Hive Pro

Threat Level

Red

HiveForce Labs

THREAT ADVISORY

ATTACK REPORT

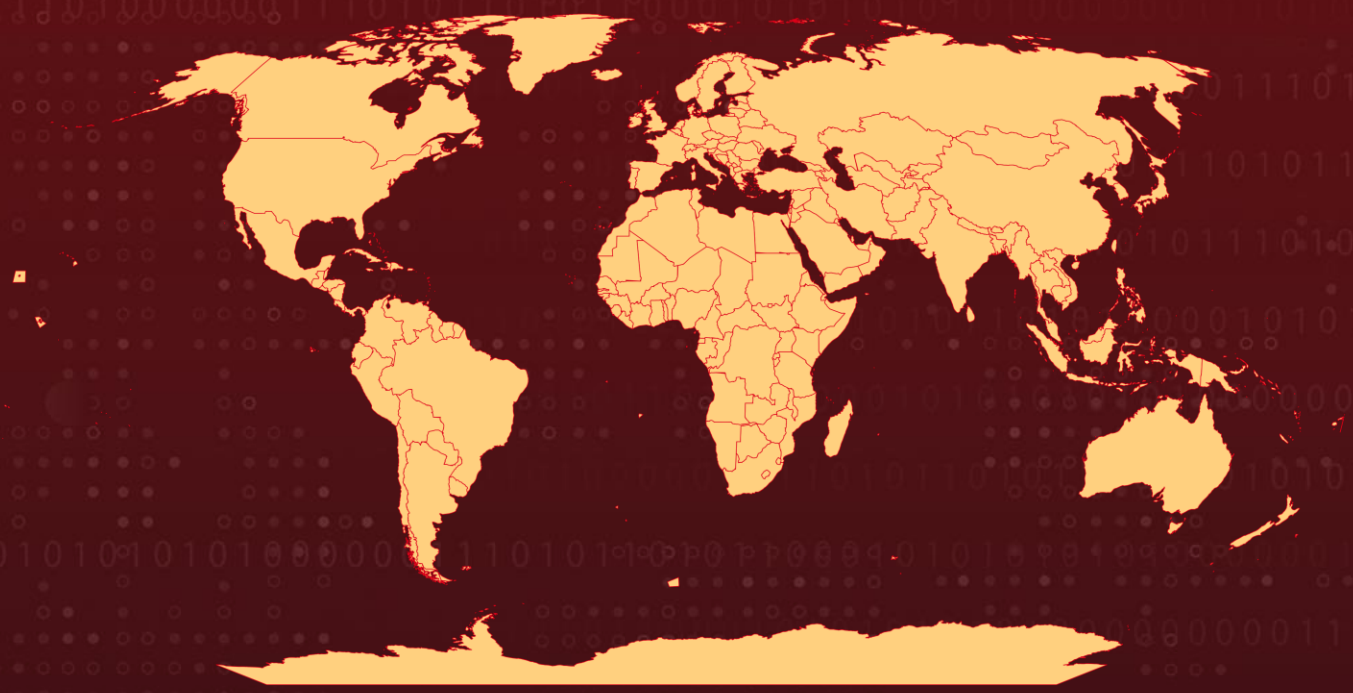
LLMjacking: An Attack Method for Stealing Cloud Credentials

Date of Publication	Admiralty Code	TA Number
May 13, 2024	A1	TA2024183

Summary

Discovered On: May 2024
Affected Region: Worldwide
Affected Platform: Cloud-based large language models (LLMs)
Attack: A recent attack, termed LLMjacking, has emerged, utilizing stolen cloud credentials to target cloud-hosted large language model (LLM) services. These credentials were acquired from a widely targeted system running a vulnerable version of Laravel (CVE-2021-3129).

Attack Regions



CVE

CVE	NAME	AFFECTED PRODUCTS	ZERO-DAY	CISA	PATCH
CVE-2021-3129	Laravel Ignition File Upload Vulnerability	Laravel Ignition	✗	✓	✓

Attack Details

#1 A recent cyberattack, named LLMjacking, has specifically targeted cloud-hosted large language model (LLM) services, utilizing stolen cloud credentials. These credentials were obtained from a system running a vulnerable version of Laravel (CVE-2021-3129). The attackers' goal is to sell LLM access to other cybercriminals while leaving the cloud account owner responsible for the associated costs.

#2 Upon gaining initial access, they proceed to exfiltrate cloud credentials and further infiltrate the cloud environment. The attacker strategically utilized the InvokeModel API call to assess the extent of their access to cloud services. By intentionally setting the max_tokens_to_sample parameter to -1 and checking the response for ValidationException, they confirmed the availability of access to the LLMs and their operational status.

#3 They also use the GetModelInvocationLoggingConfiguration for checking the service logging configuration. However, it's worth noting that certain LLM Providers don't capture specifics regarding prompt input and output, which poses challenges in effectively monitoring legitimate usage of their services.

#4 This attack also revealed a script that can check for stolen credentials across 10 different AI services. The attacker could also possibly setup an OAI Reverse Proxy, an open-source project designed to function as a reverse proxy for LLM services. Through Reverse Proxy, attackers can centrally administer access to multiple accounts without revealing compromised credentials. This capability enables attackers to potentially monetize their activities by offering access to LLM models for sale.

