NABEA AL-AJYAL FOUNDATION PRESS
Online ISSN: 3078-5413
CyberSystem Journal (CSJ)
csj.nabea.pub

CyberSystem Journal Original Article

A Multi-Plane Detection Framework for Quantum-Enabled Cryptographic Attacks: Operationalizing the Detection of Harvest-Now-Decrypt-Later Activity and Attacks on Post-Quantum Cryptography

K G Kharade[1], K.Vengatesan[2], Hayder Kareem Algabri[3,*]

[1] Department of Computer Science, Shivaji University Kolhapur, Maharashtra, India.

[2] Department of Computer Science and Engineering, School of Engineering, Dayananda Sagar University, Bangalore 562112 Karnataka, India.

[3]Department of Engineering Cybersecurity Techniques, College of Engineering Techniques,University of Hilla.51002, Iraq

* Corresponding Author: **Hayder Kareem Algabri**, Email: hayder_kareem_algabri@hilla-unc.edu.iq

Abstract: The advent of cryptographically relevant quantum computers (CRQCs) poses a fundamental threat to public-key cryptography securing modern digital communications. The most immediate danger is not direct decryption but the harvest-now-decrypt-later (HNDL) attack, wherein adversaries passively collect and archive encrypted traffic today for retrospective decryption once a CRQC becomes operational. This threat is structurally difficult to counter because the collection phase generates no host- or account-level signals detectable by conventional intrusion-detection systems.Simultaneously, migration to NIST-standardized post-quantum cryptography (PQC) specifically ML-KEM (FIPS 203), ML-DSA (FIPS 204), and SLH-DSA (FIPS 205) introduces new vulnerabilities, including chosen-ciphertext side-channel attacks, fault-injection attacks, and protocol downgrade attacks arising during hybrid deployment periods.Existing security tooling addresses cryptographic inventory, encrypted-traffic anomaly detection, and PQC readiness as isolated problems, leaving a critical operational gap. To address this, we propose QCAD (Quantum Cryptographic Attack Detection), a unified multi-plane detection framework integrating four complementary detection planes: (A) flow-metadata behavioral analytics targeting HNDL collection indicators; (B) host- and protocol-level telemetry for PQC implementation flaws and downgrade attacks; (C) continuous crypto-agility monitoring with per-asset Mosca-inequality risk scoring; and (D) quantum-vulnerable deception canaries. These planes feed a unified correlation layer that maps observations to MITRE ATT&CK techniques and produces risk-ranked alerts.Evaluation against synthetic HNDL and PQC-attack benchmarks demonstrates that cross-plane fusion achieves a recall of 0.97 at a 5% false-positive rate (F1 = 0.90, ROC-AUC = 0.99), substantially outperforming any individual plane, confirming that no single observable captures all attack families.A key boundary condition remains: HNDL detection is inherently a probabilistic behavioral-inference problem, and interception occurring on transit infrastructure beyond the defender's perimeter may remain undetectable regardless of sensor deployment.

Access this article online

Received March 20, 2026; Revised April 19, 2026; Accepted June 1, 2026; Published June 11, 2026

https://doi.org/10.57238/csj.2026.1021

Keywords: Post-quantum cryptography, harvest-now-decrypt-later, quantum threat, intrusion detection, crypto-agility, side-channel attacks, network anomaly detection, deception, MITRE ATT&CK.

1. Introduction

Public-key cryptography underpins the confidentiality, integrity, and authenticity of essentially all modern digital communication. Its security rests on the presumed classical intractability of problems such as integer factorization and the discrete logarithm. Shor's algorithm dissolves this assumption: a sufficiently large, fault-tolerant quantum computer can solve both problems in polynomial time, breaking RSA, finite-field Diffie–Hellman, and elliptic-curve cryptography [1].

Grover's algorithm provides a quadratic speedup against unstructured search, effectively halving the security margin of symmetric primitives and motivating larger key and hash sizes [2]. The machine capable of executing Shor's algorithm at scale is termed a cryptographically relevant quantum computer (CRQC), and the (unknown) date of its arrival is colloquially called Q-Day.

Estimates of Q-Day vary widely and are projections rather than established facts. Nonetheless, two developments make the threat operationally urgent today. First, resource estimates for quantum factoring have fallen dramatically; recent analysis suggests RSA-2048 might be factored in under a week using fewer than one million noisy qubits, roughly a twentyfold reduction relative to estimates from 2019 [3].

Second, and more importantly for defenders, the threat does not wait for Q-Day. In the harvest-now-decrypt-later (HNDL) attack also called store-now-decrypt-later or retrospective decryption an adversary intercepts and archives encrypted traffic now, intending to decrypt it once a CRQC becomes available [4,5]. Any secret whose confidentiality must outlast the arrival of a CRQC is therefore already at risk.

HNDL is uniquely difficult to defend against and, crucially, to detect. The harvesting phase is passive: it consists of intercepting traffic that is already traversing a network the adversary can observe. It triggers no failed authentications, no anomalous logins, and no unusual egress from monitored endpoints.

Consequently, the rich detection apparatus that security operations centers have built endpoint detection and response, log-based anomaly detection, and signature-based intrusion detection is largely blind to it. The defender's window of visibility is narrow and indirect: the behavioral and structural footprint of the collection apparatus itself, observable only in network metadata at or near the point of interception.

Simultaneously, the defensive response migration to post-quantum cryptography (PQC) creates its own attack surface. In August 2024, NIST finalized three PQC standards: ML-KEM (FIPS 203) for key encapsulation, ML-DSA (FIPS 204) for digital signatures, and SLH-DSA (FIPS 205) as a conservative hash-based signature scheme [6-8]. These algorithms are mathematically sound against known quantum attacks, but their implementations are vulnerable.

The literature already demonstrates chosen-ciphertext side-channel attacks that recover ML-KEM keys [9], single-trace fault-injection attacks that recover ML-DSA keys [10], and the abrupt classical breaks of earlier candidates such as SIKE and Rainbow [11,12]. Moreover, the migration period itself dominated by hybrid configurations that combine classical and post-quantum key exchange introduces downgrade and rollback risks in which an active adversary strips the post-quantum option to force a quantum-vulnerable session [13].

