CSAI

cloud security alliance®

CSAI Foundation

Cloud Security Alliance AI Safety Initiative

Kyber Ransomware: First Criminal Use of Post-Quantum Encryption

Unofficial AI-assisted Research

2026-04-24

Key Takeaways

- A ransomware group calling itself Kyber has become the first known criminal operation to deploy a NIST-standardized post-quantum cryptographic algorithm in production, incorporating ML-KEM-1024 (Kyber1024) alongside X25519 in its Windows encryptor to protect symmetric key material against future cryptanalytic recovery.
- The group's ESXi/Linux variant advertises post-quantum capabilities but in practice relies on classical cryptography—ChaCha8 for bulk encryption and RSA-4096 for key wrapping—a discrepancy that may reflect an incomplete port of the Windows PQC implementation, a deliberate false claim to complicate victim decryption attempts, or both.
- Kyber's first confirmed victim is a multibillion-dollar American defense contractor; the group subsequently leaked more than 141 GB of files purportedly stolen from that organization, including project files, internal builds, databases, and backup archives [1][2].
- The Windows variant systematically eliminates recovery paths by deleting shadow copies, disabling boot repair, clearing event logs, terminating database and backup services, and experimentally shutting down Hyper-V virtual machines, making offline backup integrity the primary viable recovery mechanism.
- Organizations should treat this incident as an inflection point: the criminal adoption of NIST-approved PQC algorithms creates a near-term operational threat to decryption-based recovery independent of any quantum computing timeline.

Background

In the years following NIST's initial PQC standardization process, the post-quantum cryptography transition has been broadly framed as a preparatory exercise against a future adversary: the yet-unrealized cryptographically-relevant quantum computer that would render RSA and elliptic-curve algorithms obsolete. The practical urgency of that transition sharpened in August 2024, when NIST published FIPS 203, FIPS 204, and FIPS 205—the first finalized post-quantum cryptographic standards—formalizing ML-KEM, ML-DSA, and SLH-DSA respectively for production use [3][4]. Security practitioners have since focused on migration timelines, cryptographic inventory, and the "harvest now, decrypt later" (HNDL) threat, in which nation-state adversaries collect encrypted traffic today for

decryption once quantum computing matures [5]. The emergence of Kyber ransomware in early 2026 introduces a substantially different threat model: PQC deployed offensively by criminal actors to harden ransomware encryption against the investigative and recovery techniques defenders rely on today.

Cybersecurity firm Rapid7 identified the Kyber ransomware operation in March 2026 during active incident response at a multibillion-dollar American defense contractor and IT services provider [1]. Analysts retrieved and analyzed two distinct ransomware variants that had been deployed simultaneously against the same network: one targeting Windows file servers and one targeting VMware ESXi hypervisors. Both samples share a campaign identifier and communicate through a shared Tor-based ransom infrastructure, confirming that both encryptors were operated under unified command. The group adopted its name from CRYSTALS-Kyber—the lattice-based algorithm standardized by NIST as ML-KEM—a choice that may signal deliberate alignment with the NIST PQC portfolio, a calculated attempt to associate the operation with the standard's longevity, or simply adoption of a recognizable technical term.

ML-KEM-1024, the parameter set Kyber ransomware uses, is the highest-security variant specified in FIPS 203. Its security derives from the computational hardness of the Module Learning with Errors (MLWE) problem, for which no efficient quantum algorithm is currently known [3]. NIST has encouraged organizations to begin migrating to ML-KEM as soon as possible, and adoption has begun progressing in enterprise TLS stacks, browser key exchange, and federal agency cryptographic infrastructure—a trend already underway at the time of NIST's 2024 publication [4]. The criminal appropriation of this algorithm for ransomware key protection represents a meaningful misuse of open standardized cryptography, and it underscores that the security properties NIST designed into ML-KEM—specifically, resistance to future quantum-enabled decryption—are equally available to adversaries.

Security Analysis

Windows Variant: Hybrid Post-Quantum Key Protection

The Windows encryptor is written in Rust and employs a hybrid key encapsulation scheme that represents the most technically sophisticated element of the operation. Kyber1024 and X25519 operate in parallel to protect per-file symmetric key material, while AES-CTR handles bulk data encryption [1][2]. This hybrid architecture mirrors the approach increasingly used in post-quantum TLS 1.3 deployments, where classical and quantum-resistant key exchange algorithms are run simultaneously so that security degrades gracefully if either is compromised. In the context of ransomware, this architecture has a specific defensive-denial implication: even if law enforcement later seized an RSA private key or a

classical key exchange were cryptanalyzed, the Kyber1024 wrapping of the symmetric key material would prevent recovery of encrypted files. The door that has historically been left ajar—opportunistic decryption following key seizure or algorithm weakness—is substantially narrowed.

