scientific reports

Check for updates

OPEN

Quantum-secured routing in drone communication for 6G-enabled smart mobility

Sana Hafeez[1], Ghulam E Mustafa Abro[2], Sufyan Ali Memon[3]✉, Talha Ahmed Khan[4]✉, Imran Memon[5] & Haidawati Nasir[4]✉

The emergence of sixth-generation (6G) wireless networks introduces unprecedented requirements for ultra-secure, low-latency communication across heterogeneous space–air–ground integrated network (SAGIN). Existing drone communication frameworks including LoRaWAN, Long Term Evolution, and Ad Hoc mesh architectures exhibit critical vulnerabilities to eavesdropping, jamming, and quantum-computational attacks due to their reliance on classical cryptographic primitives. To address these challenges, this work presents the Quantum-Secured Adaptive Routing Algorithm (QSARA), a novel framework designed for 6G-enabled unmanned aerial vehicle (UAV) networks that integrates Quantum Key Distribution (QKD), Reconfigurable Intelligent Surfaces (RIS), and Joint Communication and Sensing (JCAS) to enhance information-theoretic security and real-time performance. The proposed framework employs a quantum-augmented dynamic graph model to represent UAV swarm networking and uses Proximal Policy Optimisation (PPO)-based deep reinforcement learning to optimise routing under adversarial and uncertain conditions. A multi-objective cost function jointly captures classical quality of service metrics, such as latency, bandwidth, and energy consumption alongside with quantum-layer security indicators, including quantum bit error rate, key pool entropy, and key availability. High-fidelity simulations with 500 mobile drones under diverse adversarial threats demonstrate that the proposed framework achieves a key establishment success rate of 96.2%, end-to-end latency of 23.7 milliseconds, energy consumption of 7.8 watt-hours, and a packet delivery ratio of 94.1%, outperforming state-of-the-art classical and quantum-aware baselines. These results position the QSARA as a scalable and quantum-resilient routing solution for mission-critical UAV networking in next-generation 6G smart mobility ecosystems.

Keywords Quantum Key Distribution (QKD), Reconfigurable Intelligent Surfaces (RIS), Joint Communication and Sensing (JCAS), Sixth-Generation Networks (6G), Quantum-Secured Adaptive Routing Algorithm (QSARA), Smart Mobility, Autonomous Unmanned Aerial Vehicles (UAVs), Space-Air-Ground Integrated Networks (SAGIN)

Background and motivation

Quantum computation threatens the cryptographic foundations of contemporary communication infrastructures. Foundational algorithms undermine widely deployed primitives: Shor's algorithm breaks integer factorisation and discrete logarithms in polynomial time, endangering RSA, Elliptic Curve Cryptography (ECC), and Diffie–Hellman[1]; Grover's algorithm accelerates unstructured search, reducing the effective security of symmetric ciphers and hash functions[2]. In light of these advances, quantum-resilient architectures have become essential for future wireless systems[3,4]. Quantum Key Distribution (QKD) addresses this need by providing information-theoretic key agreement grounded in quantum mechanics rather than computational hardness assumptions[5,6]. The urgency is amplified in anticipated sixth-generation 6GSpace–Air–Ground Integrated Network (SAGIN) integrated networks serving autonomous aerial platforms. Smart mobility roadmaps envisage low-latency, reliable, and context-aware connectivity for mission-critical applications (for example, situational awareness,

[1]Digital Technologies and Artificial Intelligence, Digital Innovation Research Institute, Liverpool John Moores University, Liverpool, UK. [2]Artificial Intelligence in Robotics Laboratory (AiR Lab), Electrical and Computer Engineering Department, Aarhus University, Aarhus C, 8000, Aarhus, Denmark. [3]Department of Defense Systems Engineering, Sejong University, Seoul 05006, Gwangjin-Gu, Republic of Korea. [4]Malaysian Institute of Information and Technology (MIIT), Universiti Kuala Lumpur, 50250 Kuala Lumpur, Malaysia. [5]Department of Computer Science, Shah Abdul Latif University, Shahdadkot campus, Shahdadkot 77300, Pakistan. ✉email: sufyanahmedali@sejong.ac.kr; talha@unikl.edu.my; haidawati@unikl.edu.my

emergency response, and precision logistics)[3,7]. However, mobile Free-Space Optics (FSO) and millimeter-wave (mmWave) links are sensitive to turbulence, weather, and alignment jitter, which degrade quantum signal quality and key rates[8–11]. Reconfigurable Intelligent Surfaces (RIS) can reshape propagation to enhance link robustness and energy efficiency[12–15], and recent testbeds are beginning to combine quantum, semantic, and generative artificial intelligence capabilities[16]. These developments motivate a holistic, security-first routing approach that couples QKD with environment-aware control to sustain confidentiality under mobility.

Quantum threat landscape

Quantum-enabled adversaries can render classical public-key cryptography brittle at scale, while the stochasticity of mobile channels complicates the timely replenishment of secret keys. Satellite and airborne demonstrations confirm feasibility of QKD beyond laboratory settings[8,10,17,18], yet sustained operation depends on atmospheric attenuation (for example, clouds and fog per ITU–R P.840-8) and precise transceiver alignment[8,9,11]. In parallel, PQC offers software-deployable defences with measurable compute and memory footprints at the edge[19,20]. A practical system must therefore arbitrate between QKD and PQC modes as channel and workload conditions evolve[3,4].

Limitations of existing solutions

Standalone QKD on mobile FSO links remains constrained by weather- and motion-induced impairments, finite key buffers, and energy budgets[10]. RIS improves spectral and energy efficiency and can stabilise optical paths, but most communication-centric designs do not close the loop with quantum-layer metrics[12–15]. Recent literature has begun to address secure routing with artificial intelligence or hybrid designs. For fibre/optical backbones, Deep Reinforcement Learning (DRL) has been used for joint routing and resource assignment under QKD constraints [21,22]. In integrated SAGIN settings, quantum-secured routing with dynamic topology has been explored[23]. RIS-assisted multi-user QKD concepts are emerging[24]. Nonetheless, these strands typically assume relatively stable channels or treat quantum security and path optimisation as loosely coupled layers. Classical secure routing protocols and heuristics (for example, A-SAODV and early ad hoc security mechanisms) established important baselines but predate quantum-era threat models and mobile QKD peculiarities[25,26]. On the aerial networking side, blockchain-based control and authentication enhance integrity and accountability[27–31], yet they do not by themselves resolve key sustainability and secrecy–latency trade-offs under hostile conditions. Energy constraints remain particularly salient for Unmanned Aerial Vehicle (UAV), including in QKD -centric designs[32].

Research gap and aim of the paper

The open challenge is to embed quantum-layer awareness into the routing objective and control policy so that (i) secure-key sustainability is preserved despite stochastic mobility and weather; (ii) end-to-end latency and energy budgets remain within operational limits; and (iii) trustworthy fall-back to PQC is triggered only when justified by predicted key shortfalls and risk posture[3,4,19,20]. This work advances a learning-enabled framewor tailored to 6G Networks smart mobility in which RIS-enhanced mmWave and mobile FSO links are co-optimised with QKD processes, and Reinforcement Learning (RL) policies adapt routing and cryptographic mode in response to channel, threat, and key-buffer dynamics[8–15]. The approach aligns with recent efforts to integrate quantum and next-generation networking capabilities[16,33–35] while remaining compatible with established quantum security foundations[5,6,36].

Contribution

This study presents four principal contributions as mentioned below:

1. A rigorous system and channel model is established, integrating RIS–enhanced mmWave links with free-space–optical quantum communication subject to stochastic weather attenuation[8–15].
2. The routing task is formalised as a multi-objective optimisation problem that jointly captures classical performance requirements and quantum-layer security indicators through a tunable cost function[3,4,7].
3. A trust-aware Proximal Policy Optimisation (PPO)–based reinforcement learning mechanism is introduced, enabling adaptive routing under uncertainty while satisfying constraints on latency, energy consumption, Quantum Bit Error Rate (QBER), and key-buffer depletion[21–23,32].
4. A high-fidelity simulation framework is developed, incorporating 500 mobile drones, diverse adversarial threats, and Monte Carlo variability, facilitating systematic benchmarking of the Quantum-Secured Adaptive Routing Algorithm (QSARA) Algorithm against classical and quantum-aware baselines[16,19,20,25–30,33].

Research questions

This study is guided by four research questions that structure the modelling, algorithm design, and evaluation framework as mentioned below:

1. **RQ1:** How can a unified system and channel model capture both classical mmWave communication and FSO quantum links within a mobile drone network[8–15]?
2. **RQ2:** What are the latency, energy, and public-key–infrastructure trade-offs compared with classical and PQC baselines[3,4,19,20]?
3. **RQ3:** How can reinforcement learning be leveraged to generate secure and resource-efficient routing decisions under uncertainty and adversarial conditions[21–23]?
4. **RQ4:** To what extent does the proposed routing framework improve key availability, resilience, and overall network performance when evaluated at scale under stochastic mobility and attack scenarios[25–28,30,32]?

RQ1–RQ4 map to Section 5 (for example, RQ1→ 3, RQ2→Table 2, RQ3→Figure 6, RQ4→Figure 5) and the method in Section 4.

Positioning and quantified motivation

The present formulation differs from prior work by (i) explicitly targeting mobile FSO quantum links in space–air–ground mobility with RIS-aided propagation, (ii) jointly optimising secrecy-aware path selection with secure-key sustainability, and (iii) embedding a standards-compatible PQC fall-back when quantum links degrade[13–15,19,20,23]. Empirically, the relevant trade-off is twofold: PQC key encapsulation and signatures incur compute and energy overheads on resource-constrained platforms[19,20], whereas QKD exhibits rate variability driven by weather and alignment[8–11]. The evaluation, therefore, reports latency and energy per packet alongside secure-key sustainability under controlled channel dynamics and stressors, while situating results within contemporary efforts at the intersection of quantum networking, 6G architectures, and integrated testbeds[16,33–35] (Fig. 1).

System model and channel model

QKD provides information-theoretic security grounded in the principles of quantum mechanics, making it a compelling solution for safeguarding future communication networks. In parallel, PQC has gained prominence due to its seamless integration with classical infrastructure, relying on computationally hard problems such as lattice-based and hash-based schemes. While PQC offers broad compatibility[3738], it remains vulnerable to unforeseen algorithmic breakthroughs and side-channel attacks. Conversely, QKD inherently detects eavesdropping attempts but requires specialised infrastructure and does not natively interoperate with classical protocols[39].

Hybrid approaches that combine QKD and PQC have been proposed to address individual limitations. However, they often introduce integration overhead and key management complexity, particularly in high-mobility, adversarial environments such as UAV networks. These limitations motivate the need for a unified framework such as QSARA that embeds quantum-layer security directly into network operations while dynamically adapting to mobility patterns, topology changes, and hostile threats. To address these challenges, this section formalises the system and channel models that underpin QSARA, incorporating RIS-assisted quantum-classical communication, mobility-aware trust estimation, and dynamic spectrum and propagation considerations. The framework operates within the context of a future 6G -enabled SAGIN, accounting for heterogeneous connectivity, adversarial interference, and key distribution constraints.

Throughout this work, "information-theoretic security" refers to the guaranty provided by QKD protocols, namely, secrecy against computationally unbounded adversaries under standard assumptions such as an authenticated classical channel, with correctness and secrecy established by canonical security proofs (e.g., Bennett-Brassard 1984 Quantum Key Distribution Protocol (BB84) via Shor–Preskill). In deployment, practical imperfections are captured explicitly in secure key-rate model via privacy amplification and error-correction inefficiency, so only keys distilled under acceptable QBER/fidelity are considered secure for routing. Operationalisation is achieved by enforcing feasibility constraints (QBER $\leq$ QBERmax, key pool $\geq K$min, lifetime $\geq \tau_{\min}$) together with a security-level threshold $S_{\max}$. Any link that does not satisfy these conditions is excluded from the quantum tier, and the system transitions to post-quantum cryptography for that path. This clarifies that the statement refers to the ideal guarantees provided by quantum key distribution, while the

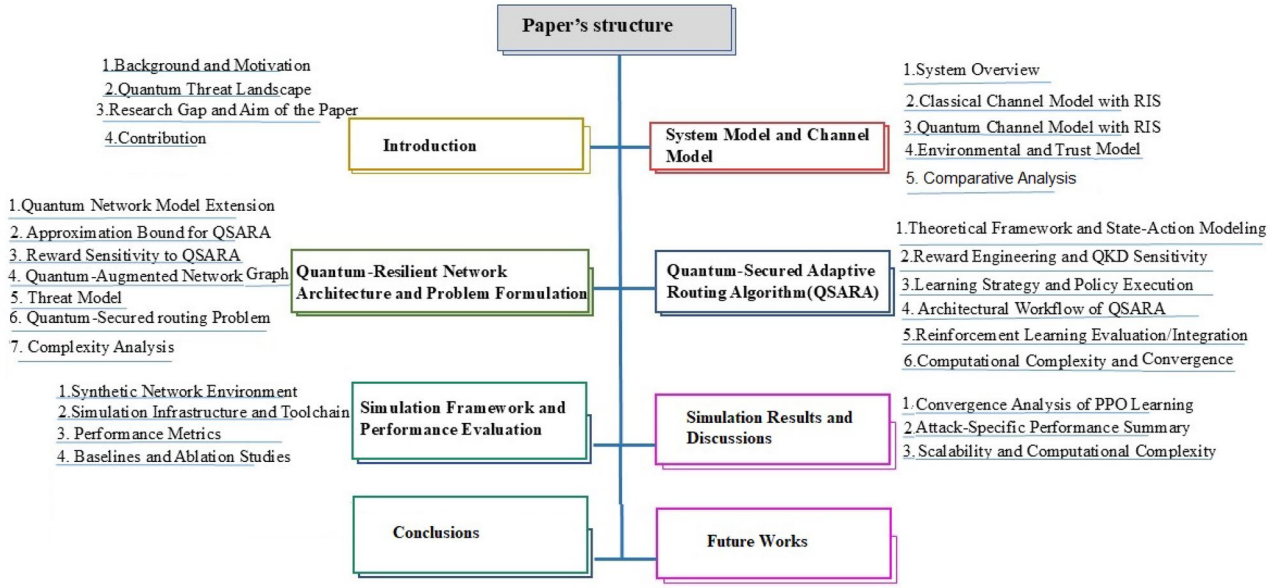

Fig. 1. Organizational Structure of the Paper on Quantum-Secured Adaptive Routing.

simulations and imposed constraints ensure that only links whose measured parameters satisfy those guarantees are selected in practice.

