VECTRA
SECURITY THAT THINKS

DOCUMENTO TÉCNICO

La IA que hay tras las soluciones de Vectra

ÍNDICE

Vectra® protege a las empresas mediante la detección y neutralización de los ciberataques.

Vectra® es líder en detección de amenazas y respuesta, para empresas híbridas y con entornos multinube. La plataforma de Vectra utiliza la IA para detectar rápidamente las amenazas en nubes públicas, identidades, aplicaciones SaaS y centros de datos. Solo Vectra optimiza la IA para detectar las técnicas de ataque (las TTP que emplean los atacantes), en lugar de alertar simplemente cuando descubre algo inusual. Dada su señal de amenaza de gran fiabilidad y el contexto claro que proporciona, los equipos de seguridad pueden responder a las amenazas en menos tiempo y detener más rápido los ataques que ya se hayan iniciado. Numerosas organizaciones de todo el mundo confían en Vectra para ganar en resiliencia ante las ciberamenazas y evitar que el ransomware, los ataques contra las cadenas de suministro, la usurpación de identidades y otros tipos de ciberataques afecten a su actividad. Para obtener más información, visite vectra.ai.

Introducción

La ciencia de datos es la brújula que guía a la inteligencia artificial de Vectra. Siempre hemos estado convencidos de que, bien empleada, la combinación de ambas disciplinas puede revolucionar la lucha contra los ciberataques y dar ventaja a los responsables de seguridad. Sin embargo, no todas las IA son iguales. En este documento, analizaremos qué es la inteligencia artificial (IA) y definiremos los términos clave relacionados con las soluciones de IA. Además, describiremos las dos principales metodologías con las que aplicar la IA a la detección de amenazas y descubriremos cómo detecta Vectra las amenazas con la IA.

Este documento es apto para todos; tanto para los escépticos que dudan de la IA como para aquellos que están profundamente fascinados por su potencial.

¿Qué es la inteligencia artificial?

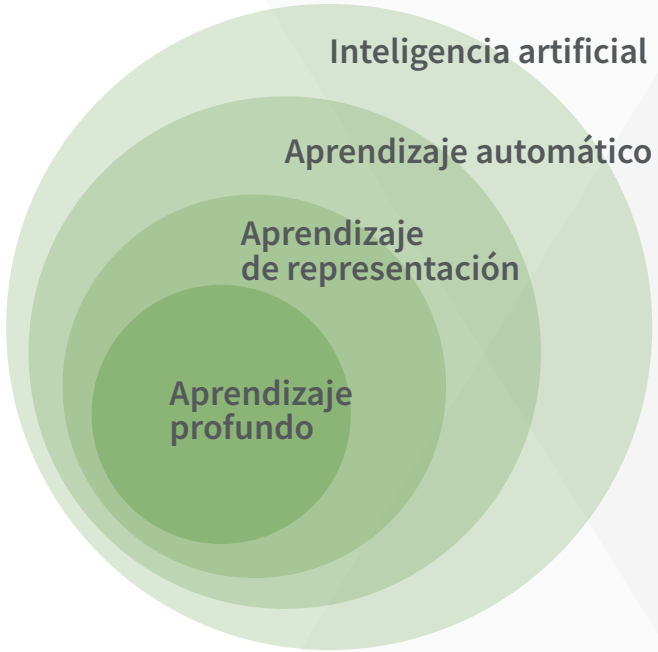
Definición de la inteligencia artificial

Se ha extendido la idea errónea de que los términos "inteligencia artificial", "aprendizaje automático" y "aprendizaje profundo" se refirieren a la misma disciplina o que actúan del mismo modo, pero no es así. Es cierto que están relacionados, pero cada uno tiene un significado propio y específico. Si conocemos el ámbito al que cada uno de estos términos se refiere, comprenderemos mejor qué hacen las herramientas que recurren a la IA.

Inteligencia Artificial (IA): se define como cualquier sistema que puede automatizar el razonamiento y aproximarse al funcionamiento de la mente humana. Es un término amplio que abarca las subdisciplinas del aprendizaje automático, el aprendizaje de representación y el aprendizaje profundo. El término "IA" se aplica a un sistema que se basa en el uso de reglas que se programan de manera expresa, así como al que ha aprendido de forma autónoma a partir de una gran cantidad de datos. Este último tipo es el que fundamenta tecnologías como las de los vehículos autónomos y los asistentes virtuales, y entra en la subdisciplina del aprendizaje automático.

Aprendizaje automático: el aprendizaje automático es una subdisciplina de la IA en la que las acciones del sistema no las dicta explícitamente un humano, sino que se aprenden a partir de datos. Estos sistemas pueden procesar desde unas decenas de puntos de datos hasta miles de millones de ellos con el fin de aprender a representar de la mejor forma posible nuevas instancias de datos y posteriormente responder a ellas.

