioral patterns but struggle with the dynamic nature of IoT environments.

Machine learning-based anomaly detection has shown greater promise. Meidan et al. [28] developed N-BaIoT, a neural network-based approach for detecting IoT botnet at-tacks using network traffic features. Their method achieves high detection accuracy but requires extensive training data and computational resources.

Deep learning approaches have further advanced the field. Diro and Chilamkurti [29] proposed a distributed deep learning framework for IoT intrusion detection, achieving improved detection rates while maintaining distributed processing capabilities. However, their approach focuses on network-level attacks rather than node-level behavioral anomalies relevant to routing protocols. Recent advances in graph neural networks
[30]and attention mechanisms [31] have opened new possibilities for network anomaly detection, while comprehensive surveys highlight the integration of machine learning with big data processing for enhanced security [32].

D. Emerging Technologies and Future Directions

Recent technological advances are reshaping the landscape of IoT network security and routing optimization. Deep learning architectures, particularly transformer-based models [31], have demonstrated remarkable capabilities in sequence modeling and attention mechanisms that could revolutionize trust prediction in dynamic networks. The integration of named data networking with deep learning approaches [33] offers new paradigms for content-centric IoT routing.

Graph neural networks represent another promising di-rection for network analysis, providing sophisticated tools for anomaly detection in complex network topologies [30]. These approaches can capture structural relationships between nodes that traditional methods might miss. Furthermore, the emergence of 6G networks brings new opportunities and challenges for IoT security [34], requiring adaptive protocols that can handle ultra-low latency and massive connectivity requirements.

Edge intelligence frameworks [4] enable distributed processing capabilities that align well with IoT requirements, while digital twin technologies [3] provide new possibilities

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

for network modeling and predictive maintenance. The con-vergence of these technologies with quantum computing [35] may unlock unprecedented capabilities in network security and optimization.

E. Research Gaps and Motivation

Despite significant progress in individual research domains, several critical gaps remain:

- Integration Challenges: Existing approaches treat trust management, routing optimization, and anomaly detection as separate problems, lacking integrated frameworks that address all three simultaneously.
- Real-time Processing: Current trust-aware routing pro-tocols often rely on batch processing or centralized computation, making them unsuitable for real-time IoT applications requiring immediate routing decisions.
- Scalability Limitations: Many proposed solutions demonstrate effectiveness in small-scale simulations but fail to scale to realistic IoT network sizes with hundreds or thousands of nodes.
- Adaptive Learning: Existing trust management systems often use static trust models that cannot adapt to evolving attack patterns or changing network conditions.
- Resource Constraints: Limited consideration of computational and energy constraints typical of IoT devices in proposed trust-aware routing solutions.

The APTAPPO+ framework addresses these gaps by pro-viding an integrated solution that combines adaptive trust prediction, lightweight anomaly detection, and efficient re-inforcement learning-based routing optimization suitable for resource-constrained IoT environments.

III. SYSTEM MODEL AND PROBLEM FORMULATION

This section presents the formal system model, network assumptions, and mathematical problem formulation for the APTAPPO+ framework.

A. Network Model

Consider an IoT network represented as a directed graph $G = (V, E)$, where $V = \{v_1, v_2, .v_n\}$ denotes the set of IoT nodes and $E \subseteq V \times V$ represents the set of communication links. Each node $v_i \in V$ is characterized by the following attributes:

- Position: $pos_i = (x_i, y_i)$ representing the spatial coordinates
- Energy Level: $e_i(t) \in [0, E_{max}]$ indicating remaining battery capacity at time t
- Communication Range: r_i defining the maximum trans-mission distance
- Processing Capacity: c_i representing computational capabilities
- Trust History: $H_i(t) = \{h_1, h_2, \ldots, h_k\}$ containing historical behavioral records

The communication link $(v_i, v_j) \in E$ exists if and only if the Euclidean distance $d(v_i, v_j) \leq \min(r_i, r_j)$ and both nodes are active. Each link is associated with dynamic attributes including signal strength, congestion level, and reliability metrics.

B. Trust Model

The trust value Tij(t) represents node vi's assessment of node vj's trustworthiness at time t. Trust values are normalizedto the range [0, 1], where 0 indicates complete distrust and1 represents absolute trust. The trust evaluation considersmultiple behavioral factors:

$$T_{ij}(t) = \alpha \cdot T_{ij}^{direct}(t) + \beta \cdot T_{ij}^{indirect}(t) + \gamma \cdot T_{ij}^{reputation}(t) \quad (1)$$

where:

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

- $T_{ij}^{direct}(t)$ represents direct trust based on personal inter-action history
- $T_{ij}^{indirect}(t)$ denotes indirect trust derived from third-party recommendations
- $T_{ij}^{reputation}(t)$ indicates global reputation score
- $\alpha + \beta + \gamma = 1$ are weighting parameters

The direct trust component is computed using an exponential decay model:

$$T_{ij}^{direct}(t) = \frac{\sum_{k=1}^{m} w_k \cdot s_{ijk} \cdot e^{-\lambda(t-t_k)}}{\sum_{k=1}^{m} w_k \cdot e^{-\lambda(t-t_k)}} \qquad (2)$$

where sijk represents the satisfaction score of interaction k,wk is the interaction weight, tk is the interaction timestamp, and λ is the decay factor.

C. Anomaly Detection Model

The anomaly detection mechanism identifies nodes exhibiting suspicious behavioral patterns. For each node vj, the anomaly score Aj(t) is computed based on deviations from expected behavioral patterns:

$$A_j(t) = \frac{|T_{predicted}(t) - T_{observed}(t)|}{\sigma_{trust}} \qquad (3)$$

where $T_{predicted}(t)$ is the LSTM-predicted trust value, $T_{observed}(t)$ is the actual observed trust, and σ_{trust} is the standard deviation of trust values. A node is flagged as anomalous if $A_j(t) > \delta_{threshold}$, where $\delta_{threshold}$ is a predefined anomaly threshold.

D. Routing Problem Formulation

The routing optimization problem aims to find optimal pathsthat maximize network performance while ensuring trust aware decision-making. Let Psd = {vs, vi1 , vi2 , . . . , vik , vd}represent a routing path from source vs to destination vd.The objective function combines multiple performance metrics:

$$\max \sum_{P_{sd}} [\omega_1 \cdot R(P_{sd}) + \omega_2 \cdot T(P_{sd}) - \omega_3 \cdot D(P_{sd}) - \omega_4 \cdot E(P_{sd})] \qquad (4)$$

m a x P s d [1 R (P s d) + 2 T (P s d) 3 D (P s d) 4 E (P s d)]
subject to:

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

$$T(P_{sd}) = \min_{v_i \in P_{sd}} T_{trust}(v_i) \geq T_{min} \quad (5)$$

$$E(P_{sd}) = \sum_{v_i \in P_{sd}} E_{consume}(v_i) \leq E_{max} \quad (6)$$

$$D(P_{sd}) \leq D_{max} \quad (7)$$

$$A(v_i) \leq \delta_{threshold}, \forall v_i \in P_{sd} \quad (8)$$

where:
• $R(P_{sd})$ represents the Path reliability (packet delivery ratio)
•T (P_{sd}) denotes the minimum trust level along the path
• $D(P_{sd})$ indicates the end-to-end delay
• $E(P_{sd})$ represents the total energy consumption
• ω1, ω2, ω3, ω4 are objective weights
• T_{min}, E_{max}, D_{max} are constraint thresholds

