ISACA Rome Chapter

NETWITNESS

Netwitness Workshop

Netwitness Education

Stefano Maccaglia
Director Incident Response and Red Team practices
Netwitness

2025

ISACA Rome Chapter

Pensieri profondi su IA e cybercrime

(senza attendere 7,5 milioni di anni)

Chi sono?

Attualmente sono il Direttore dei Servizi di Incident Response e del Red Teaming in Netwitness.

La mia carriera è iniziata nel 1997 in Digital, ma ho iniziato a crackare software da bambino con un Commodore C64…

Ho lasciato la "scena underground" nei primi anni 2000 e sono passato alla cybersecurity (nella parte offensiva) fino al 2009, anno in cui ho deciso di confrontarmi con l'Incident Response.
Da quel momento ho investigato moltissimi casi in differenti settori, dal mondo finanziario a quello militare, dalle Agenzie governative alle Telco.

Sono Direttore in Netwitness dal 2021 e ho la fortuna di guidare un gruppo di talenti internazionali incredibili.

https://it.linkedin.com/in/stefanomaccaglia

Agenda

- Oggi discuteremo di come l'Intelligenza Artificiale (IA) è attualmente utilizzata nel mondo cybercriminale, in contrasto a come è descritta nelle presentazioni di marketing e nei film di fantascienza...
- Spesso, parlando anche con «addetti ai lavori», mi trovo davanti a imbarazzanti descrizioni di fantomatiche capacità legate all'IA...
- Per alcune di queste persone l'IA è «senziente», una sorta di «onniscente super criminale» che può decidere di svolgere hacking su praticamente ogni sistema e piattaforma...
- La verità, per fortuna, è ancora lontana da queste apocalittiche previsioni... Più concretamente oggi, i cybercriminali usano l'IA come validissimo motore inferenziale per ricerche, profilazioni e automazione di attività ripetitive di un attacco informatico...

MARVIN, NON SKYNET

What Hollywood promises

- autonomous AI that finds 0-days,
- writes polymorphic malware,
- coordinates botnets.

What criminals actually use

- tools to draft phishing,
- clone voices,
- generate deepfake videos,
- automate reconnaissance and negotiation workflows

In base all'osservazione possiamo definire l'IA adottata dai cybercriminali come **"uno stagista senza etica e senza spirito di iniziativa"**:
Bravo nel comporre e sintetizzare concetti, ma incapace di operare senza il supporto di un esperto in carne e ossa.

MARVIN, NON SKYNET

COSA CI PROPONE HOLLYWOOD?

ISACA Rome Chapter

CHECK POINT

ARTIFICIAL INTELLIGENCE

EXECUTIVE INSIGHTS SECURITY

OCTOBER 24, 2025

AI 2030: The Coming Era of Autonomous Cyber Crime

By Raymond Schippers | Office Of The CTO

Organizations around the world are rapidly adopting AI, including across the enterprise, where it is already prov efficiency gains. As a result, cyber security is entering a turning point where AI fights AI. The phishing scams an are only precursors to a coming era of autonomous, self-optimizing AI threat actors. Systems that can plan, exe attacks with limited human oversight or even none at all.

FUTURECISO

Where Security is Number One

PEOPLE PROCESS TECHNOLOGY RESOURCES EVENTS

Home > Artificial Intelligence

2026: When autonomous AI transforms cyber attacks and security models

by Allan Tan – January 13, 2026

BCG

CYBERSECURITY AND DIGITAL RISK

AI Is Raising the Stakes in Cybersecurity

By Vanessa Lyon, Shoaib Yousuf, and Mahmood Serry

SLIDESHOW DECEMBER 18, 2025 5 MIN READ

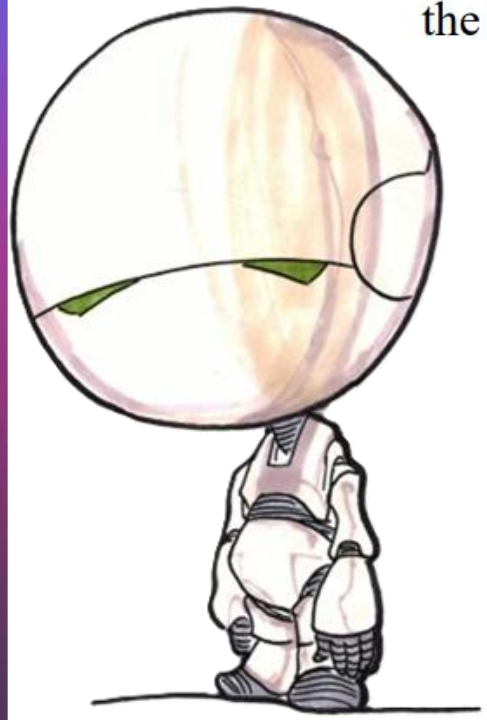
actors to automate large parts of the "cyber kill chain." Such AI-enabled attacks
ed operational shutdowns, financial losses, and regulatory penalties. The
th: Offense is scaling faster than organizations are modernizing their defenses.

"Un'IA che si sveglia, diventa malvagia e decide di hackerare il pianeta."
"0-day e malware perfetti con un solo clic."
"L'AI è in grado di creare e gestire autonomamente una botnet."

MARVIN, NON SKYNET

COSA CI PROPONE HOLLYWOOD

- Se non ne avete abbastanza…



IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

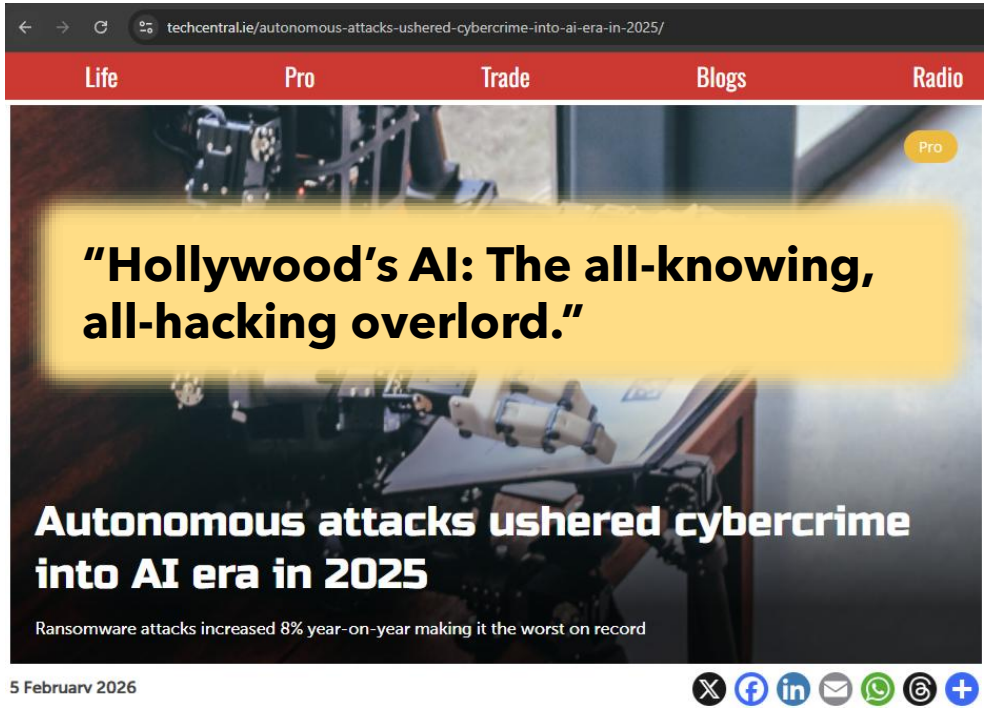
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 12, December 2024

DOI: 10.17148/IJARCCE.2024.131263

AI-Generated Cyber Threats the Rise of Autonomous Hacking Systems

Jayasudha Yedalla

Colorado Technical University, Colorado, USA



ISACA Rome Chapter

MARVIN, NON SKYNET

ALCUNI FATTI!...

Campagne di Phishing basate su IA Generativa.

- L'IA sta evolvendo rapidamente gli attacchi di phishing permettendo di migliorarne la qualità e la quantità.
- Questo sta portando alla creazione di un vero e proprio **ecosistema** in cui BEC*, furto di credenziali, vishing e frodi proliferano incontrollate.

*BEC = Business Email Compromise

Initial Reconnaissance e Social Media Profiling

- Gli attacchi basati su Email sono cresciuti del 16% (anno su anno)*, grazie all'integrazione dell'IA nei contesti della reconnaissance, dell'ingegneria sociale e delle estorsioni
- Le applicazioni di collaborazione (Teams, Slack, etc.) sono sempre più abusate e sfruttate come canali secondari di attacco e controllo.

