Energy-Efficient Multi-Hop LoRa Communication in Forested Environments via Proximal Policy Optimization

Muhd Kahfi bin Jumali[1], Lim Kit Guan[2]*, Ervin Gubin Moung[3],
Lorita Angeline[4], Tianlei Wang[5], Kenneth Teo Tze Kin[6]*
Modelling, Simulation & Computing Laboratory-Faculty of Engineering,
Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia[1, 2, 4, 6]
Data Technology and Applications Research Group-Faculty of Computing and Informatics,
Universiti Malaysia Sabah, Kota Kinabalu, Malaysia[3]
Faculty of Intelligent Manufacturing, Wuyi University, Jiangmen, China[5]

***Abstract*—Multi-hop LoRa networks extend coverage for large-scale Internet of Things deployments but are severely limited by interference-induced collisions, retransmissions, and rapid battery depletion of relay nodes. Conventional routing strategies that minimize hop count or rely on static heuristics fail to account for dynamic medium contention and its impact on energy consumption and reliability. This paper proposes a Proximal Policy Optimization (PPO)–based routing framework for multi-hop LoRa networks that learns interference-aware and energy-efficient routing policies through reinforcement learning. A discrete-event simulation framework is developed to model LoRa physical-layer behaviour, co-spreading-factor interference, adaptive data rate control, and battery-limited relay nodes under multi-source traffic. The routing problem is formulated as a Markov Decision Process (MDP) in which the PPO agent selects next-hop relays based on local topology, relay load, and channel occupancy, while physical-layer parameters are adapted independently using a standards-inspired physical-layer parameters are adapted independently using a standards-inspired ADR mechanism (ADR) mechanism. Simulation results show that the proposed approach achieves a packet delivery ratio of up to 73.7%, reduces collision rates by approximately 46% compared with Random routing, and lowers the average energy consumption per delivered packet to about 206 mJ, outperforming Shortest Path and Ad hoc On-Demand Distance Vector (AODV)-like routing. These gains are achieved by learning spatially diverse routing paths that mitigate relay congestion and reduce collision-induced retransmissions.**

***Keywords*—LoRa; Multi-Hop; PPO; ADR; MDP-Markov Decision Process; Reinforcement-Learning; PDR**

I. Introduction

The rapid growth of Internet of Things (IoT) [1] applications in smart agriculture, environmental monitoring, and industrial automation has accelerated the adoption of Low Power Wide Area Networks (LPWANs) as a cost-effective solution for long-range, energy-constrained wireless connectivity [2]. Among LPWAN technologies, Long Range (LoRa) has emerged as a particularly attractive option due to its ability to provide kilometer-scale coverage while enabling multi-year battery operation [3]. However, the star topology employed in Long Range Wide Area Network (LoRaWAN), where end devices communicate directly with a gateway, becomes ineffective in scenarios where distance, terrain, or urban obstructions prevent reliable direct links, as empirically demonstrated in forested and vegetated deployments [4], [5], [6]. To overcome this limitation, multi-hop LoRa networks introduce relay-assisted communication, extending coverage while preserving LoRa's low-power characteristics [7].

Despite their potential, multi-hop LoRa networks face significant routing challenges under multi-source traffic. Unlike traditional wireless sensor networks with centralized scheduling or orthogonal channel access, LoRa operates under pure Additive Links On-line Hawaii Area (ALOHA) and strict duty-cycle constraints [8], limiting coordination. When multiple sources independently select relays based on local metrics such as distance or link quality, traffic converges on a small subset of relays, creating persistent collision hotspots. These collisions increase retransmission, waste energy through long transmissions, and accelerate relay battery depletion. Although Adaptive Data Rate (ADR) improves link robustness, it does not mitigate contention among competing sources, leaving collision-aware routing an open challenge.

Existing routing approaches for multi-hop LoRa [9] networks fall into three broad categories. Distance-based heuristics, such as shortest path routing, offer simplicity but suffer from severe load imbalance and collision concentration. Adaptive protocols inspired by Ad hoc On-Demand Distance Vector (AODV) incorporate energy or load-aware metrics to improve fairness yet still rely on static decision rules that are insufficient under dynamic multi-source contention. More recent studies have applied reinforcement learning to LoRa networks, primarily focusing on physical-layer parameter optimization such as spreading factor selection [10]. While effective at the link level, these approaches do not address network-layer routing coordination across competing sources, leading to suboptimal performance in multi-hop scenarios.

Energy efficiency is a central concern in this context, as relay nodes are typically battery-powered and must forward traffic from multiple sources [11]. In interference-limited LoRa networks, energy consumption is dominated not by hop count

*Corresponding author.

alone but by collision-induced retransmissions and the use of high spreading factors on congested links. Consequently, routing strategies that ignore multi-source interaction often incur excessive energy expenditure and uneven battery depletion, directly reducing network lifetime [12]. The fundamental challenge is therefore to design routing mechanisms that jointly improve packet delivery reliability, minimize energy consumption, and balance relay utilization under dynamic traffic and interference conditions.

Recent advances in deep reinforcement learning [13] have shown strong potential for addressing such multi-objective optimization problems in complex networked systems. Proximal Policy Optimization (PPO) offers stable and sample-efficient policy learning in non-stationary environments, making it well suited for multi-source routing scenarios where traffic patterns and interference evolve over time [14]. When provided with appropriate observability into relay utilization and channel conditions, PPO can learn adaptive routing behaviors without requiring explicit coordination or message exchange between sources.

Motivated by this observation, this paper proposes a PPO-based routing framework for multi-hop LoRa networks that explicitly accounts for multi-source interaction. By augmenting the routing state representation with features capturing relay usage frequency and temporal recency, the proposed approach enables implicit coordination among sources, allowing routing decisions to proactively avoid congested relays. Physical-layer parameters are adapted independently using a standards-inspired ADR mechanism [15], ensuring a clear separation between routing optimization and link-level adaptation.

The main contributions of this work can be summarized as follows:

- Multi-source–aware state features are introduced to enable routing agents to detect and avoid relays that are currently utilized by competing sources, thereby reducing collision probability without incurring explicit coordination overhead.
- An energy and reliability-aware reward function is designed to balance packet delivery, energy consumption, and collision avoidance in battery-limited relay networks.
- A curriculum-based PPO training strategy is developed to accelerate convergence toward collision-aware routing policies.
- Finally, through comprehensive simulation-based evaluation, it is demonstrated that the proposed framework achieves improved packet delivery reliability, reduced collision rates, and enhanced energy efficiency compared with conventional routing approaches, while exhibiting emergent spatial path separation that mitigates relay energy hotspots.

The remainder of this paper is organized as follows. Section II reviews related work on multi-hop LoRa routing and reinforcement learning for wireless networks. Section III presents the system model, including network architecture, collision dynamics, and energy consumption. Section IV details the proposed PPO-based routing framework. Section V discusses performance evaluation and key insights. Finally, conclusions are drawn in Section VI.

II. Related Work

LPWAN [16] technologies were developed to provide long-range connectivity under stringent energy constraints, making them suitable for large-scale IoT deployments [17] where frequent battery replacement is infeasible. Comparative surveys position LoRa/LoRaWAN among the most widely adopted LPWAN options because it provides kilometers-scale coverage with low device complexity and low power operation, albeit with trade-offs in capacity and interference robustness under ALOHA-like access. In practical deployments, coverage and reliability are highly scenario-dependent (terrain, foliage, obstacles, antenna height), motivating designs that go beyond conventional single-hop LoRaWAN star topologies when direct gateway connectivity is unreliable.

A major limitation of LoRa networks especially under multi-node contention is that performance can degrade rapidly due to collisions, imperfect orthogonality among spreading factors, and capture effects [18]. Early experimental and analytical studies show that "scaling" LoRa is non-trivial because concurrent transmissions can interfere even when nominally using different Spreading Factor (SF), while co-SF collisions depend on relative received powers and overlap in time [19]. Complementary simulation-based investigations further quantify how packet delivery probability is impacted by offered load, SF distribution, and channel contention, reinforcing that interference modelling choices (co-SF and inter-SF rejection assumptions, capture threshold, traffic model) materially affect energy and reliability conclusions.

