Artificial intelligence al servizio della cyber security

Viterbo, 7 ottobre 2022

Alessio Mercuri
Security Engineer

VECTRA®
SECURITY THAT THINKS.®

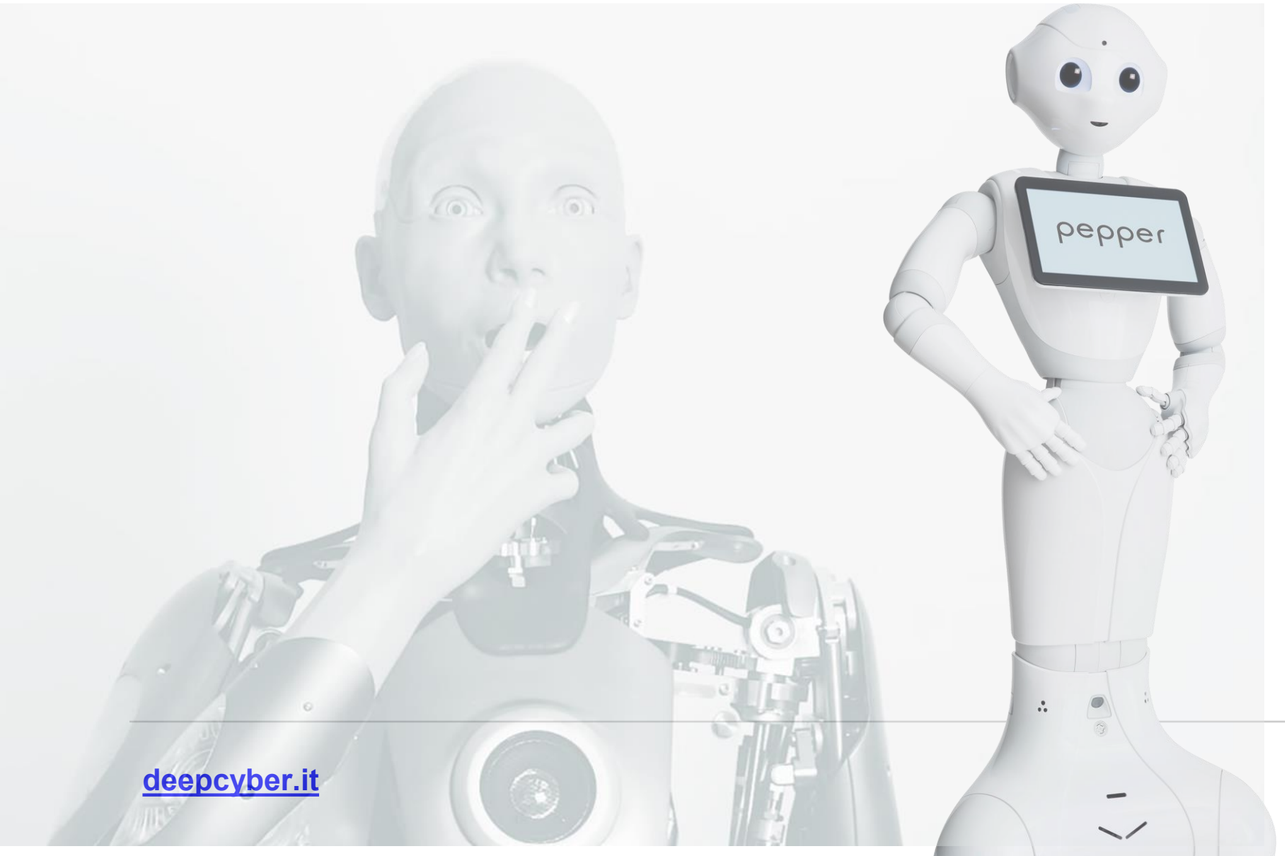
Gerardo Costabile
CEO

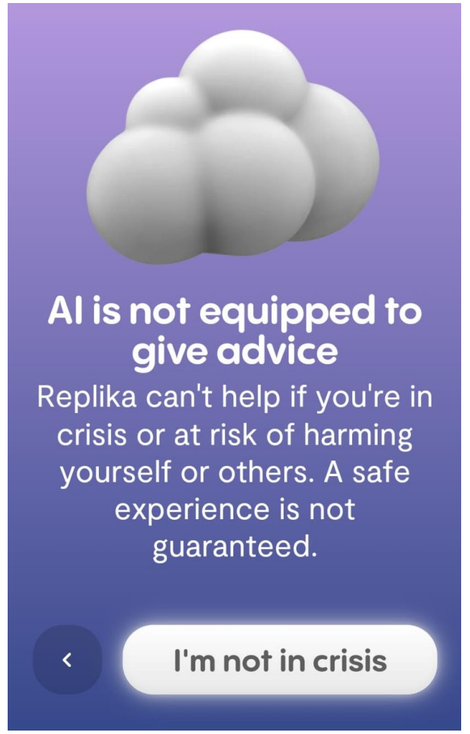
DEEPCYBER
a Maggioli Group company
Advanced Intelligence, Protection, Antifraud.

Il primo robot del III secolo a.C.

Progettato da un ingegnere e scrittore vissuto nel III sec a.C. a Bisanzio, il *Servo automatico* di Philon è il più antico robot della storia con sembianze umane. Capace di servire vino, la sua struttura si compone di più elementi: all'interno, sotto la tunica, sono posti due contenitori, rispettivamente di vino e acqua, collegati alla brocca attraverso tubi d'aria che passano lungo il braccio destro dell'automa e trasportano il liquido. La mano sinistra, dove viene poggiata la tazza, è connessa a un sistema di leve che regola l'andamento di entrambi gli arti: quando viene poggiata la tazza, il braccio sinistro scende mentre quello destro si muove per versare il vino mescolato all'acqua; man mano che viene riempita la tazza, il peso aumenta e di conseguenza il braccio scende fino a raggiungere il livello più basso e quindi il limite di capienza. Da qui il sistema si blocca, l'ospite può servirsi e le braccia ritornano al punto di partenza, pronte per ricominciare. Il Servo automa è oggi conservato al Museo Tecnologico dell'Antica Grecia Kostas Kotsanas a Katakolo nell'Elide (Grecia occidentale).

Robot umanoidi e social robots (esempio, Pepper)