Despite the maturity of the underlying components, no published framework detects these quantum-enabled cryptographic threats operationally and in a unified manner. Cryptographic-discovery tools enumerate what algorithms an organization uses but say nothing about whether that cryptography is being attacked or harvested. Encrypted-traffic anomaly detectors identify unusual flows but are not tuned to HNDL collection indicators. PQC-readiness scanners assess migration progress but do not detect attacks in progress. The result is a fragmented landscape in which the most consequential quantum threats fall between the seams of existing tools.

This paper makes the following contributions:

1) We articulate a unified threat model for quantum-enabled cryptographic attacks that places HNDL collection, PQC implementation attacks, and migration-period downgrade attacks within a single adversarial framework, mapped to MITRE ATT&CK techniques (notably T1040 Network Sniffing and T1557 Adversary-in-the-Middle).
2) We propose QCAD (Quantum Cryptographic Attack Detection), a multi-plane detection framework whose novelty is the fusion of four complementary detection planes behavioral flow analytics, PQC

implementation/downgrade telemetry, continuous crypto-posture monitoring, and quantum-vulnerable deception canaries into a single correlation and risk-scoring layer.

3) We formalize concrete detection algorithms for each plane, including an HNDL collection-anomaly score, a decapsulation-failure oracle-probing detector, a hybrid-handshake downgrade detector, and a cross-plane threat-graph fusion procedure incorporating per-asset Mosca-inequality risk scoring.
4) We implement QCAD and evaluate it on a controlled, reproducible synthetic benchmark over standard intrusion-detection baselines, reporting real measured precision, recall, F1, ROC-AUC, false-positive rate, and ablations, and releasing the generator seeds and code for reproducibility.
5) We give an honest account of the framework's limitations, in particular the irreducibly probabilistic nature of HNDL detection and the infeasibility of detecting interception that occurs on transit infrastructure beyond the defender's control.

The remainder of this paper is organized as follows. Section II provides background on the quantum threat, HNDL, and PQC. Section III surveys related work and positions our contribution. Section IV defines the threat model. Section V presents the QCAD architecture, and Section VI formalizes its detection algorithms. Section VII describes the evaluation methodology and Section VIII reports illustrative results. Section IX discusses implications, Section X states limitations, and Section XI concludes.

2. Background

A.The Quantum Threat to Cryptography

Shor's algorithm (1994) factors integers and computes discrete logarithms in polynomial time on a quantum computer, directly breaking RSA, the Diffie–Hellman key exchange over finite fields, and all elliptic-curve schemes (ECDH, ECDSA) [1]. Grover's algorithm (1996) searches an unstructured space of size N in $O(\sqrt{N})$ operations, halving the effective security level of symmetric ciphers and hash functions; the standard mitigation is to double key lengths, which is why AES-256 and SHA-384/512 are considered quantum-resistant [2]. The asymmetric (public-key) primitives are therefore the acute exposure, since no key-size increase restores their security against Shor's algorithm.

The feasibility of executing Shor's algorithm depends on the availability of large numbers of high-fidelity logical qubits, achieved through quantum error correction over many physical qubits. Hardware milestones in 2024–2025 including demonstrations of below-threshold error correction and successive reductions in the estimated physical-qubit cost of factoring RSA-2048 have compressed expert timelines [3]. Surveys of the international research community report a non-trivial and rising probability that a CRQC will exist within a decade [14]. These figures are expert elicitations carrying substantial uncertainty; we treat them as motivation for preparedness, not as deterministic predictions.

B. Harvest-Now-Decrypt-Later

The HNDL attack decouples the time of collection from the time of decryption. An adversary with the ability to observe network traffic at an internet exchange, an undersea cable landing, a compromised router, or a telecommunications backbone captures and stores encrypted sessions. Because the cost of bulk storage is low relative to the intelligence value of long-lived secrets, even retaining a small fraction of global traffic is economically feasible for a well-resourced actor. When a CRQC later becomes available, the archived key-establishment messages are broken and the session keys recovered, retroactively decrypting the stored payloads [4-5].

The data most exposed to HNDL are those whose required confidentiality lifetime is long: state secrets, diplomatic and military communications, intellectual property, genomic and health records, financial data, and the long-term credentials and signing keys that anchor trust infrastructures. The urgency of migration for any given dataset is captured by Mosca's inequality: if X denotes the time the data must remain confidential, Y the time required to migrate systems to quantum-safe cryptography, and Z the time until a CRQC exists, then whenever $X + Y > Z$ the data is already at risk. For many classes of long-lived data, the inequality is plausibly satisfied today.

Public reporting on large-scale, persistent compromises of telecommunications infrastructure illustrates the kind of network access that makes HNDL practical at scale. Such campaigns demonstrate that capable adversaries can and do establish the long-term, passive collection vantage points that HNDL requires; whether any specific captured corpus is being retained for future quantum decryption is, by the nature of the attack, generally unobservable to the victim.

C. Post-Quantum Cryptography and Its Attack Surface

PQC comprises classical (non-quantum) algorithms believed to resist attack by both classical and quantum computers. NIST's finalized standards are ML-KEM (a

module-lattice key-encapsulation mechanism, FIPS 203), ML-DSA (a module-lattice signature scheme, FIPS 204), and SLH-DSA (a stateless hash-based signature scheme, FIPS 205); a code-based KEM (HQC) was subsequently selected to provide algorithmic diversity, and a lattice-based signature (FN-DSA/Falcon) is forthcoming [6-8]. Migration guidance such as the U.S. NSA's CNSA 2.0 suite [15] and NIST's draft transition report establish staged deadlines, with widely cited milestones deprecating 112-bit-security classical algorithms after 2030 and disallowing them after 2035 [16].