In multi-hop settings, these issues can be amplified: relays introduce additional transmissions (hence additional channel occupancy), and traffic tends to concentrate around a subset of "good" relays (those that make strong progress toward the gateway), increasing the probability of repeated collisions and retransmission-driven energy waste [20]. This motivates routing policies that are explicitly collision-aware and load-aware rather than purely distance-optimal. LoRaWAN's ADR [21] is a standardized mechanism designed primarily for star topologies: the network server adjusts SF and transmits power using link-quality statistics (commonly using margin-based logic over recent Signal-to-Noise Ratio (SNR) / Received Signal Strength Indicator (RSSI) history) [22]. ADR is widely used because reducing SF and/or transmit power can substantially reduce time-on-air and energy per message when link margin is high, while increasing SF/power improves robustness when the link is marginal.

However, ADR alone does not resolve contention externalities: even when each node's physical settings are individually reasonable, collisions can still dominate under shared-medium access particularly when multiple nodes select similar SFs and transmit times. Thus, while ADR addresses the "link adaptation" dimension, it typically does not optimize network-layer forwarding decisions or coordinate multi-source relay usage in multi-hop mesh behaviours [23].

Multi-hop and mesh extensions [24] for LoRa have been explored to address coverage holes and obstacles, proposing lightweight routing protocols (often distance-vector-like, flooding-assisted discovery, or heuristic metrics such as hop count, link quality, or airtime cost). A recent comprehensive survey consolidates these directions, highlighting that many LoRa mesh proposals prioritize simplicity and feasibility over strong optimality guarantees, and often assume limited coordination under duty-cycle constraints. Empirical protocol studies also demonstrate that practical mesh routing must manage not only reachability and hop efficiency but also traffic concentration and fairness, because a small number of relays can become persistent bottlenecks [25].

Importantly, classical ad hoc routing protocols such as AODV were developed for higher-rate radios and different interference regimes, relying on route discovery and maintenance logic that may not directly translate to LoRa's long airtime, low duty cycle, and collision-dominated shared channel. Nevertheless, AODV remains a common baseline because it captures the value of adaptive path selection versus static shortest-path heuristics. Learning-based methods have been applied primarily to LoRaWAN parameter adaptation (e.g., selecting SF / Transmission Power (TP) / channel assignments to improve energy efficiency and reliability). Recent Deep Reinforcement Learning (DRL) work shows that intelligent control can outperform fixed heuristics in dynamic environments, especially when objective functions include energy and performance constraints [26]. In parallel, PPO is widely recognized as a stable policy-gradient algorithm for complex decision-making problems, due to its clipped surrogate objective that reduce destructive policy updates [27].

Despite this progress, most learning-driven LoRa papers focus on star topologies and Physical Layer (PHY) parameter selection rather than joint routing/forwarding under multi-hop mesh contention [28]. In a multi-source, battery-limited relay mesh, the key difficulty is not only choosing energy-efficient PHY settings but also preventing multiple sources from repeatedly converging on the same high-quality relay-creating collision hotspots and unfair battery depletion. Across the above literature, a consistent gap is the lack of explicit multi-source awareness in routing decisions for multi-hop LoRa meshes [29]. Existing mesh routing protocols and classical baselines typically optimize progress metrics (distance/hops/link quality) but do not directly encode "competition" from other sources (who is using which relays and how recently), even though that competition is the driver of collision synchronization and relay hotspot depletion in shared-medium LoRa. At the same time, ADR-based PHY adaptation optimizes link margin but does not solve contention-driven losses or relay fairness.

From the above, a clear gap emerges where LoRa scalability or interference studies characterize collision behaviour and airtime-driven limitations but do not provide a network-layer mechanism for multi-source relay competition in battery-limited multi-hop settings. Multi-hop LoRa routing protocols propose metrics such as Time on Air (ToA)-aware distance-vector routing and cross-layer scheduling but typically lack explicit multi-source awareness signals that help avoid relay hotspots formed by concurrent sources selecting the same "best" relays.

III. Methodology

A. Network Architecture

The considered multi-hop LoRa network is represented as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the node set $\mathcal{V}$ comprises source nodes, battery-powered relay nodes, and a gateway acting as the final data sink. A directed link $(i, j) \in \mathcal{E}$ exists if node j lies within the communication range of node i. Each source periodically generates uplink packets that must be forwarded through one or more relay hops to reach the gateway. Relay nodes operate in a store-and-forward manner and are constrained by finite energy budgets, while the gateway is assumed to have unlimited energy resources. The network topology remains static during each simulation run, whereas wireless channel conditions, interference patterns, and medium access timing evolve dynamically. LoRa Mesh Network Topology is shown in Fig. 1.

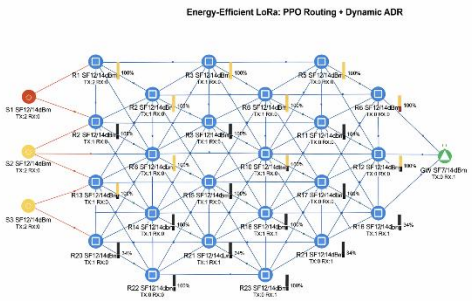

Fig. 1. LoRa Mesh Network Topology

B. Energy Consumption Model

The energy consumption of each relay node is determined by the power used for transmission and the corresponding LoRa time-on-air. Specifically, the energy consumed for transmission from node i to node j is expressed as

$$E_{i,j}^{\text{tx}} = I_{\text{tx}}(P_i^{\text{tx}})\, V\, T_{\text{air}} \tag{1}$$

where $P_{i,j}^{\text{tx}}$ denotes the transmit power selected for the link and $T_{\text{air}}(SF_{i,j})$ represents the airtime associated with the selected spreading factor. The airtime consists of preamble and payload components and increases exponentially with the spreading factor due to the reduced bit rate. Consequently, higher spreading factors result in significantly longer transmission durations and increased energy expenditure. Each relay node maintains a residual energy state that is updated after every transmission according to:

$$E_i^{\text{res}}(t+1) = E_i^{\text{res}}(t) - \sum E_{i,j}^{\text{tx}} \tag{2}$$

Equation (2) updates the residual energy of relay i after time step t by subtracting the total transmission energy spent on all forwarding events executed by that relay during the step. Here, $E_i^{\text{res}}(t)$ denotes the remaining battery energy at time t, and $E_{i,j}^{\text{tx}}$ represents the energy consumed when node i transmits a packet to next hop j. In our simulator, $E_{i,j}^{\text{tx}}$ is computed from the radio's current draw at the selected transmit power and the packet time-on-air, i.e., $E_{i,j}^{\text{tx}} = I_{\text{tx}}(P_{\text{tx}})\, V\, T_{\text{OA}}$, ensuring that higher spreading factors or power levels incur higher energy.

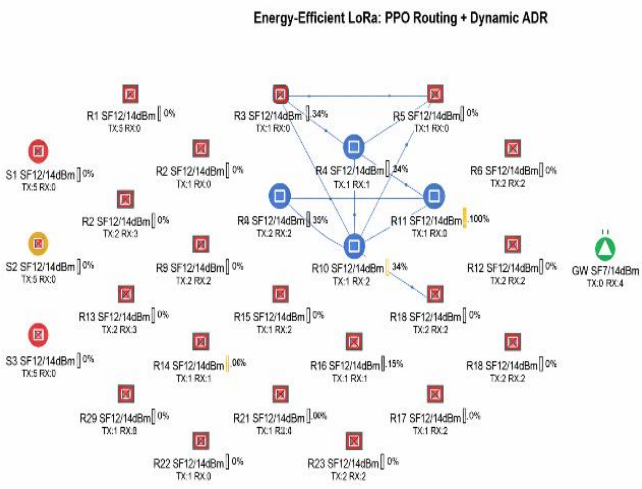

Fig. 2. Battery depletion.