*Acronis' H2 2025 report

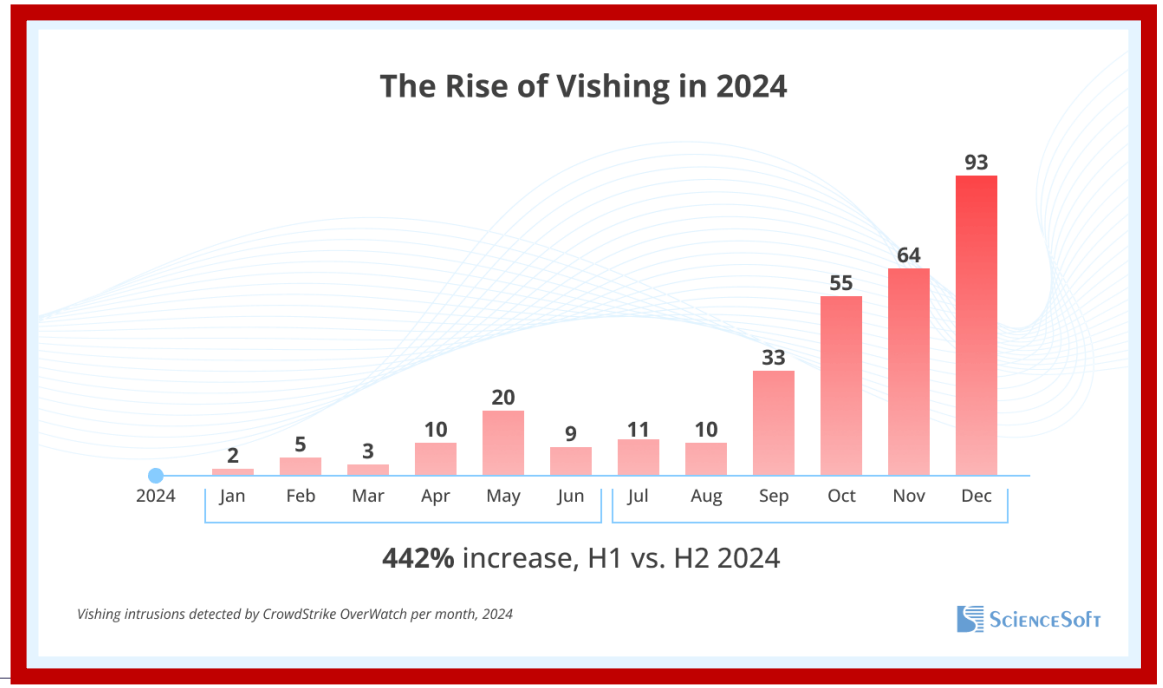
L'IA riduce le barriere lunguistiche

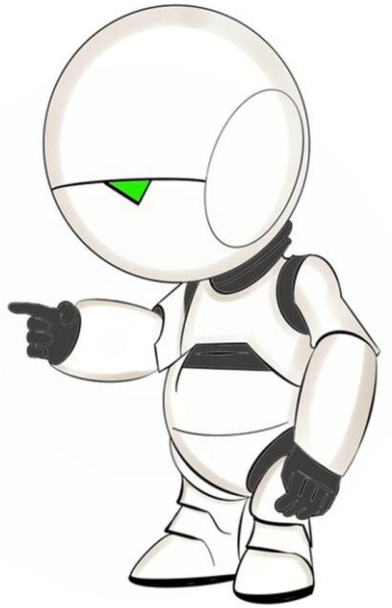
- I Cybercriminali usano l'IA per creare messaggi di phishing che evadono i filtri linguistici e che li rendono più credibili sebbene questi attacchi continuino a basarsi su links, allegati e prompt di credenziali.

MARVIN, NON SKYNET

ALCUNI FATTI!...

- Gli attacchi di tipo “Vishing” hanno avuto un balzo del 442% nel 2024 grazie ai deepfake basati su IA.
- Le frodi sulle chiamate (Deepfake) alimentano un ecosistema che muove decine di milioni di dollari.
- Più del 10% delle banche ha riportato nel 2024 frodi sulle chiamate (deepfake) nell’ordine dei milioni di dollari e questo è un fenomeno globale ormai.





RECONNAISSANCE E PROFILING

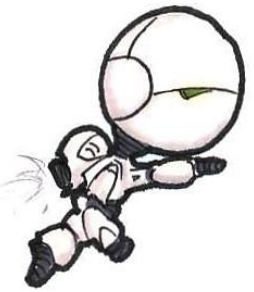
COME UTILIZZANO L'IA I CYBERCRIMINALI

- Il report di Anthropic dell'Agosto 2025 conferma l'utilizzo dell'IA in queste operazioni e più in particolare in questi ambiti:

Profilazione delle Vittime

Raccolta e selezione dei dati trafugati nelle frodi e nei data breach

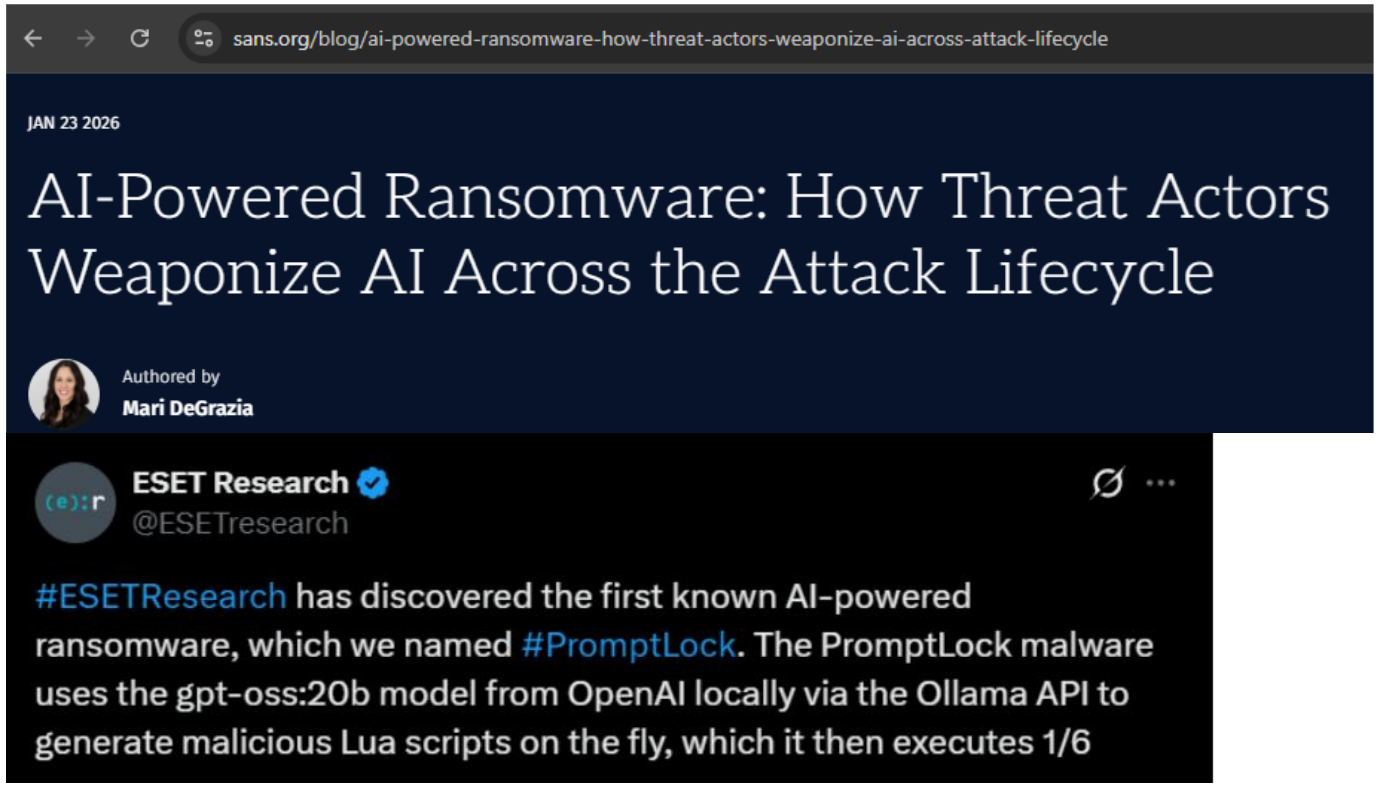
Supporto all'identificazione di asset di valore durante la reconnaissance.



RANSOMWARE OPERATIONS

COME UTILIZZANO L'IA I CYBERCRIMINALI

- Una pubblicazione SANS sul ransomware evidenzia che gli attaccanti utilizzano l'IA per generare frammenti di codice malevolo, analizzare i dati esfiltrati e gestire le negoziazioni del riscatto tramite chatbot basati su agenti-IA, operativi 24/7 in più lingue
- L'IA riduce drasticamente il tempo necessario per esaminare grandi volumi di dati sottratti, classificando automaticamente le informazioni sensibili e arrivando persino a redigere messaggi di estorsione personalizzati che elencano con precisione che cosa è stato rubato.
- Da notare che ESET ha individuato nel Settembre 2025 un malware: «PromptLock» che basa il suo comportamento su IA, ma è ancora un proof-of-concept.



ABBASSAMENTO DEL LIVELLO

COME UTILIZZANO L'IA I CYBERCRIMINALI

- OpenAI e Anthropic riportano un uso esteso dei modelli da parte di truffatori, che li impiegano come "supporto alla frode": dalla stesura di email e inserzioni truffaldine alla generazione di malware e script semplici, anche quando i sistemi di sicurezza bloccano le richieste più sofisticate.

- Il report "AI Threat 2025" di KELA evidenzia che gli attori dell'underground criminale usano strumenti di IA per automatizzare il phishing, accelerare lo sviluppo di malware e rendere più efficienti le operazioni di frode, consentendo di fatto anche a cybercriminali con competenze tecniche limitate di partecipare all'ecosistema.

AI-POWERED SOCIAL ENGINEERING 101

ISACA Rome Chapter

L'IA è diventata il complice perfetto per l'ingegneria sociale: scrive, parla e persino si presenta con l'aspetto di persone di cui le vittime si fidano già, e casi reali dimostrano che questo si è tradotto in gravi furti di dati e frodi da milioni di dollari.

Profilazione del bersaglio

- L'attaccante utilizza innanzitutto l'IA per aggregare informazioni sul proprio obiettivo

Costruzione del messaggio

- L'IA può produrre testi estremamente curati che includono il nome della vittima, dettagli rilevanti sui progetti e persino imitare il tono della leadership.

Rifinitura e localizzazione dei contenuti

- L'attaccante può tradurre con facilità il contenuto in qualsiasi lingua o dialetto, abbattendo le barriere linguistiche.

Generazione di allegati e URL

- Se l'attacco prevede l'invio di allegati o link, anche qui entra in gioco l'IA. Gli attaccanti possono usare strumenti che creano automaticamente false pagine di login o documenti contraffatti.

Scalabilità e varianti

- Infine, l'IA consente di inviare milioni di email di questo tipo, ognuna con piccole variazioni.

DEEPFAKE FRAUD IL VETTORE CYBERCRIMINALE PIU' COMUNE

Un'analisi dell'AI Incident Database condotta da Cybernews ha rilevato che, nel 2025, i deepfake (audio e video) rappresentavano circa l'81% degli incidenti di frode legati all'AI, rendendo l'impersonificazione sintetica l'arma principale del crimine informatico basato sull'intelligenza artificiale.