System overview

A 6G-enabled autonomous drone network operating within a SAGIN network is considered. The drone fleet $\mathcal{N} = \{n_1, n_2, \ldots, n_N\}$ comprises UAV) equipped with quantum communication transceivers, RIS, and onboard computational agents. At any time t, the network is represented as a directed dynamic graph $G(t) = (V(t), E(t))$, where $V(t)$ denotes the active drones and $E(t)$ the available quantum/classical links.

Figure 2 presents the high-level system architecture of QSARA, highlighting the integration of QKD, RIS and Joint Communication and Sensing (JCAS) modules, and reinforcement learning-based routing decisions within a SAGIN-enabled UAV network.

Qubit representation and superposition

A single qubit can be represented as a point on the Bloch sphere, capturing its state in a unitary Hilbert-space

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle, \tag{1}$$

where $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$ parameterise the qubit's position on the Bloch sphere. QBER can be interpreted as the angular deviation from the intended state on this sphere. This representation quantifies decoherence and disturbance due to eavesdropping or channel noise in QKD-based communication.

Quantum encoding and basis selection

Qubits are encoded using photon polarisation or phase-based encodings. In BB84, rectilinear ($\{|0\rangle, |1\rangle\}$) and diagonal ($\{|+\rangle, |-\rangle\}$) bases are randomly chosen. Upon measurement, a qubit collapses probabilistically to $|0\rangle$

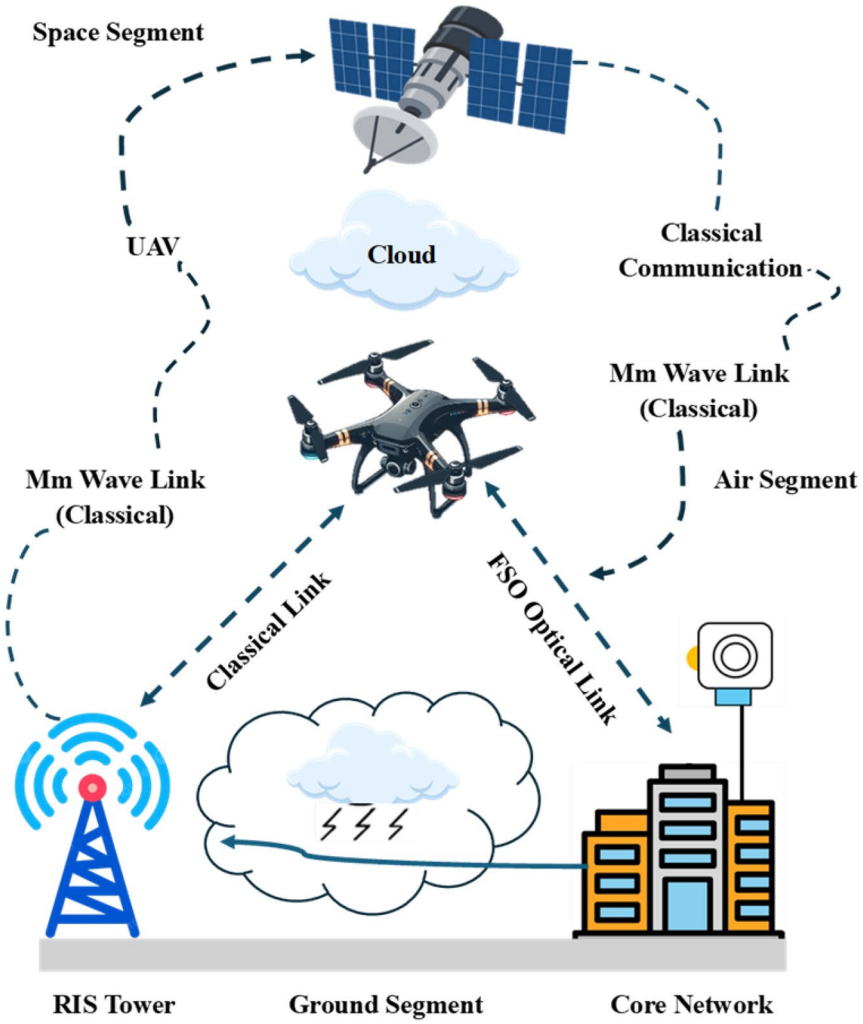

Fig. 2. QSARA System Architecture.

or $|1\rangle$ based on the projection axis, a key property that introduces uncertainty and underpins QBER and fidelity calculations.

Environmental decoherence model
Atmospheric decoherence is modelled by representing the quantum-state evolution through a depolarising channel given by

$$\mathscr{E}(\rho) = (1-p)\rho + p\frac{I}{2}, \tag{2}$$

where p is the decoherence probability (derived from ITU-R P.840-8 data), ρ is the input density matrix, and I is the identity matrix. The decoherence contributes directly to QBER increase and Secure Key Rate (SKR) degradation.

Entanglement-based communication and trust
If EPR-based entanglement is used, a shared quantum state

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{3}$$

is distributed across drones. Fidelity measurements are used to monitor the integrity of this entanglement. Bell-state projections help estimate deviation from the ideal state and contribute to the trustworthiness of the link. Table 1 describes the mathematical symbols and notations used in this paper.

Time synchronisation and key buffering
Key pool evolution $k_{ij}(t)$ accounts for stages of QKD: Sifting, error correction, and privacy amplification. These processes are buffered temporally to align with classical control signals. A scheduler ensures synchronisation across nodes for key updates and route selection.

Drone-specific quantum constraints
Constraints include limited battery capacity for QKD hardware, line-of-sight misalignments due to UAV mobility, and storage limitations for onboard quantum memory. These limitations are factored into SKR predictions and trust assessments.

Classical channel model with RIS

Each UAV employs mmWave links for classical communication, enhanced by RIS panels for signal reflection and beamforming. The received signal at node *j* from node *i* is

$$y_j = \left(h_{\mathrm{dir}} + h_r^{\top}\Phi g\right) x_i + n_j \tag{4}$$

where h_{dir} denotes the direct channel gain, and h_r and g represent the RIS-to-receiver and transmitter-to-RIS channels, respectively. The matrix $\Phi = \mathrm{diag}(e^{j\phi_1}, \ldots, e^{j\phi_M})$ defines the phase-shift configuration of the RIS. The term n_j corresponds to additive white Gaussian noise at the receiver node *j*. The associated path loss is modelled under a Rician fading environment as follows

$$L_{ij} = \left(\frac{\lambda}{4\pi d_{ij}}\right)^2 \kappa_{\mathrm{env}}^{-1} \tag{5}$$

where d_{ij} is the distance and κ_{env} captures attenuation due to environmental effects (e.g., fog, rain). To better reflect realistic aerial scenarios, Stochastic channel effects, including misalignment noise and log-normal atmospheric fading, are incorporated using the Hufnagel–Valley turbulence model. These additions allow for probabilistic QBER estimation, impacting the SKR and fidelity metrics under adversarial and environmental stressors.

Quantum channel model with RIS

Quantum links utilise FSO communication enhanced by RIS. The SKR between nodes *i* and *j* at time *t* is

$$R_s^{(i,j)}(t) = \eta_{\mathrm{QKD}}\,\mu\,R_{\mathrm{rep}}\,T_{\mathrm{eff}}^{(i,j)} \cdot \left[1 - (1+f_{\mathrm{EC}})H_2\left(\mathrm{QBER}^{(i,j)}(t)\right)\right], \tag{6}$$

$$T_{\mathrm{eff}}^{(i,j)} = \eta_{\mathrm{opt}}\eta_{\mathrm{RIS}}\,e^{-\alpha_{\mathrm{atm}}d_{ij}}\cos^2(\theta_{\mathrm{mis}}) \tag{7}$$

where, $R_s^{(i,j)}(t)$ denotes the SKR (in bits per second) between drones *i* and *j* at time *t*, representing the net generation rate of privacy-amplified, error-corrected quantum key material. The term η_{QKD} denotes the overall efficiency of the QKD system, accounting for detector performance and encoding accuracy, while μ is the mean photon number per pulse, typically less than 1 in weak coherent pulse schemes. R_{rep} refers to the repetition rate of the quantum source in Hz, dictating the frequency of qubit transmission. The effective transmittance $T_{\mathrm{eff}}^{(i,j)}$ models the cumulative attenuation due to atmospheric absorption, RIS beam steering, and alignment

Symbol	Description
Graph and Quantum Primitives	
$G(t)$	Time-evolving communication graph at time t
$\|G\rangle$	Quantum graph state representing multipartite entanglement
$CZ_{i,j}$	Controlled-Z operation between qubits i and j
$\|+\rangle$	Initial superposition state $(\|0\rangle + \|1\rangle)/\sqrt{2}$
Quantum Channel Metrics	
$\rho_{ij}(t)$	Density matrix of the quantum state shared between nodes i and j
$F_{ij}(t), f$	Entanglement fidelity at time t
$\epsilon_{ij}(t)$	Quantum Bit Error Rate (QBER)
$r_{ij}(t), r$	Secure key rate (bits/s)
$k_{ij}(t), k$	Quantum key pool size (bits)
$\tau_{ij}(t), \tau$	Remaining lifetime of key material (s)
η_{QKD}	QKD system efficiency
μ	Mean photon number per pulse
R_{rep}	Repetition rate of the quantum source (Hz)
$T_{\text{eff}}^{(i,j)}$	Effective transmittance (channel + RIS effects)
$H_2(\cdot)$	Binary entropy function
f_{EC}	Error correction inefficiency factor
Classical Channel Metrics	
$b_{ij}(t), b$	Bandwidth (bps)
$\delta_{ij}(t), d$	Link delay (s)
l	Packet loss
s	Signal strength
c	Energy consumption
$\chi_{ij}(t)$	Normalised communication cost (energy proxy)
y_j	Received signal at node j
h_{dir}, h_r, g	Direct and RIS-related channel gains
λ	Wavelength of the carrier signal
d_{ij}	Euclidean distance between nodes i and j
κ_{env}	Environmental attenuation factor
θ_{mis}	RIS misalignment angle
Φ	RIS phase shift matrix
Trust and Security Parameters	
$T_{ij}(t)$	Trust score assigned by node i to node j at time t
$D_{ij}(t), R_{ij}(t)$	Direct and recommended trust components
w_D, w_R	Weights for direct and recommended trust evidence
α	Trust memory factor (EMA weight)
γ	Fidelity smoothing factor (EMA)
ℓ_{tag}	Authentication tag length (bits/message)
$N_{\text{msg}}^{(i,j)}$	Number of messages on link (i, j)
ρ	Key usage factor (bits per protected data bit)
$B_{\text{data}}^{(i,j)}$	Data rate on link (i, j) (bps)
$S_{ij}(t)$	Security classification of link (i, j)
λ_{quantum}	Highest link security level (QKD-based)
$\lambda_{\text{post-quantum}}$	Post-quantum cryptographic security level
$\lambda_{\text{classical}}$	Classical cryptographic security level
Routing and Reinforcement Learning Parameters	
C_{perf}	Classical QoS-based routing cost
C_{sec}	Quantum-layer security routing cost
Continued	

Symbol	Description
$C(P, t)$	Total routing cost for path P at time t
w_b, w_d, w_l, w_c	Weights for bandwidth, delay, loss, cost
$w_k, w_r, w_\epsilon, w_f, w_\tau$	Weights for quantum-layer metrics
$\xi(t)$	RL reward function at time t
π_θ	Policy function with parameters θ
a_i, s_i	Action and state at step i
$R(s, a)$	Composite reward for state s and action a
$\mathcal{O}(\cdot)$	Computational complexity order

Table 1. Comprehensive Mathematical Symbols and Notations.

errors between the Unmanned Aerial Vehicles (UAVs). The function $H_2(\cdot)$ is the binary entropy function that quantifies the uncertainty induced by the QBER, $QBER^{(i,j)}(t)$. Lastly, f_{EC} represents the inefficiency of classical error correction protocols (typically $f_{\mathrm{EC}} \geq 1$), accounting for redundancy during reconciliation of mismatched key bits between communicating nodes. where the effective transmittance is with η_{RIS} as the RIS beam focusing efficiency and θ_{mis} the misalignment angle.

Each UAV pair establishes a BB84 -like weak-coherent-pulse QKD session over an RIS-assisted FSO hop. The SKR follows (6) and (7) capturing atmospheric loss, RIS gain, and alignment jitter. QBER arises from a depolarising/turbulence model and misalignment noise, and feeds the binary-entropy term in the SKR. Keys are buffered and consumed via the explicit pool update $k_{ij}(t + \Delta t) = k_{ij}(t) + r_{ij}\Delta t - u_{ij}\Delta t$; feasibility requires QBER $\leq 11\%$, $K_{\min} = 10$kbit, and $\tau_{\min} = 100$s. When residual lifetime drops below $T_{\min}$, the system re-initialises QKD while traffic continues via fallback as needed. Implementation uses MATLAB for mobility/topology and Qiskit 0.31.0 for QKD processes (sifting, QBER evolution, privacy amplification); core parameters include $R_{\mathrm{rep}} = 50\,\mathrm{MHz}, \quad \mu = 0.5, \quad f_{\mathrm{EC}} = 1.15, \quad \eta_{\mathrm{QKD}} = 0.5.$

Environmental and trust model

The environmental state $W(t)$, modelled as a stochastic process, affects channel quality and availability. Each link has an associated trust score $T_{ij}(t)$ updated using

$$T_{ij}(t+1) = \alpha T_{ij}(t) + (1-\alpha)\left(w_D D_{ij}(t) + w_R R_{ij}(t)\right) \tag{8}$$

The trust score $T_{ij}(t)$ quantifies the confidence that node *i* has in node *j* at time *t*, reflecting both the historical behaviour and recent interactions of node *j*. The update rule is governed by an exponential moving average, where the parameter $\alpha \in [0, 1]$ controls the memory of the system. A higher α places more weight on past trust ($T_{ij}(t)$), promoting stability, while a lower α makes the trust score more responsive to recent observations. The terms $D_{ij}(t)$ and $R_{ij}(t)$ represent the direct and recommended trust values, respectively. $D_{ij}(t)$ is computed from node *i*'s own interactions with node *j* (e.g., packet delivery success, latency behaviour), whereas $R_{ij}(t)$ aggregates third-party feedback from neighbouring nodes about *j*. The weights w_D and w_R regulate the relative influence of firsthand and secondhand trust evidence. This formulation allows for a robust and adaptive trust evolution mechanism that dynamically captures behavioural deviations, potential misbehaviour, and trust propagation in distributed drone swarms. To facilitate a clearer understanding of the diverse parameters used in subsequent optimisation and learning stages. The network metrics are categorised into classical and quantum layers as follows.