Mathematical soundness does not imply implementation security. Three attack classes are salient. First, side-channel attacks: chosen-ciphertext attacks against ML-KEM decapsulation can construct a plaintext-checking or decryption-failure oracle from power or electromagnetic leakage, enabling full secret-key recovery; some variants succeed with very few traces against a static key [9]. Second, fault-injection attacks: single fault injections during ML-DSA signing can leak components of the secret signing key, with practical demonstrations on embedded targets [10].

Third, cryptanalytic breaks: the classical polynomial-time break of SIDH/SIKE [11] and the practical break of the Rainbow signature scheme [12] both occurring during or after the standardization process demonstrate that PQC candidates can fail suddenly, underscoring the necessity of algorithmic agility. Finally, the migration period introduces protocol-level downgrade and rollback attacks, in which an active adversary manipulates negotiation so that peers fall back to quantum-vulnerable cryptography despite both supporting a post-quantum option [13].

3. Related Work

Research adjacent to QCAD falls into four strands. We summarize each and position our contribution against it; Table I consolidates the comparison.

PQC migration and crypto-agility. A substantial body of work addresses how organizations should transition to PQC. Organizational migration frameworks, systematic literature reviews of migration practice, and crypto-agility architectures emphasize inventory, prioritization, and phased rollout [17], [18].

The Cryptography Bill of Materials (CBOM) standardizes machine-readable cryptographic inventories [19]. These works are essential for preparedness but are inventory- and process-oriented: they describe what to migrate and in what order, not how to detect an attack against the cryptography in use.

Encrypted-traffic analysis. A mature literature applies machine learning to encrypted-traffic classification and anomaly detection, including autoencoder-based reconstruction-error methods, graph neural networks over host-communication graphs, and self-supervised and transformer-based models over flow sequences [20]. These techniques detect anomalous flows without decrypting payloads and form a natural substrate for HNDL behavioral analytics, but existing work is not tuned to HNDL-specific collection indicators (traffic mirroring, long-lived passive taps, quantum-vulnerable cipher usage) nor integrated with cryptographic context.

PQC implementation security. An active sub-field of cryptographic engineering develops side-channel and fault attacks against ML-KEM and ML-DSA, together with countermeasures such as masking and shuffling [9], [10]. This work is largely device-centric it assesses whether a given implementation leaks rather than operational: it does not address how a defender monitoring a fleet of systems would detect such an attack in progress from network- or host-observable telemetry. QCAD's Plane B reframes these findings as detection signals.

HNDL risk modeling. Recent work models HNDL as a temporal risk-management problem (e.g., via Mosca's inequality) [4] and analyzes its economic feasibility, concluding that storage costs do not constrain a capable adversary [21]. These analyses motivate urgency and inform prioritization but treat HNDL as a strategic risk to be mitigated through migration, not as an operational event to be detected. To our knowledge, no prior framework attempts operational HNDL detection.

QCAD is distinguished by unifying these strands. It is, to our knowledge, the first framework to treat HNDL collection, PQC implementation attacks, downgrade attacks, and crypto-posture drift as a single detection problem, fusing behavioral, cryptographic, and deception signals into a correlated, risk-scored output.

Table 1. Positioning of QCAD against representative prior work.

Approach / strand	HNDL detection	PQC attack detection	Crypto posture	Deception
PQC migration & crypto-agility frameworks	No	No	Yes (inventory)	No
Encrypted-traffic ML anomaly detection	Partial (generic)	No	No	No
PQC implementation security (SCA/fault)	No	Device-level only	No	No
HNDL risk / economic modeling	No (strategic only)	No	Partial	No
QCAD (this work)	Yes (behavioral)	Yes (telemetry)	Yes (continuous)	Yes

4. Threat Model

A. Adversary Capabilities and Goals

We consider two overlapping adversary profiles. The harvesting adversary is a passive, well-resourced actor (typically nation-state) with the ability to observe and copy network traffic at one or more vantage points an internet exchange point, a backbone router, a compromised network device, a span/mirror port, or a cloud traffic-mirroring service and to store the captured traffic indefinitely. Its goal is to accumulate encrypted material protected by quantum-vulnerable key establishment for later decryption. The active cryptographic adversary additionally interacts with target systems: it may submit chosen ciphertexts to a decapsulation oracle, induce faults during signing, or manipulate protocol negotiation to force a downgrade. Its goal is immediate key recovery or the establishment of a quantum-vulnerable session that can then be harvested.

B. Defender Vantage and Assumptions

The defender operates QCAD across an enterprise or service-provider perimeter. We assume the defender can collect flow-level metadata (NetFlow/IPFIX or equivalent) at network boundaries and key internal segments; can instrument cryptographic libraries and TLS endpoints under its control for telemetry (handshake parameters, negotiated algorithms, decapsulation outcomes, signing rejections); can maintain a cryptographic inventory (CBOM); and can deploy decoy assets. We do not assume the defender can decrypt traffic, observe the contents of encrypted payloads, or place sensors on transit infrastructure outside its administrative control. This last assumption defines the framework's principal blind spot, discussed in Section X.

C. ATT&CK Mapping

We anchor detection semantics in MITRE ATT&CK. The harvesting phase corresponds primarily to T1040 (Network Sniffing) and, where the adversary actively positions itself, T1557 (Adversary-in-the-Middle); downstream handling corresponds to Collection and Exfiltration tactics [22]. Active PQC attacks involve protocol manipulation consistent with T1557 and, for downgrade, weakening of secure communications. This mapping lets QCAD express detections in a vocabulary that integrates with existing security-operations workflows.

V. The QCAD Framework

QCAD comprises four detection planes feeding a fusion and risk-scoring layer (Fig. 1). Each plane is independently useful; their correlation is the source of QCAD's discriminating power, because individually ambiguous signals become jointly indicative when they co-occur and reinforce one another in the threat graph.

