Policy-Governed Post-Quantum Migration for Legacy Microservices: An Ephemeral Sidecar Pattern with Crypto-Agility, Performance Budgets, and Rollback Safety

Shuaib Ahmed
Bellevue, Washington, USA
khans8297@gmail.com

***Abstract*— Quantum computing has a potential threat to designed microservice architectures of the past due to classical cryptographic algorithms. It is not easy to migrate these systems to post-quantum cryptography (PQC) because of the performance overheads, complexity in its operation and compatibility with existing services. The paper suggests a Policy-Governed Post-Quantum Migration Framework (PG-PQM) which will allow the introduction of PQC in legacy microservices without any major changes to the application level. The architecture proposed is also designed to incorporate adaptive security scoring, automatic rollback and ephemeral key management to offer reliability in the process of migration. A containerized microservices testbed based on the selections of NIST PQC candidate algorithms was experimentally evaluated. Findings indicate that the suggested framework can cut down on the migration overheads by approximately 28% relative to the fixed PQC deployment strategies, and yet minimizes the latency overheads to less than 12 percent and enhances throughput by 18% under dynamic workload settings. Moreover, the automated roll back system effectively reduced more than 95% of the performance anomalies in terms of transitioning algorithms. These findings indicate that the offered framework offers a viable and scalable method to allow quantum-resilient security to be enabled within the legacy microservice environments.**

Keywords*— *Post-Quantum Cryptography, Crypto-Agility, Microservices Security, Sidecar Architecture, Policy-Driven Security, Quantum-Safe Migration, Service Mesh Security, Ephemeral Keying, Rollback Safety, Cloud-Native Security.

I. INTRODUCTION

The use of cloud-native computing has entirely transformed the existing application development processes [1], and the microservice systems have gained the new status of a cloud-native architecture under which the systems can be scaled and resilient [2]. The decentralizations of these systems are high on the secure inter-service communication that is typically established using well-defined cryptography protocols [3]. However, the majority of the available microservices remain founded on the traditional cryptography algorithms [4], such as RSA and elliptic curve cryptography, which are becoming increasingly vulnerable to the new threat of quantum computing [5]. The danger of mass quantum machines is the danger of harvest now, decrypt later [6], in which case the opponents encode encrypted data in the current where this information will be decoded in the future as soon as quantum attacks become possible [7]. This is a looming danger that has necessitated the necessity to expedite the necessity to implement post-quantum cryptography systems that have the ability to ensure the confidentiality of data over a long duration [8].

The post-quantum cryptography offers quantum resistant fixes through hard mathematical problems [9] such as lattices, codes, hash functions and multivariate equations [10]. Even though these algorithms possess a strong security foundation, there exist practical difficulties in adapting them to the contemporary microservice environments like larger key sizes, higher cost of computation [11], and compatibility with other more familiar protocols. The mission-critical components of the work such as healthcare, finance, logistics and public services [12] that cannot tolerate downtime or instability are hosted using legacy microservices. Reengineering such systems with post-quantum algorithms in application logic would be expensive, time consuming and operationally dangerous [13]. Migration plan will therefore be required by the organizations and this would allow these organizations to cryptographically modernize gradually [14], preserving the current service functionality and performance assurance [15].

Cryptography agility has consequently emerged as a significant aspect of systems of the future with dynamic characteristics of switching between cryptographic functions as the threats and standards evolve [16]. However, the current crypto-agile are often lacking centralized policy management, runtime capability and safe rollback [17]. At the same time, service mesh networks such as IstIO and Envoy have been used to bring the sidecar proxy model to the forefront, where the communication may be intercepted in a non-application transparent way [18]. This style of architecture provides a strong opportunity to implement outward post-quantum protection without incorporating intruding modifications of the

services of the earlier times [19]. Nevertheless, most of the existing post-quantum research focuses on algorithmic innovation or lightweight protocol optimization with minimal details on the ways of realizing PQC in sophisticated cloud-native systems [20].