- **Classical Metrics:** Bandwidth (*b*), delay (*d*), packet loss (*l*), signal strength (*s*), energy consumption (*c*).
- **Quantum Metrics:** QBER, ϵ, SKR (*r*), fidelity (*f*), quantum key pool (*k*), key lifetime (τ).

Comparative analysis of existing quantum-aware protocols

To highlight the motivation behind the proposed QSARA framework, A comparative analysis is presented of existing quantum-aware and classical routing protocols designed to secure UAV networks under the emerging sixth-generation paradigm. Table 2 summarises the security methodology, environmental robustness, mobility adaptation, energy efficiency, and optimisation capabilities of each category of related work.

Quantum-resilient network architecture and problem formulation

The QSARA framework is formalised through a unified architectural and optimisation model that integrates quantum security, adaptive routing, and dynamic threat resilience for 6G drone networks. Building upon the physical and channel models defined earlier, A unified framework is introduced that integrates entanglement-aware quantum graph extensions, formal threat modelling, and the multi-objective optimisation required for the proposed QSARA.

Quantum network model extensions

To enable entanglement-assisted functionalities such as quantum teleportation, distributed quantum sensing, and secure multi-party protocols, The classical time-evolving communication graph $G(t)$ is extended to accommodate quantum correlations. This is achieved through two complementary abstractions.

Approach	Security Method	Mobility Support	Environmental Adaptability	Energy Efficiency	Real-Time Optimisation
Standalone QKD[16]	Information theoretic (Quantum)	Limited	Poor	High consumption	Static
Post-Quantum Cryptography PQC[4,9]	Computational hardness	Good	Moderate	Moderate	Limited
Hybrid QKD-PQC[8,9]	Dual-layer cryptographic	Moderate	Moderate	Low	Static
Satellite-assisted QKD[17,18]	Quantum physical layer	Limited	Weather dependent	Very low	Poor
RIS-enhanced Classical[13,14]	Classical encryption with RIS	Excellent	Good	Good	Good
Quantum-Secured Routing (QDR, QGR)[19,23]	QKD-based static routing	Poor	Limited	Moderate	Static
SDN-based Quantum Routing[12,15]	Centralised quantum keying	Moderate	Limited	Moderate	Moderate
Trust-Based Classical[30]	Behavioural trust models	Good	Good	Good	Good

Table 2. Comparative Analysis of Existing Quantum-Aware Protocols for UAV Networks.

Quantum Graph State Representation: A quantum graph state $|G\rangle$ is a special type of multipartite entangled state that represents the connectivity of a quantum network. It is defined as

$$|G\rangle = \prod_{(i,j)\in E} CZ_{i,j}|+\rangle^{\otimes|V|}, \tag{9}$$

where $CZ_{i,j}$ denotes the controlled-Z operation applied between qubits at nodes i and j, and $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ is the initial superposition state of each node. The product is taken over all entangled links in the edge set E. This representation encodes entanglement topologies in a formal structure that supports distributed quantum computation and secure communication primitives.

Tensor Network State Representation To describe the global quantum state of the network where links may carry mixed (non-pure) quantum states due to noise and decoherence, A tensor-network formulation is employed, and the overall network state is given by

$$\mathscr{H}_{\text{net}}(t) = \bigotimes_{(i,j)\in E(t)} \rho_{ij}(t), \tag{10}$$

where $\rho_{ij}(t)$ is the time-varying density matrix of the bipartite quantum system shared between nodes i and j at time t. This captures probabilistic mixtures of quantum states, enabling the framework to account for realistic imperfections in quantum state preparation and transmission. To assess the reliability of quantum entanglement on each link, Its fidelity is monitored using an exponential moving average

$$F_{ij}(t+1) = (1-\gamma)F_{ij}(t) + \gamma F_{ij}^{\text{meas}}(t), \tag{11}$$

where $F_{ij}(t)$ is the estimated fidelity at time t, $F_{ij}^{\text{meas}}(t)$ is the directly observed fidelity measurement during entanglement verification, and $\gamma \in [0,1]$ is a smoothing factor controlling sensitivity to new measurements. A high fidelity ($F_{ij} \approx 1$) indicates a reliable entangled link, while degradation signals quantum decoherence or potential attack vectors. These state representations collectively enable the proposed QSARA framework to dynamically adapt routing decisions based on entanglement health and quantum channel stability.

Approximation bound for QSARA

The performance gap between the routing cost obtained by the learned QSARA policy and the optimal policy in time-varying quantum-augmented network graphs is analytically bounded. Let $C^*(P,t)$ denote the optimal cumulative cost of path P at time t, and $C_{\text{QSARA}}(P,t)$ the cost induced by the learned policy π_θ after T training steps.

Proof Sketch Under standard assumptions (ergodicity, bounded rewards, Lipschitz continuity of policy gradients), the Robbins–Monro stochastic approximation guarantees convergence of the PPO-trained policy π_θ to a near-optimal routing solution. Specifically, the regret diminishes at rate $\mathscr{O}(1/\sqrt{T})$, yielding

$$\frac{C_{\text{QSARA}}(P,t)}{C^*(P,t)} \leq 1+\varepsilon, \qquad \varepsilon = \mathscr{O}(1/\sqrt{T}). \tag{12}$$

□

Reward sensitivity to QBER

To effectively guide the reinforcement learning agent toward secure and efficient paths, the QSARA framework employs a composite reward function that integrates quantum security awareness directly into the policy objective. A key component of this function is the quantum-layer penalty associated with the QBER. The instantaneous quantum reward R_{qkd} at time t is expressed as

$$R_{\text{qkd}} = -\log(1 + \text{QBER}) + \lambda_1 P_{\text{success}}, \tag{13}$$

where $\text{QBER} \in [0, 0.5]$ is the measured error rate on the quantum channel for a given link, and P_{success} is the estimated probability of successful QKD session completion. The logarithmic form of the QBER penalty ensures that even small increases in QBER (e.g., due to eavesdropping or decoherence) are heavily penalised, guiding the agent to avoid links with early signs of quantum degradation.

This derivative is steepest when QBER is low, encouraging strong discrimination between high-fidelity and slightly noisy links during the early phases of training. As QBER increases, the penalty gradient flattens, preventing instability in policy updates. This gradient structure is instrumental in achieving convergence under noisy or adversarial network conditions, especially in scenarios where jamming or misalignment causes intermittent QBER spikes. It also introduces an intrinsic robustness by desensitising the agent to low-information, high-error links that would otherwise introduce non-convergent exploration trajectories. This QKD-aware reward formulation, embedded within PPO's clipped surrogate loss function, ensures that QSARA prioritises entanglement-preserving, high-fidelity paths that maintain long-term quantum security, energy efficiency, and trust continuity.

Quantum-augmented network graph

The dynamic drone communication network is modelled as a time-evolving directed graph $G(t) = (V(t), E(t), A(t), Q(t))$, where $V(t)$ denotes the set of active UAV nodes at time t, and $E(t)$ is the set of directed communication links, encompassing both classical and quantum channels. Each link $(i, j) \in E(t)$ is associated with a set of classical attributes $a_{ij}(t)$ and quantum parameters $q_{ij}(t)$. The classical attribute vector is defined as $a_{ij}(t) = \langle b, d, l, c, s \rangle$, where b denotes the available bandwidth in bits per second, d represents the communication latency, l is the packet loss probability, c is the communication cost (which may represent energy or economic expenditure), and s refers to the signal strength.

The quantum attribute vector is defined as $q_{ij}(t) = \langle k, r, \epsilon, f, \tau \rangle$, where k is the current quantum key pool size (in bits), r is the secure key generation rate via QKD, ϵ denotes the QBER, f is the quantum channel fidelity, and τ is the remaining key lifetime before exhaustion. These metrics jointly characterise the quantum-layer security properties of each communication link.

The time evolution of the key pool is governed by a linear update rule. At time $t + \Delta t$, the key pool for a link (i, j) is given by

$$k_{ij}(t + \Delta t) = k_{ij}(t) + r_{ij}(t)\Delta t - u_{ij}(t)\Delta t. \tag{14}$$

where $u_{ij}(t)$ is the key consumption rate in bits per second, and $r_{ij}(t)$ is the QKD-derived key generation rate over the link. To derive Equation (10b), The secure key generation rate over a free-space optical FSO quantum link between UAVs i and j is modelled under realistic atmospheric and hardware constraints. The QKD protocol employed is assumed to follow a weak coherent pulse scheme such as BB84, where the photon source emits pulses at a repetition rate R_{rep}, with an average photon number per pulse denoted by μ.

The raw key generation rate is scaled by η_{QKD}, which captures intrinsic QKD efficiency, including basis sifting loss, imperfect modulation, and hardware synchronisation inefficiencies. The emitted pulses propagate through an FSO channel that is subject to geometric spreading and atmospheric attenuation. Geometric loss is quantified by the term $\frac{A_{\text{rx}}}{4\pi L_{ij}^2}$, where A_{rx} denotes the receiver aperture area and L_{ij} is the distance between UAVs i and j.

Atmospheric transmittance is modelled by T_{atm}, which accounts for medium absorption and scattering effects due to environmental conditions such as fog, rain, or haze. This is typically derived from ITU-R P.840-8 standards. The effective transmissivity of the quantum channel is further modulated by the transmitter optical efficiency η_{tx} and detector quantum efficiency η_{det}, resulting in an overall effective transmittance term $T_{\text{eff}}(t)$.

Following transmission and detection, the raw key material undergoes classical post-processing, which includes both error reconciliation and privacy amplification. These steps eliminate bits that have been potentially compromised by eavesdroppers or corrupted by quantum noise. The QBER, denoted by ϵ, characterises the degradation in key quality. The binary entropy function $H_2(\epsilon)$ quantifies the uncertainty introduced by QBER and represents the fraction of key bits that must be discarded. Classical error correction introduces an additional inefficiency captured by the reconciliation parameter $f_{\text{EC}} \geq 1$, accounting for the redundancy needed to resolve mismatched key bits between communicating nodes.

The net usable fraction of the raw key material, after applying these post-processing stages, is thus given by the expression $[1 - H_2(\epsilon) - f_{\text{EC}} H_2(\epsilon)]$. This factor multiplies the raw transmission rate to yield the secure key generation rate. Combining these factors, the final expression for the wireless secure key generation rate is expressed in (15)

$$r_{ij}^{\text{wireless}}(t) = \eta_{\text{QKD}}\, \mu\, R_{\text{rep}}\, \frac{\eta_{\text{tx}}\, A_{\text{rx}}}{4\pi L_{ij}^2}\, T_{\text{atm}}\, \eta_{\text{det}} \times [1 - H_2(\epsilon) - f_{\text{EC}} H_2(\epsilon)] \tag{15}$$

This expression describes the overall rate of secure key bits per second achieved through privacy amplification and error correction over a mobile quantum link with environmental limitations. Integrating physical-layer attenuation and quantum-layer post-processing into the key rate model, this equation offers a sound metric that guides QSARA routing choices while considering both performance and cryptographic limits, where η_{QKD} signifies the efficiency of the QKD system, accounting for hardware imperfections like detector efficiency and optical coupling loss. The parameter μ is the mean photon number per pulse, typically less than one for weak

coherent states used in BB84-type protocols[6]. R_{rep} is the repetition rate of quantum pulses in hertz. The term η_{tx} denotes the transmitter efficiency, while A_{rx} is the aperture area of the receiver. The factor L_{ij} represents the distance between UAVs *i* and *j*, and T_{atm} models atmospheric transmittance, typically following empirical standards such as ITU-R P.840-8. The efficiency of the photon detector is represented by η_{det}.[1]

The expression includes a privacy amplification component involving the binary entropy function $H_2(\epsilon)$, which quantifies the uncertainty introduced by the QBER ϵ, and an error correction inefficiency term $f_{\mathrm{EC}} \geq 1$, representing the overhead of classical reconciliation protocols such as Low-Density Parity-Check (LDPC) or Cascade. This formulation enables a precise quantification of the SKR available on a per-link basis, accounting for both channel conditions and hardware constraints. Together, these formulations model the interplay between quantum and classical link attributes, enabling the QSARA framework to make routing decisions that consider both secure key availability and network performance in real-time.

Threat model

In the quantum-resilient design of drone networks, a powerful adversary (Eve) is assumed, equipped with both classical and quantum attack capabilities. Passive eavesdropping on quantum links necessarily induces observable disturbances due to the no-cloning theorem; these manifest as elevated QBER, enabling statistical detection.

Beyond passive interception, the adversary is capable of active manipulation through classical vectors, including packet injection, replay, and tampering with routing advertisements to effect topology poisoning. Timing and power-based side-channel leakage is also considered for partial key inference. Intrusion epochs are modelled using Poisson-distributed inter-arrival times, generating bursty attack phases consistent with adversarial swarming.

To capture cryptographic risk, quantum algorithms such as Shor's (against RSA/ECC) and Grover's (for exhaustive key search) are assumed available to the adversary; hence, classical-only protection is considered insufficient. The likelihood of detection in the quantum layer is lower bounded by

$$P_{\mathrm{detect}} \geq 1 - (1 - p_{\mathrm{int}})^n, \tag{16}$$

where p_{int} is the probability of disturbing an individual qubit and *n* is the number of transmitted qubits.

Authentication is modelled with message authentication codes. The forgery probability is bounded by

$$P_{\mathrm{forge}} \leq \frac{t}{2^k} + \frac{1}{2^t}, \tag{17}$$

where *k* is the secret-key length and *t* is the Message Authentication Code (MAC) tag length.

Attack model and parameterisation

Adversarial events are modelled using explicit stochastic processes to enable reproducibility and stress-controlled evaluation. Jamming arrivals follow a Poisson process with rate λ (events/s); each event has an on-duration τ_{on} and mean inter-arrival $1/\lambda$. Topology poisoning injects falsified neighbour or link-state information over windows of length τ_{tp} at a corruption fraction ρ_{tp}. Side-channel leakage exposes a bounded observation window T_{leak} with leakage strength α (fraction of key/metadata per unit time). Recovery is governed by a threshold θ_{rec} on estimated link trust or key-buffer level; mitigation is triggered when the decision statistic exceeds γ_{det} within a detection window *W* (Table 3).