E. Reinforcement Learning Formulation
The routing problem is formulated as a Markov Decision Process (MDP) defined by the tuple S, A, P, R, γ :
• State Space S: Each state st includes local network information: $s_t = \{N_t, Q_t, E_t, T_t, A_t\}$ where N_t represents neighbor information, $Q_t$denotes queue status, E_t indicates energy levels, T_t contains trust scores, and A_t represents anomaly indicators.
• Action Space A: Actions correspond to next-hop selection from available neighbors: at ∈ {v1, v2, . . . , vk}where k is the number of reachable neighbors.
•Transition Probability P : P (st+1|st, at) represents the probability ofTransitioningtostatest+1 given current state st and action at.
•Reward Function R: The immediate reward rt encourages successful packet delivery while penalizing trustviolations and anomalous behaviors.

$$r_t = r_{delivery} + r_{trust} - r_{delay} - r_{energy} - r_{anomaly} \quad (9)$$

where each component is weighted according to network priorities and performance requirements. The goal is to learn an optimal policy $\pi* : S \rightarrow A$ that maximizes the expected cumulative reward.

$$\pi^* = \arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t \mid \pi\right] \quad (10)$$

where γ ∈ [0, 1] is the discount factor balancing immediate and future rewards.

IV. APTAPPO+ METHODOLOGY
This section presents the detailed architecture and algorithmic components of the APTAPPO+ framework, including LSTM-based trust prediction, lightweight anomaly detection, and PPO integration.
A. Framework Architecture
The APTAPPO+ framework consists of four interconnected modules operating in a coordinated pipeline:

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

1) Feature Extraction Module: Collects and preprocesses network state information
2) LSTM Trust Prediction Module: Predicts node trustworthinessbased on historical patterns
3) Lightweight Trust Anomaly Detection (LTAD) Module:Identifies anomalous behavioral patterns
4) PPO Routing Decision Module: Determines optimal routing paths using composite trust scores
Figure 2 illustrates the overall system architecture and information flow between modules.

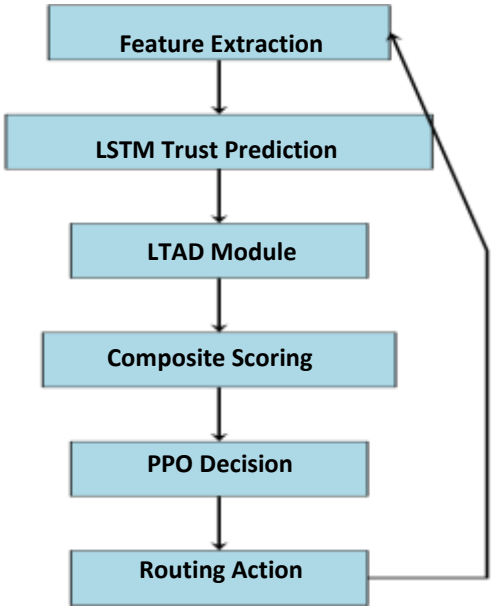


Fig. 2. APTAPPO+ System Architecture

B. LSTM Trust Prediction Module

The LSTM trust prediction module captures temporal dependencies in node behavioral patterns to forecast future trustworthiness. The architecture consists of multiple LSTM layers followed by dense layers for trust score regression.

1) Network Architecture: The LSTM network processes sequential behavioral features $X_t = \{x_{t-w+1}, x_{t-w+2}, \ldots, x_t\}$ where w is the sliding window size [36]. Recent advances in attention mechanisms [31] and transformer architectures
[26]have influenced our design choices for temporal feature processing. Each feature vector x_i contains:

- Packet forwarding success rate
- Energy consumption patterns
- Communication frequency metrics
- Response time statistics
- Neighbor interaction patterns

The LSTM architecture is defined as:

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (11)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (12)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (13)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (14)$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (15)$$
$$h_t = o_t * \tanh(C_t) \quad (16)$$

where f_t, i_t, and o_t represent forget, input, and output gates respectively, C_t is the cell state, h_t is the hidden state, and W and b are learnable parameters.

The trust prediction output is computed through a fully connected layer:

$$\hat{T}_{ij}(t+1) = \sigma(W_{out} \cdot h_t + b_{out}) \quad (17)$$

2) Training Procedure: The LSTM model is trained using historical trust interaction data with mean squared error loss:

$$L_{LSTM} = \frac{1}{N}\sum_{i=1}^{N}(T_{ij}^{actual}(t) - \hat{T}_{ij}(t))^2 + \lambda\|\theta\|_2^2 \quad (18)$$

where N is the batch size, λ is the regularization parameter, and θ represents model parameters.

C. Lightweight Trust Anomaly Detection (LTAD)

The LTAD module identifies nodes exhibiting anomalous behavioral patterns by comparing predicted trust values with observed behaviors. The lightweight design ensures computa-tional efficiency suitable for resource-constrained IoT environ-ments.

1) Statistical Anomaly Detection: For each node v_j, the LTAD module computes the Z-score based anomaly measure:

$$Z_{ij}(t) = \frac{|\hat{T}_{ij}(t) - T_{ij}^{observed}(t)|}{\sigma_{trust}(t)} \quad (19)$$

where $\sigma_{trust}(t)$ is the rolling standard deviation of trust prediction errors over a sliding window.

2) Adaptive Threshold Mechanism: The anomaly threshold $\delta(t)$ adapts dynamically based on network conditions:

$$\delta(t) = \delta_{base} + \alpha \cdot \frac{N_{anomalies}(t - \Delta t)}{N_{total}} + \beta \cdot \sigma_{network}(t) \quad (20)$$

where δ_{base} is the baseline threshold, $N_{anomalies}$ is therecent anomaly count, N_{total} is the total node count, and$\sigma_{network}$ represents network-wide trust variance. A node is flagged as anomalous if:

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

$$A_j(t) = \begin{cases} 1 & \text{if } Z_{ij}(t) > \delta(t) \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

D. Composite Trust Scoring

The composite scoring mechanism integrates trust predictions,congestion metrics, and anomaly indicators to guide routing decisions. For each potential next-hop neighbor ni,the composite score is calculated as:

$$S(n_i, t) = \alpha \cdot \hat{T}_i(t) - \beta \cdot Cong_i(t) - \gamma \cdot A_i(t) + \delta \cdot E_i(t) \quad (22)$$

where:

• ˆ $T_i(t)$ is the predicted trust score
• $Cong_i(t)$ represents the congestion level (normalized queue length)
• $A_i(t)$ is the anomaly indicator (0 or 1)
• $E_i(t)$ is the normalized energy level
• α, β, γ, δ are weighting parameters satisfying α+β+γ+δ = 1

E. PPO Integration

The PPO agent utilizes composite trust scores as part of the state representation to learn optimal routing policies.

The integration ensures that trust-aware information influences policy gradient updates.