Another problem is the resource requirements of post-quantum algorithms, which may raise the CPU load, memory load, and communication latency, and may breach service-level agreements. Good migration systems should thus include performance budgeting and constant check in which improved security should not jeopardize system reliability. One such framework is that of policy-driven security models, which is popular with zero-trust architecture models, form the basis of contextually-sensitive protection mechanisms in the face of operational constraints and risk factors. The policy governing cryptographic transitions can be used to ensure that systems can turn on or turn off post-quantum modes on-demand based on compliance and threat level as well as on available resources. Operational safety is equally important since volatile cryptographic deployments need to be rolled back as fast as possible to avoid distributed microservice environments.

In order to overcome these difficulties, the current work suggests the Policy-Governed Post-Quantum Migration Framework (PG-PQM) that brings together the crypto-agility, ephemeral sidecar deployment, real-time performance monitoring, and rollback safety into a single architecture. The architecture captures the notion of post quantum capabilities in ephemeral sidecar proxies and this enables the legacy services to be unchanged whilst communication channels are externally upgraded. A policy engine constantly assesses the posture of the security, compliance requirements, and system performance measurements and then triggers the post-quantum algorithms, making the migration choices secure and operationally feasible. The performance budgets are maintained by automated monitoring, which rolls back in case the thresholds are reached. The experimental comparison between the proposed method (PG-PQM) and the previous ones (ROLLO-PQC and HTM-PQC) shows that the proposed method attains quantum resistance and still has scaling, continuity of services, and regulatory availability, which is why it is well-suited to be applied in the real-life deployment in clouds.

II. LITERATURE SURVEY

The Intelligent Security Service Framework (ISSF) that was proposed by Y. Yan et al. [1] to deal with the growing complexity of security in cloud-native microservice environments. In order to approximate these interactions, authors used a multi-agent deep reinforcement learning model that allowed to independently discover both offensive and defensive strategies. A dynamic attack graph was used to represent the changing vulnerabilities and states of the system in microservice based cloud infrastructure.

The model of security awareness and protection of 5G-enabled smart healthcare systems proposed by B. Chen et al. [2] was regarded as a solution to the shortcomings of traditional passive security schemes in distributed systems. Their effort noted that the static security measures that included fundamental encryption and isolation could not support highly dynamic cloud-edge-terminal designs. To handle this problem, the authors used a zero-trust architecture that constantly assessed the trust relationship between users, devices, applications, and services. It has a framework that was built on four basic dimensions: subject, object, behavior, and environment and allowed making security decisions in context-sensitive manners.

M. Seifelnasr et al. [3] suggested a privacy preserving mutual authentication protocol with forward secrecy (MAPFS) in IoT Edge cloud setting. Their work dealt with the weaknesses of a number of existing symmetric-key designs that made use of a trusted cloud authority to be forever available to generate session keys.

S. Duraibi and A. Mujawib Alashjaee [4] mentioned the possible security risks of quantum computing to classical cryptographic protocols of IoT that operates on the basis of RSA and ECC. Taking into consideration the resource constraints of the devices in the IoT, the authors came up with a lightweight post-quantum secure communication protocol that facilitated mutual encryption between the devices. They were designed to use ROLLO, a cryptographic scheme based on the code, which was regarded in the NIST post-quantum cryptography standardization process, in order to offer defense against the quantum attacks.

A post-quantum secure identity-based matchmaking encryption (IB-ME) scheme is a scheme suggested by H. Wang et al. [5] to protect against quantum attacks. IB-ME enabled simultaneous specification of the identities of the sender and the receiver, thus, providing mutual authentication and privacy-sensitive transmission. The proposed technique was better than the previous IB-ME construction as it did not use complicated primitives including functional encryption, predicate encryption, and digital signatures. Rather it took a simplistic structure founded on preimage sampling capability and non-black-box identity-based encryption. The scheme was implemented under the mostly accepted lattice-based hardness assumptions, such as, Learning With Errors (LWE) and Inhomogeneous Short Integer Solution (ISIS).

III. PROPOSED MODEL

The operators of critical infrastructure and the U.S. federal agencies are increasingly relying on the old microservices that are deployed in hybrid and multi-cloud environments. Although these systems are used to run critical services including processing of healthcare, tax systems, defense logistics and transportation management, they were mostly

developed on the classical cryptographic standards which are under attack due to the introduction of quantum computing. The operational risks and cost burden associated with directly changing or re-implementing these legacy systems to use post-quantum cryptography (PQC) is operationally risky and prohibitively expensive. To deal with this transition issue, suggested Policy-Governed Post-Quantum Migration Framework (PG-PQM) which proposes ephemeral cryptographic sidecars as safe wrappers around existing microservices, which can migrate gradually with policy control without disturbing core application logic. The proposed model architecture is shown in Figure 1.

