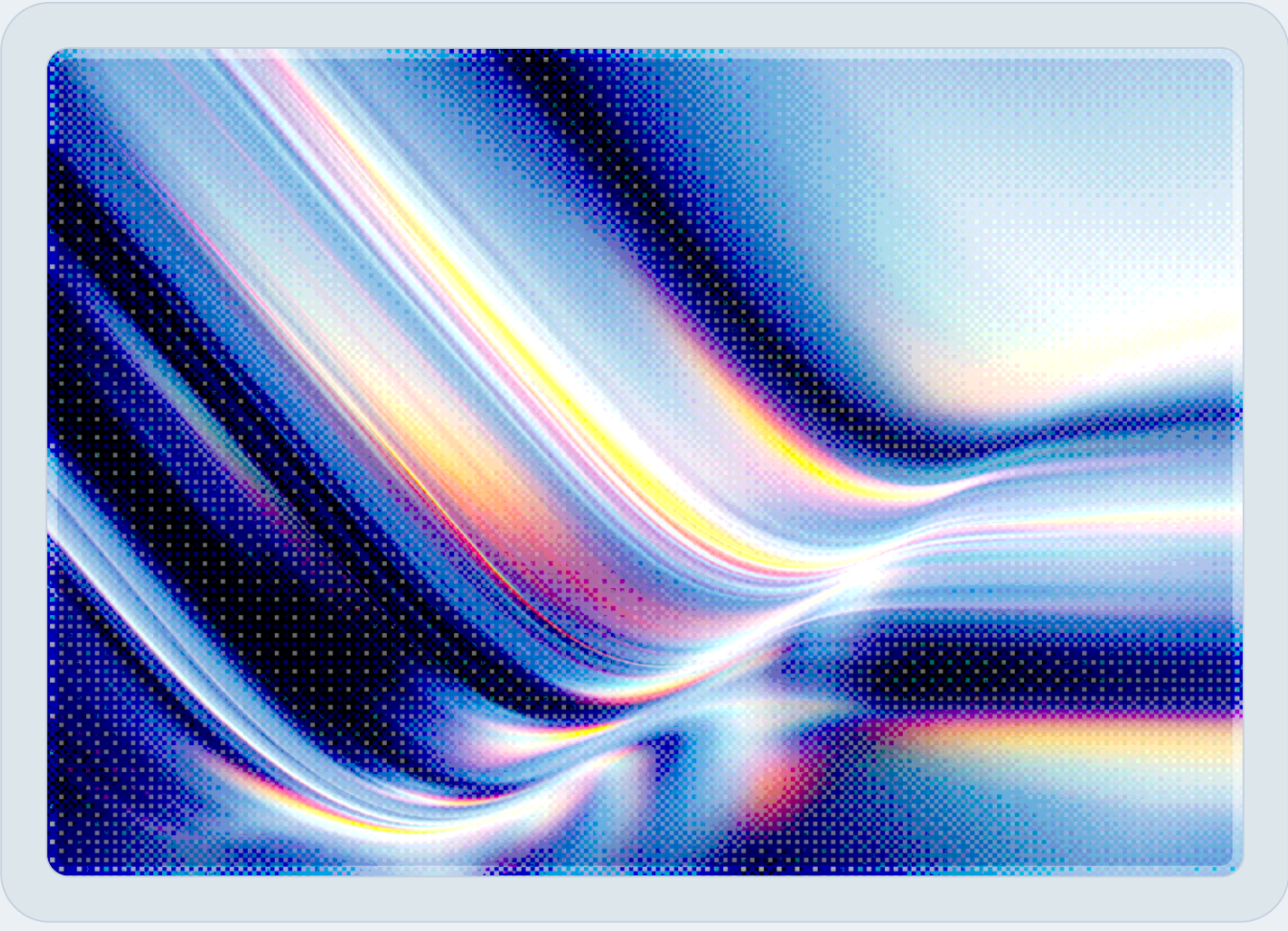
CSAI Foundation | Cloud Security Alliance

