CSAI Foundation

Cloud Security Alliance AI Safety Initiative

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

Two Unpatched Privilege Escalation Flaws and a Silent Defense-Disabling Tool Actively Exploited in the Wild

Unofficial AI-assisted Research

2026-04-18

Key Takeaways

- Three zero-day exploits targeting Windows Defender — BlueHammer, RedSun, and UnDefend — were released publicly between early and mid-April 2026 by a security researcher who publicly cited dissatisfaction with MSRC's handling of the disclosure process [1][2].
- BlueHammer (CVE-2026-33825), a local privilege escalation flaw rated CVSS 7.8, was patched in Microsoft's April 2026 Patch Tuesday update; RedSun and UnDefend remain unpatched as of April 18, 2026 [3][4].
- RedSun abuses Defender's cloud file rollback mechanism to achieve SYSTEM-level code execution on any Windows 10, Windows 11, or Windows Server 2019 and later system with Defender enabled, with reported near-100% reliability [5][6].
- UnDefend silently blocks Defender's signature update pipeline while reporting the endpoint as healthy to management consoles, effectively freezing Defender's signature-based threat intelligence without triggering alerts [7][8].
- Huntress researchers confirmed active exploitation of all three techniques in the wild, with BlueHammer observed since April 10 and RedSun and UnDefend PoC activity documented from April 16 onward [1][9].
- Organizations relying exclusively on Windows Defender for endpoint protection carry materially elevated risk from two of these three unpatched flaws. Supplemental EDR capability and Windows hardening controls represent the most actionable interim mitigations available until Microsoft issues patches [see Recommendations section].

Background

Windows Defender — formally Microsoft Defender Antivirus — serves as the primary endpoint protection layer on virtually every Windows 10, Windows 11, and Windows Server installation globally. Because Defender runs with SYSTEM privileges and performs privileged file operations as part of its core threat remediation functions, it represents a high-value attack surface: a flaw in the security tool itself can become the mechanism by which an attacker obtains the access that the tool is designed to prevent.

The BlueHammer, RedSun, and UnDefend exploit cluster did not emerge from adversary-nation research or organized criminal discovery. They were released by a single security researcher operating under the pseudonyms "Chaotic Eclipse" and "Nightmare-Eclipse," who published working proof-of-concept code on GitHub as a public protest against what the researcher described as severe deterioration in MSRC's handling of responsible vulnerability disclosure [2]. According to public statements attributed to the researcher, MSRC imposed unusual procedural requirements during the disclosure process, including a demand for a video demonstration of the exploit [2].

Whatever the merits of the researcher's grievance, the decision to publish functional exploit code before patches existed for two of the three flaws transformed a dispute about disclosure ethics into an active threat to enterprise Windows environments worldwide. Within days of the GitHub publication, Huntress threat researchers confirmed that threat actors were deploying the exploits against real endpoints [1][9].

The disclosure unfolded in stages across roughly two weeks. BlueHammer appears to have been the first technique disclosed, and Microsoft moved quickly enough to include a fix (CVE-2026-33825) in the April 2026 Patch Tuesday release [3][4]. RedSun and UnDefend were released in a second wave on approximately April 15–16, 2026, following the researcher's stated frustration that the patch for BlueHammer had been credited without any acknowledgment of the underlying systemic issues raised during the disclosure process [2]. As of the date of this note, no CVE identifiers have been assigned to RedSun or UnDefend, and Microsoft has issued no patch or official timeline for remediation.

Security Analysis

BlueHammer (CVE-2026-33825): Remediation Engine TOCTOU

BlueHammer exploits a time-of-check to time-of-use (TOCTOU) race condition embedded in Defender's threat remediation logic [3][10]. During malware cleanup, Defender performs privileged file operations — quarantining, removing, or rolling back files — with SYSTEM-level privileges. The vulnerability arises because Defender validates the target file path at the time of check but does not re-validate it at the moment of the actual write operation, creating a narrow window in which the path can be redirected without Defender's knowledge.

