CSAI

cloud security alliance®

CSAI Foundation

Cloud Security Alliance AI Safety Initiative

Microsoft Defender Triple Zero-Day: BlueHammer, RedSun, UnDefend

Three Actively Exploited Vulnerabilities in Windows Defender — Two Still Unpatched

Unofficial AI-assisted Research

2026-04-19

Key Takeaways

- A security researcher operating as "Chaotic Eclipse" released three working exploits targeting Microsoft Defender between April 3 and April 16, 2026 — one (BlueHammer, CVE-2026-33825) has been patched; two (RedSun and UnDefend) remain unpatched as of this writing with no emergency patch committed by Microsoft. [1][2]
- All three vulnerabilities are under active exploitation. Endpoint security firm Huntress documented hands-on-keyboard threat actor activity exploiting BlueHammer from April 10 onward, and observed real-world deployment of RedSun and UnDefend PoC code beginning April 16. [3]
- The three vulnerabilities compose into a coherent attack chain: UnDefend silently degrades Defender's threat intelligence while reporting the endpoint as healthy to management consoles; RedSun escalates an unprivileged user to SYSTEM; and the now-degraded Defender is unable to detect the resulting foothold. Any organization relying exclusively on Defender for endpoint protection is currently operating with a significant, unmitigated gap. [1][4]
- Organizations should apply April 2026 Patch Tuesday updates immediately to close BlueHammer, deploy a supplementary EDR capable of detecting Defender bypasses, and implement out-of-band Defender signature version checks against Microsoft's update feed rather than trusting dashboard health indicators while UnDefend remains unpatched. [5][6]

Background

Between April 3 and April 16, 2026, a security researcher using the pseudonym "Chaotic Eclipse" — also attributed in GitHub repository records to the alternate handle "Nightmare-Eclipse" [14] — publicly released exploit code for three distinct vulnerabilities in Microsoft Defender on GitHub, each accompanied by working proof-of-concept code and detailed technical write-ups. According to the researcher's public statements, the disclosures were motivated by frustration with Microsoft's Security Response Center process, which reportedly required submission of video demonstrations before

vulnerability reports would receive initial triage. Referring to a previous incident, the researcher stated, "I was not bluffing Microsoft, and I'm doing it again," signaling intent to continue releasing findings without awaiting official patches. [7]

Three separate working exploits against a single product released in thirteen days is uncommon in public disclosure history — most comparable multi-vulnerability clusters, such as the 2021 Exchange Server ProxyLogon and ProxyShell disclosures, involved multiple researchers or developed across longer windows. BlueHammer appeared on April 3, 2026, several days before any patch was available, and was actively exploited in the wild by April 10 — four days before Microsoft addressed it in its April Patch Tuesday cycle. [2][3] RedSun and UnDefend followed on April 16, with neither receiving a CVE assignment nor an emergency patch. At the time of this writing, Microsoft has not committed to an out-of-band release; the next scheduled Patch Tuesday is several weeks away. [2][8]

The vulnerabilities share a common attack surface: the privileged file operations that Defender performs during malware detection and remediation. Because Defender runs at SYSTEM integrity level during these operations, flaws in path validation and API usage create pathways for unprivileged local users to redirect or interfere with Defender's own elevated operations. The researcher identified and exploited variations on this theme three times in thirteen days, suggesting the underlying design patterns in Defender's remediation logic may harbor additional undiscovered flaws. [1][9]

Security Analysis

BlueHammer (CVE-2026-33825): Patched

BlueHammer is a local privilege escalation vulnerability rooted in a time-of-check to time-of-use (TOCTOU) race condition within Windows Defender's threat remediation engine. Its NVD description — "Insufficient granularity of access control in Microsoft Defender allows an authorized attacker to elevate privileges locally" — does not fully convey the operational impact; in practice, the exploit's demonstrated reliability on fully patched Windows 10 and Windows 11 systems suggests the prose description understates the attack's severity even though the CVSS vector accurately captures the confidentiality, integrity, and availability components. Microsoft assigned it CVSS 7.8 (High) with vector CVSS:3.1/AV:L/AC:L/PR:L/UI:N/S:U/C:H/I:H/A:H, and classified the root cause as CWE-1220 (Insufficient Granularity of Access Control). [10]

The attack works by placing a file designed to trigger Defender's malware detection, then using a batch opportunistic lock (oplock) to pause Defender's remediation thread at the moment it opens the quarantine target. During that pause — after the path has been validated but before the privileged write

executes — the attacker swaps in an NTFS junction point that redirects the target path to `C:\Windows\System32`. When Defender's remediation thread resumes with its elevated privileges, it writes attacker-controlled content into the protected system directory rather than the intended temporary location. [4][9] This technique, sometimes called an oplock-junction attack, has appeared in privilege escalation research against antivirus and EDR products for years, but its presence in Defender's production code underscores that the vendor's own security tooling has not been immune to well-documented vulnerability classes.

