Starkiller: Industrializing Real-Time MFA Bypass at Scale

Unofficial AI-assisted Research

Cloud Security Alliance AI Safety Initiative

2026-03-21

Key Takeaways

- Starkiller is a commercially operated Phishing-as-a-Service (PhaaS) framework that uses a headless browser running inside a Docker container to proxy real login pages in real time, defeating both signature-based detection and legacy multi-factor authentication mechanisms including SMS one-time passwords, TOTP authenticator apps, and push notifications [3][4].
- Adversary-in-the-Middle (AiTM) phishing attacks do not break or circumvent MFA; they relay the legitimate authentication handshake transparently and capture the resulting session cookie, giving attackers persistent authenticated access without ever needing the victim's password or OTP again [3][4][5].
- The Phishing-as-a-Service ecosystem has matured into a multi-tiered criminal industry with subscription pricing, customer support, analytics dashboards, and continuous feature development — dramatically lowering the technical barrier for enterprise-grade account takeover campaigns [1][7].
- The March 2026 international law enforcement action that seized 330 domains associated with the Tycoon 2FA platform — the dominant PhaaS operator prior to disruption — demonstrates meaningful progress in supply-side disruption, but platform fragmentation means successor tools including Starkiller, Sneaky 2FA, and Rockstar 2FA have continued filling the vacuum [1].
- FIDO2 hardware security keys and device-bound passkeys are among the authenticator categories that are cryptographically immune to AiTM relay attacks, because authentication responses are bound to the legitimate domain's origin and cannot be relayed through a proxy; certificate-based authentication provides equivalent protection in organizations with established PKI infrastructure. Migration to phishing-resistant MFA should be treated as an urgent infrastructure priority, not a roadmap item [19][17].

Background

Multi-factor authentication was designed to address password theft, and for most of the past decade it succeeded. Attackers who obtained a user's password through data breaches, infostealer malware, or credential stuffing could not complete authentication without access to the victim's second factor. That assumption has eroded as AiTM attacks have scaled from specialist tradecraft to commercially available services. The Adversary-in-the-Middle attack model, which has existed in research circles since at least 2017 [10], had by 2023 moved from requiring significant technical expertise to being available as commercial subscription services — a substantial lowering of attacker barriers.

AiTM phishing reframes the attack objective. Rather than stealing a password, the attacker inserts infrastructure between the victim and the legitimate identity provider, relaying the full authentication exchange in real time and capturing the session cookie issued at the conclusion of a successful login. Because the victim has completed genuine authentication — including any MFA challenge — the session token is valid, fully authenticated, and usable by the attacker immediately. Traditional MFA controls were not designed to prevent this class of attack because they authenticate the user, not the communication channel. SMS one-time passwords, TOTP codes from authenticator applications, and push notification approvals all operate on the same vulnerable premise: they prove the user responded to a challenge, but they do not bind that proof to the specific domain or device that will consume the resulting credential.

The ecosystem that has grown around this attack primitive has industrialized it thoroughly. Phishing kits with AiTM capabilities are now sold on underground forums with tiered pricing, feature comparison matrices, Telegram-based customer support, and monthly release cycles. Security researchers have documented a rapidly expanding catalog of distinct PhaaS platforms since 2022 [1][7], with the most prominent operators — Tycoon 2FA, EvilProxy, Rockstar 2FA, and Sneaky 2FA — collectively responsible for campaigns affecting hundreds of millions of phishing messages and hundreds of thousands of documented victims [1][2]. The March 2026 coordinated takedown of Tycoon 2FA by Europol and law enforcement from six European countries and the United Kingdom, supported by Microsoft, Proofpoint, Cloudflare, and other private sector partners, represents the most significant supply-side disruption of a PhaaS platform to date [1]. The emergence of Starkiller in the months surrounding that disruption illustrates how quickly the market reconstitutes.