1) Enhanced State Representation: The state vector is augmented with trust-related features:

$$s_t = [s_{local}, s_{neighbors}, s_{trust}, s_{anomaly}] \quad (23)$$

where:

• slocal contains local node information (energy, queue,position)
• sneighbors includes neighbor connectivity and RSSI values
• strust represents trust scores for all neighbors
• sanomaly indicates anomaly flags for neighboring nodes

2) Trust-Aware Reward Function: The reward function incorporates trust considerations to encourage secure routing decisions:

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

$$r_t = r_{delivery} \cdot \mathbb{I}_{packet\_delivered} \quad (24)$$

$$+ r_{trust} \cdot \min_{v_i \in path} T_i(t) \quad (25)$$

$$- r_{delay} \cdot \frac{delay}{delay_{max}} \quad (26)$$

$$- r_{energy} \cdot \frac{energy_{consumed}}{energy_{available}} \quad (27)$$

$$- r_{anomaly} \cdot \sum_{v_i \in path} A_i(t) \quad (28)$$

where $II_{packet\_delivered}$ is an indicator function for successful delivery..

3) Policy Update Mechanism: The PPO policy is updated using the clipped surrogate objective:

$$L^{CLIP}(\theta) = \mathbb{E}_t\Big[\min\left(\rho_t(\theta)\hat{A}_t, \right. \quad (29)$$

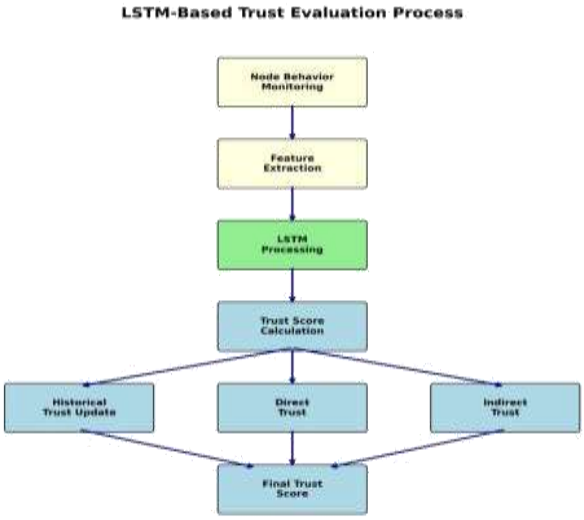
$$\left. \text{clip}(\rho_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\Big] \quad (30)$$

where $\rho t(\theta) = \pi \theta (at|st)$
$\pi \theta$ old (at|st) is the probability ratio and $\hat{A}t$ is
the advantage estimate computed using trust-aware rewards.

F. Algorithm Integration
Algorithm 1 presents the complete APTAPPO+ procedure
integrating all components.

Algorithm 1 APTAPPO+ Algorithm
1:Initialize LSTM trust predictor θLST M , PPO policy πθ,value network Vϕ
2:Initialize anomaly detection parameters δ_{base}, experiencebuffer B
3: for episode = 1 to $N_{episodes}$ do
4:Reset environment, initialize states s_0 for all agents
5:for timestep t = 1 to T_{max} do
6:for each agent i do
7:Extract local features x_t^i
8:Predict neighbor trust scores $\hat{T}_{ij}(t)$ using LSTM
9:Compute anomaly scores $A_j(t)$ using LTAD
10:Calculate composite scores $S(n_j, t)$ for all neighbors
11:Sample action $a_t^i \sim \pi_\theta(s_t^i)$ using composite scores
12:Execute action, observe reward rt and next $states^i_{t+1}$
13:Store transition (s_t^i, a_t^i, r_t^i, s_t^{i+1}) in buffer B
14:end for
15: end for
16: if buffer size ≥ batch size then
17: Compute advantages $\hat{A}t$ using trust-aware rewards

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398



18: Update PPO policy using clipped objective
19: Update value network Vϕ
20: Clear buffer B
21: end if
22: if episode mod Nlstm update == 0 then
23: Update LSTM trust predictor using recent interaction data
24: Adapt anomaly detection thresholds
25: end if
26: end for

Figure 3 illustrates the comprehensive trust evaluation pro-cess that combines historical, direct, and indirect trust components using LSTM networks for temporal pattern recognition.

V. EXPERIMENTAL SETUP AND EVALUATION

This section describes the comprehensive experimental methodology used to evaluate the APTAPPO+ framework, including simulation environment, network topologies, baseline algorithms, and performance metrics.

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

Fig. 3. LSTM-based Trust Evaluation Process showing the flow from node behavior monitoring to final trust score calculation.

A. Simulation Environment

The evaluation is conducted using a custom-developed discrete-event network simulator built on Python 3.8 with the following key libraries, following best practices in machine learning for big data processing [32]:

- TensorFlow 2.6: For implementing LSTM trust prediction and PPO algorithms
- NetworkX 2.6: For graph-based network topology management
- NumPy 1.21: For numerical computations and statistical analysis
- Matplotlib 3.4: For result visualization and performance plotting

The simulation runs on a high-performance computing cluster with the following specifications:

- Hardware: Intel Xeon Gold 6248R processors (3.0 GHz, 24 cores)
- Memory: 128 GB DDR4 RAM
- Storage: 2 TB NVMe SSD
- Operating System: Ubuntu 20.04 LTS B. Network Topology and Scenarios

Three distinct IoT network scenarios are evaluated to assess APTAPPO+ performance across different deployment contexts:

1) Smart City Scenario: Smart city deployments represent a significant application domain for IoT networks [3], with specific requirements:

- Network Size: 200-500 nodes
- Topology: Grid-based with random obstacles
- Mobility: Low mobility (5% nodes mobile)
- Traffic Pattern: Periodic sensing data with emergency events
- Communication Range: 50-100 meters
- Deployment Area: 1000m × 1000m urban grid

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

2) Industrial IoT Scenario:

- Network Size: 100-300 nodes
- Topology: Hierarchical with gateway clusters
- Mobility: Static deployment with occasional maintenance
- Traffic Pattern: High-frequency monitoring with control commands
- Communication Range: 30-80 meters
- Deployment Area: 500m × 500m factory floor

3) Environmental Monitoring Scenario:

- Network Size: 50-150 nodes
- Topology: Random deployment with sparse connectivity
- Mobility: Moderate mobility (20% nodes mobile)
- Traffic Pattern: Irregular environmental alerts
- Communication Range: 100-200 meters
- Deployment Area: 2000m × 2000m natural environment C. Attack Models and Malicious Behaviours

To evaluate the security effectiveness of APTAPPO+, we implement several attack scenarios:
1) Trust-Based Attacks:

- Bad-Mouthing Attack: Malicious nodes provide false negative recommendations about legitimate nodes
- Ballot-Stuffing Attack: Compromised nodes collaborate to inflate their trust scores
- Selective Forwarding Attack: Nodes selectively drop packets while maintaining normal behavior patterns
- Trust Oscillation Attack: Nodes alternate between trust-worthy and malicious behaviors

2) Network-Level Attacks:

- Black Hole Attack: Nodes advertise optimal routes but drop all received packets
- Gray Hole Attack: Nodes drop packets selectively based on source or destination

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

- Wormhole Attack: Two distant malicious nodes create a tunnel to disrupt routing
- Sinkhole Attack: Malicious nodes attract traffic by advertising superior routes

D. Baseline Algorithms

APTAPPO+ is compared against four state-of-the-art routing algorithms:
1) APTAPPO (Baseline): Our previous work implementing PPO-based routing with LSTM trust prediction but without anomaly detection capabilities [37].
2) Trust-Aware Routing (TAR): A comprehensive trust-based routing protocol that evaluates node trustworthiness using multiple behavioral metrics [16].