App «Replika» (con oltre 7 Mln di download).

AI e «superintelligenza»

Nel 2007 si registravano già 53 definizioni di intelligenza e 18 definizioni di intelligenza artificiale.

«L'amicizia, l'IA e molte altre cose della vita sono come la pornografia: non sono definibili ma le riconosciamo quando le incontriamo». Questo perché *«è una scorciatoia, usata per riferirsi approssimativamente a diverse discipline».* (Prof. Luciano Floridi, 2022)

Secondo Nick Bostrom la «superintelligenza» si definisce come qualunque intelletto che superi di molto le prestazioni cognitive degli esseri umani in quasi tutti i domini di interesse.

In base a questa definizione Deep Fritz, il programma che gioca a scacchi, non è una superintelligenza, poiché è intelligente solo nel dominio limitato degli scacchi.

Secondo Floridi, una lavastoviglie che lava bene i piatti o lo faccia meglio di me, non vuol dire che lo faccia «come» me o che abbia bisogno di intelligenza per farlo.

Le forme di «superintelligenza»

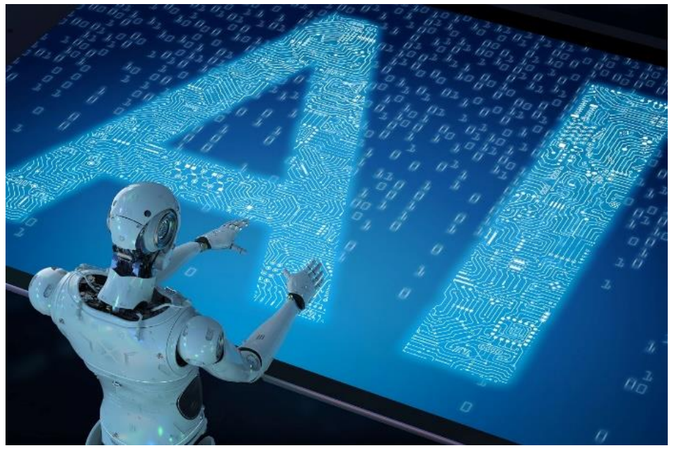
Distinguiamo 3 tipologie:

a) **Superintelligenza di grande velocità:** è un intelletto come la mente umana, ma più veloce (esempio, emulazione di un cervello 10.000 volte più veloce di uno biologico). Con un fattore di accelerazione di 1 milione, un'emulazione potrebbe realizzare un intero millennio di lavoro intellettuale in un solo giorno lavorativo. Ad una mente così veloce gli eventi del mondo esterno sembrano verificarsi al rallentatore.

