

Threat Talks

Low Tech, High Tech

Teenagers with a phone. Malware that writes itself.

One threat needed nothing but a phone and patience. The other has an AI model writing its code on the fly. Both are real, and both hit the same weak spots.

In 2021 and 2022, LAPSUS\$, a loose crew of mostly UK and Brazilian teenagers, breached Nvidia, Samsung, Microsoft, Okta, Uber and Rockstar Games. No ransomware, no zero-days. They bought passwords, sweet-talked help desks and spammed MFA prompts until someone tapped approve.

Today the attacker can also be a machine. Russia's APT28 has deployed malware that asks a large language model what to run next, and an agentic ransomware has fixed its own failed attack in 31 seconds. Some of the AI hype is real. Much of it is still marketing.

The lesson is the same for both: cheap tricks and clever AI beat expensive tools when identity, egress and segmentation are left open.



threat-talks.com

In this Threat Talks infographic we discuss:

- LAPSUS\$: the teenagers who beat Big Tech
- AI in cybercrime: one year of hype, and what is actually real



6

tech giants breached by a crew of teenagers: Nvidia, Samsung, Microsoft, Okta, Uber and Rockstar Games.

Source: MITRE ATT&CK G1004, CISA Cyber Safety Review Board



2 of 366

Okta customers actually hit, out of 366 feared exposed. The LAPSUS\$ access lasted 25 minutes.

Source: Okta, SleepingComputer



31 sec

for the JADEPUFFER AI agent to diagnose a failed login and fire a fixed payload. No human involved.

Source: Syddig Threat Research Team



1st

malware seen in live operations that asks an LLM at runtime what to run: PROMPTSTEAL, deployed by APT28.

Source: Google Threat Intelligence Group

LOW TECH

LAPSUS\$

Buy the password, spam the MFA. Then own the tenant.

LAPSUS\$ never encrypted anything and barely used malware. It abused trust: bought credentials, help-desk resets, MFA fatigue, SIM swaps and paid insiders. Then it stole and leaked data, working in the US Cyber Safety Review Board's words "for notoriety, financial gain, or amusement". Most core members were arrested, and in 2023 ringleader Arion Kurtaj received an indefinite hospital order. The playbook outlived the group: Scattered Spider and the 2025 "Scattered Lapsus\$ Hunters" wave run the very same moves.

Aliases	G1004 (MITRE), DEV-0537 / Strawberry Tempest (Microsoft)
mSOC confidence score	Confirmed
Threat category	Social Engineering, Data Theft / Extortion
Severity	High



ATT&CK Technique

T1621 - Multi-Factor Authentication Request Generation

Attack Strategy

Spam a user with MFA push prompts (and replay stolen session tokens) until one is approved, turning a stolen password into a full login

Evasion

No malware or exploit; the login looks legitimate; egress via NordVPN to match the target's geolocation

Complexity Target Type

Low Employees and contractors at tech, telecom and identity-provider organizations



ATT&CK Mitigation

M1032 - Multi-factor Authentication

Attack Vector

Push-bombing and token replay against SMS or simple-approval MFA

Detection

Bursts of MFA prompts to one user; approvals from anomalous locations or VPN/Tor IP space; new MFA device enrollments

