VECTRA
SECURITY THAT THINKS

WHITE PAPER

L'intelligenza artificiale secondo Vectra

SOMMARIO

Vectra® protegge le aziende rilevando e bloccando gli attacchi informatici.

Vectra® è un leader nelle attività di rilevamento e neutralizzazione degli attacchi per le aziende che operano con ambienti ibridi e multicloud. La piattaforma Vectra utilizza l'intelligenza artificiale per rilevare istantaneamente le minacce che colpiscono cloud pubblici, identità, applicazioni SaaS e data center. Solo Vectra ottimizza l'uso dell'IA per rilevare i metodi impiegati dagli attaccanti – cioè le tattiche, tecniche e procedure dietro gli attacchi – anziché per limitarsi a segnalare uno scostamento rispetto al funzionamento "normale". Il risultato è la segnalazione altamente affidabile delle minacce presenti e una definizione chiara del contesto grazie a cui gli esperti di sicurezza possono reagire alle minacce in tempi brevi e bloccare velocemente gli attacchi in corso. In tutto il mondo le aziende si affidano a Vectra per contrastare le minacce informatiche più pericolose e per prevenire gli attacchi ransomware, le compromissioni della supply chain, i furti di identità e qualsiasi altro incidente informatico rivolto al business. Per maggiori informazioni, visita il sito vectra.ai.

Introduzione

In fatto di intelligenza artificiale (IA), l'unica bussola a cui si affida Vectra è la data science. Da sempre sosteniamo che un utilizzo congiunto ed efficiente di data science e intelligenza artificiale ha il potenziale di trasformare la lotta agli attacchi informatici perché rende più incisivi gli interventi di chi si occupa di sicurezza informatica. Tuttavia, l'intelligenza artificiale non è tutta uguale. Questo documento spiega che cos'è l'intelligenza artificiale e quali sono i termini chiave associati alle soluzioni di IA, definisce le due principali metodologie alla base dell'applicazione dell'IA per il rilevamento delle minacce e infine esamina nel dettaglio in che modo Vectra utilizza l'IA per individuare le minacce.

Che tu sia scettico circa l'impiego dell'intelligenza artificiale o un convinto sostenitore del suo potenziale, questo documento fa per te.

Che cos'è l'intelligenza artificiale?

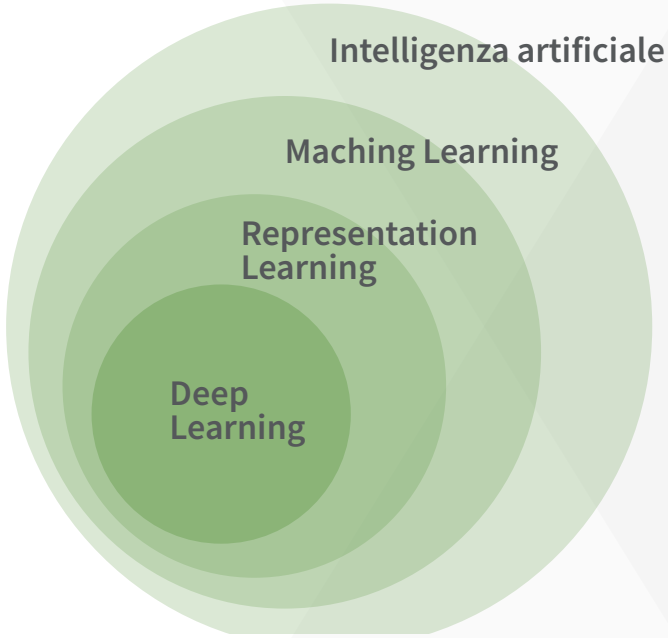
Definizione di intelligenza artificiale

I termini *intelligenza artificiale*, *machine learning* e *deep learning* sono spesso usati impropriamente come sinonimi o come diverse sfumature qualitative dello stesso concetto, invece non è così. Questi tre termini sono correlati, ma ognuno ha un significato preciso. Conoscere l'esatto significato di ogni termine aiuta a capire meglio come funzionano gli strumenti che integrano funzionalità di intelligenza artificiale.

Intelligenza artificiale (IA): qualsiasi sistema in grado di eseguire automaticamente compiti caratteristici dell'intelligenza umana. È una disciplina ampia che comprende varie branche, come il *machine learning*, il *representation learning* e il *deep learning*. Il termine intelligenza artificiale si applica indifferentemente a un sistema che si basa sull'uso di regole esplicitamente programmate e a un sistema che ha autonomamente acquisito conoscenze elaborando dei dati. La forma di IA che impara dai dati è alla base della tecnologia che fa funzionare le auto a guida autonoma e gli assistenti virtuali e rientra nel sottogruppo del machine learning.

Machine Learning (ML): significa "apprendimento automatico" ed è la branca dell'intelligenza artificiale che si concentra sulla capacità dei sistemi di apprendere autonomamente da una serie di dati pur senza essere stati programmati esplicitamente da un umano. Questi sistemi sono in grado di elaborare miliardi di dati, renderli rappresentativi e utilizzabili per rispondere a nuove istanze di dati.

Representation Learning (RL): sebbene sia una disciplina poco discussa, è essenziale per svariate tecnologie di IA comunemente usate. Si tratta di un insieme di tecniche che consente a un sistema di apprendere nuove rappresentazioni astratte da una serie di dati. Un esempio di RL è la trasformazione di immagini di dimensioni diverse in un elenco di numeri di lunghezza uniforme che rappresenta un distillato delle immagini originali. Grazie a questa astrazione, i sistemi a valle riusciranno a gestire in modo più efficace nuovi tipi di dati.



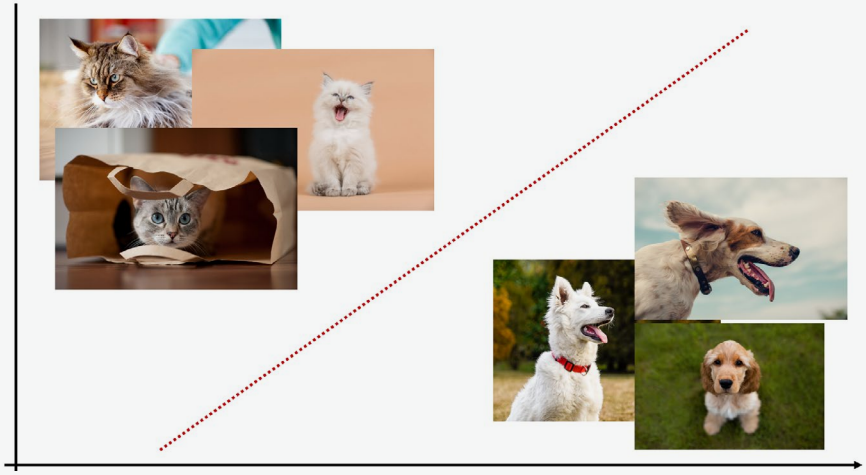
La relazione tra le varie branche dell'intelligenza artificiale.
Riferimento: "Deep Learning" Goodfellow, Bengio & Courville (2016)

Deep Learning (DL): spesso associato alle reti neurali, questo approccio coniuga ML e RL per cercare nei dati una gerarchia di astrazioni che rappresentino gli input in maniera progressivamente più complessa. Ispirandosi al funzionamento del cervello umano, i modelli di DL utilizzano più strati di neuroni il cui peso sinaptico si adatta in funzione degli input: gli strati più profondi della rete apprendono nuove rappresentazioni astratte che rendono più semplice classificare un'immagine o tradurre del testo. Il deep learning è una tecnica efficace per risolvere alcuni tipi di problemi complessi, ma non è assolutamente un passepartout per l'automazione intelligente.

