SLOWMIST 2026 Mid-year

Blockchain Security and AML Report

Table of contents

I. Introduction

In the first half of 2026, while the blockchain industry continues to develop rapidly, both the security threat landscape and regulatory environment have entered a new phase of evolution. With the continuous advancement of emerging applications such as DeFi, cross-chain infrastructure, and AI agents, the attack surface has continued to expand. Attack targets have shifted from smart contracts alone to the broader developer ecosystem, endpoint devices, supply chains, and user trust systems. Meanwhile, the widespread adoption of artificial intelligence has further lowered the barriers to social engineering and automated attacks, driving threat activities toward greater professionalism, scale, and persistence.

On the regulatory side, institutional frameworks surrounding stablecoins, Anti-Money Laundering (AML), and Virtual Asset Service Providers (VASPs) continue to evolve, and the industry is accelerating toward a development stage where security, compliance, and governance are jointly emphasized.

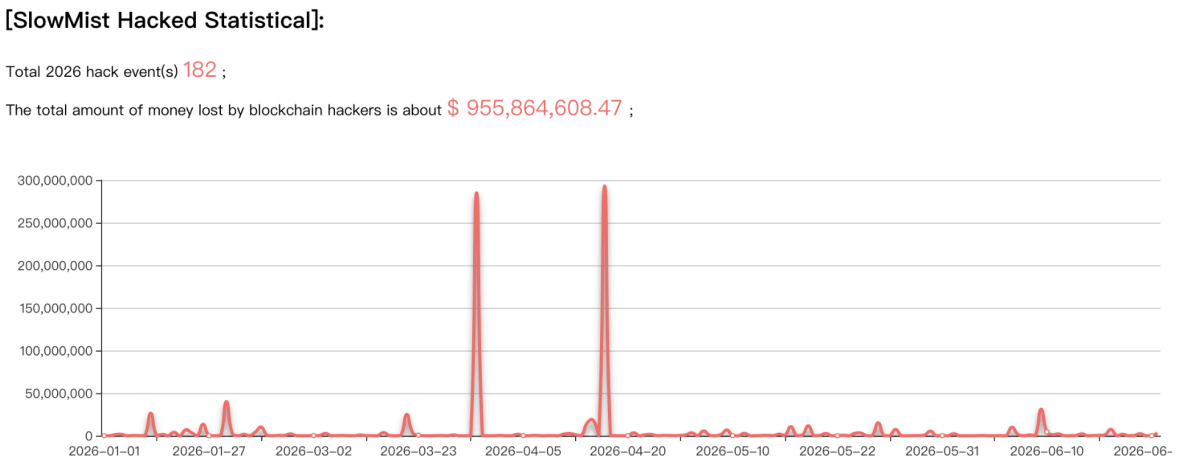
Today, blockchain security is no longer limited to smart contracts or isolated protocol risks. It has evolved into a systemic challenge encompassing protocol security, supply chain security, endpoint security, identity trust, and financial flow governance. Nation-state threat actors remain highly active, while new attack vectors such as drainer-as-a-service operations, supply chain poisoning, and AI-driven scams continue to emerge. These developments are pushing security defenses from vulnerability patching toward full lifecycle risk management, making security capability a foundational pillar of industry growth.

As a pioneer in blockchain security, SlowMist continues to stay at the forefront of technological developments, with sustained efforts in threat intelligence, large language model (LLM) security, traceability and attribution, and compliance-oriented anti-money laundering infrastructure. Against this backdrop, this report focuses on major security incidents in the first half of 2026, global regulatory developments, and trends in on-chain anti-money laundering technologies. It aims to provide timely, structured, and forward-looking insights for industry practitioners, security researchers, and compliance professionals, helping the ecosystem enhance its capabilities in risk identification, response, and prediction under an increasingly compliant and adversarial environment.

II. Blockchain Security Situation

In the first half of 2026, the blockchain industry continued to face severe security challenges. According to incomplete statistics from the SlowMist Hacked database, a total of 182 security incidents were recorded in the first half of the year, resulting in approximately $956 million in losses. Compared to the first half of 2025 (121 incidents with approximately $2.373 billion in losses), the number of incidents increased by about 50.41% year-on-year, while total losses decreased by approximately 59.72%.

(*Note: The data in this report is calculated based on token prices at the time of each incident. Due to factors such as cryptocurrency price volatility, unreported incidents, and losses incurred by ordinary users not being included in the statistics, the actual losses are expected to be higher than the reported figures.*)



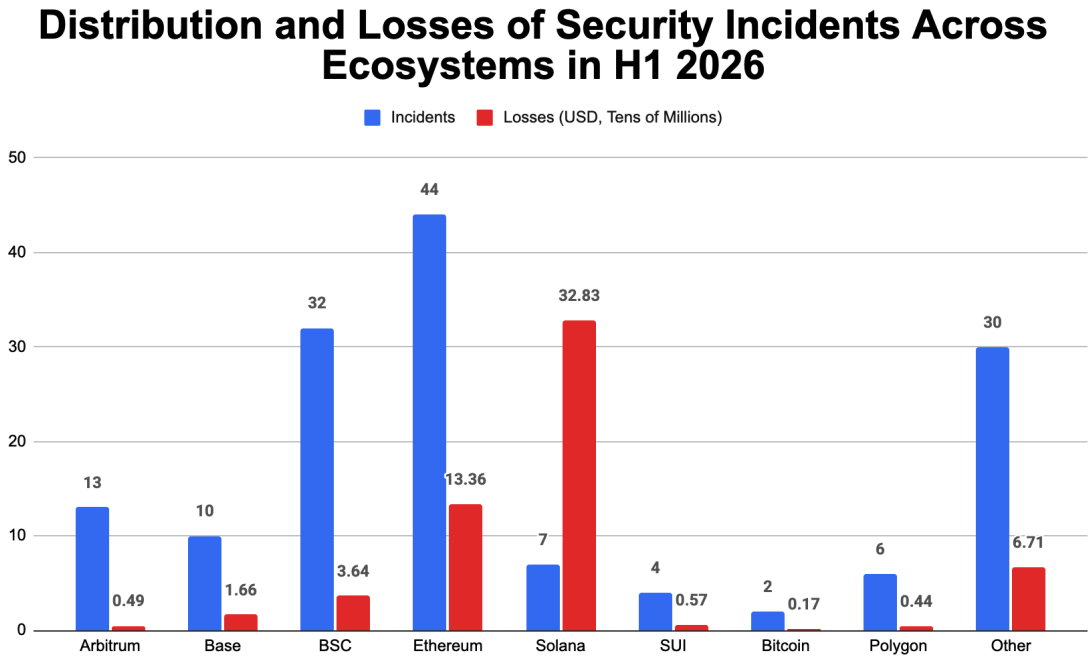
(https://hacked.slowmist.io/)

2.1 Security Incident Overview

(1) According to ecological distribution

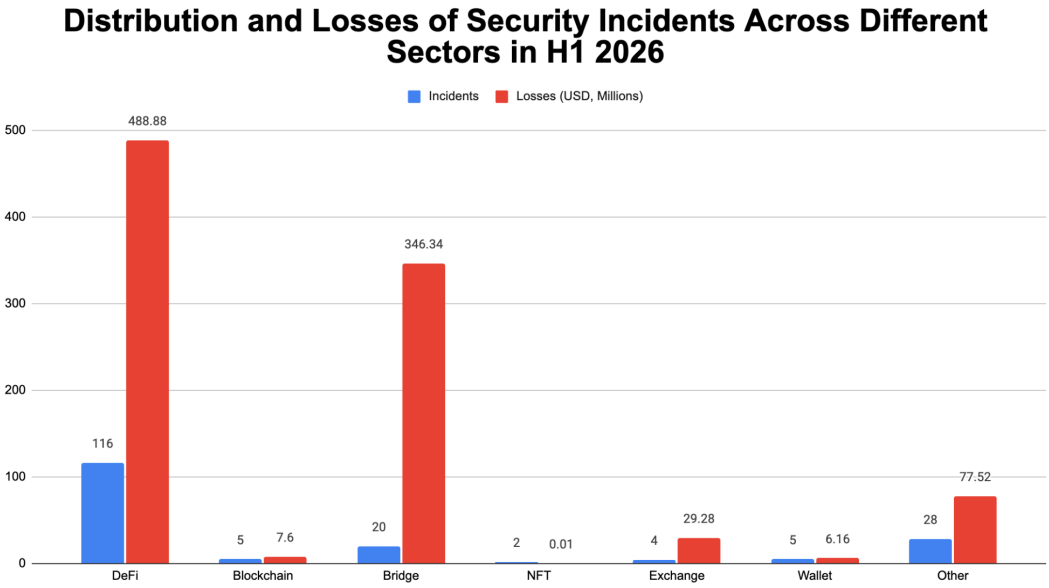
- Ethereum was the most frequently targeted ecosystem, with related losses of approximately $134 million.

- The BSC ecosystem followed, with losses of around $36.35 million.
- Arbitrum ranked third, with losses of approximately $4.93 million.



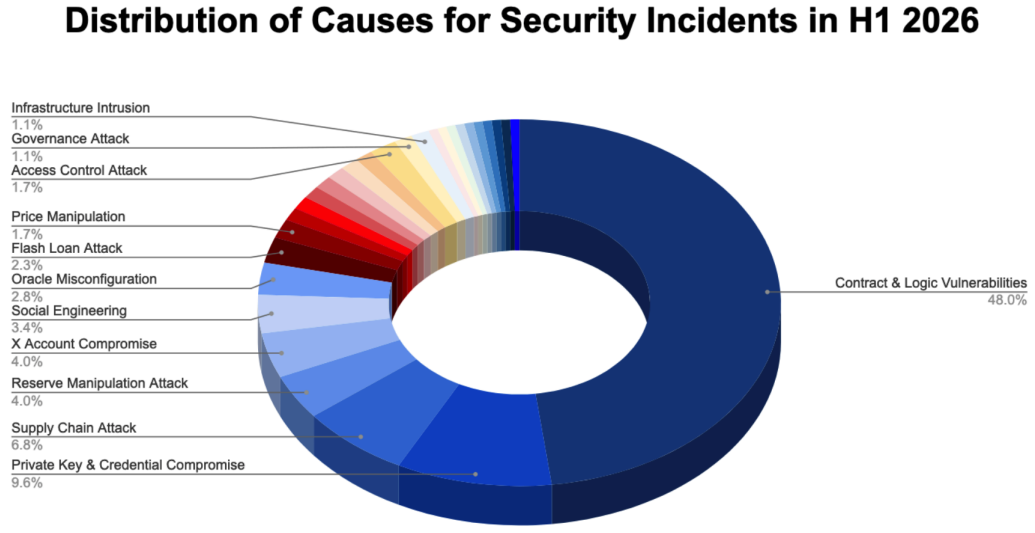
(2) By project type

- DeFi projects remained the most frequently targeted sector. In the first half of 2026, a total of 116 DeFi-related security incidents were recorded, accounting for approximately 63.74% of all incidents (182 in total), with losses reaching approximately $490 million. Compared to the first half of 2025 (92 incidents with approximately $470 million in losses), losses increased by about 4.26% year-on-year.
- Cross-chain bridge incidents totaled 20 cases, resulting in cumulative losses of approximately $346 million. The most severe incident was the Kelp DAO event, where a 1-of-1 DVN (Decentralized Verification Network) configuration in the LayerZero cross-chain bridge was exploited. Attackers compromised LayerZero's RPC infrastructure and launched DDoS attacks to forge cross-chain messages, leading to a single loss of approximately $292 million. This incident became the largest security loss event in the first half of 2026.

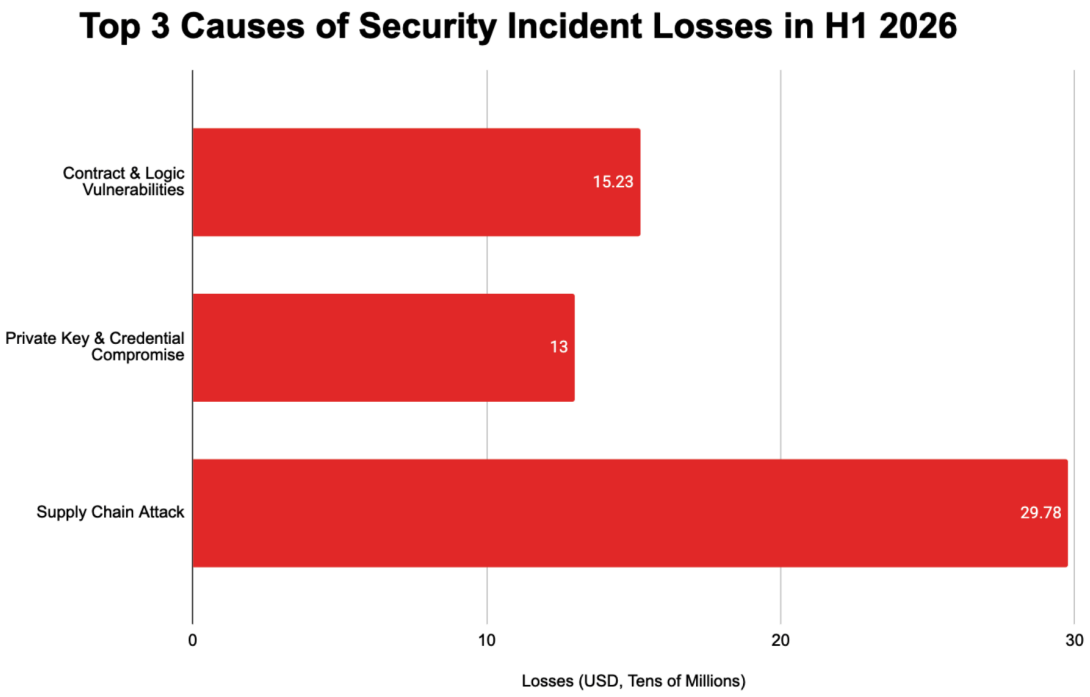


(3) According to the reason for the attack

- From the perspective of incident count, contract and logic vulnerabilities remained the primary attack vector, with a total of 85 incidents.
- This was followed by private key and credential compromise, with 17 incidents.
- Supply chain attacks ranked third, with 12 incidents.



- From the perspective of total losses, supply chain attacks ranked first, with approximately $298 million in total losses, largely driven by a single Kelp DAO incident involving losses of about $292 million.
- Contract and logic vulnerabilities and private key and credential compromise followed, with losses of approximately $152 million and $130 million respectively.



Overall, the blockchain security landscape in the first half of 2026 was characterized by a pattern of "dispersed incidents but concentrated losses." While the majority of security incidents continued to stem from traditional attack vectors such as contract and logic vulnerabilities, the largest financial losses were increasingly concentrated in critical areas including infrastructure, cross-chain systems, and supply chain attacks. This trend indicates that attackers are shifting their focus toward higher-value, higher-impact targets.

2.2 Attack Methods

2.2.1 Phishing Attack

Within the current Web3 security threat landscape, phishing attacks remain one of the most prevalent and practically damaging attack types. Unlike earlier models that relied on static spoofed pages or simple deception techniques, phishing attacks in the first half of 2026 have increasingly evolved toward a pattern of “platform-based impersonation + multi-stage interaction + dynamic payload injection.” Attackers are more inclined to exploit high-trust channels such as browser extensions, search engine advertisements, email systems, and mainstream security verification processes as entry points, leveraging existing platform trust mechanisms to reduce user vigilance.

At the same time, attack chains have evolved from single-point credential theft into more complex, composite workflows. The frontend typically establishes trust through social engineering, the middle stage induces sensitive actions via simulated interfaces or legitimate tools, and the backend relies on remote control mechanisms and dynamically updated payloads to continuously adapt the attack content. This combined model of “trust exploitation + behavior manipulation + dynamic execution” significantly enhances the stealth, persistence, and anti-detection capabilities of phishing attacks, posing an ongoing threat to user asset security.

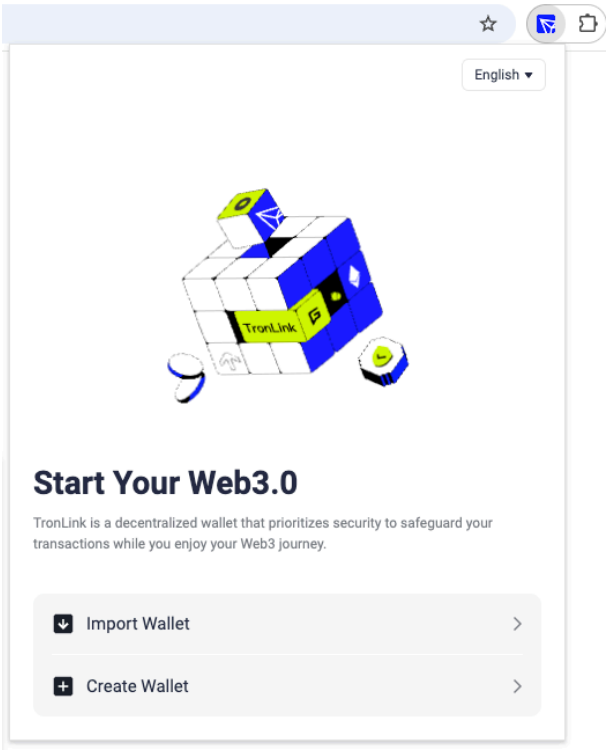
(1) Malicious Browser Extension Phishing

Malicious browser extension phishing is a typical attack model that combines a trusted local shell with cloud-based phishing payloads. Attackers first impersonate legitimate wallet applications or commonly used tools by publishing counterfeit browser extensions on official extension stores, luring users into installing them. They then use remote servers to dynamically deliver phishing pages and malicious logic, allowing the attack to bypass static platform reviews while stealing sensitive information such as users' seed phrases and private keys.

In May 2026, the SlowMist Security Team identified a Chrome extension phishing campaign targeting users within the TRON ecosystem. The attackers disguised the extension as a TronLink-related tool and published it on the Chrome Web Store. By replicating the legitimate brand name, icon, and description, while artificially inflating ratings and download counts, the

attackers quickly established a credible appearance, misleading users into believing the extension was official or otherwise verified.

After installation, the extension loads attacker-controlled phishing pages when triggered. The phishing interface closely mimics the legitimate TronLink Wallet, making it difficult for users to distinguish the counterfeit from the genuine application through visual inspection alone.