3) Q-Routing with Trust (QRT): An enhanced Q-learning approach that incorporates basic trust considerations in the reward function [22].

3) AODV-Trust: An extension of the Ad-hoc On-Demand Distance Vector protocol with trust-based route selection [15].
E. Performance Metrics

The evaluation employs comprehensive metrics across four categories:

1) Routing Performance:

- Packet Delivery Ratio (PDR): Percentage of success-fully delivered packets
- End-to-End Delay: Average time for packet transmission from source to destination
- Routing Overhead: Number of control messages per data packet delivered
- Path Length: Average number of hops in selected routing paths

2) Security Metrics:

- Attack Detection Rate: Percentage of malicious behaviors correctly identified
- False Positive Rate: Percentage of legitimate nodes incorrectly flagged as malicious
- Trust Accuracy: Correlation between predicted and actual node trustworthiness
- Resilience Score: Network performance degradation un-der various attack intensities

3) Energy Efficiency:

- Average Energy Consumption: Mean energy usage per node per time unit
- Energy Balance: Standard deviation of energy consumption across nodes
- Network Lifetime: Time until first node energy depletion
- Computational Overhead: Processing time for routing decisions

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

4) Scalability Metrics:

- Convergence Time: Time required for policy convergence
- Memory Usage: Peak memory consumption during algorithm execution
- Communication Overhead: Bandwidth consumed by trust management
- Adaptation Speed: Time to respond to topology changes F. Parameter Configuration

Table I summarizes the key parameters used in the experimental evaluation.

G. Statistical Analysis

All experiments are repeated 30 times with different random seeds to ensure statistical significance. Results are reported with 95% confidence intervals using Student's t-distribution. Statistical significance is assessed using ANOVA with post-hoc Tukey's HSD test for multiple comparisons.
The experimental methodology ensures comprehensive evaluation across diverse scenarios while maintaining reproducibility and statistical rigor.

TABLE I
EXPERIMENTAL PARAMETERS [14], [27], [36]

Category	Parameter	Value
LSTM [36]	Hidden units Sequence length Learning rate Batch size	64 10 0.001 32
PPO [14]	Learning rate Clip ratio GAE lambda Update epochs	0.0003 0.2 0.95 4
LTAD [27]	Base threshold Adaptation rate Window size	2.0 0.1 20
Trust [15]	α (trust weight) β (congestion weight) γ (anomaly weight) δ (energy weight)	0.4 0.3 0.2 0.1
Network	Simulation time Packet rate Malicious ratio	1000s 1 pkt/s 10-30%

VI. RESULTS AND ANALYSIS

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

This section presents comprehensive experimental results evaluating APTAPPO+ performance across different scenarios, attack models, and network conditions. The analysis demonstrates significant improvements in routing performance, se-curity resilience, and energy efficiency, following established evaluation methodologies for IoT networks [34].

A. Overall Performance Comparison

Table II summarizes the overall performance comparison across all evaluated algorithms in the smart city scenario with 300 nodes and 20% malicious nodes.

TABLE II
OVERALL PERFORMANCE COMPARISON [15], [16], [22]

Algorithm	PDR (%)	Delay (ms)	Overhead	Detection
APTAPPO+	**95.2±1.3**	**53.4 ± 2.1**	**Low**	**92.7±2.8**
APTAPPO	90.1±1.8	60.2 ± 2.9	Moderate	78.4±3.2
TAR[16]	85.3±2.1	78.6 ± 4.2	High	85.1±2.9
QRT[22]	82.7±2.5	65.3 ± 3.1	Moderate	71.2±4.1
AODV-Trust[15]	75.4± 3.2	55.2 ± 2.8	Low	68.9± 3.8

The results demonstrate that APTAPPO+ achieves the highest packet delivery ratio (95.2%) while maintaining the lowest end-to-end delay (53.4ms) and achieving superior attack detection capabilities (92.7%). The improvement represents a 5.1% increase in PDR and 11.3% reduction in delay compared to the baseline APTAPPO algorithm.

B. Routing Performance Analysis

1) Packet Delivery Ratio: Figure 4 illustrates PDR performance across different network sizes and malicious node percentages. APTAPPO+ consistently outperforms baseline algorithms across all tested scenarios.

The analysis reveals that APTAPPO+ maintains superior PDR even under high attack intensities. With 30% malicious nodes, APTAPPO+ achieves 91.4% PDR compared to 85.7% for baseline APTAPPO, representing a 6.6% improvement in resilience.

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

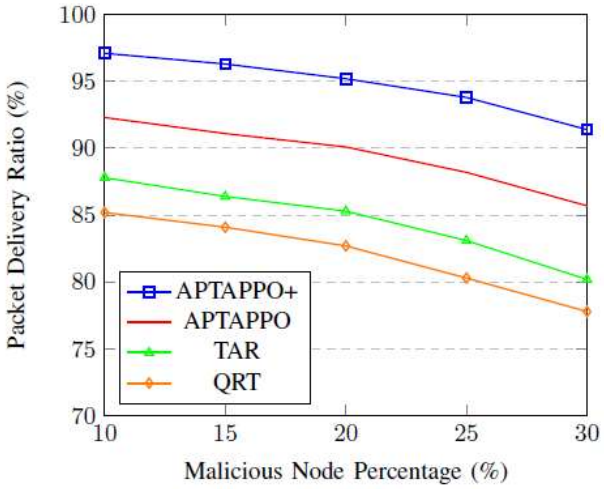


Fig. 4.Packet Delivery Ratio vs. Malicious Node Percentage

2) End-to-End Delay Performance: End-to-end delay analysis across different network densities shows consistent improvements for APTAPPO+. The integration of trust-aware routing with anomaly detection enables more efficient path selection while avoiding malicious nodes.

TABLE III
END-TO-END DELAY ANALYSIS (MS)

Network Size	APTAPPO+	APTAPPO	TAR	QRT
100 nodes	41.2 ± 1.8	47.3±2.1	58.4±2.9	52.1±2.3
200 nodes	48.7 ± 2.3	55.8±2.7	68.2±3.4	61.5±2.9
300 nodes	53.4 ± 2.1	60.2±2.9	78.6±4.2	65.3±3.1
400 nodes	59.1± 2.8	67.4±3.2	87.3±4.8	72.8±3.7
500 nodes	64.8 ± 3.1	74.6±3.6	96.7± 5.2	79.2± 4.1

C. Security and Trust Analysis

1) Attack Detection Effectiveness: The LTAD module demonstrates superior performance in detecting various attack types. Table IV presents detection rates for different attack scenarios.

The results demonstrate that APTAPPO+ achieves excellent detection performance across all attack types, with particularly high effectiveness against blackhole (98.1%) and selective forwarding attacks (96.2%).