The exploit chain begins by placing a file that triggers a Defender detection. When Defender's real-time protection engine initiates remediation, BlueHammer uses a batch opportunistic lock (oplock) to pause Defender's file operation at the critical moment between check and write. During this pause, the attacker creates an NTFS junction point — a filesystem-level directory symlink — that redirects Defender's target

path from an attacker-controlled temporary directory to `C:\Windows\System32`. When Defender resumes and completes the write operation, it follows the junction and overwrites a system binary with SYSTEM-level privileges, enabling code execution in the most privileged context on the local machine [3][10].

Microsoft addressed BlueHammer in the April 2026 Patch Tuesday cycle, delivering the fix through Defender Antimalware Platform update version 4.18.26030.3011 [3][4]. Organizations that have applied April 2026 cumulative updates should be protected against this specific technique, though the remediation does not address the remaining two exploits.

RedSun: Cloud File Rollback Weaponized as Privilege Escalation

RedSun presents greater practical risk than BlueHammer because it remains unpatched, exploits a wider attack surface through the Windows Cloud Files API rather than a traditional file operation, and achieves SYSTEM-level privilege escalation with reported near-100% reliability [5][6][11]. When Defender detects what it classifies as a cloud-tagged file — a file placeholder managed by the Windows Cloud Files API, used natively by OneDrive and other sync providers — it attempts to restore or roll back that file to its original location as part of the remediation process. Critically, Defender performs this rollback without validating whether the restoration target path remains under attacker-accessible directories, enabling path redirection through the same NTFS junction technique used in BlueHammer.

The attack sequence involves triggering a Defender detection on a crafted file, then replacing it with a cloud file placeholder using the Cloud Files API. As Defender initiates its rollback, the attacker uses filesystem manipulation — again combining oplocks and NTFS junctions — to redirect the target path to `C:\Windows\System32`. Defender follows the redirected path and writes the file with SYSTEM-level privileges, delivering arbitrary SYSTEM-level code execution to an entirely unprivileged user [5][6]. Researchers have documented the exploit functioning with near-100% reliability on fully patched Windows 10, Windows 11, and Windows Server 2019 and later systems where `cldapi.dll` is present — which describes the vast majority of current production Windows deployments [5].

No CVE has been assigned and no patch exists as of April 18, 2026. Microsoft has not publicly acknowledged the vulnerability or provided a remediation timeline. The working PoC remains publicly accessible on GitHub [6].

UnDefend: Silent Degradation of Defender's Detection Capability

While BlueHammer and RedSun are precision privilege escalation tools, UnDefend addresses a different attacker objective: rendering endpoint protection ineffective while concealing that degradation from defenders [7][8][9]. UnDefend targets Defender's signature update pipeline, exploiting logic in how Defender manages its update mechanism as an unprivileged user. In its passive mode [7], UnDefend blocks all Defender signature updates silently, freezing the endpoint's threat intelligence at whatever definitions were current at the time of execution. In its aggressive mode [7], UnDefend exploits a condition triggered by major Defender platform updates to fully disable the engine.

The strategic danger of UnDefend extends beyond simply bypassing Defender: the tool reportedly causes Defender to misreport its own status to management consoles and security tooling, indicating that definitions are current and the engine is healthy when neither is true [7][8]. An endpoint where UnDefend has been deployed will appear compliant in dashboard views while Defender's signature-based detection is blind to any malware signature added after the point of compromise. Behavioral and heuristic detection capability may be partially preserved, but the endpoint's ability to catch known threats via signature matching is entirely frozen. This characteristic makes UnDefend particularly well-suited as a persistence enabler — deployed early in an intrusion to ensure that subsequent malware payloads escape signature-based detection by the platform most widely trusted to catch them.