Gli esempi citati includono truffe «romantiche» con personaggi video generati dall'AI, truffe di tipo "emergenza familiare" con voci clonate e schemi di investimento che sfruttano falsi video "in diretta" di amministratori delegati, tutti casi che hanno portato a perdite finanziarie significative.

$25 MILLION DEEPFAKE CFO SCAM

Nel 2024, un dipendente dell'area finance di una multinazionale a Hong Kong è stato indotto a trasferire circa 25 milioni di dollari dopo aver partecipato a una videoconferenza in cui ogni altro "partecipante" – incluso il CFO – era in realtà un deepfake generato a partire da filmati pubblici.
La vittima inizialmente aveva sospettato un tentativo di phishing dopo una email anomala, ma il meeting sintetico, estremamente realistico, l'ha convinta a seguire le istruzioni di trasferimento, causando quindici pagamenti non autorizzati.

CNN World Africa Americas Asia Australia China Europe India Middle East United Kingdom

World / Asia

Finance worker pays out $25 million after video call with deepfake 'chief financial officer'

By Heather Chen and Kathleen Magramo, CNN

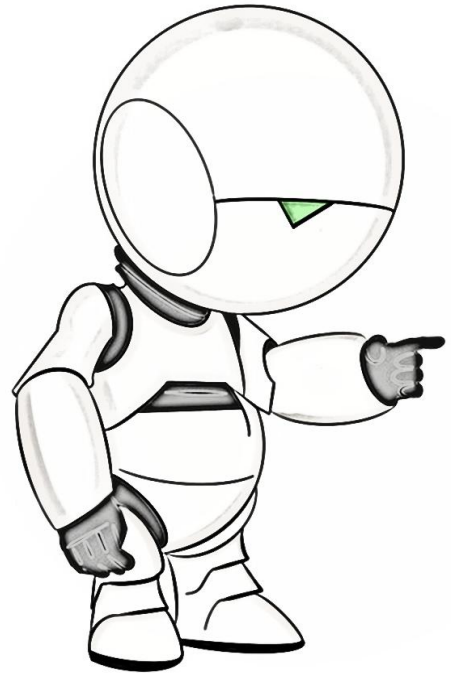
2 min read · Published 2:31 AM EST, Sun February 4, 2024

Link:
https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html

AI-GENERATED PHISHING

TOP ENTERPRISE EMAIL THREAT

Per esperienza diretta in ambito incident response, il phishing generato dall'AI è ormai la "principale minaccia via email", davanti a ransomware e insider risk, perché l'AI consente una personalizzazione di massa e la produzione rapidissima di varianti uniche di attacco.



From: support@rnicrosoft.co.uk
Sent: 16/01/2023 11:44
To: **Bob Smith** <Bob.Smith@company.com>
Subject: Urgent Action Needed!

Outlook

Microsoft Account

Verify your account

We detected some unusual activity about a recent sign in for your Microsoft account. you might be signing in from a new location app or device.

To help keep your account safe. We've blocked access to your inbox , contacts list and calander for that sign in. Please review your recent activity and we'll help you secure your account. To regain access you'll need to confirm that the recent activity was yours.

http://account.liive.com/ResetPassword.aspx

Thanks,
The Microsoft Team

From: support@microsoft.co.uk
Sent: 16/01/2023 11:44
To: **Bob Smith** <Bob.Smith@company.com>
Subject: Unusual Sign In Activity

Outlook

Microsoft Account

Verify your account

We detected some unusual activity about a recent sign in for your Microsoft account bo******@company.com. you might be signing in from a new location app or device.

To help keep your account safe. We've blocked access to your inbox, contacts list and calendar for that sign in. Please review your recent activity and we'll help you secure your account. To regain access you'll need to confirm that the recent activity was yours.

Review recent activity

Thanks,
The Microsoft Team

AUTOMATED RECONNAISSANCE

Gli attaccanti sfruttano l'IA per trasformare un'OSINT lenta e manuale in "dossier sulle vittime" costruiti in pochi minuti, pronti per alimentare direttamente campagne di phishing, BEC e flussi di intrusione.

Social networks

- LinkedIn for org charts, roles, recent job moves and connections;
- X/Facebook for personal interests, family and travel.

Corporate surface

- websites,
- press releases,
- Blogs,
- PDFs.

Technical breadcrumbs

- GitHub,
- package registries and job posts reveal technologies in use,
- Cloud and Identity providers,
- VPNs,
- EDR names,

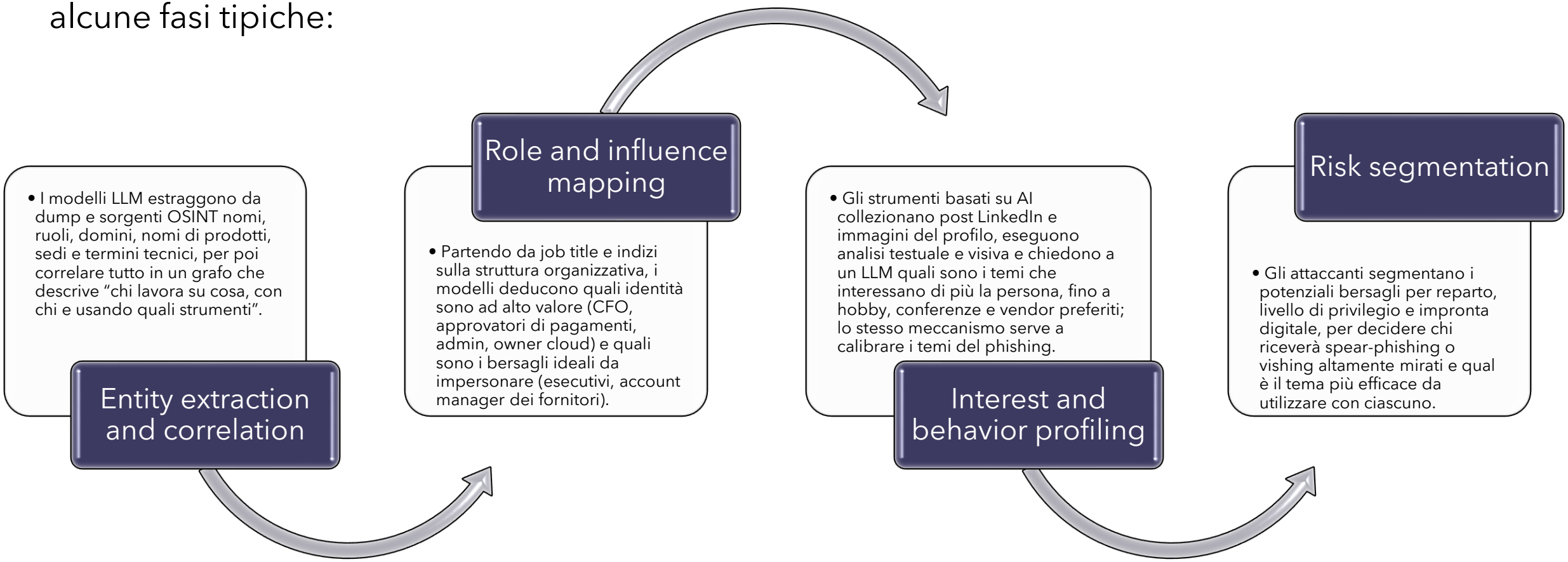
Dark-web data

- prior credential leaks,
- internal emails,
- document dumps supply usernames,
- email threading patterns and internal jargon.

Per prima cosa gli attaccanti collezionano dati pubblici e semi-pubblici, poi li passano ai modelli per profilare il bersaglio e definire il modo più efficace per aprire la "testa di ponte" dell'attacco.

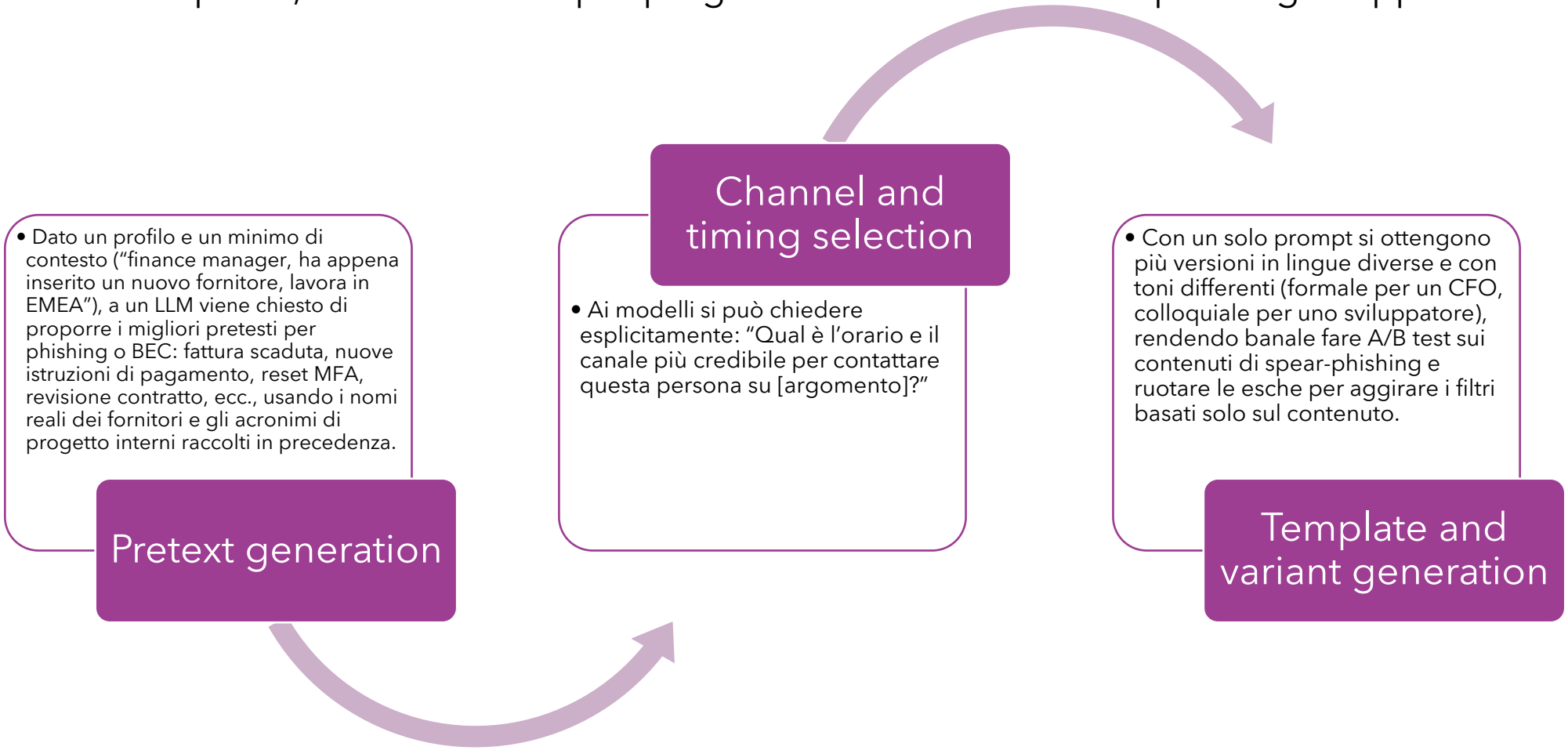
COME L'IA PROCESSA E ARRICCHISCE L'OSINT