Le diverse tecniche usate dagli algoritmi di apprendimento

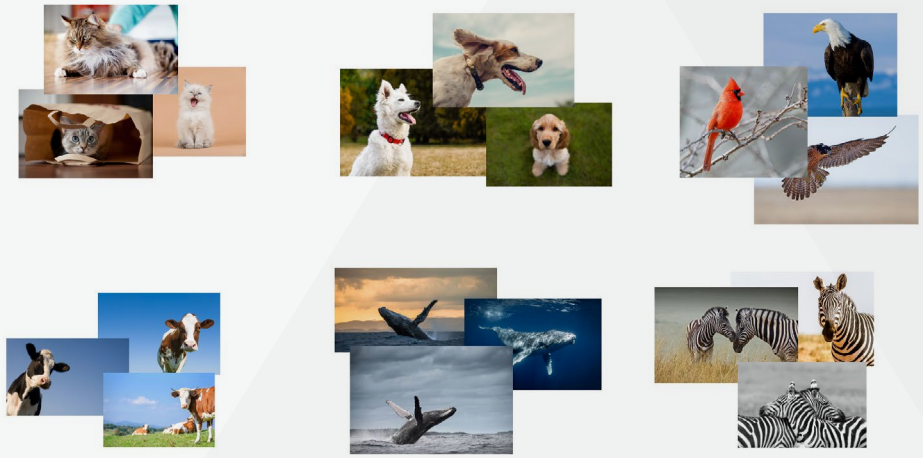
Una delle principali caratteristiche degli algoritmi di ML consiste nella loro capacità di classificare le istanze di dati in entità separate. Le due principali categorie di machine learning a supporto di questa capacità sono l'apprendimento **supervisionato** e **non supervisionato**.

L'apprendimento **supervisionato** si basa sulla capacità del modello di apprendere da una serie di dati preventivamente etichettati. Sulla scorta di quanto appreso, quando viene posto di fronte a dati nuovi, il modello riuscirà a prevedere un'etichetta. Nell'esempio che segue, se si alimenta un modello di apprendimento supervisionato con grandi quantità di immagini di cani e gatti, quando viene proposta una nuova immagine, il sistema sarà in grado di prevedere se si tratta di un cane o di un gatto. L'apprendimento supervisionato presuppone l'utilizzo di una grande quantità di dati etichettati con cui allenare il modello, il quale, una volta allenato, risulterà altamente efficace a generalizzare e quindi a etichettare correttamente nuove istanze di dati.



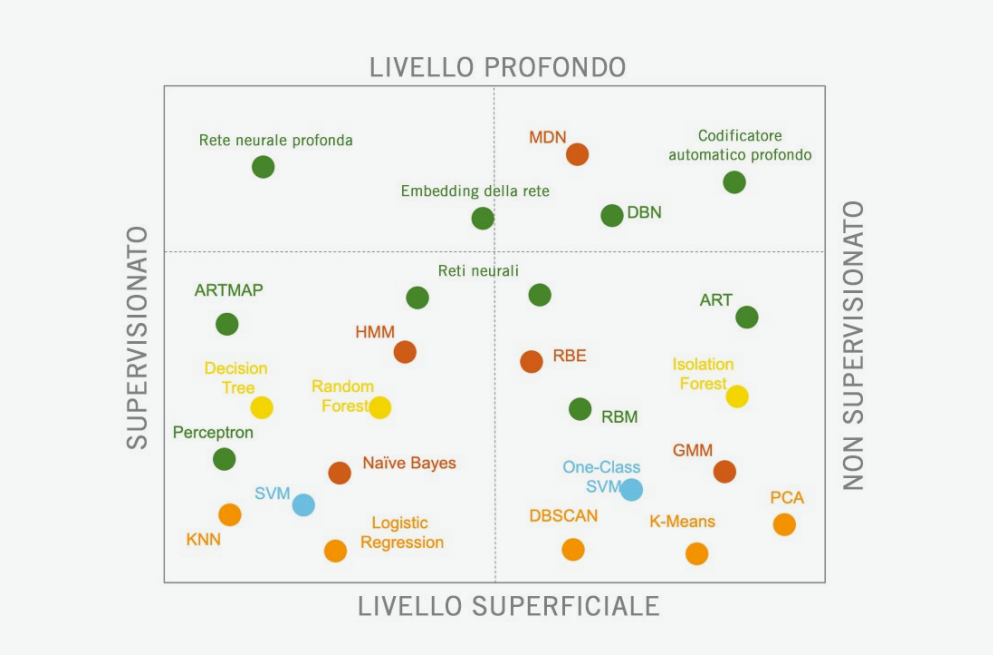
L'apprendimento supervisionato utilizza dati già etichettati per individuare i fattori che contraddistinguono le diverse etichette. I modelli che si basano su questo tipo di apprendimento sono in grado di etichettare nuovi dati.

L'apprendimento **non supervisionato** si basa sulla capacità del modello di apprendere da una serie di dati non etichettati. In questo caso il modello identifica autonomamente una struttura nei dati forniti e riesce in seguito a determinare se e come una nuova istanza di dati rientra in tale struttura. Il vantaggio dei modelli non supervisionati è che non richiedono un allenamento preventivo. È un approccio che possiede un'eccellente capacità di individuare i dati che si differenziano dagli altri, ma non è altrettanto efficace a classificare tali anomalie.



L'apprendimento non supervisionato impara la struttura sottostante dei dati non etichettati. I modelli che si basano su questo tipo di apprendimento sono in grado di determinare se e in che misura i nuovi dati rientrano in questa struttura.

Nell'ambito di questi approcci ad ampio spettro rientrano i diversi algoritmi di apprendimento riportati di seguito, ma i ricercatori ne inventano continuamente di nuovi. Inoltre, è possibile creare sistemi più complessi combinando tra loro gli algoritmi. A questo punto ci si chiede come faccia un data scientist a scegliere l'algoritmo o gli algoritmi giusti per risolvere un problema specifico. Esiste un algoritmo che garantisce le prestazioni migliori a prescindere dal problema da risolvere?



Esiste un numero elevatissimo di algoritmi di machine learning con pregi e difetti diversi che li rendono più o meno appropriati a un problema.

Il teorema "no free lunch"

Nessun singolo algoritmo di apprendimento automatico è universalmente quello che assicura le migliori prestazioni per tutti i problemi. Questo è il senso del teorema "no free lunch", cioè non c'è una soluzione unica che funziona in tutte le salse. In pratica, dato un problema, ci sarà sempre un algoritmo specifico che darà risultati migliori di un algoritmo generico. La necessità di disporre di algoritmi specifici per affrontare problemi specifici conferma la necessità primaria di avere a disposizione un numero di algoritmi in costante crescita. In alcune situazioni una rete neurale supervisionata darà risultati migliori e in altre sarà un clustering gerarchico non supervisionato ad assicurare prestazioni ottimali.

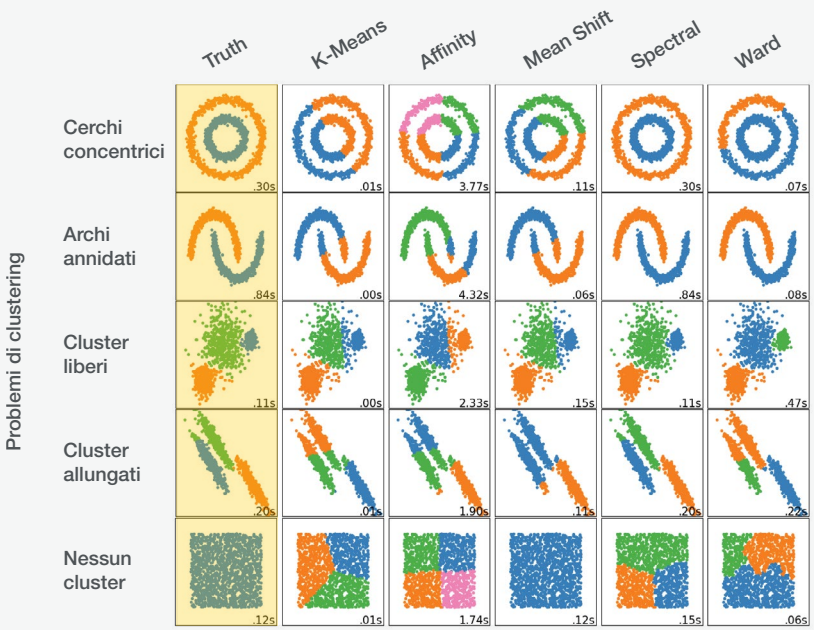
L'algoritmo utilizzato per il riconoscimento delle immagini nelle applicazioni di guida autonoma, ad esempio, non può essere applicato alle funzionalità di traduzione automatica. Ogni algoritmo ha un impiego specifico ed è ottimizzato per una determinata applicazione e per i dati che il modello elabora.