Microsoft addressed BlueHammer in its April 14, 2026 Patch Tuesday release, distributing the fix as Antimalware Platform version 4.18.26050.3011 through Defender's automatic update mechanism. Organizations that receive Defender updates automatically should verify that affected endpoints have received this platform version. Environments with restricted or delayed update policies require expedited action, as BlueHammer was confirmed exploited in the wild before the patch shipped. [5][6]

RedSun: Unpatched

RedSun is a second local privilege escalation technique targeting Defender's cloud file rollback mechanism. When Defender detects a potentially malicious file that has been tagged for cloud sync, it attempts to restore the file to its original location as part of remediation. RedSun exploits the absence of target path validation during that elevated write operation. [1][4]

The attack sequence begins with an attacker registering a fake cloud sync provider — one observed specimen used the recognizable provider name "SERIOUSLYMSFT" as an indicator of compromise — and creating a malicious cloud placeholder file using the Windows Cloud Files API (`cldapi.dll`). The attacker then triggers Defender's detection against the placeholder. When Defender attempts cloud file rollback with SYSTEM privileges, the attacker has already placed Windows reparse points that redirect the rollback destination from an innocuous path to `C:\Windows\System32`. The write lands in the privileged directory, enabling SYSTEM-level code execution. Vulnerability analyst Will Dormann independently verified the exploit's effectiveness, with reporting indicating near-100% reliability across Windows 10, Windows 11, and Windows Server 2019 and later where `cldapi.dll` is present. [14][3]

RedSun carries no assigned CVE as of April 19, 2026. Microsoft has not issued a patch or committed to an out-of-band release, meaning every Windows endpoint with Defender enabled and no supplementary protection is currently exposed to a reliable, locally-exploitable privilege escalation with a public proof-of-concept. The combination of unpatched status and documented in-the-wild exploitation makes RedSun the most immediately actionable risk in this cluster for enterprise security teams.

UnDefend: Unpatched

UnDefend is categorically distinct from the other two vulnerabilities. Rather than escalating privileges, it degrades Defender's protective capability while concealing the degradation. A standard, non-administrative user can trigger UnDefend to block Defender's signature update pipeline or, in its more aggressive mode, disable the Defender engine entirely during a major platform update cycle. The tool continues to report the affected endpoint as healthy to management consoles, meaning compliance dashboards and security information and event management systems that rely on Defender health telemetry will show a false green status on a compromised or degraded endpoint. [1][8]

The practical consequence — confirmed by Huntress intrusion data showing UnDefend deployed alongside RedSun specifically to degrade the detection baseline before the privilege escalation step [3] — is that UnDefend functions as a persistence enabler. By freezing Defender's threat intelligence at a point in time before an attack commences — or by preventing updated signatures from reaching an endpoint — UnDefend ensures that subsequent malware payloads delivered after the disable operation face a weakened detection baseline. A detectable symptom of UnDefend in passive mode is Windows Update error code 80070643 appearing during Defender signature update attempts; [1][8] organizations should alert on repeated occurrences of this error code, particularly across multiple endpoints simultaneously, as it may indicate coordinated UnDefend deployment rather than routine update failures.

The three vulnerabilities chain in sequence, each requiring only local, unprivileged access: deploy UnDefend to silently degrade detection while the endpoint appears compliant, use RedSun to escalate to SYSTEM, then establish a persistent foothold that a definition-frozen Defender cannot detect or remediate. Public proof-of-concept code is available for all three steps. Each vulnerability is independently dangerous; as a coordinated sequence, they represent an unusually complete local compromise pathway against the endpoint security product bundled with every Windows installation.

Disclosure Context and Systemic Concern

Beyond the immediate technical risk, the disclosure pattern raises a systemic concern about the economics of responsible disclosure when vendors impose friction in the triage process. The researcher's stated objection — that MSRC required video demonstrations before providing initial triage of vulnerability reports — raises friction that, in the authors' assessment, is not unreasonable to highlight, though the choice to release working exploits without coordinated disclosure creates clear harm to organizations in the affected installed base. The thirteen-day window in which three exploits were released, exploited, and only partially patched illustrates how quickly the calculus shifts when a researcher with access to working exploit code decides to operate outside coordinated disclosure norms.

