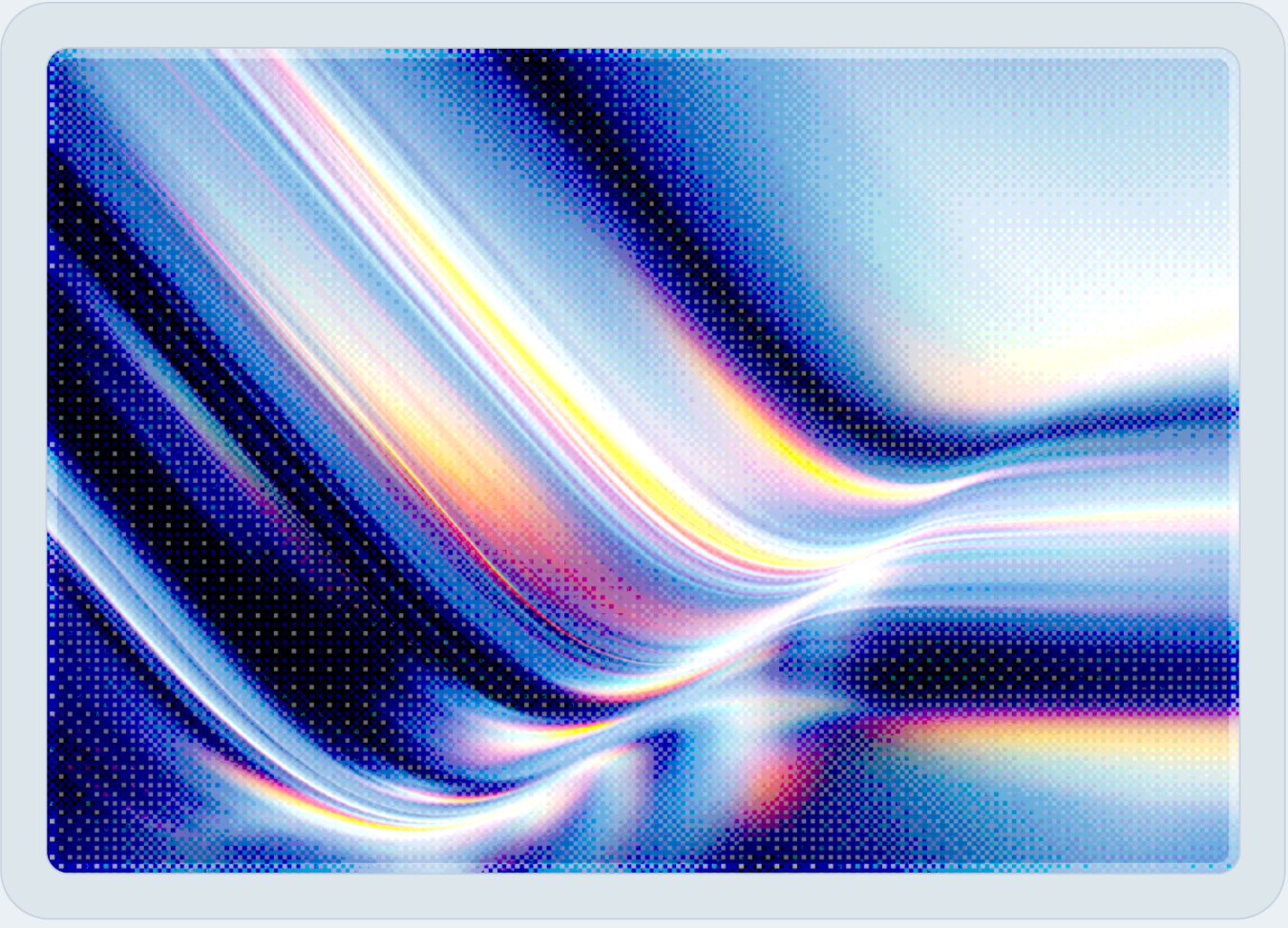
CSAI Foundation | Cloud Security Alliance