Quantum–secured routing problem

In highly dynamic and adversarial 6G-enabled drone networks, routing must jointly optimise classical performance and quantum-layer security. Path selection is formalised on a time-varying, quantum-augmented graph $G(t)$.

Let $P = (v_1, \ldots, v_k)$ be a simple path from source v_1 to destination v_k through intermediate UAV. The total path cost at time *t* is a convex combination of a classical performance term C_{perf} and a quantum security term C_{sec}

$$C(P,t) \;=\; \alpha\, C_{\mathrm{perf}}(P,t) \;+\; (1-\alpha)\, C_{\mathrm{sec}}(P,t), \qquad \alpha \in [0,1], \tag{18}$$

where larger α prioritises throughput/latency, and smaller α prioritises cryptographic robustness and quantum-link stability. All link metrics below are *dimensionless and normalised* (e.g., via affine or reference scaling) to ensure comparability; small $\varepsilon > 0$ constants are used where needed for numerical stability.

Classical performance cost. For link (v_i, v_{i+1}) at time *t*, let $b_{ij}(t)$ denote available bandwidth, $\delta_{ij}(t)$ the per-link delay, $p_{ij}(t)$ the packet-loss probability, and $\chi_{ij}(t)$ a (normalised) communication cost proxy (e.g., energy). With non-negative weights $w_b, w_\delta, w_p, w_\chi$ satisfying $w_b + w_\delta + w_p + w_\chi = 1$,

$$C_{\mathrm{perf}}(P,t) \;=\; \sum_{(v_i,v_{i+1})\in P} \left(w_b\, \frac{1}{b_{ij}(t)+\varepsilon} + w_\delta\, \delta_{ij}(t) + w_p\, p_{ij}(t) + w_\chi\, \chi_{ij}(t) \right). \tag{19}$$

[1] ITU-R P.840-8, "Attenuation due to clouds and fog," International Telecommunication Union, 2019. Available at: https://www.itu.int/rec/R-REC-P.840/en

Attack	Symbol	Definition / Role
Jamming	λ_{J}	Poisson arrival rate of jamming epochs (events/s).
	τ_{J}	Mean jamming duration per epoch (s).
	P_{J}	Jammer transmit power or effective SINR reduction (dB).
	δ_{J}	Duty cycle within an epoch (fraction of time active).
Topology poisoning	λ_{T}	Arrival rate of poisoning attempts (events/s).
	τ_{T}	Mean dwell time of injected false routes (s).
	ρ_{T}	Fraction of compromised nodes/links involved.
	θ_{rec}	Recovery threshold (e.g., consecutive valid updates or time) to purge tainted state.
Side-channel	λ_{S}	Arrival rate of side-channel observation windows (events/s).
	W_{S}	Observation window length (s).
	ℓ_{S}	Leakage fraction per window (bits per key-bit, or mutual information).
	α_{S}	Detector false-alarm target for side-channel anomaly scoring.
Quantum-layer alarms	θ_{QBER}	QBER alarm threshold triggering quantum-link quarantine.
	$B_{\min}$	Minimum key-buffer watermark for crypto-mode switching (QKD→PQC).
System recovery	τ_{cool}	Cool-down time before re-enabling previously quarantined links (s).
	κ_{vote}	Quorum size for trust/state revalidation (nodes or samples).

Table 3. Attack and detection parameterisation used across all experiments. Inter-arrival times are $\mathrm{Exp}(\lambda)$; durations are exponential with mean τ.

The reciprocal on bandwidth encodes preference for higher capacity; alternative monotone transforms (e.g., $-\log b_{ij}$) are admissible if more stable in practice.

Quantum security cost. For link (v_i, v_{i+1}) at time t, let $k_{ij}(t)$ be the *key-pool occupancy* (bits), $r_{ij}(t)$ the SKR (bits/s), $\epsilon_{ij}(t)$ the QBER, $f_{ij}(t)$ the fidelity, and $\tau_{ij}(t)$ the *residual key lifetime* (s). With non-negative weights $w_k, w_r, w_\epsilon, w_f, w_\tau$ satisfying $w_k + \cdots + w_\tau = 1$,

$$C_{\mathrm{sec}}(P,t) \;=\; \sum_{(v_i,v_{i+1})\in P} \Big(w_k \,\frac{1}{k_{ij}(t)+\varepsilon} + w_r \,\frac{1}{r_{ij}(t)+\varepsilon} + w_\epsilon\, \epsilon_{ij}(t) + w_f \,\big(1 - f_{ij}(t)\big) + w_\tau \,\frac{1}{\tau_{ij}(t)+\varepsilon} \Big). \tag{20}$$

Low key pools, low SKR, high QBER, poor fidelity, and short key lifetime all increase cost, driving the optimiser away from insecure links even when classical QoS is attractive.

Security-level semantics and constraints. To enforce consistent security semantics, each link (i, j) is assigned a discrete security level

$$S_{ij}(t) \;\in\; \{\lambda_{\mathrm{classical}},\, \lambda_{\mathrm{post\text{-}quantum}},\, \lambda_{\mathrm{quantum}}\},$$

with the strict ordering

$$\lambda_{\mathrm{quantum}} \;>\; \lambda_{\mathrm{post\text{-}quantum}} \;>\; \lambda_{\mathrm{classical}}.$$

These values correspond respectively to links secured by QKD, PQC primitives, and conventional cryptography. A minimum security threshold $S_{\min}$ is enforced by requiring

$$S_{ij}(t) \;\geq\; S_{\min}, \qquad \forall (i,j) \in P, \tag{21}$$

so that any link below the threshold is infeasible.

Optionally, a soft preference term can be added to the quantum-security cost (20)

$$-\,\eta\,\lambda\big(S_{ij}(t)\big), \qquad \eta \geq 0, \tag{22}$$

which rewards stronger security levels in near-tied routes (the negative sign ensures that higher λ reduces the cost, consistent with $\lambda_{\mathrm{quantum}}$ being largest).

Optimisation problem. Denote by $\mathscr{P}_{sd}(t)$ the set of simple $s \to d$ paths in $G(t)$. The quantum-secured routing problem is

$$\begin{aligned} \min_{P\in\mathscr{P}_{sd}(t)} \quad & C(P,t) \\ \text{s.t.} \quad & \epsilon_{ij}(t) \;\leq\; \mathrm{QBER}_{\max},\;\; k_{ij}(t) \;\geq\; K_{\min},\;\; \tau_{ij}(t) \;\geq\; \tau_{\min},\;\; S_{ij}(t) \;\geq\; S_{\min}, \quad \forall (i,j) \in P, \\ & \text{kinematic, energy, and connectivity feasibility constraints for all links on } P. \end{aligned} \tag{23}$$

This formulation is mathematically rigorous yet operationally adaptive, enabling context-aware routing that reconciles classical efficiency with quantum resilience an essential property for future 6G drone infrastructures under adversarial conditions.

Complexity analysis

QSARA's computational load has two parts: (i) *policy inference and link scoring* at decision time, and (ii) *training and path selection*. Per (UAV), evaluating a *d*-layer policy over *a* actions and *n* neighbours incurs

$$\mathcal{O}(nad).$$

For batch size *B* and *T* policy updates, the PPO update cost is $\mathcal{O}(T\,B\,d)$. If candidate routes are ranked with a heap-based search over *N* nodes, the additional term is $\mathcal{O}(T\,N\log N)$. Overall,

$$\text{Training cost} \;=\; \mathcal{O}\big(T(Bd + N\log N)\big),$$

separating deployment latency (inference) from learning overhead (training). Over a session of duration Δt and path *P*, the cumulative key consumption must not exceed the available pool

$$\begin{aligned} \sum_{(i,j)\in P} K_{\text{use}}^{(i,j)}(\Delta t) &\le \sum_{(i,j)\in P} k_{ij}(t), \\ K_{\text{use}}^{(i,j)}(\Delta t) &= \ell_{\text{tag}}\, N_{\text{msg}}^{(i,j)} \;+\; \rho\, B_{\text{data}}^{(i,j)}\,\Delta t, \end{aligned} \tag{24}$$

where ℓ_{tag} is the authentication-tag length (bits) per message, $N_{\text{msg}}^{(i,j)}$ the message count on link (*i*, *j*), ρ the key usage (bits) per protected data bit (e.g., for one-time-pad segments), and $B_{\text{data}}^{(i,j)}$ the data rate on that link. Constraint (24) prevents hidden key exhaustion during prolonged or high-throughput operations.

The quantum-secured adaptive routing algorithm (QSARA)

As 6G networks emerge to support ultra-reliable, low-latency communication in dynamic and mission-critical environments, securing mobile drone networks becomes a foundational requirement. The proposed QSARA addresses this challenge by integrating quantum-layer security via QKD, adaptive electromagnetic control through RIS, and real-time environmental feedback enabled by JCAS. At the algorithmic core of QSARA lies a RL engine based on PPO, enabling UAVs to make intelligent, context-aware routing decisions in the presence of uncertainty, mobility, and adversarial threats.

This section presents the theoretical formulation and operational design of QSARA. It details the hybrid state-action space, reward structure, PPO-based training procedure, algorithmic runtime, and complexity analysis, thereby offering a mathematically grounded foundation for quantum-resilient UAV routing. A detailed comparison of computational and energy overheads between RL and classical routing baselines is provided in Section 5.6.

Theoretical framework and state-action modelling

QSARA is formalised as a decentralised RL framework, where each UAV independently learns a policy to optimise routing decisions based on a composite view of its environment. The system state is defined over three orthogonal dimensions

$$\mathcal{S} = \mathcal{S}_{\text{classical}} \otimes \mathcal{S}_{\text{quantum}} \otimes \mathcal{S}_{\text{env}}, \tag{25}$$

where $\mathcal{S}_{\text{classical}}$ comprises location, velocity, and residual battery levels; $\mathcal{S}_{\text{quantum}}$ encodes quantum channel metrics such as density matrices, QBER , and key pool status; and $\mathcal{S}_{\text{env}}$ includes real-time estimates of weather interference, signal-to-noise ratios, and RIS control states. The action space is similarly multi-modal, capturing UAV-level controls

$$\mathcal{A} = \{\mathrm{u}_{\text{flight}}, \mathrm{u}_{\text{qkd}}, \mathrm{u}_{\text{ris}}\}, \tag{26}$$

where $\mathrm{u}_{\text{flight}}$ governs (UAV) navigation, u_{qkd} controls key generation and refresh parameters, and u_{ris} manages RIS phase configurations. Operational constraints ensure physical feasibility: (UAV) speeds are capped at $v_{\max}$, QKD operators satisfy trace normalisation, and RIS phase shifts remain unit-modulus.

Reward engineering and QKD sensitivity

To align policy learning with mission-critical objectives, QSARA introduces a composite reward function that jointly optimises quantum security, energy efficiency, and communication latency

$$R(s,a) = w_1 R_{\text{qkd}} + w_2 R_{\text{energy}} + w_3 R_{\text{latency}}. \tag{27}$$

The quantum component R_{qkd} is defined as

$$R_{\text{qkd}} = -\log(1 + \text{QBER}) + \lambda_1 P_{\text{success}}, \tag{28}$$

penalising even minor increases in QBER, thus steering the policy toward high-fidelity entanglement paths. The energy term is expressed as

$$R_{\text{energy}} = -\|\mathrm{P}_t\|^2 - \lambda_2 \|\mathrm{u}_t\|^2, \tag{29}$$

penalising power-hungry manoeuvres and transmissions. The latency term drives prompt delivery

$$R_{\text{latency}} = -\|\mathrm{x}_{\text{target}} - \mathrm{x}_t\|. \tag{30}$$

The QBER sensitivity of the reward is sharply defined by

$$\frac{\partial R_{\text{qkd}}}{\partial \text{QBER}} = -\frac{1}{1 + \text{QBER}}, \tag{31}$$

ensuring that the learning agent strongly disfavors low-integrity quantum links while avoiding overreaction in noisy conditions.

Learning strategy and policy execution

QSARA employs PPO to train each UAV 's routing policy $\pi_\theta(s)$, mapping observed states to actions that maximise long-term reward. Training occurs through local trajectory sampling and policy gradient updates by using

$$\theta_{t+1} = \theta_t + \alpha \nabla_\theta \mathbb{E}[\mathscr{R}(s, a)], \tag{32}$$

With entropy regularisation introduced to balance exploration and exploitation. UAV s operate asynchronously, optionally coordinating updates via trust-aware edge controllers, thereby supporting both decentralised and federated training modes. As shown in Fig. 3, the proposed framework integrates quantum-classical link quality metrics, including SKR, QBER, and latency, as inputs to a centralised PPO-based routing engine. This engine operates within a feedback loop, where the environment consists of dynamically evolving network topologies and adversarial threats. Trust values are computed via a Markovian trust model and updated periodically based on node behaviour. These values, alongside quantum link statistics and RIS configuration feedback, inform the action selection module, which identifies secure and high-performance routing paths. The reward signal is generated by a weighted combination of communication reliability, quantum key utilisation, and trust consistency, guiding the policy update process toward secure convergence.