Security Analysis

Starkiller: Technical Architecture

Starkiller's headless browser architecture diverges in an operationally significant way from the reverse-proxy model used by earlier AiTM toolkits. Where prior-generation phishing frameworks such as Evilginx and Modlishka operate as reverse-proxy servers that intercept and relay HTTP traffic between the victim and the legitimate identity provider, Starkiller's architecture centers on a headless Chrome browser instance running inside a Docker container [3][4]. This distinction has significant defensive implications.

Traditional reverse-proxy AiTM kits must faithfully reconstruct and relay HTTP responses from the upstream identity provider, including dynamically generated page content, JavaScript, and Content Security Policy headers. This reconstruction process creates detectable artifacts: subtle rendering differences, modified resource paths, and inconsistent header signatures that defenders and abuse detection systems can learn to recognize. Starkiller's headless browser approach eliminates this surface entirely. The victim interacts with what is, for all observable purposes, the genuine login page — because the page is being rendered live by a real Chrome process against the real upstream service. The attacker infrastructure observes the session through the browser runtime rather than through HTTP interception, capturing keystrokes, session tokens, and screen state as they occur in the live browser context [4][5].

The platform is commercially operated — attributed by researchers to a threat actor group operating under the name Jinkusu, which maintains an active subscriber forum, distributes monthly framework updates, and provides customer support through Telegram [3][4]. Pricing and access details are consistent with the broader PhaaS market, where monthly subscription costs for comparable platforms range from $200 to $350 [7]. Documented campaigns include a June 2025 credential-harvesting wave targeting financial institutions and a November 2025 campaign that added several evasion layers: spoofed `.co.com` domains, Cloudflare-branded CAPTCHA lure pages, Base64-encoded payload scripts, and referrer validation to filter automated analysis crawlers from genuine victim sessions [5].

Starkiller's capabilities extend beyond session cookie capture. The platform includes real-time keylogging across all form fields presented during the proxied session, live screen streaming of the victim's interaction, geolocation tracking and browser fingerprinting of incoming sessions, automated Telegram alerts when new credentials arrive, and campaign analytics covering visit counts and conversion rates [3][4]. The platform also incorporates a specialized module targeting 1Password users, capturing one-time passwords and account recovery codes in addition to primary credentials. Separately, operators have used the platform to exploit the OAuth 2.0 device authorization flow defined

in RFC 8628, directing victims to legitimate `microsoft.com/devicelogin` pages where they unknowingly authorize attacker-controlled OAuth applications — a design abuse that does not require any malicious domain in the visible URL [5].

The Broader PhaaS Landscape

Starkiller exists within a crowded and competitive market. The Phishing-as-a-Service ecosystem has consolidated around several dominant platforms while continuously generating new entrants. Tycoon 2FA, which researchers described as the dominant PhaaS operator during its operational period, relied on CAPTCHA lure pages, QR codes embedded in PDF attachments, SVG file lures, and URL shorteners to route victims through its reverse-proxy infrastructure targeting Microsoft 365 and Gmail accounts [1] [2]. Its disruption in March 2026 — which included seizure of approximately 330 domains and a civil lawsuit in the U.S. Southern District of New York against an alleged operator — removed a major supplier, but CrowdStrike's post-takedown analysis documented that modified Tycoon 2FA infrastructure persisted following the law enforcement action [2].

Rockstar 2FA, which researchers trace to the earlier DadSec and Phoenix phishing kit lineage, has operated since May 2024 and employs an AiTM reverse proxy targeting Microsoft 365, presenting victims with Cloudflare Turnstile CAPTCHA challenges before loading a spoofed but dynamically branded login page [7]. Sneaky 2FA, active through 2025, added Browser-in-the-Browser capability in November 2025 — presenting victims with a convincing pop-up window containing a fake browser address bar displaying a legitimate Microsoft domain, adapting its rendering to match the victim's browser and operating system [8]. Mamba 2FA targets not only Microsoft Entra ID and Active Directory Federation Services but third-party SSO providers and consumer Microsoft accounts, with instant credential relay to operator-controlled Telegram bots [9].

