ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Design of AI-Based Key Distribution and Authentication Protocols Resistant to Quantum Attacks for Next-Generation Cryptosystems

Bing Han[1,*] *and Jinze Du*[1]

[1] School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, Gansu, China

SUMMARY: *In order to solve problems including quantum threats, insufficient detection of abnormal visit, and bad adaptability of fixed strategies which traditional authentication and key agreement mechanisms encounter in open visit environments, this paper puts forward a post-quantum artificial intelligence-based key distribution and authentication agreement which is designed for new cryptographic schemes. According to ML-KEM, ML-DSA, or SLH-DSA, this protocol runs inside a cooperative structure that includes end points, gateway devices, auth servers, and one key management centre. It brings in risk-conscious mechanisms, authentication level promotion, and dynamic key re-production to realize pseudonym protection, two-way authentication, and safe session key building. Through the formal verification work, the prototype realization, and the repeated experiments on the expanded situations, the protocol's security property, authentication effect, and AI-based risk controlling abilities have been comprehensively appraised. The outcomes show that the put-forward plan reaches a high interception proportion against common attack situations like replay, man-in-the-middle, spoofing, equipment cloning, and token stealing. The artificial intelligence danger calculation apparatus obtains an average AUC of 0.9720 and an F1 value of 0.9254, hence effectively raising the capacity to recognize abnormal visit behaviors. When put beside the ECC baseline scheme and the fixed-policy post-quantum scheme, this method obtains higher security gains and environment adaptive ability, although it brings restricted extra time lags, communication expenses and calculation cost.*

KEYWORDS: *Post-quantum cryptography; Authentication; Key distribution; Risk-aware; Dynamic key renewal*

1 Introduction

Along with the non-stop enlargement of cloud-edge-device cooperative access, mobile equipment inter-connection, and Internet of Things services, identity verification and session key agreement have become foundational parts for guaranteeing the trusted access of network entities and the safe transmission of data. The great part of current access systems have as their foundation RSA, ECC, Diffie-Hellman, and the modified forms that come from them. Although these mechanisms have for a long time kept good safety and engineering usability in traditional calculation environments, their safety depends on the calculation complexity of difficult problems like whole number factorization and discrete logarithm. After quantum calculation ability attains an available magnitude, Shor's arithmetic will greatly weaken or even destroy the safe foundations of conventional public-key cryptology systems, therefore making current

*hb651114158@163.com
https://doi.org/10.65102/is2026750
Published: 30 April, 2026

authentication and key exchange frameworks face never-before-seen pressure for replacement. In the mean while, the IoT application scenes have the features of different kinds of end points, open connection lines, frequent visit actions and big environment changes, therefore they bring an attack area that is far bigger than what the traditional close networks have. This therefore causes problems like identity pretend, repeat attacks, middle-man disturbance, and session seizing to become even more notable. Current literature summaries show that quantum-safe identity verification and key interaction have already become an important research focus in cryptographic protocol study in recent years, and protocol design for IoT applications is moving from only pursuing light-weight solutions to an overall balance of safety, effectiveness, and environment compatibility [1, 2].

As the reply to this change, the post-quantum cryptography standardization work has given clearer technical limits and workable realization roads for protocol design. In 2024, NIST successively released FIPS 203,FIPS 204, and FIPS 205, respectively establishing the standard status of the Modular Key Encapsulation Mechanism (ML-KEM), the Modular Digital Signature Algorithm (ML-DSA), and the Stateful Hash Signature Algorithm (SLH-DSA). These standards supply authoritative technique bases for key building and identity verification in post-quantum circumstances [3-5]. Among these, ML-KEM provides a unified interface for symmetric key establishment; ML-DSA is suitable for digital signature and identity binding tasks for business entities; and SLH-DSA can serve as a supplementary signature path in scenarios requiring high long-term robustness. In its transition proposals, NIST has further pointed out that the construction of future cipher systems must not be confined to the substitution of single algorithms, but must at the same time take into account the speed of system transfer, interface matching, cipher flexibility, and multi-stage arrangement schemes [6]. This implies that the focus of post-quantum protocol research has gradually shifted from "which new algorithm to adopt" to "how to integrate new cryptographic primitives into real-world access workflows while maintaining acceptable authentication latency, communication overhead, and risk control capabilities in complex environments."

With regard to post-quantum identity authentication and key exchange, currently existing research has already put forward several representative schemes for mobile terminal and multi-party interaction situation. Rewal et al. proposed a three-lattice-based authentication key exchange protocol for mobile devices, with targeted enhancements in anonymity, forward secrecy, and session key protection [7]; Pursharthi and Mishra afterwards kept on putting effort into mobile situations, ameliorating post-quantum authentication and key agreement mechanisms for promoting the security characteristics and applicability of the protocols [8]; The authors Chaudhary and others further put forward a three-party quantum-safe authentication and key agreement structure that balances anonymity, hence making the protocol logic for multi-body key sharing more overall comprehensive [9]; Braeken, from a transitional deployment perspective, proposed a bidirectional multi-factor hybrid framework based on ECC and KEM, offering an engineering-feasible approach for migrating traditional systems to post-quantum architectures [10]. These efforts demonstrate that post-quantum authentication protocols have gradually evolved from theoretical constructs toward practical deployment. However, existing solutions still predominantly rely on static authentication processes, meaning that once registration parameters and authentication paths are determined, the protocol execution rules typically remain unchanged.Such designs often lack the ability to adjust risk levels in real time when faced with drastic fluctuations in access contexts, abnormal terminal behavior, or increased attacker spoofing capabilities. In the same period, the enlargement of message dimensions, calculation burden, and terminal resource use which comes from post-quantum cryptography mean that the protocol's adaptive ability in complex business surroundings still needs further checking and confirmation.

From the perspective of industrial application, medical Internet of Things, cloud-supported access, and multi-device collaboration scenarios have become key research sites for post-quantum identity authentication research. The authors including Adeli and other persons have proposed an authentication plan which satisfies post-quantum demands for Internet of Things medical systems, and hence therefore stressed the importance of maintaining light-weight safety mechanisms in sensitive working environments [11]; The PQCAIE plan put forward by Mansoor and other persons brings post-quantum identity verification into electronic medical treatment scenes, therefore further showing the application possibility of quantum-safe entrance in medical care services [12]; Bahache and other authors expanded their investigation to cloud-based medical access frameworks, having discussion about the cooperative realization of quantum-resistant identity verification and key exchanging in cloud-aided environments [13]; Ahmad and Jagatheswari advanced protocol design for Medical IoT scenarios from the perspectives of three-lattice-based authentication key exchange and cloud-assisted multi-factor user authentication, respectively [14, 15]. These investigations offer precious references for post-quantum access in environments with resource constraints, but they also expose a widespread problem: most schemes still regard "whether the entity's identity is valid" as the core authentication standard, while they give not enough attention to the fluctuation of the access behavior itself. For terminal access in open networks, relying solely on static credentials, fixed challenge-response mechanisms, and predefined key renewal cycles makes it difficult to adequately address more dynamic threats such as device cloning, credential abuse, spoofed logins, and access accompanied by anomalous traffic.

Upon the contrary, the utilization of machine learning and deep learning technologies within identity verification, entrance authority management, and abnormal situation discovery is continuously getting deeper. Research by Saleem et al. indicates that machine learning-enhanced attribute-based authentication mechanisms can improve decision resilience in IoT access control, allowing authentication policies to transcend fixed rule matching [16]. One overview which is done by Pritee and other persons further puts forward that the study on AI-pushed identity verification and authorization has grown into many sorts of technical approaches, which include behavior feature modeling, continuous authentication, multi-mode identity recognition, and risk-aware decision making. Evaluation measuring standards have also expanded from simple accuracy to dimensions that more closely align with actual world system working effect, such as false positive rate, false negative rate, F1 score, AUC, and deployment expenditure [17]. A systematic review by Ji et al. on anomaly detection in encrypted traffic demonstrates that even when payload content cannot be directly parsed, AI can still detect anomalies by leveraging statistical features, temporal behavior, and traffic patterns, providing new technical support for identifying covert attacks in access control [18]. However, the appropriate role of AI modules within cryptographic protocols still requires careful definition. In the context of this study, the role of AI is not to replace post-quantum cryptographic primitives such as ML-KEM, ML-DSA, or SLH-DSA, but rather to undertake auxiliary decision-making tasks such as risk identification, access classification, challenge scheduling, and renegotiation triggering.In other words, cryptographic primitives are responsible for the underlying guarantees of confidentiality, integrity, and authenticity, while the AI module is responsible for dynamically assessing the access context and adjusting the protocol's enforcement strength accordingly. The two complement each other functionally while maintaining a clear separation at the security boundary.