Il teorema "no free lunch": nessun singolo algoritmo è universalmente quello che assicura le migliori prestazioni per tutti i problemi.

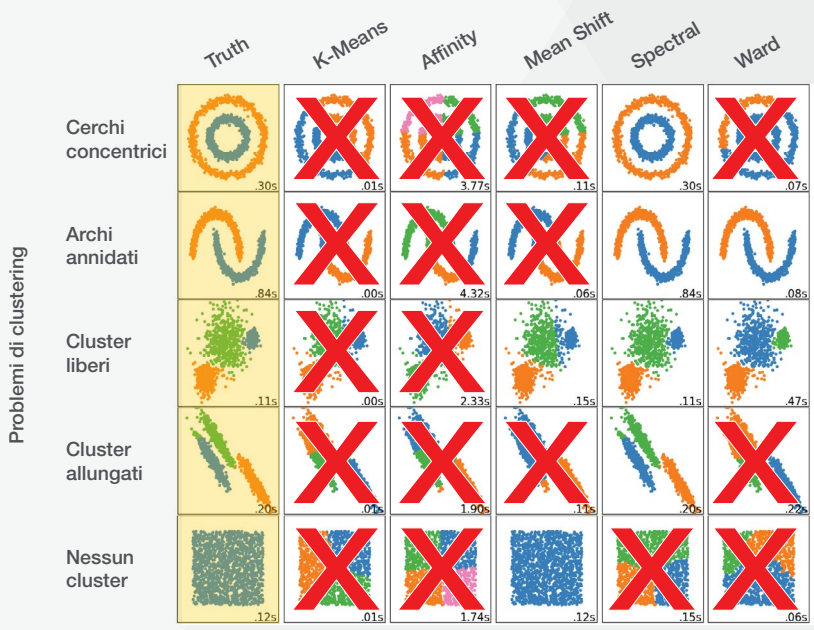
Individuare lo strumento giusto

Ma allora, come fa un data scientist a scegliere l'algoritmo giusto? Deve combinare scienza e intuizione. Combinando la definizione del problema e una profonda conoscenza dei dati, il data scientist viene indirizzato nella giusta direzione. Però attenzione: se procede nel modo sbagliato i risultati non saranno solo mediocri, saranno completamente sbagliati! L'esempio seguente lo dimostra chiaramente.

Per ogni set di dati l'algoritmo prescelto genera risultati molto diversi. Non solo per ogni problema esiste un algoritmo ottimale, ma alcune scelte causano risultati decisamente indesiderati. Ecco perché scegliere l'approccio giusto per il problema in questione è fondamentale.



Confronto tra i risultati prodotti dagli algoritmi di ML (asse X) rispetto a diversi set di dati (asse Y). Le etichette vere sono indicate in giallo. Adattato da scikit-learn.org.



Confronto dei risultati. La X indica una previsione errata che causerebbe risultati indesiderati. Nessun algoritmo è efficace per ogni set di dati. Adattato da scikit-learn.org.

Come si misura *l'efficacia*?

Per i data scientist, come misurare l'efficacia di un modello è un aspetto importante nella scelta del modello più adatto. *L'accuratezza* è un attributo molto citato.

$$Accuratezza = \frac{(Veri\ positivi + Veri\ negativi)}{(Veri\ positivi + Veri\ negativi + Falsi\ positivi + Falsi\ negativi)}$$

Come metrica l'accuratezza ha sicuramente un grande valore, anche se può falsare la verità circa le reali prestazioni di un modello. Poniamo che l'obiettivo sia classificare dei dati assegnando l'etichetta A o B. Se l'etichetta A ha una frequenza 1000 volte più alta dell'etichetta B, si può facilmente raggiungere un'accuratezza del 99,9% assegnando ai dati sempre l'etichetta A. L'accuratezza sarà elevata, ma non si riuscirà mai ad assegnare correttamente l'etichetta B. È evidente che l'accuratezza non è la metrica ideale se vogliamo individuare i dati di tipo B. Fortunatamente, i data scientist dispongono di altre metriche per misurare e ottimizzare l'efficacia di un modello.

La precisione è una di queste. La precisione misura il grado di esattezza che dimostra un modello nell'individuare una particolare etichetta rispetto al totale delle previsioni di quell'etichetta.

$$Precisione = \frac{Veri\ positivi}{(Veri\ positivi + Falsi\ negativi)}$$

I data scientist che vogliono ottenere un punteggio di precisione elevato costruiranno modelli in grado di prevedere le etichette senza generare troppi falsi allarmi. Ciò che la precisione non ci dice, però, è se il modello non è riuscito a etichettare i casi che ci interessano. Un'altra metrica, il recupero, fornisce una prospettiva diversa.

Il recupero misura la frequenza con cui un modello individua correttamente una data etichetta rispetto a tutte le istanze dell'etichetta.

$$Richiamo = \frac{Veri\ positivi}{(Veri\ positivi + Falsi\ negativi)}$$

I data scientist che vogliono ottenere un punteggio di recupero elevato costruiranno modelli a cui non sfuggiranno le istanze ricercate.

Tenendo traccia e bilanciando precisione e recupero, i data scientist riescono a eseguire misurazioni efficaci e a ottimizzare le prestazioni dei modelli.

Per i data scientist, come misurare l'efficacia di un modello è un aspetto importante nella scelta del modello più adatto.

Applicare l'intelligenza artificiale al rilevamento delle minacce

L'IA e le varie discipline che la compongono vengono ampiamente utilizzate per cercare e bloccare gli attaccanti che prendono di mira le aziende. Per l'identificazione attiva delle minacce informatiche sono emersi due diversi paradigmi: uno basato sui dati matematici e uno basato sui metodi di attacco. In questa sezione analizzeremo le differenze tra questi due paradigmi e spiegheremo perché l'IA basata sui metodi di attacco garantisce risultati ottimali.

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

Nel paradigma basato sui dati matematici, i data scientist generano semplici serie di dati statistici usando un numero limitato di algoritmi generici impostati per rilevare anomalie o cambiamenti. Combinando questi dati, i ricercatori in seguito creano centinaia di regole statistiche. Se serve un nuovo dato statistico, viene creato utilizzando il medesimo approccio generico. Le regole statistiche vengono spesso combinate a filtri di soppressione espliciti nella fase di post-elaborazione al fine di ridurre il volume dei rilevamenti generati con questo tipo di approccio generico – come enunciato dal teorema "no free lunch", gli algoritmi generici generano prestazioni mediocri.

Procedura di rilevamento con paradigma basato sui dati matematici

Immaginiamo, ad esempio, di voler rilevare un canale di comando e controllo. Gli esperti di data science potrebbero iniziare generando una statistica relativa alla rarità di tutti i domini esterni. Gli esperti di sicurezza quindi decideranno la soglia di rarità che farà scattare l'avviso circa il canale di comando e controllo. Se molti domini usati dai dispositivi IoT si trovano al di sopra di tale soglia, sarà necessario applicare un filtro di soppressione per ignorare tutti i dispositivi IoT e ulteriori filtri di soppressione sugli agent, le subnet e gli altri attributi al fine di arrivare a un volume gestibile di avvisi. La natura generica di questo approccio richiede la presenza di queste regole di soppressione, nonostante comportino il rischio di trascurare una tecnica di elusione usata da un attaccante.