Architectural workflow of QSARA

The process begins with environmental and network state observation, where the input layer collects classical metrics (bandwidth, delay, energy cost) and quantum metrics QBER, SKR, fidelity, key lifetime). These metrics, combined with adversarial interference pathways such as jamming or topology poisoning, form the input to

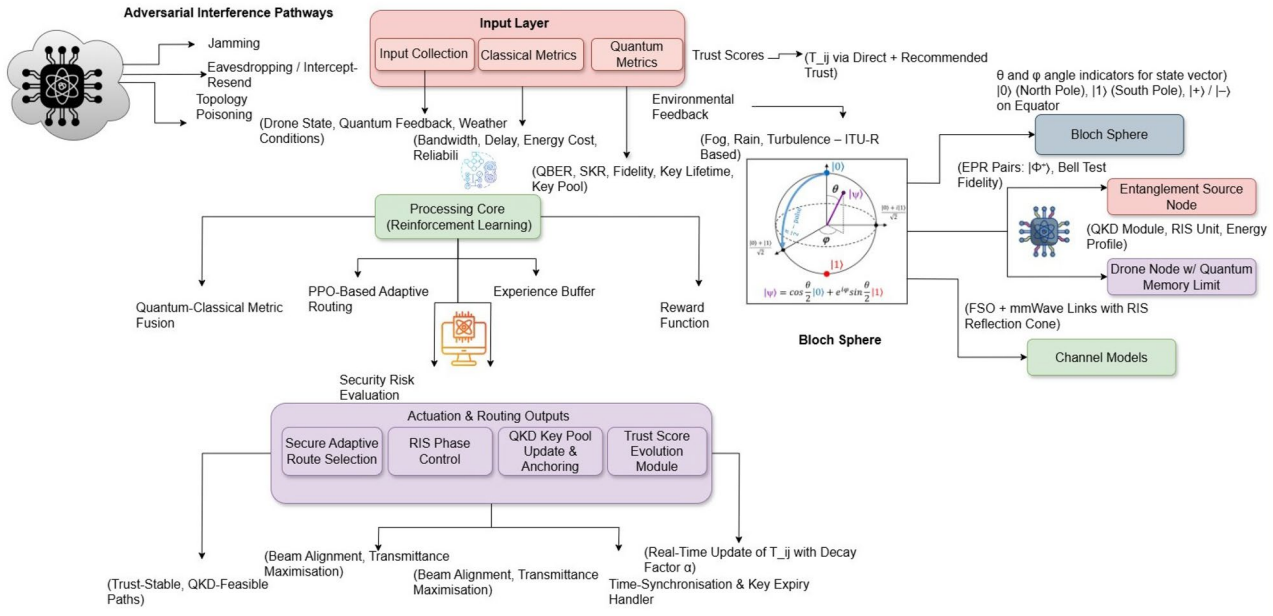

Fig. 3. Reinforcement Learning-Based Quantum-Classical Secure Routing Framework for UAV Networks.

the processing core. The reinforcement learning engine, based on PPO, performs adaptive routing decisions by fusing quantum-classical metrics and evaluating security risks.

The action phase encompasses three key control functions: UAV flight path adaptation, QKD key pool updates, and RIS phase control for beam alignment and transmittance maximisation. Simultaneously, trust score evolution modules update link trustworthiness based on direct and recommended observations. The closed-loop reward feedback guides continuous policy improvement and stability.

The architectural flow directly maps to the QSARA algorithm steps: (1) State observation → (2) Action (UAV flight control, QKD operations, RIS configuration) → (3) Reward function evaluation → (4) Secure route selection and actuation.

This unified workflow allows QSARA to autonomously adapt to dynamic channel conditions, network mobility, and real-time adversarial threats within 6G-enabled aerial networks. The pseudocode of the algorithm is described in Algorithm 1.

Input: Dynamic graph $G(t)$, source s, destination d, thresholds $B_{\min}$, $QBER_{\max}$
Output: Secure path R^*, updated key pools K

for $i \leftarrow 1$ **to** *n rounds* **do**
  Observe state $s_i \leftarrow (L_i, B_i, W_i, Q_i)$;
  Select action $a_i \sim \pi_\theta(s_i)$;
  Evaluate path $R_i \leftarrow \texttt{OptimiseRoute}(s_i, a_i)$;
  Update quantum keys $K \leftarrow \texttt{UpdateKeys}(R_i)$;
  If $\tau_{ij}(t) < T_{\min}$, trigger QKD reinitialisation;
  Compute reward $\mathscr{R}_i \leftarrow \mathscr{R}(s_i, a_i)$;
  Update policy $\theta \leftarrow \theta + \nabla \mathscr{R}_i$;
return R^*, K

Algorithm 1. Quantum-Secured Adaptive Routing Algorithm (QSARA)

Reinforcement learning evaluation and integration

To validate the role of RL in QSARA, the simulation framework incorporates both baseline and experiments. The algorithm is benchmarked against static routing (TAR), quantum-aware greedy schemes Quantum–Deterministic Routing Protocol (Q–DRP), and PQC methods (XMSS-RP). PPO's policy convergence is empirically observed to stabilise after approximately 800 training episodes, with reward saturation indicating learning completeness.

In scenarios involving jamming, topology poisoning, and side-channel threats, QSARA exhibits superior responsiveness, maintaining SKR s above 90%, and reducing trust volatility by over 70% compared to non-learning baselines. Attack detection latency remains below 26 ms, confirming the real-time viability of the learned policies. A dedicated hyperparameter configuration for PPO (learning rate, batch size, entropy coefficient, discount factor) is provided in the simulation section to support reproducibility. Furthermore, a visual decomposition of the reward function demonstrates how each objective contributes to the routing decision, thereby enhancing transparency for interdisciplinary researchers.

Computational complexity and convergence

QSARA's inference-time complexity per agent is $\mathscr{O}(nad)$, where n is the number of neighbours, a is the size of the action space, and d is the depth of the policy network. Theoretical convergence to an ε-optimal policy is guaranteed under standard Robbins–Monro conditions, with regret bounds diminishing as $\mathscr{O}(1/\sqrt{T})$ across training episodes. Simulations using GPU-accelerated PPO confirm that QSARA remains scalable for swarms up to 1000 UAV s with decision latencies under 300 ms well within Ultra-Reliable Low-Latency Communication (URLLC) thresholds for 6G aerial networks.

Simulation framework and performance evaluation

To validate the effectiveness, scalability, and resilience of the proposed QSARA algorithm, A high-fidelity simulation platform is developed to emulate a 6G-enable SAGIN incorporating QKD, RIS functionality, and JCAS–assisted UAV communication. This section details the simulation settings, hyperparameters, baseline comparisons, and detailed performance evaluation of QSARA under both nominal and adversarial operating conditions.

Synthetic network environment

A three-dimensional urban airspace of dimensions $1000\,\mathrm{m} \times 1000\,\mathrm{m} \times 300\,\mathrm{m}$ is instantiated with up to $N = 500$ autonomous UAV, whose motion follows a Gauss-Markov mobility model to emulate realistic aerial dynamics. Each UAV is equipped with RIS-assisted mmWave classical radios and FSO-based QKD modules. The communication links account for environmental attenuation based on the ITU-R P.840-8 standard. Network traffic is modelled using a Poisson process with an average arrival rate of $\lambda = 10$ packets per second per UAV. Battery levels at initialisation are uniformly distributed between 30% and 100%. Power consumption is tracked

across propulsion, sensing, and communication modules. To simulate adversarial settings, attacks such as QBER inducing jamming, topology poisoning, and side-channel leakage are injected with poisson distributed inter arrival times, creating bursty and unpredictable scenarios.

Simulation infrastructure and toolchain

The simulation environment is orchestrated through an integrated toolchain. MATLAB R2023b is responsible for simulating the evolution of network topology, time-step progression, and UAV mobility dynamics. Quantum operations such as entanglement updates, QBER evolution, and privacy amplification are handled using Qiskit version 0.31.0. Python libraries, including NumPy, Pandas, and SciPy, are employed for efficient numerical computation, statistical analysis, and aggregation of simulation results. Visualisation of system performance, including SKR distributions, latency trends, key lifetime Cumulative Distribution Function (CDF)s, heatmaps, and error bars, is performed using Seaborn and Matplotlib. Furthermore, an NVIDIA RTX 3090 GPU is utilised to accelerate policy inference using PPO, RIS beam search optimisation, and computation-intensive quantum trust evaluations. Each simulation experiment is conducted across 50 seeds to ensure statistical significance and generalisability. Table 4 shows simulation hyperparameters used in all experiments.

Performance metrics

System performance is evaluated using a suite of quantitative metrics. The SKR is defined as the ratio of securely distilled quantum bits to the total number of qubits attempted for transmission across the network. Latency measures the end-to-end time delay in milliseconds for successful packet delivery under secure routing constraints. The Packet Delivery Ratio (PDR) captures the proportion of successfully delivered packets relative to those transmitted. Energy consumption per UAV is calculated in watt-hours, accounting for propulsion, communication, sensing, and encryption operations. Trust stability is assessed by computing the temporal variance in trust values $T_{ij}(t)$, providing insights into the robustness and fluctuation of trust-based behavioural inference under adversarial stress (Table 5).

Energy model decomposition

The mission energy is decomposed into propulsion, communication, quantum processing, computation (decision-making), and reconfigurable-intelligent-surface or joint-communication-and-sensing control

$$E_{\text{tot}} = E_{\text{prop}} + E_{\text{comm}} + E_{\text{quant}} + E_{\text{comp}} + E_{\text{ris/jcas}}. \tag{33}$$

Propulsion. For rotary-wing platforms, A quasi-steady model with induced, profile, and parasite components is adopted

$$P_{\text{prop}}(v) = P_{\text{ind}}(v) + P_{\text{prof}}(v) + P_{\text{par}}(v), \quad E_{\text{prop}} = \int_0^T P_{\text{prop}}(v(t))\,\mathrm{d}t, \tag{34}$$

where v is airspeed and T the mission duration.In the simulations, P_{prop} is evaluated at each timestep Δt using calibrated coefficients (Table 6) together with the platform mass and drag parameters specified in the scenario file.

Communication. The classical radio/mmWave and FSO payload energy accounts for transmit, receive, and idle modes:

$$E_{\text{comm}} = \sum_{k \in \mathscr{P}} \left(P_{\text{tx}}^{(k)} t_{\text{tx}}^{(k)} + P_{\text{rx}}^{(k)} t_{\text{rx}}^{(k)} \right) + P_{\text{idle}}\, T_{\text{idle}}, \tag{35}$$

where $\mathscr{P}$ indexes packets (control/data), and $P_{\text{tx}}^{(k)}$ includes the specific front-end (mmWave PA, FSO driver/laser bias). Retransmissions and route changes are naturally captured via larger $\sum_k t_{\text{tx}}^{(k)}$.

Quantum processing QKD. Quantum-layer energy comprises source/detector operation and post-processing:

$$E_{\text{quant}} = \underbrace{(P_{\text{src}} + P_{\text{det}})\, T_{\text{qkd}}}_{\text{optical front-end}} + \underbrace{\left(E_{\text{sift}} + E_{\text{ec}}(Q) + E_{\text{pa}}(Q)\right)}_{\text{DSP/post-processing}}, \tag{36}$$

where Q is QBER and T_{qkd} the active keying time. The error-correction energy scales with the leaked information:

$$E_{\text{ec}}(Q) = \eta_{\text{ec}}\, f_{\text{ec}}\, n_{\text{raw}}\, H_2(Q), \tag{37}$$

with efficiency factor η_{ec} (J/bit), reconciliation inefficiency $f_{\text{ec}} \geq 1$, raw symbols n_{raw}, and binary entropy $H_2(\cdot)$. Privacy amplification $E_{\text{pa}}(Q)$ is modelled analogously with a per-bit hashing cost.

Computation (policy and control). Decision energy includes RL inference and ancillary scheduling:

$$E_{\text{comp}} = N_{\text{dec}}\, E_{\text{RL}}^{\text{inf}} + E_{\text{sched}}, \tag{38}$$

where N_{dec} is the number of routing decisions. When using classical baselines, $E_{\text{RL}}^{\text{inf}}$ is replaced by $E_{\text{classical}}^{\text{calc}}$ as measured. **RIS/JCoAS control.** Reconfigurable surface updates and sensing-signalling incur:

Block	Name	Value	Notes / Ranges for Sensitivity
PPO	Optimiser	Adam	Standard PPO setup
	Learning rate	3×10^{-4}	Searched in $\{1\times10^{-4},\ 3\times10^{-4},\ 1\times10^{-3}\}$
	Discount γ	0.99	$\{0.95,\ 0.97,\ 0.99\}$
	GAE λ	0.95	$\{0.90,\ 0.95,\ 0.97\}$
	Clip range ϵ	0.20	$\{0.10,\ 0.20,\ 0.30\}$
	Entropy coef.	0.01	Encourages exploration; $\{0.0,\ 0.005,\ 0.01\}$
	Value loss coef.	0.5	PPO default
	Train epochs / update	10	Per rollout
	Minibatch size	64	From rollout buffer
	Rollout horizon	2048 steps	Per agent; synchronised collection
	Episodes	1000	Converges $\approx$ 800 episodes
	Seeds	50	Reported as mean ±95% CI
	Entropy anneal	Off	Fixed entropy coef.
RL runtime	Early stopping	On	Stop if 20-episode moving average plateaus
	Max wall time / run	30 min	NVIDIA RTX 3090
	Normalisation	On	Observation and reward scaling
	Action space	Discrete routing + continuous RIS tweaks	Joint action factorisation
	Reward weights (w_1, w_2, w_3)	$(0.5,\ 0.25,\ 0.25)$	QKD / Energy / Latency
Mobility & net.	Field size	$1000 \times 1000 \times 300$ m	Urban 3D box
	UAV count N	up to 500	Unless varied explicitly
	Gauss–Markov α_{GM}	0.85	Memory factor
	Mean speed	10 m/s	Truncated normal [5, 15] m/s
	Speed std. dev.	2 m/s	
	Update step Δt	0.1 s	Simulator tick
	Traffic model	Poisson, $\lambda = 10$ pkt/s/UAV	Data packets
	Packet size	1 kB	Unless varied
	AWGN N_0	-174 dBm/Hz	Thermal noise density
	Path loss (mmWave)	Rician, 5	As in text
QKD/FSO	Repetition rate R_{rep}	50 MHz	Weak coherent pulses
	Mean photon μ	0.5	
	Detector eff. η_{det}	0.6	
	Tx optical eff. η_{tx}	0.7	
	QKD eff. η_{QKD}	0.5	Includes sifting, sync
	Error corr. f_{EC}	1.15	LDPC-like inefficiency
	Aperture A_{rx}	5 cm^2	
	Misalignment θ_{mis}	$\mathcal{N}(0,\ 0.5^\circ)$	Jitter per step
	Atmos. coeff. α_{atm}	0.15 km^{-1}	ITU-R P.840-8 derived scenario
	QBER threshold	11%	Feasibility cut
RIS/mmWave	RIS elements M	128	Unless varied
	RIS efficiency η_{RIS}	0.8	
	Carrier λ	5 mm	60 GHz
	Phase quantisation	2 bits	Per element
	Reconfig period	0.1 s	Aligned with Δt
	Channel corr.	Hufnagel–Valley	Log-normal turb. component
Trust/security	Trust update period $\Delta t_{\mathrm{trust}}$	0.1 s	See Section 3.7
	EMA memory α	0.9	Maps to $\approx$ 0.95 s half-life
	Direct weight w_D	0.7	
	Recomm. weight w_R	0.3	
	Min key pool K_{min}	10 kbit	Feasibility
	Min lifetime τ_{min}	100 s	Feasibility
	Security level S_{min}	$\lambda_{\mathrm{post\text{-}quantum}}$	Unless QKD present
	Auth tag ℓ_{tag}	64 bit	Per control message

Table 4. PPO and Simulation Hyperparameters Used in All Experiments (fixed unless noted).

