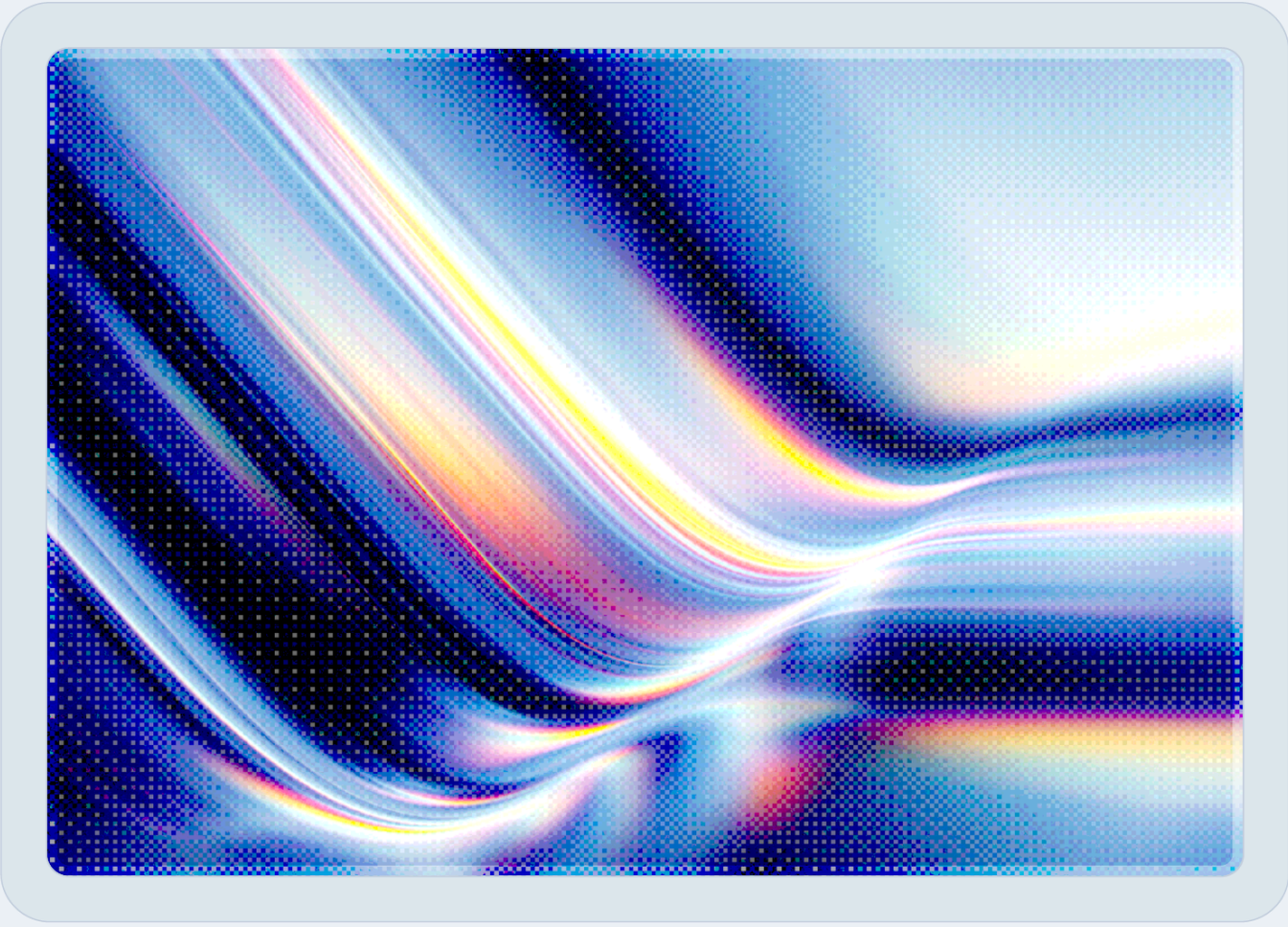
CSAI Foundation | Cloud Security Alliance