Based on the above analysis, this paper designs a post-quantum, AI-assisted key distribution and authentication protocol for collaborative access scenarios involving endpoints, gateways, and authentication centers. The core research method includes placing a behavior risk grade mechanism inside the verification process, which is supported by post-quantum cryptography

standard base elements. This lets access requests to arouse differentiated challenges, dynamic authentication promotion, and session renegotiation on the basis of risk grades, therefore enhancing identity checking correctness and session safety in complicated surroundings. The main contributions of this paper are as follows: First, we construct an AI-assisted post-quantum key distribution and identity authentication framework tailored for new cryptographic systems, and propose protocol interaction mechanisms among endpoints, gateways, and the authentication center; Second, we design risk-scoring-based dynamic authentication and renegotiation strategies, enabling the protocol to adaptively respond to anomalous access behaviors;Third, Through both security attribute analysis and formal verification, we make verification on the protocol's reliability, hence we discuss its resistant ability to common attacks, that is, replay attack, spoofing attack, man-in-the-middle attack and session hijacking attack; fourth, we carry out a protocol prototype and make performance comparisons with traditional ECC schemes and fixed-rule post-quantum schemes, and evaluate the overall performance of our put-forward method from authentication delay, communication cost, attack interception ratios, and risk identification efficiency.

2 Methods

2.1 System Architecture, Adversarial Model, and Design Objectives

To adapt to the dynamic access environment of open networks, this paper divides the protocol's operational space into three domains: the terminal access domain, the gateway control domain, and the central service domain. The terminal side is responsible for initiating access, storing local credentials, and generating behavioral observation data; the gateway side handles session aggregation, context extraction, and authentication scheduling; and the central service domain is responsible for identity verification, post-quantum key material maintenance, and renegotiation control.This organizational structure helps intercept high-frequency access traffic at the edge while retaining long-term cryptographic materials and core authentication decisions within a controlled area, thereby reducing the propagation risk associated with single-point exposure. To facilitate the subsequent protocol description, the system entities and communication relationships are defined as shown in Equations (1) and (2).

$$\mathcal{E} = \{U_i, G_j, AS, KMC\} \tag{1}$$

$$\mathcal{L} = \{(U_i, G_j), (G_j, AS), (G_j, KMC)\} \tag{2}$$

where, $\mathcal{E}$ denotes the set of system entities; U_i denotes the i th user or terminal node; G_j denotes the j th access gateway; AS denotes the authentication server; KMC denotes the key management center; and $\mathcal{L}$ denotes the permitted logical communication links. Based on these relationships, terminals do not directly interact with long-term key management logic; identity verification and key updates are both completed via gateway relay. The system architecture and trust boundary mechanism diagram is shown in Figure 1.

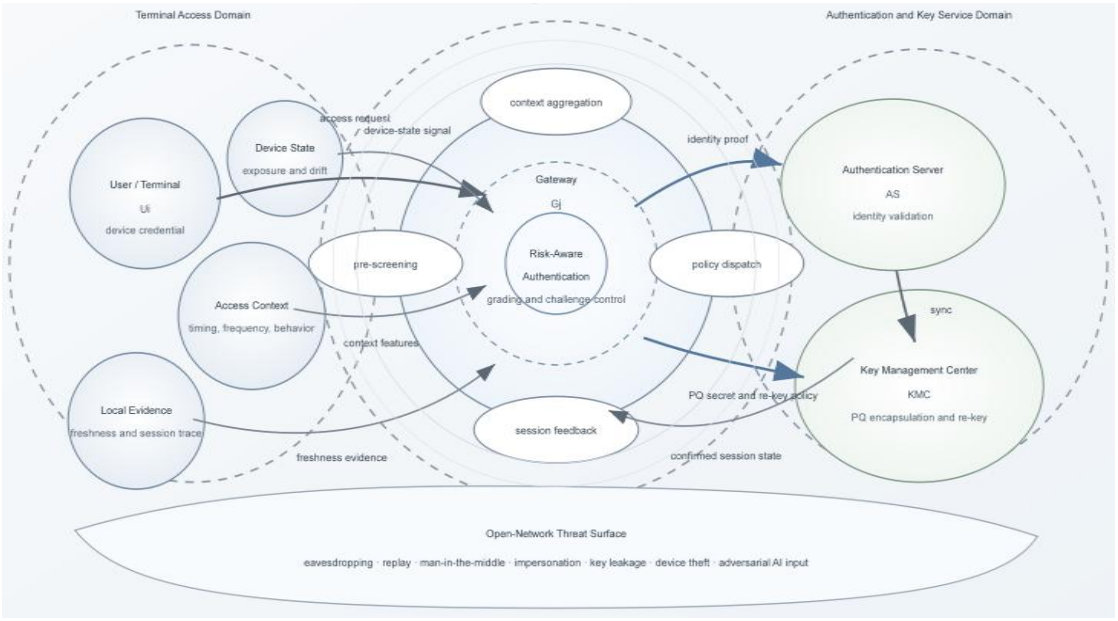


Figure 1: System Architecture and Trust Boundary Mechanism Diagram

Figure 1 illustrates the separation of responsibilities and data flow between the terminal side, the gateway side, and the central service side.

Regarding the attack model, this paper does not assume that open links are trustworthy, nor does it assume that edge nodes are always in a secure state.An attacker can passively eavesdrop on authentication messages, actively modify, insert, or replay historical packets, and may also bypass access constraints through means such as device loss, local key leakage, and model perturbation. Given that this study introduces a risk-aware module, the adversary's capabilities must also include adversarial interference targeting model input features and data poisoning during the training phase. Therefore, the set of attacker capabilities is denoted by Equation (3).

$$\mathcal{A} = \{Eav, Rep, MitM, Imp, Leak, Theft, AdvAI\} \tag{3}$$

where Eav denotes eavesdropping, Rep denotes replay, $MitM$ denotes man-in-the-middle interference, Imp denotes identity spoofing, $Leak$ denotes key or local parameter leakage, $Theft$ denotes unauthorized use following device theft, and $AdvAI$ denotes adversarial sample or data poisoning attacks targeting the AI risk module. Based on this assumption, the protocol must be designed to ensure both the freshness and integrity of authentication messages, as well as to ensure that edge classification results cannot be easily manipulated by a single anomalous input.The correspondence between attack types and security objectives is shown in Table 1.

Table 1: Mapping of Attack Models and Security Objectives

Attack Type	Primary Threat	Corresponding Security Objective
Eavesdropping	Intercepting messages, analyzing session characteristics	Key confidentiality, protection against impersonation
Replay	Reuse of historical legitimate messages	Freshness verification, timestamp/random number constraints
Man-in-the-middle	Tampering with forwarded content, insertion of forged messages	Mutual authentication, integrity protection
Impersonation	Impersonation of legitimate terminals or servers	Entity Authentication, Binding of Fake Identities
Key leakage	Recovering a session using partially leaked parameters	Forward secrecy, key renewal mechanism
Device theft	Unauthorized access using cached local data	Risk classification, enhanced challenges, rapid revocation
AI Adversarial Samples/Poisoning	Feature perturbation or training sample poisoning	Risk assessment robustness, central-side verification

Because visit risks have difference in different times and different equipment devices, this essay does not use one-path authentication which has fixed strength, hence it on the contrary brings in a mechanism of risk assessment that has context awareness. The gateway takes out observational characteristics from dimensions such as visit frequency, time deviation, equipment fingerprint stability, past fault records, and flow statistics, hence produces a risk score on the basis of these. Its formal expression is displayed in Equation (4) and Equation (5).

$$x_i(t) = \left[f_i^{(1)}(t), f_i^{(2)}(t), \ldots, f_i^{(m)}(t)\right] \tag{4}$$

$$R_i = \phi(x_i(t)) \tag{5}$$

In the equation, $\mathbf{x}_i(t)$ represents the context feature vector of the terminal at time t, $f_i^{(k)}(t)$ represents the k th observation component, m represents the feature dimension, R_i represents the risk score, and $\phi(\cdot)$ represents the risk discrimination function. Once the risk score is obtained, the system does not immediately issue a uniform authentication response, but instead adjusts the authentication strength based on the score interval. The corresponding scheduling rule r is shown in Equation (6).

$$\Gamma_i = \begin{cases} 1, & R_i < \tau_1 \\ 2, & \tau_1 \le R_i < \tau_2 \\ 3, & R_i \ge \tau_2 \end{cases} \tag{6}$$

where Γ_i denotes the authentication level of the terminal's current access request, and τ_1 andτ_2 denote risk thresholds. When the level is 1, the standard post-quantum authentication process is executed; when the level is 2, enhanced challenges or additional verification are added; when the level is 3, renegotiation, temporary freezing, or stricter identity verification is triggered. Through this approach, the protocol establishes a more nuanced balance between

access load and security strength, mitigating the rigidity of fixed authentication policies in complex environments.