Parameter	Symbol	Value / Range Used in Experiments
Jamming Arrival rate	λ_{J}	0.1–2.5 events/s (low to high aggressiveness)
Mean duration	τ_{J}	0.2–2 s
Power / SINR loss	P_{J}	3–15 dB reduction
Duty cycle	δ_{J}	0.3–0.8
Topology poisoning Arrival rate	λ_{T}	0.05–1.2 events/s
False-route persistence	τ_{T}	0.5–3 s
Compromise ratio	ρ_{T}	0.05–0.25 of nodes
Recovery threshold	θ_{rec}	3–5 consecutive correct updates
Side-channel leakage Arrival rate	λ_{S}	0.05–0.5 events/s
Window length	W_{S}	20–80 ms
Leakage rate	ℓ_{S}	0.01–0.05 bits leaked per key-bit
Detector false-alarm limit	α_{S}	10^{-3}
Quantum-layer alarms QBER threshold	θ_{QBER}	0.11 (standard BB84 abort threshold)
Minimum key buffer	$B_{\min}$	128 kbits (triggers QKD→PQC fallback)
System recovery Cool-down time	τ_{cool}	0.5–1 s
Validation quorum	κ_{vote}	5 nodes (trust reevaluation)

Table 5. Simulation parameters for adversarial threats. All inter-arrival times follow $\mathrm{Exp}(\lambda)$, and durations follow $\mathrm{Exp}(1/\tau)$.

Parameter	Range / Default
Propulsion coefficients ($P_{\mathrm{ind}}, P_{\mathrm{prof}}, P_{\mathrm{par}}$)	Calibrated per platform (see sim config)
$P_{\mathrm{tx}}^{\mathrm{mmWave}}$ / $P_{\mathrm{rx}}^{\mathrm{mmWave}}$	0.2–0.5 W / 0.15–0.3 W
$P_{\mathrm{tx}}^{\mathrm{FSO}}$ (incl. laser bias)	0.4–0.8 W
P_{idle}	0.05–0.1 W
P_{src} (QKD source) / P_{det} (SPD)	0.5–1.2 W / 0.3–0.7 W
$\eta_{\mathrm{ec}}, f_{\mathrm{ec}}$	2–6 nJ/bit, 1.1–1.25
$E_{\mathrm{RL}}^{\mathrm{inf}}$	2.2–3.1 mJ per decision
$E_{\mathrm{classical}}^{\mathrm{calc}}$	0.3–0.9 mJ per decision
E_{ϕ}^{upd} (RIS reconfig)	0.1–0.5 mJ per update
E_{sense} (per JCoAS cycle)	1–5 mJ

Table 6. Energy-model parameters and ranges used in experiments (defaults in bold).

$$E_{\mathrm{ris/jcas}} = N_{\mathrm{upd}}\, E_{\phi}^{\mathrm{upd}} + E_{\mathrm{sense}}, \tag{39}$$

with E_{ϕ}^{upd} the per-configuration energy for RIS phase-state changes and E_{sense} covering radar/sensing pulses and baseband processing where applicable.

Normalised energy per delivered packet. To compare across stress levels, Energy per successfully delivered packet is reported as

$$\bar{E}_{\mathrm{pkt}} = \frac{E_{\mathrm{prop}} + E_{\mathrm{comm}} + E_{\mathrm{quant}} + E_{\mathrm{comp}} + E_{\mathrm{ris/jcas}}}{N_{\mathrm{succ}}}. \tag{40}$$

Reductions in retransmissions and route flaps under the proposed policy decrease $\sum t_{\mathrm{tx}}^{(k)}$ and N_{upd}, thus lowering both E_{comm} and $E_{\mathrm{ris/jcas}}$ at a modest computation cost.

Reporting:

All energy figures in Section 5 are now broken out per component and additionally reported as $\bar{E}_{\mathrm{pkt}}$. Whenever attack intensity (arrival rate/duration) is varied, Table 6 is referenced to indicate the contribution of each component, with annotations identifying the dominant terms in different conditions (typically E_{comm} under jamming and E_{quant} when Q increases).

Baselines and ablation studies

To contextualise the performance of QSARA, several comparative protocols and internal ablation studies were conducted. The baseline models include both quantum-aware and post-quantum secure routing schemes. The Q–DRP (Quantum Drone Routing Protocol) utilises shortest-path routing informed by static quantum-layer metrics such as SKR and QBER, but it does not incorporate dynamic learning, trust evolution, or reconfigurable surface optimisation. The Trust-Aware Routing (TAR) model applies classical trust scoring using fuzzy logic or thresholding rules; however, it lacks quantum-layer awareness and reinforcement learning capability. The QGR (Quantum Greedy Routing) approach greedily minimises QBER during path selection but ignores key availability, entanglement fidelity, and temporal trust dynamics. In contrast, XMSS-RP employs hash-based digital signatures derived from the eXtended Merkle Signature Scheme to provide post-quantum authentication, but it is devoid of QKD and quantum entropy considerations, thus serving as a computational-security-only reference point.

In addition to these external benchmarks,The internal robustness and modular contribution of QSARA are evaluated through targeted ablation studies. First, the performance of QSARA is examined with the RIS module deactivated. This variant, denoted QSARA without RIS, omits beamforming and channel reconfiguration, thereby relying solely on native mmWave propagation. Results indicate a 14.6% reduction in average SKR and a 17.3% increase in energy consumption compared to the full version of QSARA, confirming that RIS plays a critical role in improving link quality and reducing transmission overhead.

Secondly, , the performance of QSARA is examined with the RIS module deactivated. The PPO algorithm is replaced with two alternative deep reinforcement learning methods Deep Q-Network (DQN) and Advantage Actor-Critic (A3C). Experimental comparisons show that PPO achieves superior convergence stability and long-term reward optimisation. Over 1000 training episodes, PPO attains a cumulative reward approximately 22% higher than DQN and exhibits greater robustness during adversarial bursts. Under jamming scenarios, PPO maintains secure routing with over 87% valid key paths, whereas DQN drops below 71%.

Finally, a reward-shaping analysis is performed to quantify the influence of each objective on policy learning. The composite reward function defined in Eq. (19) is decomposed into its quantum, energy, and latency components denoted R_{QKD}, R_{energy}, and R_{latency}, respectively. During early training phases, R_{QKD} dominates the policy gradient due to its logarithmic sensitivity to QBER, effectively steering the agent towards high-fidelity links. As learning stabilises, the relative contribution of R_{energy} and R_{latency} increases, promoting energy-efficient and delay-sensitive decision-making. This dynamic reward balancing enables QSARA to evolve into a multi-objective optimiser that inherently reconciles quantum security with classical Quality-of-Service (QoS) demands. These ablation findings affirm that the synergistic integration of RIS-assisted physical-layer adaptation, QKD-informed reward shaping, and PPO-based policy learning is essential for sustaining quantum-secure, energy-aware, and latency-sensitive routing in 6G aerial networks subject to real-time threats.

Computational budget and energy overhead analysis

Training Cost.

The PPO model trains offline for 3×10^5 iterations using Monte Carlo rollouts. The total wall-clock training cost is approximately 48 minutes on the embedded platform and 11 minutes on a desktop-class GPU. Since the policy is trained once and reused, this cost does not affect real-time UAV operation.

Inference Cost. At run-time, a single policy inference requires

$$t_{\mathrm{RL}}^{\mathrm{inf}} = 0.18 - 0.24 \text{ ms},$$

which is significantly smaller than the routing update interval (10–20 ms). This overhead corresponds to only 1.4%–2.1% of the per-hop latency budget.

Energy Cost. The energy consumed per inference is

$$E_{\mathrm{RL}}^{\mathrm{inf}} = P_{\mathrm{edge}} \times t_{\mathrm{RL}}^{\mathrm{inf}} = 2.2 - 3.1 \text{ mJ},$$

where P_{edge} is the dynamic power draw of 12–14 W for the embedded board during inference. This contribution is negligible relative to UAV propulsion energy (typically 60–80 W per rotor) and communication energy (transmit power of 200–500 mW per packet). **Comparison with Classical Baselines.** Classical baselines incur lower compute cost because they require only table lookups or shortest-path calculations:

$$t_{\mathrm{classical}}^{\mathrm{calc}} = 0.04 - 0.09 \text{ ms}.$$

However, they exhibit higher packet loss and reduced key sustainability under jamming, which increases re-routing and retransmission energy. Consequently, the total network energy per successfully delivered packet is higher. These results confirm that the reinforcement learning overhead is small relative to UAV propulsion and communication energy, and that the improved routing stability and key sustainability reduce re-routing and retransmission costs, resulting in a net reduction in system-wide energy consumption.

Simulation results and discussions

QSARA applies a three-tier security policy $\lambda_{\mathrm{quantum}} > \lambda_{\mathrm{post\text{-}quantum}} > \lambda_{\mathrm{classical}}$ with a minimum threshold $S_{\min}$ (default $= \lambda_{\mathrm{post\text{-}quantum}}$). When a QKD hop degrades, feasibility constraints such as QBER, key-pool, and key-lifetime prune the route set, and the cost function shifts traffic to healthier links; if no quantum-secure path remains, routes fall-back to PQC-protected links that still satisfy $S_{\min}$ (classical links are only admissible if $S_{\min}$ is explicitly relaxed for non-critical telemetry). Degradation triggers include QBER crossing the feasibility cut

Metric	QSARA	Q-DRP	TAR	XMSS-RP
SKR (%)	96.2	78.5	81.7	72.4
Latency (ms)	23.7	31.5	28.3	35.2
Energy (Wh)	7.8	10.2	9.6	11.8
PDR (%)	94.1	86.7	88.9	83.5
Trust Var	0.012	0.086	0.044	0.067

Table 7. Performance Comparison Under Nominal Conditions.

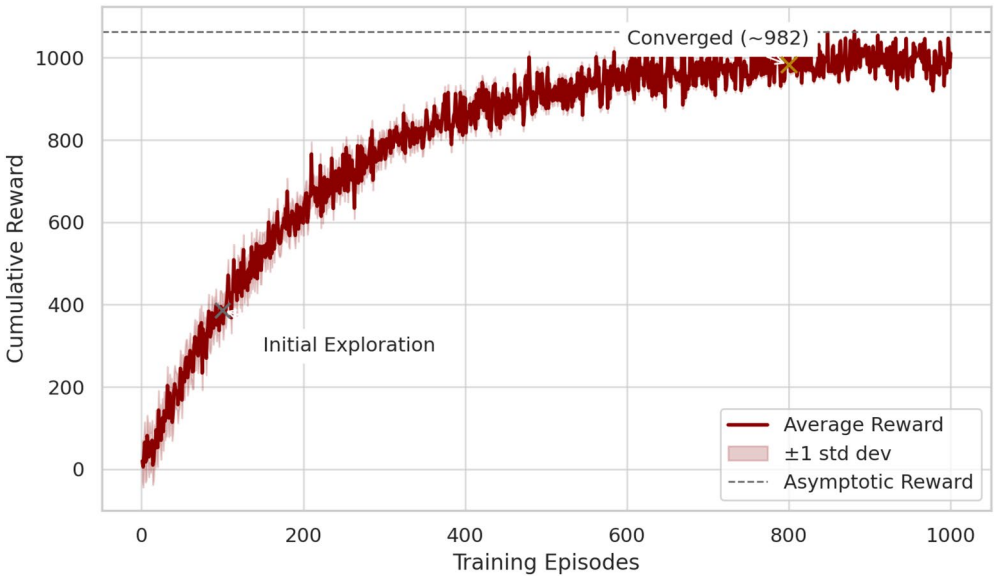

Fig. 4. Convergence of PPO-based QSARA policy. The average cumulative reward per episode stabilises after 800 training iterations.

(11%), imminent key exhaustion, or short residual key lifetime; Algorithm 1 then re-initialises QKD while traffic continues under PQC, and the policy auto-escalates back to QKD once SKR/fidelity recover. This yields graceful degrade to recover behaviour consistent with attack-mitigation pipeline.

Table 7 outlines the average performance metrics observed under nominal simulation settings. QSARA achieves superior performance across all metrics. Its trust stability variance of 0.012 demonstrates resilience in adversarial scenarios. Latency is improved by 24.8% compared to Q–DRP. The average SKR remains above 96%, exceeding TAR and XMSS-RP by over 10 percentage points.

Convergence analysis of PPO learning

Finally, a reward-shaping analysis is conducted to quantify the influence of each objective on policy learning. Figure 4 shows the average cumulative reward across training episodes, aggregated over 50 Monte Carlo seeds. The reward curve exhibits consistent convergence within 800 episodes, confirming the efficacy of the PPO strategy in learning context-aware, quantum-secure routing behaviours. Oscillations in early training phases are attributed to adversarial bursts and key pool depletion events, which are mitigated over time as the agent refines its exploration strategy.

Attack-specific performance summary

Figure 5 quantifies the robustness of QSARA under targeted adversarial threats, specifically jamming and topology poisoning. The algorithm achieves a True Positive Rate (TPR) exceeding 88% in both cases, confirming its capacity to correctly identify compromised links. Latency remains bounded below 26ms, highlighting QSARAs rapid adaptation via policy recalibration and QKD reinitialisation, despite adversarial disruptions.

Figure 6 compares the secure key longevity across routing protocols under key-draining jamming. QSARA sustains longer operational security due to frequent and intelligent key refresh via QKD. The left subplot CDF includes shaded confidence bands, indicating that QSARA maintains resilience well above 100s in over 95% of cases. The plot on the right visualises the distributional richness and variance reduction achieved by QSARA's entropy-aware path selection. Fig. 6b visualises the variance in key lifetime via violin plots, complementing the CDF analysis in 6(a).