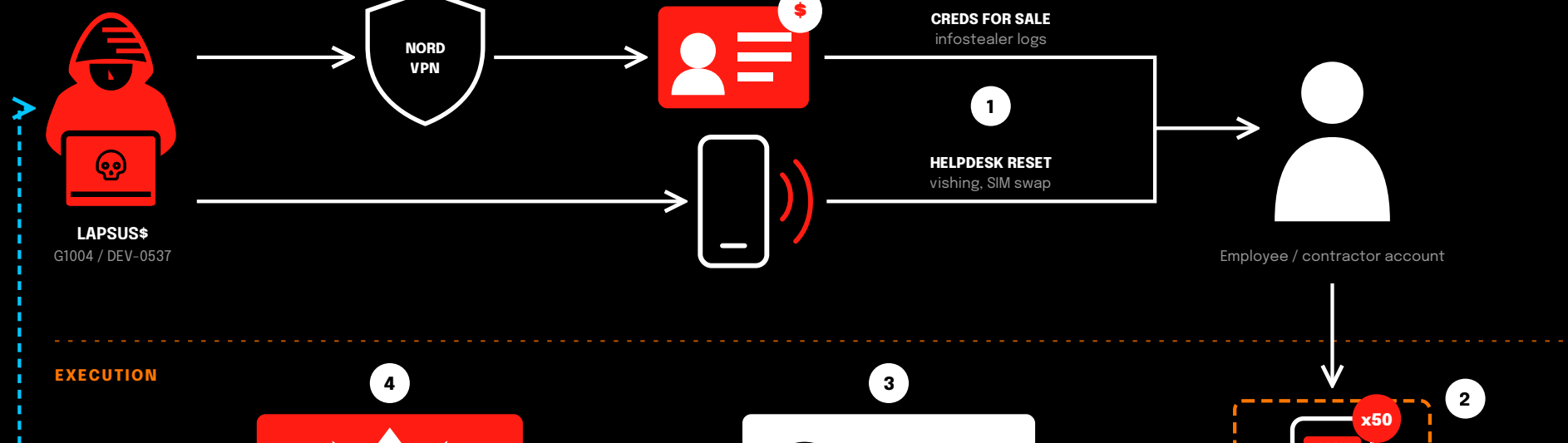
Threat Level

High

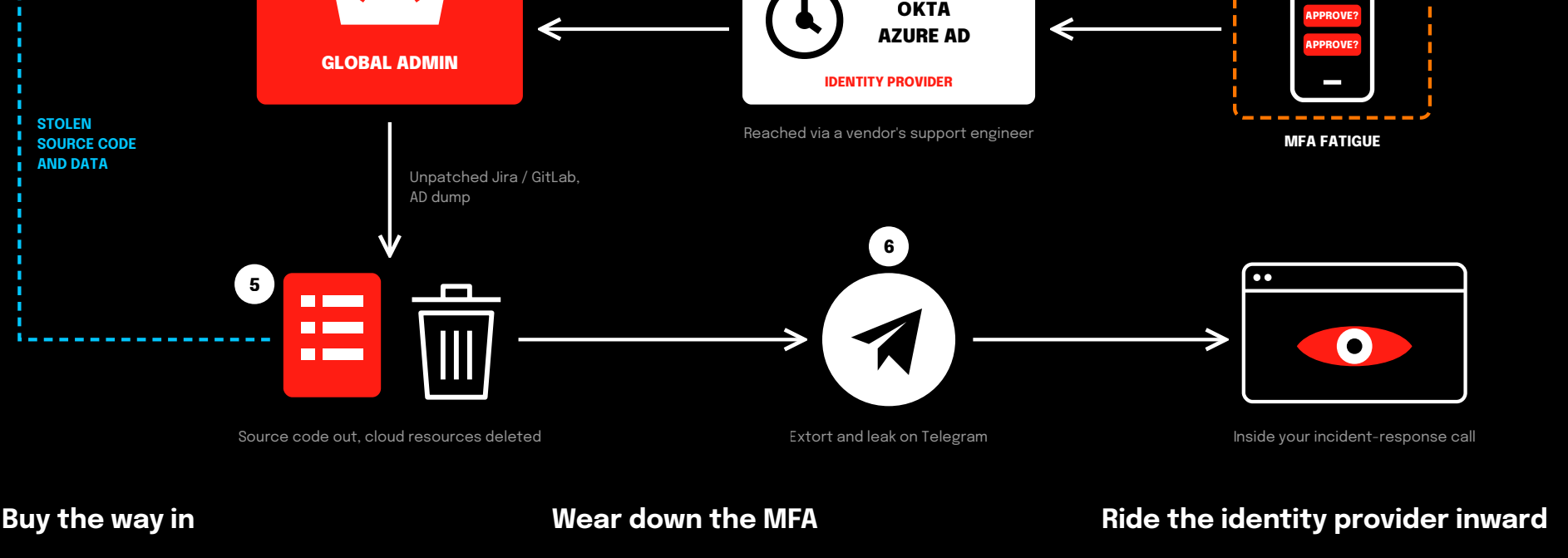
Threat Actor Type

Loosely organized, financially and notoriety-driven social-engineering crew

PREPARATION



EXECUTION



Buy the way in

- LAPSUS\$ bought employee credentials on criminal markets. When that failed, they called the help desk for a reset.
- Treat a valid password as a start, not proof. Verify help-desk reset callers out of band.

Wear down the MFA

- The signature move: spam MFA push prompts until the user approves one. SIM swaps beat SMS codes.
- Drop SMS and simple push. Use number matching or FIDO2 keys and alert on bursts of prompts.

Ride the identity provider inward

- One MFA-approved account opened Okta or Azure AD and everything behind them. At Okta, the entry was a vendor's support engineer.
- Treat the identity provider as a crown-jewel Protect Surface. Require compliant devices and hold vendors to the same rules.