The open-source tools from which many commercial PhaaS platforms derive their architecture — Evilginx, Modlishka, and Muraena — remain freely available and widely deployed by both threat actors and penetration testers. Evilginx is particularly capable among the open-source options, hosting its own DNS server, auto-provisioning Let's Encrypt TLS certificates, and supporting simultaneous configuration of multiple phishing targets [10]. Its legitimate red-team origins have made it a common starting point for actors entering the AiTM space.

Why MFA Doesn't Protect Against This Class of Attack

The critical insight is that AiTM attacks do not attack MFA — they use it. When a victim navigates to an AiTM proxy, the following sequence occurs: the victim submits credentials, which are relayed to the real identity provider; the identity provider issues an MFA challenge, which is relayed back to the victim; the

victim satisfies the challenge (whether by entering an OTP or approving a push notification), and that response is relayed forward to the identity provider; the identity provider issues a session cookie confirming successful, MFA-verified authentication; and the proxy intercepts that cookie before it reaches the victim's browser. The attacker now holds a valid authenticated session. The victim may or may not notice that no expected post-authentication page appeared; many AiTM deployments show the victim a plausible completion screen or redirect them to a benign landing page.

This attack chain is equally effective against all forms of knowledge-based or possession-based MFA: SMS codes, TOTP authenticator app codes, and push notification approvals all operate on a challenge-response model that an in-path attacker can relay without modification. Documented breach campaigns attributed to groups such as Scattered Spider — responsible for attacks on MGM Resorts and Caesars Entertainment in September 2023 and a wave of UK retailer compromises in April–May 2025, among others — demonstrate that this technique scales to high-profile enterprise environments with mature security programs [11][12].

Detection Signals

AiTM infrastructure does produce observable signals in enterprise telemetry, though they require behavioral correlation rather than signature matching. Successful logins from IP addresses with no prior history for that user, particularly when the login occurs within seconds of a link click in email, warrant investigation. Concurrent authenticated sessions from geographically separated locations may indicate that a session token was captured and is being used from attacker infrastructure. High-privilege actions — inbox forwarding rule creation, application password registration, OAuth consent grants — occurring within minutes of authentication are frequently post-compromise automation. Microsoft Sentinel includes an analytic rule (`PossibleAiTMPhishingAttemptAgainstAAD`) that correlates sign-in anomalies consistent with AiTM activity, and Microsoft Defender Threat Intelligence provides IOC enrichment for known phishing infrastructure [13].

QR codes embedded in PDF attachments and SVG file attachments containing embedded URLs have emerged as common delivery mechanisms in PhaaS campaigns specifically because they evade URL scanning in email security gateways [2]. Conditional Access policies that require device compliance prior to token issuance represent a defensive control that breaks AiTM at the infrastructure level: an attacker's proxy cannot satisfy device compliance requirements, and the identity provider will therefore not issue a usable session token even if the victim completes authentication.

Recommendations

Immediate Actions

Enterprises should audit their Conditional Access policies to ensure that sensitive workloads — at minimum, administrative interfaces, email platforms, and financial applications — require either a compliant device or a phishing-resistant authenticator before issuing tokens. This significantly raises the bar for post-authentication token capture, as an attacker's proxy infrastructure cannot satisfy device compliance requirements and the identity provider will decline to issue a token — provided device compliance policy is correctly scoped and enforced across all access paths. Organizations running Microsoft Entra ID should evaluate Microsoft Token Protection, which cryptographically binds Primary Refresh Tokens to the registered device and renders intercepted tokens non-portable [18].

Email security controls should be updated to scan QR codes in PDF attachments and evaluate SVG file attachments for embedded URLs, which are delivery vectors specifically designed to bypass conventional URL scanning. Where business justification for SVG file delivery does not exist, blocking these attachment types reduces attack surface.

Phishing simulation programs should be updated to include AiTM-style lures. Traditional phishing awareness training focused on identifying fake login pages does not prepare users to recognize an AiTM attack, where the page they interact with is the genuine article. Training should address the concept that a convincing login experience does not guarantee the absence of an attacker.