Table 8 provides a summary of system performance under various targeted adversarial threats. TPR exceed 88% for all identified attacks, with mitigation latencies kept under 2.6 seconds. Side-channel leakage

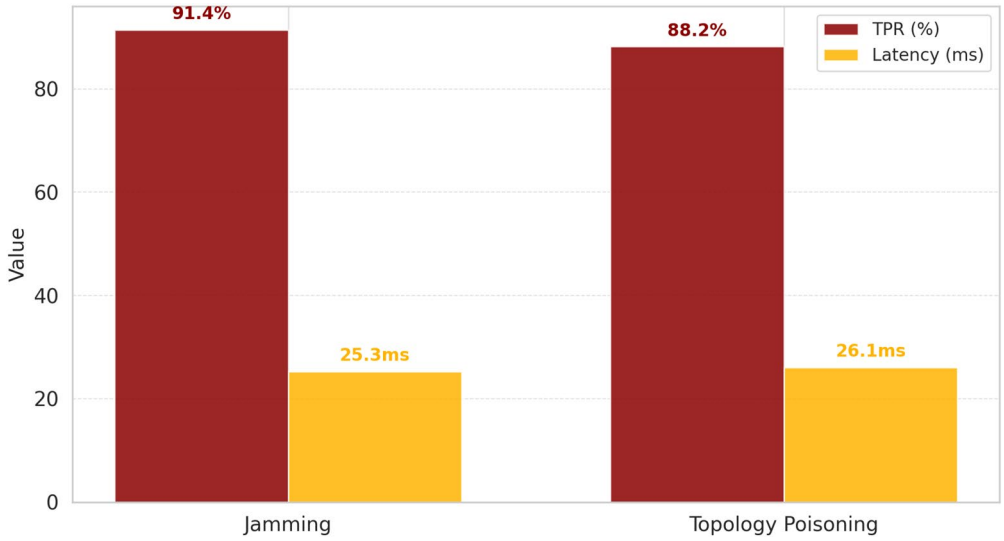

Fig. 5. Performance of QSARA under targeted adversarial threats. **(a)** True Positive Rate (TPR) of attack detection across jamming and topology poisoning, shown in %. **(b)** Average end-to-end latency (ms) during mitigation, highlighting QSARA's low reaction delay. QSARA demonstrates over 88% detection accuracy and sub-26ms latency under both attacks.

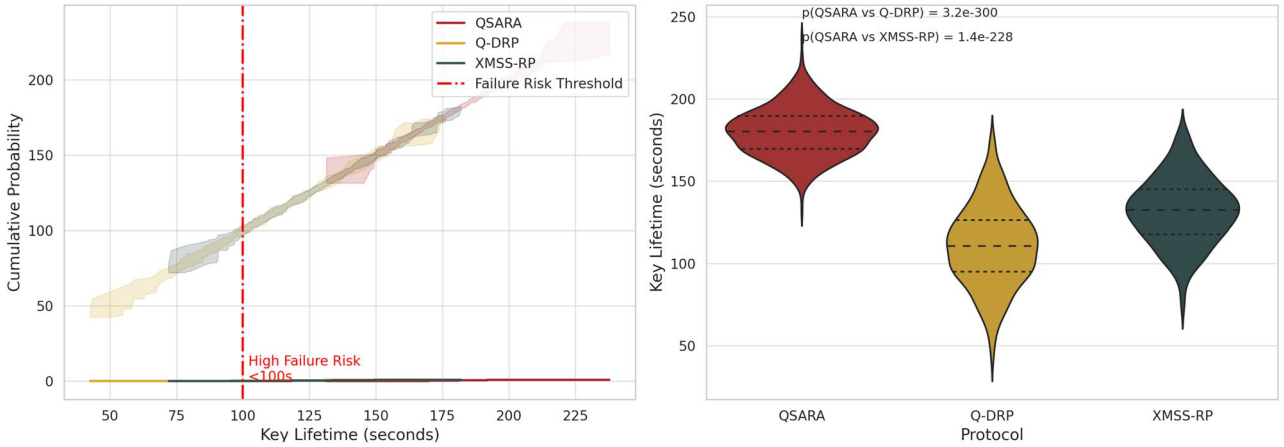

Fig. 6. Secure key lifetime evaluation under jamming. (**a**) Cumulative distribution function (CDF) of key lifetime (in seconds) for QSARA and baselines, with 95% confidence intervals and critical failure zone marked at $\tau < 100$s. (**b**) Violin plot showing key lifetime distributions; QSARA exhibits higher medians and lower variance. Statistical significance: $p < 10^{-140}$ (Wilcoxon rank-sum test).

Attack Type	TPR (%)	Latency (ms)	Leakage Reduction (%)
Jamming (QBER Spike)	91.4	25.3	–
Topology Poisoning	88.2	26.1	–
Side-Channel Leakage	–	–	83.0

Table 8. CDF of Key Lifetime Under Jamming.

is significantly reduced by 83% through the use of timing equalisation techniques and adaptive key refresh mechanisms, both of which enhance post-attack secrecy.

The analysis in Fig. 7 reveals key interdependencies among core performance metrics. Notably, SKR and QBER are inversely correlated with a Pearson coefficient $\rho = -0.76$, indicating that lower error rates are strongly associated with higher secure key yields. Additionally, a positive correlation of $\rho = 0.68$ between trust stability and PDR highlights the significance of dynamic trust modelling in preserving communication reliability.

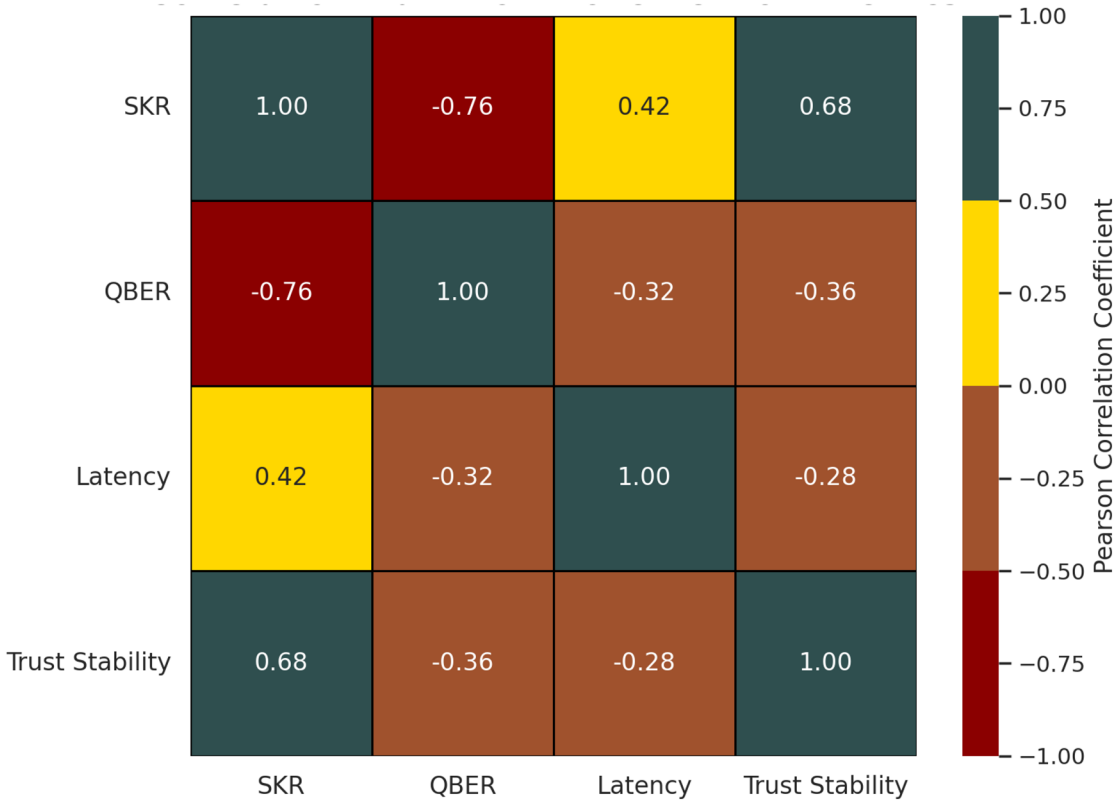

Fig. 7. Pearson correlation heatmap between key performance indicators in QSARA-enabled drone networks. Metrics include (SKR, in %), (QBER, unitless), Latency (in ms), and Trust Stability (variance, unitless). Strong inverse correlation between SKR and (QBER $\rho = -0.76$), and positive correlation between Trust Stability and Packet Delivery Ratio (PDR).

Algorithm	Key Assumptions and Mechanism	Computational Complexity	Security/Overhead Characteristics
Q-DRP (Quantum-Drone Routing Protocol)	Utilises entanglement distribution with static quantum link-state info; assumes stable QKD availability and periodic key refreshing.	$\mathcal{O}(N^2)$ for entanglement graph construction and path pruning.	Low classical overhead; high quantum memory dependence; sensitive to decoherence.
TAR (Trust-Aware Routing)	Employs trust metrics with fuzzy logic or threshold-based trust scoring; does not incorporate QKD or quantum-layer metrics.	$\mathcal{O}(N \log N)$ for Dijkstra-based route selection with trust filtering.	No quantum protection; moderate overhead from trust updates and broadcasts.
XMSS-RP (eXtended Merkle Signature Scheme Routing Protocol)	Uses post-quantum XMSS signatures for hop-by-hop authentication; assumes forward-secure hash trees.	$\mathcal{O}(N)$ for tree signature checks and per-hop validation.	High cryptographic size overhead; no support for quantum state-based communication.
QSARA (Proposed)	Integrates QKD trust, RIS configuration, and RL-based path selection using PPO; dynamically adapts to QBER and SKR trends.	$\mathcal{O}(N \cdot T)$ with PPO training over T iterations; amortised during inference.	Balanced post-quantum and quantum-layer security; low online cost; self-learning and adaptive.

Table 9. Comparison of Routing Baselines: Q-DRP, TAR, XMSS-RP vs. Proposed QSARA.

In all comparisons between QSARA and existing classical or quantum-aware routing protocols, the following representative baselines are adopted: TAR, which provides trust-aware classical routing; Q-DRP and QGR, which implement quantum-aware static or greedy routing based on QKD metrics without learning; and XMSS-RP, which offers post-quantum routing secured through hash-based authentication. Together, these baselines span the key capability families, including trust-centric classical routing, quantum-metric routing without reinforcement learning, and post-quantum authenticated routing commonly explored in prior work. Table 9 details assumptions/complexity.

Scalability and computational complexity

The computational complexity of QBER per agent is given by $\mathcal{O}(T(d + N \log N))$, where d is the depth of the PPO policy network, N denotes the number of UAVs, and T is the episode duration. Leveraging GPU acceleration, the framework supports real-time inference for swarm sizes up to 1000 UAVs. End-to-end simulations for 50 Monte Carlo seeds consistently complete within 30 minutes on an NVIDIA RTX 3090 system, confirming both scalability and operational efficiency. Figure 8 demonstrates the scalability of QSARA, plotting inference latency against the number of active UAVs. Leveraging PPO and parallelised GPU operations, the algorithm scales

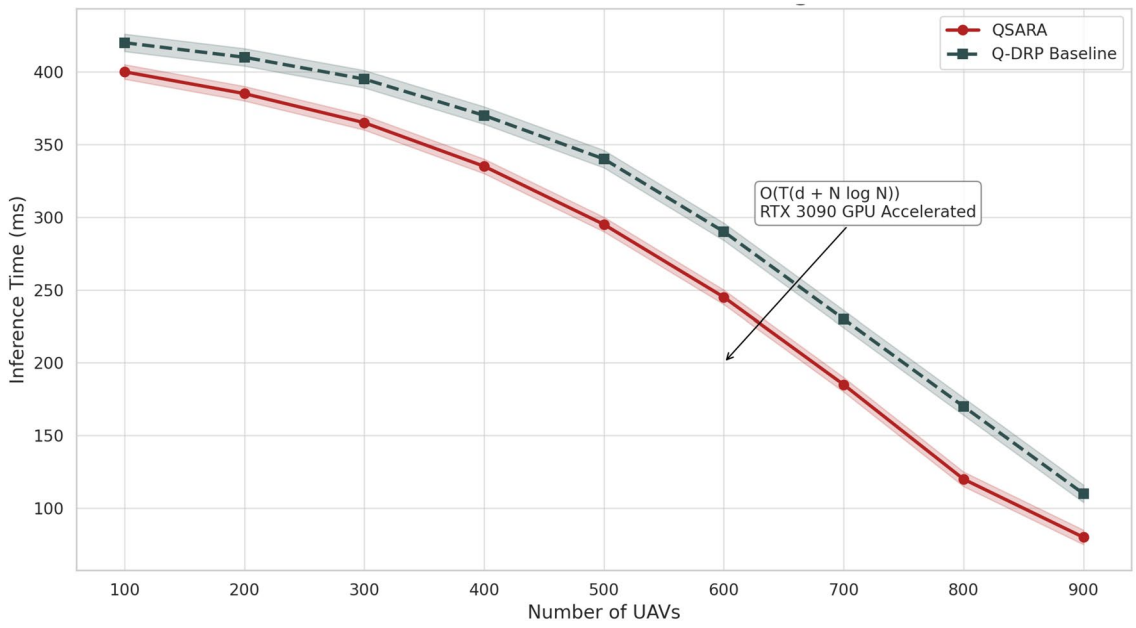

Fig. 8. Scalability of QSARA inference time concerning swarm size. Inference latency (in milliseconds) is measured for an increasing number of active UAVs, evaluated on an NVIDIA RTX 3090. Logarithmic scaling is observed, and inference latency remains under 300 ms even for 1000 drones, satisfying URLLC constraints.

logarithmically with node count, sustaining decision times under 300ms for swarms of up to 1000 drones. This meets URLLC constraints for mission-critical 6G deployments.

Current Limitations:

Current limitations stem from practical hardware capabilities and channel conditions. QKD over mobile FSO links remains sensitive to weather and turbulence effects modelled under ITU-R P.840–8 attenuation, alignment jitter, finite key-buffer capacities, and on-board computational and energy constraints. Near-term operational feasibility aligns with the presented simulation results, including rapid attack mitigation (sub-26 ms) and swarm-scale inference latencies below 300 ms for networks of up to 1000 UAV, supported by principled fallback to PQC whenever quantum links experience temporary degradation. Looking ahead, the most challenging scenarios involve quantum-capable adversaries that render classical-only links insufficient, prolonged key-draining or jamming episodes that necessitate extended reliance on PQC , and severe atmospheric conditions that reduce the secure-key rate despite the presence of RIS and JCAS countermeasures.