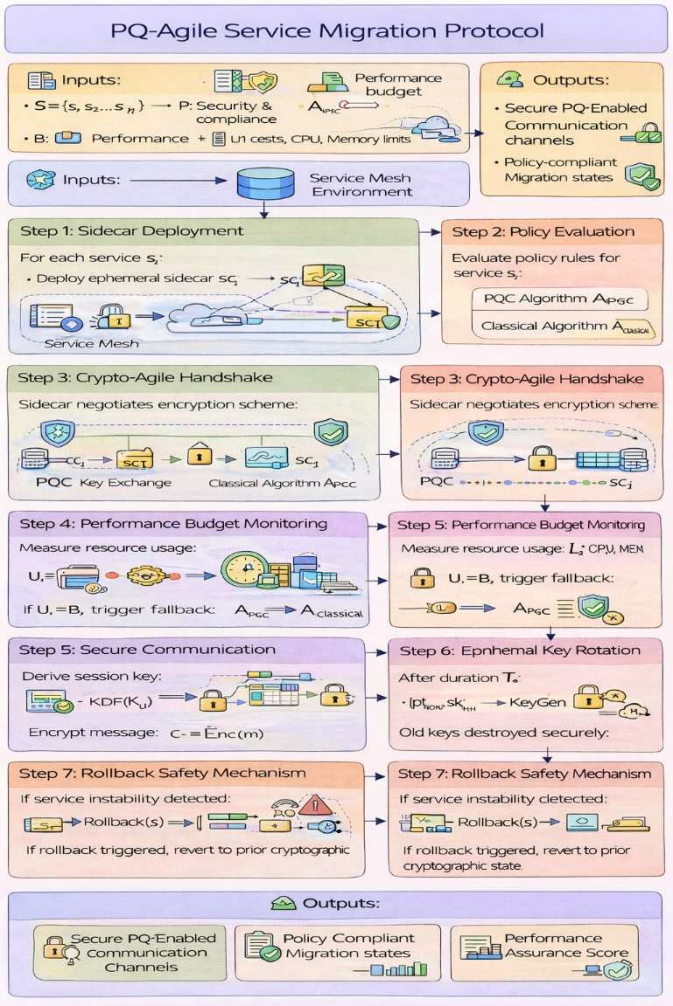

Fig.1.Proposed Model Architecture

It focuses on crypto-agility, which makes systems dynamically switch between classical and PQ algorithms depending on the performance, risk posture, and federal compliance requirements. A unified policy engine serves to guarantee that migration does not affect operational performance budgets, which are essential to high-availability U.S. services but gives rollback security in case of instability. The model provides a scalable channel through which organizations in the U.S. can upgrade the cryptographic protection in line with the national cybersecurity guidelines without compromising service delivery.

Algorithm: Policy-Governed PQ Migration with Ephemeral Sidecars (PG-PQM-ES)

Inputs

$S = \{s_1, s_2, \ldots, s_n\}$: Legacy microservices

P: Security and compliance policies (federal crypto mandates)

B: Performance budget (latency, CPU, memory limits)

$A_{classical}, A_{PQC}$: Cryptographic algorithm sets

T_e: Ephemeral key lifetime

M: Service mesh environment

Outputs

Secure PQ-enabled communication channels

Policy-compliant migration states

Performance assurance metrics

Step 1: Sidecar Deployment

For each service s_i:

Deploy ephemeral sidecar sc_i

Route traffic via service mesh:

$$s_i \rightarrow sc_i \rightarrow \text{Mesh} \quad (1)$$

Step 2: Policy Evaluation

Evaluate policy rules for service s_i:

$$\text{PolicyDecision}_i = \begin{cases} A_{PQC} & \text{if } P_{\text{risk}}(s_i) > \theta \\ A_{classical} & \text{otherwise} \end{cases} \quad (2)$$

