WHITE PAPER

CYTEX™
AI-Powered Resilience Command Center

600+ FortiGate Firewalls Hacked by an Amateur With AI

hello@cytex.io | https://cytex.io

AI-Powered Attacks Hack 600+ FortiGate Firewalls

600+ Fortinet FortiGate firewall instances across 55 countries have been compromised in a campaign that demonstrates a troubling new reality: unsophisticated threat actors can now execute large-scale attacks using commercial generative AI. The hackers behind this operation leveraged at least two LLMs to plan their campaign, generate custom tools, and assess success rates, all while possessing relatively low technical expertise. The result is a blueprint for how AI is democratizing cyberattacks at scale.

AI-Powered Attack Chain: Planning with AI

The threat actor used commercial LLMs to design the entire operation. The AI generated technically accurate command sequences, referenced academic research on offensive AI agents, and provided guidance on timing and success probability. However, when conditions deviated from the plan, the human operator struggled to adapt, a telling sign of limited technical depth.

AI-Generated Tooling

Researchers identified multiple custom scripts on the attacker's infrastructure that were almost certainly AI-generated. These scripts handled:

- Parsing firewall configurations.
- Extracting credentials.
- Automating VPN connections.
- Performing mass scanning.
- Aggregating results across compromised devices.

The volume and variety of custom tooling would typically indicate a well-resourced development team. Instead, a single actor or very small group generated the entire toolkit through AI assistance.

The Target: Exposed FortiGate Firewalls

The attackers did not exploit zero-day vulnerabilities. They focused on fundamental security gaps:

- Scanning for management interfaces exposed on ports 443, 8443, 10443, and 4443.
- Using common and weak credentials for initial access.
- Exploiting single-factor authentication across VPN and administrative accounts.

Scale

Compromised devices spanned 55 countries across Africa, Asia, Latin America, North America, and Europe. Some organizations had multiple devices breached, pointing to managed service provider deployments or large enterprise networks.

Post-Compromise Activity

Once inside, the attackers leveraged open-source offensive tools to:

- Extract NTLM password hashes.
- Obtain complete domain credential databases.
- Move laterally through pass-the-hash and pass-the-ticket attacks.
- Target Veeam Backup & Replication servers, likely to extract credentials and destroy backups in preparation for ransomware.

The Attacker Profile: An amateur with an AI assistant

Researchers assess with moderate confidence that this is a financially motivated, Russian-speaking threat actor with low-to-medium technical capability. The extensive reliance on AI across all operational phases: planning, tool generation, and execution, is the tell.

The Takeaway

- Commercial AI services are enabling even unsophisticated actors to conduct cyberattacks at scale.
- This campaign succeeded not through novel exploits, but through a combination of exposed management interfaces, weak credentials, and single-factor authentication, fundamental security gaps that AI helped an amateur exploit.

Immediate Actions for FortiGate Users

- Ensure management interfaces are not exposed to the internet.
- Change all default and common credentials on FortiGate appliances, including VPN user accounts.
- Rotate all SSL-VPN user credentials, especially if any management interface was internet-accessible.
- Implement MFA for all administrative and VPN access.
- Review configurations for unauthorized administrative accounts or policy changes.
- Audit VPN connection logs for logins from unexpected locations. → Isolate backup servers from general network access.
- Monitor for unauthorized PowerShell module loading on backup servers.

Post-exploitation detection

Organizations that may have been affected should monitor for:

- Unexpected DCSync operations (Event ID 4662 with replication-related GUIDs)
- New scheduled tasks named to mimic legitimate Windows services
- Unusual remote management connections from VPN address pools
- LLMNR/NBT-NS poisoning artifacts in network traffic
- Unauthorized access to backup credential stores
- New accounts with names designed to blend with legitimate service accounts

Indicators of compromise (IOCs)

IOC Value	IOC Type	First Seen	Last Seen	Annotation
212[.]11.64.250	IPv4	1/11/2026	2/18/2026	Threat actor infrastructure used for scanning and exploitation operations
185[.]196.11.225	IPv4	1/11/2026	2/18/2026	Threat actor infrastructure used for threat operations

Strong security fundamentals remain the most powerful defense against AI-augmented threats.