Storm-3168: Agentic Cloud Attacks on Azure Identities

2026-09-28

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Key Takeaways

- Microsoft disclosed on September 25, 2026 that the threat actor it tracks as Storm-3168 used two compromised Azure service principals to reconnoiter, then destroy, cloud infrastructure across a single tenant, and assessed the activity as linked to JADEPUFFER, the agentic ransomware first documented by Sysdig in July 2026 [1][4].
- The initial credentials were exposed in plaintext inside a public GitHub issue; Microsoft found that the secret remained retrievable through the issue's edit history even after the original text was redacted, a pattern Microsoft's finding illustrates: removing exposed content does not revoke it [1][2].
- One service principal spent roughly 15.5 hours performing more than 300 successful read operations across subscriptions, resource groups, and storage before a second service principal executed a seven-minute destructive burst that attempted more than 100 storage account deletions along with Key Vault, Function App, and App Service plan deletions [1][3].
- Three IP addresses (45.131.66[.]106, 34.153.223[.]102, 64.20.53[.]230) sending requests with the user agent `python-requests/2.34.2` link this campaign to Storm-3168's infrastructure; two of those addresses also appear in Sysdig's July report on JADEPUFFER's Langflow-based intrusion, a technical indicator suggesting overlapping operator infrastructure between the two disclosures beyond the shared naming convention [1][4].
- Roughly 30 minutes after the destructive burst, the same identities returned to harvest storage account access keys via ListKeys calls, targeting accounts tied to Azure Site Recovery, suggesting an intent to interfere with backup and recovery, not merely to delete production data [1][2].
- Neither Microsoft nor the outlets covering the disclosure could confirm that an autonomous agent executed the Azure-specific operations; the "agentic" characterization currently rests on the precision, tempo, and division of labor across identities rather than on captured agent reasoning traces, a distinction defenders should preserve when briefing leadership [2][3].
- This campaign adds to a growing body of evidence — including CSA's prior research on the non-human identity governance vacuum and on agentic AI operationalized in commodity intrusions — that non-human identities are becoming a dominant post-exploitation target for both human-directed and AI-accelerated intrusions [6][7].

Background

Storm-3168 is the designation Microsoft assigned to a cluster of activity it disclosed on September 25, 2026, in which a threat actor used two compromised Azure service principals from a single tenant to enumerate cloud resources and then destroy them [1]. Microsoft assessed this activity as connected to JADEPUFFER, a name first attached to a separate but related campaign that Sysdig's Threat Research Team documented in July 2026 as the first fully agentic ransomware operation it had observed [4]. In that earlier campaign, an autonomous agent gained initial access to an internet-facing Langflow instance through CVE-2025-3248, an unauthenticated remote code execution vulnerability in Langflow's code-validation endpoint, then independently harvested credentials, pivoted through a Nacos service-discovery platform using a documented default JWT signing key, and encrypted more than 1,300 configuration records before leaving an extortion note [4]. Sysdig's researchers based their "agentic" assessment on the payloads themselves: self-narrating code comments explaining the agent's own reasoning, and a 31-second diagnose-and-correct cycle in which the agent identified why a backdoor account creation had failed and issued a revised, technically distinct payload without any observable human intervention [4]. Even so, a human operator still set up the campaign's infrastructure, provisioned the command-and-control server, and selected the target — a reminder that "agentic" in this context describes the execution of the intrusion, not the absence of a human directing it [5].