Invece di avere una persona che passa ore su ogni fonte, gli attaccanti lasciano che siano i modelli a digerire e strutturare le informazioni. In questo processo possiamo distinguere alcune fasi tipiche:



TRASFORMARE PROFILI IN PIANI DI ATTACCO

Una volta costruiti i profili, l'IA viene usata per progettare e calibrare nel tempo i singoli approcci.



Un caso di esempio

UN MESSAGGIO IMPORTANTE

La mattina di lunedì 17 febbraio, dopo una pausa caffè, Amir Fakhry, dirigente di una filiale di una primaria banca del Medio Oriente, ricevette un'email molto delicata da Ahmed Abbas, Responsabile Affari Finanziari di una importante compagnia petrolifera.

Subject: Urgent Wire Transfer Required for Project Continuation
From: ahmed.abbas@petroleumcompany.com
To: amir.fakhry@primarybank.com
Date: February 17, 2025

Dear Amir Fakhry,
I hope this message finds you well. I am writing to inform you of an urgent financial matter that requires immediate attention to ensure the continuation of our critical operations.
As you are aware, our ongoing collaboration with International Drilling Co. in Frankfurt is pivotal to our expansion plans. Due to unforeseen circumstances, we urgently need to expedite the payment process to avoid costly project delays.
Details for the Wire Transfer:

- **Beneficiary Name:** International Drilling Co. GmbH
- **Bank Name:** Frankfurt Bank AG
- **Account Number:** DE89370400440532013000
- **SWIFT Code:** FBAGDEFF
- **Amount:** €2,500,000
- **Purpose of Transfer:** Project Code DCO250411 Nabir Offshore drill

We count on your swift action to process this transfer by the end of the business day today. Please confirm once the transfer has been successfully initiated.
Thank you for your immediate attention to this matter and for your continued support.
Best regards,
Ahmed Abbas
General Manager of Financial Affairs
Petroleum Company
[Contact Information]

TROPPO OCCUPATO PER RISPONDERE...

Mentre leggeva l'email, Amir ricevette una telefonata: l'interlocutore parlava di un servizio telefonico che poteva essere attivato sulla sua linea.
La chiamata durò circa due minuti.
Chiusa bruscamente la telefonata, infastidito dal servizio non richiesto, Amir notò che un altro numero aveva chiamato lasciando un messaggio in segreteria telefonica..

Il messaggio proveniva dalla Compagnia Petrolifera e sembrava che la chiamante fosse la segretaria del Manager.

Amir provò a richiamare, ma la telefonata venne deviata con il messaggio che l'utente era occupato.

Ascoltando il messaggio vocale, notò che la donna aveva lasciato un messaggio molto pressante, spiegando che la situazione era critica e che era necessario procedere immediatamente con il bonifico.

UNA CHIAMATA DALL'ALTO...

Mentre stava ancora decidendo come muoversi, Amir ricevette un'altra chiamata, questa volta dal Direttore della Filiale che dichiarò chiaramente di aver ricevuto una telefonata da Ahmed Abbas, che conosceva personalmente.
La Compagnia Petrolifera era troppo importante per ritardare una transazione di tale rilievo, quindi diede il via libera ad Amir per procedere con il bonifico.

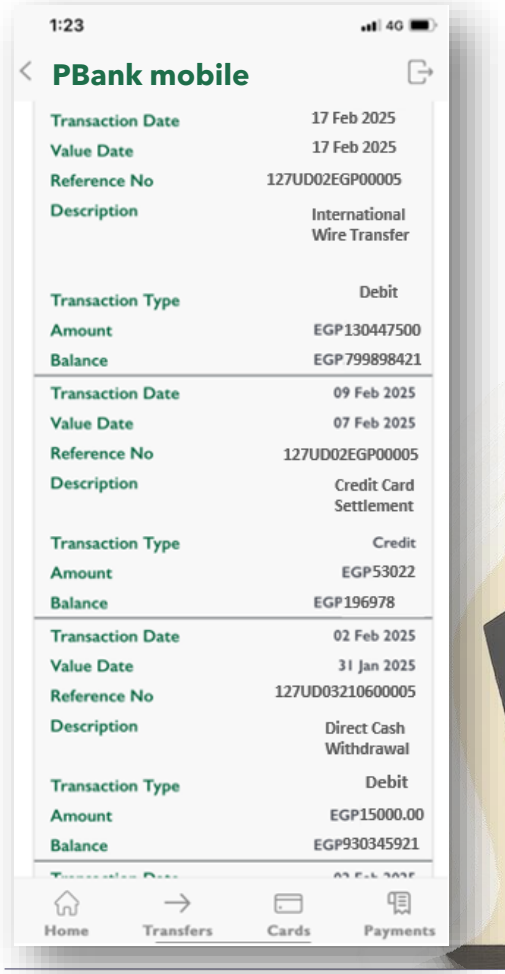
Mentre stava preparando l'operazione, un'ennesima chiamata lo raggiunse.

L'interlocutrice era di nuovo la segretaria del Manager, che ribadiva la necessità di procedere con urgenza all'operazione a causa del forte impatto su un progetto molto delicato.

La donna parlava con tono fermo ma cortese e sembrava a tutti gli effetti la vera segretaria di Ahmed Abbas; lasciò inoltre un indirizzo email al quale chiese che venisse inviata la ricevuta del bonifico.

NUMERO SBAGLIATO?...

Alle 15:21 del giorno successivo, un dirigente dell'Ufficio Finanziario della Compagnia Petrolifera chiese chiarimenti riguardo a una transazione eseguita il giorno precedente.



Sembrava che il bonifico non riportasse un numero di progetto corretto, cosa che risultava strana al dirigente.
Informato della richiesta, Amir contattò direttamente l'ufficio del Manager e riferì quanto accaduto.

Ahmed Abbas reagì subito in modo aggressivo e disse ad Amir che era stato ingannato da un truffatore.

Imbarazzato per la situazione, Amir informò immediatamente il Direttore di Filiale e tentò di bloccare il bonifico, ma non fu possibile recuperare il denaro.

THE “SCHEME”…

Alle 16:43 dello stesso giorno, il Direttore e Amir chiamarono la Sede Centrale della Banca per segnalare la frode e chiedere supporto nella gestione dell’incidente.
Sulla base delle prime analisi, emerse che il denaro era stato trasferito durante la notte su un conto di Francoforte, dal quale era stato immediatamente girato su tre diversi conti operativi alle Isole Vergini Britanniche e a Vanuatu.

I beneficiari dei bonifici erano piccole società locali attive in quei territori, senza alcuna presenza nota su Internet.

Quando fu convocato dalla Security locale, il responsabile diretto di Amir dichiarò chiaramente che la telefonata ricevuta il giorno prima, apparentemente dalla Compagnia Petrolifera per sollecitare il bonifico, apparteneva senza dubbio ad Ahmed Abbas, che conosceva personalmente.
Le indagini successive chiarirono che i truffatori avevano raccolto campioni vocali di Abbas da interventi e conferenze pubbliche.

PERCHE' MI STA CHIAMANDO DI NUOVO?...

Due giorni dopo, in mattinata, Amir ricevette una chiamata diretta da un numero sconosciuto.
Rispondendo, riconobbe immediatamente la falsa segretaria.
La donna gli disse che il suo computer era sotto il controllo della sua banda e che, se voleva evitare ulteriori problemi, avrebbe dovuto collaborare fornendo l'elenco dei conti correnti intestati alla Compagnia Petrolifera e le transazioni eseguite dal Cliente negli ultimi 12 mesi.

La donna fu molto chiara: se non avesse rispettato la richiesta, la situazione si sarebbe evoluta in modo molto negativo per lui.

Spaventato dalla minaccia, ma intenzionato ad aiutare a risolvere il caso, Amir riferì subito la chiamata al proprio responsabile e insieme decisero di scalare immediatamente il caso alla struttura di Cybersecurity della banca.

RISULTATO

Le analisi condotte sul desktop di Amir, sul suo telefono personale e sul laptop aziendale non mostrarono alcun segno di compromissione.
I log delle chiamate del sistema telefonico aziendale permisero al team di Incident Response di individuare il numero nascosto che aveva chiamato il sistema, riconducendolo a un call center.

Quando le Forze dell'Ordine fecero irruzione nel call center e chiesero chiarimenti, il responsabile del call center spiegò che l'azienda era stata vittima di un attacco informatico il mese precedente, ma che l'incidente non aveva apparentemente impattato le operazioni.