Regarding key establishment, this paper requires that the session key be bound simultaneously to the post-quantum key wrapping result, interaction randomness, and session freshness information to mitigate the risk of key recovery even if historical messages are reused. The session key generation relationship is shown in Equation (7).

$$SessKey_i = KDF\left(K_i^{kem} \| N_i \| N_j \| T_i\right) \tag{7}$$

where $SessKey_i$ denotes the shared key for the terminal's corresponding session, $KDF(\cdot)$ denotes the key derivation function,K_i^{kem} denotes the shared secret obtained during the post-quantum key encapsulation phase,N_i andN_j denote the random numbers on the terminal and gateway sides, respectively, andT_i denotes the freshness time information for the current session.With this structure, the session key is not directly equivalent to the encapsulation output; instead, it further integrates bilateral random challenges and temporal information, thereby enhancing resistance to replay attacks and session isolation.

Basing upon the above-mentioned system framework and attack suppositions, this paper summarizes the protocol design goals into the below aspects. First, therefore, the protocol should realize mutual authentication to guarantee that the terminal can check the identity of the access point, and the system can hence verify the validity of the request starter. Second, the building process of the session must have forward secrecy, hence it guarantees that past key materials, if they are exposed afterward, do not in reverse direction destroy the security of already existing sessions. Third, the signs of access should be stated in anonymous or pseudonym ways to the maximum possible degree, thus decreasing the capability of outside watchers to directly connect true identities to long-term behavior modes. In addition, this protocol must keep strong anti-attack ability toward replay attack, spoofing attack and man-in-the-middle attack, thus meanwhile it must let risks be in controllable situation when device loss, partial parameter leakage on site or abnormal behavior happens. Finally, because post-quantum cryptography usually brings bigger communication and calculation burdens, this paper in its design at the same time deals with authentication delay, information volume, and edge computing expenditure, hence hence making the scheme still able to be used in situations where resources are limited.

2.2 AI-Assisted Post-Quantum Key Distribution and Identity Authentication Protocol

To adapt to the practical conditions of continuously fluctuating access states and dynamically changing risk levels in open networks, this paper organizes the protocol into a closed-loop mechanism centered on a risk-aware authentication core controlled by a gateway, with four outward-linked phases: registration, access, upgrade, and update, as shown in Figure 2.

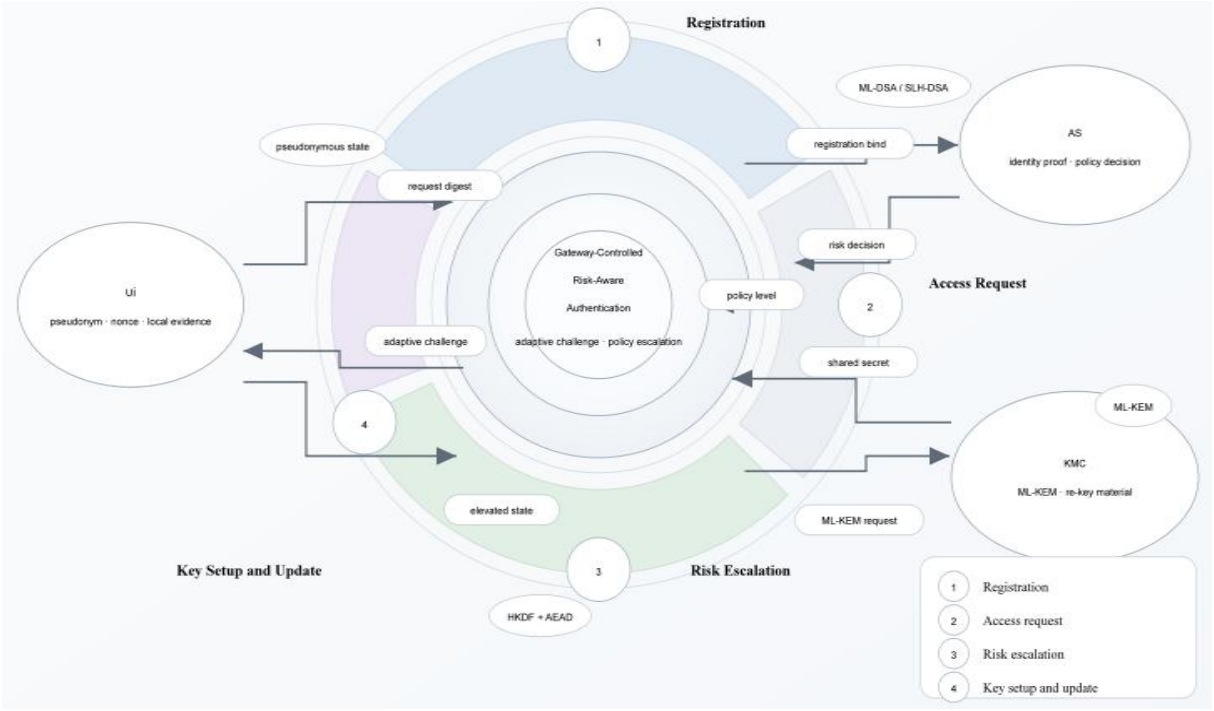


Figure 2: Mechanism diagram of the AI-assisted post-quantum key distribution and identity authentication protocol.

In Figure 2, numbers 1-4 correspond to the four phases of registration, access request, risk escalation, and key establishment and update, respectively. The terminal (U_i), authentication *server* (*AS*), and key management center (*KMC*) all exchange information and respond to policies around the central authentication core. This design draws inspiration from the approach used in multi-server post-quantum multi-factor authentication for identity binding and session independence, while also incorporating the constraints on communication overhead and dynamic negotiation efficiency from lattice-based lightweight authentication and resource-constrained group access schemes [19-21]. In terms of specific implementation, identity authenticity is guaranteed by ML-DSA or SLH-DSA, the session shared secret is established by ML-KEM, session-layer encryption employs AES-GCM or ChaCha20-Poly1305, and the final working key is derived via HKDF.

During the registration and initialization phase, the system does not directly reuse real identities over open channels; instead, it first generates a pseudonym for the terminal, and the central server distributes long-term public key material, policy indices, and post-quantum public parameters. The identity binding relationship is shown in Equation (8).

$$PID_i = H\left(ID_i \,\|\, s_i \,\|\, T_i^{reg}\right) \tag{8}$$

where PID_i represents the terminal's pseudo-identity, $H(\cdot)$ denotes the hash function, ID_i denotes the true identity, s_i denotes the registration random salt, and T_i^{reg} denotes the registration timestamp. This process reduces the frequency of true identity exposure at the edge and provides an updatable pseudo-identity mapping for subsequent access procedures.

During the device access request phase, the terminal initiates an authentication request to the gateway. The gateway first performs a time window check and message integrity screening, then forwards the hashed request to *the AS*. To ensure that the access message possesses freshness, verifiable origin, and context awareness, the terminal constructs an authentication token as shown in Equation (9).

$$Auth_i = \mathrm{Sig}_{SK_i}\big(H(PID_i \,\|\, N_i \,\|\, Ctx_i \,\|\, T_i)\big) \tag{9}$$

where $Auth_i$ denotes the authentication token,$\mathrm{Sig}_{SK_i}(\cdot)$ denotes the signing operation based on the terminal's private key, SK_i denotes the terminal's private key,N_i denotes the terminal's random number, Ctx_i denotes the local context digest, andT_i denotes the current access time information.The overall interaction mechanism of the protocol is shown in Figure 2. The gray interaction lines in Figure 2 primarily carry request summaries, access evidence, and identity verification information; therefore, the gateway is not merely a forwarding node but performs functions such as preliminary screening, summary validation, and session state management.