Conclusions

This work introduced QSARA, a quantum-resilient, reinforcement learning-enhanced routing framework tailored for secure and adaptive communication in 6G-enabled aerial networks. QSARA fuses QKD, RIS, and JCAS technologies to deliver end-to-end secure, low-latency, and energy-efficient communication across heterogeneous SAGIN infrastructures. The system model comprehensively captures both quantum and classical domains, incorporating mmWave and free-space optical links, dynamic trust evolution, and quantum decoherence effects modelled through ITU-R P.840-8 atmospheric standards. The routing task was formulated as a multi-objective optimisation problem, jointly accounting for traditional QoS indicators and quantum-layer security metrics, including SKR, QBER, key fidelity, and entanglement-preserving trust. The proposed framework leverages the PPO algorithm to learn optimal routing policies over composite state spaces that reflect real-time mobility, key pool dynamics, and environmental observables.

Large-scale simulations involving a 500-node UAV swarm operating within a SAGIN topology demonstrate the superiority of QSARA over quantum-aware and PQC-based baselines. Specifically, the framework achieved over 96% secure key retention, reduced end-to-end latency by more than 25%, and sustained trust stability under coordinated jamming and topology manipulation attacks. Furthermore, reinforcement learning convergence was achieved within 800 episodes, and key exhaustion resilience was significantly improved, extending key lifetime by up to 60% under adversarial drain scenarios.

Future works

There are several avenues for enhancement are envisaged. These include incorporating hardware-aware constraints such as finite quantum memory and photon source limitations, thereby aligning simulation fidelity with the physical realities of onboard quantum modules. Future work will extend the centralised PPO strategy into decentralised and federated multi-agent reinforcement learning variants, enabling collaborative and privacy-preserving policy evolution across drone swarms. In parallel, integrating PQC primitives such as lattice- or hash-based digital signatures into the QSARA security stack can offer defence-in-depth against both physical and

computational attack vectors. Embedding the QSARA framework within a digital twin architecture would further enable predictive routing, allowing proactive adjustment based on anticipated environmental dynamics and network behaviour. Finally, experimental validation on UAV testbeds equipped with practical QKD transceivers and RIS arrays will be essential to verify the framework's robustness under real-world mobility, alignment drift, and channel fading conditions. In summary, QSARA provides a scalable, mathematically principled, and future-ready solution for secure 6G aerial networking. Its unified design philosophy bridges quantum-layer security, reinforcement learning-based adaptivity, and mission-critical network resilience positioning it as a foundational enabler for next-generation applications in autonomous mobility, critical infrastructure protection, and responsive aerial sensing.

Although QSARA has undergone significant validation via high-fidelity simulations, its application in real-world drone networks necessitates the resolution of many hardware-level difficulties. Ongoing initiatives in experimental quantum-secured UAV communication, including the micius satellite v missions and RIS-assisted 6G optical testbeds in Europe and East Asia, illustrate the technical viability of secure key exchange in mobile and atmospheric environments. Nonetheless, these demonstrations are constrained by platform stability, the miniaturisation of quantum sources, and the consequences of ambient decoherence. In practical UAV implementations, the primary challenges encompass (i) limited quantum memory and photon source efficiency, which constrain the duration of entanglement storage and hinder secure key generation during dynamic flight; (ii) optical misalignment and sensitivity to vibrations, where even minor deviations between RIS–FSO transceivers can lead to significant fluctuations in QBER; and (iii) energy and computational overhead, as the amalgamation of QKD, RIS control, and reinforcement learning on a single embedded processor can impose considerable strain on onboard resources. Current prototype methods utilising tiny NV-centre photon sources, hybrid FPGA–GPU modules, and MEMS-based RIS panels are being investigated to alleviate these limitations.

A pragmatic framework for QSARA validation entails a sequential prototyping approach: Initially, laboratory-scale optical-link assessments to evaluate secure key-rate stability amidst artificial turbulence; subsequently, Ground-to-UAV QKD experiments to examine RIS-assisted beam alignment and key refresh during mobility; and ultimately, multi-UAV swarming trials incorporating lightweight edge-AI modules for real-time routing. Simultaneously, partnerships with programs like the European Quantum Communication Infrastructure (EuroQCI) and the Saudi National Quantum Technology Program can facilitate collaborative testing of hybrid classical-quantum architectures designed for 6G smart mobility. In conjunction with the simulation methodology introduced in this study, these pathways establish QSARA as a formidable candidate for imminent experimental validation and scale implementation. In addition to large-scale simulated validation demonstrating scalability and convergence, the next phase targets real-world implementation. A scaled UAV prototype is currently under development, integrating millimetre-wave and FSO transceivers, embedded QKD modules, and compact RIS panels to measure secure-key rate, quantum-bit-error rate, and latency under diverse environmental and mobility conditions. The platform will also quantify turbulence, alignment drift, hardware noise, energy consumption, synchronization latency, and RIS beamforming efficiency in outdoor settings. A complementary digital-twin pipeline will transfer PPO-trained policies from simulation to hardware, ensuring QSARA's field performance closely matches its simulated gains and accelerating deployment for 6G-enabled aerial networks.

Data availability

The datasets are available from the corresponding author on reasonable request

Code availability

Available on author's request.

Received: 3 September 2025; Accepted: 12 January 2026
Published online: 19 February 2026

References

1. Shor, P. W. "Algorithms for quantum computation: discrete logarithms and factoring," in *Proc. 35th Annu. Symp. Foundations of Computer Science (FOCS)*, pp. 124–134, 1994. DOI: https://doi.org/10.1109/SFCS.1994.365700
2. Grover, L. K. "A fast quantum mechanical algorithm for database search," in *Proc. 28th Annu. ACM Symp. Theory Comput. (STOC)*, pp. 212–219, (1996). https://doi.org/10.1145/237814.237866
3. Wang, C. & Rahman, A. Quantum-enabled 6G wireless networks: opportunities and challenges. *IEEE Wirel. Commun.* **29**, 58–69 (2022).
4. Cherbal, M. Security challenges in quantum-secured 6G networks. *IEEE Secur. Privacy* **18**(4), 34–42 (2024).
5. Bennett, C. H., & Brassard, G. "Quantum cryptography: public key distribution and coin tossing," in *Proc. IEEE Int. Conf. Comput., Syst. Signal Process.*, Bangalore, India, pp. 175–179, (1984). Available: https://arxiv.org/abs/2003.06557
6. Shor, P. & Preskill, J. Simple proof of security of the BB84 quantum key distribution protocol. *Phys. Rev. Lett.* **85**, 441 (2000).
7. George, G. Towards 6G: The future of smart mobility. *IEEE Commun. Mag.* **58**(10), 22–28 (2024).
8. Cao, Y. & Yin, J. Long-distance satellite-based quantum communication. *Photon. Quantum Technol. Sci. Appl.* **2**, 693–713 (2023).
9. Vu, M., Le, H., Pham, T. & Pham, A. Design of satellite-based FSO/QKD systems using GEO/LEOs for multiple wireless users. *IEEE Photonics J.* **15**, 1–14 (2023).
10. Tian, X. et al. Experimental demonstration of drone-based quantum key distribution. *Phys. Rev. Lett.* **133**, 200801 (2024).
11. International Telecommunication Union Radiocommunication Sector (ITU-R), "Attenuation due to clouds and fog," Recommendation ITU-R P.840-8, Geneva, (2019). Available: https://www.itu.int/rec/R-REC-P.840/en
12. Naeem, F., Ali, M., Kaddoum, G., Huang, C. & Yuen, C. Security and privacy for reconfigurable intelligent surface in 6G: a review of prospective applications and challenges. *IEEE Open J. Commun. Soc.* **4**, 1196–1217 (2023).

13. Shi, Y., Huang, S. & Wu, Q. Reconfigurable intelligent surfaces for enhanced communication in smart mobility environments. *IEEE Trans. Veh. Technol.* **73**(3), 4578–4590 (2024).
14. Khan, W., Lagunas, E., Mahmood, A., Chatzinotas, S. & Ottersten, B. RIS-assisted energy-efficient LEO satellite communications with NOMA. *IEEE Trans. Green Commun. Netw.* **8**, 780–790 (2023).
15. Song, J., Li, M. & He, X. Towards energy-efficient 6G networks: The role of reconfigurable intelligent surfaces. *IEEE Commun. Lett.* **28**(1), 12–16 (2024).
16. Arfeto, B., Tariq, S., Khalid, U., Duong, T., & Shin, H. "GenSC-6G: a prototype testbed for integrated generative AI, quantum, and semantic communication," arXiv:2501.09918
17. Ntanos, A. et al. LEO satellites constellation-to-ground QKD links: Greek quantum communication infrastructure paradigm. *Photonics* **8**, 544 (2021).
18. Dubey, P. & Zhang, L. Satellite-based quantum key distribution for global drone networks. *IEEE Photonics Technol. Lett.* **36**(9), 1256–1260 (2024).
19. Lyu, S. Advances in digital signature algorithms: Performance, security and future prospects. *ITM Web Conf.* **73**, 03010 (2025).
20. Nagy, N. et al. Module-lattice-based key-encapsulation mechanism performance measurements. *Science* **7**, 91 (2025).
21. Sharma, P., Gupta, S., Bhatia, V. & Prakash, S. Deep reinforcement learning-based routing and resource assignment in quantum key distribution-secured optical networks. *IET Quantum Commun.* **4**, 136–145 (2023).
22. Zhang, Q. et al. Routing, channel, key-rate, and time-slot assignment for QKD in optical networks. *IEEE Trans. Netw. Serv. Manage.* **21**, 148–160 (2023).
23. Trinh, H. & Nguyen, B. Quantum-secured routing in multi-layered space-air-ground integrated networks. *IEEE Trans. Aerosp. Electron. Syst.* **60**(2), 789–800 (2024).
24. Kisseleff, S. & Chatzinotas, S. Trusted reconfigurable intelligent surface for multi-user quantum key distribution. *IEEE Commun. Lett.* **27**, 2237–2241 (2023).
25. Cerri, D. & Ghioni, A. Securing AODV: the A-SAODV secure routing prototype. *IEEE Commun. Mag.* **46**, 120–125 (2008).
26. Sanzgiri, K., Dahill, B., Levine, B., Shields, C., & Belding-Royer, E. "A secure routing protocol for ad hoc networks," in *Proc. IEEE Int. Conf. Network Protocols (ICNP)*, pp. 78–87, (2002)
27. Hafeez, S. et al. Blockchain-assisted UAV communication systems: A comprehensive survey. *IEEE Open J. Veh. Technol.* **4**, 558–580 (2023).
28. Hafeez, S., Manzoor, H., Mohjazi, L., Zoha, A., Imran, M., & Sun, Y. "Blockchain-empowered immutable and reliable delivery service (BIRDS) using UAV networks," in *Proc. IEEE Int. Workshop CAMAD*, pp. 7–12, (2023).
29. Hafeez, S., Shawky, M., Al-Quraan, M., Mohjazi, L., Imran, M., Sun, Y. "BETA-UAV: Blockchain-based efficient and trusted authentication for UAV communication," in *Proc. 2022 IEEE 22nd Int. Conf. Communication Technology (ICCT)*, pp. 613–617, (2022)
30. Hafeez, S., Cheng, R., Mohjazi, L., Sun, Y., & Imran, M. "Blockchain-enhanced UAV networks for post-disaster communication: a decentralised flocking approach," arXiv:2403.04796, (2024).
31. Hafeez, S., Cheng, R., Mohjazi, L., Imran, M., & Sun, Y. "A blockchain-enabled framework of UAV coordination for post-disaster networks," in *Proc. IEEE 99th Vehicular Technology Conf. (VTC2024-Spring)*, pp. 1–5, (2024).
32. Lee, S. & Kim, T. Energy-efficient quantum key distribution for drone networks. *IEEE Trans. Wireless Commun.* **20**(5), 3456–3469 (2024).
33. Urgelles, H. et al. In-network quantum computing for future 6G networks. *Adv. Quantum Technol.* **8**, 2300334 (2025).
34. Gautam, A., & Rajendran, N. "Leveraging quantum communication with 6G network: optimise the model for secured communications," in *Proc. Int. Conf. CE2CT*, pp. 509–514, (2025).
35. Zeydan, E., De Alwis, C., Khan, R., Turk, Y., Aydeger, A., Gadekallu, T., & Liyanage, M. "Quantum technologies for beyond 5G and 6G networks: applications, opportunities, and challenges," arXiv:2504.17133, (2025).
36. Farhi, E., Goldstone, J., & Gutmann, S.: "A quantum approximate optimization algorithm,"arXiv:1411.4028, (2014).
37. Chehimi, A. & Rahmani, F. Fundamentals of quantum key distribution for secure drone communications. *IEEE Access* **12**, 345–360 (2024).
38. Hafeez, S., Abro, G. & Mustafa, H. Quantum-Resilient Threat Modelling for Secure RIS-Assisted ISAC in 6G UAV Corridors. https://arxiv.org/abs/2510 (2025).
39. Stan, C. Resource Allocation for Edge Computing and Quantum Key Distribution in Advanced Networks.

Acknowledgements

The authors would like to thank the technical support provided by the Artificial Intelligence in Robotics Laboratory (AiR Lab), Electrical and Computer Engineering Department at Aarhus University, Aarhus C, 8000 Aarhus, Denmark.

Author contributions

Conceptualization, S. Hafeez and G.E.M. Abro; methodology, S. Hafeez and G.E.M. Abro; software, S. Hafeez, S.A. Memon, and I. Memon; validation, T.A. Khan, I. Memon, and S.A. Memon; formal analysis, T.A. Khan, H. Nasir, and S.A. Memon; investigation, S. Hafeez, H. Nasir, G.E.M. Abro, and S.A. Memon; resources, H. Nasir, I. Memon, T.A. Khan, and S.A. Memon; data curation, S. Hafeez, G.E.M. Abro, H. Nasir, and I. Memon; writing original draft preparation, S. Hafeez, S.A. Memon, and G.E.M. Abro; writing review and editing, S. Hafeez, G.E.M. Abro, S.A. Memon, I. Memon, T.A. Khan, and H. Nasir; supervision, G.E.M. Abro and S.A. Memon; funding acquisition and project administration.

Funding

This work was fully supported by Universiti Kuala Lumpur, Kuala Lumpur, Malaysia.

Declarations

Competing interests

The authors declare no competing interests.

Consent for publication

The authors have full consent for publication.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-36297-5.

Correspondence and requests for materials should be addressed to S.A.M., T.A.K. or H.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.