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 integra la definizione del problema, cioè il metodo scelto dall'attaccante, e l'individuazione del modello corretto. Gli esperti di sicurezza definiscono il problema identificando il metodo usato dall'attaccante nel senso più ampio – senza limitarsi al tool o alla vulnerabilità che è stata sfruttata – e i data scientist trovano l'algoritmo appropriato per identificare quel metodo e lavorano a stretto contatto con gli esperti di sicurezza nel valutare le successive iterazioni della soluzione. Questo procedimento consente di individuare direttamente il metodo scelto dall'attaccante e non solo di rilevare le anomalie superficiali segnalate dagli approcci basati sui dati matematici.

L'approccio basato sul metodo di attacco facilita il lavoro dei team di sicurezza perché assicura prestazioni migliori in termini di recupero e precisione, garantisce la resilienza ai cambiamenti negli strumenti di aggressione e genera un numero inferiore di tipi di rilevamento. Quando un particolare metodo di attacco inizia a farsi più frequente, viene avviato un nuovo processo di rilevamento. Nonostante la natura più sofisticata di questo approccio richieda tempi di sviluppo maggiori, i metodi degli attaccanti cambiano molto lentamente e quelli nuovi compaiono sistematicamente accanto a quelli più vecchi già coperti.

Procedura di rilevamento con paradigma basato sul metodo di attacco

Come funziona 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. Nelle prossime sezioni analizzeremo la portata della copertura offerta da Vectra e come avviene lo sviluppo delle procedure di rilevamento, il motore che raccoglie e genera i dati per i rilevamenti, in che modo i singoli eventi vengono correlati per individuare gli incidenti di sicurezza su cui intervenire e i meccanismi interni di due procedure di rilevamento di Vectra.

In che modo Vectra sviluppa le procedure di rilevamento

Le procedure di rilevamento di Vectra hanno l'obiettivo di individuare gli attaccanti e i metodi di attacco impiegati, non semplicemente le anomalie. La copertura è assicurata da ricercatori specializzati in sicurezza informatica con competenze diversificate e data scientist che vantano una profonda conoscenza di come estrarre valore da enormi quantità di dati complessi. Nel corso degli ultimi dieci anni, queste due squadre hanno messo a punto una strategia per il rilevamento delle minacce altamente collaborativa e scalabile in funzione dei domini di sicurezza e dei tipi di dati, che risulta molto efficace nell'individuare i comportamenti di attacco senza generare un volume di traffico enorme.

Gli esperti di sicurezza di Vectra sono presenti per l'intera durata dello sviluppo delle procedure di rilevamento, gestiscono il processo e monitorano costantemente i metodi di attacco in circolazione. Non cercano specifici tool o gruppi di attacco bensì i metodi generici a cui ricorrono gli attaccanti. Ad esempio, poniamo che gli esperti di sicurezza rilevino la funzionalità beacon usata da Cobalt Strike per gli attacchi ransomware.

> 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.

In che modo Vectra crea gli algoritmi IA basati sul metodo di attacco.

Anziché concentrarsi esclusivamente sui beacon di Cobalt Strike, estraggono gli elementi significativi di questa tecnologia e studiano il metodo di *controllo* dell'attaccante. In questo modo Vectra riesce a estendere la sua copertura ai due tool che a oggi sono utilizzati per eseguire questo metodo, ma anche a tutti quelli che verranno sviluppati in futuro.

Una volta identificato il metodo dell'attaccante, gli esperti di sicurezza lavorano alla raccolta di un corpus di campioni di dati innocui e nocivi. I dati nocivi provengono da varie fonti, tra cui metadati anonimizzati condivisi volontariamente dai clienti, algoritmi per la creazione sintetica di dati, incidenti informatici documentati pubblicamente e attacchi studiati nei nostri lab interni. Gli esempi di dati innocui provengono dall'ampio bacino di metadati anonimizzati dei clienti di Vectra.

Una volta che conoscono il metodo di attacco e i dati di supporto, gli esperti di sicurezza e i data scientist collaborano allo sviluppo di un prototipo di modello con una soglia ottimale per il rilevamento dei metodi di attacco. Il prototipo viene distribuito in modalità beta silenziosa, eseguito in background e utilizzato per restituire informazioni schematiche prelevate da una base volontaria di clienti più ampia. Per fare in modo che il modello finale abbia la maggiore efficacia possibile, il prototipo segnala ogni istanza di metodo di attacco osservata e ogni istanza sospetta, vale a dire tutti gli eventi che sono appena al di sotto della soglia. L'inclusione degli eventi appena al di sotto della soglia permette ai data scientist di mettere ulteriormente a punto il modello e assicurarsi che non sfugga alcun comportamento. A questo punto il modello viene sottoposto a iterazione, finché i severi standard di qualità non attestano la sua capacità di rilevare i metodi di attacco nel mondo reale.

Gli step finali dello sviluppo delle procedure di rilevamento prevedono la creazione di un'interfaccia utente dedicata che comunica tutti i dati contestuali del metodo di attacco identificato e altre informazioni supplementari sui sistemi in questione, se rilevanti. Il modello viene infine distribuito in produzione, dove diventa operativo e inizia a segnalare gli incidenti ai clienti. La medesima pipeline di prototipi usata per raccogliere i dati viene impiegata per monitorare l'efficacia del modello a livello generale e, se necessario, per apportare ulteriori miglioramenti alla procedura di rilevamento.

L'esito di questo lungo processo è che i modelli non richiedono frequenti messe a punto perché risultano efficaci per contrastare sia gli strumenti di attacco attuali sia quelli di prossima generazione. L'approccio di Vectra basato sul metodo di attacco si distingue per la sua capacità di rilevare le azioni degli attaccanti, e non solo gli eventi inconsueti.

Il motore di streaming in tempo reale che genera dati operativi

La velocità con cui avvengono le operazioni di rilevamento ha un peso perché qualsiasi ritardo offre agli attaccanti l'opportunità di far progredire l'attacco. Anziché utilizzare batch di dati periodici, gli algoritmi di Vectra utilizzano dati in streaming, scelta che consente di scovare istantaneamente gli attaccanti e che assicura un ampio margine di tempo per bloccare l'intrusione.

Anche la scala operativa conta perché l'estensione delle reti aziendali, delle implementazioni cloud e dei servizi SaaS è in costante crescita, il che comporta un aumento corrispondente della mole di dati che Vectra deve analizzare per le operazioni di rilevamento. Il motore di streaming in tempo reale di Vectra supporta grandi corporation internazionali estraendo volumi di dati significativi da utilizzare come una solida base di conoscenze per il lungo termine.

La quantità di dati storici disponibile ha un grosso impatto sull'efficacia degli algoritmi, in particolare di quelli basati sull'apprendimento non supervisionato. L'esecuzione delle procedure di rilevamento con batch periodici limita la quantità di dati che possono essere elaborati in un lasso di tempo ragionevole. La scelta di Vectra di usare dati in streaming consente agli algoritmi di estrarre le informazioni significative dagli eventi e di aggiungerli alle nuove baseline di dati dei modelli. Poiché vengono compilate con dati in streaming, le baseline comprendono mesi di dati, milioni di eventi e generano avvisi di qualità eccellente.