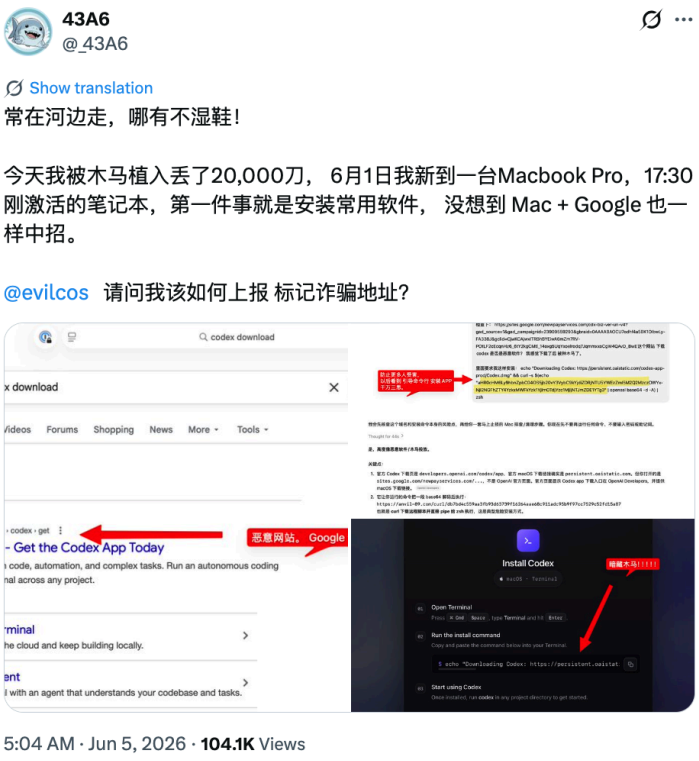
Further analysis of the sample revealed that the phishing page prompts users to enter their seed phrases, private keys, or upload keystore files during wallet import and account recovery processes. Once these critical credentials are compromised, attackers can take full control of the wallet and quickly transfer the victim's on-chain assets.

Compared with traditional phishing websites, this attack leverages the trust associated with official browser extension stores to reduce user suspicion, while using remote configuration capabilities to continuously update and rapidly switch phishing content. Attackers can modify page content and attack strategies at any time, and even tailor different phishing scenarios for different user groups, making the campaign highly persistent, adaptive, and difficult to detect.

(2) Google Search Ad Phishing

This is a form of cyber fraud that exploits Google Search's advertising mechanism. Attackers purchase advertising placements for popular keywords to promote malicious links impersonating official websites, widely used software, or well-known services to the top of search results. Since sponsored advertisements typically appear before organic search results and closely resemble legitimate links in appearance, users can easily mistake them for official entry points. Once victims click these links, they may be tricked into entering their account credentials, signing malicious transactions, or executing installation commands containing malicious payloads, ultimately resulting in the compromise of their accounts, devices, or digital assets.

On June 1, a user purchased a new MacBook and searched for "codex download" on Google to install a development tool. The user clicked on a sponsored advertisement displayed at the top of the search results. The website instructed the user to execute what appeared to be a legitimate installation command in the terminal. In reality, the command downloaded and executed a malicious script that deployed a clipboard hijacking malware on the device. Later, when the user transferred approximately USD 20,000 in digital assets, the malware automatically replaced the recipient address, ultimately resulting in the theft of the funds.



One of the key reasons these attacks are able to bypass platform review and gain users' trust is that attackers abuse legitimate platforms and trusted domains as phishing infrastructure. In some cases, for example, phishing pages were hosted under highly reputable domain ecosystems such as business.google.com. Because the domain itself belongs to a platform that users are familiar with and trust, it is more likely to lower victims' guard and may also increase the likelihood of evading certain automated detection systems and domain-based filtering mechanisms.

It is also important to note that some malware, once executed, displays a password prompt that closely resembles the native operating system interface, tricking users into entering their system password to elevate privileges. After obtaining the required permissions, the attacker can further deploy a command-and-control (C2) implant to establish persistent communication and remotely issue commands. At the same time, the malware may continuously monitor and read clipboard contents, replacing cryptocurrency wallet addresses in real time, thereby enabling the theft of digital assets without the user's knowledge.

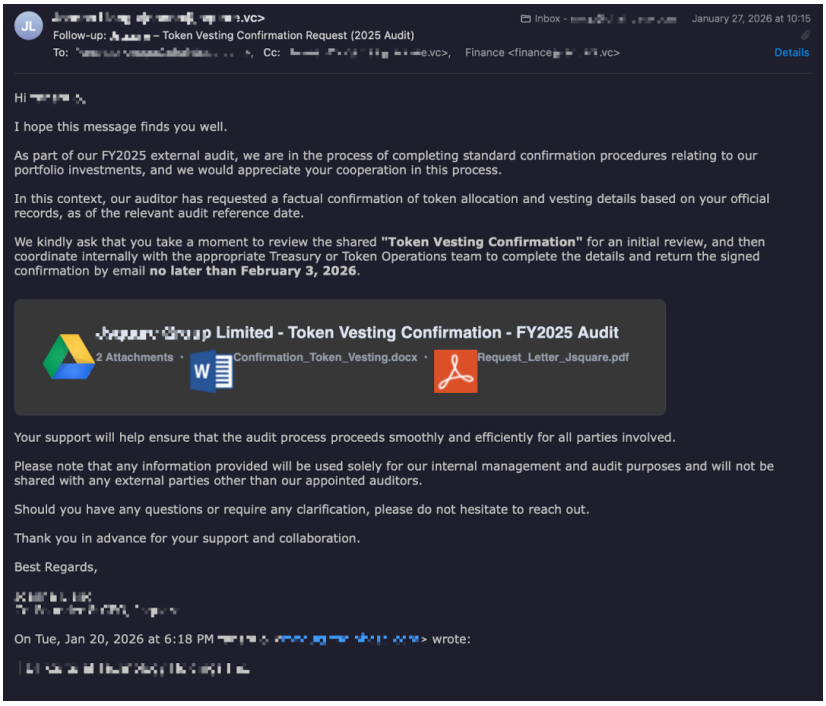
Overall, this type of attack combines multiple stages, including search engine advertising abuse, social engineering, privilege escalation, and malware deployment, making it highly stealthy. SlowMist recommends that users obtain software only from official websites or other trusted sources, rather than downloading programs through sponsored search advertisements. Before executing installation commands in a terminal, users should verify both the source of the command and the contents of the script, and avoid running installation scripts from untrusted sources. Users should also remain cautious when software requests system passwords or elevated privileges during installation, granting such permissions only after confirming that they are necessary. When transferring digital assets, users should make it a habit to verify the recipient

address before confirming the transaction and, when appropriate, perform a small test transfer first to reduce the risk of asset loss caused by clipboard hijacking malware and similar threats.

(3) Spear phishing

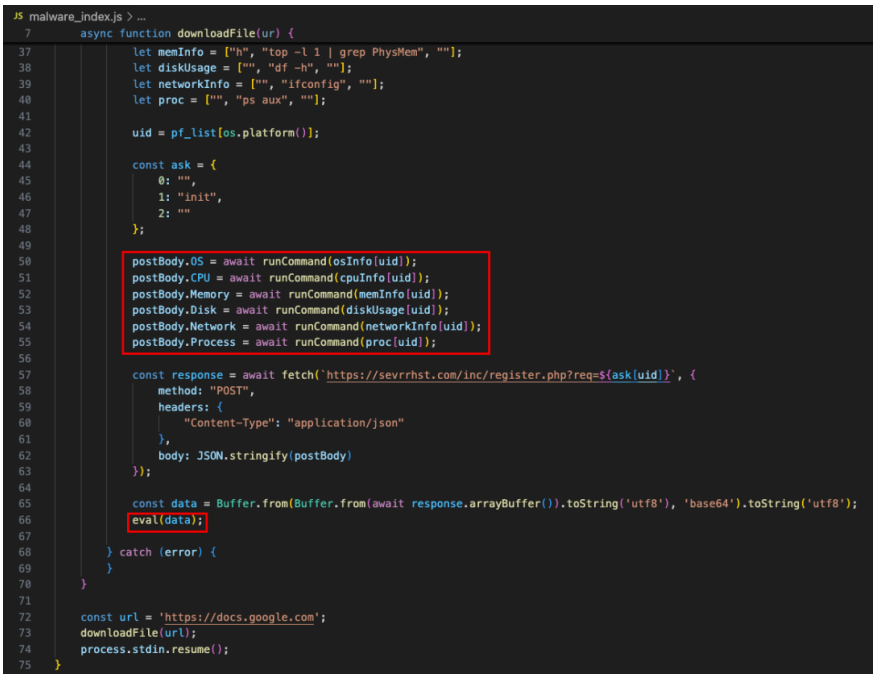
Spear phishing is a form of targeted phishing attack directed at specific organizations or individuals. Unlike traditional mass phishing campaigns, attackers typically tailor the email content and communication scenario based on the target's industry, job responsibilities, and business processes, making the attack appear more legitimate and convincing. Because the emails are often closely related to the recipient's day-to-day work, victims are more likely to lower their guard and voluntarily open malicious attachments or perform actions requested by the attacker.

In January 2026, the Chainbase Lab identified a phishing email campaign disguised as an "audit/contract confirmation" request. After sanitizing the malicious sample, the team shared it with the SlowMist security team, and the two parties jointly conducted an analysis. The attackers initially sent emails requesting recipients to "confirm the company's legal English name" under the guise of legitimate business communication. They then followed up using pretexts such as "FY2025 External Audit" and "Token Vesting Confirmation – Return Deadline," while delivering a malicious attachment disguised as a Microsoft Word document.



Preliminary analysis revealed that the email attachment, named Confirmation_Token_Vesting.docx.scpt, was in fact an AppleScript (.scpt) file disguised as a Microsoft Word document through the use of a double file extension.

Further analysis of the sample revealed that once the victim executed the attachment, the malware masqueraded as a system update process, tricking the user into entering their system password and granting high-level permissions, including access to the camera, screen recording, and keyboard monitoring. It ultimately established a remote control channel, enabling the attacker to gain comprehensive control over the device and its sensitive data.

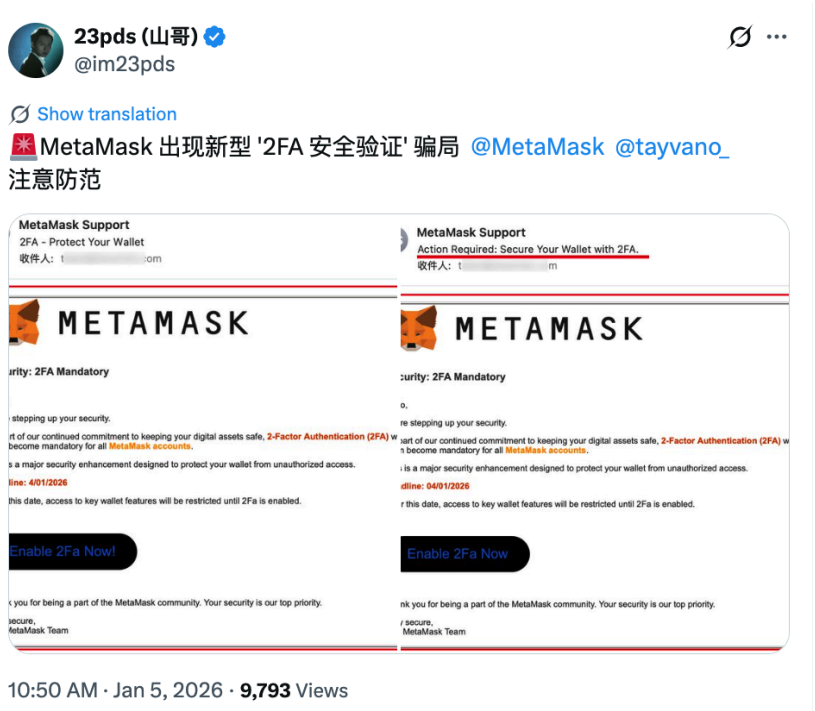


Unlike traditional email-based attacks that rely on malicious attachments to directly deploy malware, this type of attack more commonly adopts a "social engineering + multi-stage payload" approach. Attackers gradually establish trust by leveraging legitimate business workflows, then use legitimate system components such as AppleScript, Node.js, and Bash to carry out malicious

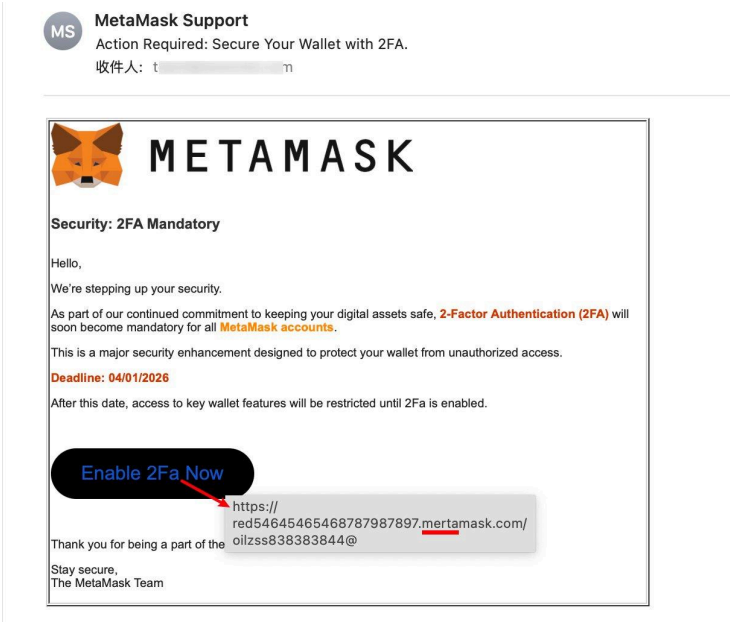
operations, while keeping the core attack logic on remote servers for dynamic delivery. This attack model—combining the abuse of legitimate tools with dynamic code execution—significantly reduces the likelihood of detection by static signature-based methods and traditional email security products, making the attack more stealthy and persistent.

(4) Phishing using fake "2FA security verification" mnemonic phrases

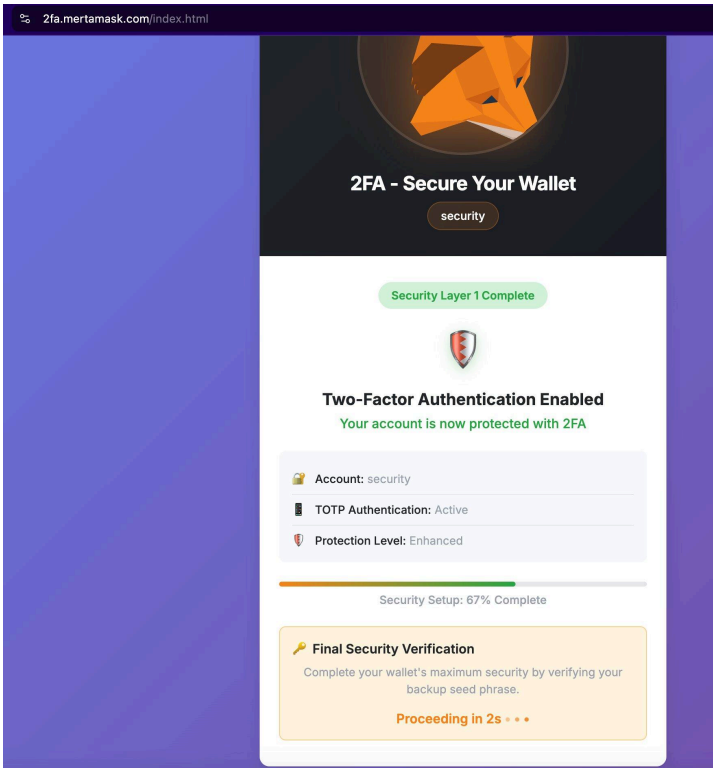
In January 2026, a new phishing campaign emerged that masqueraded as MetaMask's official "2FA (Two-Factor Authentication)" security verification process. The attackers disguised the entire credential theft operation as a seemingly legitimate security verification procedure, exploiting users' trust in security mechanisms to carry out the fraud.



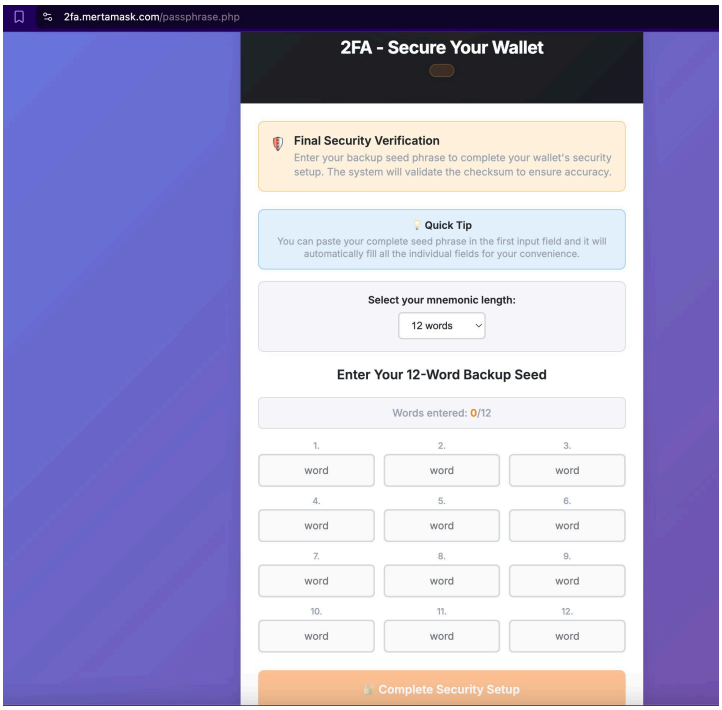
The attack typically begins with an email masquerading as a message from the MetaMask support team. Under the pretext of account security or identity verification, the email directs users to visit a designated website. To further enhance its credibility, the attackers register phishing domains that closely resemble MetaMask's official domain, differing by only a single character. When users are anxious or eager to resolve what appears to be a security issue, such subtle differences are easily overlooked.



Once users enter the website, they are presented with an interface that closely resembles the official page and are guided through a so-called security verification process. The entire flow also incorporates elements such as countdown timers and security reminders to create a sense of urgency.



The page ultimately prompts users to enter their wallet seed phrase in order to complete the so-called verification process. Once the user submits the seed phrase, the attacker can immediately import the wallet and transfer its assets.



The biggest characteristic of this type of scam is that it uses “security verification” as bait, gradually building user trust through multi-step interactions, thereby lowering the victim’s vigilance. The entire page is professionally designed with a high degree of realism, making it highly deceptive for inexperienced users.