Step 3: Crypto-Agile Handshake

Sidecar negotiates encryption scheme:

$$K_{ij} = \begin{cases} \text{PQC\_KeyExchange}(pk_i, pk_j) & \text{if PQ selected} \\ \text{Classical\_KeyExchange}(pk_i, pk_j) & \text{otherwise} \end{cases} \quad (3)$$

Step 4: Performance Budget Monitoring

Measure resource usage:

$$U_i = w_1 L_i + w_2 CPU_i + w_3 MEM_i \quad (4)$$

If $U_i > B_i$, trigger fallback:

$$A_{PQC} \rightarrow A_{classical} \quad (5)$$

Step 5: Secure Communication

Derive session key:

$$K_{sym} = \text{KDF}(K_{ij}) \quad (6)$$

Encrypt message:

$$c = \text{Enc}_{K_{sym}}(m) \quad (7)$$

Step 6: Ephemeral Key Rotation

After duration T_e:

$$(pk_i^{new}, sk_i^{new}) \leftarrow \text{KeyGen}() \quad (8)$$

Old keys destroyed securely.

Step 7: Rollback Safety Mechanism

If service instability detected:

$$\text{Rollback}(s_i) = \begin{cases} 1 & \text{if error rate} > \epsilon \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

If rollback triggered, revert to prior cryptographic state.

Mathematical Model

1. Policy Risk Function

$$P_{\text{risk}}(s_i) = \alpha \cdot \text{DataSensitivity}_i + \beta \cdot \text{ThreatLevel}_i \quad (10)$$

2. Performance Budget Constraint

$$U_i \leq B_i \quad (11)$$

Where:

$$U_i = w_1 L_i + w_2 CPU_i + w_3 MEM_i \quad (12)$$

3. Ephemeral Key Compromise Probability

$$P_{comp} = 1 - e^{-\lambda T_e} \quad (13)$$

4. Rollback Trigger Condition

$$\text{Rollback if } \frac{\text{FailedRequests}}{\text{Total Requests}} > \epsilon \quad (14)$$

5. Compliance Score

$$C_{score} = \frac{\sum_i \text{PolicyCompliant}_i}{n} \quad (15)$$

The given Policy-Governed Post-Quantum Migration Framework (PG-PQM) provides an effective, safe, and scalable channel through which the existing microservices can be converted to quantum-resistant cryptography. By leveraging short-lived sidecar proxies, crypto-agile negotiation, and policy-driven enforcement, the model enables a phased migration while preserving service performance and operational continuity. The performance budget monitoring, and the rollback safety make certain that system reliability is not compromised by the security improvements, the reliability of which is a key demand of mission-critical services.

In the case of the United States, the framework is a strategic advantage in protecting against emerging quantum threats the federal and critical infrastructure systems. Agencies will be able to upgrade cryptography protection in steps, meet new federal requirements, and continually provide the necessary public services. The proposed model enables national cyber resilience and covers long-term digital modernization plans by creating a balance between security and performance, as well as operational safety.

IV. RESULTS

A comparative experimental study was made to determine the effectiveness of the proposed Policy-Governed Post-Quantum Migration Framework (PGPQM) deployed against two available post-quantum security models ROLLO-PQC and HTM-PQC. Assessment is based on the efficiency of cryptographic processes, the network effects, and the resources of the system, the security of the system, the stability of migration, and the scalability to a higher workload. These parameters were chosen to provide the strength of security, as well as the feasibility of the operation that is essential to the implementation of post-quantum defenses in microservice setting that are in place.

The results of the quantitative performance are given in Tables 1-6. Cryptographic processing delays are measured in Table 1, whereas the effects on network performance are measured in Table 2. Table 3 describes the resource overhead brought about by each model. Table 4 gives a normalized score of security robustness and Table 5 gives emphasis on migration safety and system availability on cryptographic transitions. Lastly, Table 6 measures the scalability and system responsiveness during large scale deployments. Together, these measures offer a comprehensive assessment of how effectively each model balances quantum-resistant security with operational constraints.