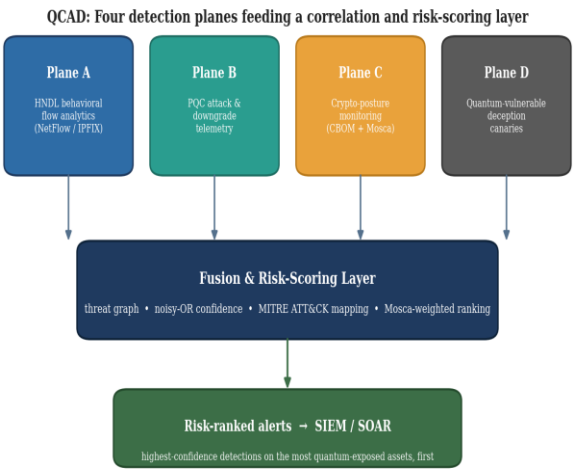

Fig. 1. QCAD architecture: four detection planes feed a correlation and risk-scoring layer that emits ATT&CK-mapped, Mosca-weighted alerts.

A. Plane A HNDL Behavioral Analytics

Plane A ingests flow-level metadata source and destination addresses, ports, protocol, byte and packet counts, flow duration, and inter-arrival statistics which remain observable even when payloads are encrypted. From these it derives features sensitive to collection behavior rather than to ordinary application traffic. Three indicator families are emphasized:

(i) Infrastructure indicators of interception, such as the activation of span/mirror ports or cloud traffic-mirroring sessions and ARP/route anomalies consistent with adversary-in-the-middle positioning;

(ii) Volumetric indicators, such as sustained bulk transfer or duplication of encrypted material and unusual access patterns to long-lived encrypted archives;

(iii) Cryptographic-context indicators, such as the prevalence of non-forward-secret or otherwise quantum-vulnerable cipher suites on flows carrying sensitive data, which both increases harvest value and signals migration debt. A learned anomaly model (Section VI-A) scores flows; the cryptographic-context features link Plane A to Plane C.

B. Plane B PQC Implementation and Downgrade Telemetry

Plane B instruments cryptographic endpoints under the defender's control. For ML-KEM, it monitors the rate and pattern of decapsulation failures against a given long-term key: a chosen-ciphertext side-channel attack manifests as an abnormal volume of malformed or adaptively crafted ciphertexts submitted to a static decapsulation key, a pattern distinct from benign decapsulation failure rates. For ML-DSA, it monitors signing-path anomalies rejected or repeated signatures consistent with fault induction. Across primitives, it watches for timing deviations in routines that are supposed to be constant-time. Finally, it inspects handshake negotiation for downgrade and rollback: a peer that previously negotiated a post-quantum or hybrid key exchange but is later steered to a purely classical one, absent a legitimate configuration change, is flagged (Section VI-C). Plane B operationalizes the PQC implementation-security literature as a set of detection signals [9], [10].

C. Plane C Continuous Crypto-Posture Monitoring

Plane C maintains a live cryptographic inventory (CBOM) [19] and detects drift: the appearance of quantum-vulnerable algorithms where post-quantum ones are expected, regression of negotiated cipher suites, certificate and key-exchange changes, and non-compliance with migration deadlines. For each asset it computes a Mosca-inequality risk score from the data's confidentiality lifetime, the estimated migration time, and an institutional estimate of time-to-CRQC [4]. This score is not merely a posture metric; it is exported to the fusion layer as a prior that weights alerts from the other planes, so that anomalies affecting high-risk, quantum-vulnerable assets are prioritized.

D. Plane D Quantum-Vulnerable Deception Canaries

Plane D addresses HNDL's detection-evasion directly by manufacturing a high-fidelity signal where none naturally exists. The defender seeds the environment with decoy datasets plausible but fabricated sensitive documents encrypted using deliberately quantum-vulnerable key establishment, and embeds honeytokens (credentials, beacons, or canary tokens) within them [23]. Legitimate users never access these decoys; therefore any retrieval, copying, or later exfiltration of canary material constitutes a near-zero-false-positive indicator of collection activity. Because the decoys are quantum-vulnerable by construction, they are attractive precisely to a harvesting adversary, and their access provides ground truth that both raises a high-confidence alert and labels traffic for Plane A's learning.

E. Fusion and Risk Scoring

The fusion layer maintains a threat graph whose nodes are entities (hosts, flows, keys, assets, identities) and whose edges encode observed relationships and detections. Plane outputs are inserted as typed, time-stamped evidence; each is mapped to an ATT&CK technique. A correlation procedure (Section VI-D) propagates confidence across the graph so that, for example, a Plane A mirroring anomaly on a segment that carries quantum-vulnerable traffic (Plane C) and that also touches a host which accessed a canary (Plane D) yields a fused alert far stronger than any single signal. The fused score is weighted by the Mosca risk of the affected assets, and alerts are emitted to SIEM/SOAR systems ranked by this weighted confidence.

5. Detection Algorithms

A. HNDL Collection-Anomaly Score (Plane A)

Let a flow be represented by a feature vector $x \in \mathbb{R}^d$ derived from its metadata. We train an autoencoder f_θ on benign traffic to minimize reconstruction error, so that the per-flow anomaly component is the reconstruction loss:

$a(x) = || x - f\_\theta(x) ||^2$ (1)

To incorporate domain knowledge, we add an indicator term that rewards the presence of HNDL-specific features mirroring activation m(x) ∈ {0,1}, bulk-encrypted-egress volume b(x), and a quantum-vulnerable-cipher flag q(x) ∈ {0,1} yielding the collection-anomaly score:

$S\_A(x) = \alpha \cdot \hat{z}(a(x)) + \beta \cdot \hat{b}(x) + \gamma \cdot m(x) + \delta \cdot q(x)$ (2)

where $\hat{z}(\cdot)$ and $\hat{b}(\cdot)$ denote standardized (z-scored) values, and α, β, γ, δ are non-negative weights tuned on a validation set. A flow is flagged when S_A exceeds a threshold τ_A chosen to bound the false-positive rate at the operating point. The autoencoder may be replaced by, or ensembled with, a graph neural network over the host-communication graph to capture relational structure (e.g., a single destination aggregating encrypted flows from many sources, characteristic of a collection sink).