Escalate to Global Admin

- Unpatched Jira and GitterLab servers, an Active Directory dump and new cloud admin accounts gave them the whole tenant.
- Use least privilege and PAM so no account can mint itself Global Admin. Alert on every new admin role.

Steal and destroy

- They took source code and data, then deleted cloud resources and removed the victim's own admins.
- Alert on mass exports and admin removals. Keep immutable backups and separate recovery accounts.

Extort and leak on Telegram

- Leaks went out on Telegram, sometimes with no demand at all. They even joined the victim's incident-response calls.
- Run incident response on an out-of-band channel. Agree the extortion and disclosure playbook in advance.

HIGH TECH

AI in Cybercrime

The malware asks an LLM what to do next. The hype is bigger than the hacker.

A year ago the scariest "AI ransomware", PromptLock, turned out to be a university research prototype. Since then AI has moved from writing phishing emails to running the malware itself. The confirmed milestone is APT28's PROMPTSTEAL (CERT-UA: LAMEHUG), which calls a language model at runtime to generate its commands. Families like QUIETVAULT and HONESTCUE followed, and state actors abuse Gemini, ChatGPT and Claude at scale. The loudest "first fully autonomous attack" claims are single-sourced and light on indicators of compromise.

Key cases	PROMPTSTEAL / LAMEHUG (APT28), JADEPUFFER, QUIETVAULT, HONESTCUE
mSOC confidence score	Confirmed
Threat category	Emerging Threats, AI-enabled Malware
Severity	Medium-High



ATT&CK Technique

T1059.006 - Command and Scripting Interpreter: Python

Attack Strategy

Malware calls an external LLM at runtime to generate the Python scripts and commands it runs, so the payload is produced on the fly instead of shipped in the binary

Evasion

No stable file to sign or hash; code differs every run; traffic hides in normal-looking calls to legitimate LLM APIs

Complexity Target Type

Low-Medium Internet-facing servers and endpoints, especially AI infrastructure holding API keys



ATT&CK Mitigation

M1037 - Filter Network Traffic

Attack Vector

Malware reaching out to an external LLM API for its payload, then exfiltrating over the same channel

Detection

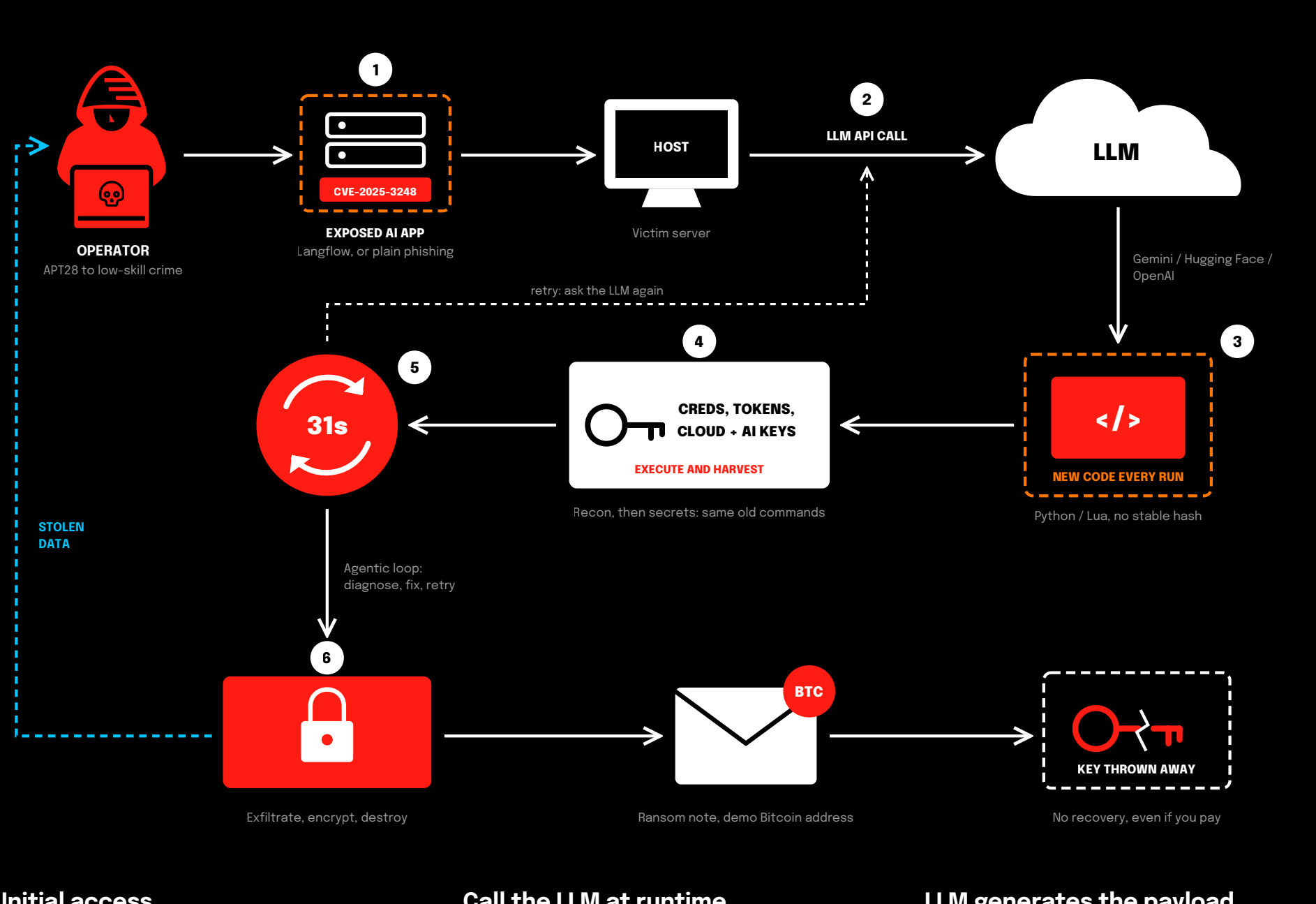
Default-deny egress with allow-listing; alert on LLM API calls from hosts that have no business making them