Costs through longer airtime and larger current consumption. Once $E_i^{\text{res}}(t) \le 0$, relay i is treated as depleted (dead) and is removed from candidate next-hop sets, which can force route reconfiguration and directly impacts network lifetime. Fig.2 shows the Battery depletion.

C. Signal Propagation Model

Wireless signal propagation is modelled using a forest path-loss [30] formulation augmented with log-normal shadow fading to capture the effects of vegetation and terrain irregularities. To capture realistic rural/forested deployments, the wireless channel is modelled using a log-distance path-loss formulation with an additional vegetation attenuation term and log-normal shadow fading, consistent with empirical LoRaWAN path-loss characterizations in forested and vegetated environments reported in [5], [6], [31], [32]. The vegetation attenuation term κd represents the specific attenuation through foliage as standardized in ITU-R Recommendation P.833-10 [33]. For a transmitter–receiver separation distance $d_{i,j}$ (in meters), the path loss is expressed as:

$$PL(d) = PL(1\text{ m}) + 10n\log_{10}(d) + \kappa d + X_\sigma \quad (3)$$

where $PL(1\text{ m})$ denotes the free-space path loss at 1 meter, n is the path-loss exponent, κ is a vegetation attenuation coefficient (in dB/m), and $X_\sigma \sim \mathcal{N}(0,\sigma^2)$models log-normal shadow fading. In our implementation, $PL(1\text{ m})$ is obtained from the standard free-space expression at carrier frequency f,

$$RSSI_{i,j} = P_i^{\text{tx}} + G_t + G_r - PL(d_{i,j}) \quad (4)$$

where P_i^{tx} is the transmit power (dBm) and G_t and G_r are antenna gains (dBi). This formulation directly matches the simulator's computation $RSSI = P^{tx} + 2G - PL$.

$$SNR_{i,j} = RSSI_{i,j} - N_0 \quad (5)$$

$$N_0 = -174 + 10\log_{10}(BW) + NF \quad (6)$$

with N_0 denoting the noise floor. Successful packet reception requires both the RSSI and SNR to exceed the minimum demodulation thresholds determined by the selected spreading factor [34]. If either condition is violated, the packet is considered lost due to weak signal conditions [35].Demodulation Criteria in Table I

TABLE I. DEMODULATION CRITERIA

SF	Receiver Sensitivity Threshold (dBm)	Min SNR (dB)
7	-123	-7.5
8	-126	-10
9	-129	-12.5
10	-132	-15
11	-134.5	-17.5
12	-137	-20

Transmission duration is modeled using the LoRa time-on-air expression, which directly influences both channel occupancy (collision probability) and energy consumption. The symbol duration is

$$T_{\text{sym}} = \frac{2^{SF}}{BW} \quad (7)$$

$$T_{\text{preamble}} = (N_{\text{preamble}} + 4.25)T_{\text{sym}} \quad (8)$$

$$Tpayload = Npayload_symb \cdot Tsym \quad (9)$$

$$N_{\text{payload}} = 8 + \max\left(\left\lceil \frac{8PL - 4SF + 28 + 16CRC - 20IH}{4(SF - 2DE)} \right\rceil (CR + 4), 0\right) \quad (10)$$

where N_{preamble} is the programmed preamble length. The payload duration depends on the number of payload symbols N_{payload} and is payload length (bytes), $CRC \in 0,1$ indicates CRC enablement, $IH \in 0,1$ indicates implicit header mode, and $DE \in 0,1$ is the low-data-rate optimization flag (typically enabled for $SF \ge 11$). This expression is consistent with the simulator's airtime computation used for every transmission event.

D. Co-SF interference & Collision Decision

Collisions [36] are modeled using co-spreading-factor overlaps, consistent with the simulator's event-based tracking of concurrent transmissions at each SF. During the reception of packet p transmitted from node i to node j, the set of overlapping interferers using the same SF is denoted $\mathcal{J}_j^{(SF)}$. Interferer received powers are accumulated in the linear domain:

$$P_{k,j}^{(\text{lin})} = 10^{\frac{RSSI_{k,j}}{10}} \quad (11)$$

where $RSSI_{k,j}$ is expressed in dBm. The aggregate interference power at node j is then obtained by summing the linear powers of all interfering transmitters,

$$I_j = \sum_{k \in \mathcal{J}_j^{(SF)} \setminus i} P_{k,j}^{(\text{lin})} \quad (12)$$

The signal-to-interference ratio at the receiver is computed as:

$$SIR_{i,j} = 10\log_{10}\left(\frac{10^{RSSI_{i,j}/10}}{I_j + \epsilon}\right) \quad (13)$$

where ϵ is a small constant introduced to avoid numerical instability when no concurrent interferers are present. A packet

collision is declared when the computed Signal-to-Interference Ratio (SIR) falls below a predefined capture threshold γ_{th}

$$\text{Collision occurs if } SIR_{i,j} < \gamma_{\text{th}} \quad (14)$$

In this study, a conservative capture threshold is employed, reflecting the imperfect orthogonality of LoRa spreading factors under real-world synchronization and fading conditions. When a collision occurs, the packet is considered lost and removed from the reception buffer, triggering retransmission attempts and additional energy consumption. This collision model directly links medium access contention, routing decisions, and physical-layer parameter selection, thereby allowing the reinforcement learning agent to implicitly learn collision-aware routing behavior.

E. Dynamic Adaptive Data Rate (ADR) Mechanism

To adapt physical-layer parameters in response to time-varying channel conditions, a Dynamic ADR mechanism is employed for all routing schemes [37]. ADR operates independently of the routing algorithm and is applied uniformly to ensure a fair comparison between PPO-based routing and baseline methods. For each transmitter–receiver link (i, j), the ADR controller maintains a sliding window of recent received signal strength and signal-to-noise ratio measurements. Let $\mathcal{H}_{i,j} = \{SNR_k\}_{k=1}^{W}$ denote the most recent W SNR observations collected at the receiver. The average link margin is computed as:

$$M_{i,j} = \text{SNR}_{i,j} - \text{SNR}_{min}(\text{SF}) \quad (16)$$

where $\bar{SNR}_{i,j}$ is the mean SNR over the observation window and $SNR_{\min}(SF)$ is the minimum demodulation threshold for the currently selected spreading factor. When the link margin exceeds a predefined upper threshold, the ADR controller attempts to reduce energy consumption by decreasing the spreading factor or transmit power, prioritizing spreading factor reduction due to its exponential impact on time-on-air. Conversely, when the margin falls below a lower threshold, the controller increases the spreading factor or transmits power to restore link reliability. If neither adjustment is feasible due to parameter bounds, the current configuration is retained. This adaptive process balances robustness and energy efficiency while avoiding excessive packet loss. ADR updates are performed periodically and only after sufficient measurement samples are available, preventing unstable oscillations in parameter selection.

Algorithm 1 Dynamic Adaptive Data Rate (ADR) Control

```
INPUT: Node ID, Current SF, Current TX Power, Link Quality History
OUTPUT: Optimized SF and TX Power

FUNCTION UpdateLinkQuality (nodeID, RSSI, SNR):
   history_RSSI [nodeID]. append (RSSI)
   history_SNR [nodeID]. append (SNR)
   Keep only the most recent N measurements
END FUNCTION

CalculateADR (nodeID, SF_current, P_current):

   IF length (history_RSSI [nodeID]) < 10 THEN
      RETURN (SF_current, P_current)
   END IF

   RSSI_avg ← (1 / N) × Σ (i = 1 to N) RSSI_i
   SNR_avg  ← (1 / N) × Σ (i = 1 to N) SNR_i

   Margin_RSSI ← RSSI_avg – Sensitivity [SF_current]
   Margin_SNR  ← SNR_avg  − MinSNR [SF_current]

   SF_new ← SF_current
   P_new  ← P_current

   IF (Margin_RSSI > 15 dB) AND (Margin_SNR > 10 dB)
THEN
      IF SF_current > SF_min THEN
         SF_new ← SF_current − 1
      END IF
   ELSE IF (Margin_RSSI < 5 dB) OR (Margin_SNR < 2 dB)
THEN
      IF SF_current < SF_max THEN
         SF_new ← SF_current + 1
      END IF
   END IF

   Margin_RSSI′  ← RSSI_avg − Sensitivity [SF_new]

   IF (Margin_RSSI′  > 20 dB) AND (P_current > P_min)
THEN
      P_new ← max(P_min, P_current − 3 dB)
   ELSE IF (Margin_RSSI′  < 8 dB) AND (P_current <
P_max) THEN
      P_new ← min(P_max, P_current + 3 dB)
   END IF

   RETURN (SF_new, P_new)
```