Correlazione delle minacce tramite l'intelligenza artificiale

L'IA di Vectra non serve esclusivamente a identificare i metodi impiegati dai singoli attaccanti ma anche per correlare tra loro le azioni allo scopo di individuare e classificare gli attacchi in corso e definirne la priorità. L'attività di correlazione è indispensabile perché, per raggiungere il loro obiettivo finale, i criminali informatici svolgono sequenze di azioni in domini diversi. Un algoritmo di correlazione dedicato analizza i comportamenti di account e host, reti e cloud per generare una segnalazione chiara che si è verificato un incidente di sicurezza.

I comportamenti rilevati vengono quindi assegnati a riferimenti stabili come account o sistemi host.

Poniamo, ad esempio, che in una rete o in un ambiente cloud ibrido vengano attribuiti IP transitori a host stabili sulla base degli artefatti osservati da un algoritmo che chiameremo host-id. Gli artefatti vengono raccolti dai metadati della rete – come informazioni su entità di protezione host di Kerberos, DHCP indirizzi MAC e cookie – e da integrazioni API quali EDR, vCenter, Azure e AWS. Una volta che a un host sono stati attribuiti degli artefatti, la presenza di un determinato artefatto in un IP genererà l'attribuzione del flusso di metadati e dei comportamenti di attacco associati all'host in questione, e non solo all'IP.

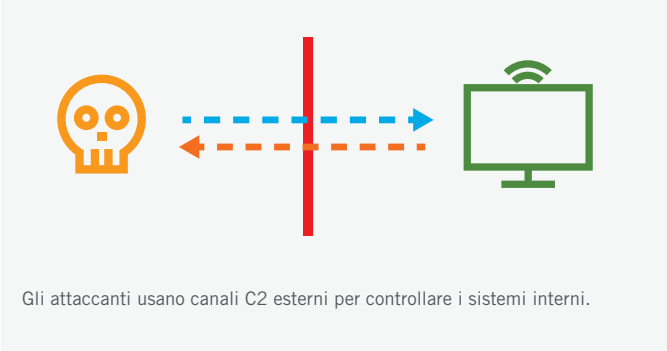
> L'IA di Vectra non serve esclusivamente a identificare i metodi impiegati dai singoli attaccanti ma anche per correlare tra loro le azioni allo scopo di individuare e classificare gli attacchi in corso e definirne la priorità.

In AWS, l'attribuzione è più macchinosa a causa del modo con cui gli eventi vengono registrati nel piano di controllo AWS. Essi vengono infatti associati ai ruoli assunti e non agli account utente sottostanti. Conoscere l'effettivo utente IAM o SAML che ha assunto il ruolo è fondamentale per rispondere a un attacco, tuttavia un numero imprecisato di account può assumere uno stesso ruolo.

Gli attaccanti più sofisticati riescono a complicare ulteriormente il lavoro degli addetti alla sicurezza concatenando i ruoli nel tentativo di occultare l'origine dell'attacco. Ricorrendo a una speciale tecnologia denominata Kingpin, Vectra riesce a risalire la catena dei ruoli e ad attribuire gli attacchi osservati all'utente sottostante, e non semplicemente a un ruolo.

Una volta che i comportamenti dell'attaccante sono stati attribuiti a un indicatore stabile, possono essere studiati insieme e utilizzati per individuare il profilo comportamentale del sistema sottostante, passaggio che consente quindi di etichettare e classificare le minacce attive in base alla gravità. L'algoritmo di correlazione è stato progettato per replicare le azioni di analisi delle minacce che svolgono gli analisti e i ricercatori di Vectra e riesce a scoprire gli scenari di attacco avanzati che richiedono un intervento immediato in quanto attribuibili ad attaccanti esterni o a minacce interne di livello amministrativo.

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 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.

I tool di comando e controllo sono presenti in framework di attacco preconfezionati come Cobalt Strike e Metasploit ma possono anche essere sviluppati internamente dai gruppi di attacco più avanzati. In ogni caso, tutti questi framework supportano la criptazione del canale, il domain fronting o altre misure elusive che consentono all'attaccante di passare inosservato.

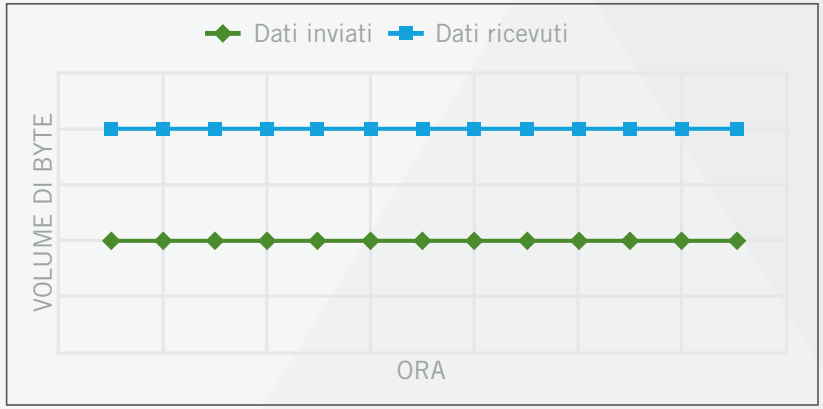
Vectra rileva i canali C2 a prescindere che siano criptati o che siano state impiegate altre tecniche di elusione.

Metodologia di rilevamento

Vectra rileva i canali C2 a prescindere che siano criptati o che siano state impiegate altre tecniche di elusione. Deriva questa capacità dall'approccio basato sul metodo di attacco descritto in precedenza (mentre l'approccio basato sui dati matematici non offre gli stessi risultati).

Quando gli esperti di sicurezza di Vectra hanno estratto le caratteristiche essenziali del comportamento di un canale C2, si sono accorti che gli indicatori più determinanti del metodo non erano gli aspetti circostanziali del traffico, come la presenza di domini rari o di agent, ma piuttosto la rappresentazione grafica del traffico di rete nel tempo.

Consideriamo la seguente rappresentazione del traffico innocuo proveniente da un sistema esterno.

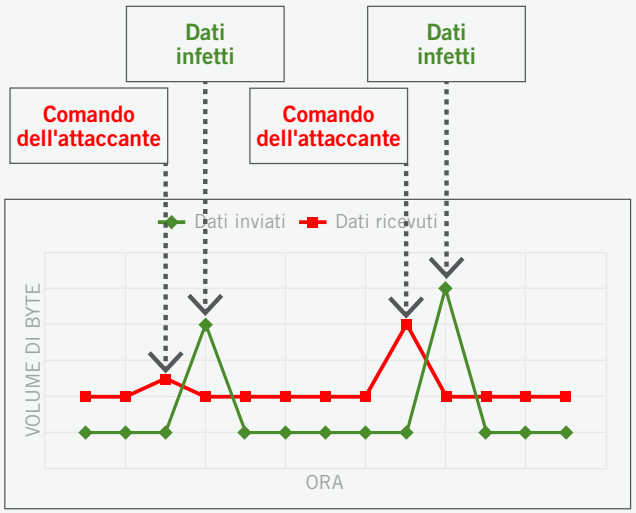


Traffico innocuo

Esempio di traffico dati con beacon innocuo.

Questo esempio illustra il traffico prodotto da un host che comunica tramite beacon con un server esterno. I beacon sono funzioni di rete molto comuni utilizzate da servizi come chat, stock ticker e ad tracker per mantenere sincronizzati i sistemi locali e remoti e permettere loro di comunicare. Questa stessa funzionalità viene purtroppo utilizzata anche dai canali C2 nocivi.

Esiste però una sottile differenza tra l'aspetto di un beacon usato da uno stock ticker e quello di un beacon usato da un canale nocivo. Questa è la rappresentazione grafica di un tunnel criptato nocivo.