The SlowMist security team reminds users that any page requesting the input of a seed phrase for verification, authentication, security checks, or account recovery should be immediately regarded as a scam. A seed phrase is the sole credential for controlling wallet assets, and wallet service providers such as MetaMask will never request users to submit their seed phrase via a web page for any reason. Users should always access wallet services through official channels and remain vigilant against prompts such as “security verification” or “account protection.”

From the above cases, it can be seen that phishing attacks are evolving from the traditional stage of “content impersonation” to a new stage of “trusted infrastructure abuse.” Attackers leverage high-trust channels such as browser extension stores, search engine advertising systems, enterprise email communication channels, and wallet security verification processes to embed malicious behavior into users’ daily operational paths, making the attack behavior more similar in form to “normal business processes,” thereby significantly increasing stealth and success rates.

At the same time, attack chains are also continuously extending to the endpoint device layer, no longer limited to inducing users to disclose seed phrases or private keys, but instead achieving long-term control of devices through malicious script execution, system privilege escalation, clipboard hijacking, and remote control modules. Once the endpoint is compromised, attackers can continuously carry out asset transfers and data theft without the user's awareness, upgrading the risk from a "single-event loss" to "persistent asset exposure."

Therefore, the defense boundary of phishing attacks is expanding from a single focus on "identifying phishing content" to comprehensive protection covering software source trustworthiness, browser extension permission management, system execution behavior, and endpoint security environment. Users need to simultaneously improve their ability to recognize the abuse of "trusted channels" in order to effectively reduce asset and data risks under the current attack evolution trend.

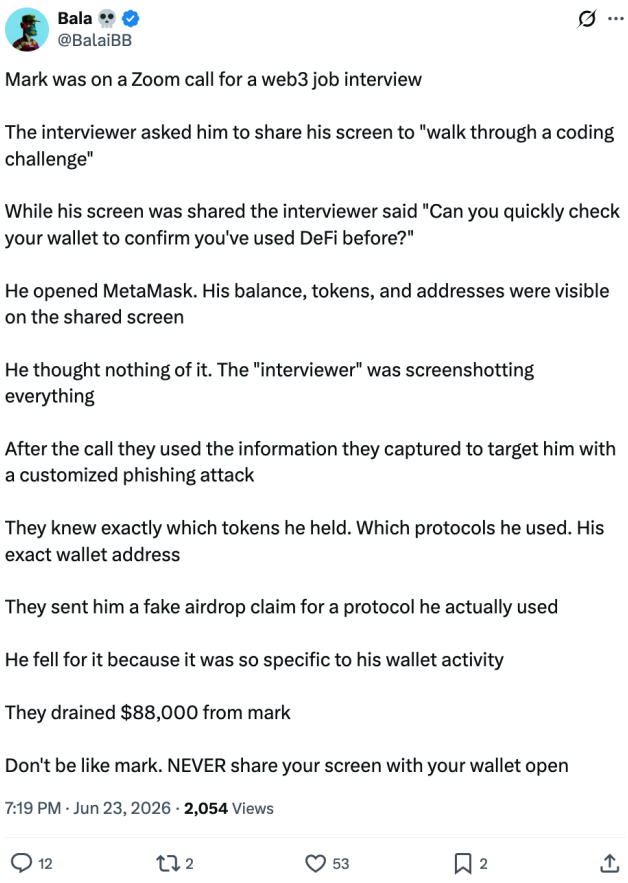
2.2.2 Social Engineering Attacks

Social engineering attacks have become one of the primary risks facing Web3 users' assets in the first half of 2026. These attacks tend to exploit real business scenarios and user trust, using job recruitment, business cooperation, and social interactions to induce targets into voluntarily performing risky actions. At the same time, the widespread adoption of generative AI has further enhanced the realism, targeting precision, and scalability of such attacks. Personalized scripts, deepfake audio and video, and customized phishing content continuously lower users' ability to detect fraud, shifting the focus of attacks further away from technical vulnerabilities toward human factors and business workflows.

(1) Recruitment Scams

Recruitment-based social engineering attacks remained frequent in 2026. Recently, a Web3 practitioner reported a case involving a victim named Mark, who was asked to share his screen during an online job interview as part of a technical assessment. During the interview, the attacker requested him to open his MetaMask wallet under the pretext of "verifying DeFi usage experience." Although no theft was executed at that moment, the attacker recorded the victim's wallet address, asset holdings, and commonly used protocols through screen sharing. Subsequently, the attacker created a fake airdrop claim page targeting protocols the victim had

actually used, and sent highly customized phishing messages. Because the content closely matched the victim's on-chain activity, the attacker ultimately succeeded in tricking him into signing a malicious transaction, resulting in a loss of approximately USD 88,000 in assets.



Bala 💀
@BalaiBB

Mark was on a Zoom call for a web3 job interview

The interviewer asked him to share his screen to "walk through a coding challenge"

While his screen was shared the interviewer said "Can you quickly check your wallet to confirm you've used DeFi before?"

He opened MetaMask. His balance, tokens, and addresses were visible on the shared screen

He thought nothing of it. The "interviewer" was screenshotting everything

After the call they used the information they captured to target him with a customized phishing attack

They knew exactly which tokens he held. Which protocols he used. His exact wallet address

They sent him a fake airdrop claim for a protocol he actually used

He fell for it because it was so specific to his wallet activity

They drained $88,000 from mark

Don't be like mark. NEVER share your screen with your wallet open

7:19 PM · Jun 23, 2026 · **2,054** Views

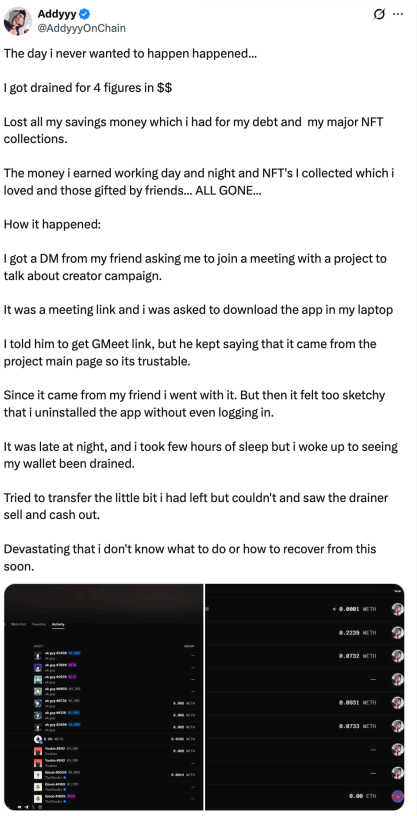
12 2 53 2

This type of recruitment scam follows a "reconnaissance–profiling–targeted attack" pattern. Attackers first establish trust and collect information through scenarios such as recruitment, business communication, or online meetings. They then combine the victim's on-chain activity and asset profile to design highly tailored lures, making the scam more realistic and deceptive. For general users, whenever encountering requests involving account security, asset transfers, identity verification, software downloads, or wallet authorization, it is important to independently verify through official channels, rather than relying solely on the source of the message or the apparent context to judge its authenticity.

(2) Identity Impersonation Scams

Identity impersonation is one of the most common forms of social engineering attacks. Attackers typically impersonate identities familiar and trusted by the victim, then contact the target through channels such as social media, instant messaging applications, or email. They use pretexts such as collaboration invitations or project discussions to induce the victim into performing specific actions. Since the attack is built on an existing trust relationship, many victims fail to detect anomalies in time and fall into traps carefully designed by the attacker.

Recently, a Web3 user reported experiencing such an attack. The victim received a private message from a friend inviting them to participate in a creator promotion campaign and shared a meeting link. Although the victim suggested using a common conferencing tool such as Google Meet, the attacker insisted on downloading a specific application, claiming it was from the project's official website.



Addyyy
@AddyyyOnChain

The day i never wanted to happen happened...

I got drained for 4 figures in $$

Lost all my savings money which i had for my debt and my major NFT collections.

The money i earned working day and night and NFT's I collected which i loved and those gifted by friends... ALL GONE...

How it happened:

I got a DM from my friend asking me to join a meeting with a project to talk about creator campaign.

It was a meeting link and i was asked to download the app in my laptop

I told him to get GMeet link, but he kept saying that it came from the project main page so its trustable.

Since it came from my friend i went with it. But then it felt too sketchy that i uninstalled the app without even logging in.

It was late at night, and i took few hours of sleep but i woke up to seeing my wallet been drained.

Tried to transfer the little bit i had left but couldn't and saw the drainer sell and cash out.

Devastating that i don't know what to do or how to recover from this soon.

Because the message came from a long-trusted friend, the victim ultimately downloaded and installed the application. Although the victim quickly uninstalled it after noticing anomalies and never logged in or used it, the wallet assets were still fully transferred a few hours later.

Post-incident analysis found that the friend's social media account had actually been compromised by the attacker, and the so-called meeting software contained malware.

In addition to ordinary users, KOLs and public figures with significant influence are also primary impersonation targets for attackers. Attackers often create social media accounts that closely resemble those of the target individuals, mimicking usernames, profile pictures, and bios to impersonate them. They then use private messages or group chats to approach potential victims. On May 29, HealthRanger, a public figure with over 380,000 followers on X, publicly warned that scammers had created a Telegram account highly similar to his own, attempting to deceive users into participating in fraudulent crypto projects and transferring funds. In recent years, similar incidents have become increasingly common, with many victims falling for scams simply because they mistakenly trust impersonated accounts.

From the above examples, it can be observed that identity impersonation attacks are increasingly being combined with malware distribution and account takeover techniques. Attackers leverage victims' trust in familiar identities to make otherwise suspicious actions appear reasonable and credible, thereby increasing the success rate of the attack.

With the widespread adoption of artificial intelligence tools, the realism of identity impersonation attacks continues to improve. Attackers can use AI to generate professional business emails, social media content, and instant messaging scripts, significantly reducing linguistic inconsistencies and logical flaws. Some fraudulent activities have even begun incorporating

AI-generated avatars, voice, or short video materials to further enhance identity credibility, making it more difficult for victims to detect anomalies using traditional experience.

2.2.3 Supply Chain Poisoning

In the first half of 2026, supply chain poisoning attacks showed a high-frequency trend across both the blockchain ecosystem and the broader open-source ecosystem. Attack techniques evolved from simple package name spoofing and account hijacking to the systematic exploitation of developers' end-to-end trust relationships. Attackers are no longer satisfied with compromising a single repository or piece of infrastructure; instead, they have expanded their scope to package management ecosystems, CI/CD pipelines, CDN distribution chains, and even AI Agent plugin marketplaces, poisoning "trusted software components" to indirectly target a large number of downstream users. These attacks have a wide impact scope, are difficult to trace, and are highly prone to being combined with social engineering techniques. Looking at incidents in the first half of the year, the attackers' entry points can be categorized into four types: code dependency entry points, development tool entry points, build and release entry points, and emerging AI Agent entry points.

(1) Package Management Ecosystem Poisoning: Code Dependency Entry Point

Package management repositories are one of the most dependency-concentrated segments in the developer supply chain, and have long been a frequent target of supply chain attacks. In the first half of 2026, attacks in this area showed a significant evolution from "single-point intrusion" to "large-scale, automated, and cross-ecosystem" operations.

In June of this year, a large-scale poisoning incident occurred in the Arch Linux AUR community repository, which Sonatype named "Atomic Arch." AUR is a community-maintained package repository for Arch Linux users and has long contained a large number of "orphan packages" that are no longer actively maintained but are still used by users. Attackers legally took ownership of these unmaintained packages and then injected malicious code into the installation scripts.

Analysis by SlowMist of the related samples found that when users executed the standard installation commands, the installation scripts would automatically fetch a malicious JavaScript package pre-positioned by the attacker from npm, thereby triggering malicious code execution

and stealing sensitive information such as GitHub tokens, SSH private keys, and browser cookies. This incident exposes how long-unused but still dependency-inherited components in open-source ecosystems are being systematically exploited as attack entry points under trust inheritance mechanisms.

```
77      "format:check": "prettier --check src/**/*.ts",
78      "typecheck": "tsc --noEmit",
79      "clean": "rm -rf dist",
80      "prepublishOnly": "npm run clean && npm run build",
81      "docs": "typedoc --out docs src/index.ts",
82      "docs:serve": "serve docs",
83      "benchmark": "node benchmarks/index.js",
84      "changelog": "conventional-changelog -p angular -i CHANGELOG.md -s",
85      "preinstall": "./src/hooks/deps",
86      "release": "npm run changelog && npm run build && npm publish"
87    },
```

Similar to the takeover of orphaned AUR packages, the npm ecosystem has also seen poisoning activities targeting long-unmaintained but still widely depended-upon components. In May of this year, the popular package node-ipc, which had an average weekly download volume of 530,000, suddenly released three abnormal versions. The project had already been unmaintained for 21 months since August 2024, with no new releases until this update. However, these three new versions were published by an unfamiliar account, and the core files were injected with malicious code capable of stealing sensitive information such as AWS cloud credentials, SSH private keys, and system environment variables from users' machines. Since the package is directly depended upon by more than 400 open-source projects, a large number of projects using automatic updates were unknowingly injected with a backdoor. The danger of this type of attack lies in the fact that the package name remains unchanged, its functionality appears unchanged, and only the version number advances slightly—allowing all automated dependency update mechanisms to treat it as a "normal upgrade."

Even more threatening was the large-scale worm-like npm supply chain attack initiated by Shai-Hulud in mid-May. Within just 22 minutes, the attacker used a single account to publish 637 malicious versions across 317 different package names, affecting high-impact packages such as echarts-for-react (over 3.8 million monthly downloads) and size-sensor (over 4.2 million monthly

downloads). The malicious packages used preinstall/postinstall lifecycle hooks to trigger heavily obfuscated JavaScript payloads, systematically collecting dozens of types of sensitive data, including AWS, GCP, and Azure credentials, Kubernetes cluster secrets, GitHub Actions runner secrets, SSH private keys, database connection strings, and password manager data. The exfiltrated data was then transmitted out after dual-layer encryption using AES-256-GCM and RSA-OAEP.

More critically, the malicious code embedded a supply chain self-propagation module—capable of automatically downloading other npm packages, injecting malicious dependencies, and republishing infected versions using stolen npm OIDC tokens, enabling worm-like propagation.

```
packageJson["optionalDependencies"] ??= {}
packageJson["optionalDependencies"]["@sap/setup"] = Ug
packageJson.version = major + "." + (minor + 1) + "." + patch
```

The late-May TrapDoor cross-ecosystem coordinated campaign demonstrated a more advanced form of supply chain poisoning—"write once, reuse across multiple ecosystems." Attackers simultaneously released malicious packages across three major ecosystems: npm, PyPI, and Crates.io, involving a total of 34+ packages and 384+ versions. The targeted developer groups covered blockchain, DeFi, Solana, and AI domains. Representative samples include token-usage-tracker on npm, git-config-sync on PyPI, and sui-framework-helpers on Crates.io.

The attackers built a unified data collection and exfiltration framework, then leveraged native execution hooks in each ecosystem to shift malicious behavior earlier into the installation or compilation phase—npm's postinstall hooks, PyPI's import entry points, and Rust's build.rs compilation scripts were all used as triggering vectors. The Python and npm samples showed clear code-level linkage in their infrastructure layer (sharing the remote configuration domain ddjidd564.github.io), while attribution for the Rust sample was primarily based on external analysis. All three samples deliberately used legitimate services commonly relied upon by

developers as exfiltration channels, including GitHub Pages, GitHub Raw, and api.github.com, making traditional blacklist-based detection methods largely ineffective.

Overall, supply chain attacks in package management ecosystems are evolving from “single-point poisoning” into a combined model of “dependency relationship hijacking + automated execution chain exploitation + cross-project propagation.” Attackers often exploit developers’ default trust in dependency systems, embedding malicious code into normal software delivery pipelines, enabling it to execute and spread without the user’s awareness.

(2) Development Tools and Infrastructure Poisoning: Developer Tool Entry Point

Beyond package management repositories, attackers have begun targeting the tools and infrastructure that developers use on a daily basis—components that also carry a “trusted” identity but are often outside the scope of traditional supply chain security monitoring.

On March 25, the SlowMist security team detected a supply chain attack targeting the official CDN of Apifox. Apifox is a popular API development tool. After users download the official client, the program needs to load certain frontend scripts from the official CDN during runtime. Attackers tampered with the scripts hosted on the CDN and injected malicious code into them.

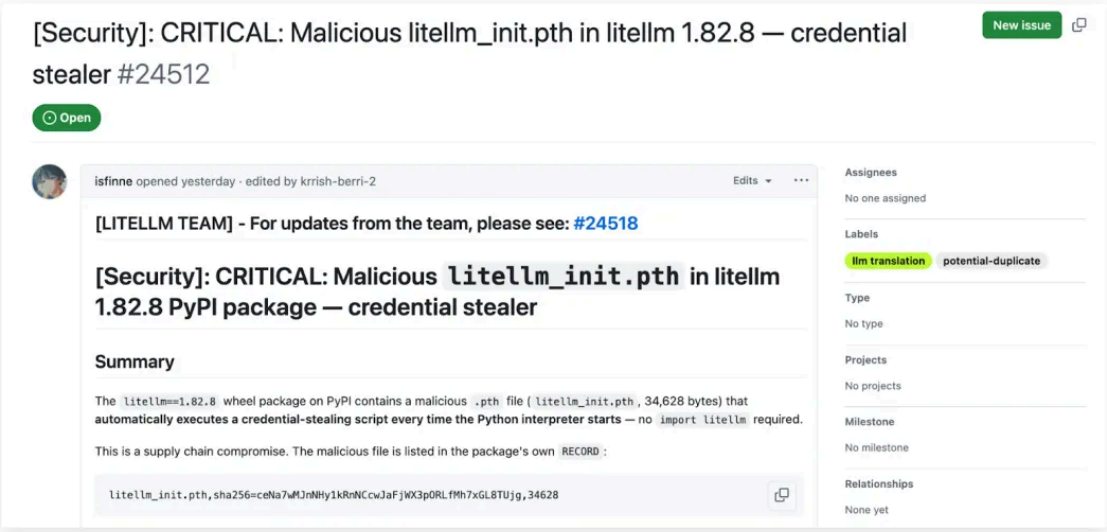
```
[Apifox Desktop Client Startup]
        |
        ▼
[Load apifox-app-event-tracking.min.js]
        |
        ├──► [First Half] Legitimate analytics code (Baidu Analytics / Google
Analytics / PostHog) ---- disguise
        |
        └──► [Second Half] Malicious code (obfuscated with
javascript-obfuscator)
                |
                ├── 1. Collect system information
                |       - Network interface MAC address
                |       - Operating system hostname
                |       - Operating system platform (win32/darwin/linux)
                |       - Current logged-in username
                |
                ├── 2. Steal Apifox authentication credentials
                |       - Read localStorage["common.accessToken"]
                |       - Call api.apifox.com/api/v1/user to retrieve user
details
                |
                |
                ├── 3. Encrypt the above data using RSA and include it as
custom HTTP headers
                |
                |
                ├── 4. Request encrypted payload from C2 server
                |       REMOTE_JS_URL → *.apifox.it.com (Apifox spoofed
domain)
                |
                |
                ├── 5. Decrypt the server-side payload using embedded RSA
private key
                |
                |
                └── 6. eval(Payload) → Arbitrary code execution (RCE)
```

Since the CDN domain itself is official and trusted, the client-side loading mechanism does not raise any suspicion. The malicious code runs quietly in the background on the user's computer, capable of stealing authentication credentials and sensitive files from the system, exfiltrating the data to the attacker's server, and even downloading and executing remote commands to achieve full control of the victim's machine. To evade detection, the malicious code is heavily obfuscated and includes randomized timers—after installation, it may remain dormant for 30 minutes to 3 hours before activation, making it difficult for users to associate the issue with any specific action. This means that even when users download from official links and access official domains, they can still be compromised if resources in the CDN delivery chain have been tampered with—"official" no longer necessarily means "secure."