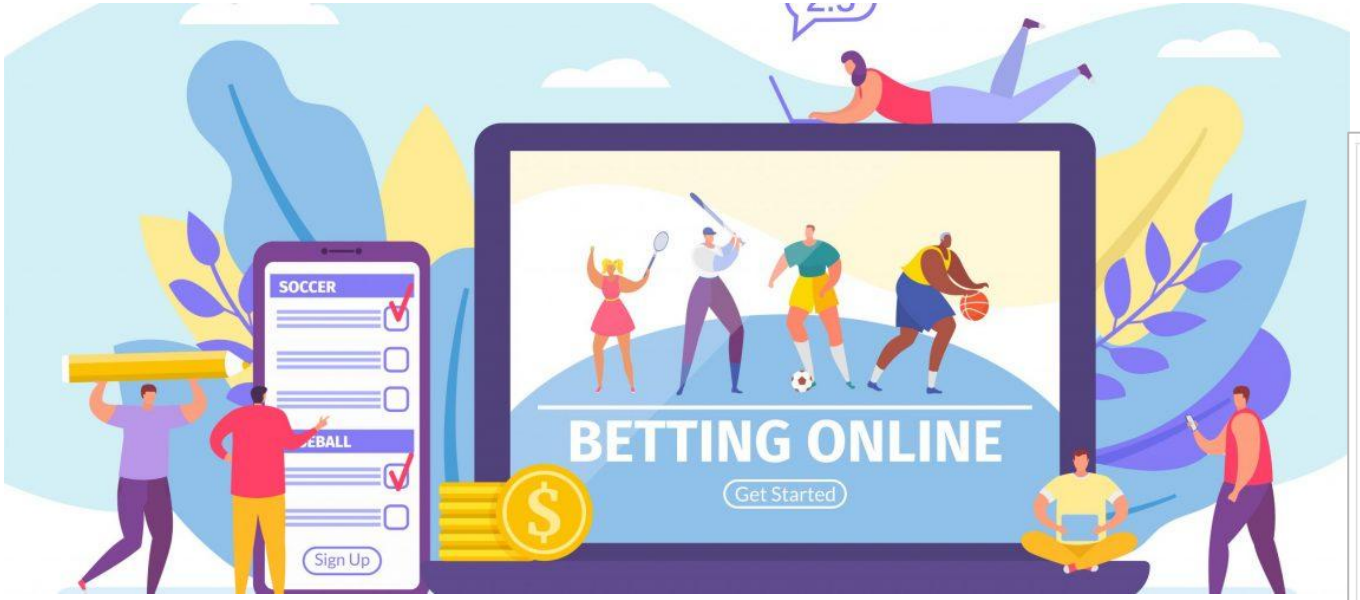
Invitato a collaborare con gli investigatori e dopo aver ascoltato la registrazione della donna che parlava con Amir, il responsabile si rese conto che la voce era un falso sintetico basato sulle registrazioni del risponditore automatico della propria azienda.

La nostra indagine ha infine accertato che lo Unified Communications Manager (CCM) installato nel call center era stato raggiunto da remoto via SSH utilizzando le credenziali di default.

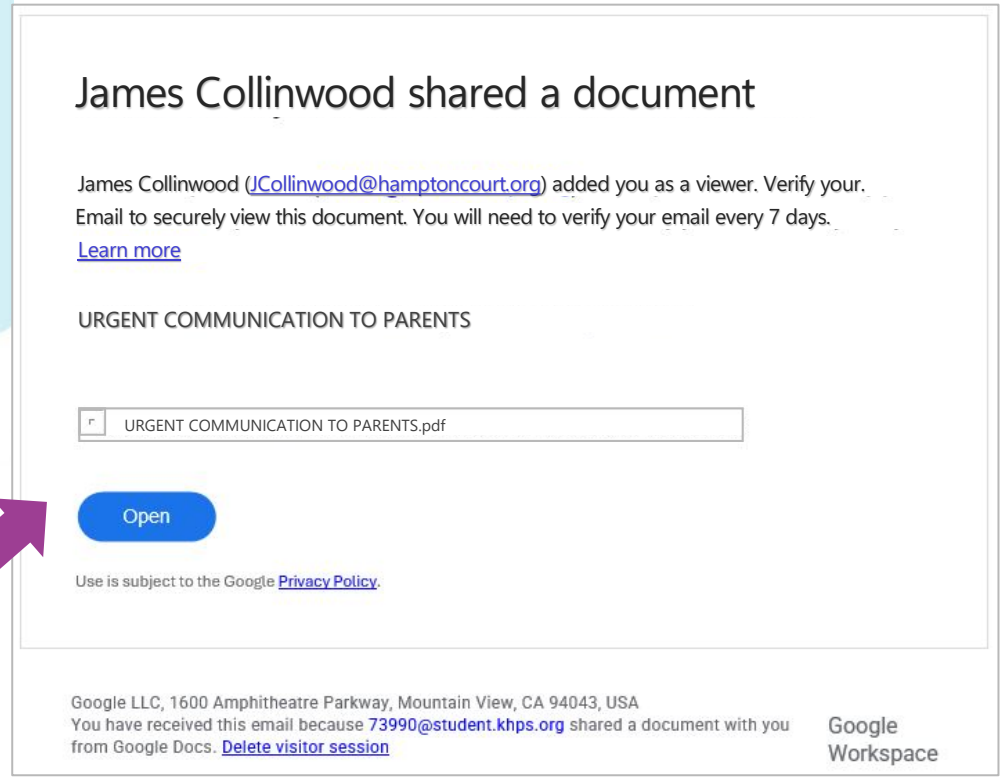
Un secondo esempio

UN BERSAGLIO PERFETTO

- In questo secondo caso, una grande società di scommesse online è stata presa di mira da un attaccante RaaS sofisticato e ben strutturato.



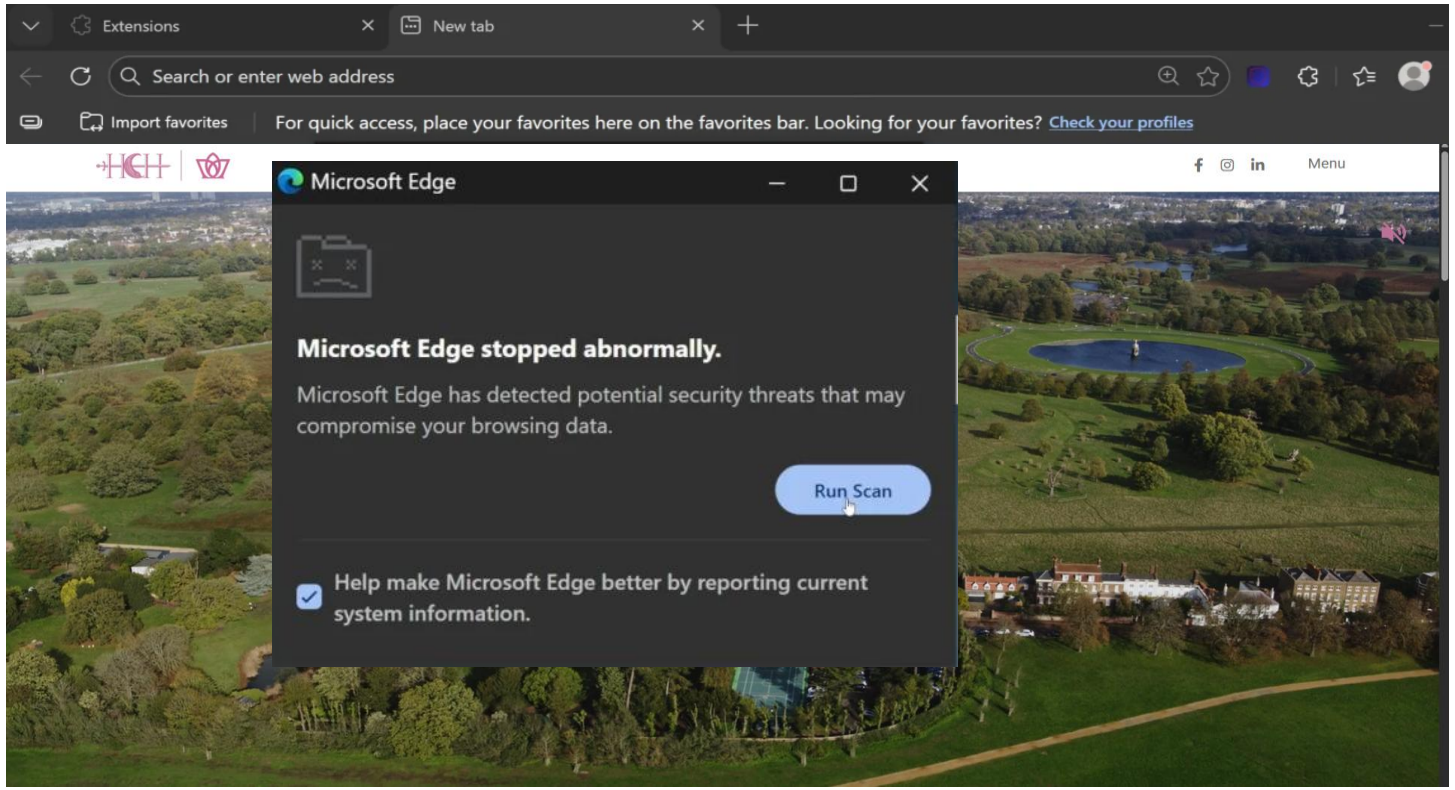
La gang ha inizialmente scelto come bersaglio il Business Continuity Manager (BCM) dell'azienda di scommesse.

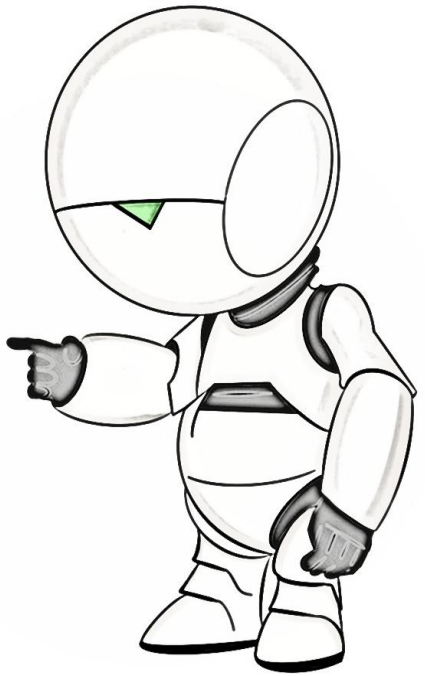


- Hanno raccolto dati da LinkedIn, interventi a conferenze e post pubblici sui social per comprendere situazione familiare, interessi sportivi e progetti in corso. Molto probabilmente hanno poi riversato questi dati in un LLM per generare un pretesto credibile ("questione personale urgente").