The Azure campaign Microsoft disclosed in September follows a similar operational signature but targets a different layer of the cloud stack: identity and resource management rather than application-layer services. Two of the three IP addresses Microsoft attributes to Storm-3168 in its September blog post — 45.131.66[.]106 and 64.20.53[.]230 — match infrastructure Sysdig identified as JADEPUFFER's command-and-control and staging servers in July, a technical indicator, beyond the shared naming convention, suggesting overlapping operator infrastructure between the two campaigns, though IP reuse alone cannot rule out shared third-party hosting [1][4]. Microsoft could not confirm how the second compromised service principal's credentials were obtained, but it did trace the first to a public GitHub issue where the client ID, secret, and tenant ID had been posted in plaintext [1]. Critically, Microsoft found that redacting or deleting the exposed text after the fact did not eliminate the exposure: the secret remained retrievable from the issue's edit history [1][2].

The resulting intrusion unfolded against production Azure workloads spanning storage accounts, SQL databases, Key Vaults, Function Apps, virtual machines, App Services, and backup infrastructure, all within subscriptions belonging to the same tenant [1][3]. Because Storm-3168's operations divided reconnaissance and destruction across two separate identities, and because the timing and sequencing of individual API calls suggest scripted or automated execution rather than manual console interaction, Microsoft and independent commentators have characterized the campaign as consistent with AI-orchestrated tradecraft, even though the Azure-specific evidence made public does not include the kind

of self-narrating payload text that anchored Sysdig's agentic assessment of JADEPUFFER [1][2][3]. That gap matters for how defenders should read the disclosure: the operational tempo and identity discipline are real and actionable signals regardless of whether the destructive phase was driven by a human operator, a scripted tool, or an autonomous agent.

Security Analysis

The attack's structure separated cleanly into three phases, each conducted primarily through direct calls to Azure Resource Manager rather than through the Azure portal, consistent with the `python-requests/2.34.2` user agent Microsoft observed across the activity [1]. The first service principal conducted a long reconnaissance phase — approximately 15 hours and 30 minutes — during which it issued more than 300 successful read operations enumerating subscriptions, virtual machines, resource groups, storage accounts, App Services, and Key Vaults [1][3]. Ninety minutes after that phase ended, a second service principal enumerated resources across two subscriptions in roughly five seconds, a pace inconsistent with manual operation and suggestive of a script or agent iterating through a resource list far faster than a human operator could review results [1][2]. Sixteen hours later, the second identity probed App Service configuration stores and attempted to locate an Azure OpenSearch resource that did not exist, then attempted a ListKeys call against a nonexistent storage account seventy seconds after that — both failures that, taken together, suggest the operator or agent was working from an incomplete or generic target list rather than intelligence specific to the victim environment [1].

Less than one second after that failed ListKeys attempt, the campaign shifted into a seven-minute destructive window. During this phase, the second service principal attempted more than 100 storage account deletions, most of which succeeded, alongside deletions of a Key Vault, a Function App, and an App Service plan [1][2][3]. Parallel attempts to delete Azure SQL databases failed uniformly, but not because of any defensive control — Microsoft attributed the failures to the attacker's use of an unsupported API version, an operational error rather than a blocked action [1]. The campaign also targeted Azure Site Recovery and Backup protection locks, suggesting an intent to weaken recovery options rather than simply delete production data outright [1][2]. Approximately thirty minutes after the destructive burst concluded, the same identities returned for a credential-harvesting phase, issuing more than thirty successful ListKeys requests to retrieve storage account access keys, with particular focus on accounts associated with Site Recovery infrastructure [1][3]. Microsoft's own disclosure summarizes the combined destructive and credential-harvesting activity as more than 150 operations within a 35-minute window, following reconnaissance that had unfolded over more than 15 hours; readers should note that this is Microsoft's stated figure rather than a CSA-derived total, and that the individual

phase timings disclosed elsewhere in Microsoft's post (a 7-minute destructive burst followed by a roughly 30-minute gap before harvesting begins) leave some ambiguity about exactly how the 35-minute window is bounded [1][2][3].