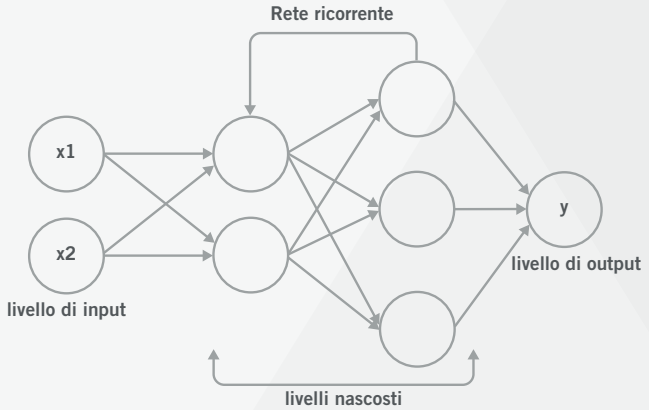


Esempio di traffico dati con comando e controllo nocivo.

La presenza di picchi regolari è evidente. Si verificano quando l'attaccante invia un comando al quale il sistema infetto risponde. Il primo picco di dati giunge spontaneamente ed è seguito a breve dalla risposta del sistema infetto.

Studiando schemi di questo tipo, i data scientist di Vectra sono riusciti a elaborare una strategia ottimale per individuare questo comportamento. Poiché i dati temporali che caratterizzano il comportamento del canale C2 sono molto simili ai dati del riconoscimento vocale e dell'elaborazione del linguaggio naturale, il team Vectra ha deciso di implementare un modello di deep learning.

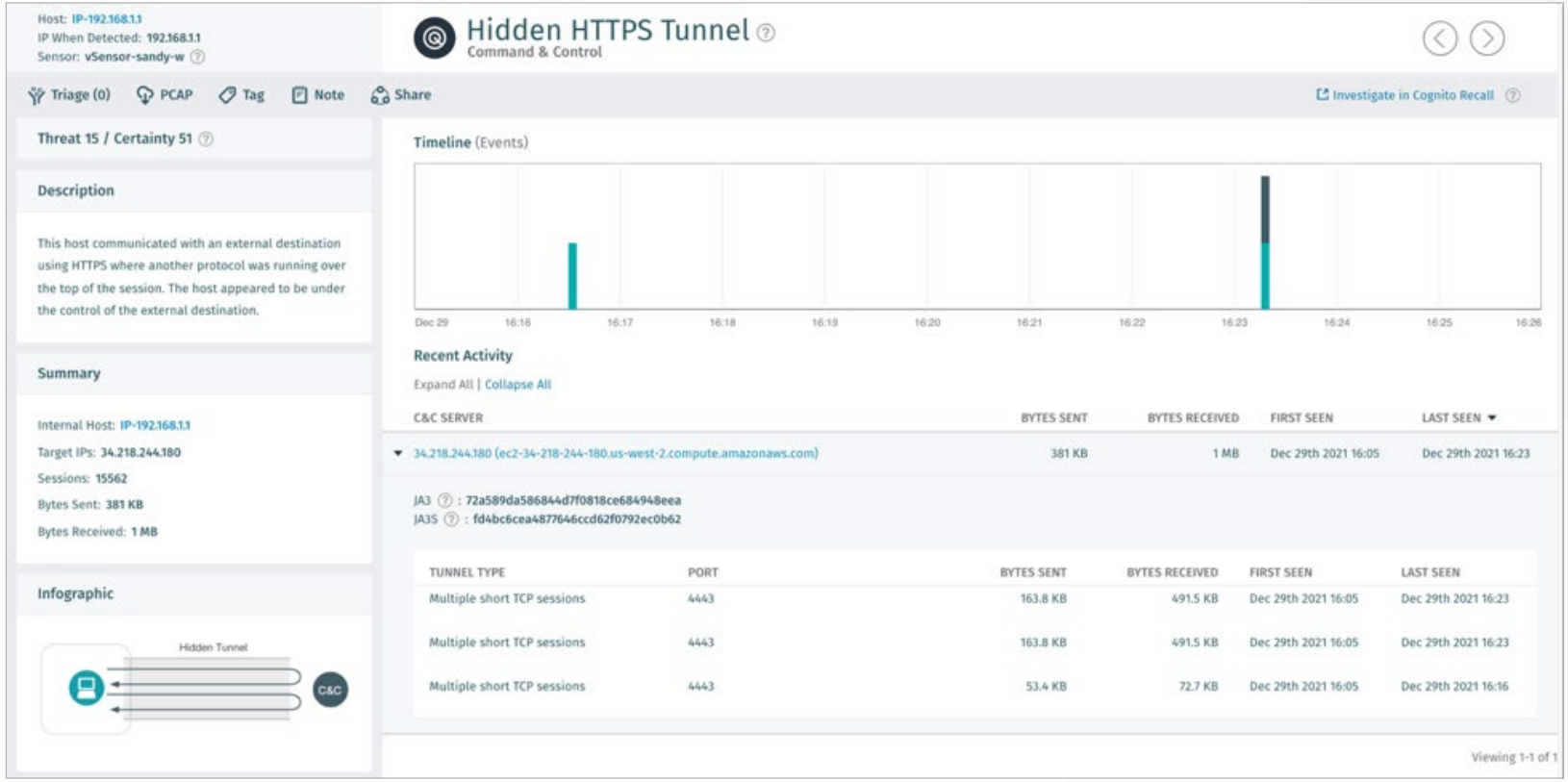
Pertanto, per identificare i comportamenti di attacco Vectra utilizza una rete neurale ricorrente specifica denominata LSTM (Long Short-Term Memory). Questo tipo di algoritmo è particolarmente abile nel riconoscere eventi con molte tempistiche differenti, e questa proprietà è critica per capire a fondo la natura dei dati trasmessi dal canale C2. L'allenamento dell'algoritmo LSTM avviene con dati reali e campioni generati da altri algoritmi. Il set di dati copre un'ampia gamma di scenari, tool, configurazioni e ambienti affinché il modello impari a identificare il segnale che indica la presenza di un canale di controllo a prescindere dal tool utilizzato per l'attacco.



Vectra usa reti neurali ricorrenti per distinguere le comunicazioni C2 nocive dalle comunicazioni innocue.

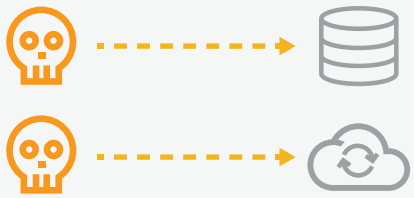
È importante sottolineare anche che questo processo basato su algoritmi è reso possibile dal modo con cui Vectra formatta i dati delle sessioni di rete. Vectra è in grado di produrre metadati di tipo Zeek, tuttavia il suo parser personalizzato fornisce metadati con un livello di affidabilità superiore a quelli standard di Zeek perché analizza le comunicazioni di rete con una frequenza inferiore al secondo. Questa visibilità granulare su tutte le comunicazioni innocue e nocive consente ai team di data science di Vectra di usare algoritmi che assicurano la migliore copertura possibile per un ampio ventaglio di problemi.

Questa speciale strategia basata sull'impiego di metadati unici e algoritmi sofisticati si traduce in una straordinaria capacità di scovare gli attaccanti. La scelta di concentrarsi sui dati delle comunicazioni anziché sui segnali superficiali consente di individuare anche il traffico criptato e garantisce resilienza nel tempo perché le eventuali modifiche apportate ai tool non avranno impatti. Il segnale inequivocabile della presenza di un dato comportamento elimina inoltre la necessità di implementare dei filtri di soppressione, che potrebbero potenzialmente far passare inosservati i canali usati per il fronting o le azioni furtive perpetrate dagli attaccanti.



Vectra rileva un canale C2 criptato.