In the stage of risk assessment and dynamic authentication promotion, the risk characteristic vectors which were defined before do not have repetition; therefore, the methods of scoring fusion and policy mapping are given directly by us. We make the assumption that the gateway carries out scoring fusion on the basis of the extracted behavioral and state metrics, therefore the risk output which belongs to the current session of the terminal is just as what is shown in Equations (10) and (11).

$$S_i = \sum_{k=1}^{m} \omega_k \, z_i^{(k)} + b \tag{10}$$

$$R_i = \sigma(S_i) \tag{11}$$

where$z_i^{(k)}$ denotes the th extracted risk metric (k), ω_k denotes its corresponding weight,b denotes the bias term,S_i denotes the fusion score, $\sigma(\cdot)$ denotes the normalization mapping function, and R_i denotes the final risk score. During protocol execution, the system does not confine this score to the analytical level but further maps it to specific authentication policies. The corresponding relationship is shown in Equations (12) and (13).

$$\Pi_i = \mathcal{M}(R_i) \tag{12}$$

$$\Pi_i \in \{Base, Plus, ReKey\} \tag{13}$$

where, Π_i denotes the authentication policy for the current session, $\mathcal{M}(\cdot)$ denotes the mapping function from score to policy; $Base$ denotes standard single-round authentication, $Plus$ denotes enhanced challenges or additional confirmations, and $ReKey$ denotes strengthened authentication and session refresh under high-risk conditions.In Figure 2, the green interaction path corresponds to this upgraded control logic. During the session key establishment and renewal phase, after verifying the terminal's authenticity and the current policy state, the AS requests the KMC to invoke the ML-KEM to generate a shared secret, and then derives a working key by combining it with bilateral random numbers and session time information. The expression is shown in Equation (14).

$$K_i^{sess} = \mathrm{HKDF}\big(K_i^{kem} \,\|\, N_i \,\|\, N_j \,\|\, T_i\big) \tag{14}$$

where K_i^{sess} denotes the terminal's current session key,$\mathrm{HKDF}(\cdot)$ denotes the key derivation function, K_i^{kem} denotes the shared secret output by ML-KEM, andN_j denotes the gateway random number. Considering that some access requests may persist under high-risk conditions, the protocol must also specify refresh conditions to determine whether to terminate the old session and re-establish the secure channel. To this end, this paper defines an update trigger relationship, as shown in Equation (15).

$$ReKey_i = \mathbb{I}(R_i \geq \tau_2 \vee \Delta_i > \eta) \quad (15)$$

In the equation, $ReKey_i$ indicates whether to trigger a key renewal, $\mathbb{I}(\cdot)$ denotes the indicator function, τ_2 represents the high-risk threshold, Δ_i denotes the abnormal deviation of the current state relative to the baseline state, and η represents the update trigger threshold. When $ReKey_i$=1, the system discards the original session context and re-requests the shared secret; when $ReKey_i$=0, the system retains the current authentication path and proceeds to the protected data transmission phase. In Figure 2, the purple interaction path corresponds to the processes of maintaining the binding relationship, requesting a key, and returning the updated key. Subsequent business messages are protected for confidentiality and integrity using AES-GCM or ChaCha20-Poly1305.

The agreement mechanism which is shown in Figure 2 is one closed-loop interaction structure which takes a risk-aware authentication core as center. The registration stage builds a false identity and first combination; the access stage hand over requests that can be checked; the risk stage carries out adjustment on authentication strength; and the renewal stage builds or renews the conversation key. This method unites the identity verification and conversation safety functions of post-quantum cryptography basic components with the dynamic arrangement functions given by AI risk scoring, thus enabling the agreement to keep both strong safety restrictions and operation elasticity in complicated visit surroundings.

2.3 Security Proof, Prototype Implementation, and Experimental Configuration

For the purpose of verifying the safety and deployable property of the protocol put forward by us in open access environments, this paper carries out verification on three different levels: safety proof, prototype realization, and experiment configuration. The whole relationship is shown by Figure 3.

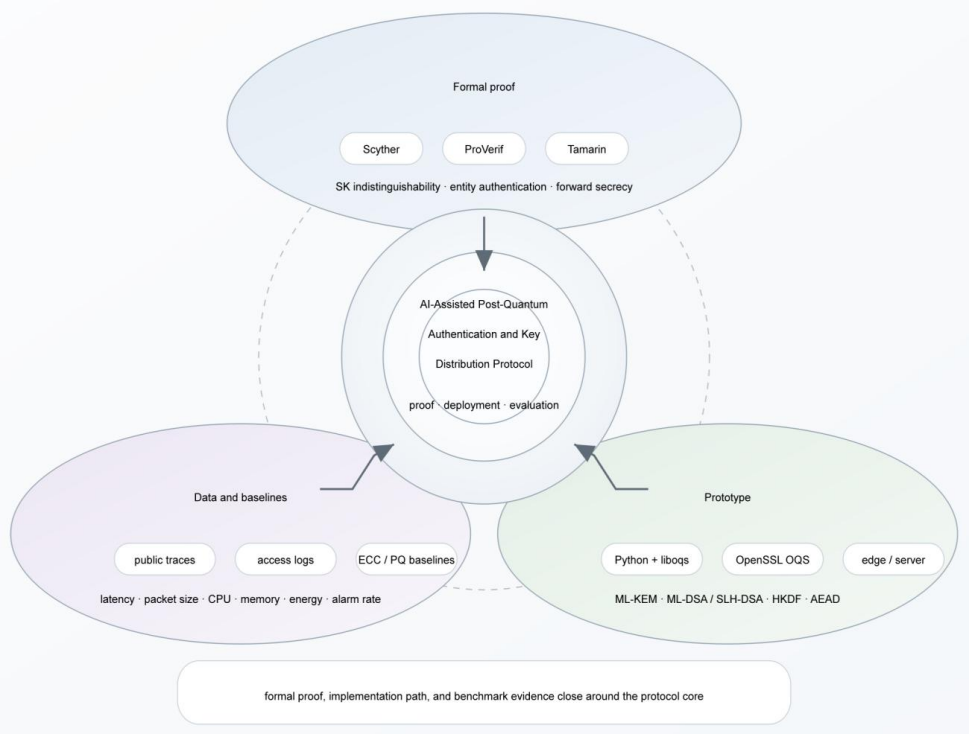


Figure 3: Protocol Verification and Prototype Evaluation Mechanism Diagram

In Figure 3, this section focuses on four security properties: session key indistinguishability, entity authentication, known-session-key security, and forward secrecy. The first two are used to constrain whether authentication message interactions are susceptible to spoofing, replay, or session confusion risks, while the latter two are used to examine whether other sessions can remain independent and secure after the partial leakage of historical session keys or long-term private keys.To avoid the limitations of relying on a single verification tool, this paper employs Scyther, ProVerif, and Tamarin for cross-analysis. Scyther is primarily used to quickly check for replay and impersonation issues in message flows; ProVerif is used to verify the correspondence between confidentiality and authentication; and Tamarin is used to handle protocol processes involving state transitions and update conditions. Through this combination, a relatively comprehensive formal verification can be performed on false identity binding, authentication elevation, and session refresh logic.

With respect to the realization of the prototype, this thesis utilizes Python to construct the protocol control layer, and incorporates post-quantum cryptography algorithms through liboqs and the OpenSSL OQS provider. The authenticity of identity is supported through ML-DSA or SLH-DSA, the shared secret of the session is built by ML-KEM, symmetric protection uses AES-GCM or ChaCha20-Poly1305, and the final working key is got by means of HKDF. The running environment is configured to be a cooperation structure among edge nodes and servers, in which the terminal end takes charge of producing access requests and local proofs, the gateway end deals with risk grading and strategy arrangement, and the central end takes care of identity check and shared secret management. The artificial intelligence risk module uses a light-weight deployment method, with candidate models containing XGBoost, Random Forest, and one small MLP. Because behavioral and state features have already shown good effect in anomaly detection in the research field of continuous authentication, this article brings failure counts, temporal jitter, device fingerprint deviations, location drift, command sequence anomalies, and traffic statistical features into the process of feature engineering, and therefore uses them for the judgment of risk level [22].