Upon execution, the Windows variant conducts a methodical anti-recovery campaign. It terminates SQL Server, Microsoft Exchange, and common backup processes; deletes all Volume Shadow Copy Service (VSS) snapshots via `vssadmin`; disables Windows boot repair; clears all Windows event logs; and empties the Recycle Bin [1]. A nascent capability to shut down Hyper-V virtual machines was also observed in the sample, indicating the operators are actively extending the encryptor's disruption surface to include guest virtual machines running on Windows Server hosts. Encrypted files are appended with the `.#~~~` extension. Rapid7 noted an unusual mutex string in the sample that appears to reference a track on the Boomplay music streaming platform—a minor operational security artifact with potential attribution value [1].

ESXi/Linux Variant: Partial PQC Claims and Classical Implementation

Despite prominently advertising "post-quantum encryption using Kyber1024" in its ransom note and ESXi interface defacement, the Linux encryptor in fact relies on classical cryptography: ChaCha8 for bulk file encryption and RSA-4096 for key wrapping [1]. The discrepancy between stated and actual cryptographic implementation warrants careful examination.

The ESXi variant enumerates all virtual machines on the targeted infrastructure and encrypts datastore files using a size-tiered approach: files smaller than 1 MB are encrypted in full; files between 1 MB and 4 MB have only their first megabyte encrypted; and files larger than 4 MB are intermittently encrypted based on operator configuration [1]. Encrypted files receive the `.xhsyw` extension. The partial-encryption strategy, now common in modern ransomware operations, prioritizes breadth and speed over completeness, ensuring that large datastores are rendered inaccessible quickly even when full encryption would be time-prohibitive.

The gap between the ESXi variant's PQC marketing and its classical implementation could reflect an operational timeline in which post-quantum integration began with the Windows encryptor and has not yet been ported to the Linux build. Alternatively, it may represent a deliberate false claim intended to discourage victims from attempting decryption, complicate forensic characterization, or amplify the perceived sophistication of the operation in ransom negotiations. In either interpretation, the discrepancy reveals an important principle: defenders should not treat ransomware operators' cryptographic claims as technically accurate without independent verification, and the presence of PQC marketing should not prevent forensic examination of actual algorithm use.

Targeting Profile and Operational Maturity

The selection of a defense contractor as the first confirmed victim, combined with the technical sophistication of the Windows encryptor, is consistent with a group that possesses prior operational maturity rather than one conducting its initial campaign. Defense industrial base organizations represent attractive ransomware targets in part because they hold both sensitive intellectual property and operationally critical data, creating multiple leverage points for extortion: payment to obtain a decryption key, payment to prevent data publication, and in some cases payment to prevent notification of government clients. The subsequent publication of more than 141 GB of allegedly stolen files—spanning project files, internal builds, databases, and backup archives—confirms that the Kyber operation incorporates data exfiltration alongside encryption, consistent with the double-extortion model that has been prevalent in sophisticated ransomware operations [2].

The coordinated simultaneous deployment of Windows and ESXi encryptors against the same network, sharing a campaign identifier and command infrastructure, reflects a level of pre-compromise access and planning that goes beyond opportunistic exploitation. Security researchers have assessed that the operational infrastructure and dual-variant capability are consistent with groups that are positioned to transition toward a Ransomware-as-a-Service (RaaS) model, in which cryptographic sophistication—including PQC integration—would be made available to affiliate operators who need not possess independent cryptographic expertise [2].

Implications for Recovery and Incident Response

The most operationally consequential implication of Kyber ransomware is not the theoretical quantum-resistance of its encryption, but the practical narrowing of non-payment recovery paths. Ransomware recovery has historically relied on one of several avenues: restoration from uncompromised backups, law enforcement seizure of decryption keys from threat actor infrastructure, identification of implementation flaws in the ransomware's cryptography (weak key generation, key reuse, or algorithm misimplementation), and—in some cases—purchase of decryption keys from third-party intermediaries. The Kyber operation directly addresses or complicates each of these. The anti-recovery measures are designed to eliminate online backup accessibility, and will do so in most standard configurations where backup services run under domain credentials or shadow copies are not protected by additional controls. The Kyber1024 component significantly reduces the probability that key seizure would enable retrospective decryption. And the use of Rust with a hybrid KEM scheme, while not precluding implementation flaws, suggests more deliberate cryptographic design than is common in opportunistic ransomware operations—though this assessment would benefit from comparative analysis of peer families.

This pattern suggests that incident response planning for ransomware should no longer assume that technical recovery pathways will be available in lieu of—or alongside—offline backup restoration. The survivability of verified, offline, tested backups must now be treated as the primary recovery strategy rather than one option among several.