OAuth Consent Abuse: ConsentFix v3 and the SaaS Extortion Economy

AiTM Phishing, MFA Bypass, and Persistent SaaS Access Compromise

2026-05-05

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Key Takeaways

- ConsentFix v3, documented in active circulation as of May 2026, automates OAuth authorization-code phishing against Microsoft Azure and Microsoft 365, bypassing multi-factor authentication entirely by abusing pre-consented first-party applications rather than stealing passwords.
- Adversary-in-the-Middle (AiTM) phishing campaigns are converging with OAuth consent abuse, enabling real-time session token interception at scale. Microsoft reported a single campaign in April 2026 that targeted more than 35,000 users across over 13,000 organizations in 26 countries [1].
- For SaaS-native OAuth token intrusions, the extortion model no longer depends on ransomware encryption—persistent access to cloud data provides equivalent leverage through the threat of regulatory exposure under GDPR, HIPAA, and state breach notification statutes.
- Microsoft's July 2025 change to default Entra ID consent policies—restricting end-user consent to high-risk scopes such as `Files.ReadWrite.All` and `Sites.ReadWrite.All`—mitigates one class of consent abuse but does not address the authorization-code phishing used by ConsentFix v3 [2].
- Phishing-resistant MFA (FIDO2 hardware keys, device-bound passkeys) defeats AiTM attacks in their standard form but offers no protection against ConsentFix, where the victim completes a genuine authentication flow. Organizations should deploy phishing-resistant MFA broadly while adding token-lifecycle controls for the ConsentFix vector.
- Security teams must shift from credential-centric controls to token-lifecycle management, combining conditional access policies with continuous monitoring of OAuth application grants and refresh token activity.

Background

OAuth 2.0 is the foundational authorization protocol underpinning nearly every enterprise SaaS integration. When a user clicks "Sign in with Google" or "Allow access" for a third-party productivity application, they are executing an OAuth consent flow that grants a token—often including long-lived

refresh tokens—enabling the application to act on their behalf. This design is intentional and necessary for modern SaaS interoperability, but it creates a persistent authorization surface that attackers have learned to exploit independently of passwords or MFA challenges.

Consent phishing as a category emerged from early campaigns against Microsoft 365 tenants, in which attackers registered malicious applications in Microsoft Entra ID and socially engineered users into granting them delegated permissions. Microsoft and the security research community documented financially motivated actors using this approach to automate cryptocurrency mining, business email compromise, and data exfiltration as early as 2023 [3]. The key insight that made consent phishing dangerous was the same one that makes it difficult to eradicate: a legitimately completed OAuth flow produces tokens that are, from the platform's perspective, entirely valid. No credential theft occurs, and no password is ever transmitted to an attacker-controlled server.

The technique's attack surface expanded substantially in late 2025 when Push Security documented a variant they named ConsentFix, a browser-native technique that fused the social engineering mechanics of the ClickFix malware distribution style with OAuth authorization-code interception [4]. Rather than directing users to authorize a malicious application, ConsentFix abused Microsoft's Azure CLI—a first-party, pre-trusted application that requires no admin consent in most tenant configurations—to complete a real authentication flow and then capture the resulting authorization code before Microsoft's token exchange service could consume it. Because the victim is completing a genuine Microsoft authentication, MFA challenges are satisfied by the user themselves, and the attack succeeds regardless of authentication strength, including hardware tokens and passkeys [5].

Two subsequent iterations refined the user experience of the attack. ConsentFix v2, attributed to security researcher John Hammond who developed it as a proof-of-concept, replaced the original copy-paste step—where users were instructed to paste a localhost URL into an attacker-controlled field—with a drag-and-drop interaction designed to reduce victim abandonment by simplifying the interaction step [16]. By early May 2026, a third iteration had emerged on underground forums and was confirmed in active use by multiple threat intelligence vendors.

Security Analysis

ConsentFix v3: Automation and Scalability as a Service

ConsentFix v3 preserves the core of its predecessors—abusing the OAuth 2.0 authorization code flow against pre-consented Microsoft first-party apps—but introduces an automated backend pipeline that transforms the technique from a manually operated attack into an operationally scalable one [6]. The

toolkit has been advertised on underground forums with promotional materials [6][7], reflecting the broader commoditization of offensive tooling into Phishing-as-a-Service models.

The attack chain begins with automated target profiling. The toolkit verifies Azure tenant presence, harvests employee names, roles, and email addresses from public and semi-public sources, and constructs highly personalized phishing emails. To increase delivery rates and evade link-scanning defenses, malicious links are embedded inside PDF documents hosted on legitimate document-sharing platforms—BleepingComputer has documented DocSend in particular as a hosting platform favored in these campaigns [7], chosen in part because such services may be allowlisted by enterprise email security gateways due to their widespread legitimate use. The phishing page itself is hosted on Cloudflare Pages, again leveraging a trusted infrastructure provider to defeat URL-reputation controls.