IV. Proposed Learning Model

A. Problem Formulation

Routing in a multi-hop LoRa mesh network is inherently sequential and stochastic due to time-varying interference, shared medium access, and battery depletion of relay nodes. At each forwarding step, a relay must select a next-hop neighbor without full knowledge of future channel conditions or competing transmissions from other sources. This makes classical shortest-path or greedy routing suboptimal in the presence of contention and energy imbalance. To address this challenge, the routing problem is formulated as a Markov Decision Process (MDP) [38], enabling the use of reinforcement learning to derive an adaptive routing policy. The MDP is defined by the tuple (S, A, P, R, γ) where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P}$ denotes the state transition dynamics, $\mathcal{R}$ is the reward function, and $\gamma \in (0,1]$ is the discount factor. The objective is to learn a routing policy $\pi_\theta(a \mid s)$, parameterized by θ, that maximizes the expected cumulative discounted reward over an episode, corresponding to energy-efficient and reliable end-to-end packet delivery.

1) State space design. The state observed by the PPO agent encodes both topological progress and medium access conditions relevant to routing decisions [39]. For each forwarding node, a fixed-length state vector is constructed by

aggregating features associated with up to K candidate next-hop relays. In this study, K = 10 is used to bound the action space. Let $\mathbf{s}_t \in \mathbb{R}^{16K}$ denote the state at decision step t. For each candidate relay k, the following normalized features are included: relative distance to the gateway, hop distance from the current node, normalized progress toward the gateway, residual energy ratio, historical load ratio, averaged RSSI and SNR statistics, channel occupancy at the candidate's spreading factor, node-level and link-level collision rates, local traffic density, spreading-factor congestion, transmission recency, and relay load indicators. To explicitly address multi-source contention, two additional features are incorporated: the number of other sources recently using the same relay and the recency of such usage. The state vector is expressed as

$$st = [f1, t, f2, t, \dots, fK, t], fk, t \in R16 \quad (17)$$

where zero-padding is applied when fewer than K candidates are available. This formulation ensures a fixed-dimensional input suitable for neural network training while preserving local routing context and contention awareness.

2) Action space & feasibility Masking. The action space corresponds to the selection of the next-hop relay among the candidate neighbors [40]. At each decision step t, the agent selects an action

$$at \in 1, 2, \dots, K, \quad (18)$$

where each index corresponds to a specific candidate relay in the ordered neighbor list. Actions associated with invalid or non-existent candidates are masked during policy evaluation to prevent infeasible selections. Importantly, the PPO agent optimizes routing decisions only; physical-layer parameters such as spreading factors and transmit power are adapted independently by the ADR controller.

The reward function is designed to jointly incentivize reliable packet delivery, energy efficiency, collision avoidance, and load balancing across relay nodes [41]. Let r_t denote the reward assigned after an episode completes. A large positive reward is granted when a packet is successfully delivered to the gateway, scaled by the current packet delivery ratio to encourage global reliability. Conversely, strong penalties are imposed for packet loss due to collisions or weak signal conditions, reflecting the high energy cost of retransmissions. Energy efficiency is explicitly encouraged by penalizing excessive transmission energy, while hop-count penalties discourage unnecessarily long routes. To mitigate relay hotspots and promote fairness, additional penalties are applied when routing paths overlap significantly with those of other concurrent sources or when relay load imbalance is detected. Formally, the reward can be expressed as:

$$r = R_s ucc - \alpha E - \beta H - \delta O, if delivered$$
$$r = -R_f ail - \alpha E - \beta H, otherwise \quad (19)$$

Here, $R_s ucc$, and $R_f ail$ are positive constants rewarding successful delivery and penalizing packet failure, respectively. The coefficients α, β, and δ control the relative importance of energy consumption, hop count, and path overlap. By penalizing overlapping relay usage, the reward function discourages routing decisions that create relay hotspots and persistent collisions. This formulation enables the agent to learn routing strategies that trade minimal hop count for reduced contention and energy waste when necessary.

B. Proximal Policy Optimization (PPO) Architecture

The routing policy is optimized using PPO [42], which improves training stability by constraining policy updates. PPO maximizes a clipped surrogate objective defined as

$$L^{PPO}(\theta) = E_t[\min(r_t(\theta) A_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon) A_t)] \quad (20)$$

where,

$$r_t(\theta) = \frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{old}}(a_t \mid s_t)} \quad (21)$$

is the probability ratio between the current and previous policies, $\hat{A}_t$ is the estimated advantage function, and ϵ is a clipping parameter that limits the magnitude of policy updates.

This clipped formulation prevents excessively large policy updates that could otherwise destabilize the learning process, particularly in stochastic and highly constrained environments such as multi-hop LoRa networks. By bounding the probability ratio $r_t(\theta)$ within the interval $[1-\epsilon, 1+\epsilon]$, PPO ensures that newly updated policies do not deviate significantly from the previously deployed policy, thereby preserving stable and incremental policy improvement. This property is especially important in routing optimization, where abrupt changes in forwarding decisions may lead to sudden congestion, increased collision rates, or uneven energy depletion among relay nodes. Fig 3 PPO architecture.

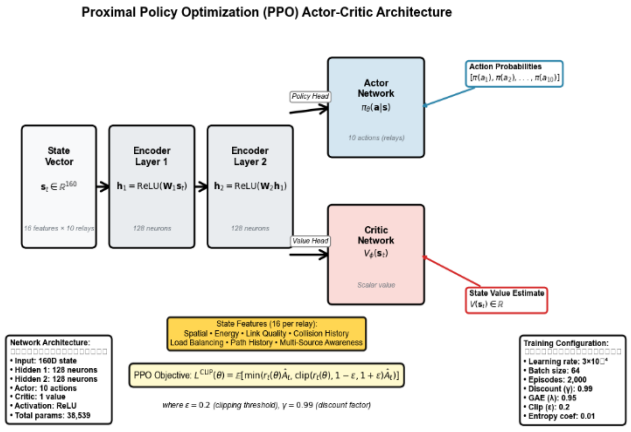

Fig. 3. Proximal Policy Optimization (PPO) architecture.

Algorithm 2 Proximal Policy Optimization (PPO)

INPUT: State s ∈ R^160, Actions A = {a1, ..., a10}, Learning rate α = 3×10^-4
OUTPUT: Optimized routing policy π_θ

NETWORKS:
Actor: π_θ(a|s) = Softmax(f_θ(s))
Critic: V_φ(s) = g_φ(s)

TRAINING (Per Episode):

Action Selection with Temperature Scaling

p(a|s) = (π _ θ (a|s) ⊙ M) / Σ(π _ θ (a|s) ⊙ M)
a_t ~ Categorical(p(a|s)^(1/T))
where M is an action mask, T is the temperature

Discounted Returns and Normalization
G_t = Σ(τ = t to T_end) γ^(τ−t) r_τ
Ĝ_t = (G_t − μ(G)) / (σ(G) + ε_norm)

PPO Update (K = 4 epochs, using collected trajectories)
ratio_t = π_θ(a_t|s_t) / π_θ_old(a_t|s_t)
A_t = Ĝ_t − V_φ(s_t)

L^CLIP = −E[min(ratio_t·A_t , clip(ratio_t, 1−ε, 1+ε)·A_t)]
L^VF = E[(V_φ(s_t) − Ĝ_t)^2]
L^ENT = −E[Σ_a π_θ(a|s_t) log(π_θ(a|s_t))]

L_total = L^CLIP + 0.5·L^VF − 0.01·L^ENT

Update parameters:
θ, φ ← θ, φ − α ∇_{θ,φ} L_total

Gradient Clipping
||∇_{θ,φ}|| ⩽ 0.3

PARAMETERS:
γ = 0.99, ε = 0.2, K = 4,
T = max(0.01, 0.6 − (episode / N_max)·0.59)

C. Training Procedure

Training is conducted over multiple episodes, each consisting of concurrent packet transmissions from multiple source nodes. During training, stochastic action sampling is employed to promote exploration, while temperature scaling is used to avoid premature convergence. Experience tuples (s_t, a_t, r_t, s_{t+1}) are collected across episodes and used to update the policy and value networks in mini batches. To reduce synchronization-induced collisions and encourage path diversity during early learning, guided exploration is applied in initial training phases. As training progresses, the agent gradually transitions to fully autonomous decision-making based on learned policies.