b) **Superintelligenza collettiva:** è un sistema composto da un gran numero di intelletti minori tale che le sue prestazioni complessive del sistema, in numerosi domini molto generali, superano di gran lunga quelle di qualsiasi sistema cognitivo esistente. L'intelligenza collettiva eccelle nella risoluzione di problemi facilmente scomponibili (es. costruzione navetta spaziale etc) e può, nel tempo, avere una evoluzione in termini di integrazione, fino a diventare un unico grande intelletto, una singola «grande mente».

c) **Superintelligenza di qualità:** un sistema che è almeno altrettanto veloce di una mente umana e qualitativamente molto più intelligente.

Artificial intelligence black hat & white hat



AI, investigazioni e giustizia: alcuni esempi negli ultimi anni

- **ROSS**, il robot avvocato sviluppato grazie a Watson di IBM oppure il canadese **KYRA** o ancora **LUMINANCE** (Università di Cambridge), premiata come Best Artificial Intelligence Product in Legal durante il convegno londinese CogX.

- **“X-LAW”**, è una soluzione ideata dalla questura di Napoli (e utilizzata anche a Venezia) sulla «predizione» dei furti in città (parafrasando il famoso film “minority report”).

Moltissime sono le applicazioni commerciali, ormai, per utilizzare AI nell’antifrode e nella cyber security.

I diversi vantaggi da valutare in ambito «digital»

- **Intelligenza di velocità o collettiva**, ovvero legata alla mole di dati e quindi alla possibilità di velocizzare l'analisi, automatizzando quanto più possibile l'analisi e l'identificazione dei pattern di interesse;

- **Orchestrazione o automatismi di contrasto**, per ridurre il tempo di reazione ed anche la manualità;

- **Unknown unknowns**, ovvero individuazione di modelli e pattern di frode/illecito/attacco, ancora sconosciuti. In tal caso, il modello si baserà su entità matematiche complesse e funzioni non lineari, ovvero su un metodo non parametrico. A seconda della configurazione, tali modelli possono operare in presenza di supervisione o meno.

Alcuni punti di attenzione

- **Opacità:** L'AI è – se davvero AI - un'entità "black box": sebbene la "decisione" prodotta sia osservabile, risulta difficile capirne i motivi e/o estrarli sotto forma di semplici regole.

- **Eccessiva delega e fiducia alla tecnologia/algoritmi**

DARPA Cyber Grand Challenge 2016

Nell'estate 2016, sette squadre finaliste nella Cyber Grand Challenge organizzata dal DARPA degli Stati Uniti, hanno «combattuto» con sistemi di IA in grado di scansionare autonomamente i server nella rete dei rivali alla ricerca di exploit e proteggere i propri server trovando e correggendo attivamente i bug del software. (Con premi di 2 milioni di dollari per il primo posto, 1 milione di dollari per il secondo e 750.000 dollari per il terzo)

Ma sul piano più industriale, come si sono mosse le soluzioni? Con quali paradigmi? Vediamo uno degli esempi più interessanti.

Applicare l'intelligenza artificiale al rilevamento delle minacce cyber

Per l'identificazione attiva delle minacce informatiche sono emersi due diversi paradigmi: uno basato sui dati matematici e uno basato sui metodi di attacco.