Once a victim reaches the phishing page and authenticates through the embedded Azure CLI authorization flow, their browser is redirected to a localhost URL containing the OAuth authorization code. The page instructs the user to paste or drag this URL into a form field presented as a required verification step. The moment the URL is submitted, the page forwards it to a Pipedream webhook, and an automated backend immediately exchanges the authorization code for a full token set—including an access token and a long-lived refresh token—before Microsoft's authorization server has the opportunity to detect anomalous redemption patterns. The harvested tokens are then imported into a post-exploitation platform described in underground forum communications reported by BleepingComputer [7] as "Specter Portal," which provides a management interface for interacting with compromised Microsoft 365 and Azure environments.

The persistence risk introduced by refresh tokens is particularly significant. Access tokens typically expire within an hour, but refresh tokens may remain valid for days, weeks, or—in certain application configurations—indefinitely. An attacker holding a valid refresh token can continuously mint fresh access tokens without triggering a new authentication event, enabling sustained access to email, SharePoint, OneDrive, and any other resource within the victim's Microsoft 365 tenancy that the Azure CLI is authorized to access.

AiTM Campaigns: Token Interception at Scale

Concurrent with the ConsentFix lineage, a separate class of OAuth abuse operates through Adversary-in-the-Middle phishing infrastructure. AiTM attacks differ from ConsentFix in mechanism but converge on the same outcome: session token capture that bypasses MFA. In an AiTM scenario, the phishing page acts as a reverse proxy, relaying the victim's real Microsoft authentication to Microsoft's servers while simultaneously capturing the session cookie or token material returned to the browser.

On May 4, 2026, Microsoft's Defender Research team published an analysis of a multi-stage AiTM campaign that had operated between April 14 and 16, targeting more than 35,000 users across more than 13,000 organizations in 26 countries, with approximately 92 percent of targeted accounts belonging to organizations headquartered in the United States [1]. The campaign used a pretextual "code of conduct review" theme, delivering phishing lures as PDF attachments with filenames evoking disciplinary and HR processes. Victims were directed to click a "Review Case Materials" link within the PDF, landing on a page that displayed a Cloudflare CAPTCHA—functioning as an anti-analysis gate to impede sandbox detonation—before initiating the AiTM session interception. The sectors most heavily targeted were healthcare and life sciences, financial services, professional services, and technology, reflecting attacker prioritization of data-rich environments with high regulatory exposure and therefore high extortion leverage.

A parallel campaign documented by Microsoft in January 2026 targeted energy sector organizations using a multi-stage AiTM flow that abused SharePoint for intermediate document delivery, adding an additional hop designed to exploit the implicit trust organizations extend to SharePoint-hosted content [8]. These campaigns collectively illustrate how AiTM infrastructure has matured from experimental proof-of-concept into production-grade, multi-stage delivery chains capable of operating at enterprise scale.

The SaaS Extortion Economy

The downstream monetization of OAuth token abuse has diverged meaningfully from the encryption-ransomware model that dominated cybercriminal revenue streams throughout the 2020s. Rather than deploying destructive payloads that lock systems and demand cryptocurrency for decryption keys, actors operating against SaaS environments increasingly favor pure extortion: exfiltrating sensitive data held in email, CRM, HR, and collaboration platforms, then threatening to publish it on data-leak sites unless a payment is made. This model imposes lower operational risk on attackers—no malware implant to detect, no lateral movement through endpoint controls—while generating substantial leverage by targeting data whose exposure triggers regulatory penalties under GDPR, HIPAA, and state breach notification statutes.

The extortion infrastructure supporting this model has grown substantially. ReliaQuest's Q1 2026 threat intelligence report recorded 2,638 posts to ransomware and extortion data-leak sites during the quarter, a 22 percent increase over the same period in 2025, with 91 active leak sites at the end of the quarter—a record high [9]. The proliferation of active sites suggests that barriers to entry for new extortion groups have fallen, with threat intelligence consistently documenting the increasing commoditization of supporting infrastructure [9]. Groups such as ShinyHunters have demonstrated publicly that identity-first

intrusions leveraging SaaS-native access can achieve major operational impact without any endpoint malware, suggesting that endpoint-centric detection alone is insufficient for identity-first SaaS intrusions of this type.

Supply chain amplification compounds the individual account compromise risk significantly. In August 2025, threat actor UNC6395 exploited OAuth tokens from Salesloft's Drift AI Chat integration to access customer Salesforce environments across more than 700 organizations, including prominent technology and security vendors [10]. The mechanism was straightforward: Drift, like many SaaS productivity tools, held OAuth refresh tokens granting it broad Salesforce API access. Compromising that OAuth credential store yielded transitive access to every connected customer environment. This supply chain vector—where a single OAuth integration breach propagates access across an entire ecosystem—is precisely the risk model that conventional third-party security assessments are not designed to surface, because the permissions are legitimate and the integration is authorized.