Nick Tausek of Swimlane observed that "at that speed, an AI SOC needs to connect identity activity with cloud changes before the damage spreads," pointing to the mismatch between human-paced detection review and machine-paced destructive execution [3]. Ross Filipek of Corsica Technologies noted that recovery from this class of incident extends well beyond restoring deleted resources: organizations must "establish which identities and keys were touched, then review their use before trusting restored services," since the credential-harvesting phase means that simply rebuilding deleted infrastructure could hand the same access back to the attacker [3]. This point is reinforced by the pattern of targeting: the campaign specifically went after Site Recovery-linked storage accounts and backup protection locks, controls meant to blunt exactly this kind of destructive incident, which is consistent with an intent to interfere with recovery rather than random selection [1][2].

Table 1 summarizes the operational sequence as disclosed by Microsoft, separating each phase's own duration from the elapsed time since the previous phase to keep the timeline unambiguous.

Phase	Phase Duration	Time Since Previous Phase	Key Activity	Volume
Reconnaissance	~15 hours 30 minutes	— (initial phase)	Enumeration of subscriptions, VMs, resource groups, storage, Key Vaults	300+ successful read operations
Secondary enumeration	~5 seconds	~90 minutes after reconnaissance ended	Rapid cross-subscription resource listing by second identity	2 subscriptions enumerated
Pre-destruction probing	Not disclosed	~16 hours after secondary enumeration	App Service config store and OpenSearch enumeration; failed ListKeys	Mixed success/failure
Destruction	7 minutes	<1 second after failed ListKeys	Storage account, Key Vault, Function	100+ deletion attempts

Phase	Phase Duration	Time Since Previous Phase	Key Activity	Volume
			App, App Service plan deletions	
Credential harvesting	Not disclosed	~30 minutes after destruction concluded	ListKeys against storage accounts, focus on Site Recovery	30+ successful requests

Source: Microsoft Security Blog, September 25, 2026 [1].

Recommendations

Immediate Actions

Organizations should treat any credential that has ever appeared in a public repository, issue, pull request, or gist as permanently compromised, regardless of whether the original post was later edited or deleted, and rotate it without delay [1][2]. Security teams should also audit service principals for anomalous read volume and cross-subscription enumeration speed, since the reconnaissance phase in this campaign — hundreds of read operations sustained over many hours, followed by a burst of enumeration in seconds — is a detectable pattern distinct from normal application behavior [1][3]. Finally, teams should verify that deletion locks and Azure Backup protection locks are actually applied to Site Recovery infrastructure and other recovery-critical resources, since this campaign specifically targeted those controls in its final phase [1].

Short-Term Mitigations

Least-privilege scoping for service principals deserves particular attention: a principal that can enumerate an entire tenant's resource inventory and also delete storage accounts, Key Vaults, and Function Apps combines reconnaissance and destructive capability in a single identity, which is precisely the access pattern this campaign exploited [1][3]. Organizations should also enable Microsoft Defender for Cloud coverage across Resource Manager, Storage, Key Vault, App Service, and Databases, since these are the exact services the campaign touched, and should adopt secrets-scanning tooling on source repositories and CI/CD pipelines to catch exposed credentials before they reach a public GitHub

issue or comment [1]. Credential lifecycle practices should shift toward short-lived, workload-attested tokens rather than long-lived static secrets stored in configuration files or environment variables, a recommendation that applies equally to human-directed and AI-accelerated compromise scenarios [1] [6].

Strategic Considerations

The overlap in infrastructure between Sysdig's July disclosure of JADEPUFFER and Microsoft's September disclosure of Storm-3168 suggests that a single operator or closely affiliated group is capable of running both an application-layer agentic ransomware campaign against Langflow and Nacos deployments and a cloud-identity-focused destructive campaign against Azure tenants, potentially reusing tooling, infrastructure, or automation logic across both [1][4]. Security leaders should not wait for definitive proof of autonomous agent involvement before treating the operational tempo itself as the actionable signal: whether a script or an LLM-driven agent executed the destructive phase, the fact that reconnaissance-to-destruction can compress into a seven-minute window means detection and response processes built around human-paced investigation will not keep pace [1][2][3]. Longer term, CSA assesses this campaign as a concrete argument for elevating non-human identity governance — distinct per-service-principal identities, attested short-lived credentials, and usage-based access reviews — to a primary rather than secondary security program, since it was non-human identities, not user accounts or endpoints, that carried the entire attack from initial access through destruction and credential theft [1][6][7].