Short-Term Mitigations

Organizations should accelerate deployment of FIDO2 security keys or device-bound passkeys for any population of users with elevated privilege, access to sensitive data stores, or administrative rights over identity infrastructure. FIDO2 authentication is cryptographically bound to the requesting domain's origin at the hardware level; a relay proxy receives a response that is valid only for the legitimate domain and cannot be forwarded to complete an authentication attempt against attacker-controlled infrastructure [14]. Google has publicly attributed the absence of phishing-related account compromises among its approximately 85,000-person workforce to universal security key deployment — a case study that provides substantial real-world evidence for this control [19][17]. Microsoft reached general availability for passkeys in Microsoft Authenticator in January 2025 and announced Windows Hello-bound passkeys in Entra ID for worldwide preview in March 2026 [15][16].

Identity governance review processes should be tuned to detect and alert on OAuth application consent grants, particularly for applications requesting access to mail, calendar, and file storage resources. AiTM campaigns that exploit the OAuth device authorization flow leave a consent grant record in the identity provider that may be the only audit artifact of the compromise.

Strategic Considerations

The PhaaS market's trajectory suggests that AiTM capability will become more widely available, more reliable, and more automated over the planning horizon. Starkiller's headless browser architecture is not unique to this platform — similar approaches appear across the PhaaS landscape — and the pattern of rapid reconstitution following law enforcement actions suggests that supply-side disruption alone is unlikely to be sufficient. Enterprise identity security strategy should treat phishing-resistant authentication as a baseline infrastructure requirement rather than an aspirational control, and migration timelines should reflect that distinction.

Federated identity architectures deserve particular scrutiny. Compromised Okta or Entra ID administrator accounts can be used to register attacker-controlled identity providers through inbound federation, enabling impersonation of arbitrary users across the entire tenant without any further credential theft. The MGM Resorts attack in September 2023 illustrated this path: a compromised Okta super-administrator account was leveraged to establish federated authentication trust with attacker infrastructure, providing persistent access to downstream applications [12]. Privileged identity management controls — including just-in-time access, hardware-bound administrator credentials, and continuous monitoring of federation trust configuration — should be treated as coequal with perimeter MFA deployment.

CSA Resource Alignment

This analysis intersects with several active CSA frameworks and research initiatives. The CSA Cloud Controls Matrix (CCM), and more comprehensively the AI Controls Matrix (AICM) as a superset addressing AI-adjacent identity risks, includes Identity and Access Management control domains covering multi-factor authentication requirements, session management, and privileged access governance. The controls at stake in AiTM campaigns — IAM-02 through IAM-09 in CCM v4.0 — directly address the authentication chain integrity and session lifecycle management that AiTM attacks exploit.

The CSA Zero Trust Advancement Center's continuous verification model is directly implicated. Zero trust architecture assumes that a valid session credential does not imply the legitimacy of the ongoing session, and requires continuous re-evaluation of device posture, behavioral signals, and access context throughout a session's life. This assumption aligns with the detection approach most effective against AiTM: behavioral correlation of post-authentication activity rather than reliance on the authentication event itself as a trust boundary.

The CSA MAESTRO framework for agentic AI threat modeling is increasingly relevant to PhaaS as AI-assisted automation enters both the attack and defense sides of this domain. Emerging PhaaS platforms are beginning to incorporate automated follow-on capabilities — inbox rule creation, data exfiltration scripting, and lateral movement preparation — that execute immediately after session token capture. MAESTRO's threat model for autonomous agent behavior and trust boundary enforcement applies to these automated post-compromise chains.

CSA STAR program participants should evaluate their attestations in IAM and incident response domains against the AiTM threat model specifically, as the control evidence expected for traditional MFA deployments does not necessarily reflect the controls required to address phishing-resistant authentication.