Huntress researchers observed UnDefend deployed alongside RedSun in active attacks, consistent with a combined attack chain in which an attacker first degrades Defender's detection coverage, then uses RedSun to escalate privileges, and subsequently uses the resulting SYSTEM access to establish persistent footholds that a degraded Defender cannot detect [1][9].

The Combined Attack Chain

Considered individually, each of these tools presents a serious but bounded threat. Considered together, they form a coherent intrusion chain that is available to any threat actor willing to pull working code from GitHub. An attacker with local, unprivileged access — such as a phishing victim who executes a malicious attachment — can deploy UnDefend to silently degrade Defender, use RedSun to escalate to SYSTEM, and then operate with the highest local privileges while Defender's signature-based detection remains degraded. The April Patch Tuesday update removes BlueHammer as an option but does nothing to constrain either of the remaining two techniques, and the combined RedSun-plus-UnDefend chain is fully operational on every supported Windows version where `cldapi.dll` is present, as of this writing [5][8].

Recommendations

Immediate Actions

Organizations should apply the April 2026 Microsoft Patch Tuesday cumulative updates across all Windows endpoints and servers as their first priority. The update includes the BlueHammer fix (Defender Platform 4.18.26030.3011) and 163 additional security patches [3][4]. Eliminating BlueHammer reduces the available tooling from three to two techniques, though the remaining pair present sufficient risk to warrant additional action regardless.

Until Microsoft issues patches for RedSun and UnDefend, security teams should supplement Defender with a third-party EDR capable of detecting Defender bypasses and anomalous filesystem operations. Huntress and other managed detection providers have demonstrated detection capability for these specific techniques and are the more reliable source of real-time protection for the unpatched flaws [1][9].

Defenders should implement immediate detection logic for the behavioral signatures associated with these attacks. Monitoring should include: creation of NTFS junction points in user-writable directories (`%USERPROFILE%\Downloads`, `%USERPROFILE%\Pictures`, `%TEMP%`); unusual use or registration of cloud sync providers via the Cloud Files API on endpoints that do not legitimately use cloud file synchronization; privileged Defender file write operations targeting `C:\Windows\System32`; and anomalous execution of binaries named `RedSun.exe`, `FunnyApp.exe`, or variations thereof from user-accessible paths [1][8].

Short-Term Mitigations

Application control enforcement using Windows Defender Application Control (WDAC) or AppLocker can materially raise the bar for these exploits by preventing execution of unsigned or unexpected binaries from user-writable directories including `Downloads` and `Pictures`. These controls do not patch the underlying vulnerabilities but do constrain the environments in which the exploits can successfully run.

Attack Surface Reduction (ASR) rules should be reviewed and applied where compatible with operational requirements. Relevant rules include those blocking executable content from email and webmail, blocking process creation from PSExec and WMI commands, and blocking untrusted and unsigned processes from running from USB or user-writable paths.

Organizations operating internet-facing or high-risk endpoints should consider running Defender in passive mode alongside a fully active third-party AV/EDR. This configuration preserves Defender's visibility and logging while removing it from the primary enforcement path, which reduces the value of UnDefend as a capability-degradation tool against those specific endpoints.

Review Defender health reporting in SIEM and endpoint management tooling for discrepancies between reported status and actual definition currency. UnDefend's capability to misreport Defender's health means that dashboard-level compliance checks are insufficient; out-of-band verification of definition version against Microsoft's published update feed is recommended [7][8].

Strategic Considerations

This incident exposes a structural risk that extends beyond the specific vulnerabilities: Windows environments that rely on a single endpoint security product — particularly the built-in one — carry concentrated exposure when that product becomes the attack vector. Defense-in-depth at the endpoint layer, including behavioral detection capability independent of Defender's threat intelligence, provides resilience that signature-based approaches alone cannot offer.