By formulating multi-hop LoRa routing as a reinforcement learning problem and incorporating collision-aware state features and energy-sensitive rewards, the proposed PPO-based framework learns routing policies that mitigate relay hotspots, reduce collision-induced retransmissions, and improve the overall energy–reliability trade-off.

D. Simulation Setup & Configuration

This work is simulation-based rather than hardware-deployed a deliberate methodological choice given that deploying a 27-node multi-hop LoRa mesh (3 sources, 23 relays with 90000mJ battery in a 100 m hexagonal grid, 1 gateway) across a 1.9 km × 1.1 km forested area with a maximum link range of 300 m and a source-to-gateway distance of 1,100 m for controlled, repeatable experiments is logistically and financially impractical within a single study. Simulation allows systematic isolation of individual variables (SF assignment, relay energy, channel occupancy, multi-source contention) under conditions grounded in empirically validated forest propagation parameters. All experiments are implemented in Python 3.11 using PyTorch for the PPO agent, LoRaSim for discrete-event network dynamics, SimPy for real-time scheduling, and NumPy/Pandas for data processing (Table II).

TABLE II. SIMULATION PARAMETERS

LoRa Parameters	**Values**
Gain	2 dBi
CR, c	1
BW	125kHz
Duty Cycle, n	10%
Length of payload, npl	100 bytes
References distance, d0	1m
Path loss index, y	3.8
Shadow fading, x0 `~N	N(0, 8^2) dB
SF	{7, 8, 9, 10, 11, 12}
TP	{8, 11, 14, 17, 20} dBm
Training Parameters	**Values**
State Dimension	160
Action Dimension	10
Learning Rate, α	3×10^{-4}
Total Training Episodes	2000
Discount Factor, γ	0.99
Clip Epsilon, ε	0.2

E. Baseline Routing Methods

To objectively evaluate the effectiveness of the proposed PPO-based routing framework, its performance is compared against three conventional routing strategies: Shortest Path routing, Random routing, and an AODV-like routing method. These baselines are selected to represent commonly used heuristic, stochastic, and protocol-inspired approaches in multi-hop wireless networks.

1) Shortest Path Routing selects the next-hop relay that minimizes the Euclidean distance to the gateway [43]. At each forwarding decision, the relay j^* is chosen such that

$$j^* = \arg\min_{j \in \mathcal{N}_i} d(j, \text{GW}) \quad (22)$$

where $\mathcal{N}_i$ denotes the set of candidates neighboring relays of node i, and $d(j, \text{GW})$ is the Euclidean distance between relay j and the gateway. This strategy minimizes hop distance toward the destination but does not consider relay energy, traffic load, or channel interference. As a result, it often concentrates traffic on a small subset of relays, leading to energy hotspots and increased collision probability.

2) Random Routing selects the next-hop relay uniformly at random from the set of reachable neighboring relays [44]. If N candidate relays are available, the selection probability for each relay j is given by

$$P(j) = \frac{1}{|\mathcal{N}_i|} \quad (23)$$

where $|\mathcal{N}_i|$ is the number of available candidate relays. This baseline does not exploit any network state information and serves as a lower-bound reference for routing performance. While Random routing can occasionally provide path diversity, it generally leads to inefficient routing paths, excessive hop counts, and elevated energy consumption.

3) AODV-like Routing is inspired by the AODV protocol [45] and selects the next-hop relay based on a composite routing metric that accounts for distance to the gateway, residual energy, and relay load. For each candidate relay j, a routing cost function is defined as

$$C(j) = \alpha\ d(j, \mathrm{GW}) + \beta\left(1 - \frac{E_j}{E_{max}}\right) + \gamma\, L_j \qquad (24)$$

where E_j denotes the residual energy of relay j, E_{max} is the initial battery capacity, and L_j represents the relay load, modeled as the ratio of forwarded packets to received packets. The weighting coefficients $\alpha, \beta,$ and γ balance progress toward the gateway, energy awareness, and load distribution, respectively. The relay with the minimum cost $C(j)$ is selected as the next hop. Although the AODV-like method introduces energy and load awareness, it relies on static weighting factors and lacks adaptability to dynamic interference and collision conditions. Consequently, it remains susceptible to suboptimal routing decisions under time-varying traffic and channel states.

V. Simulation Results

This section presents a comprehensive evaluation of the proposed PPO-based multi-hop LoRa routing framework with dynamic ADR and compares its performance against conventional baseline routing strategies, namely Shortest Path, Random Routing, and AODV-like routing. The evaluation focuses on energy efficiency, packet delivery reliability, collision behaviour, and path diversity under realistic propagation and interference conditions.

A. Training Convergences and evaluation

All the figures below illustrate the learning dynamics and performance evolution of the proposed PPO-based routing framework in a multi-hop LoRa network. The results demonstrate how reinforcement learning progressively improves routing decisions by reducing collisions, increasing packet delivery reliability, and reducing energy consumption.

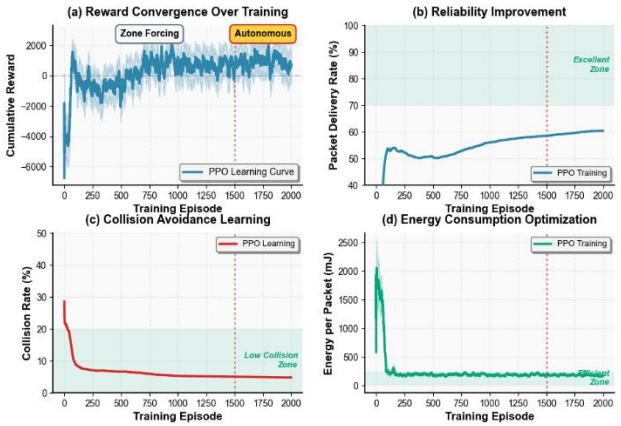

Fig. 4. PPO Learning curve & training optimization.

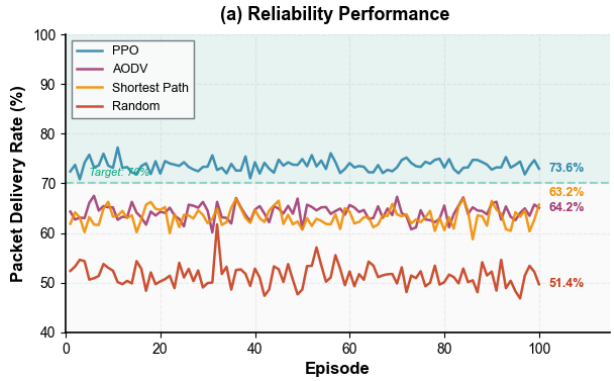

Fig. 5. Packet delivery reliability.

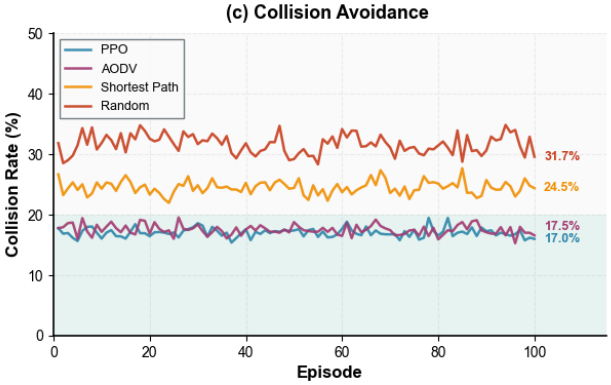

Fig. 6. Collision avoidance.