References

[1] Proofpoint Threat Research, "Disruption Targets Tycoon 2FA, Popular AiTM PhaaS," Proofpoint Blog, March 2026. https://www.proofpoint.com/us/blog/threat-insight/disruption-targets-tycoon-2fa-popular-aitm-phaas

[2] CrowdStrike Intelligence, "Tycoon2FA Phishing-as-a-Service Platform Persists Following Takedown," CrowdStrike Blog, March 2026. https://www.crowdstrike.com/en-us/blog/tycoon2fa-phishing-as-a-service-platform-persists-following-takedown/

[3] Krebs on Security, "Starkiller Phishing Service Proxies Real Login Pages to Bypass MFA," February 2026. https://krebsonsecurity.com/2026/02/starkiller-phishing-service-proxies-real-login-pages-mfa/

[4] Abnormal AI, "Starkiller Phishing Kit: Technical Analysis," Abnormal AI Blog, 2026. https://abnormal.ai/blog/starkiller-phishing-kit

[5] The Hacker News, "Starkiller Phishing Suite Uses AiTM Headless Browser to Defeat MFA," March 2026. https://thehackernews.com/2026/03/starkiller-phishing-suite-uses-aitm.html

[6] Sekoia Threat Intelligence, "Sneaky 2FA: Exposing a New AiTM Phishing-as-a-Service," Sekoia Blog, 2025. https://blog.sekoia.io/sneaky-2fa-exposing-a-new-aitm-phishing-as-a-service/

[7] The Hacker News, "Phishing-as-a-Service: Rockstar 2FA Targets Microsoft 365," November 2024. https://thehackernews.com/2024/11/phishing-as-service-rockstar-2fa.html

[8] The Hacker News, "Sneaky 2FA Phishing Kit Adds Browser-in-the-Browser Pop-Up Capability," November 2025. https://thehackernews.com/2025/11/sneaky-2fa-phishing-kit-adds-bitb-pop.html

[9] CYFIRMA Research, "Mamba Phishing-as-a-Service Kit: How Modern AiTM Attacks Operate," CYFIRMA, 2025. https://www.cyfirma.com/research/mamba-phishing-as-a-service-kit-how-modern-adversary-in-the-middle-aitm-attacks-operate/

[10] Push Security, "Detecting Evilginx, EvilnoVNC, Muraena, and Modlishka," Push Security Resources, 2025. https://pushsecurity.com/resources/phishing-detecting-evilginx-evilnovnc-muraena-and-modlishka

[11] Push Security, "Scattered Spider TTP Evolution in 2025," Push Security Blog, 2025. https://pushsecurity.com/blog/scattered-spider-ttp-evolution-in-2025

[12] CISA / FBI / MS-ISAC, "Scattered Spider: Joint Cybersecurity Advisory (AA23-320A)," November 2023. https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-320a

[13] Microsoft Security, "Identifying Adversary-in-the-Middle (AiTM) Phishing Attacks Through Microsoft Sentinel," Microsoft Tech Community, 2024. https://techcommunity.microsoft.com/blog/microsoftsentinelblog/identifying-adversary-in-the-middle-aitm-phishing-attacks-through-3rd-party-netw/3991358

[14] FIDO Alliance, "Passkeys: Phishing-Resistant Authentication," FIDO Alliance, 2025. https://fidoalliance.org/passkeys/

[15] Microsoft, "Passkeys in Microsoft Authenticator — General Availability," January 2025. https://learn.microsoft.com/en-us/entra/identity/authentication/concept-authentication-passkeys-fido2

[16] WinBuzzer, "Microsoft Entra Passkeys for Windows — Phishing-Resistant Sign-In," March 2026. https://winbuzzer.com/2026/03/11/microsoft-entra-passkeys-windows-phishing-resistant-sign-in-xcxwbn/

[17] Canadian Centre for Cyber Security, "Defending Against Adversary-in-the-Middle Threats: Phishing-Resistant MFA (ITSM.30.031)," 2025. https://www.cyber.gc.ca/en/guidance/defending-against-adversary-middle-threats-phishing-resistant-multi-factor-authentication-itsm30031

[18] Microsoft Learn, "Token Protection — Conditional Access," Microsoft, 2025. https://learn.microsoft.com/en-us/entra/identity/conditional-access/concept-token-protection

[19] Krebs on Security, "Google: Security Keys Neutralized Employee Phishing," July 2018. https://krebsonsecurity.com/2018/07/google-security-keys-neutralized-employee-phishing/