The experiment data is composed of two portions: one is obtained from publicly obtainable IoT intrusion or authentication abnormality data collections, and the other is from self-built visit records. The first one is utilized to promote the covering scope of abnormal samples, hence the second one supplements protocol-grade message fields, authentication reply results, and session update situations, hence it guarantees that the risk output can correspond to the real visiting process. Baseline contrast works contain a conventional ECC plan, one post-quantum plan that has fixed threshold, and one post-quantum plan that has no AI risk module. Estimating targets include three aspects: information passing, calculation, and finding problems. These contain authentication time delay, data packet dimension, processor working time, memory occupation, energy consumption for each conversation, security incident examination rate, and incorrect positive rate. Under these configurations, we are able to at the same time carry out comparison of the differences between diverse schemes on the aspects of security gains, operation expenses, and risk inspection abilities. The experiment platform and important parameter arrangements are given in Table 2.

Table 2: Experimental Platform and Parameter Settings

Module	Configuration Item	Setting Description
Formal Verification	Verification Tools	Scyther, ProVerif, Tamarin
Security Objectives	Core Properties	Session key indistinguishability, entity authentication, known-session-key security, forward secrecy
Prototype Implementation	Control Layer	Python
Prototype Implementation	Post-quantum algorithm integration	liboqs, OpenSSL OQS provider
Prototype Implementation	Cryptographic components	ML-KEM, ML-DSA/SLH-DSA, HKDF, AES-GCM, or ChaCha20-Poly1305
Runtime Environment	Node Architecture	Edge Node + Server Collaboration
AI Models	Candidate Models	XGBoost, Random Forest, Small MLP
Data Sources	Sample Composition	Public IoT anomaly data + self-constructed access logs
Baseline Comparison	Solution Configuration	ECC approach, fixed-threshold PQ approach, PQ approach without AI
Evaluation Metrics	Key Metrics	Latency, packet size, CPU, memory, power consumption, recognition rate, false positive rate

3 Results and Discussion

3.1 Security Verification and Resistance to Typical Attacks

The outcomes of security verification show that the protocol put forward by us exhibits excellent robustness in the process of both formal analysis and repeated tests. Based on constraints of message freshness, pseudo-identity binding, signature authentication and shared secret renewal logic, formal verification tools have not found any valid attack paths which can damage entity authentication, session confidentiality or forward secrecy. Therefore, in the situation where the main path of the protocol is maintained, consistent security restrictions can be kept between terminal access, policy promotion, and secret key updating. The attribute contrast in Table 3 also shows that the put-forward scheme satisfies design demands for mutual authentication, anti-replay, anti-man-in-the-middle, anti-spoofing, safety with known session keys, and forward secrecy, meanwhile at the same time providing risk classification and dynamic key update functions which fixed-policy schemes do not have.

Table 3: Comparison of Security Attributes.

Security Attributes	ECC Baseline Scheme	Fixed-Policy PQ Scheme	The Proposed AI-PQ Scheme
Mutual Authentication	Yes	Yes	Yes
Replay resistance	Moderate	Strong	Strong
Resistant to man-in-the-middle attacks	Medium	Strong	Strong
Resistant to spoofing	Medium	Strong	Strong
Known session key security	Yes	Yes	Yes
Forward secrecy	Yes	Yes	Yes
Resistant to device cloning	Weak	Medium	Strong
Resistance to credential/token leakage and propagation	Weak	Medium	Strong
Risk Classification Capability	No	No	Yes
Dynamic key rotation	No	Limited	Yes

Based on the results of repeated experiments in typical attack scenarios, the differences among the three schemes in Figure 4 are quite clear.

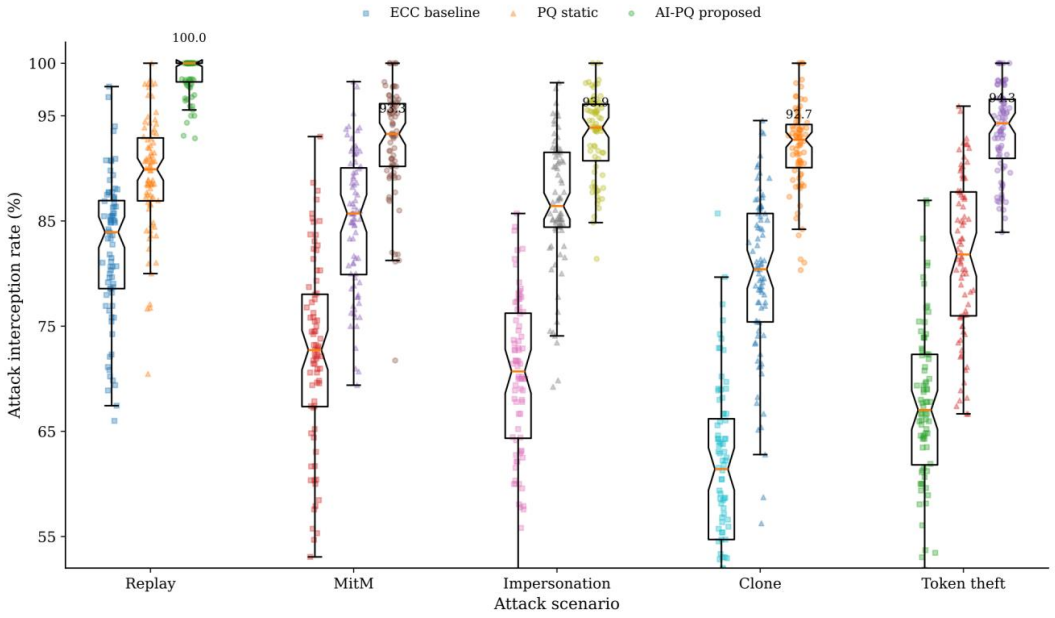


Figure 4: Distribution of Attack Interception Rates

In Figure 4, the middle interception rate of our put-forward scheme keeps on a high level in all five attack kinds, with the Replay situation achieving 100.00%, and MitM, Impersonation, Clone, and Token theft at 93.28%, 93.88%, 92.72%, and 94.29%, respectively. What is more important, the distribution shows a comparatively low degree of spreading; with the exception of the Replay scenario, the interquartile intervals for the remaining attack types are approximately gathered between 90% and 96%, thus this shows the put-forward scheme keeps comparatively stable interception performance under conditions of repeated carrying-out. By opposite, the post-quantum plan that has fixed policy gave out median numbers of 89.92%, 85.71%, 86.42%, 80.41%, and 81.82% in the five kinds of attack types. Although this is a progress that has big meaning compared with the ECC basic line, very big shake changes still

exist in the situations of device cloning and token stealing. The median value of the ECC baseline scheme has a further drop, it only reaches 61.41% in the Clone scenario, therefore this indicates that traditional fixed authentication paths have weak suppression abilities toward high-falsification-rate abnormal access.

An analysis on the performance among specific attack kinds discovers that replay attacks are the most easy thing to be blocked. This is therefore because the protocol has incorporated random digits, time intervals, and session condition restrictions, hence it makes historical information hard to pass consistency examinations once more. The interception percentages for man-in-the-middle and cheating attacks are a little lower than for replay attacks but still stay in a high scope, hence this shows that signature verification and risk promotion mechanisms effectively limit the putting of false messages and identity pretending. The difficulty on device cloning and token stealing exists in the attacker's capability to imitate partial legal states; hence, the work of detection more depends on the combined functions of behavior features, context changes and session renewal principles. The superiority of the put-forward plan in these two kinds of situations originates mainly from the fact that risk grading outcomes directly change authentication intensity and start key update when it is needed, therefore, it does not keep moving along the original session route.