Why Existing Controls Fall Short

The July 2025 Microsoft Entra ID default consent policy change, which required admin approval for any application requesting `Files.Read.All`, `Files.ReadWrite.All`, `Sites.Read.All`, or `Sites.ReadWrite.All` scopes, represents a meaningful step toward reducing the blast radius of malicious application registrations [2]. However, ConsentFix v3 does not rely on registering a new third-party application or requesting these high-risk permissions. It abuses the Azure CLI, a Microsoft first-party application that is pre-consented across virtually all Entra ID tenants and requires no user-facing consent prompt. The managed consent policy therefore does not impede the ConsentFix attack path, and organizations that have implemented these controls may have an incomplete picture of their OAuth exposure.

Conditional access policies offer stronger protection but require careful configuration to address token-based attacks specifically. Policies that enforce compliant-device checks for token redemption, bind access tokens to specific IP ranges or named locations, and enable continuous access evaluation can reduce the dwell time and lateral mobility of a stolen token. However, these controls can be bypassed if the attacker operates from an infrastructure IP that falls within policy tolerances, and they do not address the authorization-code interception window in which ConsentFix v3 operates.

Recommendations

Immediate Actions

Security teams commonly lack consolidated visibility into the OAuth grants active in their Microsoft 365 and Azure environments—a gap that extends to grants made by former employees whose accounts have since been deprovisioned but whose delegated tokens may remain active. Administrators should audit Entra ID enterprise application consent grants immediately, revoke tokens associated with unfamiliar applications, and implement an admin consent workflow that routes all new application consent requests through a review queue.

Organizations should also enumerate all first-party Microsoft applications that are pre-consented in their tenancy—particularly the Azure CLI, Azure PowerShell, and Microsoft Graph Explorer—and evaluate whether end users have legitimate operational need for these applications. Where feasible, restricting the Azure CLI to specific named user accounts or enforcing compliant-device binding for its token redemptions should reduce the ConsentFix attack surface meaningfully, limiting the population of potential victims to accounts and devices that meet organizational policy. Microsoft's Entra ID provides application-level conditional access policies that can enforce these constraints without requiring broader policy changes.

Short-Term Mitigations

Refresh token lifetime management should be reviewed and tightened across all Microsoft 365 tenancy configurations. Microsoft allows organizations to configure token lifetime policies that reduce how long refresh tokens remain valid for non-compliant devices or high-risk sign-in contexts. Shortening refresh token windows for unmanaged devices and enforcing continuous access evaluation policies—which allow Microsoft Entra ID to revoke tokens in real time when anomalous activity is detected—limits the persistence window available to an attacker who has captured a token via ConsentFix or AiTM techniques.

Email security controls should be updated to treat PDF attachments containing links to document-sharing platforms with additional scrutiny. ConsentFix v3 and related AiTM campaigns consistently use legitimate document-hosting infrastructure as an intermediate delivery layer to evade URL reputation defenses. Security awareness training should be updated to emphasize that legitimate employer HR, compliance, and legal communications will not require employees to authenticate through unfamiliar web pages or copy URL strings from their browser.

Detection engineering teams should add alerting for OAuth authorization code redemptions that occur from IP addresses, ASNs, or geographic locations inconsistent with the user's normal authentication pattern. Authorization codes have a short validity window by design, but the exchange event produces a log entry in Entra ID's sign-in logs that can be correlated against the preceding authentication event's client characteristics. Anomalies in this correlation—where the code was generated from one location and redeemed from another—are a strong candidate indicator of ConsentFix-style interception that warrants investigation, particularly where other user risk signals are present.

Strategic Considerations

The SaaS supply chain OAuth risk illustrated by the UNC6395 campaign requires a governance response, not only a technical one. Organizations should implement a formal SaaS application review process that includes an OAuth scope inventory for every integrated application, periodic refresh token rotation requirements for third-party integrations, and contractual language requiring SaaS vendors to notify customers promptly of any OAuth credential compromise. CSA's SaaS Security Capability Framework (SSCF) provides a structured controls vocabulary for codifying these requirements across vendor relationships [11].