#5 In an LLMjacking attack, the damage manifests in the form of escalated costs to the victim, given the substantial expenses associated with using an LLM, which can accumulate rapidly. If left undetected, the attack could result in over $46,000 of LLM consumption costs per day for the victim. It's crucial for organizations to monitor their utilization of LLM services, and cloud providers offer tools to facilitate this task.

Recommendations

Vulnerability Management: Implement a robust vulnerability management process to ensure that software and systems are regularly assessed for vulnerabilities and updated with the required security patches. Prioritize critical vulnerabilities identified by security advisories and vendors to mitigate the risk of exploitation by threat actors.

Use a Dedicated Secrets Management Solution: Employ a specialized secrets management solution, these tools are designed specifically for securely storing and managing sensitive information like passwords, API keys, and certificates.

Detailed Logging: Enabling detailed logging is crucial, especially since the log of the InvokeModel command does not inherently capture details about the prompt input and output. Adopt comprehensive logging mechanisms ensuring that all facets of model usage are effectively monitored and archived for subsequent analysis and compliance purposes.

Comprehensive Cloud Security Coverage: CSPM solutions offer holistic visibility into cloud environments, identifying security risks, misconfigurations, and compliance gaps across multiple cloud platforms. Similarly, CIEM tools focus on managing permissions and access entitlements, mitigating the risk of unauthorized access or misuse.

Potential MITRE ATT&CK TTPs

TA0042 Resource Development	TA0001 Initial Access	TA0002 Execution	TA0003 Persistence
TA0006 Credential Access	TA0007 Discovery	TA0010 Exfiltration	T1588 Obtain Capabilities
T1588.006 Vulnerabilities	T1588.007 Artificial Intelligence	T1190 Exploit Public-Facing Application	T1098 Account Manipulation
T1098.001 Additional Cloud Credentials	T1059 Command and Scripting Interpreter	T1212 Exploitation for Credential Access	T1586 Compromise Accounts
T1580 Cloud Infrastructure Discovery			

Indicators of Compromise (IOCs)

TYPE	VALUE
IPv4	83[.]7[.]139[.]184, 83[.]7[.]157[.]76, 73[.]105[.]135[.]228, 83[.]7[.]135[.]97

Patch Link

https://raw.githubusercontent.com/projectdiscovery/nuclei-templates/master/cves/2021/CVE-2021-3129.yaml

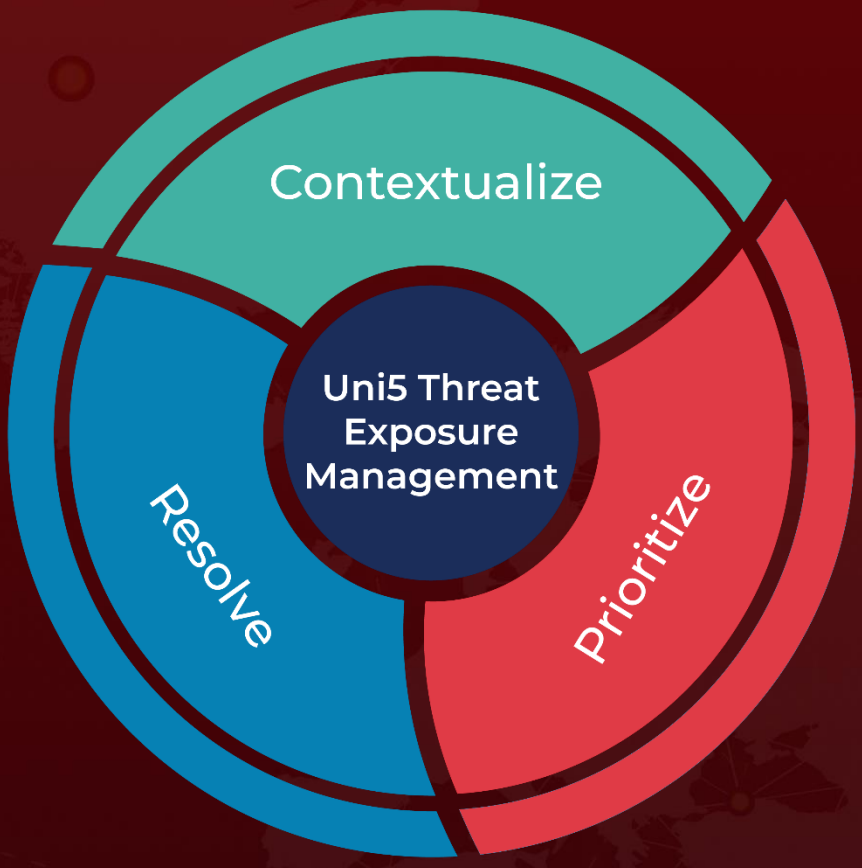
References

https://sysdig.com/blog/llmjacking-stolen-cloud-credentials-used-in-new-ai-attack/

What Next?

At **Hive Pro**, it is our mission to detect the most likely threats to your organization and to help you prevent them from happening.

Book a free demo with **HivePro Uni5**: Threat Exposure Management Platform.



REPORT GENERATED ON

May 13, 2024 • 12:00 AM

More at www.hivepro.com