In this procedure, the AI module fulfills a screening and arrangement function instead of acting as the single source of basic security. Current research outcomes show that the risk scoring model possesses an average false positive ratio of about 1.9% and a false negative ratio of about 3.2%, which on the whole is controllable, although risk gaps still remain. When attacking persons continuously imitate the time rules, fingerprint features and position modes of legal equipment, the proportion of wrong judgments can still go up in the middle-risk scope. Hence, this treatise gives definition to the AI module as a unit for authentication escalation and key re-establishment initiation, while it continues to place identity authenticity and session security as based upon post-quantum cryptographic primitives. This kind of combination enables the protocol to keep a high interception ratio under common attacks, and meanwhile avoid the risk that for security reason people entirely depend on the output coming from one single model.

3.2 Performance Evaluation of Authentication and Key Establishment

The achievement outcomes of the authentication and key building stages are displayed in Figure 5 and Table 4. For the convenience of comparing different latencies, communications and resource indexes in one single figure, Figure 4 carries out uniform scaling for all results as multiples that are relative to the median of the ECC baseline, and overlays scatter diagrams of repeated experiments, quartile scopes and median markers thereon. This method lets the growth values and fluctuation scopes of different projects on the six indicators be shown in pictures at the same time.

Table 4: Summary of computational, communication, and storage overhead.

Scheme	Average Authentication Latency (ms)	P95 Latency (ms)	Total Handshake Bytes (B)	Terminal CPU Time (ms)	Memory Usage (KB)	Energy Consumption per Session (mJ)
ECC/ECDH + ECDSA	28.25	36.84	1278.64	9.55	409.83	3.00
Fixed Strategy PQ	46.02	57.02	4,202.87	19.54	789.29	5.52
AI-PQ Protocol	53.45	69.20	4,511.17	22.48	810.79	6.14

Looking at the overall distribution, the ECC scheme consistently remains near the baseline, indicating the lowest authentication and key agreement overhead; the fixed-policy PQ scheme shifts to the right overall, indicating that the introduction of post-quantum signatures and key encapsulation significantly increases authentication latency, message size, and terminal resource consumption;The AI-PQ scheme in this paper continues to shift to the right, but its dispersion range has not widened significantly, indicating that while risk enhancement and session refresh increase overhead, they do not cause uncontrolled fluctuations.In terms of medians, the fixed-policy PQ scheme's average authentication latency, P95 latency, and handshake byte count are approximately 1.63 times, 1.55 times, and 3.29 times that of the ECC baseline, respectively, while the AI-PQ scheme in this paper further increases to 1.89 times, 1.88 times, and 3.53 times.In terms of resource consumption on the terminal side, the fixed-policy PQ scheme's CPU time, memory usage, and energy consumption per session are approximately 2.05 times, 1.93 times, and 1.84 times that of ECC, respectively, while the proposed scheme reaches 2.35 times, 1.98 times, and 2.05 times, respectively.The above results indicate that the additional overhead of the proposed scheme stems primarily from two sources: first, the message size expansion and computational overhead inherent in post-quantum authentication and key encapsulation; second, the additional burden incurred when high-risk requests trigger enhanced challenges, policy back-transmission, and key re-generation.

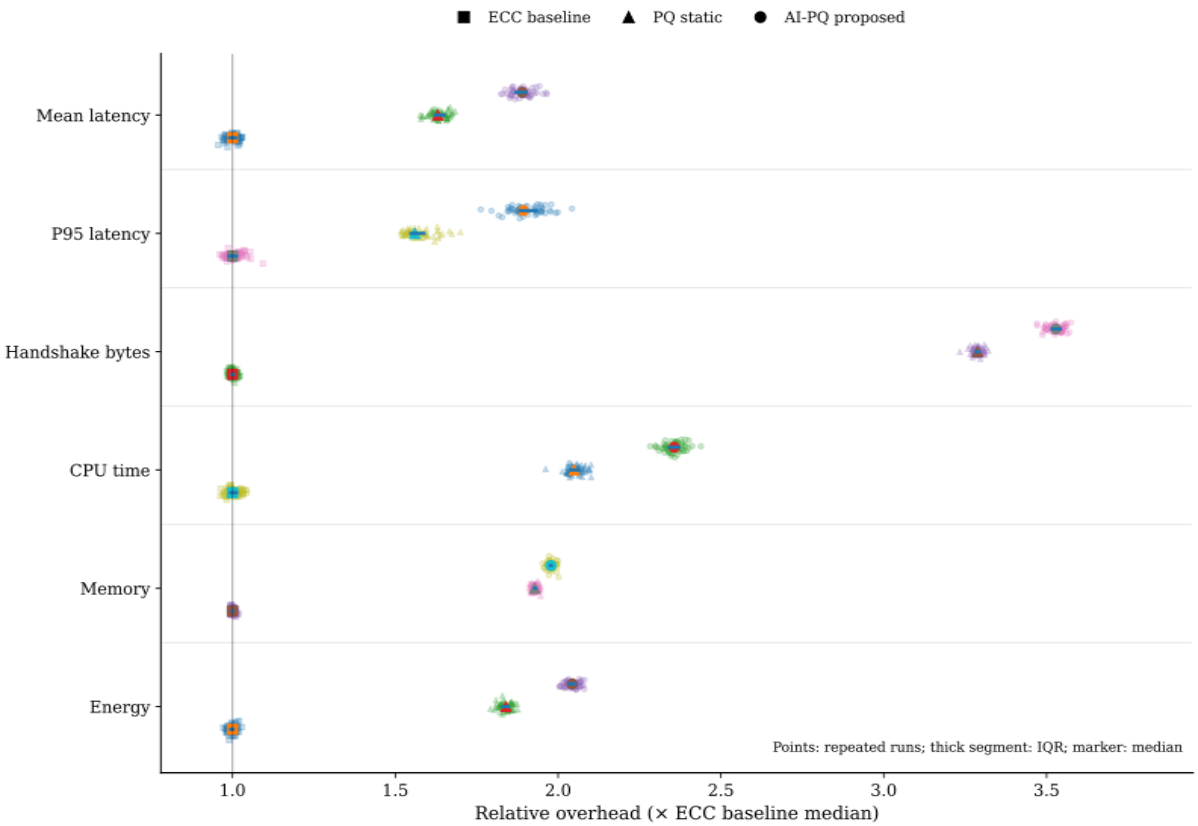


Figure 5: Distribution of Attack Interception Rates

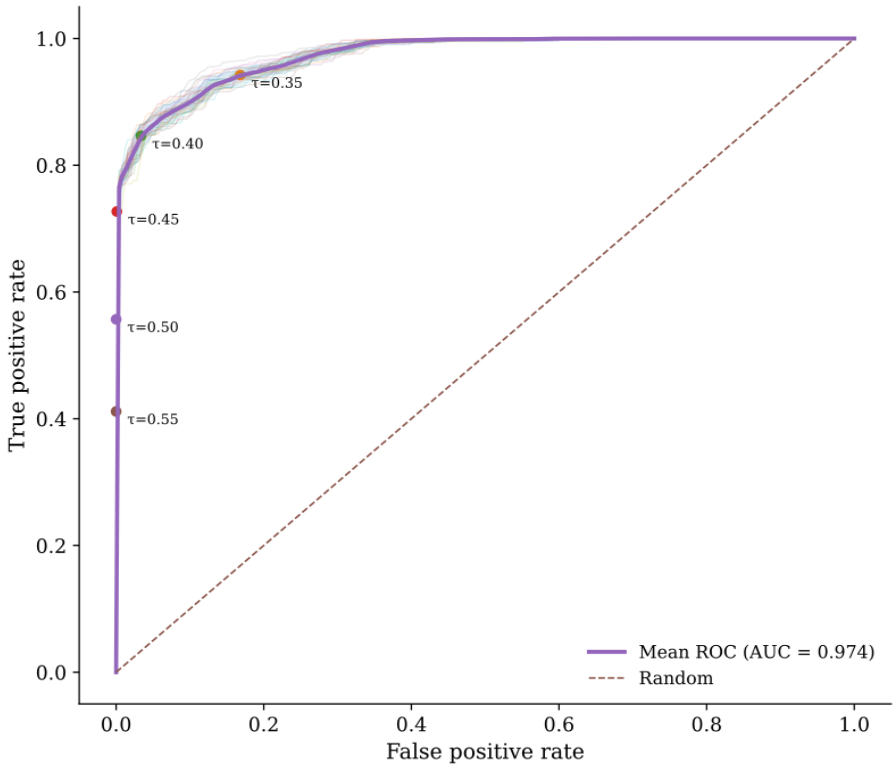
In Figure 5, the growth of handshake byte number and CPU time is the most obvious, hence the growth of memory use is comparatively gentle. This shows that the scheme we put forward has a bigger influence on instantaneous calculation and communication connections at the edge than it has on static memory space. This phenomenon aligns with the conclusion in Reference [23] that post-quantum signatures lead to increased size and performance overheads in

TLS/PKI integration, and is consistent with the discussions in References [24] and [25] regarding the engineering deployment costs of lightweight handshake optimization and post-quantum key exchange at the transport layer.