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

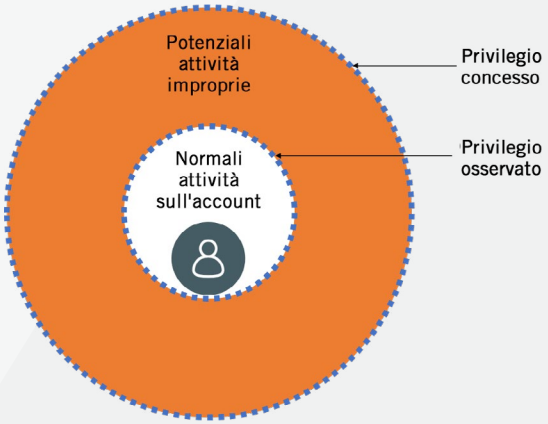
Gli attaccanti che riescono a procurarsi credenziali con privilegi si assicurano un accesso incontrastato a risorse di rete e cloud. Le credenziali sono un vettore di accesso "pulito" che consente di non utilizzare payload con exploit e malware, cioè strumenti che lasciano tracce e generano avvisi. L'imposizione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

Quando si verifica un furto di credenziali, prevenirne un uso improprio è complicato. Riuscire a rilevare la presenza di un attaccante che assume il controllo di un account dopo aver rubato delle credenziali di accesso presenta numerose difficoltà. Ogni azione che l'attaccante svolge è esplicitamente consentita dalle autorizzazioni di cui dispone. La definizione di avvisi che segnalano interazioni nuove o recenti non avrà alcuna efficacia perché tutti gli utenti operano in ambienti dinamici dove l'accesso a nuove risorse è parte integrante delle normali attività quotidiane. Un attaccante che è riuscito a violare un ambiente tenterà di passare inosservato eseguendo operazioni generiche non dirette a un account in particolare per evitare di destare sospetti. Per riuscire a individuare in maniera efficace un abuso di credenziali con privilegi serve un approccio basato sul metodo di attacco perché è l'unico che consente di risalire alle intenzioni dell'attaccante.

Metodologia di rilevamento

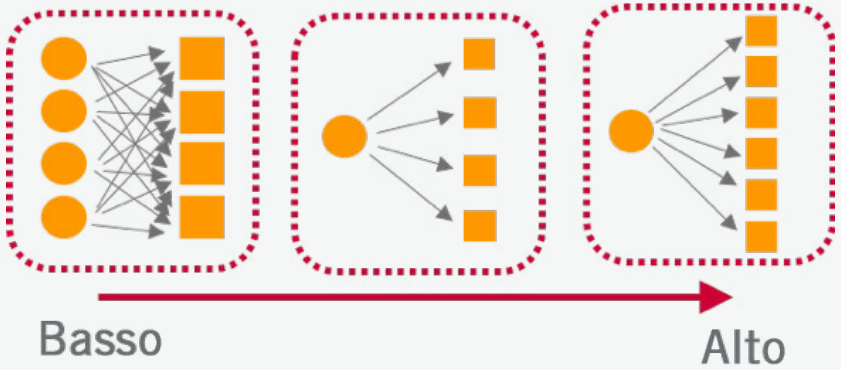
Vectra è in grado di individuare gli abusi di credenziali rubate sia negli ambienti di rete sia negli ambienti cloud. Un caratteristica fondamentale dell'approccio basato sul metodo di attacco è la conoscenza dell'impiego che fanno gli attaccanti delle credenziali rubate. Le credenziali con privilegi sono preziose per i criminali informatici perché consentono di accedere a servizi e funzionalità protette.

Gli esperti di sicurezza di Vectra sono arrivati alla conclusione che se conoscessimo il livello di privilegio di ogni account, host, servizio e cloud potremmo stilare una mappa di tutte le risorse più preziose che esistono. Il concetto del *privilegio concesso* è ben noto, ma questa rappresentazione permette di aumentare l'ambito di impiego effettivo del privilegio rispetto al privilegio minimo necessario. Gli esperti di sicurezza e i data scientist di Vectra hanno escogitato un modo nuovo di rappresentare il valore dei sistemi di un ambiente basandosi su quanto osservato nel tempo. Il valore dinamico di base è definito *privilegio osservato* ed equivale a un uso delle credenziali con approccio zero-trust che non richiede alcuna configurazione manuale.



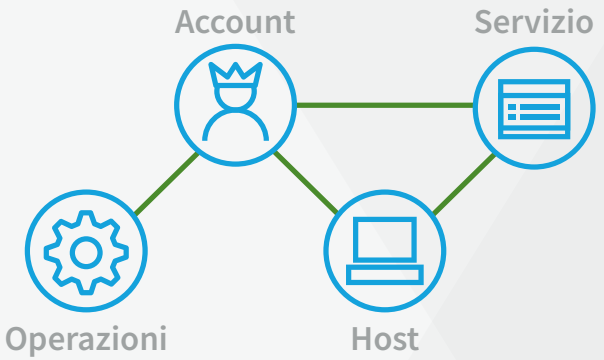
Il privilegio osservato è una versione zero-trust dei normali privilegi che servono a un utente per svolgere il proprio lavoro. L'uso di privilegi superiori a quelli necessari richiede un approfondimento.

Per calcolare il *privilegio osservato*, l'IA di Vectra non si basa sul privilegio definito dall'IT ma considera le interazioni storiche tra le entità monitorate. La varietà e la specificità degli accessi e degli utilizzi contribuiscono fortemente ai punteggi di privilegio. Un sistema che accede sistematicamente ad altri sistemi, a loro volta collegati regolarmente ad altri, avrà un punteggio di privilegio basso mentre un sistema che accede a molti sistemi a cui non se ne collegano altri avrà un punteggio di privilegio elevato. In questo modo Vectra riesce a riconoscere gli account degli amministratori di dominio e gli account degli utenti normali.



Vectra impara i livelli di privilegio osservato basandosi sui comportamenti degli utenti. Un account che accede a molti servizi comuni ha un livello di privilegio inferiore rispetto a uno che accede a servizi più esclusivi.

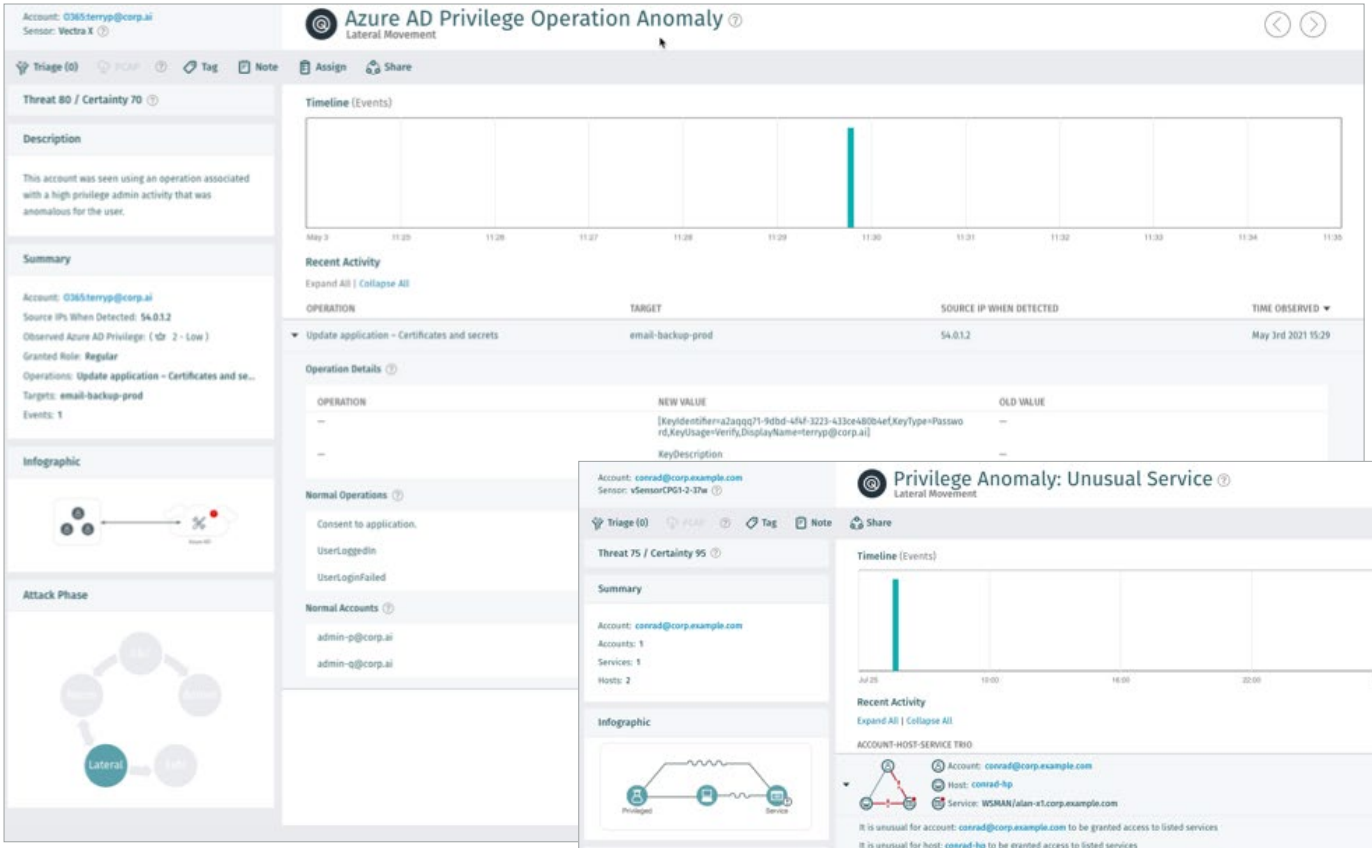
Una volta calcolato il punteggio del privilegio osservato, si procede alla mappatura di tutte le interazioni tra account, servizi, host e cloud per comprendere le normali interazioni storiche tra i sistemi. Quindi, un pacchetto di algoritmi non supervisionati – che include algoritmi personalizzati per il rilevamento di anomalie e implementazioni HDBSCAN – verifica i punteggi e identifica i casi anomali di uso improprio dei privilegi.