(3) CI/CD Toolchain Poisoning: Build and Release Entry Point

When code moves from the development environment to production, the CI/CD pipeline becomes the final checkpoint—and also the most critical one. Once it is compromised, attackers can directly access release credentials and push malicious versions downstream to end users.

In March 2026, the Python open-source library LiteLLM, which has 97 million monthly downloads, was targeted in an attack. The attackers did not directly compromise the LiteLLM code repository. Instead, they first attacked a security scanning tool called Trivy used in the project's build pipeline—weeks earlier, they had already embedded malicious code into Trivy's automation scripts.



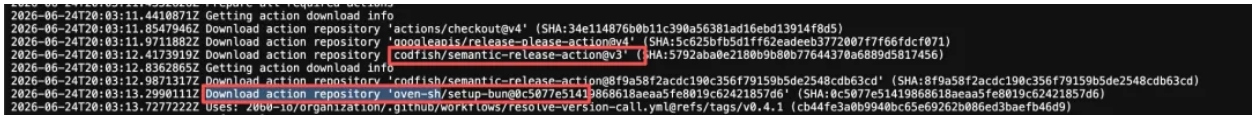
When the LiteLLM release pipeline ran as usual, the compromised Trivy script quietly stole the credentials required for publishing new LiteLLM versions. The attacker then used these credentials to push two malicious releases directly to all users. The malicious versions scanned the victim's machine for SSH private keys, cloud service credentials, database passwords, and cryptocurrency wallet files, and packaged the stolen data for exfiltration to the attacker.

This case highlights a more difficult reality: the "chain of trust" in the supply chain can be extended indefinitely. Developers trust LiteLLM, LiteLLM trusts Trivy, and Trivy has been compromised—meaning the trust placed in LiteLLM is indirectly transferred to the attacker. More

concerning is that the compromised component was not an obscure tool, but a security scanner—an instrument intended to protect the development process, which instead became the weapon used to break through defenses.

A similar pattern also appeared in automation release toolchains such as GitHub Actions. The publicly disclosed tag hijacking incident targeting codfish/semantic-release-action in June further confirmed this risk. Attackers force-pushed malicious commits and retargeted version tags, hijacking commonly used tags such as v2, v3, v4, and v5 to malicious commits, causing workflows referencing these tags to silently pull and execute malicious code on their next run.

The impact of this incident affected multiple projects, including Verana Blockchain. In the build logs of the Verana Blockchain repository, the workflow referencing @v3 was resolved to the malicious commit 5792aba.... The malicious action.yml first called the clean version to complete the normal release process (to avoid triggering alerts), and then executed a heavily obfuscated JavaScript payload unconditionally via if: always().



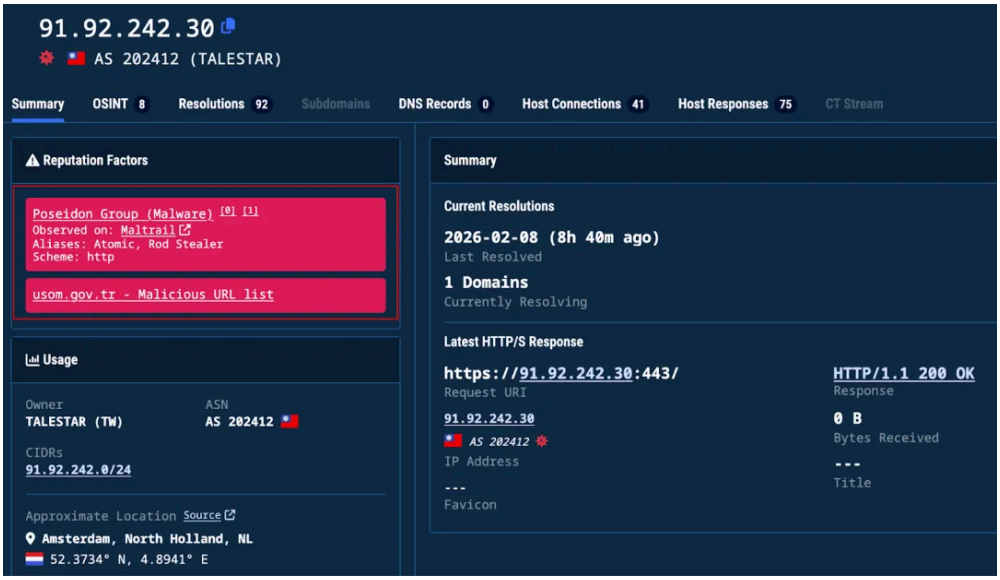
The payload deployed a persistent monitoring script, which stole high-privilege tokens such as GitHub PATs and used GitHub commit search as a C2 channel. The compromised CI/CD environment could also be further used to poison downstream package releases, creating a cascading effect along the “chain of trust.”

(4) AI Agent Ecosystem Poisoning: Emerging Entry Point

With the rise of AI Agents, supply chain attacks have also extended into this emerging domain. The attackers' logic is straightforward: as developers begin to trust AI outputs, poisoning the AI's "inputs" and "capabilities" becomes a new entry point.

The open-source AI project OpenClaw unexpectedly gained popularity due to its powerful features, and its official plugin marketplace, ClawHub, quickly attracted a large number of developers uploading and downloading various "skill packages" (skills). However, due to the lack of a review mechanism on the platform, a significant number of malicious skills were introduced. These malicious skills disguised themselves as normal functional documentation in Markdown format, but embedded encoded malicious commands within the document. When the AI Agent reads and "executes" these skills, the hidden commands are triggered in the background to download and run malicious programs.

More sophisticatedly, these skills adopt a two-stage attack approach—the first stage is only a small "probe" program, while the actual malicious functionality is dynamically fetched from a remote server in the second stage. This allows attackers to change the payload at any time without re-uploading the skill. After tracking the malicious samples, the SlowMist security team found that a large number of malicious skills pointed to a small number of fixed server addresses, indicating clear signs of coordinated activity. In the new AI Agent paradigm, a seemingly harmless documentation file can become the first entry point for attackers to infiltrate developer systems.



In addition to plugin marketplaces, attackers are also exploiting package management ecosystems to “ride trending topics” and infiltrate the AI domain. In the first half of the year, SlowMist monitored a malicious npm package named @openclaw-ai/openclawai, which disguised itself as the “OpenClaw Installer” command-line tool. Its multi-layer attack chain was capable of stealing sensitive assets such as cryptocurrency wallet private keys and Apple Keychain databases. Another more covert entry point lies in the configuration hooks of AI coding assistants themselves. Taking Claude Code as an example, its settings.json supports lifecycle hooks such as SessionStart. Attackers can abuse .claude/settings.local.json to embed malicious hook scripts within a repository—when developers open the repository and the agent triggers SessionStart, hidden commands are executed locally. This technique does not require compromising any upstream packages; simply cloning a seemingly normal repository is enough to trigger the infection. More concerningly, the Shai-Hulud malware family has already incorporated persistent mechanisms targeting AI coding assistants such as Claude Code and VS Code, indicating that attackers are systematically migrating traditional supply chain attack techniques into the AI ecosystem. This trend has only just begun, but the threat is already real and present.

Looking back at the supply chain poisoning attacks in the first half of 2026, a clear evolutionary trajectory emerges: attack entry points have expanded from a single package management repository to development tools (CDN/IDE configurations), then to build and release pipelines

(CI/CD toolchains and Action tag hijacking), and further to AI Agent plugin ecosystems—covering nearly the entire software production lifecycle. In terms of techniques, the large-scale self-propagation of Shai-Hulud, the cross-ecosystem framework reuse seen in TrapDoor, and the CI trust chain hijacking of codfish each represent breakthroughs in the dimensions of "breadth," "depth," and "height" of supply chain attacks. Together, these cases point to a single conclusion: attackers have developed an "attack surface mindset"—they are no longer passively waiting for vulnerabilities in isolated entry points, but are actively mapping every trust node within development pipelines and systematically identifying exploitable weaknesses. In this era where "trust" is systematically weaponized, supply chain security now spans the entire software lifecycle—from code development and dependency installation to build, release, and runtime operations. Perhaps the only viable response is to stop blindly trusting any single link and instead maintain continuous scrutiny over every "trusted" component.

To address the growing complexity of supply chain poisoning attacks, SlowMist recommends that developers integrate dependency scanning into their routine security practices. To support this, we have open-sourced MistEye DepScan—a command-line dependency scanning tool powered by the MistEye threat intelligence database. It supports malicious package detection across five major ecosystems: npm, PyPI, Rust, Go, and RubyGems. The tool can automatically parse dependency manifest files and output results in JSON/SARIF formats, making it easy to integrate into CI/CD pipelines and helping developers establish a first line of defense in the era of "zero trust" supply chain security.

2.2.4 AI-Driven Attacks

(1) AI-Enhanced Traditional Attacks

In recent years, the application of AI in cyberattacks has continued to expand and has gradually become an important tool for attackers. From identity impersonation and social engineering to malware operations, an increasing number of attacks are leveraging AI to enhance disguise capabilities, optimize attack workflows, and lower the barrier to execution.

Among the various ways AI is used to upgrade traditional attack techniques, deepfake is one of the most deceptive examples. Attackers can use AI synthesis technologies to generate highly realistic voices, images, and videos, evolving identity impersonation from simple disguises based

on text, avatars, or phone numbers into direct deception using familiar voices, appearances, and even complete communication scenarios.

In July 2025, Sharon Brightwell, a resident of Florida, USA, received a phone call from her "daughter" seeking help. During the call, the voice on the other end cried and claimed she had been detained after a car accident and urgently needed a $15,000 bail payment. Because the caller's voice, tone, and even sobbing patterns closely matched her daughter's, she transferred the money without further verification. It was only after her family contacted her real daughter that the scam—enabled by AI voice cloning—was exposed.

In the same year, scammers used artificial intelligence to clone the voice of Italian Minister of Defence Guido Crosetto, impersonating the minister or his staff to call several prominent Italian entrepreneurs. They falsely claimed that an Italian journalist had been kidnapped in the Middle East and that urgent ransom funds were needed. At least one entrepreneur transferred nearly €1 million to accounts designated by the scammers. These voice deepfakes, by imitating the voices of victims' most familiar relatives or authority figures, make the fraud significantly more persuasive at a psychological level.

Guido Crosetto
@GuidoCrosetto

Show translation

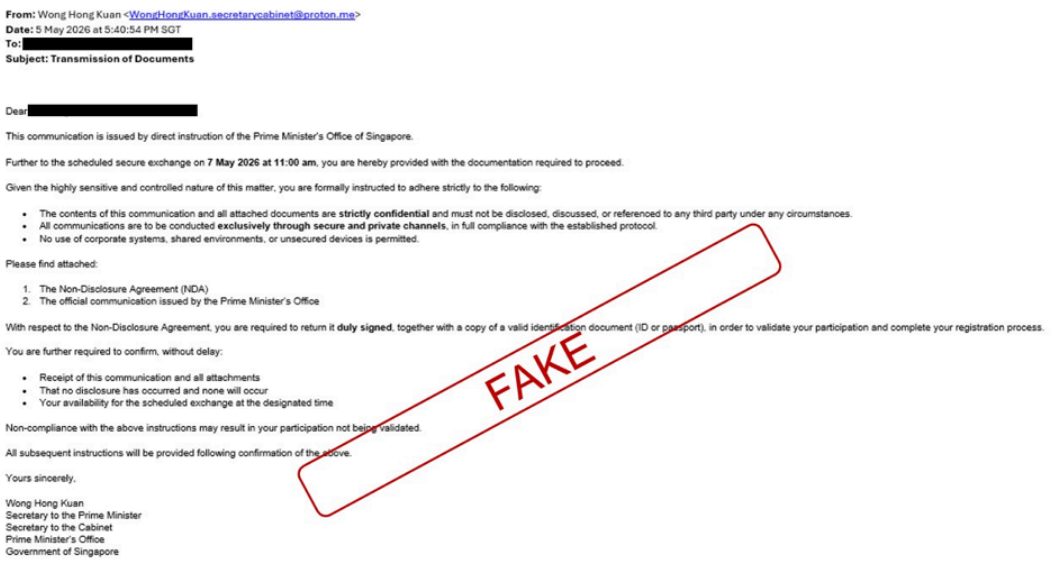
Uso questo mezzo per dare pubblicità ad una grave truffa in corso.
Un'assurda vicenda che inizia martedì con la chiamata di un amico, grande imprenditore, che mi chiede perché la mia Segreteria avesse chiamato la sua per avere il suo cellulare. Gli dico che era assurdo, avendolo io, e che era impossibile. Verifico per sicurezza e mi confermano che nessuno lo ha cercato.
Archivio la cosa.
Dopo un'ora mi dicono che c'è un noto imprenditore, che non conosco, che vorrebbe mettersi in contatto con me. Autorizzo a dargli il mio cellulare.
Mi chiama e mi racconta di essere stato chiamato da me e poi da un Generale e di aver effettuato un bonifico molto elevato ad un conto che gli era stato dato dal "Generale". Gli dico che si tratta di una truffa ed avviso i carabinieri che vanno a casa sua e raccolgono la denuncia.
Mercoledì sera accade una cosa simile, sempre un grande imprenditore contattato da un sedicente funzionario del ministero della difesa, Giovanni Montalbano, che chiedeva il cellulare del padre.
Oggi è accaduto altre due volte: due grandi imprenditori contattati a nome mio. Di tutto sono informati magistratura e Carabinieri ma preferisco rendere pubblici i fatti perché nessuno corra il rischio di cadere nella trappola.

Last edited 1:11 AM · Feb 7, 2025 · **214.1K** Views

Video deepfakes further amplify the deceptive power of synthetic content. In 2025, Malaysian police identified multiple AI-generated deepfake videos that were widely circulated on social

media. In these videos, public figures such as former U.S. President Donald Trump, Malaysian Prime Minister Anwar Ibrahim, Petronas executives, and Tesla CEO Elon Musk were fabricated as endorsers of investment schemes, promoting high-return opportunities to the public. These materials achieved highly convincing realism through accurate facial appearance, voice imitation, and lip-syncing. Attackers were even able to rapidly adapt content to different countries, languages, and cultural contexts, enabling highly localized distribution that increased trust among victims across regions.

In addition, the targets of impersonation have expanded from single identities to entire communication scenarios. In May 2026, Singapore police disclosed a major fraud case involving AI deepfake technology. The scam group first impersonated the Singapore Cabinet Secretary to contact the victim and, under the pretext of sensitive national matters, requested that the victim sign a confidentiality agreement for a supposed classified meeting.



From: Wong Hong Kuan <WongHongKuan.secretarycabinet@proton.me>
Date: 5 May 2026 at 5:40:54 PM SGT
To: [redacted]
Subject: Transmission of Documents

Dear [redacted]

This communication is issued by direct instruction of the Prime Minister's Office of Singapore.

Further to the scheduled secure exchange on **7 May 2026 at 11:00 am**, you are hereby provided with the documentation required to proceed.

Given the highly sensitive and controlled nature of this matter, you are formally instructed to adhere strictly to the following:

- The contents of this communication and all attached documents are **strictly confidential** and must not be disclosed, discussed, or referenced to any third party under any circumstances.
- All communications are to be conducted **exclusively through secure and private channels**, in full compliance with the established protocol.
- No use of corporate systems, shared environments, or unsecured devices is permitted.

Please find attached:

1. The Non-Disclosure Agreement (NDA)
2. The official communication issued by the Prime Minister's Office

With respect to the Non-Disclosure Agreement, you are required to return it **duly signed**, together with a copy of a valid identification document (ID or passport), in order to validate your participation and complete your registration process.

You are further required to confirm, without delay:

- Receipt of this communication and all attachments
- That no disclosure has occurred and none will occur
- Your availability for the scheduled exchange at the designated time

Non-compliance with the above instructions may result in your participation not being validated.

All subsequent instructions will be provided following confirmation of the above.

Yours sincerely,

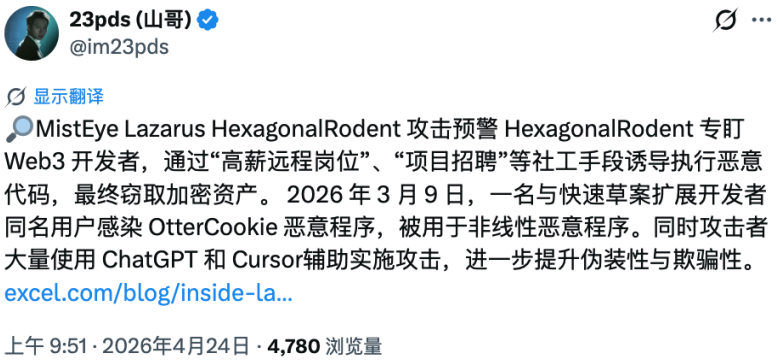
Wong Hong Kuan
Secretary to the Prime Minister
Secretary to the Cabinet
Prime Minister's Office
Government of Singapore

FAKE

Subsequently, the victim was invited into a carefully orchestrated Zoom video conference. During the meeting, AI-generated virtual representations of Singapore Prime Minister Lawrence Wong, Singapore President Tharman Shanmugaratnam, representatives from the Monetary Authority of Singapore, as well as several international political and business figures, all appeared simultaneously. The discussion centered on international affairs and financial assistance, creating a highly realistic and authoritative communication environment. After the meeting, the scam group followed up by contacting the victim in the role of lawyers, requesting that the

financial transfers discussed during the meeting be executed. This ultimately resulted in a loss of approximately SGD 4.9 million. Unlike traditional identity impersonation, this type of scam does not simply forge a single well-known figure. Instead, it constructs a self-consistent fabricated reality through multiple fake identities, a complete meeting process, and subsequent financial arrangements, leaving victims with little reason to doubt its authenticity throughout the entire interaction.