Il paradigma basato sui dati matematici è imperfetto per il rilevamento delle minacce

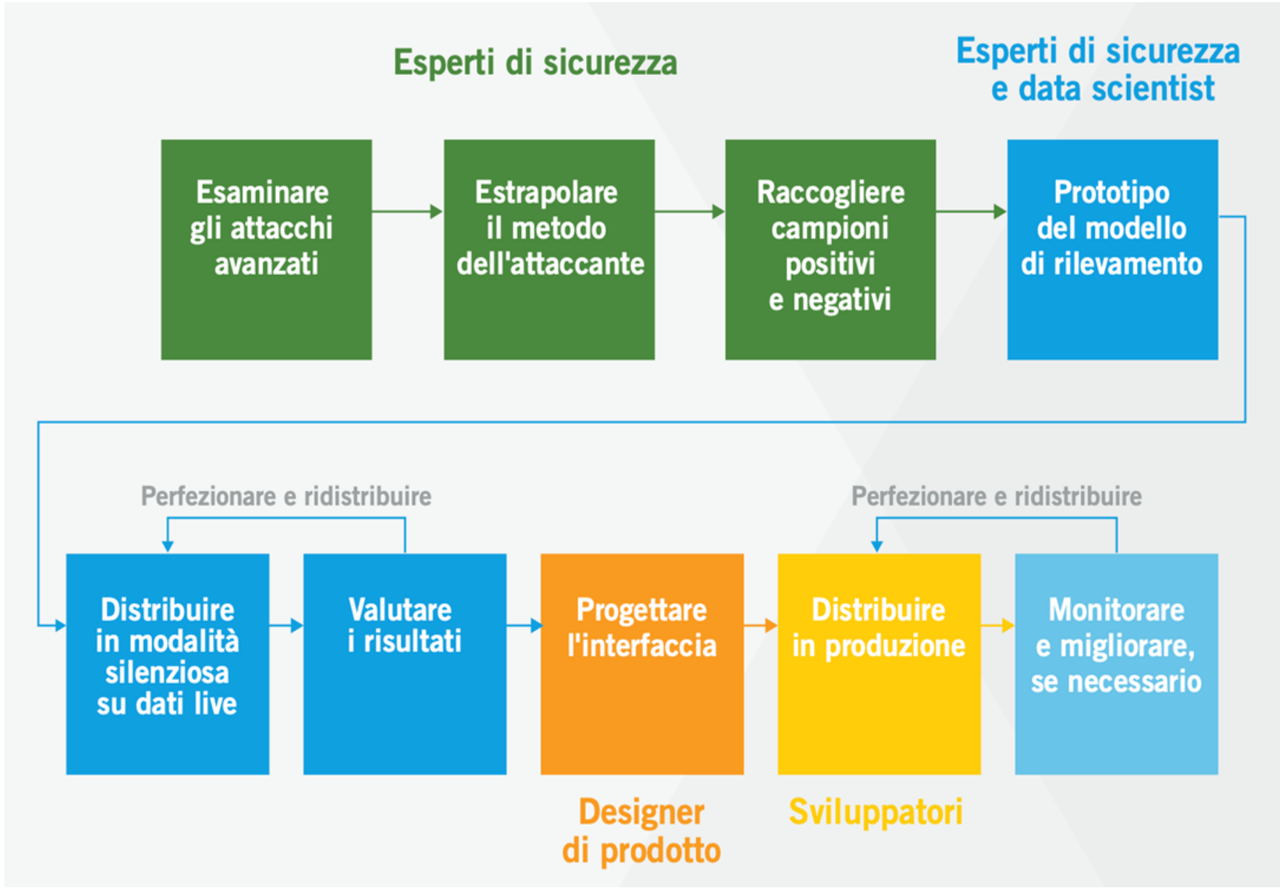
- Serie di dati statistici che usano un numero limitato di algoritmi generici impostati per rilevare anomalie o cambiamenti.
- Centinaia di regole statistiche combinate a filtri di soppressione espliciti per ridurre il numero di alert derivati da questo approccio generico.