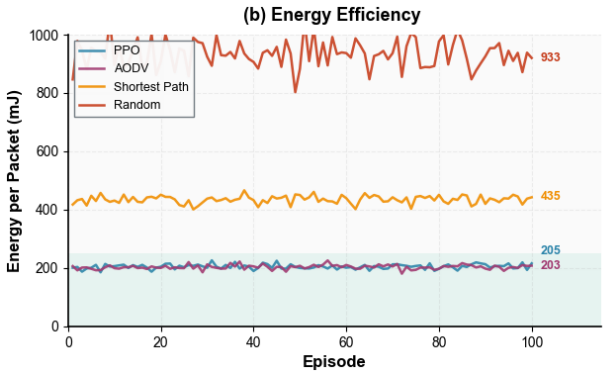

Fig. 7. Energy consumption optimization.

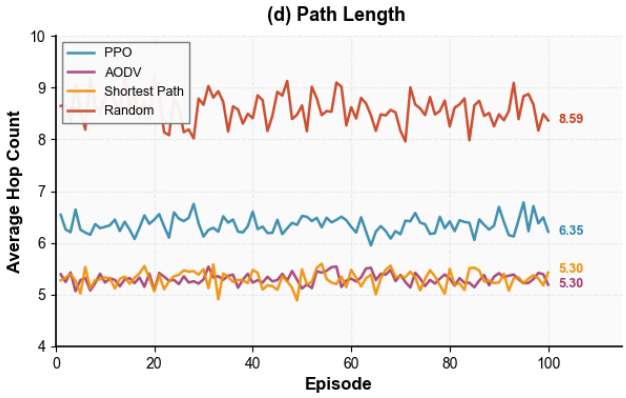

Fig. 8. Average hop count.

The cumulative reward in Fig. 4. represents the expected discounted return $E[\sum_t \gamma^t r_t]$, where r_t penalizes transmission energy E, hop count H, and path overlap O while rewarding successful packet delivery. During early training, the reward is highly negative and exhibits large variance due to exploratory routing decisions that frequently select congested relays, leading to collisions and retransmissions. As training progresses, the PPO policy updates stabilize through clipped probability ratio optimization, resulting in steadily increasing and eventually convergent reward values. The reduction in reward variance indicates policy convergence and reduced exploration.

Fig. 5. shows the evaluation of packet delivery ratio. In LoRa networks, packet decoding succeeds only when the signal-to-interference-plus-noise ratio satisfies $\text{SINR} = \frac{P_r}{\sum I+N} \geq \text{SINR}_{\text{min}}(\text{SF})$ here $\sum I$ represents aggregated co-spreading-factor interference. At early training stages, frequent overlapping transmissions violate this condition, causing packet loss. As the PPO agent learns to avoid relays with high channel occupancy and recent usage, $\sum I$ decrease, allowing more packets to satisfy the SINR constraint. Consequently, the PPO-based approach achieves a higher delivery ratio than Random, AODV-like, and Shortest Path routing.

Average Collision rate is illustrated in Fig. 6. Baseline routing methods, particularly Shortest Path routing, repeatedly select the same relays, increasing the probability of overlapping transmissions and persistent collision hotspots. In contrast, the PPO agent minimizes path overlap O by incorporating relay usage history and collision statistics into the state representation. This leads to a rapid reduction in collision rate during training and a significantly lower steady-state collision level compared with all baseline methods.

Fig. 7. presents the average energy consumption per delivered packet. The transmission energy is modeled as $E_{tx} = I_{tx} \times V \times T_{air}$, where T_{air} increases exponentially with the spreading factor. During early training, frequent retransmissions and ADR-triggered increases in spreading factors result in excessive energy consumption. As collision probability decreases under PPO routing, retransmissions are reduced and ADR operates at lower spreading factors and transmit power levels, thereby reducing both T_{air} and transmission current. The combined effect yields the lowest energy consumption per delivered packet among all evaluated methods.

Average Hop count in Fig. 8 shows PPO routes are longer than Shortest Path and AODV, but far shorter than Random routing. PPO learns a controlled trade-off: it accepts a modest hop-count increase (about 6.35 hops on average) to bypass relays that are frequently selected by other sources, which are more likely to be congested. In a contention-limited LoRa mesh, avoiding these hotspot relays reduces co-SF airtime overlap, lowering collision probability and the need for energy-costly retransmissions. Illustrate in Fig. 10, PPO still preserves forward progress toward the gateway, so the additional hops reflect purposeful path diversification rather than inefficient wandering.

Fig. 4 also shows the PPO training convergence across 2,000 episodes in two phases. The agent initially produces highly negative cumulative rewards (−6,000) due to random exploration, then improves rapidly between episodes 50–500 as the policy learns interference avoidance and load balancing. Stable performance is reached at approximately episode 500 and sustained through the Autonomous phase (episodes 1,500–2,000) without performance collapse, confirming the policy generalizes beyond the structured curriculum. Training is performed entirely offline; deployed relay nodes execute only the inference step a single forward pass through a compact neural network making the approach fully compatible with resource-constrained LoRa hardware. Overall, these results confirm that energy efficiency and reliability in multi-hop LoRa networks are dominated by collision-induced retransmissions rather than hop count alone, and that the PPO-based framework outperforms conventional heuristic routing across all performance metrics by learning interference-aware, load-balanced policies that cooperate implicitly with ADR.

B. Path Diversity and Spatial Traffic Separation

Figures below illustrate the routing behaviour of multiple source nodes before and after PPO learning, highlighting the impact of learned routing policies on spatial traffic separation and collision reduction. Prior to PPO learning, routing paths from different sources converge toward the same set of relays that are geographically closest to the gateway. This convergence creates persistent collision hotspots, as multiple packets are forwarded simultaneously through the same relays. In LoRa networks, when multiple transmissions using the same spreading factor overlap in time, the aggregated interference power ∑I increases, and packet decoding fails if the signal-to-interference-plus-noise ratio, defined as $SINR = P_r/(sumI + N)$, falls below the spreading-factor-dependent threshold SINR_{min} a result, the convergent routing behavior observed before learning leads to a high collision rate of approximately 34.7%. As shown in Fig. 9.

After PPO learning as shown in Fig. 10, the routing paths exhibit clear spatial separation, with each source predominantly selecting relays in distinct regions of the network. This behaviour emerges from the PPO reward formulation, which penalizes path overlap O and collision-induced retransmissions while rewarding successful packet delivery. By observing relay usage history, channel occupancy, and recent transmission activity in the state representation, the PPO agent learns to minimize O, defined as the number of concurrent routing decisions that reuse the same relay within a short time window. Reducing O directly lowers the probability of overlapping transmissions, thereby decreasing the aggregated interference ∑I and increasing the likelihood that the SINR condition for successful decoding is satisfied.

The spatial separation of routing paths also improves energy efficiency and relay fairness. Since the energy consumed per transmission is modelled $E_{tx} = I_{tx} \times V \times T_{air}$, and retransmissions incur the same energy cost as successful transmissions, reducing collision probability significantly lowers total energy consumption. Furthermore, by distributing traffic across multiple relays rather than repeatedly using the same nodes, PPO prevents energy hotspots and slows battery depletion at individual relays. This balanced relay utilization contributes to improved network sustainability and longer operational lifetime.

Overall, the results demonstrate that path diversity is not achieved through random exploration but through learned, interference-aware routing decisions. The PPO-based framework effectively transforms routing behaviour from convergence-driven interference patterns into spatially distributed paths that reduce collisions, improve reliability, and enhance energy efficiency in multi-hop LoRa networks.