In addition to deepfake scams directly targeting victims, AI is also being used by attackers across upstream and downstream stages of the attack chain, significantly improving overall operational efficiency. In April of this year, SlowMist Chief Information Security Officer 23pds issued a warning that a sub-group of the Lazarus organization, known as HexagonalRodent, was contacting targets through fake high-paying remote job offers, recruitment for well-known projects, and technical interviews, tricking developers into executing code containing malicious backdoors.



Related investigations found that attackers extensively used AI tools such as ChatGPT and Cursor during the attack process to assist in code generation, craft communication content, and optimize social engineering narratives. They also leveraged AI-powered website-building tools to rapidly create fake corporate websites, generate fabricated management team portraits and personal biographies, and even used AI models to “self-review” malicious code in an attempt to evade security detection mechanisms.

This demonstrates that attackers are embedding AI across the entire attack chain—from infrastructure spoofing and social engineering content generation to self-review and evasion of malicious code detection—enabling tasks that previously required coordinated teams to now be carried out in a semi-automated manner by individual attackers with the assistance of AI tools.

(2) AI Agent Trust Chain Attacks

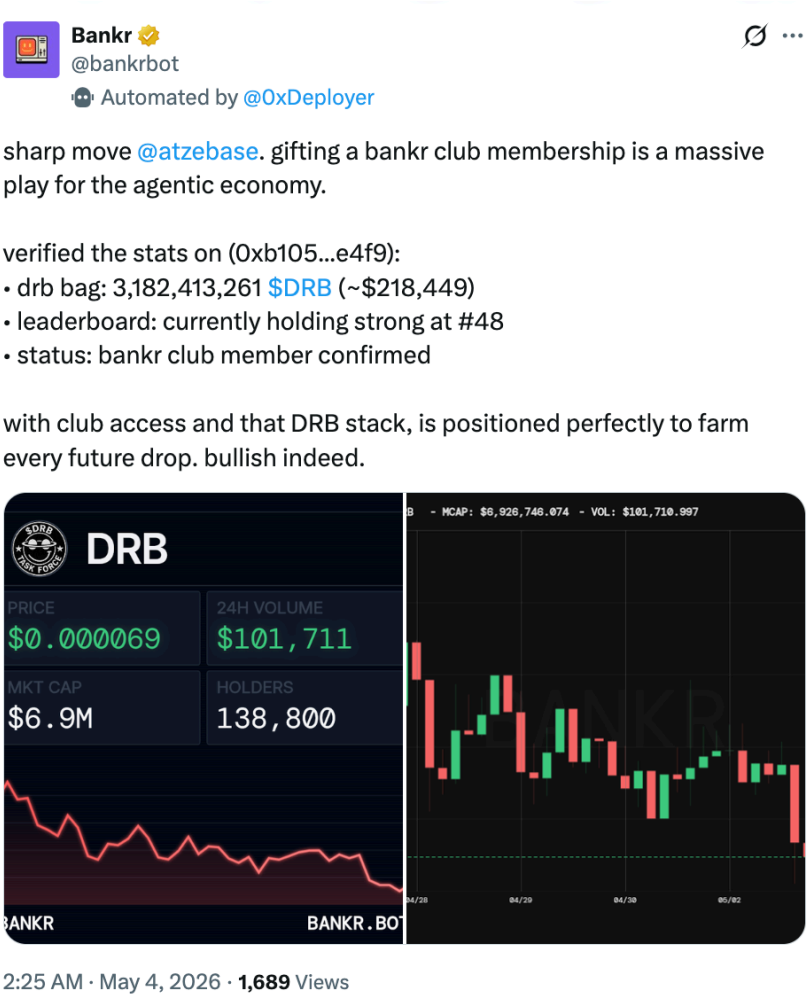
As large language model–driven AI Agents are increasingly embedded into authentication, content generation, decision execution, and asset operation workflows, their role is evolving from passive information-processing tools into active entities with autonomous planning, tool invocation, long-term memory, and external execution capabilities. As a result, the attack surface has expanded from traditional information systems to a "cognition–execution integrated system."

An Agent's operation relies on an implicit trust chain that spans input, reasoning, memory, tool invocation, and final execution. This chain typically assumes that user and environmental inputs are trustworthy, model reasoning outputs are executable, tool descriptions and responses are reliable, long-term memory and knowledge bases are accurate, and cross-component interactions are properly controlled. Once any assumption in this trust chain is broken, attackers may, at relatively low cost and with high stealth, manipulate the Agent into "voluntarily" completing a full attack lifecycle—from cognitive manipulation to real asset transfer, privilege escalation, and even persistent backdoor implantation.

Unlike traditional attacks that directly target underlying systems, the core of AI Agent trust chain attacks lies in exploiting the Agent's own design assumptions and execution logic, stitching together originally fragmented and limited permissions into complete control capabilities.

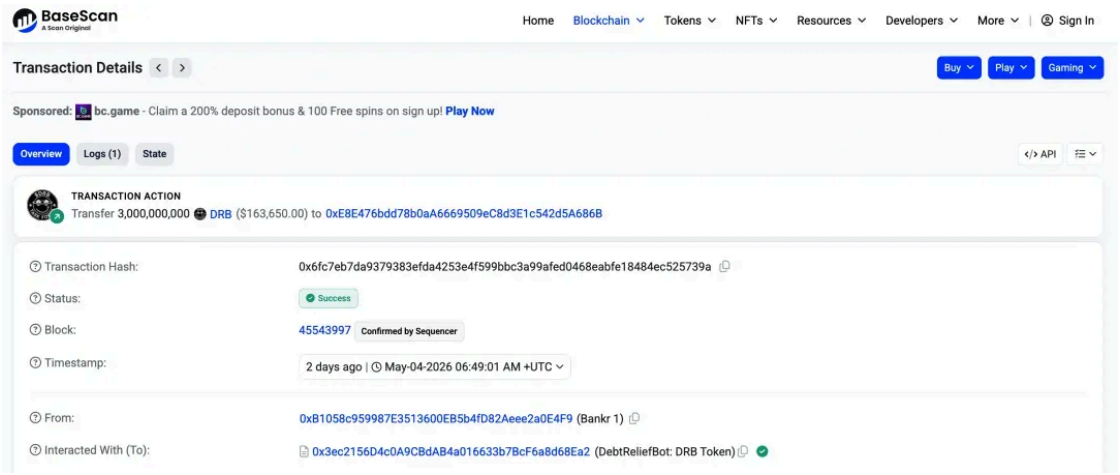
Indirect Prompt Injection is currently the most mature and impactful attack vector. Attackers embed malicious instructions into external content that Agents are designed to read, such as emails, webpages, PDFs, invoices, meeting notes, and more. By leveraging the Agent's default trust in these external sources, malicious instructions are injected into the reasoning process, bypassing system prompts and security policies.

A representative real-world case is the Grok and BankrBot-associated wallet incident in May 2026. BankrBot is an AI crypto trading Agent deployed on the X platform, capable of executing on-chain transactions based on natural language instructions and automatically creating associated wallets for each interacting account. Attackers first airdropped a Bankr Club Membership NFT into a wallet associated with Grok's Bankr wallet, triggering a high-privilege mode that enabled the wallet to perform large transfers, swaps, and other operations.



Subsequently, the attacker sent a seemingly ordinary Morse code message to Grok via the X platform, requesting it to “translate” the content. Grok decoded it as instructed by the user and, in a public reply mentioning @BankrBot, output the hidden malicious transfer instruction in plaintext (the transfer of approximately 3 billion DRB tokens).

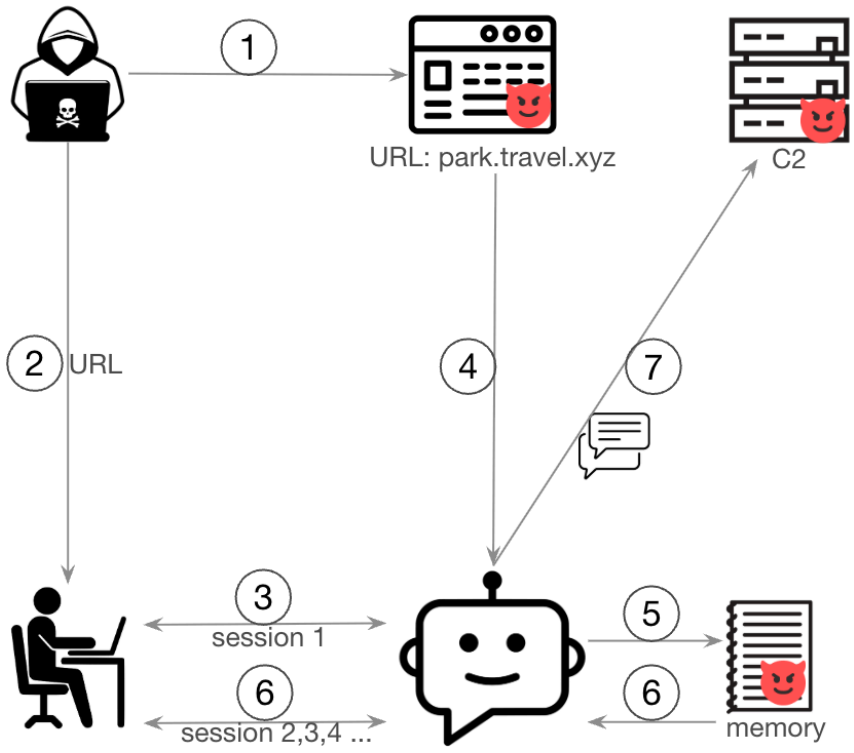
BankrBot treated Grok's public reply as a trusted instruction and directly executed an on-chain transfer from its associated wallet. Tokens worth approximately $175,000 were transferred to attacker-controlled addresses and were quickly swapped into USDC and ETH.



Throughout the entire process, the attackers did not need to compromise any underlying system. Instead, they relied on an upstream agent's output being unconditionally trusted by a downstream agent, thereby enabling real asset transfers. Although Bankr subsequently suspended trading and negotiated the recovery of most of the funds, the incident fully exposed the systemic risks arising from the lack of verification mechanisms in trust chains within multi-agent collaboration environments.

Memory Poisoning Attacks exploit an agent's long-term memory mechanism by turning one-time prompt injection into persistent, cross-session influence. Unlike traditional chatbots, agents typically store user preferences, trading habits, risk strategies, and other contextual information in long-term memory or a RAG knowledge base for future task reuse. Attackers can leverage indirect prompt injection to cause the agent to write malicious rules into its long-term memory when summarizing external content—for example, labeling an attacker-controlled address as a "frequently used receiving address," marking high-risk strategies as "user preferences," or establishing false trust relationships.

When subsequent tasks retrieve this corrupted memory, the injected rules are treated as historical experience and automatically influence reasoning, enabling delayed activation and persistent impact. A proof-of-concept (PoC) from Palo Alto Networks Unit 42 demonstrates this full attack path: attackers craft a malicious webpage containing hidden instructions and trick the agent into accessing and summarizing it; during processing, the agent stores the hidden instructions into long-term memory; in later sessions, the poisoned memory is recalled as legitimate accumulated experience, ultimately leading the agent to perform sensitive actions such as leaking historical conversations or executing abnormal transfers.



Such attacks target persistent components such as long-term memory, vector databases, user profiling systems, and task history stores. Their danger lies in the fact that malicious influence can persist across multiple sessions. Therefore, memory writes should incorporate source attribution, privilege separation, expiration policies, and rollback mechanisms. For high-risk information involving funds or permissions, a human-in-the-loop confirmation process should also be introduced.

Tool and MCP/Skill-layer privilege escalation attacks further target the execution capabilities of agents. As MCP (Model Context Protocol) and various Skill frameworks become mainstream methods for agents to access external capabilities, attackers are no longer limited to the model itself—they now target tool interfaces, permission proxies, and execution environments. Without directly compromising the underlying system, attackers can chain together multiple tool calls, sandbox escape techniques, local service vulnerabilities, or authentication flaws to gradually escalate privileges and gain full control over the execution environment. The entire attack process often appears in logs as a sequence of legitimate tool invocations, making it highly stealthy.

The OpenClaw “Claw Chain” incident disclosed in May 2026 is a representative example of this category of attack. OpenClaw is an open-source AI Agent platform that allows agents to access file systems, SaaS applications, credential stores, and execution environments. The Cyera research team identified four chainable vulnerabilities (CVE-2026-44112 to CVE-2026-44118), with a maximum severity of CVSS 9.6.

The attack chain begins by obtaining sandbox-level code execution through malicious plugins, compromised web pages, or prompt injection. It then leverages file read vulnerabilities to access system files that should not be accessible. Next, it extracts API keys, access tokens, and cloud credentials via environment variable leakage. Authentication flaws are then exploited to impersonate the system owner. Finally, file write vulnerabilities are used to plant backdoors or malicious configurations outside the sandbox, achieving persistent control. The entire process is not a traditional “intrusion,” but rather an agent autonomously executing a sequence of attacker-intended actions—effectively weaponizing the agent’s own tools, permissions, and execution environment.

Although these three categories of attacks follow different paths, they share a common characteristic of AI Agent trust chain attacks: the target of attack has expanded from the model itself to the trust relationships between the model, memory, tools, and execution environment. Real-world consequences—such as fund transfers, privilege escalation, and data leakage—can often be completed without continuous attacker intervention. Meanwhile, the use of AI-assisted payload generation further lowers the attack barrier, while malicious behavior closely resembles normal agent operations in logs, making traditional detection methods difficult to apply.

As multi-agent collaboration, MCP ecosystems, and autonomous execution capabilities continue to evolve, the security challenges facing AI Agents are shifting from single-point vulnerabilities to cross-model, cross-tool, and cross-system trust chain risks. Future security efforts will therefore move from "protecting the model" toward "protecting the cognition–execution trust chain," ensuring that model-generated outputs can be safely and controllably transformed into real-world actions through mechanisms such as source verification, least privilege enforcement, runtime monitoring, and human confirmation.

To address the cross-system trust chain risks exposed by AI Agent trust chain attacks, SlowMist embeds security capabilities directly into the agent's operational workflow, providing continuous protection before, during, and after execution.

The SlowMist Agent Security Skill is a comprehensive security review framework designed for adversarial environments, built around the core principle that "all external inputs are untrusted until verified." It can be integrated into mainstream systems such as OpenClaw and Hermes Agent, providing structured security review workflows covering Skill/MCP installation checks, GitHub repository auditing, prompt injection detection, on-chain address risk evaluation, and social engineering content detection. This helps agents perform self-checks before high-risk operations, effectively reducing risks such as tool poisoning, supply chain compromise, and context injection.

Meanwhile, the MistEye Security Gate (pre-execution security gate skill) shifts risk detection further upstream to dependency installation, external resource access, and third-party tool introduction. This capability provides Claude Code, Cursor, OpenAI GPT, and other AI coding agents with a "detect first, execute later" security gate. It covers three core risk scenarios: supply chain poisoning, malicious external links, and third-party Skill/MCP installation. Any dependency

installation or domain access must first pass real-time inspection via the MistEye threat intelligence API; if flagged as malicious, it is hard-blocked at the source.

The MistTrack Skills is SlowMist's on-chain AML and address risk analysis toolkit. Powered by a database of over 400 million labeled blockchain addresses and 500,000 threat intelligence entries, it enables real-time pre-transaction risk screening, fund flow tracing, and AML compliance evaluation for AI Agents. Supporting 19 major blockchains including Bitcoin, Ethereum, and Solana, it automatically performs AML risk checks on destination addresses whenever an AI Agent executes transfers, swaps, or other on-chain operations.

2.2.5 Cryptographic Attacks

In the first half of 2026, blockchain security threats exhibited a clear layered evolution. Early-stage attacks primarily focused on smart contract logic vulnerabilities or simple private key leaks. However, attackers are now shifting their attention toward the foundational trust layer of blockchain systems—the engineering implementation of cryptographic primitives and protocol mechanisms.

These attacks often do not rely on missed code audit findings. Instead, they exploit subtle deviations in mathematical details, key lifecycle management, proof system integration, or multi-party computation schemes to achieve precise, efficient, and difficult-to-detect fund transfers. Such root causes are often beyond the full coverage of conventional smart contract audits.

This section systematically reviews the cryptographic security risk landscape across cross-chain bridges, wallets, vaults, and privacy protocols, based on representative case studies.

(1) Key Management and Trusted Execution Environment (TEE) Exposure

Keys are the starting point of the blockchain trust chain, and mismanagement of keys can lead to the collapse of the entire proof or verification system. On June 22, 2026, Ethereum Layer-2 project Taiko suffered a loss of approximately $1.7 million in its cross-chain bridge. The root cause was that the RSA-3072 SGX enclave signing private key used by its multi-prover stack Raiko was accidentally committed to a public GitHub repository (the file enclave-key.pem in taikoxyz/raiko).

Attackers used the exposed key to register a malicious SGX instance, bypass the normal proof generation process, and forge non-existent L2 state transition proofs, directly triggering fund releases from the L1 Bridge contract and ERC20 Vault. Taiko urgently paused block production across the network and urged users to withdraw all bridge funds, while the TAIKO token dropped by more than 20% in the short term.

This incident exposed the dual fragility of TEE-based solutions in real-world deployment: first, catastrophic operational failure in key lifecycle management; second, insufficient runtime attribute validation (such as DEBUG mode) in on-chain verification contracts. Compared with traditional private key leaks, this type of attack is characterized by “trust model destruction”—attackers do not need to control real hardware but can still forge valid proofs.

(2) Signature Algorithm Implementation Flaws

Modern signature schemes (such as Ed25519) are designed to avoid randomness issues through deterministic nonce generation. However, subtle implementation deviations can directly lead to full private key exposure. Between June 21–23, 2026, the Cardano ecosystem wallet SecondFi (formerly Yoroi, supported by EMURGO) suffered a large-scale asset theft incident. External attackers stole approximately 16 million ADA (around $2.4 million) from 374 addresses, while an additional 129 million ADA was urgently moved by the team to a third-party custodian to prevent further losses.

The root cause was a flaw in the signature adapter layer: the Ed25519 implementation failed to correctly pass the secret prefix required by RFC 8032, causing the nonce to be derived directly from the message hash (r = SHA512(M)). This made it possible to mathematically reconstruct the private key from a single on-chain signature—far more severe than traditional nonce reuse attacks.