PAN-OS Zero-Day: Unauthenticated Root RCE Under Active Exploitation

CVE-2026-0300 Buffer Overflow in Captive Portal – No Patch Available

2026-05-07

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Key Takeaways

- CVE-2026-0300 is a critical (CVSS 9.3) buffer overflow vulnerability in the User-ID Authentication Portal of Palo Alto Networks PAN-OS that allows an unauthenticated remote attacker to execute arbitrary code with root privileges on PA-Series and VM-Series firewalls [2][4].
- Active exploitation is confirmed. CISA added CVE-2026-0300 to its Known Exploited Vulnerabilities (KEV) catalog on May 6, 2026, requiring Federal Civilian Executive Branch agencies to apply mitigations by May 9, 2026 [1].
- No patch exists as of this writing (May 7, 2026). Palo Alto Networks plans a first wave of fixes beginning May 13, 2026, with a second wave on May 28, 2026 [2].
- Unit 42 has attributed active exploitation to threat cluster CL-STA-1132, which it assesses as likely state-sponsored. Post-exploitation behavior includes credential harvesting from the compromised firewall, Active Directory enumeration, deployment of tunneling tools, and systematic destruction of forensic evidence [3].
- Organizations should immediately restrict or disable the User-ID Authentication Portal on all perimeter-facing interfaces and audit exposure across their PAN-OS estate.

Background

Palo Alto Networks PAN-OS is the operating system powering the company's Next-Generation Firewalls (NGFWs), widely deployed at enterprise network perimeters, data center edges, and cloud on-ramps worldwide. PAN-OS runs on PA-Series hardware appliances and VM-Series virtual firewalls. Its User-ID feature enables policy enforcement tied to specific user identities rather than IP addresses alone; the Authentication Portal (also called Captive Portal) is the web-based component that authenticates users who have not yet been identified by other means—typically via a browser redirect to a login page hosted on the firewall itself.

The Authentication Portal's position in the network architecture is consequential for risk analysis. When organizations configure it to accept connections from untrusted networks or directly from the internet—for example, to authenticate guest Wi-Fi users or remote workers who cannot reach an internal identity provider—the portal becomes an externally reachable HTTP service running on the same device that

enforces all perimeter security policy. That architectural choice transforms what is an internal convenience feature into an internet-facing attack surface on one of the most privileged devices in the network.

CVE-2026-0300 was publicly disclosed by Palo Alto Networks on or around May 6, 2026, simultaneously with Palo Alto's acknowledgment that it had already observed exploitation in the wild [2]. The vulnerability is classified under CWE-787 (Out-of-bounds Write), the top-ranked entry in MITRE's CWE Top 25 Most Dangerous Software Weaknesses list and the most common root cause category in memory-corruption vulnerabilities [9]. The weakness exists in the code that processes incoming authentication requests within the Captive Portal service; a specially crafted network packet triggers an out-of-bounds write that corrupts adjacent memory in a way that ultimately yields control of execution flow. Because the portal service runs with root privileges, successful exploitation immediately grants the attacker the highest level of access on the operating system.

Security Analysis

Technical Vulnerability Profile

The out-of-bounds write in CVE-2026-0300 requires no prior authentication and no user interaction: an attacker can send specially crafted packets directly to the Authentication Portal's listening port and trigger the buffer overflow remotely. The vulnerability's exploitability is rated as "network-accessible" with low attack complexity, reflecting that no credentials, physical access, or social engineering are required to initiate exploitation [2][4].

The CVSS base score is 9.3 when the Authentication Portal is reachable from the internet or from untrusted network zones. That score decreases to 8.7 when portal access is restricted to trusted internal IP addresses. Organizations with internet-facing portals are therefore operating at the higher severity rating and should prioritize mitigation accordingly [2]. In either scenario, successful exploitation results in complete compromise of the firewall's operating system: the attacker gains arbitrary code execution running as root on the appliance.