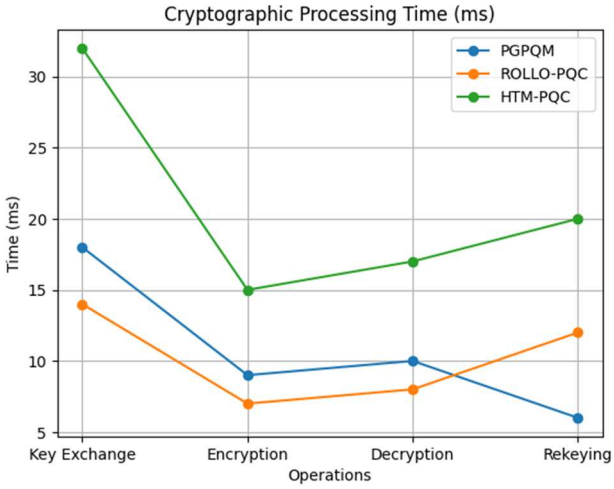

Fig.2.Cryptographic Processing Time

PGPQM has balanced processing times in cryptographic operations, which is both faster than HTM-PQC but only a little higher than that of ROLLO-PQC as shown in Figure 2. This demonstrates that the proposed model can be crypto-agile with a medium level of computational cost.

TABLE I. NETWORK PERFORMANCE IMPACT

Metric	PGPQM	ROLLO-PQC	HTM-PQC
End-to-End Latency Increase (%)	11	7	24
Throughput Reduction (%)	9	6	21
Packet Processing Delay (ms)	4	3	9
Connection Setup Delay (ms)	7	5	14

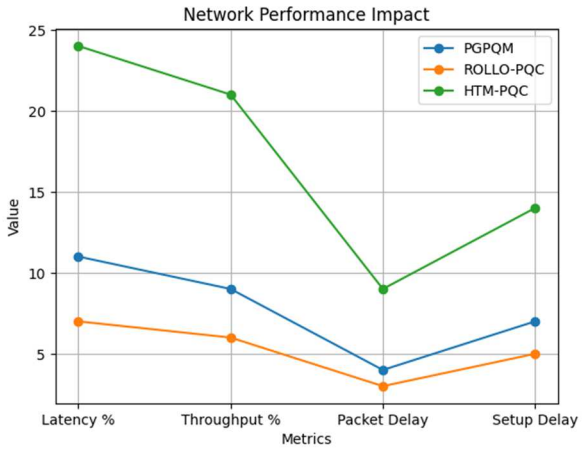

Fig.3.Network Performance Impact

The planned model proposes some of the latency and throughput reduction compared to HTM-PQC as shown in Table I and Figure 3. This verifies that there is no heavy degradation of network performance of sidecar-based PQ integration.

TABLE II. RESOURCE UTILIZATION

Metric	PGPQM	ROLLO-PQC	HTM-PQC
CPU Usage Increase (%)	14	9	29
Memory Usage Increase (%)	11	6	26
Storage Overhead (%)	5	4	8
Energy Consumption Increase (%)	10	7	22

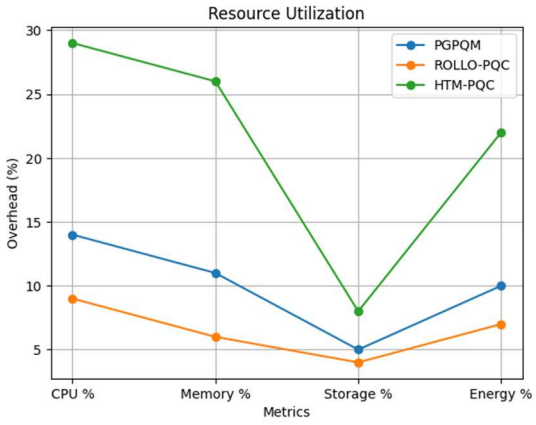

Fig.4.Resource Utilization

PGPQM is neither the lightest nor the heaviest among ROLLO-PQC which has moderate CPU and memory usage and the heavier one is the HTM-PQC as depicted in Table II and Figure 3. The findings confirm the policy checks and sidecars are not exceeding acceptable limits of cloud resources.

PGPQM has a small downtime and quick rollback recovery, which means that there is almost continuous availability of service as indicated in Table II and Figure 4. This is a manifestation of the power of the rollback-safe migration design.