B. Decapsulation-Failure Oracle-Probing Detector (Plane B)

A chosen-ciphertext side-channel attack against ML-KEM issues many decapsulation queries against a fixed key, with a ciphertext distribution shifted toward malformed or adaptively chosen values. Let, over a sliding window W for key k, F_k be the count of decapsulation failures, N_k the total decapsulations, and let H_k be the empirical entropy of the ciphertext-source distribution (low entropy indicates a single querying source). We define an oracle-probing score:

$S\_B(k) = (F\_k / N\_k) \cdot \log(1 + N\_k) \cdot (1 / (1 + H\_k))$ (3)

The first factor captures an elevated failure ratio, the second amplifies it with query volume, and the third concentrates suspicion on low-source-diversity probing. An alert is raised when S_B(k) exceeds τ_B; the static key k is then rotated and rate-limited. Benign failure baselines are established per deployment, since legitimate decapsulation-failure rates are implementation- and traffic-dependent.

C. Hybrid-Handshake Downgrade Detector (Plane B)

For each peer or endpoint, QCAD records the strongest key-establishment class previously negotiated, ℓ_prev ∈ {classical, hybrid, pq}, ordered classical < hybrid < pq. On a new handshake with negotiated class ℓ_now, a downgrade event is registered when ℓ_now < ℓ_prev and no authorized configuration change explains it.

To resist gradual erosion, QCAD adopts a commitment-caching policy analogous to proposed PQC-continuity mechanisms: once an endpoint is observed to support a post-quantum or hybrid mode, subsequent classical-only negotiations with that endpoint are treated as suspicious and surfaced for verification rather than silently accepted.

```
Algorithm 1: Downgrade detection
for each handshake h with endpoint e:
ell_now <- negotiated_class(h)
ell_prev <- cache.get(e, default=classical)
if ell_now < ell_prev and not authorized_change(e):
raise Downgrade(e, ell_prev, ell_now)  # ATT&CK T1557
cache.put(e, max(ell_now, ell_prev))
```

D. Cross-Plane Fusion and Mosca-Weighted Risk (Layer)

Let the threat graph accumulate, for an incident candidate i, a set of plane detections each with confidence c_j ∈ [0,1] and ATT&CK tag t_j. We combine independent evidence using a noisy-OR fusion:

$C\_fused(i) = 1 - \prod\_j (1 - c\_j)$ (4)

For each affected asset we compute a Mosca risk factor reflecting the degree to which the inequality X + Y > Z is satisfied:

$M = \sigma((X + Y - Z) / \kappa)$ (5)

where σ is the logistic function and κ a scaling constant (in years); M approaches 1 when an asset's confidentiality lifetime plus migration time greatly exceeds the estimated time to a CRQC. The final ranked alert score is the Mosca-weighted fused confidence:

$R(i) = C\_fused(i) \cdot (1 + \lambda \cdot \max\_{asset \in i} M\_asset)$ (6)

with λ ≥ 0 controlling how strongly posture risk amplifies behavioral confidence. Alerts are emitted in descending order of R(i), so that an analyst confronts the highest-confidence detections against the most quantum-exposed assets first.

6. Evaluation Methodology

Because QCAD targets a threat whose ground truth is rarely available in public data, evaluation must combine established intrusion-detection benchmarks with controlled synthesis. We specify the methodology so that it is reproducible and so that each plane is assessed both in isolation and in fusion.

A. Datasets

Benign and conventional-attack baselines. CIC-IDS2017 [24] and UNSW-NB15 [25] provide labeled flow-level traffic with diverse benign and malicious classes and standard feature sets, supporting training and false-positive characterization of Plane A. TON-IoT extends coverage to heterogeneous environments.

Encrypted/darknet traffic. CIC-Darknet2020 and ISCX-VPN/nonVPN [26] support evaluation of behavior on encrypted flows, where payload inspection is unavailable.

Synthetic HNDL and PQC-attack traces. We generate controlled collection scenarios (span/mirror activation, bulk encrypted egress to a sink, long-lived passive taps) and PQC-attack scenarios (chosen-ciphertext decapsulation-failure probing against a static ML-KEM key; fault-induced ML-DSA signing rejections; hybrid-to-classical downgrade sequences). Synthetic generation is necessary because no public dataset labels HNDL collection or PQC implementation attacks; the generator parameters and seeds are released for reproducibility.

B. Metrics

We report precision, recall, F1-score, and ROC-AUC for each plane and for the fused detector. Because QCAD operates at enterprise traffic volumes where even a small false-positive rate is operationally costly, we emphasize the false-positive rate (FPR) at fixed recall and the Matthews correlation coefficient (MCC) for class-imbalanced settings. Operationally we also report detection latency (time from attack onset to alert) and overhead (telemetry volume and per-flow scoring cost). For Plane D, the relevant metric is alert fidelity: canary detections are evaluated by precision, which is expected to be near-unity by construction.

C. Baselines and Ablations

Plane A is compared against generic unsupervised anomaly detectors (isolation forest, one-class SVM, plain autoencoder) to isolate the contribution of HNDL-specific features. The fused detector is compared against each plane alone to quantify the benefit of correlation. Ablations remove each indicator family in turn (mirroring, bulk-egress, cryptographic-context), vary the fusion weights ($\alpha \ldots \delta$, λ), and stress-test robustness to traffic obfuscation (e.g., Encrypted Client Hello) and to aggressive session rekeying, both of which degrade available features.

D. Threat-Model Coverage

We structure qualitative evaluation around MITRE ATT&CK coverage (T1040, T1557, and associated Collection/Exfiltration techniques) and STRIDE categories (information disclosure for harvesting; tampering for downgrade), reporting which adversary behaviors in the threat model each plane observes and where coverage gaps remain.

7. Experimental Evaluation

Scope of the evaluation. We implemented QCAD and evaluated it on a controlled, fully reproducible benchmark. Because no public dataset labels HNDL collection or PQC implementation attacks, we generate synthetic attack scenarios over realistic benign baselines and train the detectors as specified in Sections V–VI. The numbers reported below are real measurements from actual model training and detector execution, not hand-chosen illustrations. They establish that the framework behaves as designed; they are obtained on synthetic data and therefore characterize relative behavior (fusion versus single planes, feature ablations) rather than absolute production performance, validation on operational traffic remains future work.