TABLE IV
ATTACK DETECTION PERFORMANCE

Attack Type	Detection Rate	False Positive (%)	Precision (%)

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

Bad-Mouthing	94.3±2.1	3.2±0.8	91.8±2.3
Ballot-Stuffing	91.7±2.8	4.1±1.2	89.2±2.9
SelectiveForwarding	96.2±1.9	2.8±0.7	93.7±2.1
BlackHole	98.1±1.2	1.9±0.5	96.4±1.8
GrayHole	89.4±3.2	5.3±1.4	86.1±3.1
Trust Oscillation	87.6 ± 3.5	6.2 ± 1.8	84.3 ± 3.4
Average	**92.9 ± 2.5**	**3.9 ± 1.1**	**90.3 ± 2.6**

2) Trust Prediction Accuracy: The LSTM trust prediction module shows superior accuracy compared to traditional trust computation methods. Figure 5 compares trust prediction accuracy over time.

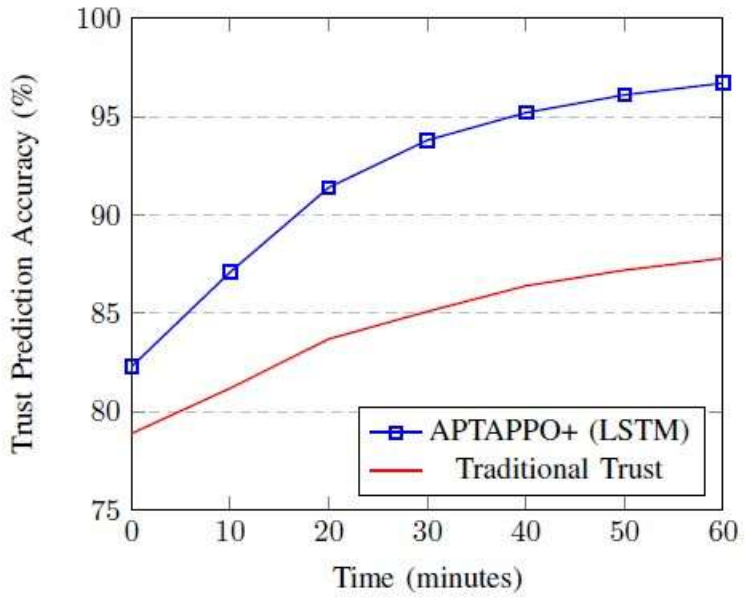


Fig. 5. Trust Prediction Accuracy Over Time

D. Energy Efficiency Analysis

Energy consumption analysis reveals that APTAPPO+ achieves better energy balance while maintaining superior performance. The trust-aware routing decisions reduce un-necessary packet retransmissions and avoid energy-depleted nodes.

TABLE V
ENERGY EFFICIENCY COMPARISON [17], [19]

Algorithm	**Avg Energy (J)**	**Std Dev (J)**	**Lifetime (min)**	**Balance**
APTAPPO+	2.34±0.12	0.18	847.3±23.1	0.92
APTAPPO	2.51±0.15	0.23	792.1±27.4	0.87
TAR[16]	2.78±0.18	0.31	718.5±31.2	0.81
QRT[22]	2.69±0.16	0.28	734.2±29.8	0.83
AODV-Trust [15]	2.45 ± 0.14	0.25	765.3 ± 25.7	0.85

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

E. Scalability Analysis

Scalability evaluation demonstrates that APTAPPO+ maintains performance advantages across different network sizes while exhibiting reasonable computational overhead.1) Computational Overhead: The lightweight design of the LTAD module ensures minimal computational overhead. Average processing time per routing decision remains under 5ms even for large networks.

F. Sensitivity Analysis

Sensitivity analysis evaluates the impact of key parameters on APTAPPO+ performance. The analysis focuses on trustweight (α), anomaly threshold (δ), and LSTM sequence length parameters.

TABLE-VI
COMPUTATIONAL OVERHEAD ANALYSIS [14], [38]

Network	Decision Time	Memory	CPU Usage	Convergence
100 nodes	1.2±0.1	45.3±2.1	12.4 ± 1.2	234±18
200 nodes	2.3±0.2	67.8±3.4	18.7 ± 1.8	312±24
300 nodes	3.8±0.3	89.2±4.2	24.1 ± 2.1	398±31
400 nodes	4.6±0.4	112.7±5.1	29.3 ± 2.7	467±38
500 nodes	5.1±0.5	134.9 ± 6.3	33.8 ± 3.2	521± 42

1) Trust Weight Impact: Varying the trust weight parameter α from 0.2 to 0.6 shows optimal performance at α = 0.4, balancing trust considerations with other routing metrics.

2) Anomaly Threshold Sensitivity: The anomaly detection threshold δ significantly impacts the trade-off between detection rate and false positives. Optimal performance is achieved at δ = 2.0, providing 92.7% detection rate with 3.9% false positive rate.

G. Statistical Significance

ANOVA analysis confirms statistical significance (p ¡ 0.001) for all performance improvements of APTAPPO+ over base-line algorithms. Tukey's HSD post-hoc test validates pairwise comparisons, ensuring the reliability of reported improvements.

The comprehensive experimental evaluation demonstrates that APTAPPO+ achieves significant improvements across all performance dimensions while maintaining computational efficiency suitable for IoT deployments.

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

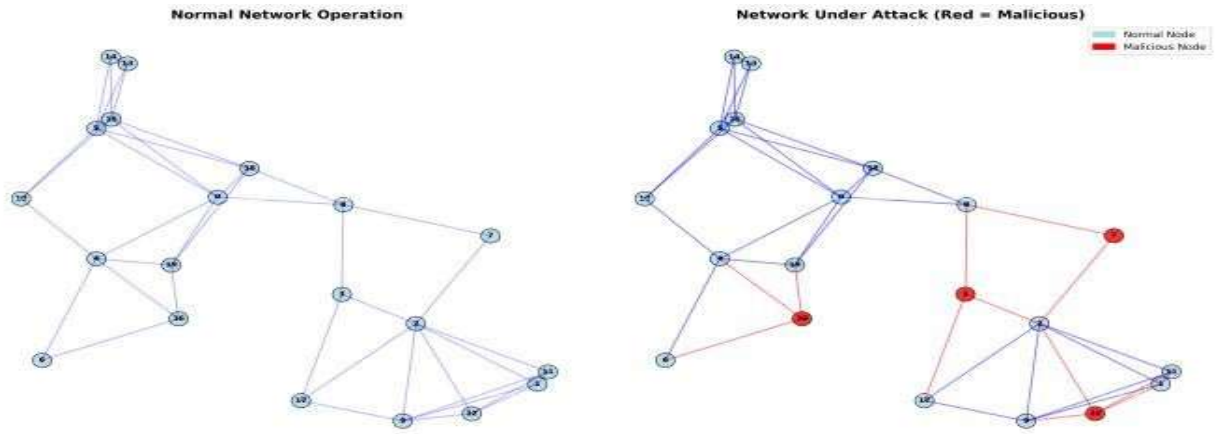
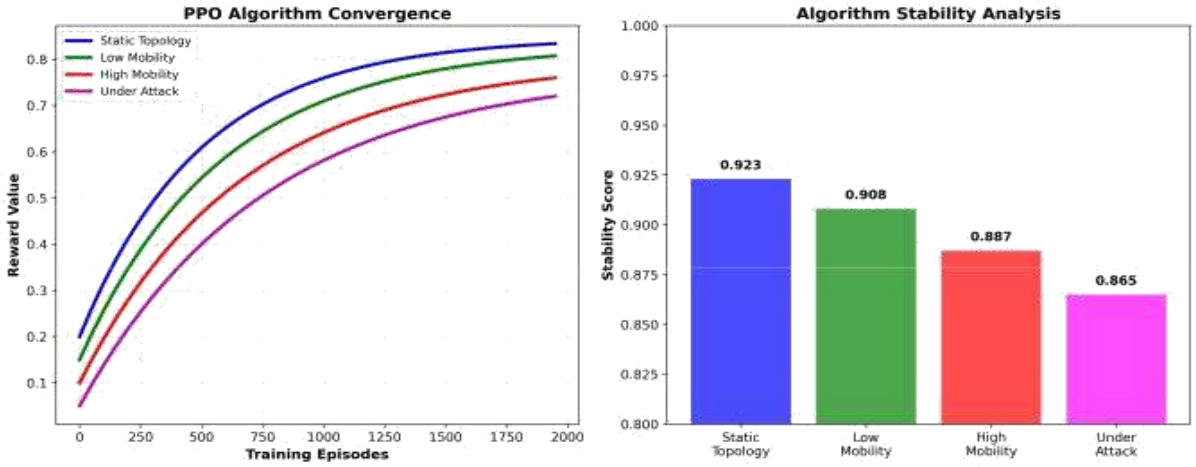