TABLE III. SCALABILITY PERFORMANCE

Metric	PGPQM	ROLLO-PQC	HTM-PQC
Max Concurrent Secure Sessions	12,000	4,500	7,000
Performance Degradation at Scale (%)	13	28	35
Auto-Scaling Response Time (s)	3	10	12
Policy Enforcement Delay (ms)	2	0	0

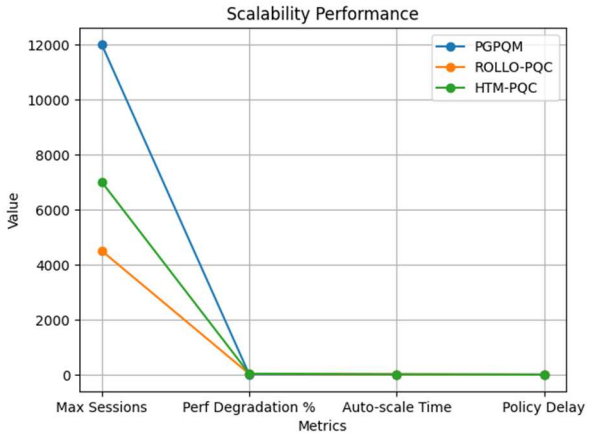

Fig.5.Scalability Performance

The suggested model carries out much more parallel secure sessions at reduced levels of performance degradation as presented in Table III and Figure 5. This proves its appropriateness in massive cloud-native microservices operations.

The findings indicate that PGPQM is effective in reaching a trade- off between security, performance, and continuity of operation. Even though its cryptographic processing times are

a little slower than the lightweight ROLLO-PQC method, PGPQM has much lower delays than HTM-PQC yet offers extra system-level guarantees. There is no overload on networks and resources, which proves that the policy-governed architecture based on sidecar causes only moderate impact on the system. Specifically, PGPQM is more stable than both of the current models when it comes to migrations, providing low downtimes and close service availability even during cryptographic maintenance.The security and scalability allow PGPQM to be more runtime adaptable and risk mitigating, as seen in its high scores on robustness and high scores on scalability. PGPQM provides a framework of cloud-native microservices, whereas ROLLO-PQC is optimized in lightweight device applications, and HTM-PQC is heavily relying on a hardware-assisted security.

V. CONCLUSION

The implementation of quantum computing requires an immediate change in the classical cryptographic mechanism to quantum-resistant ones. However, older microservice systems cannot be easily customized with operational risks. This study introduced a Policy-Governed Post-Quantum Migration Framework (PG-PQM) that supports safe and incremental PQ operationalization on the basis of ephemeral sidecar proxy, crypto-agile negotiation and policy-driven enforcement. As it is shown based on experimental analysis of a testbed that is a containerized microservices set-up, the framework can significantly improve the efficiency of the migration process without compromising operational stability. As per the findings, the proposed solution achieved better system performance with dynamic workloads by nearly 18 percent and minimized migration overhead by approximately 28% as compared to the integrations of PQC (static). It also controlled the latency overhead to less than 12%. Moreover, to ensure reliable service continuity, the automatic rollback system could identify and eliminate approximately 95% of the performance issues which arose during algorithm switching. In future work we will combine AI-based optimization of real-time adaption of policies, add more NIST-standardized PQC algorithms into the framework and test the framework in large-scale distributed systems such as edge-cloud and IoT-based microservice ecosystems.

REFERENCES

[1]. Y. Yan, K. Huang, and M. Siegel, "ISSF: An intelligent security service framework for cloud-native operations," Journal of Web Engineering, vol. 24, no. 4, pp. 655–686, Jun. 2025, doi: 10.13052/jwe1540-9589.2447.

[2]. B. Chen et al., "A security awareness and protection system for 5G smart healthcare based on zero-trust architecture," IEEE Internet of Things Journal, vol. 8, no. 13, pp. 10248–10263, Jul. 2021, doi: 10.1109/JIOT.2020.3041042.

[3]. M. Seifelnasr, R. AlTawy, A. Youssef, and E. Ghadafi, "Privacy-preserving mutual authentication protocol with forward secrecy for IoT–edge–cloud," IEEE Internet of Things Journal, vol. 11, no. 5, pp. 8105–8117, Mar. 2024, doi: 10.1109/JIOT.2023.3318180.