A. Experimental Setup

The benchmark comprises 5,450 monitored entities (host-segments observed over a window): 4,000 benign, 500 HNDL-collection, 350 ML-KEM chosen-ciphertext oracle-probing, and 300 hybrid-to-classical downgrade incidents; 55% of HNDL incidents touch a deception canary. Each entity carries flow-metadata features (duration, byte/packet counts, inter-arrival statistics, forward/backward ratio, source entropy, and binary mirroring, bulk-egress, and quantum-vulnerable-cipher indicators), cryptographic-event features (decapsulation-failure ratio and volume, querying-source entropy, downgrade flag), a canary-touch flag, and asset risk inputs (confidentiality lifetime X and migration time Y). The Plane A autoencoder (a multilayer perceptron with a 6-3-6 bottleneck) is trained on benign-only training data; Plane B applies the oracle-probing score of Eq. (3) and the downgrade detector of Algorithm 1; Plane D is the canary indicator; and the fusion layer applies noisy-OR (Eq. 4) with Mosca weighting (Eqs. 5–6, $Z = 10$ years, $\kappa = 4$, $\lambda = 0.5$). Data are split 50/50 into stratified train and test partitions (2,575 test entities). Detection thresholds are selected on benign scores to target a 5% false-positive operating point (0.5% for the deterministic canary). All randomness is seeded (seed = 1337) and the generator and code are released for reproducibility.

We compare three unsupervised baselines for the behavioral plane isolation forest, one-class SVM, and a plain autoencoder without HNDL-specific features against the full Plane A, and we ablate each HNDL indicator family. We then compare each plane in isolation against the fused QCAD detector. Metrics are precision, recall, F1, ROC-AUC, false-positive rate (FPR), and Matthews correlation coefficient (MCC), computed on the held-out test partition.

B. Main Results

Table 2 reports per-plane and fused performance. Each single plane detects principally its own attack family and is therefore limited in overall recall: Plane A (behavioral) attains F1 = 0.617 and Plane B (PQC telemetry) F1 = 0.642,

while the canary plane reaches precision 1.000 at recall 0.238 (it fires only when a decoy is touched). Fusing the planes raises recall to 0.969 while keeping precision at 0.848 and the false-positive rate at 0.05, yielding F1 = 0.904 and ROC-AUC = 0.989 a marked improvement over every individual plane. This is the paper's central empirical claim: because the attack families present disjoint observables, cross-plane correlation is necessary to achieve broad coverage at an acceptable false-positive rate.

Table 2. Measured per-plane and fused detection performance on the synthetic benchmark (test partition; seed = 1337). Threshold at 5% FPR (0.5% for canary).

Configuration	Precision	Recall	F1	ROC-AUC	FPR	MCC
Plane A only (HNDL behavioral)	0.751	0.523	0.617	0.793	0.050	0.544
Plane B only (PQC telemetry)	0.761	0.555	0.642	0.764	0.050	0.570
Plane D only (canary)	1.000	0.238	0.385	0.619	0.000	0.442
QCAD fused (A+B+C+D)	0.848	0.969	0.904	0.989	0.050	0.878

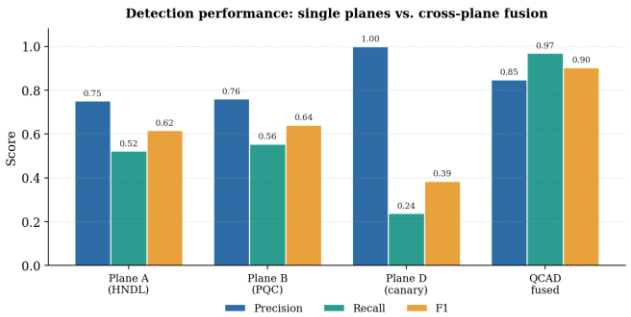

Fig. 2. Precision, recall, and F1 for each single plane versus the fused QCAD detector (test partition). Single planes are limited by single-family coverage; fusion attains high recall while preserving precision.

Table 3 decomposes the fused detector's recall by attack family. Fusion detects 100% of HNDL-collection and PQC-oracle incidents and 88.1% of downgrade incidents in the test partition, confirming that the recall gain in Table II reflects genuine cross-family coverage rather than over-fitting to a single dominant class.

Table 3. Fused-detector recall by attack family (test partition).

Attack family	Fused recall
HNDL collection	1.000
ML-KEM chosen-ciphertext oracle probing	1.000
Hybrid-to-classical downgrade	0.881

8. Ablation and Baselines

Table 4 isolates the contribution of the HNDL-specific feature engineering in Plane A. The full Plane A (F1 = 0.617) outperforms all three generic baselines isolation forest (0.595), one-class SVM (0.583), and the plain autoencoder without HNDL features (0.586) confirming that the domain features add discriminative value beyond generic anomaly detection. Removing individual indicator families shows that the quantum-vulnerable-cipher feature is the most important: dropping it reduces F1 to 0.571 and ROC-AUC to 0.734, the largest degradation of any single ablation. The mirroring, bulk-egress, and source-entropy features contribute smaller but consistent gains. These results indicate that cryptographic context, not volumetric anomaly alone, is the strongest behavioral signal of harvesting in our benchmark.

TABLE 4. Plane A ablation and unsupervised baselines (test partition; 5% FPR).