ACCESSO INIZIALE

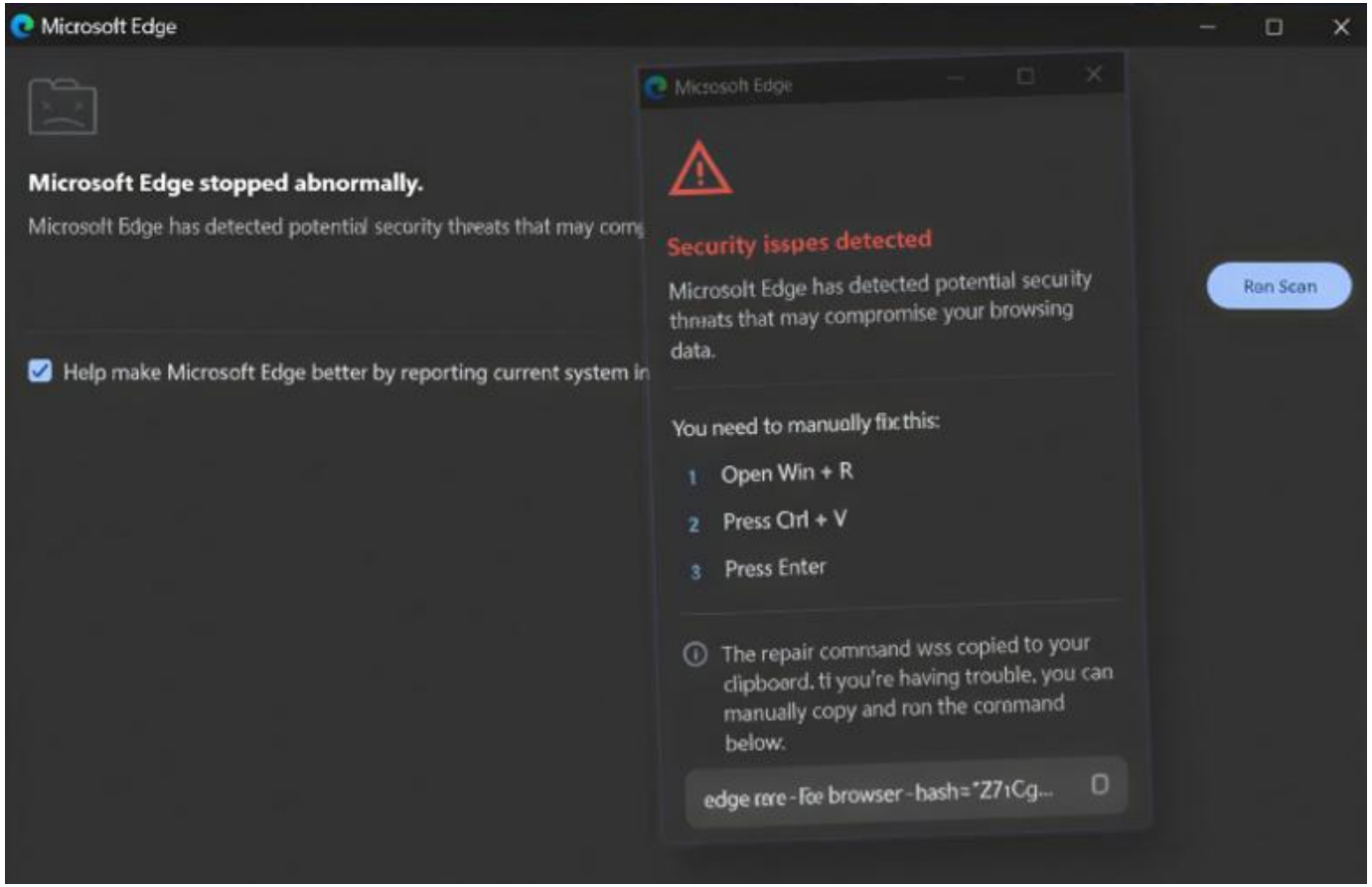
- Il link invia la vittima a un finto portale scolastico graficamente perfetto, ma il vero obiettivo è spingere l'installazione di un componente "di sicurezza" del browser che in realtà è un'estensione Edge malevola in stile CrashFix.





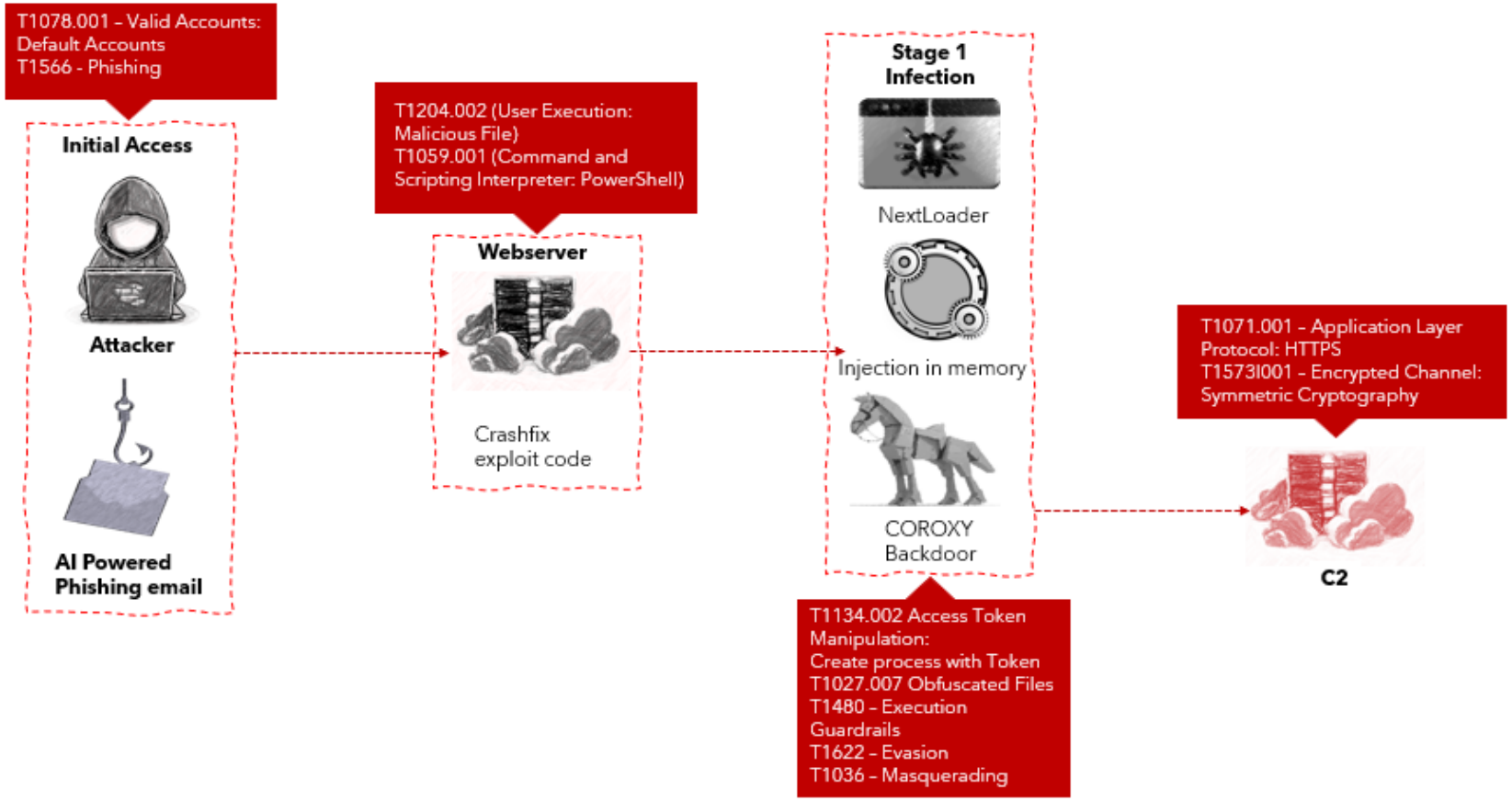
AI-CRAFTED PHISHING + CRASHFIX

- L'exploit funziona così:
  - Il payload inizialmente attende, per evitare correlazioni temporali, quindi manda intenzionalmente in crash il browser mostrando un pop-up "CrashFix" fasullo che dichiara di poter risolvere il problema.
  - A quel punto pre-carica un comando PowerShell negli appunti e istruisce l'utente, tramite un testo di social engineering rifinito con l'AI, a premere Win+R e incollare il comando per eseguirlo "così da risolvere il crash".
  - Quel comando PowerShell scarica ed esegue un secondo script che preleva e distribuisce NetLOADER.



COMPROMISSIONE DELL'ENDPOINT : NETLOADER

- All'esecuzione, il malware contatta un server di Command-and-Control (C2) per recuperare ulteriori binari malevoli.
- Nel nostro caso, il backdoor COROXY viene scaricato ed eseguito.
- Una volta attivato, COROXY inizia a fare beaconing verso il proprio C2 tramite HTTP/S, utilizzando traffico cifrato con RC4 e una logica di backoff per ridurre connessioni rumorose e regolari.