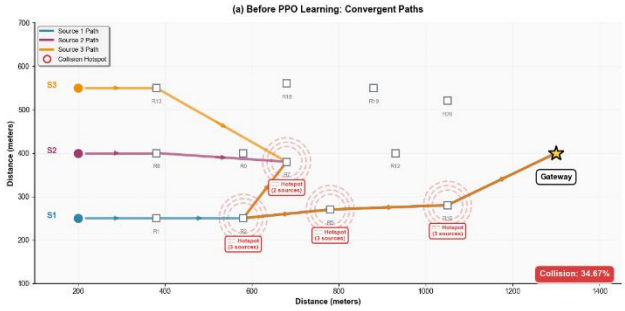

Fig. 9. Before spatial traffic separation.

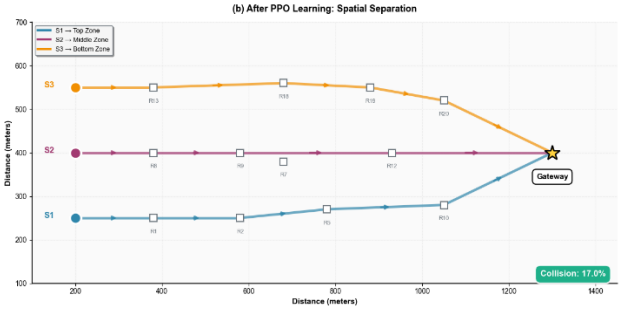

Fig. 10. After spatial traffic separation.

VI. Conclusion

This paper proposes an energy-efficient and reliable routing framework for multi-hop LoRa networks using PPO with dynamic ADR. A discrete-event simulation models realistic LoRa physical-layer effects, co-spreading-factor interference, battery-limited relays, and multi-source contention. Unlike conventional hop-count or heuristic routing, the method formulates routing as a sequential decision problem and learns interference-aware, load-balanced policies via reinforcement learning. Results show significant improvements in packet delivery reliability, collision reduction, and energy consumption per delivered packet compared with Shortest Path, Random routing, and an AODV-like baseline. The learned policy sacrifices minimal hop count to reduce interference and retransmissions, promotes spatial path diversity, mitigates relay energy hotspots, slows battery depletion, and enhances network sustainability.

Overall, energy efficiency is driven primarily by collision-induced retransmissions rather than hop count alone. By accounting for interference, relay load, and transmission history, the PPO-based framework demonstrates the promise of reinforcement learning for interference-limited IoT mesh routing, with physical hardware validation remaining an important direction for future work.

A key limitation of this work is the sim-to-real gap. The propagation model captures the dominant effects of forest path loss, vegetation attenuation, and log-normal shadow fading based on empirically validated parameters [5], [6], [32], [46] however, real deployments additionally involve hardware specific noise figures, adjacent-channel interference, oscillator frequency drift, and the discrete timing constraints of the LoRa duty-cycle and Channel Activity Detection (CAD) mechanism. To reduce sensitivity to these differences, the PPO policy operates on normalised relative state features residual energy ratios, channel-occupancy indicators, and hop-count differences rather than raw absolute signal values, which improves robustness to hardware-to-hardware calibration offsets. Nevertheless, transfer validation on physical LoRa nodes in a controlled forest testbed remains a necessary step before operational deployment.

Acknowledgment

The authors would like to acknowledge Universiti Malaysia Sabah (UMS), Ministry of Higher Education Malaysia (KPT), and Ministry of Science, Technology and Innovation (MOSTI) for supporting this research under Fundamental Research Grant Scheme - Early Career Researcher (FRGS-EC), grant no. FRGS-EC/1/2024/TK07/UMS/02/3 (university code: FRGC056-2024) and Strategic Research Fund (SRF) – Programme Space based Technology, grant no. LPK2414.

References

[1] K. Shafique, B. A. Khawaja, F. Sabir, S. Qazi, and M. Mustaqim, "Internet of things (IoT) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5G-IoT Scenarios," 2020, Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ACCESS.2020.2970118.

[2] 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, Mar. 2019, doi: 10.1016/j.icte.2017.12.005.

[3] F. Yao, Y. Ding, S. Hong, and S.-H. Yang, "A Survey on Evolved LoRa-Based Communication Technologies for Emerging Internet of Things Applications," International Journal of Network Dynamics and Intelligence, pp. 4–19, Dec. 2022, doi: 10.53941/ijndi0101002.

[4] A. I. Griva et al., "LoRa-Based IoT Network Assessment in Rural and Urban Scenarios," Sensors, vol. 23, no. 3, Feb. 2023, doi: 10.3390/s23031695.

[5] G. Callebaut and L. Van Der Perre, "Characterization of LoRa Point-to-Point Path Loss: Measurement Campaigns and Modeling Considering Censored Data," IEEE Internet Things J., vol. 7, no. 3, pp. 1910–1918, Mar. 2020, doi: 10.1109/JIOT.2019.2953804.

[6] R. El Chall, S. Lahoud, and M. El Helou, "LoRaWAN network: Radio propagation models and performance evaluation in various environments in Lebanon," IEEE Internet Things J., vol. 6, no. 2, pp. 2366–2378, Apr. 2019, doi: 10.1109/JIOT.2019.2906838.

[7] A. Lahoua, M. Younes, and L. Touseau, "Improving Energy-Efficiency In Lora Networks Thanks to a Holistic Transmission Model," pp. 469–476, 2024, doi: 10.1109/WiMob61911.2024.10770422ï.

[8] D. Zorbas, C. Caillouet, K. A. Hassan, and D. Pesch, "Optimal data collection time in lora networks— a time-slotted approach," Sensors (Switzerland), vol. 21, no. 4, pp. 1–22, Feb. 2021, doi: 10.3390/s21041193.

[9] D. L. Mai and M. K. Kim, "Multi-hop LOra network protocol with minimized latency," Energies (Basel)., vol. 16, no. 3, Mar. 2020, doi: 10.3390/en13061368.

[10] R. Airiyoshi, M. Hasegawa, T. Ohtsuki, and A. Li, "Energy Efficient Transmission Parameters Selection Method Using Reinforcement Learning in Distributed LoRa Networks," Jan. 2025, [Online]. Available: http://arxiv.org/abs/2410.11270

[11] A. Lahoua, M. Younes, and L. Touseau, "Improving Energy-Efficiency In Lora Networks Thanks to a Holistic Transmission Model," pp. 469–476, 2024, doi: 10.1109/WiMob61911.2024.10770422ï.

[12] M. Alkhayyal and A. M. Mostafa, "Enhancing LoRaWAN Sensor Networks: A Deep Learning Approach for Performance Optimizing and Energy Efficiency," Computers, Materials and Continua, vol. 83, no. 1, pp. 1079–1100, 2025, doi: 10.32604/cmc.2025.061836.

[13] X. Chen, Y. Mao, Y. Xu, W. Yang, C. Chen, and B. Lei, "Energy-efficient multi-hop LoRa broadcasting with reinforcement learning for IoT networks," Ad Hoc Networks, vol. 169, Mar. 2025, doi: 10.1016/j.adhoc.2024.103729.

[14] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," Aug. 2017, [Online]. Available: http://arxiv.org/abs/1707.06347

[15] R. Serati, B. Teymuri, N. A. Anagnostopoulos, and M. Rasti, "ADR-Lite: A Low-Complexity Adaptive Data Rate Scheme for the LoRa Network," Oct. 2022, [Online]. Available: http://arxiv.org/abs/2210.14583

[16] N. Izzeldin, M. Osman, and E. B. Abbas, "Performance Evaluation of LoRa and Sigfox LPWAN Technologies for IoT," Academic Journal of Research and Scientific Publishing |, vol. 4, pp. 5–6, [Online]. Available: www.ajrsp.comwww.ajrsp.com

[17] M. Lauridsen, H. Nguyen, B. Vejlgaard, I. Z. Kovacs, P. Mogensen, and M. Sorensen, "Coverage Comparison of GPRS, NB-IoT, LoRa, and SigFox in a 7800 km Area," in IEEE Vehicular Technology Conference, Institute of Electrical and Electronics Engineers Inc., Nov. 2017. doi: 10.1109/VTCSpring.2017.8108182.