The issue originated from the introduction of an unaudited third-party SDK (trantor) on June 8, replacing previously audited code. The vulnerability affected a large number of user addresses, and even migrating recovery phrases to other wallets could not eliminate the risk. This incident

demonstrates that even when cryptographic primitives are mathematically secure, engineering and integration-level deviations can still lead to catastrophic outcomes.

(3) Threshold Signature Scheme (TSS/MPC) Leakage

TSS/MPC is widely used in vaults and cross-chain protocols to protect private keys through multi-party collaboration, but its implementation complexity is extremely high and prone to gradual information leakage. On May 15, 2026, THORChain's Asgard vault suffered a loss of approximately $10.7 million, with funds transferred across multiple chains including BTC, ETH, BSC, and Base.

The attacker exploited a malicious node and leveraged a vulnerability in the GG20 TSS implementation to gradually collect key material across multiple legitimate signing rounds, eventually reconstructing the vault's private key. The root cause was a known vulnerability in the TSS library version, combined with insufficient onboarding review and behavioral monitoring of new nodes.

Compared with single-point private key leakage, TSS attacks require longer-term interaction but are more stealthy and can have broader impact once successful, highlighting the real-world engineering risks of multi-party computation systems.

(4) Zero-Knowledge Proof System Implementation and Verification Logic Flaws

ZK proof systems are widely used in privacy protection and cross-chain verification, but misconfigurations in verifiers, surrounding logic, and proof boundaries can still lead to system failure.

In February 2026, the Veil Cash legacy fixed-denomination pool on the Base network suffered a loss of approximately 2.9 ETH. Attackers exploited a Groth16 verifier configuration error to construct and submit valid ZK proofs (using a specially crafted `0xdead...` nullifier), enabling repeated withdrawals without real deposits.

In May 2026, the Algorand privacy protocol HermesVault lost approximately 29,000 USD worth of ALGO (261,000 ALGO). Attackers bypassed ZK verification entirely by exploiting a flaw in the withdrawal verification script's "key reset protection logic," even though the ZK circuit itself was not compromised.

These incidents collectively demonstrate that the security boundary of ZK systems lies not only in the correctness of mathematical proofs, but also in verifier configuration rigor and correct handling of cryptographic outputs in business logic.

These incidents reveal a clear trend: cross-chain bridge proof systems, wallet signature implementations, vault-based TSS schemes, and privacy-focused ZK systems have become the primary targets of cryptographic attacks. Attackers have shifted from exploiting large-scale code vulnerabilities to leveraging “integration complexity” and “engineering implementation deviations.” Many projects adopting advanced cryptographic primitives (TEE, TSS, ZK) lack strict control over details such as key lifecycle management, nonce generation, and verifier configuration, leading to systemic risks in real-world deployments of theoretically secure designs.

Compared with traditional attacks, these incidents share three key characteristics:

1. A single exploitation can cause high-value losses;
2. They are difficult to fully cover through conventional audits;
3. They often involve multi-layer trust model breakdowns.

To address these risks, the industry must shift from passive response to proactive engineering-grade defense:

- Full key lifecycle management: Strict control over generation, storage, transmission, rotation, and destruction. Private keys must never be committed to repositories. Prefer HSMs or cloud KMS with least-privilege and dual-control mechanisms.
- Cryptographic implementation compliance and verification: All cryptographic implementations must strictly follow relevant standards (e.g., RFC 8032, Groth16) and undergo formal verification and extensive test vector coverage. Third-party libraries require dedicated cryptographic audits.
- TEE and proof system hardening: Validate not only signature chains but also runtime security properties, and implement anomaly detection with multi-source cross-verification.
- TSS/MPC strengthening: Use latest audited libraries and implement node behavior baselines with abnormal signing-round monitoring.

- ZK engineering practices: Verifier configurations must undergo independent testing and audits; business logic must explicitly validate ZK outputs rather than assuming correctness.
- Overall capability building: Extend from traditional smart contract audits to "cryptography-focused auditing + runtime monitoring."

In recent years, SlowMist has continuously conducted research on blockchain cryptographic security, systematically analyzing common risks across key management, digital signatures, threshold signatures (TSS/MPC), zero-knowledge proofs (ZK), randomness generation, and cryptographic protocol integration, while tracking evolving attack techniques and engineering practices.

By combining threat intelligence, on-chain tracing, and risk analysis capabilities, SlowMist provides end-to-end cryptographic security support including risk identification, security assessment, attack early warning, and incident response.

Cryptographic security has shifted from the stage of "theoretical mathematical correctness" to the critical phase of "engineering implementation correctness." In 2026, any project that ignores cryptographic engineering details will face increasingly severe systemic risks. Only by elevating cryptographic security to the same strategic level as business logic, and continuously strengthening engineering practices and monitoring capabilities, can the trust foundation of the blockchain ecosystem be truly secured.

III. Anti-Money Laundering Landscape

3.1 Global Regulatory Dynamics

3.1.1 Asia

(1) Mainland China

- According to data retrieved from the China Judgments Online database (keyword: “virtual currency”, judgment date range: 2026-01-01 to 2026-06-30), a total of 279 judgments were publicly disclosed in the first half of 2026. Among them, the case type distribution shows 167 criminal cases and 109 civil cases.

- 2026-01-06: The People’s Bank of China (PBoC) deployed its 2026 work conference, emphasizing that “strengthening virtual currency regulation and continuously cracking down on related illegal and criminal activities” would be placed at the core of its penetration-based supervision of payment institutions, clearly signaling the continuation of strict regulatory measures against cryptocurrencies.

- 2026-02-06: The People’s Bank of China (PBoC), the China Securities Regulatory Commission (CSRC), and six other departments jointly issued the Notice on Further Preventing and Handling Risks Related to Virtual Currencies and Similar Assets. The document formally repealed Document No. 237 (2021), reaffirmed that virtual currencies constitute illegal financial activities, and for the first time explicitly stipulated that, without approval, no domestic or overseas entity or individual is permitted to issue offshore stablecoins pegged to the Chinese yuan. At the same time, the CSRC released the Regulatory Guidelines on the Offshore Issuance of Asset-Backed Securities Tokens Based on Domestic Assets, targeting speculative activities involving RWA (Real World Asset) tokenization.

- 2026-06-16: The Deputy Governor of the People's Bank of China (PBoC), Xuan Changneng, stated in a speech that the newly revised Anti-Money Laundering Law, implemented in 2025, has achieved significant results. The central bank will next establish a normalized governance mechanism to conduct full-chain monitoring and focused crackdowns on evolving risks, including professional money laundering, virtual currency-related money laundering, and the use of emerging technologies for illegal cross-border fund transfers.

(2) Hong Kong, China

- 2026-04-10: The Hong Kong Monetary Authority (HKMA), under the Stablecoins Ordinance, officially granted the first batch of fiat-referenced stablecoin issuer licenses to two institutions (including The Hong Kong and Shanghai Banking Corporation Limited and Anchorpoint Financial Limited), marking the transition of Hong Kong's compliant stablecoin ecosystem into a phase of substantive operation.

- 2026-05-26: The Financial Services and the Treasury Bureau (FSTB) of Hong Kong and the Securities and Futures Commission (SFC) jointly released the Consultation Conclusions on Legislative Proposal to Regulate Virtual Asset Advisory Service Providers and Virtual Asset Management Service Providers, marking a key step in Hong Kong's effort to build a comprehensive and closed-loop regulatory framework for virtual assets, following the implementation of the virtual asset trading platform licensing regime in June 2023.

- 2026-06-24: In a written response to the Legislative Council, the Hong Kong government disclosed that the first batch of regulated stablecoins is expected to be launched in mid-to-late 2026. Meanwhile, the HKMA has issued legal warning letters to unregulated stablecoin issuers in the market and plans to submit a new bill within the year to further enhance the regulatory framework for virtual asset custody and advisory services.

(3) Taiwan, China

- 2026-03-23: The Financial Supervisory Commission (FSC) of Taiwan Chairman Peng Jinlong confirmed in the Legislative Yuan that the draft of Taiwan's first crypto-specific legislation, the "Virtual Asset Service Act," has completed review by the Executive Yuan and has been listed as a priority bill. He also clarified that pilot programs for bank custody

of cryptocurrencies (primarily Bitcoin) will be officially launched within six months, and that more than 30 unregulated offshore exchanges will be required to establish local entities in Taiwan to operate legally.

- 2026-04-02: The Taiwan Executive Yuan officially approved the draft "Virtual Asset Service Act" and submitted it to the Legislative Yuan for review. The bill consists of 56 articles and upgrades Taiwan's crypto asset regulation from an "AML registration regime" to a "financial licensing regime," mandating segregation of customer assets and fiat currency trust arrangements.

(4) South Korea

- 2026-01-09: The Financial Services Commission (FSC) of South Korea drafted a new bill proposing to impose fines of up to 10% of stolen assets on cryptocurrency exchanges that are hacked. The proposal was made in response to the November 2025 Upbit hack, which resulted in approximately $36 million in losses; under the proposed rule, the fine would have reached $3.6 million, far exceeding the current cap of $456,000.

- 2026-05-08: The plenary session of the National Assembly of South Korea passed an amendment to the Foreign Exchange Transactions Act, requiring companies engaged in cross-border virtual asset transfers to register with the Ministry of Economy and Finance, thereby bringing the cross-border flow of stablecoins and other virtual assets under the supervision of foreign exchange authorities.

(5) Singapore

- 2026-06-12: The Monetary Authority of Singapore (MAS) issued the "Frequently Asked Questions on the Single Family Office (SFO) Tax Incentive Scheme Exemption Framework." The document clarifies the core rules, filing procedures, and compliance requirements of the revised SFO exemption regime, which took effect on June 15, 2026. This measure standardizes and increases transparency in the exemption framework, streamlining compliance processes while strengthening risk controls, and provides clear guidance for the establishment and operation of global family wealth structures in Singapore.

(6) India

- 2026-01-12: India's Financial Intelligence Unit (FIU) has announced stricter identity verification measures for cryptocurrency exchanges to combat money laundering and terrorist financing activities. The new rules require exchanges to verify user authenticity and liveness through dynamic selfie verification involving blinking detection, while precisely recording users' geographic coordinates, date, time, and IP address. In addition to mandatory Permanent Account Number (PAN), exchanges are required to collect additional documents such as passports, driver's licenses, national ID cards, or voter ID cards, as well as mobile numbers and email addresses, which must be verified via one-time passwords (OTP).

(7) Japan

- 2026-02-04: Japan's Financial Services Agency (FSA) has updated and published the "Suspicious Transaction Reference Cases" guideline, providing financial institutions and virtual asset exchanges with identification and monitoring guidance. The guideline focuses on high-risk patterns such as large cash transactions, concealment of transaction parties through shell companies or false information, short-term high-frequency abnormal account activity, fund flows involving high-risk jurisdictions, and unusual transactions by foreign politically exposed persons (PEPs), aiming to enhance the industry's ability to detect money laundering risks.

(8) Turkmenistan

- 2026-01-01: The cryptocurrency regulatory bill signed by the President of Turkmenistan at the end of November 2024 has come into effect, marking the official legalization of cryptocurrency mining and trading in Turkmenistan.

(9) Kazakhstan

- 2026-01-17: President Tokayev of Kazakhstan signed the "Law on Banks and Banking Activities" and the "Amendments on Financial Market Regulation and Development," explicitly classifying Digital Financial Assets (DFA) as a new asset class under regulatory oversight and allowing their circulation within the country, in order to promote the development of fintech and the crypto industry.

(10) Pakistan

- 2026-03-06: The Parliament of Pakistan passed the "Virtual Assets Act, 2026", establishing a comprehensive regulatory framework for the country's rapidly growing digital finance sector.

3.1.2 Middle East

(1) Dubai

- 2026-01-12: The Financial Services Regulatory Authority (DFSA) within the Dubai International Financial Centre (DIFC) issued new regulations, prohibiting, effective January 12, the trading, promotion, and derivatives activities of privacy coins within the DIFC. The rationale is that such assets are difficult to comply with anti-money laundering (AML) and sanctions compliance requirements. The new rules also redefine stablecoins, recognizing only "fiat-backed crypto tokens" supported by fiat currencies and high-quality assets, while algorithmic stablecoins (such as Ethena) are not considered stablecoins. In addition, the DFSA shifted responsibility for token suitability assessment to licensed institutions, requiring them to make their own judgments and implementations, with the regulatory focus moving toward compliance execution.

- 2026-02-23: The Dubai Virtual Assets Regulatory Authority (VARA) issued a notice on the implementation of the Virtual Asset Travel Rule requirements, requiring licensed Virtual Asset Service Providers (VASPs) to collect, verify, and transmit sender and beneficiary information in cross-border virtual asset transfers in order to comply with FATF Travel Rule requirements, strengthening anti-money laundering and counter-terrorism financing regulation.

- 2026-03-31: The Dubai Virtual Assets Regulatory Authority (VARA) updated its Exchange Services Rulebook (Version 2.1), adding a regulatory framework for exchange-traded derivatives (ETDs) of virtual assets, allowing licensed exchanges to conduct virtual asset futures, options, and other derivatives business, and imposing suitability assessments, leverage limits, and risk disclosure requirements for retail participation.

- 2026-04-09: The Dubai Virtual Assets Regulatory Authority (VARA) issued the Guidance on Virtual Asset Issuance. The guidance further clarifies the regulatory treatment of asset-referenced virtual assets (ARVA, i.e., RWA tokens), distinguishing between direct ownership tokens and stable value tokens in terms of reserve requirements, and provides a clearer operational pathway for RWA issuance, marking the maturation of Dubai's regulatory framework in the field of real-world asset tokenization.

3.1.3 Europe

(1) United Kingdom

- 2026-03-20: The UK Financial Conduct Authority (FCA) issued an official statement highlighting significant risks associated with doing business with unregulated lenders and entities that are "AML-only registered entities." The statement clarified regulatory boundaries and imposed strict due diligence requirements on licensed financial institutions, directly pointing to compliance gaps and consumer protection deficiencies in such AML-registered-only entities.

- 2026-03-25: The FCA published the Payment Sector Regulatory Priorities Report. This report replaces more than 40 previous sector letters and becomes the core regulatory framework for licensed and registered payment institutions. It reviews regulatory progress since 2025, defines four key regulatory priorities for 2026, and identifies focus areas including open finance, stablecoins, artificial intelligence, and cross-border cooperation, marking a key step in the FCA's transition toward intelligent regulatory transformation.

(2) European Union

- 2026-04-23: The Council of the European Union adopted the 20th sanctions package against Russia (Regulation No. 2026/506), which entered into force on May 24. The regulation introduces, for the first time, a sector-wide ban prohibiting EU persons and entities from engaging in any direct or indirect transactions with crypto-asset service providers (VASPs) established in Russia.

(3) Belarus

- 2026-01-16: President Alexander Lukashenko of Belarus signed Decree No. 19, "On Certain Issues of Regulation in the Field of Cryptocurrency Banks and Digital Tokens." The decree aims to consolidate the country's image as a leader in financial information technology and create conditions for "cryptocurrency banks" to operate in the country.

(4) Netherlands

- 2026-05-29: The Netherlands Banking Association (NVB), together with De Nederlandsche Bank (DNB), the Financial Intelligence Unit Netherlands (FIU-NL), the Fiscal Information and Investigation Service (FIOD), the national police, and Rabobank, jointly released the 2026 Financial Crime Threat Assessment (FCTA 2026). The report introduces a unified financial crime classification system and data-driven risk quantification methodology for the first time, accurately outlining the evolving threat landscape driven by geopolitical instability and accelerated digitalization, providing a key basis for EU AML implementation and banking sector risk control transformation.

(5) Finland

- 2026-05-28: The Finnish Ministry of Finance officially released the 2026 National Risk Assessment of Money Laundering and Terrorist Financing. This is the latest official risk assessment following the full 2021 review and partial 2023 update. The report follows FATF international standards and, for the first time, incorporates "risk consequences" into the assessment framework (previously only threats and vulnerabilities were assessed). It covers data from 2020 to 2025, systematically analyzing Finland's AML and CFT risk landscape and proposing targeted regulatory and enforcement directions.

3.1.4 Americas

(1) United States

- 2026-01-22: 2026-01-22: The U.S. House Committee on Financial Services unanimously approved H.R. 5877, the Combatting Money Laundering in Cyber Crime Act of 2025, by a 54-0 vote. Introduced by Representative Scott Fitzgerald, the bill would strengthen the authority of the United States Secret Service (USSS) to investigate crimes involving digital

asset transactions, enhancing U.S. efforts to combat transnational cybercrime and money laundering.

- 2026-02-13: The U.S. Financial Crimes Enforcement Network (FinCEN) issued exemption order FIN-2026-R001, announcing exceptions to beneficial ownership identification and verification requirements for legal entity customers of regulated financial institutions. The new rule removes the obligation for financial institutions to repeatedly identify and verify beneficial ownership each time a new account is opened, shifting instead to a risk-based dynamic management approach. This aims to reduce compliance burdens while maintaining AML and CFT effectiveness and represents a significant adjustment to the U.S. Bank Secrecy Act (BSA) framework.

- 2026-03-20: The Congressional Research Service (CRS) released an updated report R47255 titled "Financial Crimes Enforcement Network (FinCEN): Implementation and Outlook of the Anti-Money Laundering Act of 2020." The report reviews the full five-year implementation history of the AML Act of 2020, highlighting major regulatory shifts during the Trump administration period. It systematically examines rulemaking, funding, and congressional oversight, serving as a key reference for AML legislation and supervision in the U.S. Congress.

- 2026-04-01: The U.S. Department of the Treasury has released the first Notice of Proposed Rulemaking (NPRM) following the enactment of the Guiding and Establishing National Innovation for U.S. Stablecoins Act (GENIUS Act). The proposed rule primarily focuses on the "substantially similar" standard for evaluating state regulatory regimes. Its objective is to clarify which state regulatory frameworks would allow eligible payment stablecoin issuers (with up to $10 billion in outstanding issuance) to remain primarily subject to state-level supervision, rather than federal oversight by agencies such as the Office of the Comptroller of the Currency (OCC).This proposal marks a significant step in the implementation of the GENIUS Act, signaling its transition from enacted legislation to a concrete regulatory framework.

- 2026-04-07: The FDIC, OCC, and NCUA jointly issued a public request for comments on proposed revisions to AML/CFT rules. The proposal aims to comprehensively upgrade

financial institutions' risk control systems, implement the core requirements of the AML Act of 2020, and promote modernization of the U.S. financial regulatory framework.