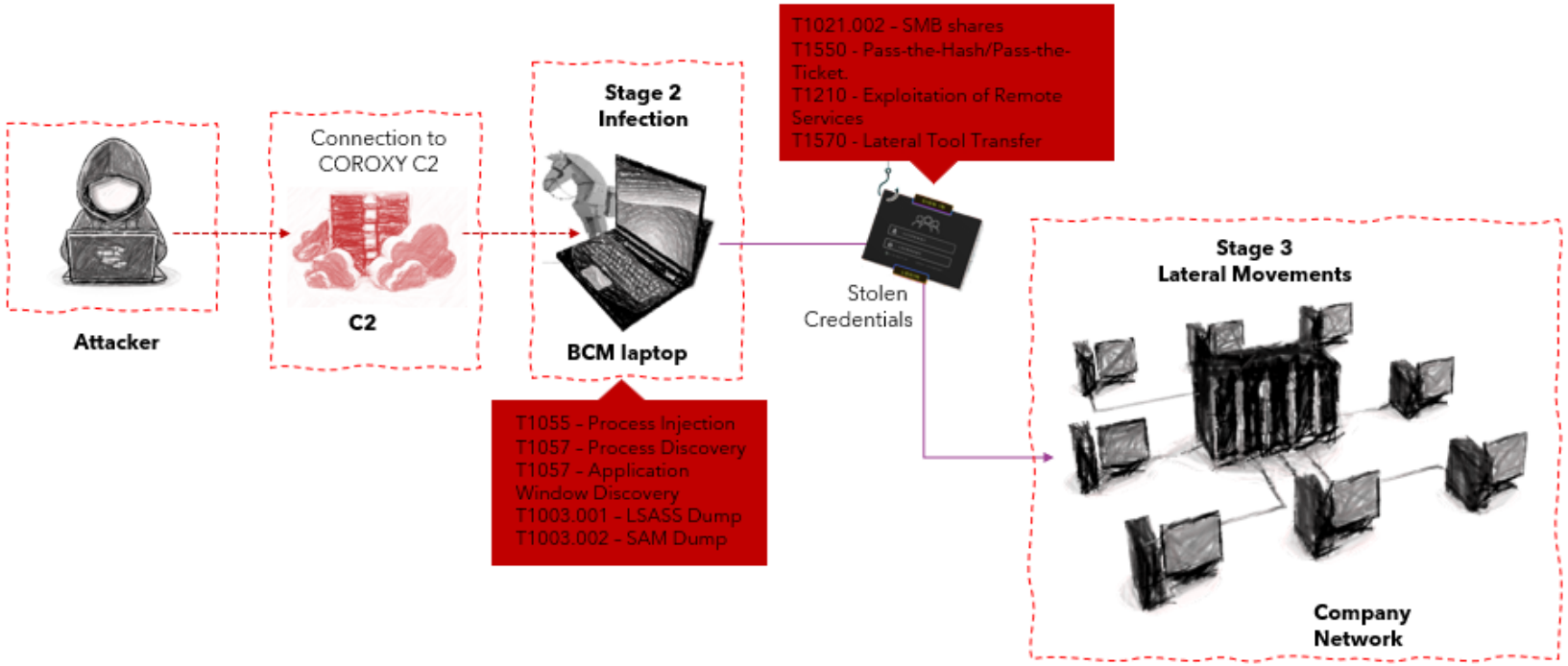


NetLOADER impiega hook Just-In-Time (JIT) e chiamate dinamiche alle API di Sistema per bypassare i controlli delle soluzioni di sicurezza quali Antivirus e EDR.

FURTO DI CREDENZIALI

Durante la prima fase l'attaccante si concentra su:

- Sulla workstation del BCM (membro del dominio), enumera i processi in esecuzione, i profili del browser e le credenziali salvate, oltre agli archivi locali di password, per raccogliere username e password aziendali, token VPN e cookie SSO.
- Estrare tutte le credenziali legate a portali, condivisioni SMB e strumenti di accesso remoto.
- Distribuire ulteriore toolset per effettuare il dump di LSASS, degli archivi password del browser e di eventuali credenziali RDP memorizzate in locale.
- Tentare i primi movimenti laterali verso la rete interna sfruttando le credenziali rubate.



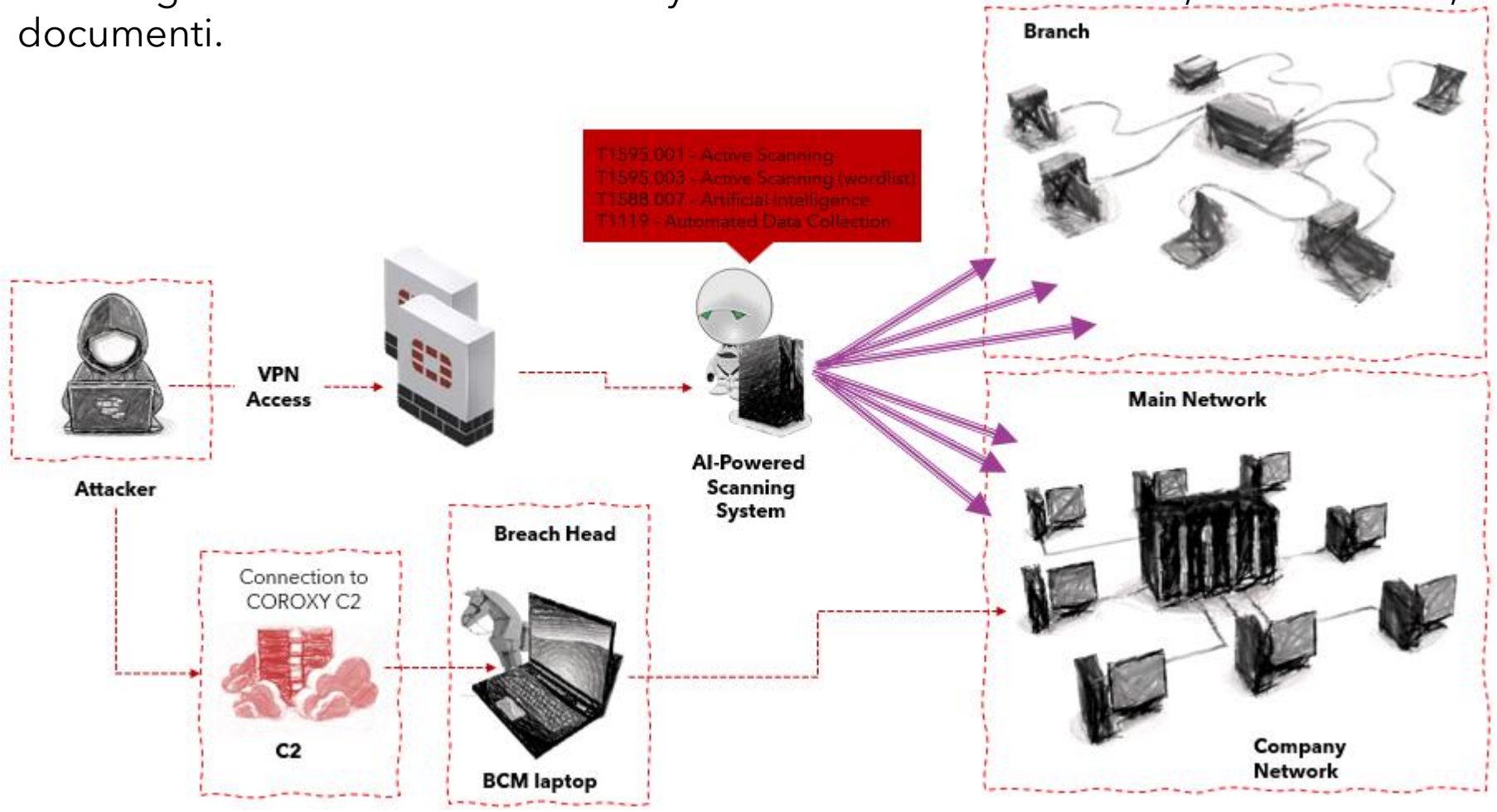
SISTEMARE IL PERIMETRO

Utilizzando le credenziali raccolte e le debolezze pubblicamente documentate nello stack di autenticazione e SSO di Fortinet, l'attaccante prende di mira il perimetro FortiGate dell'azienda accedendo alla sua interfaccia dalla rete interna.

Sfrutta un authentication-bypass (CVE-2024-55591) che consente all'attore malevolo di accedere alla console di amministrazione di FortiOS senza credenziali valide e creare nuovi account admin locali ("sec4dmin", "b4ckup", "remoteadmin1").

INTERNAL RECONNAISSANCE E MAPPATURA DEGLI ASSET CON L'IA

Avendo un accesso stabile sia alla workstation del BCM sia al firewall, l'attaccante predispone un sistema di scanning e lancia attività di discovery interna semi-automatizzate, assistite da IA, alla ricerca di share, file e documenti.

