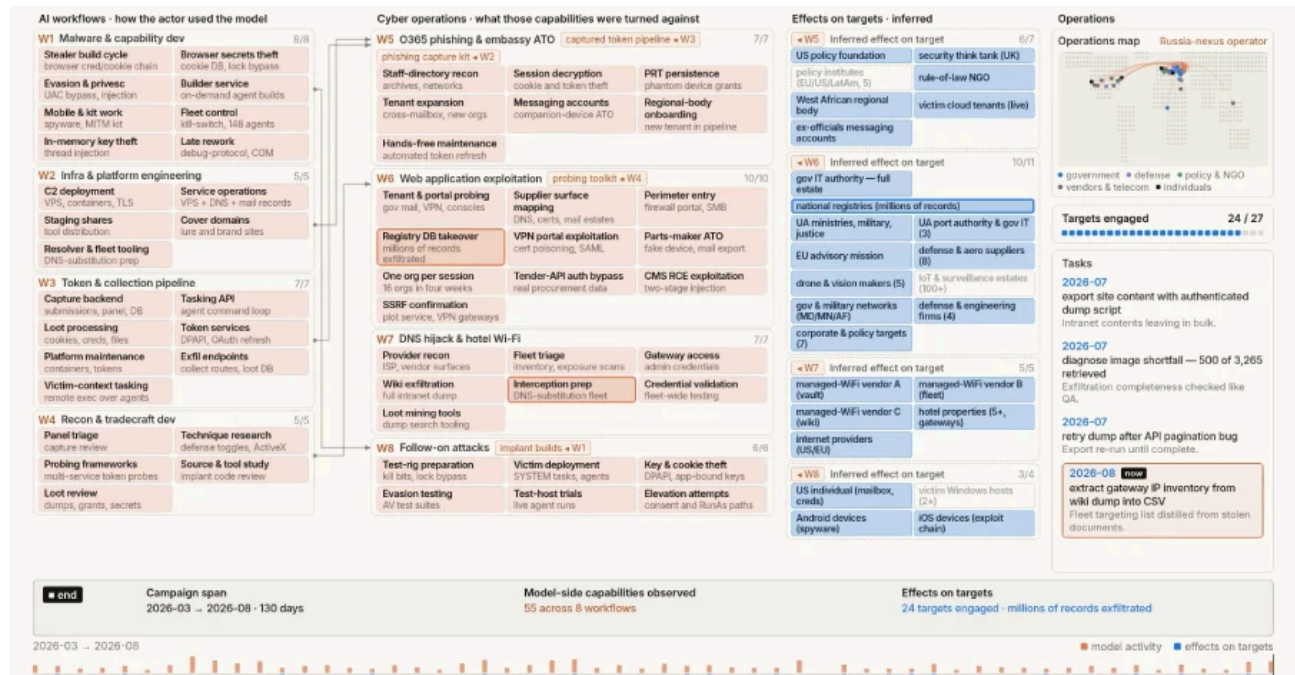


Russian State-Sponsored Hackers Use Claude to Rebuild Malware After Detection

By The Hacker News

Published: 2026-09-11 · Archived: 2026-09-19 02:00:15 UTC



Anthropic on Thursday [revealed](#) it disrupted a campaign mounted by a Russian state-sponsored threat actor that abused Claude for developing an AI-assisted workflow to get ahead of the detection curve.

The operation has been attributed to a cyber espionage group it calls **GTG-20006** (where "GTG" stands for Generative Threat Group), which aligns with broader reporting linking the cluster to [Midnight Blizzard](#) (aka APT29 and Cozy Bear).

This actor is said to have developed an AI-driven process to automatically rebuild and re-deploy their toolkit if it was detected by security products, thereby undermining defenders' ability to block the artifacts via static detections.

Attacks mounted by GTG-20006 have targeted military intelligence targets in Ukrainian and European governments, along with diplomatic and defense organizations and individuals connected to U.S. foreign policy.

The toolkit includes a number of programs -

- Two Windows-based implants
- A mobile exploitation kit
- A credential stealing tool that targets browser password stores
- A phishing platform designed to mimic priority targets like government organizations, and

- An administrative console used to manage compromised accounts

"The actor also used AI to monitor how well their tools evaded detections from known security defenses," Anthropic explained. "If their monitoring AI agents identified that any of their deployed malware was detected by a security product, agents would then set about the process of autonomously modifying and rebuilding the malware to evade the existing detections."