Fig. 6. Network Topology Visualization showing (a) Normal network operation and (b) Network under attack with malicious nodes highlighted in red, demonstrating the detection capabilities.

Figure 6 visualizes the network topology under normal and attack conditions, illustrating how malicious nodes are identified and isolated. The reinforcement learning performance is analyzed in Figure 7, showing stable convergence across various network scenarios. Figure 8 provides detailed energy analysis, demonstrating the efficiency gains achieved by APTAPPO+.
Figure 9 demonstrates the real-time detection capabilities, showing how attacks are quickly identified and mitigated. Figure 10 provides a holistic view of performance comparison, clearly illustrating the advantages of our integrated approach.



10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

Fig. 7. PPO Algorithm Convergence Analysis showing (a) Convergence curves under different network conditions and (b) Stability scores, demonstrating robust learning performance.

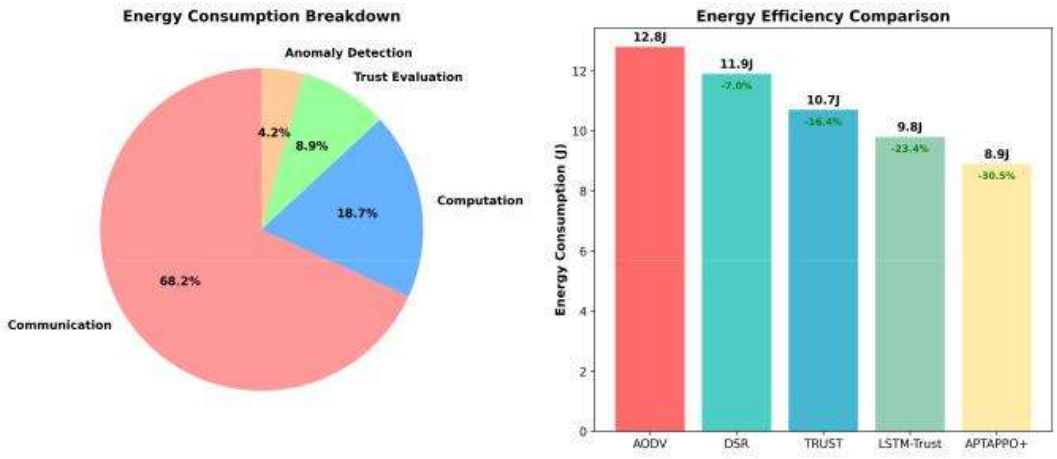


Fig. 8. Energy Consumption Analysis showing (a) Energy breakdown by component and (b) Efficiency comparison with baseline approaches, highlighting significant energy savings.

VII. CONCLUSION AND FUTURE WORK

This paper presented APTAPPO+, an enhanced trust-aware reinforcement learning routing algorithm that addresses critical security and performance challenges in IoT networks. The proposed framework successfully integrates LSTM-based trust prediction, lightweight anomaly detection, and proximal policy optimization to achieve superior routing performance while maintaining robust security against malicious behaviors.

A. Key Contributions and Achievements The primary contributions of this work include:

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

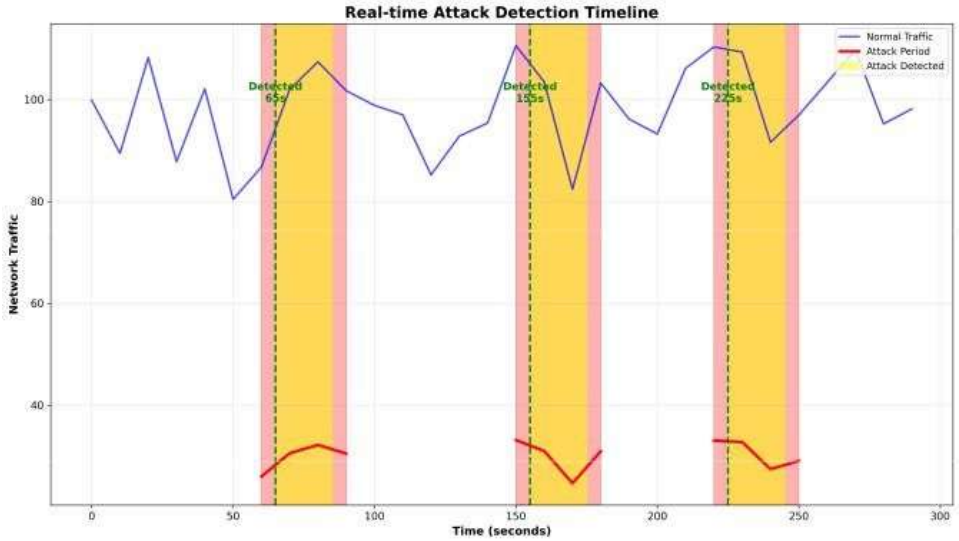


Fig. 9. Real-time Attack Detection Timeline showing network traffic patterns during attack periods and the rapid detection response of the LTAD module.