The researcher's stated motivations — frustration with a vulnerability disclosure process that the researcher experienced as dismissive [2] — echo concerns that have been raised more broadly in the security research community about the consistency of coordinated vulnerability disclosure practices at major platform vendors, though the specific experiences described here represent the researcher's own account. Organizations should monitor the MSRC advisory channel for patch availability for RedSun and UnDefend, as Microsoft's response timeline here will bear on the duration of enterprise exposure. At the time of this writing, no out-of-band patch has been released, and neither vulnerability has been formally acknowledged.

CSA Resource Alignment

The Defender triple zero-day cluster maps directly to several core CSA frameworks and guidance areas. Organizations using CSA's AI Controls Matrix (AICM) and Cloud Controls Matrix (CCM) should assess controls in the **Threat & Vulnerability Management** and **Endpoint & Mobile Security** domains, specifically those requiring timely patch deployment, defense-in-depth, and independent validation of security tooling health.

CSA's **Zero Trust** guidance is directly applicable to this incident. UnDefend's ability to silently degrade Defender while reporting a healthy status illustrates precisely the threat scenario that the Zero Trust principle of "assume breach" is designed to address: treating any endpoint as potentially compromised regardless of its apparent AV health. Organizations following CSA's Zero Trust Advancement Center guidance should treat endpoint AV status as one signal among many rather than as a binary health indicator, and should architect detection capabilities that do not depend on the health of the primary endpoint protection product [12].

CSA's Zero Trust for Critical Infrastructure guidance [13], which addresses defense-in-depth in high-stakes operational environments, is applicable to this threat pattern. The RedSun-plus-UnDefend chain — in which the security product is weaponized to achieve SYSTEM access — exemplifies the threat model that Zero Trust architectures are designed to resist by ensuring that no single control is relied upon as the sole barrier.

CSA's **STAR** (Security, Trust, Assurance, and Risk) program provides a framework for organizations to assess and attest to the effectiveness of their endpoint security controls. The degraded-but-healthy-appearing state that UnDefend creates highlights the importance of STAR's continuous assurance model over point-in-time compliance attestation.

Finally, CSA's work on agentic AI security through the **MAESTRO** threat modeling framework and the AI Safety Initiative is relevant for organizations deploying AI agents on Windows infrastructure. An agent operating on a compromised Windows host — one where an attacker has achieved SYSTEM access via RedSun and masked the compromise using UnDefend — may execute with an implicitly trusted context that has been fundamentally undermined. AI governance architectures must account for the possibility that the underlying infrastructure on which agents operate has been compromised at the OS level.

References

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

[2] BleepingComputer. "Disgruntled researcher leaks 'BlueHammer' Windows zero-day exploit." BleepingComputer, April 2026.

[3] NIST National Vulnerability Database. "CVE-2026-33825." NVD, April 14, 2026.

[4] Field Effect. "Microsoft April 2026 Patch Tuesday fixes two zero days, including BlueHammer." Field Effect Blog, April 2026.

[5] BleepingComputer. "New Microsoft Defender 'RedSun' zero-day PoC grants SYSTEM privileges." BleepingComputer, April 2026.

[6] CloudSEK. "RedSun: Windows 0day when Defender becomes the attacker." CloudSEK Blog, April 2026.

[7] Ampcus Cyber. "RedSun and UnDefend Zero-Day Exploits Hit Defender." Ampcus Cyber Shadow Ops Intel, April 2026.

[8] Blackswan Cybersecurity. "THREAT ADVISORY: RedSun Zero-Day (Windows Defender), April 17, 2026." Blackswan Cybersecurity, April 2026.

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

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

[11] CSO Online. "Caught, Quarantined, Re-Installed: RedSun Turns Microsoft Defender on Itself." CSO Online, April 2026.

[12] Cloud Security Alliance. "Zero Trust Advancement Center." CSA, accessed April 2026.

[13] Cloud Security Alliance. "Zero Trust for Critical Infrastructure Security." CSA, 2024.