Once the artifacts can bypass detection, they are staged on disposable hosting servers to which victims are redirected to so as to retrieve the malware via phishing, [ClickFix](#), and DNS hijacking schemes.

The threat actor has also been observed using AI workflows to register domains, set up the hosting infrastructure used to send phishing emails, as well as to deliver the messages and monitor command-and-control (C2) channels for successful compromises.

More than 20 distinct organizations were singled out over the course of the reconnaissance and live operations. This included government ministries, defense and intelligence bodies, embassies and diplomatic missions, think tanks, and defense-industrial companies, mainly in Ukraine and Europe. The attacks also extended to the Middle East and maritime-related government agencies in Asia.

These efforts also overlapped with a campaign dubbed [CaptiveCrunch](#) that was documented in July and August 2026 by ReliaQuest, Microsoft, Google, and Lumen Black Lotus Labs.

"The actor compromised at least three hospitality vendors that operate hotel guest Wi-Fi," Anthropic said. "They used compromised admin credentials to modify DNS records so that they pointed to services owned by the actor (a technique known as DNS hijacking). Guests of hotels using the compromised vendors who connected to the hotel Wi-Fi had their traffic, device identifier, and IP address sent to the actor's servers."

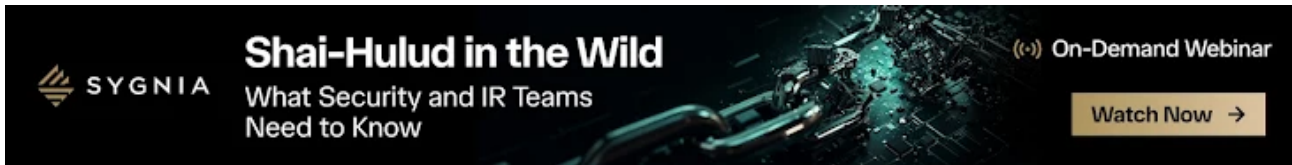
In the next stage, victims were served ClickFix-style lures to deliver Windows, Android, and iOS malware tailored to their device -

- Windows - PowerChrome, WUEngine, Shadow C2, MiniPlasma, CloudSyncSvc
- Android - GiftDrop, a rebranded version of GiftsExpress Android surveillance RAT
- iOS - [DarkSword](#)

Furthermore, the threat actor has been found to use data stolen from the hotel management systems and the individual guests' devices to identify additional targets, particularly individuals associated with Ukraine, such as government officials and drone manufacturers.

This is complemented by attempts to take over victims' WhatsApp accounts using headless browsers to link victim accounts as companion devices and ultimately bulk-exporting Russian and Ukrainian language conversations from them while suppressing read receipts.

"The actor also targeted surveillance platforms," Anthropic said. "They found authorization flaws in the application interface of camera streaming services, and from there they enumerated users and harvested tokens that granted them access to the victims' live camera streams."



GTG-20006 has been attributed to an intrusion targeting a North African government technology authority, leveraging credentials to a VPN appliance to hijack the central account server and exfiltrate the entire credential database consisting of over 300,000 national identity records and the commercial registry data of more than half a million companies operating in the country.

Also developed by the threat actor is a cloud email espionage platform, which used a device code phishing framework codenamed Embassy Kit to orchestrate a Microsoft 365 token theft campaign targeting diplomatic and government personnel, resulting in the unauthorized access and exfiltration of mail records from at least eight organizations, including a national prosecutor's office, a military education institute, and a regional intergovernmental organization.

The threat actor has also been observed delivering Windows credential stealers via fake update-themed social engineering lures, along with auxiliary tools for facilitating remote access and tampering with the victim machine's security updates so that the artifacts remain undetected.

"The actor used AI at every point in their operations," Anthropic said. "In on-premises environments, the actor used AI to monitor the stealth and persistence of their implants. "The result of the above is that AI has inverted the cost back onto defenders. Previously, defenders might have been able to slow an attacker's operational tempo via the deployment of a new detection."

Found this article interesting? Follow us on [Google News](#), [Twitter](#) and [LinkedIn](#) to read more exclusive content we post.

Source: <https://thehackernews.com/2026/09/russian-state-sponsored-hackers-use.html>