Il paradigma basato sul metodo di attacco garantisce copertura massima e meno falsi positivi

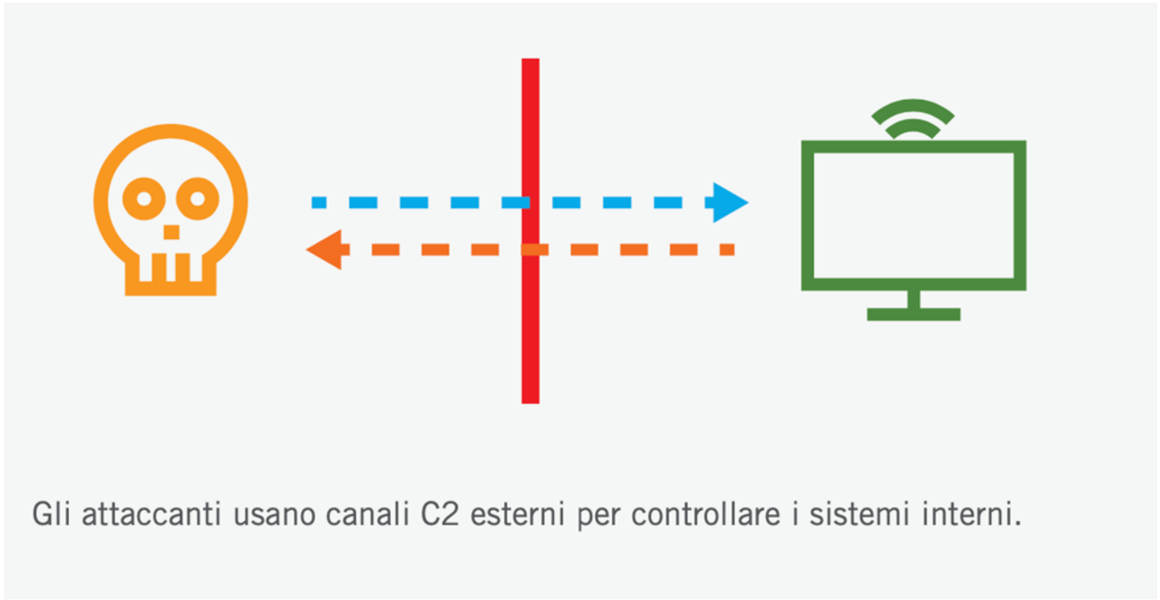
- Il paradigma basato sul metodo di attacco è un approccio che punta al metodo scelto dall'attaccante, e l'individuazione del modello corretto.
- Gli esperti di sicurezza definiscono il problema identificando il metodo usato dall'attaccante e i data scientist trovano l'algoritmo appropriato per identificare quel metodo

L’approccio di Vectra

- Vectra ha introdotto una nuova strategia per rilevare le minacce che colpiscono reti, cloud pubblici, app SaaS e identità, che si basa sull'analisi del metodo di attacco.
- Le procedure di rilevamento di Vectra hanno l'obiettivo di individuare gli attaccanti e i metodi di attacco impiegati, non semplicemente le anomalie.
- Ricercatori in ambito sicurezza e data scientist collaborano costantemente da oltre 10 anni attuando una strategia di rilevamento che risulta molto efficace nell'individuare i comportamenti di attacco senza generare un volume di traffico enorme.



Case study: applicazione dell'IA su canali C2 criptati



Gli attaccanti usano canali C2 esterni per controllare i sistemi interni.

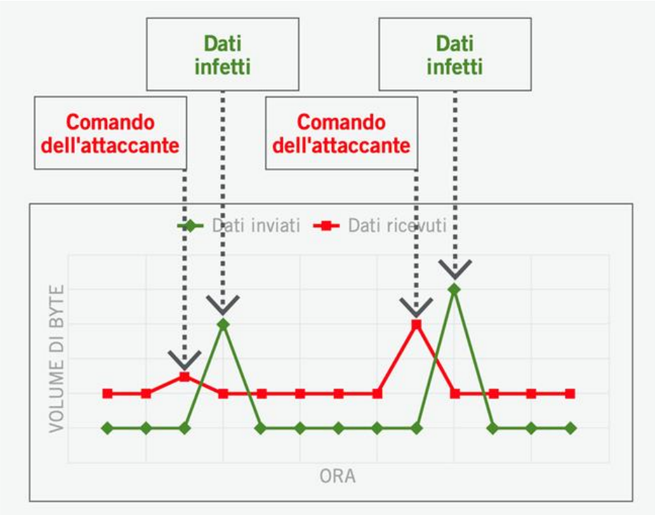
Metodo di attacco

- Ogni attacco esterno che avviene tramite la rete prevede la presenza di un canale C2, cioè di comando e controllo.
- L'attaccante che riesce a violare un host diffonde un software malevolo che stabilisce un contatto con un server esterno.
- L'host interno contatta il server esterno, che risponde diffondendo istruzioni che l'host infetto esegue, aprendo la porta all'attacco vero e proprio.
- Framework di attacco, come Cobal Strike e Metasploit o sviluppati ad-hoc.