Security teams should factor researcher frustration with vendor disclosure processes into their monitoring strategy: when researchers publicly announce dissatisfaction with a vendor's MSRC or PSIRT, the probability of forthcoming uncoordinated disclosures against that vendor increases materially.

This incident suggests that security teams may benefit from monitoring public researcher-vendor friction as a leading indicator of potential uncoordinated disclosure risk — a hypothesis worth validating against broader historical patterns in coordinated vulnerability disclosure research. Organizations that have cultivated relationships with the security research community, through bug bounty programs or threat intelligence partnerships, may receive earlier warning of forthcoming disclosures than those who rely entirely on vendor patch cycles. [7][8]

Recommendations

Immediate Actions

Organizations should treat RedSun and UnDefend as active threats requiring immediate compensating controls, not deferred items pending the next Patch Tuesday.

Apply the April 2026 cumulative security updates from Microsoft immediately across all Windows endpoints. This closes the BlueHammer vector. Defender's automatic update channel delivers the fix as platform version 4.18.26050.3011; verify this version is present on all managed endpoints rather than relying solely on compliance dashboard status, which UnDefend can falsify. [5]

Supplement Microsoft Defender with a third-party EDR or endpoint protection product capable of detecting Defender bypass techniques. This is the most durable compensating control available for RedSun and UnDefend while patches remain unavailable; the remaining actions in this section are complementary and should be implemented in parallel, not deferred in favor of this single step. The security tool itself has become an attack vector; defense-in-depth at the endpoint layer is not optional during this window.

Enable out-of-band Defender definition version monitoring. Check the actual signature timestamp and version present on each endpoint against Microsoft's published update feed, rather than trusting the health status Defender reports to management consoles. UnDefend's silent-degradation capability means dashboard compliance is not a reliable indicator of actual protection level.

Hunt for known indicators of compromise: files named `RedSun.exe`, `UnDefend.exe`, or `FunnyApp.exe` in user-accessible directories such as Downloads, Pictures, and Desktop; cloud sync provider registrations using the string "SERIOUSLYMSFT"; repeated Windows Update error code 80070643 during Defender signature updates; and anomalous writes or modifications to `C:\Windows\System32\TieringEngineService.exe`. [4][6]

Short-Term Mitigations

Enable Windows Defender Application Control (WDAC) or AppLocker policies to prevent unsigned binary execution from user-writable directories. The observed attack pattern stages exploit binaries in Downloads and Pictures folders precisely because those paths are user-writable; WDAC policies that block executable content from those paths would interrupt the observed deployment pattern even without blocking the underlying vulnerability. [3][6]

Enable kernel-level logging for Cloud Files API calls — specifically `CfRegisterSyncRoot`, `CfCreatePlaceholders`, and `CfConnectSyncRoot` invocations. Registrations of these functions from non-standard applications are anomalous and constitute a strong RedSun pre-exploitation signal. Alert on `FSCTL_SET_REPARSE_POINT` with `IO_REPARSE_TAG_MOUNT_POINT` targeting Cloud Files sync roots; this combination does not occur during normal operations and is highly specific to the RedSun technique. [6]

Establish SHA-256 hash baselines of `C:\Windows\System32\TieringEngineService.exe` across managed endpoints and monitor for deviations. Modifications to this binary can indicate successful RedSun exploitation and serve as a post-compromise detection signal for incident response triage. [6]

Alert on MsMpEng.exe failures to load or unexpected stops, particularly during scheduled platform update windows. This behavioral signal is consistent with UnDefend aggressive-mode activity. Correlate these signals with definition version staleness to distinguish UnDefend from ordinary update failures. [6]

Strategic Considerations

This incident demonstrates a pattern that security architects should incorporate into long-term endpoint strategy: when a security tool's privileged operations are implemented without sufficient path validation, the tool's own elevated access becomes a local privilege escalation primitive. Organizations should use this incident as a trigger to assess whether their EDR and AV vendors have published security advisory histories reflecting similar vulnerability classes — TOCTOU in file remediation, oplock-junction attacks,

cloud file callback misuse — as a proxy for whether comparable design patterns may be present in their deployed products. Vendors with a proactive record of identifying and patching such issues before public exploitation represent lower residual risk than those whose advisory histories are sparse.

At the policy level, the broken responsible disclosure that produced these exploits highlights the need for enterprises to maintain relationships with the security research community through channels such as their own bug bounty programs and threat intelligence partnerships. When researchers are blocked from effective engagement with a vendor, they sometimes engage with the community directly — and organizations that have cultivated researcher relationships may receive earlier warning than those who rely entirely on vendor patch cycles.