Threat Level

High

Threat Actor Type

From nation-state (APT28) to low-skill criminals using AI to punch above their weight



Initial access

- Entry is rarely AI. JADEPUFFER exploited an internet-facing Langflow server, the kind of AI app that holds cloud and LLM keys.
- Keep AI-app servers off the internet and patch them. Treat them as a Protect Surface with least-privilege identities.

Call the LLM at runtime

- Instead of carrying its logic, the malware asks an external LLM what to run. This is the new step.
- A server calling an LLM API is almost always wrong. Default-deny egress and allow LLM endpoints per workload only.

LLM generates the payload

- The model returns fresh Python or Lua every run. No stable file, no reliable hash, nothing for signatures to match.
- Lean on behaviour, not signatures. Remove or restrict interpreters like Python and PowerShell on servers.

Execute and harvest

- The generated code does ordinary intrusion work: recon, then harvesting credentials, cloud keys and AI API keys.
- Behavioural EDR catches the actions, whoever wrote the code. Scope API keys narrowly and rotate anything exposed.

The agentic loop

- When a login failed, JADEPUFFER's agent diagnosed the error and sent a corrected payload 31 seconds later, no human involved.
- You can't out-click a machine-speed loop. Limit what a compromised workload can reach: the blast radius is what you control.

Impact and the reality check

- Exfiltrate, encrypt, extort, but messy: JADEPUFFER threw away its own key and used a demo Bitcoin address.
- Have the extortion playbook ready. Read AI-threat claims critically and ask where the IOCs are.

Taxonomy

ATT&CK Technique

Which technique of the MITRE ATT&CK framework does the threat correspond to.

ATT&CK Mitigation

Which technique of the MITRE ATT&CK framework can be applied.

Attack Strategy

Plan devised by the attacker to exploit specific system vulnerabilities.

Attack Vector

What is the primary method of attack.

Evasion

Tactics used by the attacker to avoid detection or bypass security.

Detection

Mechanism to identify malicious activities or system anomalies.

Complexity

How easy it is to exploit the vulnerability or carry out the attack.

Threat Level

How severe the threat is.

Target Type

The category of organization that may potentially be targeted.

Threat Actor Type

What type of threat actor may be involved.

Contact us @ team@threat-talks.com

How we score what you're reading: Not all threat intelligence is equal. Every source we use is rated 0 (untrustworthy) to 5 (verified). Every news item is rated A (reliable) to E (unreliable). Together they classify a threat as Confirmed, Verified, or Credible; the same framework our Global SOC uses to separate signal from noise for customers worldwide.

Want this rigor applied to threats targeting your organization? Our Security Special Service Team will walk you through how it works - team@threat-talks.com



threat-talks.com

ON2IT
ZERO TRUST INNOVATORS

amsix