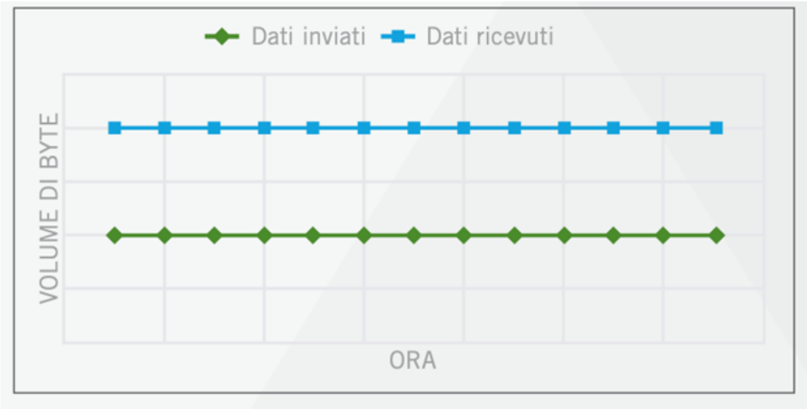
Case study: applicazione dell'IA su canali C2 criptati

Metodologia di rilevamento

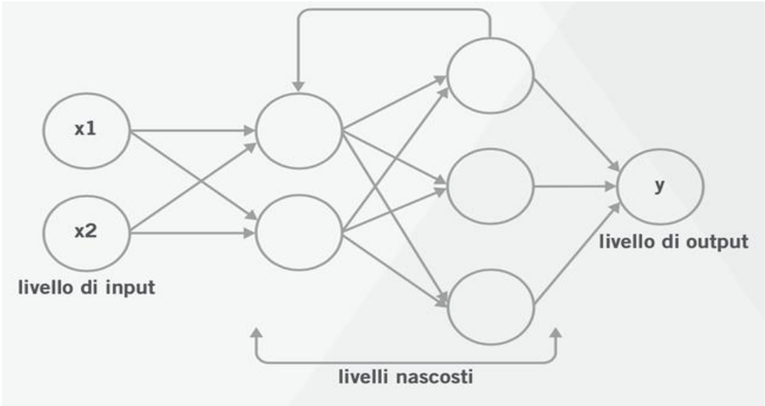
- Vectra rileva i canali C2 a prescindere che siano criptati o che siano state impiegate altre tecniche di elusione.



Per identificare i comportamenti di attacco Vectra utilizza una rete neurale ricorrente specifica denominata LSTM (Long Short-Term Memory).



- Gli indicatori più determinanti del metodo non sono gli aspetti circostanziali del traffico, come la presenza di domini rari o di agent, ma la rappresentazione grafica del traffico di rete nel tempo



Case study: applicazione dell'IA per contrastare l'abuso di credenziali con privilegi in rete e nel cloud

Gli attaccanti usano credenziali con privilegi per spostarsi lateralmente ed eseguire altre attività malevole nella rete e nel cloud.