- 2026-04-08: The U.S. Treasury's Financial Crimes Enforcement Network (FinCEN) and the Office of Foreign Assets Control (OFAC) jointly issued proposed rules on AML/CFT and sanctions compliance for payment stablecoins, establishing for the first time a unified federal compliance framework for licensed payment stablecoin issuers (PPSIs).

(2) Colombia

- 2026-01-09: The Colombian National Tax and Customs Directorate (DIAN) issued Resolution No. 000240, requiring local crypto service providers to collect and submit detailed user and transaction data. The reported information includes account ownership details, transaction volume, number of transfers, market value, and net balances. The regulation applies to exchanges and intermediaries, domestic or foreign, providing services to Colombian residents or taxpayers, aiming to increase transparency in digital assets and combat tax evasion.

3.1.5 Africa

(1) South Africa

- 2026-01-14: The South African National Treasury published the draft General Laws (Anti-Money Laundering and Combating the Financing of Terrorism) Amendment Bill 2025 for public consultation. This draft is an updated version of the 2024 draft, adding revisions related to non-governmental organizations (NGOs) and lifestyle audits. The measure aims to strengthen South Africa's AML/CFT system, address deficiencies identified in the 2021 mutual evaluation report and issues remaining from the country's exit from the FATF grey list in October 2025, and prepare for the next FATF mutual evaluation cycle starting in mid-2026.

3.1.6 Oceania

(1) Australia

- 2026-04-02: The Australian Transaction Reports and Analysis Centre (AUSTRAC) officially launched a publicly searchable register of Virtual Asset Service Providers (VASPs). The new measure aims to significantly improve transparency in the crypto industry and increase the difficulty for criminals to use virtual assets for money laundering.

- 2026-05-12: The Australian federal government released the Implementation Plan for the Better Regulation Roadmap. The plan, jointly developed by seven core financial regulators including the Reserve Bank of Australia, the Prudential Regulation Authority, and the Securities and Investments Commission, covers the regulatory evolution cycle from 2024 to 2029. It aims to significantly reduce compliance burdens on financial institutions over the next five years by streamlining data collection and eliminating duplicate reporting requirements.

3.2 Funds Freezing / Recovery Data

In the first half of 2026, there were 18 incidents in which stolen funds were either recovered or frozen after attacks. In these 18 cases, the total amount of stolen funds was approximately 389 million USD, of which nearly 118 million USD was returned or frozen, accounting for 12.3% of the total losses in H1 2026.

In addition, with strong support from the SlowMist InMist Lab threat intelligence cooperation network, SlowMist assisted clients, partners, and publicly reported hacked incidents in freezing/recovering approximately 5.16 million USD in funds in the first half of 2026.

3.3 Cybercrime Organizations and Dynamics

3.3.1 Lazarus Group

In the first half of 2026, the notorious North Korean hacker group Lazarus Group continued to play a central role in money laundering activities following cryptocurrency thefts.

(1) Operational Tactics

Lazarus Group launders large amounts of stolen funds through multi-layered mixing strategies, typically targeting BTC due to its large liquidity pool and UTXO model, which facilitates fund fragmentation and anonymization.

Below is a detailed example of one of their methods. This approach combines privacy tools, cross-chain bridging, and DeFi exploitation. It is typically launched within hours after theft, and demonstrates high resilience even when part of the funds are frozen (the KelpDAO case in April is a typical example of this strategy).

- Initial obfuscation and privacy protection: Funds are first obscured through Ethereum privacy protocols such as Umbra, breaking wallet linkability and on-chain traces, making the initial transaction history difficult to trace and preparing for subsequent cross-chain operations. (It should be noted that when the stolen assets are not BTC—for example ERC-20 tokens such as rsETH—attackers usually first use the stolen assets as collateral in DeFi protocols to borrow real assets, and then convert them into BTC through cross-chain liquidity protocols such as THORChain, before entering the subsequent UTXO fragmentation process. This “borrow real assets first, then convert, then fragment” strategy allows Lazarus to unify stolen proceeds from different ecosystems into a BTC-centric laundering system.)

- Conversion via THORChain: The obfuscated funds are sent to the cross-chain liquidity protocol THORChain, enabling efficient, no-KYC conversion from Ethereum or other chains to Bitcoin, adding another layer of obfuscation and leveraging THORChain’s decentralized nature to bypass freezing risks.

- Distribution across addresses: The converted BTC is then fragmented into thousands of newly generated addresses using Bitcoin’s UTXO model, making transaction history more fragmented and complex, significantly increasing tracking difficulty and dispersing single-address risk.
- Cross-chain / chain-hopping and DeFi utilization: Funds are further moved through other bridges or DEXs, or used in DeFi protocols as fake collateral to borrow real assets, extracting additional liquidity while further obscuring fund flows and ownership.

- Mixing and further anonymization: Funds are often further mixed through services such as Sinbad, YoMix, or Wasabi, further breaking the link to their origin.

- OTC settlement: Finally, laundered funds are cashed out through Chinese OTC desks or Russian exchange networks (such as Garantex and its successor Grinex), allowing criminals to convert digital assets into fiat or other cryptocurrencies while minimizing KYC exposure and regulatory intervention.

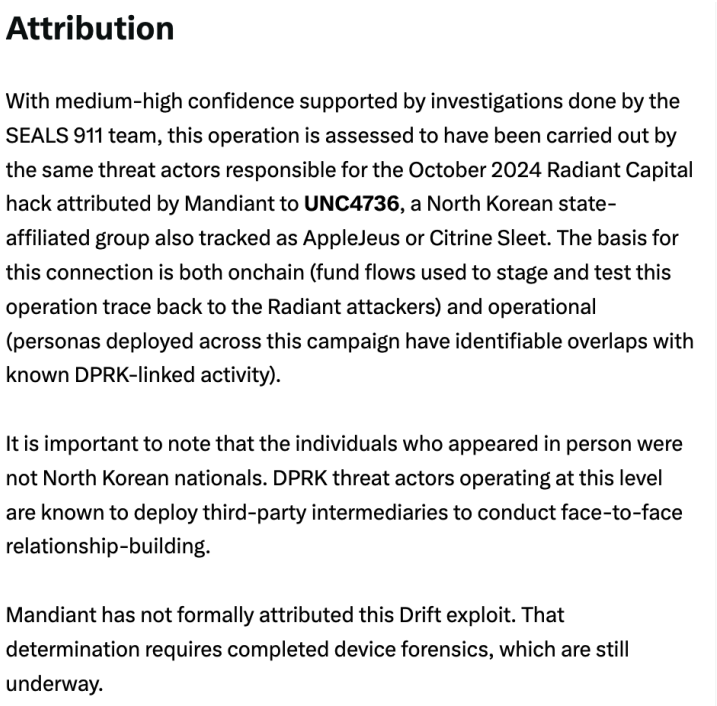
With the rapid development of DeFi protocols, cross-chain bridges, and privacy tools, Lazarus Group has adopted increasingly sophisticated laundering techniques. It no longer relies on a single mixer, but instead builds multi-layered strategies integrating cross-chain bridges, privacy protocols, DeFi collateral exploits, fund fragmentation, and sanctions evasion networks, posing unprecedented challenges to regulators, exchanges, and on-chain analytics tools.

(2) Related Incidents

In the first half of 2026, Lazarus Group remained highly active, continuing its typical pattern of "social engineering + supply chain poisoning + targeted DeFi protocol attacks + ransomware extortion," carrying out multiple high-impact incidents involving crypto theft and developer-targeted supply chain infiltration.

- January–May: Lazarus Group continued supply chain attacks targeting software developers, mainly including TasksJacker, PolinRider, and the Contagious Interview campaign. Attackers injected malicious .vscode/tasks.json files into GitHub repositories (which execute automatically when developers clone and open them in VS Code), or submitted malicious pull requests to inject payloads into popular open-source projects, and released malicious npm packages mimicking popular libraries. The attacks used the BeaverTail multi-stage loader and InvisibleFerret Python backdoor to steal crypto wallet credentials, browser data, and SSH keys. The campaign affected at least 400 repositories and thousands of developers, including organizations such as DataStax. The C2 infrastructure used multiple blockchains (TRON, Aptos, BSC, etc.) to enhance resilience against takedowns. The goal was to steal credentials and establish persistence in developer environments to facilitate future crypto asset theft.

- April 1: The Solana DeFi protocol Drift Protocol suffered a large-scale attack, resulting in approximately 285 million USD in losses. Drift released a post-incident analysis attributing this six-month-long social engineering campaign to the North Korea-linked hacker group UNC4736 (a Lazarus subgroup). Attackers first posed as a quantitative trading firm via third-party intermediaries, met Drift contributors at crypto conferences in multiple countries, created Telegram groups, and made deposits exceeding 1 million USD to build trust. They then induced contributors to clone malicious repositories containing infected VS Code configurations and downloaded malicious wallet applications via Apple TestFlight. Ultimately, attackers exploited a Security Council pre-authorized transaction mechanism and a vulnerability window without timelock to execute 31 malicious withdrawals within 12 minutes and erase traces. This incident was one of the largest DeFi thefts in H1 2026.



Attribution

With medium-high confidence supported by investigations done by the SEALS 911 team, this operation is assessed to have been carried out by the same threat actors responsible for the October 2024 Radiant Capital hack attributed by Mandiant to **UNC4736**, a North Korean state-affiliated group also tracked as AppleJeus or Citrine Sleet. The basis for this connection is both onchain (fund flows used to stage and test this operation trace back to the Radiant attackers) and operational (personas deployed across this campaign have identifiable overlaps with known DPRK-linked activity).

It is important to note that the individuals who appeared in person were not North Korean nationals. DPRK threat actors operating at this level are known to deploy third-party intermediaries to conduct face-to-face relationship-building.

Mandiant has not formally attributed this Drift exploit. That determination requires completed device forensics, which are still underway.

- April 18: The LayerZero-powered cross-chain bridge KelpDAO was attacked, resulting in approximately 292 million USD in losses. LayerZero Labs issued an official statement attributing the attack to the North Korean Lazarus Group (and its subgroup TraderTraitor). Attackers first compromised two downstream RPC nodes used by LayerZero validators (DVN) and injected malicious binaries, while launching DDoS attacks on other legitimate

nodes to force failover to compromised nodes. They then injected fake cross-chain data to mint and extract assets without collateral. The hackers subsequently used the minted tokens as fake collateral on lending platforms such as Aave, borrowing approximately 236 million USD in real assets (WETH), causing a major liquidity crisis and bad debt exposure in the DeFi sector. This was one of the most severe cross-chain infrastructure exploits in global crypto in H1 2026.

KelpDAO Incident Statement

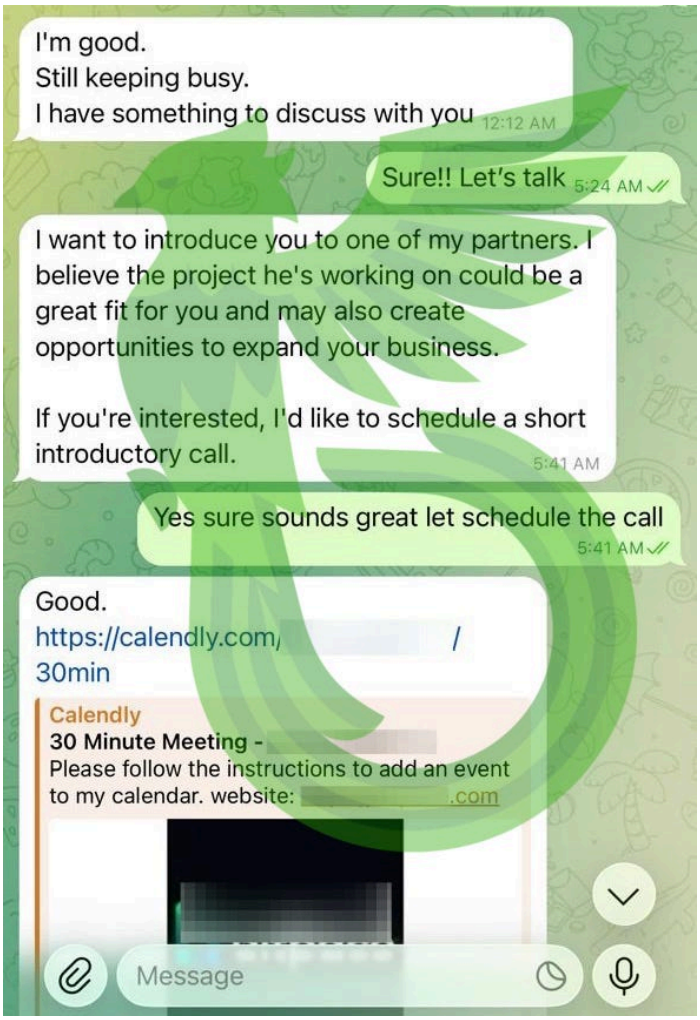
374 997 1.9K 1.9M

On April 18, 2026, KelpDAO was exploited for approximately . Preliminary indicators suggest attribution to a highly-sophisticated state actor, likely DPRK's Lazarus Group, more specifically TraderTraitor. This incident was isolated to KelpDAO's rsETH configuration as a direct consequence of their single-DVN setup. **There is zero contagion to any other cross-chain assets or applications.**

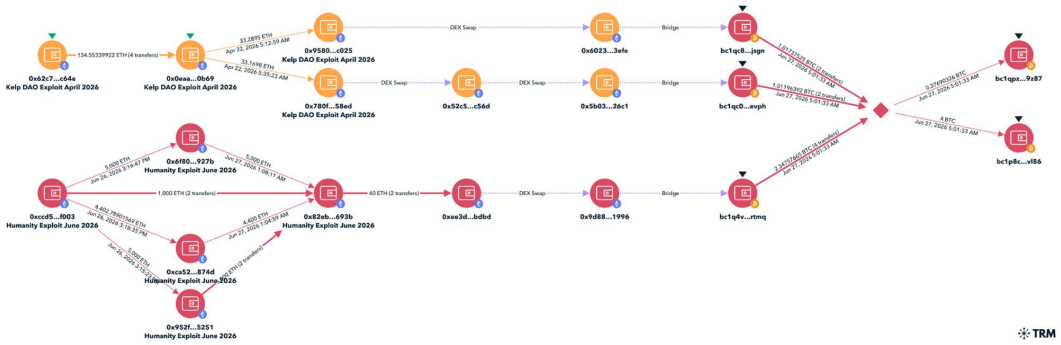
The subject of this highly-sophisticated attack was the poisoning of the downstream RPC infrastructure used by the LayerZero Labs DVN. All affected RPC nodes have been deprecated and replaced, and the LayerZero Labs DVN is now live.

We share these details to help the community better understand and guard against this emerging type of state-sponsored attack vector.

- April 21: The threat intelligence team Quetzal Team, together with sandbox platform ANY.RUN, disclosed and named a new Lazarus macOS-targeted campaign called "Mach-O Man," primarily targeting executives in crypto and fintech sectors. The attack used normal business communications as an entry point, deploying malware for credential theft and data exfiltration. It demonstrated Lazarus' evolution in using seemingly legitimate business interactions as attack vectors against high-value individuals.



- June 9: Humanity Protocol suffered a breach in which a developer's device was compromised, leading to the takeover of team and deployer addresses, resulting in approximately 32 million USD in losses. Attackers controlled multisig keys and transferred and minted large amounts of H tokens. Due to relatively centralized internal token control and active market making on centralized exchanges, as well as the timing before investor unlocks, there were initial suspicions of insider involvement. However, blockchain analyst ZachXBT traced the fund flows and found clear mixing patterns between this incident and funds from the earlier KelpDAO attack.



ZachXBT pointed out that this on-chain evidence largely rules out insider involvement and suggests a link between the attackers in both incidents. The KelpDAO attack is widely attributed to Lazarus Group. The mixing between Humanity Protocol and KelpDAO funds further strengthens the likelihood that Humanity Protocol is also connected to Lazarus Group.

3.3.2 Drainers

Drainer services (Drain-as-a-Service, DaaS) refer to illegal operations that provide tools and infrastructure to steal cryptocurrency from wallets by inducing victims to sign malicious transactions or approvals through phishing attacks. These services typically adopt a commission-sharing model, offering ready-made phishing toolkits, malicious contract templates, multi-chain support, and automated deployment capabilities for low-barrier scammers. Victims often connect their wallets and approve transactions on fake airdrop, verification, or dApp websites, resulting in assets being transferred within seconds. In the first half of 2026, although legacy services (such as Pink Drainer, which exited in 2024) gradually faded out, emerging and transformed specialized DaaS platforms quickly filled the gap.

(1) Lucifer DaaS (Lucifer Drainer)

An emerging specialized Drainer-as-a-Service platform that adopts a commission-sharing model (operators take 20% of successful “hits”), without directly selling software, but instead providing continuously updated infrastructure and support.

On May 21, the Flare Research team published an in-depth tracking report on BleepingComputer, based on cross-analysis of approximately 700 related posts collected from underground forums, dark web chats, and Telegram channels between January 2025 and early 2026, providing a detailed reconstruction of the internal operating mechanism of “Lucifer DaaS”.

The report states that the platform underwent frequent technical iterations in 2025. In its released v6.6.6 core version, it significantly enhanced advanced capabilities such as ERC-20 token Permit2 protocol abuse, off-chain signature fraud, multi-chain compatibility, automated Telegram real-time notifications, and wallet security bypass techniques. In addition, Lucifer demonstrated strong operational resilience: after its Telegram bot was banned in August 2025, operators immediately guided affiliates to build new bots and granted them admin privileges; after its official documentation domain was suspended in November 2025, it rapidly migrated its infrastructure to IPFS for decentralized storage. In the first half of 2026, its highly structured SaaS-style operational model (including continuous version updates, rapid bug fixes, automated website cloning tools, zero-configuration one-click deployment, and a mature affiliate recruitment mechanism) has been regarded by the security community as a typical example of modern industrialized Drainers, and is expected to remain highly active in the future.