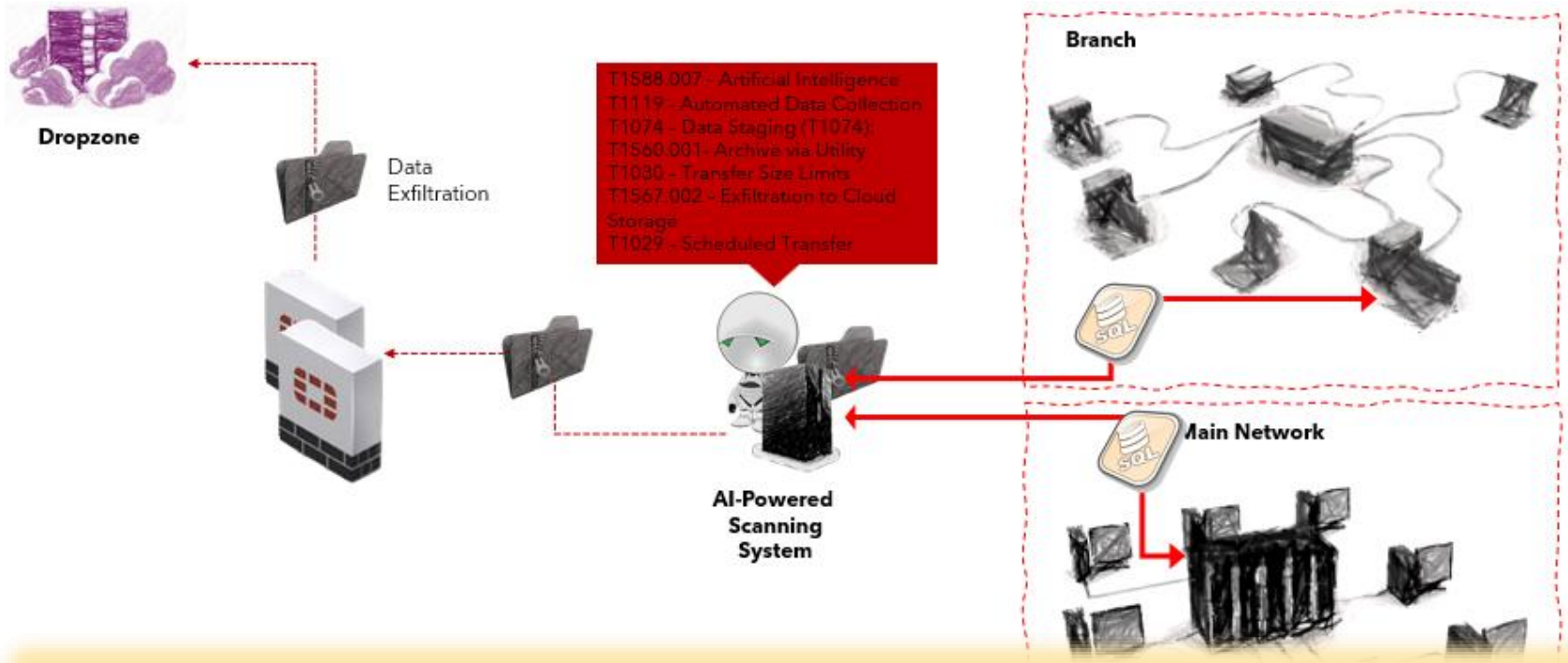


Nel giro di poche ore l'attaccante dispone di un grafo che include:

- Jump server utilizzati da DBA e fornitori per gestire i sistemi di betting in produzione.
- Applicativi che ospitano PII, dati di pagamento e storici delle scommesse.
- Account di servizio con ampi privilegi che, una volta compromessi, consentono movimenti laterali senza attività vistosa.

COLLEZIONAMENTO ED ESFILTRAZIONE DEI DATI

Guidato da questa mappa generata dall'IA, l'operatore può ora automatizzare la ricerca dei file e l'esfiltrazione dei dati:



- L' IA può eseguire query SQL ed esportare blocchi di dati utente: profili di account, storici delle scommesse, token di pagamento, eventuali numeri di carta completi presenti e documenti di identità scansionati usati per il KYC.
- I dati PII e finanziari esfiltrati vengono organizzati in archivi cifrati, che vengono poi esfiltrati attraverso la VPN del firewall (che ora appare come traffico amministrativo "legittimo").

A questo punto l'attaccante dispone sia di dati di grande valore (da vendere o usare per estorsione) sia di una comprensione profonda della topologia operativa della piattaforma di betting.

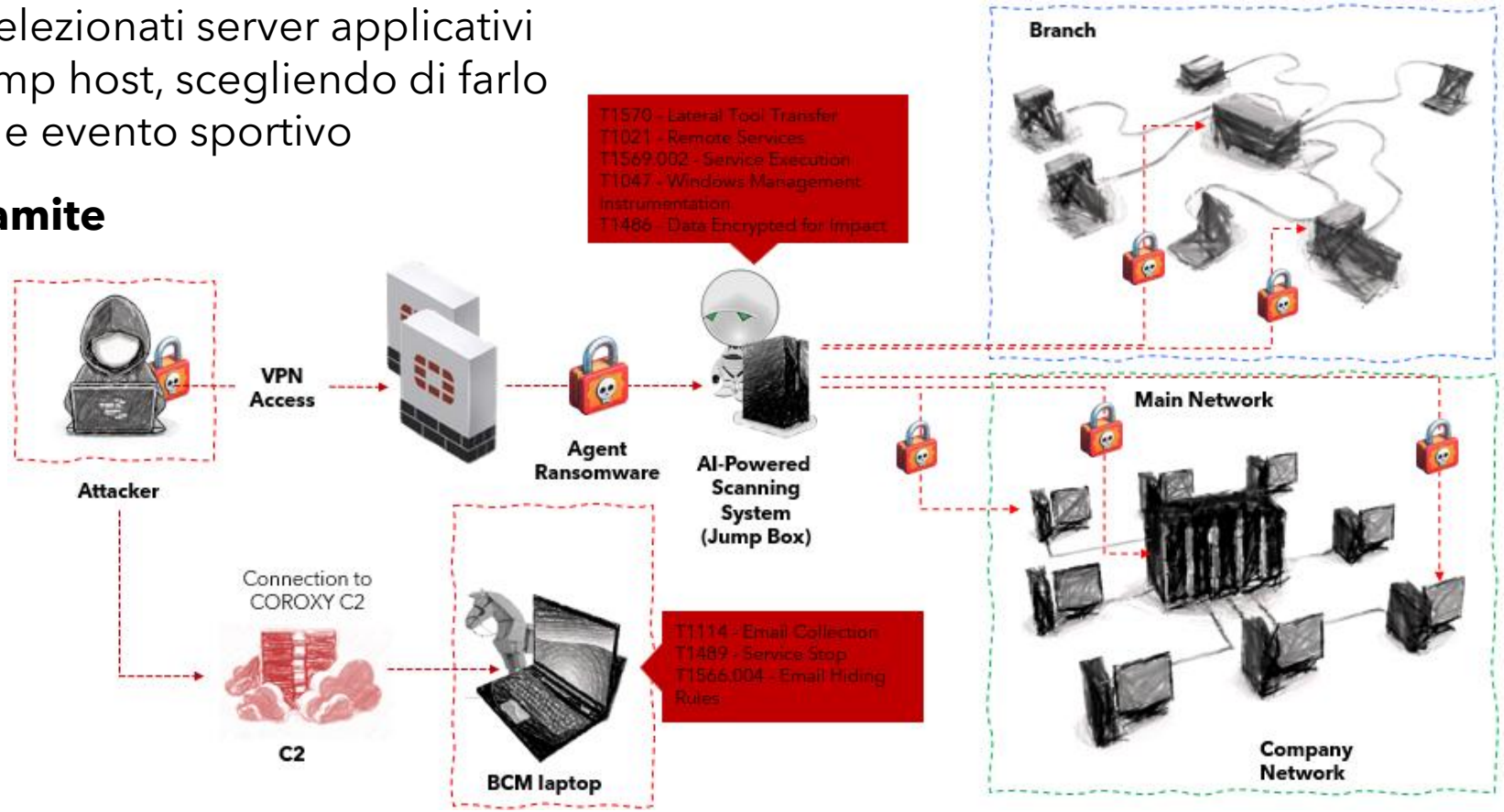
IMPATTO: RANSOMWARE, SABOTAGGIO E BLOCCO DEL BUSINESS

Per massimizzare leva e impatto, l'attaccante esegue il Ransomware operando su due fronti:

- **Distruzione operativa:**
  Distribuisce il ransomware su selezionati server applicativi e database raggiungibili dai jump host, scegliendo di farlo in contemporanea ad un grande evento sportivo

- **Sabotaggio della continuità tramite l'endpoint del BCM:**

  Rallenta il coordinamento dell'incidente: corrompendo dati e inviando comunicazioni interne fuorvianti.



DOVE HA FATTO LA DIFFERENZA L'IA

In questi casi, l'IA agisce come acceleratore in tre punti:

- **Targeting e qualità del phishing:**
  L'attività di ricognizione e redazione basata su LLM trasforma social media pubblici e LinkedIn in esche altamente personalizzate e cariche dal punto di vista emotivo, che superano senza problemi la classica formazione "cerca gli errori di ortografia".
- **Analisi dei percorsi di attacco:**
  Le configurazioni interne e le scansioni rapide vengono convertite dall'IA in un grafo dei percorsi più probabili verso i jump server e i tier dei database, comprimendo in poche ore quello che a un red-teamer umano richiederebbe giorni.
- **Gestione della campagna:**
  Gli agenti IA curano i dati rubati, etichettano i record ad alto valore (carte di credito, documenti di identità) e redigono note di riscatto e informative di data breach in più lingue, riducendo il carico di lavoro dell'operatore e permettendo una monetizzazione più rapida.

COSA ASPETTARCI NEI PROSSIMI 12 MESI

IN THE RANSOMWARE WORLD

Il ransomware tradizionale ha un ciclo di vita predicibile: gli attaccanti lo sviluppano, lo compilano, lo distribuiscono e quando le tecnologie lo identificano e contrastano ne sviluppano un altro.

L'IA stravolge questo ciclo:

Malware can evolve in real time:	**Barrier to entry is lowered**:	**Cross-platform attacks become easier**	**Detection and response must evolve**
• If ransomware dynamically generates parts of its logic, defenders can't rely on static signatures or known IoCs (Indicators of Compromise).	• With GenAI assistance, attackers don't need advanced coding skills. Even novice hackers could build functional ransomware by leaning on AI prompts.	• Golang plus AI-driven adaptability means one ransomware family could hit multiple operating systems without major rewrites.	• Security teams will need to focus more on **behavioral analysis** and **AI red-teaming**, rather than chasing signatures.

Ai miei occhi PromptLock non è una minaccia al momento, ma è importante capirne il senso, per me è il futuro dei ransomware.

Threat Type	Primary Defensive Control	Validation Method
AI-generated phishing	Phishing-resistant MFA, secure email controls, user training	Phishing simulations and red teaming
Deepfake identity fraud	Out-of-band verification, strong approval workflows	Executive fraud tabletop exercises
AI-assisted malware	Behavioral detection, EDR/XDR, sandboxing	Adversary emulation and malware response testing
Automated vulnerability discovery	Continuous scanning, patch management, exposure monitoring	Continuous penetration testing
Cloud AI model hijacking	Cloud IAM hardening, API access control, logging	Cloud security assessments and attack path validation

ISACA Rome Chapter

Grazie!

If you like this, follow me:

https://www.netwitness.com/resource/webinars-on-demand/

Youtube: Hack Talkz Channel
https://www.youtube.com/playlist?list=PLxPfQ4Dcn_BG5lgdfVYyT5TTm4i11OqYz