Recommendations

Immediate Actions

Security operations teams should prioritize the following responses in the near term. Backup architecture should be audited immediately to confirm that at least one offline or air-gapped copy exists for all critical systems and applications; given Kyber's systematic deletion of shadow copies and termination of backup services, any backup that is reachable from a compromised host should be considered at risk. VMware ESXi management interfaces should be reviewed to confirm they are not exposed to the corporate network or internet without authentication and network-level access controls. EDR coverage on Windows file servers should be evaluated for behavioral detection of shadow copy deletion (`vssadmin delete shadows /all`), Windows event log clearing, and bulk file rename activity consistent with `.#~~~` extension appending. Threat hunting teams should search for the `.#~~~` and `.xhsyw` extensions, Kyber-associated Tor onion infrastructure, and the campaign mutex string identified in Rapid7's technical analysis [1].

Short-Term Mitigations

Over the next sixty to ninety days, organizations should implement targeted hardening against the specific techniques this operation employs. VMware ESXi and Hyper-V management plane access should be restricted to dedicated privileged access workstations with enforced multi-factor authentication, isolating those management surfaces from the domain-joined Windows environment a ransomware encryptor would traverse. Backup service accounts should be audited for unnecessary administrative privileges, with the principle of least privilege enforced to prevent a ransomware process running under domain credentials from terminating backup agents. Network segmentation between backup infrastructure and production networks—enforced at the network layer rather than through software access controls alone—reduces the blast radius of credential theft and lateral movement. Incident response playbooks should be updated to account for dual-platform ransomware scenarios, with explicit sequencing for simultaneous Windows server and ESXi containment, evidence preservation, and restoration activities.

Strategic Considerations

At a strategic level, Kyber ransomware should serve as a catalyst for accelerating cryptographic modernization roadmaps rather than initiating them. Organizations that have not yet begun cryptographic inventory work—documenting which systems rely on RSA or ECC for key protection, and assessing whether those protections are relevant to ransomware recovery scenarios—should treat this incident as a trigger. The U.S. government has directed federal agencies to complete PQC migration by 2035, but the criminal adoption of ML-KEM illustrates that the threat landscape is not calibrated to that deadline [6]. Hybrid cryptographic schemes combining classical key exchange with ML-KEM are already deployed in TLS 1.3 implementations and provide a practical template for key management infrastructure migration. Incident response retainers and business continuity plans should explicitly account for scenarios in which technical decryption is not feasible and verified offline backup restoration is the sole recovery mechanism.

CSA Resource Alignment

This incident intersects with several CSA frameworks and initiatives. The CSA AI Controls Matrix (AICM), which encompasses and supersedes the Cloud Controls Matrix (CCM) for AI-integrated environments, provides the governance structure for documenting cryptographic algorithm choices and migration timelines as part of a formal cryptographic policy. Organizations pursuing STAR certification should treat FIPS 203 adoption planning and cryptographic inventory completion as auditable evidence within the encryption and key management control domain.

The CSA MAESTRO framework for agentic AI threat modeling is relevant where AI systems communicate over encrypted channels or where AI agents interact with key management infrastructure. If the key protection for AI workloads relies on classical algorithms, those workloads now face the same PQC migration urgency as any other system. CSA Zero Trust guidance on management plane isolation, network segmentation, and privileged access controls directly addresses the lateral movement and ESXi access patterns the Kyber operation exploited; organizations that have implemented Zero Trust principles are better positioned to contain a dual-variant ransomware attack before it reaches virtualization infrastructure.

More broadly, the AI Safety Initiative's ongoing monitoring of adversarial AI capabilities is relevant here: the use of modern systems programming languages such as Rust—and the integration of NIST-standardized cryptographic primitives—reflects a level of software engineering sophistication that may

be increasingly accessible to criminal operators through AI-assisted development tooling. The convergence of AI-accelerated capability development with maturing PQC standards is a trajectory that warrants sustained attention in future threat modeling cycles.

References

[1] Rapid7. "Kyber Ransomware Double Trouble: Windows and ESXi Attacks Explained." Rapid7 Blog, April 2026.

[2] BleepingComputer. "Kyber ransomware gang toys with post-quantum encryption on Windows." BleepingComputer, April 2026.

[3] NIST. "Federal Information Processing Standard (FIPS) 203, Module-Lattice-Based Key-Encapsulation Mechanism Standard." NIST CSRC, August 2024.

[4] NIST. "NIST Releases First 3 Finalized Post-Quantum Encryption Standards." NIST, August 13, 2024.

[5] Palo Alto Networks. "Harvest Now, Decrypt Later (HNDL): The Quantum-Era Threat." Palo Alto Networks Cyberpedia, accessed April 2026.

[6] White House. "National Security Memorandum on Promoting United States Leadership in Quantum Computing While Mitigating Risks to Vulnerable Cryptographic Systems (NSM-10)." White House, May 4, 2022.