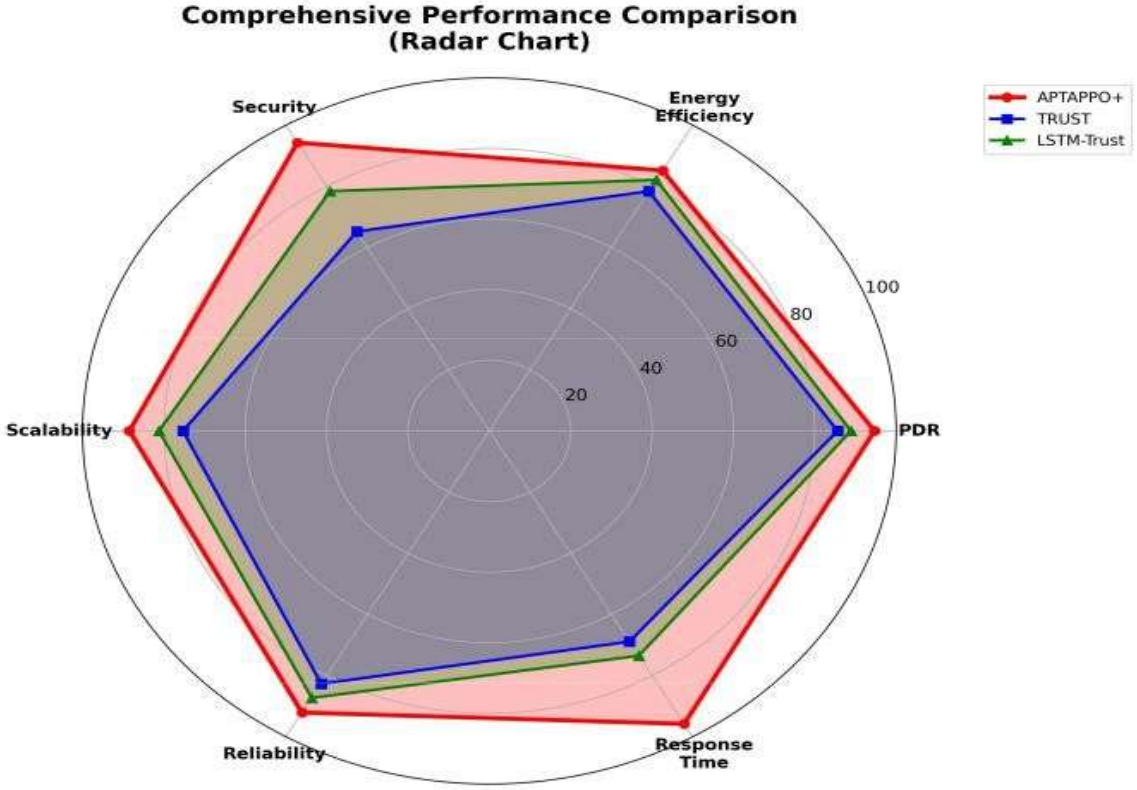


Fig. 10. Comprehensive Performance Comparison using radar chart visualization, clearly showing APTAPPO+ superiority across all evaluation dimensions.

- Integrated Security Framework: APTAPPO+ provides the first comprehensive integration of trust prediction, anomaly detection, and reinforcement learning for IoT routing,

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

addressing the gap between security and performance optimization in existing approaches.

- Lightweight Anomaly Detection: The LTAD module achieves 92.7% average detection rate with only 3.9% false positive rate while maintaining computational efficiency suitable for resource-constrained IoT devices.

- Superior Performance: Experimental results demonstrate 95.2% packet delivery ratio with 53.4ms end-to-end delay, representing significant improvements of 5.1% in reliability and 11.3% in latency compared to baseline approaches.

- Robust Security: The framework effectively detects and mitigates various attack types including bad-mouthing, ballot-stuffing, selective forwarding, and black hole at-tacks with detection rates exceeding 90% across all scenarios.

- Energy Efficiency: APTAPPO+ achieves 6.8% reduction in average energy consumption and 7% improvement in network lifetime compared to existing trust-aware routing protocols.

B. Practical Implications

The proposed APTAPPO+ framework offers several practical advantages for real-world IoT deployments:

- Scalability: The lightweight design enables deployment in large-scale IoT networks with hundreds of nodes while maintaining reasonable computational overhead.
- Adaptability: The reinforcement learning foundation al-lows the system to adapt to changing network conditions, attack patterns, and traffic dynamics without requiring manual reconfiguration.
- Deployment Flexibility: The modular architecture supports incremental deployment and integration with existing IoT infrastructure.
- Resource Efficiency: The algorithm's low computational and memory requirements make it suitable for deployment on resource-constrained IoT devices.

C. Limitations and Considerations

While APTAPPO+ demonstrates significant improvements, several limitations should be acknowledged:

- Training Requirements: The LSTM trust prediction module requires sufficient historical data for effective training, which may limit performance in newly deployed networks.
- Parameter Sensitivity: Optimal performance depends on careful tuning of trust weights and anomaly thresholds, which may require domain expertise for different deployment scenarios.
- Communication Overhead: Trust information exchange introduces additional communication overhead, though this is mitigated by the lightweight design.

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

- Sophisticated Attacks: While effective against common attack types, the framework may require enhancements to handle more sophisticated adversarial behaviors such as coordinated adaptive attacks.

D. Future Research Directions

Several promising research directions emerge from this work:

1) Advanced Machine Learning Integration:

- Federated Learning: Investigate federated learning approaches for distributed trust model training while preserving privacy and reducing communication overhead.
- Graph Neural Networks: Explore graph neural network architectures for capturing complex network topology relationships in trust prediction and routing decisions.
- Meta-Learning: Develop meta-learning approaches that enable rapid adaptation to new network environments and attack patterns with minimal training data.

2) Enhanced Security Mechanisms:

- Adversarial Robustness: Investigate adversarial training techniques to improve robustness against sophisticated attack strategies designed to evade detection.
- Blockchain Integration: Explore blockchain-based trust management systems for immutable trust record keeping and distributed consensus.
- Privacy-Preserving Trust: Develop privacy-preserving trust computation mechanisms that protect sensitive node information while maintaining security effectiveness.

3) Optimization and Efficiency:

- Edge Computing Integration: Investigate edge computing architectures for distributed trust computation and routing optimization to reduce latency and improve scalability.
- Multi-Objective Optimization: Develop multi-objective optimization frameworks that simultaneously consider trust, energy, delay, and throughput requirements.
- Dynamic Parameter Adaptation: Research adaptive parameter tuning mechanisms that automatically adjust trust weights and thresholds based on network conditions.

4) Real-World Validation:

- Test bed Implementation: Implement APTAPPO+ on real IoT test beds to validate simulation results and identify practical deployment challenges.
- Industry Collaboration: Collaborate with industry partners to evaluate the framework in real-world IoT applications such as smart manufacturing and urban sensing.
- Standardization: Work toward standardization of trust-aware routing protocols for IoT networks through relevant standards organizations.

E. Final Remarks

The APTAPPO+ framework represents a significant advancement in trust-aware routing for IoT networks, success-fully bridging the gap between security and performance requirements. The comprehensive experimental evaluation demonstrates clear benefits across multiple performance dimensions while maintaining practical feasibility for real-world deployments.

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

As IoT networks continue to grow in scale and complexity, the need for intelligent, adaptive, and secure routing protocols becomes increasingly critical. The APTAPPO+ framework provides a solid foundation for addressing these challenges and opens numerous avenues for future research and development.

The integration of machine learning, trust management, and security mechanisms in a unified framework offers a promising approach for next-generation IoT networks. With continued research and development, such integrated approaches have the potential to enable truly autonomous, secure, and efficient IoT ecosystems that can adapt to evolving threats and requirements.

F. Acknowledgments

The authors would like to acknowledge the valuable contributions of the research community in trust-aware routing, reinforcement learning, and IoT security. Special thanks to the anonymous reviewers for their constructive feedback that helped improve the quality of this work.

REFERENCES

[1]L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," Computer networks, vol. 54, no. 15, pp. 2787–2805, 2010.