Phishing-resistant authentication remains a necessary long-term investment. FIDO2 hardware security keys and device-bound passkeys are not defeated by AiTM proxy attacks in their standard form, because the cryptographic binding is to the specific origin the user is authenticating to, and a proxy presents a different origin than Microsoft. Organizations that have deployed phishing-resistant MFA broadly should verify that their conditional access policies require it uniformly, without fallback paths that an attacker could force a user down. Gaps in enforcement—such as legacy authentication protocols, shared mailboxes, or service accounts using password-based flows—represent bypass paths that sophisticated actors routinely enumerate.

CSA Resource Alignment

The threat landscape described in this research note maps directly to guidance and frameworks maintained by the Cloud Security Alliance across multiple programs.

CSA's **Cloud Controls Matrix (CCM) v4.1** addresses these risks through its Identity and Access Management (IAM) domain, which specifies controls for credential management, access authorization review, and third-party access governance [12]. The IAM-07 control requiring periodic review and

recertification of access rights applies directly to OAuth application grant management, and the IAM-09 control on user access provisioning and de-provisioning should be extended to cover delegated application authorizations.

The **SaaS Security Capability Framework (SSCF)**, published by CSA in collaboration with the SaaS security research community, provides a structured control vocabulary for assessing SaaS provider security posture and customer-side configuration requirements [11]. Supply-chain OAuth risk—where a compromised integration provider exposes connected customer environments—falls within the SSCF's third-party access and integration security capability domain.

CSA's **Zero Trust guidance**, including the published white paper "Using Zero Trust to Counter Identity Spoofing and Abuse," provides foundational principles directly applicable to token-based attacks: never-trust-always-verify semantics for token redemption, continuous session validation rather than one-time authentication, and micro-segmentation of access that limits the blast radius of any single token compromise [13]. The principle of least-privilege access is particularly relevant to OAuth scope management, where applications routinely request broader permissions than their core function requires.

The **MAESTRO agentic AI threat modeling framework** offers a forward-looking dimension to this discussion. As AI agents are increasingly granted OAuth credentials to operate autonomously within SaaS environments—reading email, updating CRM records, scheduling calendar events—they represent non-human identities that are subject to the same consent phishing and token theft vectors described here [14]. MAESTRO's Layer 4 (Tool and API Integration) specifically addresses the risks of delegated agent access, and its adversarial threat models for credential hijacking and authorization abuse should be incorporated into any AI agent deployment that operates with OAuth-scoped SaaS access. CSA's companion publication on Agentic AI Identity and Access Management provides detailed architectural guidance for implementing least-privilege, auditable delegation chains for AI agents [15].

References

[1] Microsoft Defender Research Team. "Breaking the Code: Multi-stage 'Code of Conduct' Phishing Campaign Leads to AiTM Token Compromise." Microsoft Security Blog, May 4, 2026.

[2] Microsoft Entra Blog. "OAuth Consent Phishing Explained and Prevented." Microsoft Community Hub, 2025.

[3] Microsoft Security Blog. "Threat Actors Misuse OAuth Applications to Automate Financially Driven Attacks." Microsoft, December 12, 2023.

[4] Push Security. "ConsentFix: Browser-native ClickFix Hijacks OAuth Grants." Push Security Blog, December 2025.

[5] Arctic Wolf. "New Attack Technique 'ConsentFix' Hijacks OAuth Consent Grants." Arctic Wolf Blog, 2025.

[6] Push Security. "ConsentFix v3: Analyzing a new criminal toolkit." Push Security Blog, May 2026.

[7] Bleeping Computer. "ConsentFix v3 Attacks Target Azure with Automated OAuth Abuse." BleepingComputer, May 2026.

[8] Microsoft Security Blog. "Resurgence of a Multi-stage AiTM Phishing and BEC Campaign Abusing SharePoint." Microsoft, January 21, 2026.

[9] ReliaQuest. "Ransomware and Cyber Extortion in Q1 2026." ReliaQuest, 2026.

[10] Google Cloud Threat Intelligence. "Data Theft of Salesforce Instances via Salesloft Drift Integration." Google Cloud Blog, 2025.

[11] Cloud Security Alliance. "SaaS Security Capability Framework (SSCF)." CSA, September 2025.

[12] Cloud Security Alliance. "Cloud Controls Matrix v4.1." CSA, 2021.

[13] Cloud Security Alliance. "Using Zero Trust to Counter Identity Spoofing and Abuse." CSA, 2024.

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

[15] Cloud Security Alliance. "Agentic AI Identity and Access Management: A New Approach." CSA, 2025.

[16] Push Security. "ConsentFix Debrief." Push Security Blog, 2025.