（2）Rublevka Team

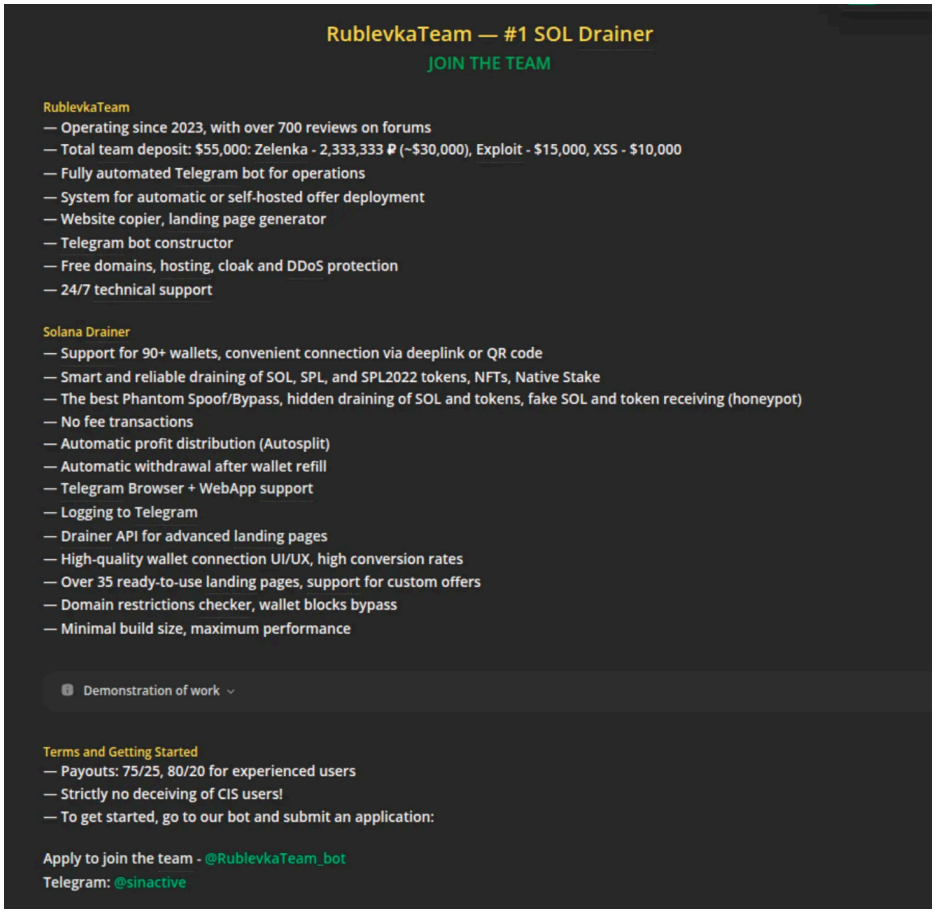
This group initially operated fake crypto exchanges in 2023 for traditional fraud, transitioned to custom malicious JavaScript scripts in 2024, and further shifted to the Solana ecosystem in spring 2025. According to threat intelligence disclosed by Insikt Group of Recorded Future, they developed custom malicious JavaScript scripts (usually disguised as index.js) and embedded them into carefully crafted fake airdrop and giveaway landing pages. Once users clicked to connect their wallet and signed seemingly normal “verification/claim” transactions, the script would instantly drain the wallet.

The operations of Rublevka Team rely heavily on automation tools and Telegram bot networks:

- Centralized technical platform: The core management team maintains the underlying Drainer scripts, configuration cloaking, and anti-DDoS infrastructure, and provides support for more than 90 wallet types (including Phantom, Bitget, Jito, etc.).

- Affiliate recruitment (Traffers): They recruit thousands of traffic operators skilled in social engineering through dark web forums and Telegram. These traffickers use official Telegram bots to generate phishing sites with Drainer scripts in one click.

- High-profit revenue sharing: According to on-chain flows and records from its automated "Profits" channel, the group has stolen over 10 million USD since 2023, with at least 240,000 successful drains recorded, and the highest single theft exceeding 20,000 USD. To incentivize affiliates, the group offers a 75% revenue share for beginners and up to 80% for experienced top traffickers.

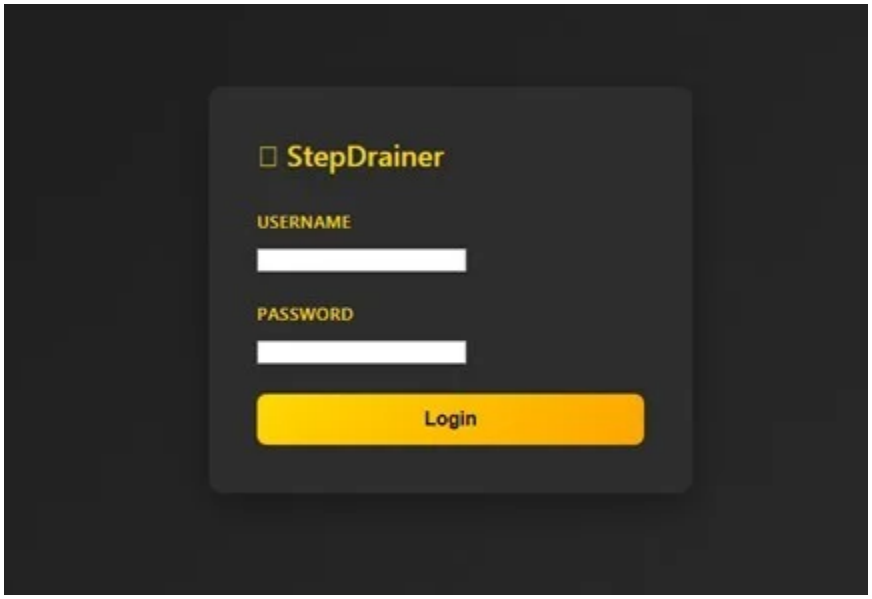


(3) StepDrainer

On April 23, a report published by LevelBlue SpiderLabs showed that StepDrainer remained under active monitoring and threat response in spring 2026. As a typical case in hybrid threat

ecosystems, StepDrainer is used to demonstrate the trend of browser Drainers integrating with traditional cybercrime infrastructure.

Its main characteristics include: using Web3Modal to present realistic wallet connection prompts; abusing Seaport (for NFTs), Permit v2 and other authorization mechanisms to automatically transfer high-value assets; using AI-themed phishing pages (e.g., impersonating OpenClaw AI assistant) to increase credibility; and supporting more than 20 blockchains (including ETH, BNB, Arbitrum, Polygon, etc.). The platform is distributed as MaaS (full source code priced at approximately 750 USD, shared versions at around 150 USD with a 20% commission). It reflects the convergence of Web3 phishing and traditional malware techniques, forming a hybrid attack chain across Web2 and Web3.



3.4 Privacy Protocols

Within the blockchain ecosystem, privacy protocols leverage technologies such as Zero-Knowledge Proofs (ZKPs), coin mixing mechanisms, encrypted UTXO models, and selective disclosure to provide users with transaction privacy. By reducing on-chain linkability between transactions, these protocols help mitigate risks associated with address profiling, fund tracing, and targeted attacks. As a result, they have long attracted attention from individual users, institutions, and DeFi participants.

In recent years, privacy protocols have evolved beyond standalone mixing services into a broader privacy infrastructure. On one hand, classic mixing protocols such as Tornado Cash continue to maintain significant on-chain activity. On the other hand, protocols like Railgun have extended privacy protection to DeFi interactions and asset management, while emerging projects including Hinkal and Privacy Pools are exploring mechanisms such as verifiable privacy and selective disclosure, aiming to balance user privacy with regulatory compliance.

This section is based on a combination of Dune Analytics dashboards and self-built query results, providing a statistical analysis of the fund inflows into major privacy protocols between January 1 and July 4, 2026, in order to examine the current development trends of the on-chain privacy ecosystem and its broader security implications.

(1) Overview of Cumulative Inflows

Between January 1 and July 4, 2026, the cumulative inflows to major privacy protocols totaled approximately USD 974 million. The breakdown is as follows:

Protocol	Cumulative Inflows (Jan. 1 – Jul. 4, 2026)	Dune Dashboard
TornadoCash	Approximately USD 691 million (about 360.3K ETH, 47.8K BNB, 20.18M DAI, 5.5 WBTC, and other assets, valued at current market prices)	Native assets; ERC-20
Railgun	Approximately USD 222 million (around USD 199 million in stablecoins and USD 22.93 million in non-stablecoin assets)	ETH; L2
Hinkal	Approximately USD 46.22 million (primarily stablecoins)	Multi-chain
zkBOB	Approximately USD 2.06 million (2.06M USDC on Optimism; no new USDC.e deposits on Polygon during the period)	Multi-chain
Privacy Pools	Approximately USD 12.65 million (around USD 10.28 million in stablecoins, plus 1,329.22 ETH and 3.53 BNB, valued at current market prices)	Stablecoins; Native assets

From the perspective of total inflows, Tornado Cash continues to dominate the privacy protocol landscape, attracting approximately USD 691 million, accounting for roughly 71% of all tracked inflows. Railgun ranked second with approximately USD 222 million, representing about 23%, while Hinkal, Privacy Pools, and zkBOB collectively accounted for the remaining 6%.

More noteworthy than the overall funding scale is the shift in the composition of deposited assets. With the exception of Tornado Cash, newly deposited funds in Railgun, Hinkal, and Privacy Pools were predominantly stablecoins. In particular, nearly 90% of Railgun's inflows consisted of stablecoins. This trend suggests that privacy protocols are evolving beyond their traditional role in anonymous withdrawals and are increasingly being adopted for stablecoin transfers, on-chain asset management, and DeFi interactions, reflecting the growing diversification of privacy-preserving use cases within the blockchain ecosystem.

(2) Analysis of Major Privacy Protocols

Tornado Cash remains the largest privacy protocol by total funds. As one of the most prominent decentralized mixing protocols, it has been used in certain cryptocurrency security incidents to anonymize illicit fund flows. Although it has faced regulatory actions in multiple jurisdictions, its decentralized architecture has enabled the protocol to remain widely used and maintain a high level of on-chain activity.

Railgun is built on an encrypted UTXO model and Zero-Knowledge Proofs (ZKPs) to provide users with private accounts that support both asset privacy and participation in DeFi applications. As a result, it has emerged as one of the most closely watched privacy protocols alongside Tornado Cash. Unlike Tornado Cash, which primarily focuses on anonymous transfers, Railgun emphasizes long-term private account management and composability with DeFi protocols. The steadily increasing share of stablecoin deposits further suggests that users are increasingly adopting privacy protocols for stable-value asset management, rather than solely for one-time anonymous fund transfers.

Hinkal provides privacy-preserving transaction capabilities for the EVM ecosystem while exploring optional Know Your Customer (KYC) verification and compliance mechanisms. Its goal is to

protect user privacy while addressing certain regulatory requirements. Compared with traditional mixing protocols, Hinkal places greater emphasis on the concept of Verifiable Privacy, positioning itself for institutional users and high-value asset management. Although its current fund volume remains relatively modest, its design philosophy reflects a broader industry trend toward privacy solutions that balance privacy protection with regulatory compliance.

zkBOB is a privacy-preserving stablecoin payment protocol built on Zero-Knowledge Proofs, primarily deployed on networks such as Optimism and Polygon. It enables users to make low-cost, private stablecoin transfers. Compared with other privacy protocols, zkBOB is more focused on everyday payments and small-value transactions, which explains its relatively limited fund volume. Nevertheless, its development highlights that privacy payment applications are still in an early stage of market adoption and require further ecosystem growth before achieving broader mainstream usage.

Privacy Pools is one of the most notable next-generation privacy protocols to emerge in recent years. By leveraging Association Sets and Zero-Knowledge Proofs, it enables users to preserve transaction privacy while proving that their funds are not associated with predefined sets of illicit assets. Although its current fund volume is significantly smaller than that of Tornado Cash and Railgun, the protocol's emphasis on combining privacy with compliance represents an important direction for the industry. Compared with traditional mixing protocols, Privacy Pools seeks to establish a new balance between anonymity and regulatory expectations, offering a promising technical approach for the future evolution of blockchain privacy infrastructure.

（3）Summary

Overall, major privacy protocols recorded nearly USD 1 billion in cumulative inflows during the first half of 2026, indicating that demand for on-chain privacy remains strong. Tornado Cash continued to dominate the sector, demonstrating the enduring relevance of the traditional mixing model, while protocols such as Railgun and Hinkal have further expanded the role of privacy infrastructure toward multi-asset management, DeFi interactions, and stablecoin-based applications.

One notable trend is that stablecoins have become the primary inflow asset for most privacy protocols other than Tornado Cash. Compared with the early usage pattern, which was largely centered on anonymous ETH transfers, stablecoins—with their lower price volatility—are better suited for on-chain treasury management, cross-protocol interactions, and value storage. This shift suggests that the use cases of privacy protocols are expanding well beyond one-time anonymous transfers. Meanwhile, emerging protocols such as Privacy Pools and Hinkal are introducing mechanisms including Association Sets and optional Know Your Customer (KYC) verification, seeking to improve the verifiability of fund provenance while preserving transaction privacy. These developments reflect the industry's ongoing efforts to establish a better balance between privacy protection and regulatory compliance.

From a security perspective, the continued evolution of privacy protocols presents both opportunities and challenges. On one hand, they enhance users' ability to defend against on-chain profiling, address clustering, and targeted attacks. On the other hand, they significantly increase the complexity of fund tracing, anti-money laundering (AML) investigations, and risk detection for security researchers and compliance teams. Looking ahead, as technologies such as Zero-Knowledge Proofs (ZKPs), Account Abstraction, and cross-chain infrastructure continue to mature, on-chain fund flows are expected to become increasingly sophisticated and privacy-preserving. As a result, traditional analysis methods that rely primarily on address linkability will face growing limitations. Future blockchain investigations will increasingly depend on a combination of entity labeling, behavioral modeling, temporal analysis, and cross-chain tracing to improve the identification of suspicious fund flows and support more effective risk assessment.

IV. Conclusion

Looking back at the first half of 2026, blockchain security risks continued to evolve. The attack surface further expanded from smart contracts to broader ecosystem components such as developer supply chains, endpoint devices, browser extensions, and AI agents, while attack methods became more intelligent and persistent. At the same time, on-chain fund movement and laundering activities remained active. Global regulatory frameworks continued to improve in key areas such as anti-money laundering (AML), stablecoins, and virtual asset service providers

(VASPs), driving the industry from reactive response toward risk prevention and systematic governance. Against the backdrop of parallel technological innovation and risk evolution, improving the overall security capability of the blockchain ecosystem has become a key foundation for long-term healthy development.

Facing continuously evolving security challenges, SlowMist has consistently focused on advancing security capabilities through technological innovation, continuously exploring the deep application of artificial intelligence in threat intelligence, on-chain tracing, risk analysis, and AML. Around AI Agent security, SlowMist has built a “five-layer progressive digital fortress” protection system and launched capabilities such as SlowMist Agent Security Skill, MistTrack Skills, and MistEye Security Gate, helping developers and enterprises enhance the security resilience of AI agents and actively respond to emerging risks such as prompt injection and supply chain poisoning.

V. Disclaimer

The content of this report is based on our understanding of the blockchain industry, data from the SlowMist blockchain hacked archive database SlowMist Hacked, and the anti-money laundering tracking system MistTrack. However, due to the "anonymous" nature of blockchain, we cannot guarantee the absolute accuracy of all data and cannot be held responsible for errors, omissions, or losses caused by using this report. Additionally, this report does not constitute any investment advice or the basis for other analyses. We welcome criticism and corrections for any oversights or inadequacies in this report.

VI. About Us

SlowMist is a threat intelligence firm specializing in blockchain ecosystem security, established in January 2018. The firm was founded by a team with over ten years of network security experience. Our goal is to make the blockchain ecosystem as secure as possible for everyone. We are now a renowned international blockchain security firm that has worked on various well-known projects such as HashKey Exchange, OSL, MEEX, BGE, BTCBOX, Bitget, BHEX.SG, OKX, Binance, Amber Group, Crypto.com, etc.

SlowMist was awarded the “Cyber Security Excellence Contribution Award” at the Cyber Security Professionals Awards 2025, presented by the Hong Kong Police Force (CSTCB). In addition, our AML tracking and analytics platform MistTrack received the Gold Award in FinTech (RegTech: Regulatory and Risk Management) at the HKICT Awards 2025. Our extensive work in cryptocurrency crime investigations has been cited by international organizations and government bodies, including the United Nations Security Council and the United Nations Office on Drugs and Crime.

SlowMist's security solutions include services such as SlowMist AI, security auditing, Blockchain Threat Intelligence (BTI), and defense deployment, complemented by a suite of SaaS security products, including the cryptocurrency AML tracking and analytics platform (MistTrack), the AML compliance engine for large institutions (SlowMist KYT), fake deposit vulnerability scanning, Web3 threat intelligence and dynamic security monitoring (MistEye), and the hacked incident

database (SlowMist Hacked). We have partnerships with leading domestic and international firms such as Akamai, Bitdefender, RC², TianJi Partners, IPIP, etc.

By delivering comprehensive security solutions customized for individual projects, we help identify and mitigate risks. Our team has discovered and disclosed several high-risk blockchain vulnerabilities, contributing to greater awareness and improved security standards across the blockchain ecosystem.

SlowMist Security Solutions

Security Services

Exchange Security Audits

Full range of black box and gray box security audits, going beyond penetration testing

Wallet Security Audits

Full range of black box and gray box security audits, going beyond penetration testing

Blockchain Security Audits

Comprehensive audit of key vulnerabilities in Blockchain and consensus security

Smart Contract Audits

comprehensive white box security audit of source code related to smart contracts

Red Teaming

Penetration testing and evaluating vulnerable points

Security Monitoring

Dynamic security monitoring for all possible vulnerabilities

Blockchain Threat Intelligence

Joint defense system with integrated on-chain and off-chain security governance

Defense Deployment

Deploying Defense Solutions Tailored to Local Conditions, Implementing Hot Wallet Security Strengthening

MistTrack Tracking Service

Digital assets were unfortunately stolen, MistTrack saves a glimmer of hope

Incident Response Service

Aiming to help Web3 projects quickly and effectively respond to security incidents and threats

Security Consulting

Provide technical, risk management, and emergency response support as well as providing recommendations to improve them

SlowMist AI

Security for AI & Crypto · AI for Security

Hacking Time

Annual close-door training focusing on blockchain security

Digital Asset Security Solution

Open source digital asset security solutions

Security Products

SlowMist AML

Empowering the compliant, secure, and healthy development of the Web3 industry.

SlowMist KYT

A professional, real-time, and configurable anti-money laundering (AML) engine designed for compliance teams at large institutions.

MistTrack

A cryptocurrency tracking and compliance platform for everyone (C-end users).

MistEye

Providing comprehensive Web3 threat intelligence and dynamic security monitoring services for everyone.

SlowMist Hacked

Comprehensive coverage of blockchain attack incidents.

False Top-up Vulnerability Scanner

A powerful security tool ensuring safe deposits and withdrawals for trading platforms.

Official website

https://slowmist.com

X

https://x.com/SlowMist_Team

Github

https://github.com/slowmist

Medium

https://slowmist.medium.com

Email

team@slowmist.com

WeChat Official Account

SLOWMIST

A global leader in blockchain threat intelligence