Claude Code e DeepSeek: spionaggio di sospetta matrice cinese diviso fra due modelli



Claude Code e *DeepSeek-v4-pro* hanno lavorato in tandem dentro la stessa catena d'attacco: operatori di sospetta matrice cinese hanno affidato al primo l'esecuzione e al secondo il ragionamento offensivo per colpire sistemi governativi in tre Paesi. È il secondo caso pubblicamente documentato, dopo GTG-1002 nel novembre 2025, di impiego di *Claude Code* in operazioni offensive attive riconducibili a operatori cinesi sospetti; l'elemento che la redazione considera più significativo è però la ripartizione dei compiti fra i modelli di due *vendor* differenti, uno occidentale e uno cinese.

Il caso è documentato da Hunt.io, che ha avvisato le organizzazioni colpite e i CERT nazionali il 6 luglio e ha pubblicato il report il 14, al termine di una finestra di *responsible disclosure* di sette giorni. La ripresa internazionale è maturata nei giorni seguenti: le prime riprese il 15 luglio, quella di Security Affairs, a firma Pierluigi Paganini, il 16. Non si tratta di *ransomware* né di estorsione: gli indicatori raccolti descrivono un'operazione di raccolta informativa e accesso persistente, con obiettivi coerenti con priorità di *intelligence* statale.

Come è emersa

I ricercatori sono arrivati alla campagna per *pivoting*: seguendo l'impronta di un *header* HTTP sulla porta 1111
, associata all'infrastruttura *command-and-control* TencShell, hanno individuato una rete di server a Hong Kong. TencShell

non è un dettaglio marginale: è un impianto in Go derivato dal *framework open source* Rshell, documentato per la prima volta da Cato CTRL nel maggio 2026 e lì già valutato come attività di sospetta matrice cinese. È la premessa dell'intera catena investigativa. Che le due ricerche insistano sulla stessa infrastruttura è confermato dalla sovrapposizione degli indicatori: un IP della lista IOC di Cato, 45.64.52[.]242
, ricompare nella tabella TencShell di Hunt.io, che lo etichetta proprio come *Cato Networks IoC*.

La ricerca ha portato a 13 server ospitati a Hong Kong e distribuiti su quattro sistemi autonomi distinti (
VMISS Inc.
, MEGA-II IDC
, CTG Server Limited
, Antbox Networks Limited
). Su uno di questi, all'indirizzo 112.213.124[.]132
, la porta 8888
esponeva un server Python SimpleHTTP
con una directory aperta: 2.431 file e 80 sottocartelle contenenti codice sorgente delle vittime, *web shell* mascherate da immagini, script di *exploit* su misura, pagine di *login* clonate, *output* di scansione e *log* operativi con annotazioni in cinese semplificato. L'errore di *OPSEC* degli operatori, non la telemetria di un *vendor*, ha aperto la porta all'analisi.

La misura reale della campagna va oltre i tre Paesi effettivamente compromessi: fra i cinque *key finding*, Hunt.io segnala la scansione di oltre 5.890 *host* governativi in dieci Paesi, ordinati con una scala di *scoring* automatico sviluppata in Python.

La divisione del lavoro fra i due modelli

L'elemento che distingue questa campagna è l'architettura del *tooling*. Dai *log* recuperati risulta che Claude Code 2.1.165

ha gestito l’esecuzione: comandi Bash, sessioni persistenti, parallelizzazione dei compiti e allestimento dell’infrastruttura di *phishing*. DeepSeek-v4-pro ha invece assolto la parte di pianificazione, generando script, scegliendo le tecniche e individuando nuovi modi per aggirare le difese quando i tentativi precedenti fallivano. Nella sintesi di Hunt.io, la logica offensiva viene instradata su un *LLM* domestico cinese mentre l’esecuzione *agentic* poggia sull’infrastruttura di Anthropic. Un file CLAUDE.md recuperato dalla directory conteneva istruzioni per creare, testare e migliorare automaticamente pagine di *phishing* clonate per più bersagli.

Le sessioni di Claude Code coprono un arco di attività fra l’8 e il 12 giugno 2026, con tre server che condividono chiavi SSH mantenuti operativi almeno fino al 18-19 giugno. Vale la pena circoscrivere la portata del dato di novità: la prima campagna di spionaggio orchestrata da AI a livello statuale resta quella attribuita da Anthropic all’attore GTG-1002 nel novembre 2025, esito di una traiettoria da consulente a operatore che l’AI offensiva ha attraversato in pochi mesi. L’inedito qui non è l’automazione in sé, già inquadrata nel dibattito su AI agentic e cybercrime, ma la ripartizione dei compiti e la conferma che lo schema è replicabile combinando i modelli di due *vendor* differenti, uno occidentale e uno cinese.

I bersagli

L’attività di compromissione confermata riguarda tre amministrazioni governative. In Thailandia gli operatori hanno sfruttato una *SQL injection* contro un sistema amministrativo pubblico, ottenendo accesso al pannello di *admin* e collocando una *web shell* camuffata da file GIF per l’esecuzione persistente di comandi; il *database* esfiltrato conteneva nomi, numeri di identità nazionale e qualifiche di dipendenti pubblici.

L’accesso al pannello risultava attivo almeno fino al 9 giugno 2026 e la sola directory riferita a questo sistema conteneva 980 file, segno di una compromissione prolungata. In Afghanistan è stata colpita un’applicazione web per la raccolta di segnalazioni dei cittadini, da cui gli attaccanti hanno estratto codice sorgente, credenziali del *database*, chiavi di cifratura e il codice di gestione della posta di un’installazione Laravel 5.8.38
, poi riusati per costruire un *exploit* Python mirato ai meccanismi di deserializzazione.