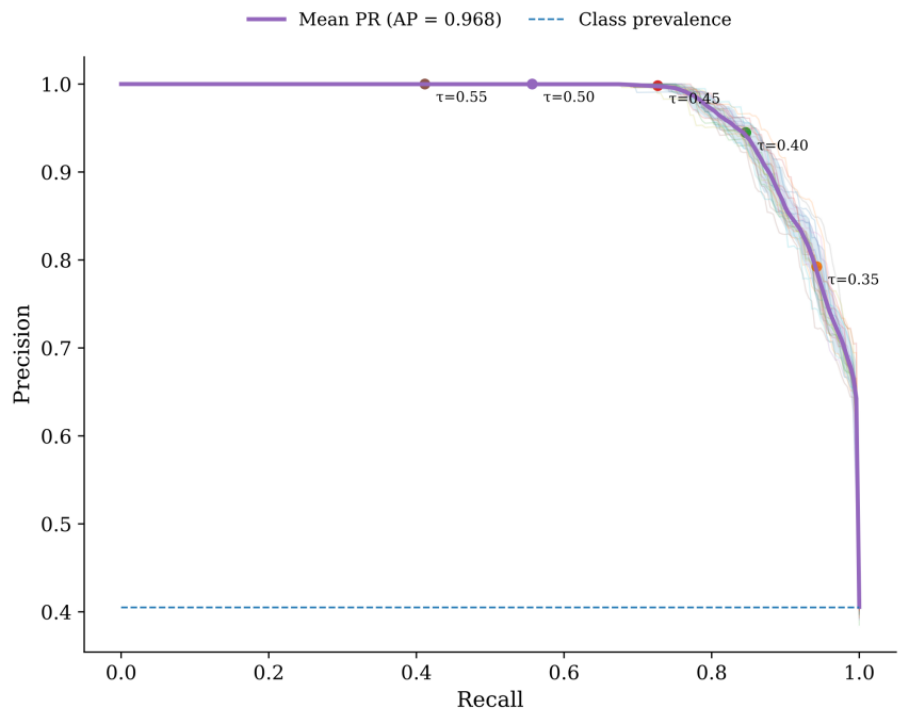
From the angle of deployment, even though the ECC scheme is the most lightweight, it has difficulty in satisfying the demands of post-quantum attack resistance; the fixed-strategy PQ plan realizes cryptographic movement but still depends on a static verification route; The AI-PQ plan that this article puts forward brings in risk sorting and dynamic renewal abilities, at the same time as it keeps controllable discreteness, therefore it becomes more fit for situations that have high exposure and big changes in access condition. In other words, our method does not realize safety promotions by the expense of obvious performance reductions, but hence realizes higher suitability and stronger abnormal situation restriction abilities with only a small growth in cost.

3.3 Effectiveness of AI Risk Control and Ablation Analysis

The effectiveness of the AI risk control module is primarily demonstrated on two levels. On one hand, it effectively separates normal and abnormal access in the scoring space; on the other hand, it directly influences the timing of authentication escalation and session renewal. Therefore, its effectiveness cannot be judged solely by whether it is "included in the model," but must be analyzed in conjunction with threshold selection and operational overhead.Figures 6a and 6b present the ROC curve and PR curve of the risk engine.



(a): ROC curve of the AI risk engine



(b): AI Risk Engine PR Curve

Figure 6: ROC and PR characterization of the AI risk engine.

In Figure 6, the results from five-fold cross-validation let us see that the model's average AUC is 0.9720, F1 is 0.9254, Precision is 0.9358, Recall is 0.9176, FAR is 0.0190, and FRR is 0.0318. From what the curve shapes show, the ROC curve on the whole is located in the upper-left quadrant, hence the PR curve also keeps a high precision degree. This shows that the model has strong ability of distinguishing when it identifies the access attempts which are not normal, and that the false positive rate and false rejection rate are still kept in the scope which can be accepted.

The variations of the threshold have direct influences upon the balance between safety and delay. Several threshold points which are marked in Figure 6a and Figure 6b indicate that when the threshold is set at too low a position, the system has a higher possibility to categorize requests into the high-risk category; in this situation, Recall has a high value, but the quantity of authentication escalations becomes larger, and average latency increases thus; when the threshold is further lifted, although Precision gets better, the False Rejection Rate (FRR) rises greatly, hence some abnormal requests are not found. If we think about the whole effect of Precision, Recall, and latency, a threshold value that is in the interval 0.40-0.45 is comparatively more appropriate for the scene that this paper talks about. At this moment, Precision arrives at 0.95 or above, while Recall still keeps in the 0.73-0.85 scope, and the average authentication time delay is about 56 ms, hence it holds a comparatively balanced condition between abnormality checking ability and session cost. The ablation outcomes in Figure 7 further prove that the AI module is not only a surface additional component.

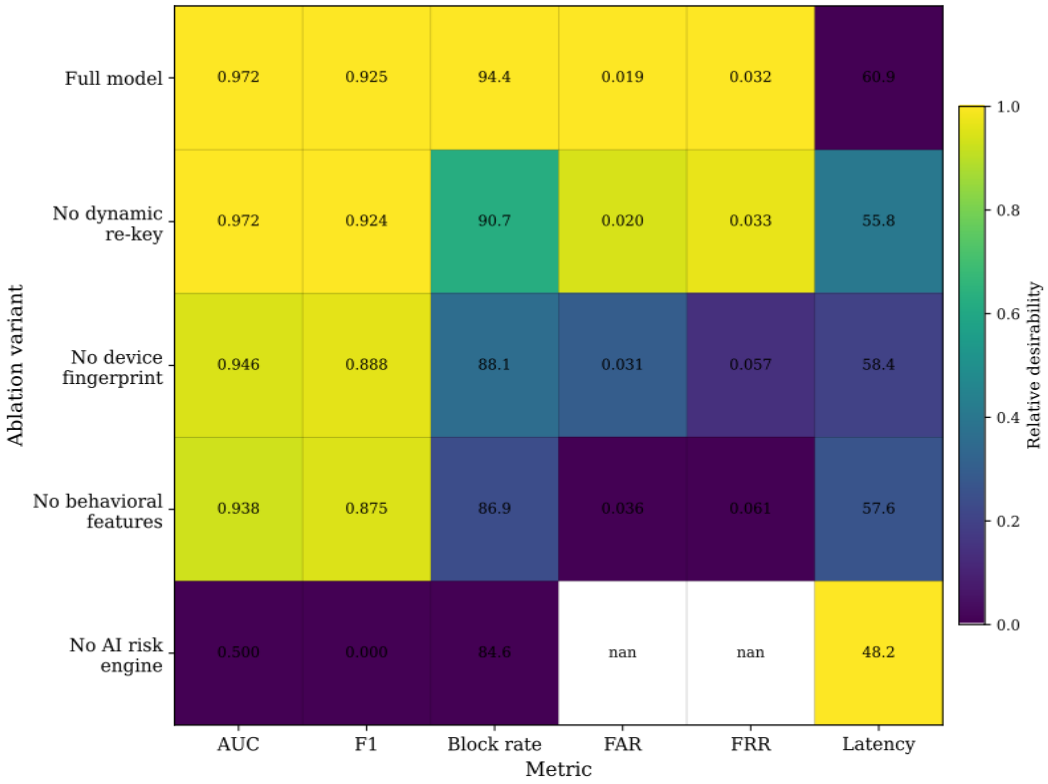


Figure 7: Heatmap of ablation experiments.