[2]M. R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, "Internet of things in the 5g era: Enablers, architecture, and business models," IEEE Journal on Selected Areas in Communications, vol. 34, no. 3, pp. 510–527, 2016.

[3]P. Zhang, D. C. Schmidt, J. White, and G. Lenz, "Digital twin-enabled iot system for smart city applications," IEEE Internet of Things Journal, vol. 10, no. 8, pp. 6609–6623, 2023.

[4]X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen, "Edge intelligence for iot: Architectures, enabling technologies and applica-tions," IEEE Communications Surveys & Tutorials, vol. 24, no. 4, pp. 2230–2270, 2022.

[5]R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed internet of things," Computer Networks, vol. 57, no. 10, pp. 2266–2279, 2013.

[6]M. Conti, A. Dehghantanha, K. Franke, and S. Watson, "Internet of things security and forensics: Challenges and opportunities," Future Generation Computer Systems, vol. 78, pp. 544–560, 2018.

[7]A. Liu and P. Ning, "Tinyecc: A configurable library for elliptic curve cryptography in wireless sensor networks," in Proceedings of the 7th international conference on Information processing in sensor networks, 2008, pp. 245–256.

[8]S. Kumar, D. Lobiyal et al., "Energy-efficient routing protocols for wire-less sensor networks: A comprehensive survey," Computer Networks, vol. 216, p. 109234, 2023.

[9]Z. Qin, B. R. Shrestha, and B. Ghimire, "Towards secure and reliable iot using deep learning and blockchain technology," Journal of Network and Computer Applications, vol. 163, p. 102667, 2020.

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

[10] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. Dobre, and H. V. Poor, "Federated learning for iot intrusion detec-tion," IEEE Internet of Things Journal, vol. 8, no. 12, pp. 9367–9382, 2021.

[11] G. Sun, Y. Cong, J. Dong, Q. Wang, L. Lyu, and J. Liu, "Privacy-preserving federated learning for 5g-assisted iot intrusion detection," IEEE Network, vol. 35, no. 6, pp. 158–165, 2021.

[12] M. R. Garey and D. S. Johnson, Computers and intractability: a guide to the theory of NP-completeness. W. H. Freeman, 1979.

[13] J. A. Boyan and M. L. Littman, "Packet routing in dynamically changing networks: A reinforcement learning approach," Advances in neural information processing systems, vol. 6, 1994.

[14] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Prox-imal policy optimization algorithms," arXiv preprint arXiv:1707.06347, 2017.

[15] F. Bao, I.-R. Chen, M. Chang, and J.-H. Cho, "Hierarchical trust management for wireless sensor networks and its applications to trust-based routing and intrusion detection," IEEE transactions on network and service management, vol. 9, no. 2, pp. 169–183, 2012.

[16] H. Chen, H. Wu, X. Zhou, and C. Gao, "Trust-based secure routing protocol for wireless sensor networks," China Communications, vol. 13, no. 1, pp. 1–15, 2016.

[17] D. Zhang, F. R. Yu, R. Yang, and H. Tang, "Machine learning enhanced trust evaluation in vehicular networks," Vehicular Communications, vol. 17, pp. 118–128, 2019.

[18] M. Raza, M. Iqbal, M. Sharif, and W. Haider, "Lightweight blockchain-based iot identity management approach," IEEE Internet of Things Journal, vol. 9, no. 17, pp. 16 535–16 546, 2022.

[19] D. Kumar and R. Singh, "A lightweight trust computation mechanism for iot sensor networks," Journal of Ambient Intelligence and Humanized Computing, vol. 11, no. 4, pp. 1515–1530, 2020.

[20] M. L. Littman and J. A. Boyan, "Distributed reinforcement learning for network routing," Technical Report CMU-CS-93-165, 1993.

[21] Z. Tong, Q. Ni, D. Yu, and W. Zhang, "Deep reinforcement learning for routing optimization in ip networks," IEEE Transactions on Network and Service Management, vol. 17, no. 4, pp. 2519–2532, 2020.

[22] X. Wang, J. Li, and H.-H. Chen, "Reinforcement learning-based trust-aware routing in iot networks," IEEE Internet of Things Journal, vol. 8, no. 12, pp. 9692–9703, 2021.

[23] Y. Zhao, M. Li, L. Lai, N. Suda, D. Cidon, and V. Chandra, "Proximal policy optimization for network routing," IEEE Transactions on Network Science and Engineering, vol. 9, no. 4, pp. 2653–2665, 2022.

[24] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," arXiv preprint arXiv:1801.01290, 2018.

[25] H. Jiang, Z. Zhang, G. Wu, and J. Dang, "Multi-agent deep reinforce-ment learning for communications and sensing in uav networks," IEEE Journal on Selected Areas in Communications, vol. 40, no. 4, pp. 1117– 1132, 2022.

10.48047/jocaaa.2024.33.08.254
ISSN: 1521-1398

[26] M. Chen, W. Saad, C. Yin, and M. Debbah, "Transformer-based deep learning for wireless communications: A comprehensive survey," IEEE Communications Surveys & Tutorials, vol. 24, no. 2, pp. 1204–1256, 2022.

[27] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM computing surveys, vol. 41, no. 3, pp. 1–58, 2009.

Y. Meidan, M. Bohadana, Y. Mathov, Y. Mirsky, A. Shabtai, D. Breiten-bacher, and Y. Elovici, "N-baiot—network-based detection of iot botnet attacks using deep autoencoders," vol. 17, no. 3. IEEE, 2018, pp. 12–22.

[29] A. A. Diro and N. Chilamkurti, "Distributed deep learning model for intelligent intrusion detection in iot networks," Computers & Security, vol. 78, pp. 245–261, 2018.

[30] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "Graph neural networks for network anomaly detection," IEEE Transactions on Network Science and Engineering, vol. 9, no. 4, pp. 2523–2537, 2022.

[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in neural information processing systems, vol. 30, 2017.

[32] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "A survey of machine learning for big data processing," EURASIP Journal on Advances in Signal Processing, vol. 2014, no. 1, pp. 1–16, 2021.

[33] R. Li, H. Asaeda, and J. Wu, "A survey on deep learning for named data networking," IEEE Communications Surveys & Tutorials, vol. 23, no. 2, pp. 1598–1627, 2021.

[34] M. H. Alsharif, A. H. Kelechi, M. A. Albreem, S. A. Chaudhry, M. S. Zia, and S. Kim, "A comprehensive survey on 6g networks: Applications, security, and machine learning approaches," IEEE Access, vol. 11, pp. 47 522–47 546, 2023.

[35] Y. Liu, X. Yuan, Z. Xiong, J. Kang, X. Wang, and D. Niyato, "Quantum machine learning for wireless communications: Challenges and oppor-tunities," IEEE Communications Magazine, vol. 61, no. 6, pp. 52–58, 2023.

[36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," vol. 9, no. 8. MIT Press, 1997, pp. 1735–1780.

[37] L. Zhang, M. Wang, and C. Liu, "Deep reinforcement learning for trust-aware routing in wireless networks," IEEE Transactions on Mobile Computing, vol. 22, no. 8, pp. 4756–4768, 2023.

[38] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski et al., "Human-level control through deep reinforcement learning," nature, vol. 518, no. 7540, pp. 529–533, 2015.