[18] A. Loubany, S. Lahoud, A. E. Samhat, and M. El Helou, "Improving Energy Efficiency in LoRaWAN Networks with Multiple Gateways," Sensors, vol. 23, no. 11, Jun. 2023, doi: 10.3390/s23115315.

[19] M. Bor, U. Roedig, T. Voigt, and J. M. Alonso, "Do LoRa low-power wide-area networks scale?," in MSWiM 2016 - Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, Association for Computing Machinery, Inc, Nov. 2016, pp. 59–67. doi: 10.1145/2988287.2989163.

[20] S. Feng, J. Chen, and Z. Zhao, "Cost Effective Routing in Large-scale Multi-hop LoRa Networks," in INFOCOM WKSHPS 2022 - IEEE Conference on Computer Communications Workshops, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/INFOCOMWKSHPS54753.2022.9798302.

[21] F. Alghamdi and F. Bajaber, "Adaptive Real-Time Channel Estimation and Parameter Adjustment for LoRa Networks in Dynamic IoT Environments," Sensors, vol. 25, no. 7, Apr. 2025, doi: 10.3390/s25072121.

[22] L. Networks, st Yi Jiang, nd Minghua Wang, and rd Xingbin Wang, "A Efficient Adaptive Data Rate Algorithm in," 2023.

[23] M. Coutinho, J. A. Afonso, and S. F. Lopes, "An Efficient Adaptive Data-Link-Layer Architecture for LoRa Networks," Future Internet, vol. 15, no. 8, Aug. 2023, doi: 10.3390/fi15080273.

[24] M. Misbahuddin, M. S. Iqbal, and L. A. S. I. Akbar, "A Multi-Hop Routing Solution for Low Latency and Energy Efficiency in Large-Scale LoRa IoT," International Journal of Computers, Communications and Control, vol. 20, no. 3, 2025, doi: 10.15837/ijccc.2025.3.6501.

[25] R. Pueyo Centelles, R. Meseguer, F. Freitag, R. Baig Viñas, and L. Navarro, "Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000. A minimalistic distance-vector routing protocol for LoRa mesh networks", doi: 10.1109/ACCESS.2017.DOI.

[26] L. Acosta-Garcia, J. Aznar-Poveda, A. J. Garcia-Sanchez, J. Garcia-Haro, and T. Fahringer, "Dynamic transmission policy for enhancing LoRa network performance: A deep reinforcement learning approach," Internet of Things (Netherlands), vol. 24, Dec. 2023, doi: 10.1016/j.iot.2023.100974.

[27] J. Haxhibeqiri, F. Van den Abeele, I. Moerman, and J. Hoebeke, "LoRa scalability: A simulation model based on interference measurements," Sensors (Switzerland), vol. 17, no. 6, Jun. 2017, doi: 10.3390/s17061193.

[28] I. Urabe, A. Li, M. Fujisawa, S. J. Kim, and M. Hasegawa, "Combinatorial MAB-Based Joint Channel and Spreading Factor Selection for LoRa Devices," Sensors, vol. 23, no. 15, Aug. 2023, doi: 10.3390/s23156687.

[29] F. Alghamdi and F. Bajaber, "Adaptive Real-Time Channel Estimation and Parameter Adjustment for LoRa Networks in Dynamic IoT Environments," Sensors, vol. 25, no. 7, Apr. 2025, doi: 10.3390/s25072121.

[30] A. E. Ferreira et al., "A study of the LoRa signal propagation in forest, urban, and suburban environments", doi: 10.1007/s12243.

[31] Y. Wu, G. Guo, G. Tian, and W. Liu, "A Model with Leaf Area Index and Trunk Diameter for LoRaWAN Radio Propagation in Eastern China Mixed Forest," J. Sens., vol. 2020, 2020, doi: 10.1155/2020/2687148.

[32] M. R. Ansah, R. A. Sowah, J. Melià-Seguí, F. A. Katsriku, X. Vilajosana, and W. O. Banahene, "Characterising foliage influence on LoRaWAN pathloss in a tropical vegetative environment," Oct. 01, 2020, Institution of Engineering and Technology. doi: 10.1049/iet-wss.2019.0201.

[33] S. Rougerie, J. Israel, K. Tomoshige, S. Rougerie, J. Israel, and T. Kan, "Validation of ITU-R P.833-9 tree attenuation model for Land Mobile Satellite propagation channel at Ku/Ka band." [Online]. Available: https://hal.science/hal-03231346v1

[34] Z. Xu, S. Tong, P. Xie, and J. Wang, "From Demodulation to Decoding: Toward Complete LoRa PHY Understanding and Implementation," ACM Trans. Sens. Netw., vol. 18, no. 4, Jan. 2023, doi: 10.1145/3546869.

[35] H. Yang, H. Ji, Z. Huang, and X. Wu, "A GNN-Based Learning Approach for Energy Optimization in Relay-Assisted IoT Networks," in IEEE Wireless Communications and Networking Conference, WCNC, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/WCNC61545.2025.10978775.

[36] N. Abdoun, S. Abboud, H. Altaleb, Z. Rajnai, and B. Donát, "Collision Detection in LoRaWAN Using Machine Learning," pp. 1–6, 2024, doi: 10.1109/sisy62279.2024.10737620ï.

[37] Y. A. Al-Gumaei, N. Aslam, M. Aljaidi, A. Al-Saman, A. Alsarhan, and A. Y. Ashyap, "A Novel Approach to Improve the Adaptive-Data-Rate Scheme for IoT LoRaWAN," Electronics (Switzerland), vol. 11, no. 21, Nov. 2022, doi: 10.3390/electronics11213521.

[38] F. Bizzarri, C. Mocenni, and S. Tiezzi, "A Markov Decision Process with Awareness and Present Bias in Decision-Making," Mathematics, vol. 11, no. 11, Jun. 2023, doi: 10.3390/math11112588.

[39] M. Löppenberg, S. Yuwono, M. R. Diprasetya, and A. Schwung, "Dynamic robot routing optimization: State–space decomposition for operations research-informed reinforcement learning," Robot. Comput. Integr. Manuf., vol. 90, Dec. 2024, doi: 10.1016/j.rcim.2024.102812.

[40] Z. Sun, H. Tian, H. Wang, S. Yan, T. Han, and H. Rong, "Utilising a two-dimensional action space as the basis for the PPO algorithm, job shops can jointly schedule AGVs and machinery," Sep. 24, 2025. doi: 10.21203/rs.3.rs-7413549/v1.

[41] M. Zheng, J. Zhang, C. Zhan, X. Ren, and S. Lü, "Proximal policy optimization with reward-based prioritization," Expert Syst. Appl., vol. 283, Jul. 2025, doi: 10.1016/j.eswa.2025.127659.

[42] K. Sun, J. Yang, J. Li, B. Yang, and S. Ding, "Proximal Policy Optimization-Based Hierarchical Decision-Making Mechanism for Resource Allocation Optimization in UAV Networks," Electronics (Switzerland), vol. 14, no. 4, Feb. 2025, doi: 10.3390/electronics14040747.

[43] H. M. Hadi and I. M. Ibrahim, "A Comprehensive Review of Shortest Path Algorithms for Network Routing," Asian Journal of Research in Computer Science, vol. 18, no. 3, pp. 152–175, Feb. 2025, doi: 10.9734/ajrcos/2025/v18i3584.

[44] D. C. Dhanapala, A. P. Jayasumana, and Q. Han, "On random routing in wireless sensor grids: A mathematical model for rendezvous probability and performance optimization," J. Parallel Distrib. Comput., vol. 71, no. 3, pp. 369–380, Mar. 2011, doi: 10.1016/j.jpdc.2010.10.016.

[45] S. Liu, Y. Yang, and W. Wang, "Research of AODV Routing Protocol for Ad Hoc Networks1," AASRI Procedia, vol. 5, pp. 21–31, 2013, doi: 10.1016/j.aasri.2013.10.054.

[46] A. E. Ferreira et al., "A study of the LoRa signal propagation in forest, urban, and suburban environments", doi: 10.1007/s12243.