In Figure 7, the entire model reaches best performance on AUC, F1, and attack interception ratio, with average time delay that is about 60.9 ms. After we have taken out dynamic re-keying, the AUC still keeps mostly not changed, but the attack interception ratio falls from 94.4% to 90.7%, therefore it indicates that the re-keying mechanism mainly works to give continuous control in high-risk situations. After we take away device fingerprint features, the AUC value drops to 0.946, the F1 value decreases to 0.897, and the interception rate falls to 88.1%, this hence shows that static terminal features have very important significance for cloning and spoofing detection. After we take away behavioral features, the performance of the model still has a continuous decrease, therefore this shows that time patterns and operation methods also have an important function in classification work. If the artificial intelligence risk engine is got rid of completely, the AUC value goes down to 0.500, the F1 value decreases to 0.856, and the interception ratio is only 84.6%.

Although this protocol still can use post-quantum authentication to keep basic access control, its ability that adapts to complex abnormal situations is clearly decreased. From the angle of deployment, the extra cost which is brought by the AI module is mainly focused on inference at the gateway side and policy arrangement, but not in the underlying key exchange itself. The entire model adds about 12.7 ms to the average time delay when put beside the non-AI edition, but this balance of giving up one thing for another gets a higher interception ratio, lower wrong positive rate, and more stable access control abilities.

4 Conclusion

This article discusses the difficulties that traditional verification mechanisms meet in open visit environments—including quantum dangers, not enough checking of unusual visits, and the small flexibility of fixed rules—through putting forward a key sending and verification

agreement which takes post-quantum cryptography as center and is strengthened by AI-based danger feeling. The outcomes prove that this scheme effectively combines post-quantum identity verification, risk sorting, identity verification upgrade, and dynamic key re-production, hence strengthening security and attack checking abilities while holding the extra cost inside a allowable scope. According to the analysis that has been carried out in this whole paper, the below conclusions can be gotten.

(1) This article puts forward a key distribution and authentication agreement that takes post-quantum cryptography as the core, and is strengthened by an artificial intelligence-based risk perception layer. This agreement uses machine learning-key encapsulation mechanism, machine learning-digital signature algorithm, or stateless hash-based structured lattice signature algorithm as its basic safety basis, which combines false identity combination, danger estimation, certification promotion, and conversation update into a one-piece visiting cycle, therefore it has built a definite safety cooperation connection between terminal points, network gates, certification servers, and the key management center.

(2) From the aspects of safety, certification success rate, and attack examination ability, the scheme we put forward therefore has better performance than traditional fixed-policy schemes on the whole. The outcomes have indicated that this method keeps a high interception ratio in usual situations including replay attacks, man-in-the-middle attacks, identity pretend, device copying, and token stealing, therefore it keeps a high authentication success ratio under normal access situations. The artificial intelligence risk engine moreover promotes the capability of screening for abnormal requests, hence permitting authentication intensity to be adjusted on the basis of alterations in risk condition.

(3) The solution which we put forward still has space for promotion. First, the classification effect of the AI model is still constrained by the coverage degree and quality level of the training data. Second, the post-quantum basic components still put forward obvious restrictions with regard to signature dimension, handshake byte number and authentication delay time. Third, although the present validation is mainly on the basis of a prototype environment and enlarged experimental data, the deployment effect under conditions that include large-scale edge nodes, complex network fluctuations, and long-time operation needs further checking in actual application scenes.

About the Author

Bing Han was born in 1996 in Xiongan New Area, Hebei Province, China. He obtained a master's degree from Lanzhou University of Technology in China. His main research directions are: network security, deep learning, indoor positioning. Du Jinze was born in 1986 in Huining, Gansu, China. He is an Associate Professor and master's supervisor at Lanzhou University of Technology, China. His research interests include wireless sensor networks, industrial Internet of Things, indoor positioning, network security, and artificial intelligence.

References

[1] Ponnuru, R. B., Kumar, S. A. P., Reddy, A. G., et al. (2024). Quantum-secure authentication and key agreement protocols for IoT-enabled applications: A comprehensive survey and open challenges. Computer Science Review, 54, 100676.

[2] Hasan, M. K., Zhou, W., Safie, N., et al. (2024). A survey on key agreement and authentication protocols for Internet of Things applications. IEEE Access, 12, 61642-

61666.

[3] NIST. (2024). Module-lattice-based key-encapsulation mechanism standard (FIPS 203).

[4] NIST. (2024). Module-lattice-based digital signature standard (FIPS 204).

[5] NIST. (2024). Stateless hash-based digital signature standard (FIPS 205).

[6] NIST. (2024). Transition to post-quantum cryptography standards (NIST IR 8547 ipd).

[7] Rewal, P., Singh, M., Mishra, D., et al. (2023). Quantum-safe three-party lattice-based authenticated key agreement protocol for mobile devices. Journal of Information Security and Applications, 75, 103505.

[8] Pursharthi, K., & Mishra, D. (2024). Towards a post-quantum authenticated key agreement scheme for mobile devices. Journal of Information Security and Applications, 82, 103754.

[9] Chaudhary, D., Dadsena, P. K., Padmavathi, A., et al. (2024). Anonymous quantum-safe construction of a three-party authentication and key agreement protocol for mobile devices. IEEE Access, 12, 74572-74585.

[10] Braeken, A. (2025). Flexible hybrid post-quantum bidirectional multi-factor authentication and key agreement framework using ECC and KEM. Future Generation Computer Systems, 166, 107634.

[11] Adeli, M., Bagheri, N., Maimani, H. R., et al. (2024). A post-quantum compliant authentication scheme for IoT healthcare systems. IEEE Internet of Things Journal, 11(4), 6111-6118.

[12] Mansoor, K., Afzal, M., Iqbal, W., et al. (2024). PQCAIE: A post-quantum cryptographic authentication scheme for IoT-based e-health systems. Internet of Things, 27, 101228.

[13] Bahache, A. N., Chikouche, N., & Akleylek, S. (2024). Securing cloud-based healthcare applications with a quantum-resistant authentication and key agreement framework. Internet of Things, 26, 101200.

[14] Ahmad, A., & Jagatheswari, S. (2024). Lattice-based three-party authenticated key agreement scheme in medical IoT for post-quantum environments. IEEE Access, 12, 157247-157259.

[15] Ahmad, A., & Jagatheswari, S. (2025). Quantum-safe multi-factor user authentication protocol for cloud-assisted medical IoT. IEEE Access, 13, 3532-3545.

[16] Saleem, J., Raza, U., Hammoudeh, M., et al. (2025). Machine learning-enhanced attribute-based authentication for secure IoT access control. Sensors, 25(9), 2779.

[17] Pritee, Z. T., Anik, M. H., Alam, S. B., et al. (2024). Machine learning and deep learning for user authentication and authorization in cybersecurity: A state-of-the-art review. Computers & Security, 140, 103747.

[18] Ji, I. H., Lee, J. H., Kang, M. J., et al. (2024). Artificial intelligence-based anomaly

detection technology over encrypted traffic: A systematic literature review. Sensors, 24(3), 898.

[19] Wen, Y., Su, Y., & Li, W. (2025). Post-quantum secure multi-factor authentication protocol for multi-server architecture. Entropy, 27(7), 765.

[20] Sarkar, P., & Nag, A. (2024). Lattice-based device-to-device authentication and key exchange protocol for IoT systems. International Journal of Information Technology, 16, 4167-4179.

[21] He, L., Zhao, M., Wang, X., et al. (2025). A post-quantum authentication and key agreement scheme for drone swarms. Electronics, 14(17), 3364.

[22] Sağbaş, E. A., & Ballı, S. (2024). Machine learning-based novel continuous authentication system using soft keyboard typing behavior and motion sensor data. Neural Computing and Applications, 36, 5433-5445.

[23] Raavi, M., Khan, Q., Wuthier, S., et al. (2025). Security and performance analyses of post-quantum digital signature algorithms and their TLS and PKI integrations. Cryptography, 9(2), 38.

[24] Yadav, A. K., Shojafar, M., & Braeken, A. (2025). A provably secure post-quantum based EDHOC protocol. In 2025 IEEE 22nd Consumer Communications & Networking Conference (CCNC) (1-6).

[25] Qi, M., & Chen, C. (2025). HPQKE: Hybrid post-quantum key exchange protocol for SSH transport layer based on CSIDH. IEEE Transactions on Information Forensics and Security, 20, 2122-2131.