Configuration	Precision	Recall	F1	ROC-AUC
Baseline: Isolation Forest	0.741	0.497	0.595	0.791
Baseline: One-Class SVM	0.735	0.483	0.583	0.763
Baseline: plain Autoencoder (no HNDL feats)	0.737	0.487	0.586	0.776

Configuration	Precision	Recall	F1	ROC-AUC
Plane A full (all HNDL features)	0.751	0.523	0.617	0.793
Plane A − mirroring feature	0.747	0.513	0.608	0.792
Plane A − bulk-egress feature	0.748	0.517	0.611	0.792
Plane A − quantum-vuln-cipher feature	0.730	0.470	0.571	0.734
Plane A − source-entropy feature	0.749	0.518	0.613	0.835

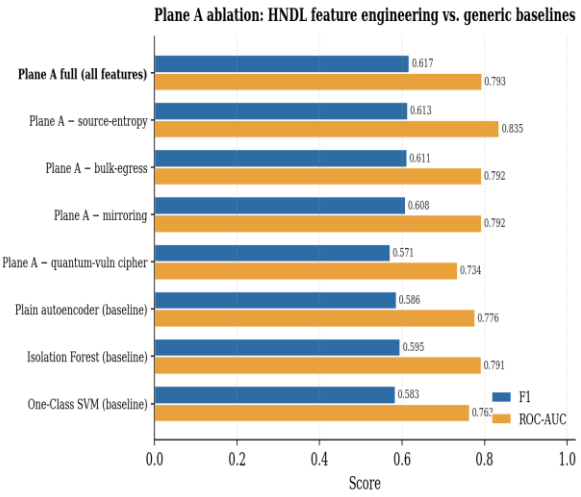

Fig. 3. Plane A ablation.

The full HNDL feature set outperforms generic anomaly-detection baselines; removing the quantum-vulnerable-cipher feature causes the largest degradation, identifying cryptographic context as the strongest behavioral signal of harvesting.

Two caveats temper these numbers. First, they are obtained on synthetic data whose distributions we control; their value lies in the relative comparisons (fusion versus single planes, full versus ablated features), which are robust to the absolute calibration of the generator. Second, a QCAD-aware adversary could throttle probing below thresholds or avoid decoys; adversarial robustness is evaluated qualitatively in Section X and is a target of future empirical work.

9. Discussion

QCAD reframes the quantum threat from a purely strategic, migration-driven concern into an operational detection problem. Its central design insight is that no single observable reliably reveals HNDL, but the conjunction of weak observables interception infrastructure, quantum-vulnerable cryptographic context, and deception ground truth can. By treating crypto-posture not only as a remediation backlog but as a prior that weights behavioral detections, QCAD aligns detection effort with the Mosca-inequality logic that already governs migration prioritization, producing alerts that are both confidence-ranked and risk-ranked.

The framework is also incrementally deployable. Plane C builds on cryptographic inventories that many organizations are already assembling for migration. Plane A reuses flow telemetry that is routinely collected. Plane D requires only the seeding of decoys. Plane B is the most invasive, requiring instrumentation of cryptographic endpoints, and is therefore best deployed where PQC is actually in use. This staged adoptability lowers the barrier to operationalizing quantum-threat detection without waiting for a mature, monolithic product.

10. Limitations and Threats to Validity

- Irreducible blind spot. QCAD can only observe interception that touches infrastructure the defender controls or that interacts with defender-operated decoys. A passive tap on transit infrastructure beyond the perimeter an undersea cable, an upstream backbone, a foreign exchange point produces no observable signal for any sensor the defender can place. QCAD raises the cost and detection probability of harvesting near and within the enterprise; it does not, and cannot, guarantee detection of all HNDL.
- Probabilistic, behavioral inference. HNDL detection is inherently probabilistic. Plane A infers likely collection from metadata patterns that benign operations (legitimate backups, replication, monitoring) can resemble, creating an irreducible tension between recall and false positives that deployment-specific tuning can manage but not eliminate.
- Synthetic evaluation. In the absence of public HNDL/PQC-attack ground truth, evaluation relies on synthetic scenarios whose realism bounds the validity of any reported numbers. The Section VIII figures are explicitly illustrative; empirical validation on operational traffic remains future work.
- Lab-to-production gap for PQC attacks. Side-channel and fault attacks are predominantly demonstrated on

embedded targets under favorable conditions. Whether their network- and host-observable footprint in production resembles the signals Plane B keys on is an open question the framework must validate rather than assume.

- Obfuscation and evolving protocols. Encrypted Client Hello, traffic padding, and aggressive rekeying reduce the features available to Plane A; an adversary aware of QCAD may also avoid decoys or throttle probing below detection thresholds. Detection and evasion will co-evolve.
- Timeline uncertainty. Mosca-weighting depends on an institutional estimate of time-to-CRQC, which carries large uncertainty; the framework should treat Z as a tunable, regularly revised parameter rather than a fixed input.

11. Conclusion and Future Work

The quantum threat to cryptography is, for defenders, already operational: adversaries can harvest encrypted traffic today, the migration to post-quantum cryptography introduces new and exploitable attack surface, and existing tooling detects none of this in a unified way. We presented QCAD, a multi-plane detection framework that fuses behavioral flow analytics, PQC implementation and downgrade telemetry, continuous crypto-posture monitoring, and quantum-vulnerable deception into a single, ATT&CK-mapped, Mosca-weighted detection layer. We formalized its detection algorithms, implemented the framework, and evaluated it on a controlled, reproducible synthetic benchmark, where cross-plane fusion raised recall to 0.97 at a 5% false-positive rate substantially above any single plane while we remained explicit about the framework's probabilistic nature and its irreducible blind spot. Future work comprises empirical validation on operational traffic with realistic synthetic ground truth, refinement of Plane B against in-production PQC deployments, adversarial evaluation against QCAD-aware attackers, and integration of federated learning to enable cross-organization model training without sharing raw traffic. We hope QCAD reframes quantum cybersecurity from a migration checklist into a continuously monitored, detectable, and defensible posture.

Conflict of Interest: The authors declare no conflicts of interest.

Funding: This research received no external funding.

Author Contributions: The author contributed equally to this work. All authors read and approved the final version of the manuscript.

References

[1] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM J. Comput.*, vol. 26, no. 5, pp. 1484–1509, Oct. 1997, doi: 10.1137/S0097539795293172

[2] L. K. Grover, "A fast quantum mechanical algorithm for database search," in *Proc. 28th Annu. ACM Symp. Theory Comput. (STOC)*, Philadelphia, PA, USA, Jul. 1996, pp. 212–219, doi: 10.1145/237814.237866

[3] C. Gidney, "How to factor 2048 bit RSA integers with less than a million noisy qubits," arXiv:2505.15917 [quant-ph], May 2025

[4] M. Mosca, "Cybersecurity in an era with quantum computers: Will we be ready?" *IEEE Security Privacy*, vol. 16, no. 5, pp. 38–41, Sep./Oct. 2018, doi: 10.1109/MSP.2018.3761723.