Affected versions span three major PAN-OS branches. PAN-OS 11.2 is vulnerable in versions prior to 11.2.4-h17, 11.2.7-h13, 11.2.10-h6, and 11.2.12. PAN-OS 11.1 is affected in versions prior to 11.1.4-h33, 11.1.6-h32, 11.1.7-h6, 11.1.10-h25, 11.1.13-h5, and 11.1.15. PAN-OS 10.2 is affected in versions prior to 10.2.7-h34, 10.2.10-h36, 10.2.13-h21, 10.2.16-h7, and 10.2.18-h6. PAN-OS 12.1.x is also affected [4][5]. Cloud NGFW, Panorama, and Prisma Access are not impacted [2].

Active Exploitation and Threat Actor Attribution

Unit 42, Palo Alto Networks' threat intelligence team, is tracking exploitation under the designation CL-STA-1132—a cluster of activity it assesses with moderate confidence as likely state-sponsored [3]. The observed exploitation sequence follows a pattern consistent with deliberate, targeted intrusion operations rather than opportunistic scanning campaigns. After triggering the buffer overflow, the attacker injected shellcode into an nginx worker process running on the firewall, effectively turning the web server component into a persistent execution environment under attacker control [3].

Post-exploitation behavior observed by Unit 42 suggests a deliberate effort to evade detection and preserve operational security. The cluster deployed EarthWorm and ReverseSocks5—publicly available but operationally mature tunneling tools frequently associated with espionage-oriented intrusions [3]—to establish covert command-and-control channels through the compromised firewall. Attackers then leveraged credentials likely obtained from the firewall itself to conduct Active Directory enumeration of the victim's internal network. The intrusion sequence concluded with systematic destruction of logs and other forensic artifacts on the compromised device, complicating incident response and attribution [3].

This post-exploitation behavior pattern—credential harvest, AD reconnaissance, covert tunneling, evidence destruction—is consistent with objectives focused on persistent access and lateral movement into the internal environment rather than immediate ransomware deployment or data exfiltration. Organizations that discover a compromised firewall should treat the entire network segment behind it as potentially exposed.

Attack Surface Exposure

The risk surface for CVE-2026-0300 is bounded but significant. The vulnerability is only triggerable when the User-ID Authentication Portal is enabled and reachable by an attacker. Organizations that have never enabled the Captive Portal, or that have restricted it to internal trusted zones, face substantially lower risk than those with internet-exposed portals. In complex network environments, however, misconfigurations of this kind are not uncommon, and many organizations may lack a current inventory of which interfaces have portal response pages enabled. Palo Alto Networks specifically noted that even interfaces where the full portal is not intentionally enabled may still present exposure if the Interface Management Profile includes Response Pages—a subtler attack surface that warrants auditing [2].

Rapid7 estimated approximately 225,000 internet-facing PAN-OS instances at risk at the time of disclosure [5], while Shadowserver tracked over 5,800 exposed VM-series firewalls globally as of early May 2026 [8]. PAN-OS firewalls occupy perimeter positions in enterprises, government agencies, healthcare systems, and critical infrastructure globally, making the affected population substantial. The KEV listing and the rapid regulatory response—a three-day remediation deadline for federal agencies—

signal that authorities assess the real-world risk as acute. CERT-EU published a coordinating advisory on May 6, 2026 [6], and the Cyber Security Agency of Singapore issued an alert the same day [7]—signals of concern extending across multiple international jurisdictions.

Implications for Defenders

A compromised perimeter firewall is among the most severe incident types an organization can face, for several reasons. First, the firewall is typically the most trusted device in the network—its traffic inspection and policy enforcement decisions are accepted by all downstream systems. An attacker with root access can read decrypted traffic, extract VPN credentials, modify security policies, and pivot laterally with the implicit trust that the network extends to the firewall's IP addresses. Second, Unit 42's observation that attackers obtained credentials from the firewall suggests that PAN-OS may retain authentication material as part of the User-ID workflow [3], meaning a compromise may yield credentials for Active Directory or other identity systems without requiring a separate phishing or credential-stuffing campaign. Third, PAN-OS firewalls do not run standard endpoint detection and response agents, so the initial compromise will be invisible to EDR tools monitoring servers and workstations. Defenders must rely on network telemetry, firewall logs, and external traffic monitoring.