CSA Resource Alignment

The Microsoft Defender zero-day cluster maps directly to several areas of CSA guidance. The AI Controls Matrix (AICM) and its predecessor the Cloud Controls Matrix (CCM) both contain Threat and Vulnerability Management domain controls requiring timely patch application, continuous vulnerability monitoring, and compensating controls when patches are unavailable. The present situation — two unpatched zero-days under active exploitation with the next Patch Tuesday weeks away — is precisely the scenario those controls are designed to address, and organizations with mature vulnerability management programs should be activating their compensating control procedures now.

CSA's Zero Trust Advancement Center guidance is directly applicable to the UnDefend threat. The Zero Trust "assume breach" principle holds that no single security control, including endpoint protection, should be treated as a binary trust indicator. UnDefend's ability to report a healthy endpoint status to management consoles while silently degrading protection is a concrete illustration of why endpoint AV health status alone cannot be a trust signal for network access decisions. Organizations applying Zero Trust principles will route endpoint health signals through multiple independent verification paths — definition version checked against an authoritative feed, behavioral telemetry from an independent sensor, and network-layer anomaly detection — rather than relying on a single console dashboard that can be falsified. [11]

For organizations deploying AI agents on Windows infrastructure, the MAESTRO (Multi-Agent Environment, Security, Threat, Risk, and Outcome) threat modeling framework highlights a layered risk dimension specific to this attack chain. In environments where the BlueHammer-RedSun-UnDefend chain has been deployed, an AI agent running on the affected endpoint operates on an OS with falsified security posture reports and SYSTEM-level attacker presence. Any trust assertions the agent makes about its execution environment — for example, that code it is running is signed, that its file system

accesses are unobserved, or that its model output is not being intercepted by a local keylogger or memory inspector — are invalid under these conditions. Organizations running agentic AI workloads on Windows endpoints should treat endpoint integrity verification as a prerequisite for the trust claims those agents are permitted to make, and MAESTRO's Layer 0 (Infrastructure) and Layer 1 (Model) threat categories provide structure for modeling these dependencies. [12]

The CSA STAR Program's continuous assurance philosophy, as distinct from point-in-time compliance certification, is also implicated. An endpoint's Defender signature version from yesterday may be accurate; its signature version today, after UnDefend deployment, may be three weeks stale while reporting as current. STAR-level continuous assurance, applied to endpoint protection health, would catch this drift through independent telemetry rather than relying on vendor-reported status — a monitoring architecture that proves its value precisely in incidents like this one. [13]

References

[1] The Hacker News. "Three Microsoft Defender Zero-Days Actively Exploited; Two Still Unpatched." The Hacker News, April 17, 2026.

[2] Born's Tech and Windows World. "Windows Defender 0-days: BlueHammer (patched) and RedSun (unpatched)." borncity.com, April 17, 2026.

[3] Help Net Security. "Researcher drops two more Microsoft Defender zero-days, all three now exploited in the wild." Help Net Security, April 17, 2026.

[4] Picus Security. "BlueHammer & RedSun: Windows Defender CVE-2026-33825 Zero-day Vulnerability Explained." Picus Security, April 2026.

[5] Field Effect. "Microsoft April 2026 Patch Tuesday fixes two zero days, including BlueHammer." fieldeffect.com, April 2026.

[6] Ampcus Cyber. "RedSun and UnDefend Zero-Day Exploits Hit Defender." ampcuscyber.com, April 2026.

[7] gblock.app. "One Researcher Dropped Three Windows Defender Zero Days in 13 Days — Microsoft Has Only Patched One." gblock.app, April 2026.

[8] Network Security Magazine. "Three Microsoft Defender Zero-Days Actively Exploited; Two Still Unpatched." Network Security Magazine, April 2026.

[9] Cyderes. "BlueHammer: Inside the Windows Zero-Day That Turns Defender Against Itself." cyderes.com, April 2026.

[10] NIST National Vulnerability Database. "CVE-2026-33825 Detail." NVD, April 17, 2026.

[11] Cloud Security Alliance. "Zero Trust Advancement Center." cloudsecurityalliance.org.

[12] Cloud Security Alliance. "MAESTRO: Agentic AI Threat Modeling Framework." CSA Blog, February 2025.

[13] Cloud Security Alliance. "CSA STAR Program." cloudsecurityalliance.org.

[14] BleepingComputer. "New Microsoft Defender RedSun Zero-Day PoC Grants SYSTEM Privileges." BleepingComputer, April 2026.