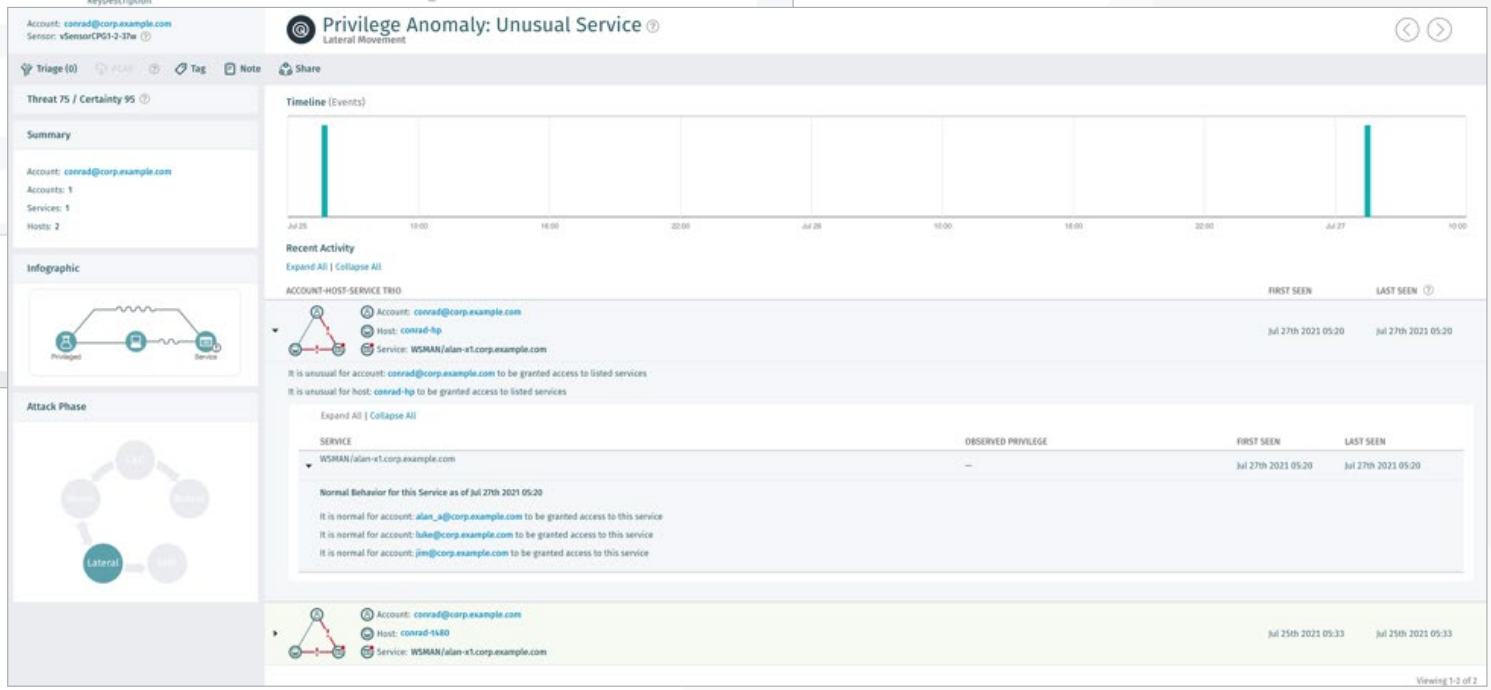


Per individuare l'abuso di credenziali Vectra applica l'apprendimento non supervisionato considerando il privilegio osservato e le interazioni tra account, host, servizi e cloud.

Questo sofisticato approccio basato sul metodo di attacco consente di individuare le credenziali rubate di cui è stato fatto un uso improprio nel cloud e nelle reti on-premise. La metrica del *privilegio osservato* attira l'attenzione delle procedure di rilevamento sulle principali azioni anomale e assicura un livello di precisione e richiamo superiori.

Per calcolare il *privilegio osservato*, l'IA di Vectra non si basa sul privilegio definito dall'IT ma considera le interazioni storiche tra le entità monitorate.





Vectra rileva gli account che fanno un uso improprio dei privilegi.

Conclusioni

Gli attaccanti diventano sempre più sofisticati, e gli addetti alla sicurezza devono evolvere di pari passo. Nel corso degli anni Vectra non ha mai smesso di introdurre innovazioni per rendere più efficace possibile la sua piattaforma di rilevamento e neutralizzazione delle minacce rivolte alle risorse on-premise e cloud.

Vectra ha sviluppato oltre un centinaio di procedure di rilevamento basate sull'IA e sul principio del metodo di attacco ed è riuscita a identificare... Ogni procedura di rilevamento è stata creata pensando a come agiscono gli attaccanti e utilizzando alcune delle tecniche di ML più avanzate disponibili. Vectra ha ottenuto ben 33 brevetti per la tecnologia che utilizza in queste procedure di rilevamento.

Oltre alla copertura assicurata dalla tecnologia brevettata di Vectra, siamo orgogliosi anche di essere l'azienda più citata dalla NSA e nel framework MITRE D3FEND, lo schema di MITRE che definisce le contromisure che devono adottare gli addetti alla sicurezza per proteggere gli ambienti informatici. Il framework D3FEND spiega come bloccare gli attacchi e contrastare le tecniche degli attaccanti definite nel framework MITRE ATT&CK. Il framework D3FEND cita ben 12 misure difensive brevettate da Vectra.

In Vectra lavoriamo per rendere il mondo più giusto e più sicuro. Per questo motivo, continueremo a usare il principio del metodo di attacco e l'intelligenza artificiale per impedire agli attaccanti di raggiungere i loro scopi.

Per maggiori informazioni, contatta info@vectra.ai.

E-mail: info@vectra.ai vectra.ai