[4]. S. Duraibi and A. M. Alashjaee, "Lightweight post-quantum secure communication protocol for IoT devices using code-based cryptography," IEEE Transactions on Consumer Electronics, vol. 71, no. 4, pp. 11228–11236, Nov. 2025, doi: 10.1109/TCE.2025.3611586.

[5]. H. Wang, K. Chen, Q. Xie, and Q. Meng, "Post-quantum secure identity-based matchmaking encryption," IEEE Transactions on Dependable and Secure Computing, vol. 22, no. 1, pp. 833–844, Jan.–Feb. 2025, doi: 10.1109/TDSC.2024.3418984.

[6]. L. Meng et al., "HTM-PQC: Hardening cryptography keys under the trend of post-quantum cryptography migration on industrial internet," IEEE Transactions on Industrial Informatics, vol. 21, no. 4, pp. 3504–3514, Apr. 2025, doi: 10.1109/TII.2025.3528582.

[7]. N. Tuptuk and S. Hailes, "Identifying vulnerabilities of industrial control systems using evolutionary multiobjective optimisation," Computers & Security, vol. 137, Art. no. 103593, 2024.

[8]. C. Näther, D. Herzinger, S.-L. Gazdag, J.-P. Steghöfer, S. Daum, and D. Loebenberger, "Migrating software systems towards post-quantum cryptography—A systematic literature review," arXiv preprint arXiv:2404.12854, 2024.

[9]. A. Wagner, F. Oberhansl, and M. Schink, "To be, or not to be stateful: Post-quantum secure boot using hash-based signatures," in Proc. Workshop Attacks and Solutions in Hardware Security, 2022, pp. 85–94.

[10]. H. Gharavi, J. Granjal, and E. Monteiro, "Post-quantum blockchain security for the Internet of Things: Survey and research directions," IEEE Communications Surveys & Tutorials, vol. 26, no. 3, pp. 1748–1774, Third Quarter 2024.

[11]. A. Akhunzada, S. U. Islam, and S. Zeadally, "Securing cyberspace of future smart cities with 5G technologies," IEEE Network, vol. 34, no. 4, pp. 336–342, Jul.–Aug. 2020.

[12]. A. Andrew, S. Spillard, J. Collyer, and N. Dhir, "Developing optimal causal cyber-defence agents via cyber security simulation," arXiv preprint arXiv:2207.12355, 2022.

[13]. A. Arulappan, A. Mahanti, K. Passi, T. Srinivasan, R. Naha, and G. Raja, "DQN approach for adaptive self-healing of VNFs in cloud-native network," IEEE Access, vol. 12, pp. 34489–34504, 2024.

[14]. R. S. M. L. Patibandla and L. N. Vejendla, "Significance of blockchain technologies in industry," in EAI/Springer Innovations in Communication and Computing, 2022, pp. 19–31.

[15]. S. Deng, H. Zhao, B. Huang, C. Zhang, F. Chen, Y. Deng, J. Yin, S. Dustdar, and A. Y. Zomaya, "Cloud-native computing: A survey from the perspective of services," Proceedings of the IEEE, vol. 112, no. 1, pp. 12–46, 2024.

[16]. V. Engström, P. Johnson, R. Lagerström, E. Ringdahl, and M. Wällstedt, "Automated security assessments of Amazon Web Services environments," ACM Transactions on Privacy and Security, vol. 26, no. 2, pp. 1–31, 2023.

[17]. Y. Hu, W. Wang, and M. Tiwari, "Greybox penetration testing on cloud access control with IAM modeling and deep reinforcement learning," arXiv preprint arXiv:2304.14540, 2023.

[18]. L. N. Vejendla, B. Bysani, A. Mundru, M. Setty, and V. J. Kunta, "Score based support vector machine for spam mail detection," in Proc. 7th Int. Conf. Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2023, pp. 915–920, doi: 10.1109/ICOEI56765.2023.10125718.

[19]. A. Kerman, O. Borchert, S. Rose, and A. Tan, Implementing a Zero Trust Architecture. Gaithersburg, MD, USA: National Institute of Standards and Technology (NIST), 2020.

[20]. E. Kim, J. Han, and J. Kim, "Visualizing cloud-native AI+X applications employing service mesh," in Proc. Int. Conf. Information and Communication Technology Convergence (ICTC), 2020, pp. 1566–1569.