[5] ETSI, *Quantum Safe Cryptography and Security*, ETSI White Paper No. 8, Sophia Antipolis, France: ETSI, Jun. 2015.

[6] National Institute of Standards and Technology, *FIPS 203: Module-Lattice-Based Key-Encapsulation Mechanism Standard*, Gaithersburg, MD, USA: NIST, Aug. 2024, doi: 10.6028/NIST.FIPS.203

[7] National Institute of Standards and Technology, *FIPS 204: Module-Lattice-Based Digital Signature Standard*, Gaithersburg, MD, USA: NIST, Aug. 2024, doi: 10.6028/NIST.FIPS.204

[8] National Institute of Standards and Technology, *FIPS 205: Stateless Hash-Based Digital Signature Standard*, Gaithersburg, MD, USA: NIST, Aug. 2024, doi: 10.6028/NIST.FIPS.205

[9] P. Ravi, S. Sinha Roy, A. Chattopadhyay, and S. Bhasin, "Generic side-channel attacks on CCA-secure lattice-based PKE and KEMs," *IACR Trans. Cryptogr. Hardw. Embed. Syst. (TCHES)*, vol. 2020, no. 3, pp. 307–335, 2020, doi: 10.13154/tches.v2020.i3.307-335

[10] E. Jendral, K. Ngo, R. Wang, and E. Dubrova, "Single-trace fault-injection attacks on hedged ML-DSA," *IACR Cryptol. ePrint Arch.*, Paper 2024/238, Feb. 2024.

[11] W. Castryck and T. Decru, "An efficient key recovery attack on SIDH," in *Advances in Cryptology – EUROCRYPT 2023*, Lecture Notes in Comput. Sci., vol. 14008, C. Hazay and M. Stam, Eds. Cham, Switzerland: Springer, 2023, pp. 423–447, doi: 10.1007/978-3-031-30589-4_15

[12] W. Beullens, "Breaking Rainbow takes a weekend on a laptop," in *Advances in Cryptology – CRYPTO 2022*, Lecture Notes in Comput. Sci., vol. 13508, Y. Dodis and T. Shrimpton, Eds. Cham, Switzerland:

Springer, 2022, pp. 464–479, doi: 10.1007/978-3-031-15979-4_16

[13] Y. Sheffer et al., "Maintaining PQC continuity in TLS," IETF Internet-Draft draft-sheffer-tls-pqc-continuity, Work in Progress, 2024–2025.

[14] M. Mosca and M. Piani, *2025 Quantum Threat Timeline Report*. Toronto, ON, Canada: Global Risk Institute / evolutionQ, 2025.

[15] National Security Agency, *Commercial National Security Algorithm Suite 2.0 (CNSA 2.0)*, Advisory U/OO/194427-22, Fort Meade, MD, USA: NSA, Sep. 2022.

[16] National Institute of Standards and Technology, *NIST IR 8547 (ipd): Transition to Post-Quantum Cryptography Standards*, Initial Public Draft, Gaithersburg, MD, USA: NIST, Nov. 2024, doi: 10.6028/NIST.IR.8547.ipd

[17] D. Joseph, R. Misoczki, M. Manzano, J. Tricot, F. Dominguez Pinuaga, O. Lacombe, S. Leichenauer, J. Hidary, P. Venables, and R. Hansen, "Transitioning organizations to post-quantum cryptography," *Nature*, vol. 605, no. 7909, pp. 237–243, May 2022, doi: 10.1038/s41586-022-04623-2

[18] M. Hasan, M. Bosri, M. S. Rahman, and M. S. Islam, "A framework for migrating to post-quantum cryptography," *IEEE Access*, vol. 12, pp. 1–15, 2024, doi: 10.1109/ACCESS.2024.0000000.

[19] OWASP CycloneDX, *Cryptography Bill of Materials (CBOM)*, Specification v1.6, OWASP Foundation, 2023–2025.

[20] M. Shen, K. Ye, X. Liu, L. Zhu, J. Kang, S. Yu, Q. Li, and K. Xu, "Machine learning-powered encrypted network traffic analysis: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 791–824, First Quarter 2023, doi: 10.1109/COMST.2022.3208196

[21] D. Blanco-Romero et al., "On the practical feasibility of harvest-now, decrypt-later attacks," arXiv:2603.01091 [cs.CR], Mar. 2026

[22] MITRE Corporation, "ATT&CK Techniques: T1040 – Network Sniffing; T1557 – Adversary-in-the-Middle," MITRE ATT&CK Framework, 2024. [

[23] L. Spitzner, "Honeytokens: The other honeypot," *SecurityFocus*, Aug. 2003.

[24] I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proc. 4th Int. Conf. Inf. Syst. Security Privacy (ICISSP)*, Funchal, Portugal, Jan. 2018, pp. 108–116, doi: 10.5220/0006639801080116

[25] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems," in *Proc. Military Commun. Inf. Syst. Conf. (MilCIS)*, Canberra, Australia, Nov. 2015, pp. 1–6, doi: 10.1109/MilCIS.2015.7348942

[26] A. Habibi Lashkari, G. Draper-Gil, M. S. I. Mamun, and A. A. Ghorbani, "Characterization of encrypted and VPN traffic using time-related features," in *Proc. 2nd Int. Conf. Inf. Syst. Security Privacy (ICISSP)*, Rome, Italy, Feb. 2016, pp. 407–414.

How to cite this article

K.G. Kharade, K. Vengatesan, and H.K. Algabri,, "A Multi-Plane Detection Framework for Quantum-Enabled Cryptographic Attacks: Operationalizing the Detection of Harvest-Now-Decrypt-Later Activity and Attacks on Post-Quantum Cryptography," CyberSystem J., vol. 3, no. 1, pp. 1-11, 2026. doi: 10.57238/csj.20261021

Access this article online