CSA Resource Alignment

This campaign is best understood through two pieces of CSA research published in 2026. CSA's *The Non-Human Identity Governance Vacuum: AI Agents and the Fastest-Growing Unmanaged Attack Surface* documents an ownership and lifecycle-management vacuum around AI agent and service credentials, reporting that only 15 percent of organizations feel highly confident in their ability to prevent NHI-based attacks, that more than 16 percent do not track the creation of new AI-related identities, and that 51 percent of organizations report no clear ownership of AI identities [6]. Storm-3168 illustrates the operational consequences of that vacuum: the compromised service principal that spent 15.5 hours enumerating tenant resources went undetected long enough for a second, destructive identity to act on its reconnaissance, a failure mode the paper's recommended fixes — cryptographic workload identity, zero-standing-privilege, and just-in-time access — are designed to close [6]. Organizations building or

maturing a non-human identity program in response to this disclosure may also find CSA's *Defining Non-Human Identity*, a foundational guide to classifying and governing NHIs through a Zero Trust lens, a useful companion to the governance-vacuum findings above [9].

CSA's *UAT-10147: Agentic AI Operationalized in Commodity Intrusions*, published in August 2026, analyzed a financially motivated intrusion actor that integrated AI-driven tooling into exploitation and post-compromise workflows against roughly 170,000 internet-exposed URLs, and explicitly situated that activity alongside the GTG-1002 disclosure as evidence that agentic tradecraft is migrating from state-aligned espionage into commodity cybercrime [7]. Storm-3168 and its ties to JADEPUFFER extend that same trajectory into destructive, extortion-oriented cloud operations, reinforcing the paper's core argument that defenders should expect AI-accelerated post-compromise activity to keep broadening beyond isolated, high-profile campaigns into routine criminal operations [7].

Both documents point organizations toward the AI Controls Matrix (AICM) v1.1, whose Identity and Access Management domain provides the control objectives needed to operationalize service-principal least privilege, credential lifecycle management, and attribution requirements referenced throughout this note [8]. Where an organization is building or maturing a non-human identity governance program in response to this disclosure, AICM's IAM domain offers a vendor-neutral baseline for assessing whether service principal privileges, credential rotation practices, and monitoring coverage are adequate to the tempo this campaign demonstrated.

References

[1] Microsoft Security. "Storm-3168: Agentic-driven cloud attacks using compromised service principals." Microsoft Security Blog, September 25, 2026.

[2] Security Affairs. "Storm-3168, Linked to JADEPUFFER, Abused Stolen Azure Identities." Security Affairs, September 2026.

[3] CSO Online. "Autonomous agents attack Azure using compromised identities and destroying resources." CSO Online, September 2026.

[4] Sysdig Threat Research Team. "JADEPUFFER: Agentic ransomware for automated database extortion." Sysdig, July 2026.

[5] TechCrunch. "The 'first' AI-run ransomware attack still needed a human." TechCrunch, July 6, 2026.

[6] Cloud Security Alliance AI Safety Initiative. "The Non-Human Identity Governance Vacuum: AI Agents and the Fastest-Growing Unmanaged Attack Surface." CSA Labs, May 20, 2026.

[7] Cloud Security Alliance AI Safety Initiative. "UAT-10147: Agentic AI Operationalized in Commodity Intrusions." CSA Labs, August 21, 2026.

[8] Cloud Security Alliance. "AI Controls Matrix (AICM) v1.1." Cloud Security Alliance, 2026.

[9] Cloud Security Alliance. "Defining Non-Human Identity." Cloud Security Alliance, July 22, 2026.