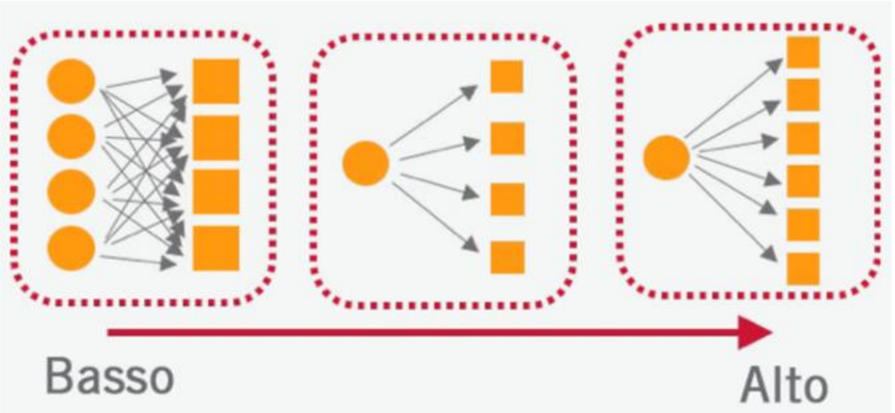
Metodo di attacco

- Gli attaccanti che riescono a procurarsi credenziali con privilegi si assicurano un accesso incontrastato a risorse di rete e cloud.
- Le credenziali sono un vettore di accesso "pulito" che consente di non utilizzare payload con exploit e malware, cioè strumenti che lasciano tracce e generano avvisi.
- L'imposizione del principio di accesso con privilegi minimi per gli utenti contribuisce a mitigare alcuni attacchi, ma la definizione dell'esatto privilegio minimo per gli utenti resta problematica.

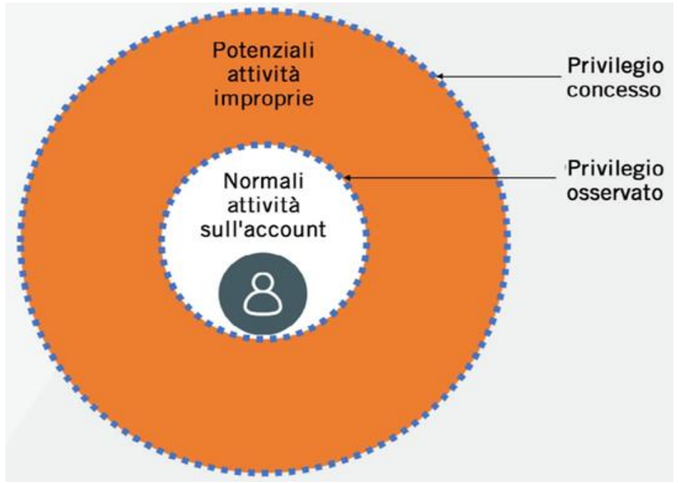
Case study: applicazione dell'IA per contrastare l'abuso di credenziali con privilegi in rete e nel cloud

Metodologia di rilevamento

- Vectra rileva gli abusi credenziali rubate sia negli ambienti di rete sia negli ambienti cloud.



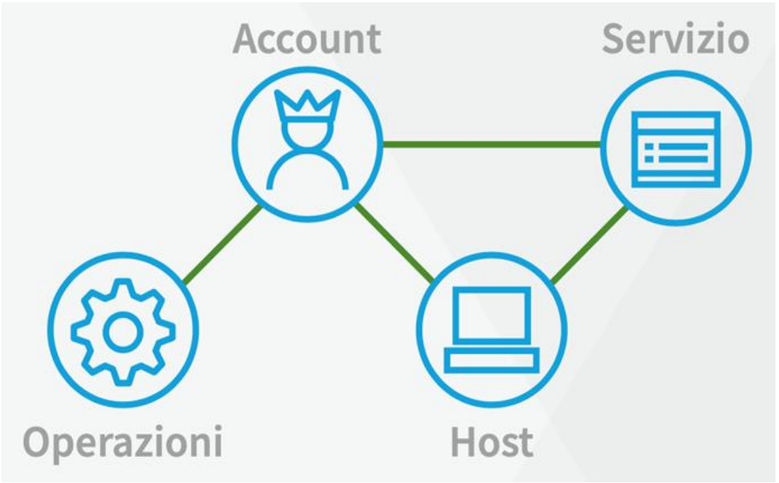
- Mappatura di tutte le interazioni tra account, servizi, host e cloud.
- Algoritmi unsupervsied e algoritmi personalizzati per il rilevamento di anomalie e implementazioni HDBSCAN, identificano i casi di uso improprio dei privilegi



Privilegio osservato

≠

Privilegio garantito



Grazie per l’attenzione!

Gerardo Costabile
gerardo.costabile@deepcyber.it

Alessio Mercuri
amercuri@vectra.ai

CYBER ACT FORUM 2022