Recommendations

Immediate Actions

The most effective immediate mitigation is to eliminate the attack surface entirely by disabling the User-ID Authentication Portal on all L3 interfaces that receive untrusted or internet traffic. For organizations that require Captive Portal functionality, Palo Alto Networks advises restricting portal access to only trusted IP zones through the Interface Management Profile configuration [2]. As a secondary measure, organizations should disable Response Pages in the Interface Management Profile on every L3 interface in any zone where untrusted or internet traffic may arrive—this addresses the subtler exposure path in which the portal's HTTP response mechanism can still be reached even when the portal login page itself is not intentionally published [2].

Any organization that has operated an internet-exposed Authentication Portal should treat its PAN-OS appliances as potentially compromised until a thorough investigation rules out intrusion. Incident response teams should acquire forensic images of firewall configurations and logs immediately, before any device reboot or configuration change, and should specifically examine nginx worker process activity, outbound tunneling connections, and any changes to security policy rules or administrator accounts.

Given that CL-STA-1132 is known to destroy logs on compromised devices, external network flow logs, DNS query logs, and upstream traffic captures may be more reliable evidence sources than the firewall's own logs.

Short-Term Mitigations

Organizations should accelerate patching schedules to apply fixed PAN-OS versions as soon as they become available beginning May 13, 2026. Patching should be treated as a critical-priority, out-of-cycle maintenance event rather than deferred to a regular maintenance window. The affected version ranges span 10.2, 11.1, 11.2, and 12.1 branches, so organizations should inventory all managed appliances against these ranges and establish a prioritized patching sequence that addresses internet-facing firewalls first.

In parallel, security operations teams should implement detection coverage for the post-exploitation indicators associated with CL-STA-1132. Network-based detection should focus on anomalous outbound connections from firewall management IP addresses, particularly connections to unusual external hosts on non-standard ports consistent with SOCKS tunneling. Identity and access management teams should audit Active Directory event logs for unusual enumeration activity—particularly LDAP queries or directory replication requests originating from firewall IP addresses—that could indicate an attacker has pivoted to internal reconnaissance.

Organizations should also review administrative access to Palo Alto Networks' Panorama management platform and connected cloud services. If an attacker gains root access to a managed firewall, they may be able to extract Panorama credentials stored in the local configuration, potentially pivoting to management-plane access across the entire firewall fleet.

Strategic Considerations

CVE-2026-0300 illustrates a recurring pattern in perimeter security: features designed to extend network access to users—such as Captive Portal authentication—require the most-trusted network device to expose additional HTTP services to untrusted networks. Similar vulnerabilities in Citrix ADC (CVE-2019-19781) and Palo Alto's own GlobalProtect portal (CVE-2019-1579) reflect the same structural risk, demonstrating that this is an architectural challenge that persists across product generations and vendors. Security teams should conduct a systematic review of all management and authentication services enabled on perimeter devices—including any interface management profiles, GlobalProtect portals, and web-based administration interfaces—and apply the principle of least exposure by disabling any service not actively required.

More broadly, this incident reinforces the strategic case for Zero Trust network architectures that decouple user authentication from the perimeter firewall's control plane. When user authentication flows through dedicated identity providers and application proxies rather than directly through the firewall operating system, a compromise of the firewall's user-facing services does not simultaneously yield access to the policy enforcement engine. Zero Trust is not immune to vulnerability, but it reduces the blast radius of a single perimeter device compromise. Organizations undertaking network modernization initiatives should weigh this architectural benefit as they evaluate their reliance on PAN-OS Captive Portal and similar perimeter-hosted authentication services.

State-sponsored exploitation of perimeter network devices is a well-documented pattern that predates this vulnerability. CL-STA-1132's post-exploitation behavior—credential harvesting, AD enumeration, covert tunneling, evidence destruction—represents a mature tradecraft playbook. Organizations in sectors that are frequent targets of espionage campaigns (government, defense, healthcare, energy, financial services) should treat this vulnerability disclosure as an occasion to assess not just their patching posture but their broader capability to detect and respond to firewall-level compromise.