A Taiwan la ricognizione ha interessato otto organizzazioni fra *supply chain* e settori contigui alla difesa, sottoposte a sola attività di *reconnaissance* e *fingerprinting*; a queste si aggiungono due organizzazioni effettivamente compromesse: un produttore chimico via *SQL injection* e un fabbricante di apparati *telecom* ed *edge*, violato dopo il ritrovamento, in file JavaScript pubblicamente accessibili, di chiavi anon
di Supabase
e *token* SAS di Azure Logic App
lasciati *hardcoded*, che hanno dato accesso diretto all’infrastruttura *cloud* di *backend*.

Gli Stati Uniti figurano solo in fasi preliminari, ed è la ragione per cui il report Hunt.io parla di quattro Paesi mentre le compromissioni confermate sono tre: *host* NASA (
launchpad.nasa[.]gov
e ngis.nasa[.]gov
) compaiono nell’*output* di scansione ma non risultano perseguiti, mentre erano in preparazione pagine clone del D.C. Council e della contea di Delaware, in Pennsylvania. In parallelo alle attività governative, gli operatori hanno condotto una campagna contro società di servizi finanziari in Europa, Australia e Asia: una pagina di *exploit* CORS ospitata sull’infrastruttura ha estratto i dati di quattro *account* amministratore WordPress da una grande piattaforma di *payment processing*, con nomi coincidenti con quelli di dipendenti

della società.

Infrastruttura e *malware*

Oltre a TencShell, Hunt.io valuta con confidenza moderata la presenza di un secondo *framework* C2, finora non documentato, che si autodefinisce Gshell C2
e che opererebbe in parallelo dai medesimi cluster. Il *malware* recuperato dalle porte di *delivery* è un binario Linux/ARM a 32 bit, mai osservato prima, denominato HSEWH-Ur
: compilato in Go, comunica via *WebSocket* verso la porta 4081 dello stesso *hub* e si autentica con *header* HTTP X-NITRO-USER e X-NITRO-PASS
, normalmente usati da API legittime di gestione infrastrutturale.

È in grado di sottrarre credenziali di messaggistica Tencent QQ (inclusi identificativi SDK e chiavi crittografiche), *token* di piattaforme di *enterprise messaging* e chiavi di accesso a servizi *cloud*. Una variante Linux/x86, recuperata da un diverso indirizzo (
38.55.105[.]143:8088
), impiega il *tool* di offuscamento Go garble
per rimuovere i nomi di funzione; entrambe condividono una chiave di cifratura identica da 80 byte, segno di una *codebase* comune fra architetture.

Hunt.io si ferma prima di qualificare questi binari come build Linux di TencShell, non avendo confermato una corrispondenza a livello di codice con il campione Windows documentato da Cato. Il filo che lega i due report resta il progetto Reacon: Cato ne aveva trovato la struttura nei *path* Go di TencShell, e la variante x86 analizzata da Hunt.io impiega garble
proprio per cancellarne i riferimenti.

Perché conta per chi difende

L’attribuzione resta prudente: Hunt.io non nomina un gruppo e si ferma a una valutazione di attività riconducibile alla Cina, sostenuta dalla presenza costante di cinese semplificato nel codice e nelle note, dal *clustering* a Hong Kong e dal profilo dei bersagli, orientato a *procurement*, relazioni con i fornitori e visibilità sulle politiche pubbliche.

Per un *SOC*, il dettaglio più istruttivo è che l’intera catena d’attacco poggiava su strumenti pubblici e leciti assemblati su un’unica macchina: ARL
sulla porta 5003
per la ricognizione, DeepAudit
sulla 3000
per l’*audit* del codice, Vshell
sulla 8084
per il *command-and-control*, tutti disponibili apertamente e con usi di *testing* legittimi. È la loro combinazione, accanto alla directory aperta di dati delle vittime, a distinguere un’intrusione attiva da un normale *red team*.

Sul piano dei modelli, va evitata un’inferenza affrettata: il report non documenta alcun *bypass* dei *guardrail*; gli agenti commerciali, del resto, applicano controlli propri sull’esecuzione, indipendenti da dove nasce il ragionamento.

Lo schema osservato pone piuttosto la questione di come presidiare un uso in cui la componente decisionale viene delegata a un modello terzo, fuori dal perimetro di monitoraggio del *vendor* che fornisce l’esecuzione. Resta anche una domanda che il report non affronta e che nessuna delle riprese chiarisce: come operatori attivi da Hong Kong, con note in cinese semplificato, abbiano ottenuto accesso a Claude Code
.

Anthropic, nel proprio rapporto di febbraio 2026 sulle

distillation attack, dichiarava di non offrire accesso commerciale a Claude in Cina, e descriveva il ricorso di laboratori cinesi a servizi *proxy* commerciali che rivendono l’accesso ai modelli di frontiera tramite reti di account fraudolenti; il report Hunt.io non collega la campagna a questo canale, e va detto che non lo fa.

A oggi, inoltre, non risultano dichiarate azioni di Anthropic o di DeepSeek sugli *account* impiegati. I vettori sfruttati, del resto, restano ordinari e prevenibili (*SQL injection*, segreti *hardcoded* in JavaScript pubblico, deserializzazione insicura, *token cloud* non ruotati), e su questi l’igiene applicativa di base continua a fare la differenza.

Il *set* completo di indicatori, con *hash* dei file e infrastruttura di rete, è pubblicato nel report originale.

Condividi sui Social Network:
https://www.ictsecuritymagazine.com/notizie/cina-claude-code-deepseek-spionaggio-ai/