CSA Resource Alignment

The CSA Zero Trust guidance portfolio directly addresses the architectural conditions that amplify the risk of CVE-2026-0300. CSA's published guidance on Zero Trust—including the Software-Defined Perimeter Architecture Guide [10] and the Zero Trust Guidance for Achieving Operational Resilience [11]—advocates for removing implicit trust from perimeter devices and implementing identity verification that does not depend on the security of any single network appliance. The fundamental Zero Trust principle of "never trust, always verify" applies with particular force to Captive Portal architectures, where user identity verification is delegated to the same appliance enforcing perimeter policy.

The CSA Cloud Controls Matrix (CCM) [12] and the AI Controls Matrix (AICM) [13] both address infrastructure security controls and third-party risk management that are relevant to the assessment and response disciplines triggered by this vulnerability. CCM control domains covering network security (NET), infrastructure and virtualization security (IVS), and security incident management (SEF) provide a structured framework for organizations inventorying their exposure and establishing response workflows. The AICM's guidance on shared security responsibility is relevant for organizations that deploy PAN-OS in managed service or MSSP contexts, where the responsibility boundary for patching and configuration hardening may be distributed across internal teams and external providers.

CSA's STAR (Security Trust Assurance and Risk) program [14] provides a mechanism for organizations to assess vendor security posture and track the patch response timelines of critical technology providers—a relevant capability given the compressed timeframe between CVE disclosure and expected patch availability for CVE-2026-0300. Organizations conducting vendor risk reviews of Palo Alto Networks or their MSSP partners in the wake of this vulnerability may find the STAR registry and CAIQ questionnaire framework useful for documenting and tracking remediation commitments.

CSA's AI Safety Initiative [15] research on the "AI Vulnerability Storm" and AI-accelerated vulnerability discovery is also germane: as AI-assisted vulnerability research tools lower the barrier to finding and weaponizing memory-corruption vulnerabilities like CVE-2026-0300, the window between public disclosure and broad exploitation is likely to further compress. Security programs that have not yet adopted AI-augmented patch prioritization and threat intelligence workflows may find themselves structurally unable to respond within the increasingly short timelines that critical infrastructure vulnerabilities demand.

References

[1] CISA. "Known Exploited Vulnerabilities Catalog." CISA, continuously updated; CVE-2026-0300 added May 6, 2026.

[2] Palo Alto Networks. "CVE-2026-0300 PAN-OS: Unauthenticated user initiated Buffer Overflow Vulnerability in User-ID™ Authentication Portal." Palo Alto Networks Security Advisories, May 2026.

[3] Unit 42. "Threat Brief: Exploitation of PAN-OS Captive Portal Zero-Day for Unauthenticated Remote Code Execution." Palo Alto Networks Unit 42, May 2026.

[4] Tenable. "Palo Alto Networks PAN-OS 10.2.x / 11.1.x / 11.2.x / 12.1.x Vulnerability." Tenable, May 2026.

[5] Rapid7. "Critical Buffer Overflow in Palo Alto Networks PAN-OS User-ID Authentication Portal (CVE-2026-0300)." Rapid7 Blog, May 2026.

[6] CERT-EU. "Critical Vulnerability in PAN-OS." CERT-EU Security Advisory 2026-006, May 2026.

[7] Cyber Security Agency of Singapore. "Active Exploitation of Palo Alto Networks PAN-OS software." CSA Singapore Alert AL-2026-048, May 6, 2026.

[8] BleepingComputer. "Palo Alto Networks warns of firewall RCE zero-day exploited in attacks." BleepingComputer, May 2026.

[9] MITRE. "CWE Top 25 Most Dangerous Software Weaknesses." MITRE CWE Program, 2024.

[10] Cloud Security Alliance. "Software-Defined Perimeter Working Group." CSA, 2024.

[11] Cloud Security Alliance. "Zero Trust Advancement Center." CSA, 2024.

[12] Cloud Security Alliance. "Cloud Controls Matrix (CCM)." CSA, v4.0.

[13] Cloud Security Alliance. "AI Controls Matrix (AICM)." CSA, 2024.

[14] Cloud Security Alliance. "STAR: Security, Trust, Assurance, and Risk." CSA, continuously updated.

[15] Cloud Security Alliance. "AI Safety Initiative." CSA, 2024.