Aprendizaje de representación: aunque no se suele hablar de él, el aprendizaje de representación es crucial en muchas de las tecnologías de IA que se emplean actualmente. Esta subdisciplina se fundamenta en el aprendizaje de nuevas representaciones abstractas a partir de datos; por ejemplo, transformar imágenes de distintos tamaños en una lista de números con una longitud uniforme que represente las imágenes originales. Esta capacidad de abstracción permite sobre todo que los sistemas que se fundamentan en este aprendizaje puedan trabajar mejor con nuevos tipos de datos.



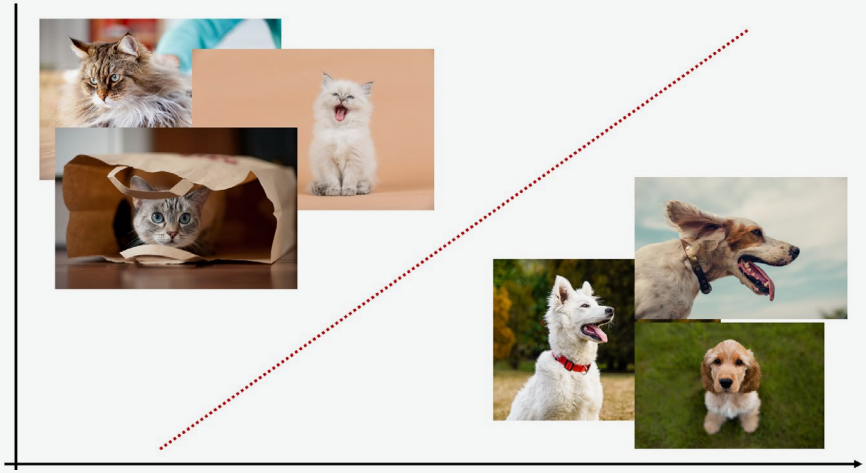
Relación entre las diferentes subdisciplinas de IA.
Referencia: "Deep Learning", Goodfellow, Bengio & Courville (2016)

Aprendizaje profundo: a menudo se asocia a las redes neuronales y se basa en el aprendizaje automático y el aprendizaje de representación, ambas subdisciplinas más amplias, para descubrir entre los datos jerarquías de abstracciones que representan entradas de datos cada vez con más complejidad. Inspirándose en el cerebro humano, los modelos de aprendizaje profundo utilizan capas de neuronas cuyos pesos sinápticos se adaptan en respuesta a las entradas de datos. En ellos se disponen capas de red más profundas que aprenden nuevas representaciones abstractas para hacer más sencillas tareas como clasificar imágenes o traducir textos. Si bien el aprendizaje profundo puede emplearse como técnica eficaz para resolver ciertos problemas complejos, no es en absoluto la solución definitiva para automatizar la inteligencia.

Tipos de técnicas de aprendizaje mediante algoritmos

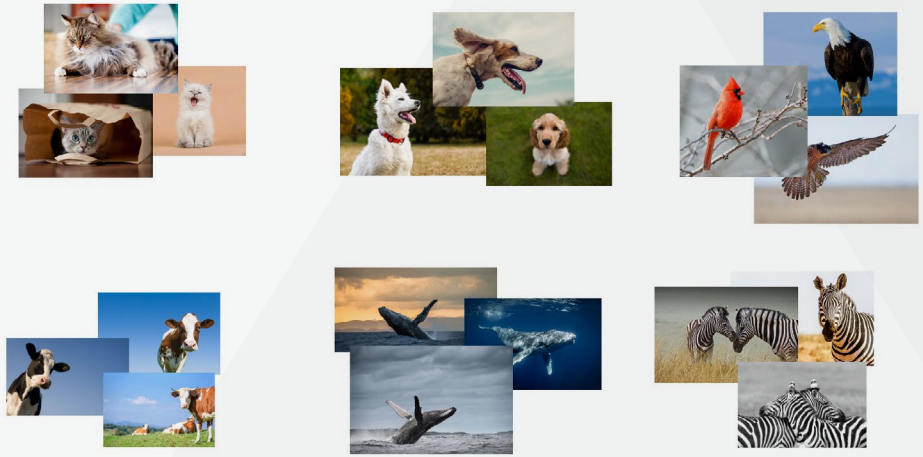
Una de las principales opciones que ofrecen los algoritmos de aprendizaje automático es clasificar instancias de datos por clases. Aunque esta capacidad comprende varias categorías de aprendizaje amplias, las más habituales son **con supervisión** y **sin supervisión**.

El aprendizaje **con supervisión** es aquel en el que el modelo aprende de un conjunto de datos etiquetados. Una vez aprendidos estos datos, cuando se le presentan otros nuevos, el modelo puede asignarles una etiqueta. Pensemos en este ejemplo: si suministramos a un modelo de aprendizaje supervisado un gran número de imágenes de perros y gatos y más tarde le mostramos una nueva imagen de un perro o un gato, será capaz de discernir qué animal es. El aprendizaje supervisado requiere un gran corpus de datos de entrenamiento etiquetados de los que el modelo pueda aprender, pero una vez entrenado, puede ser muy eficaz a la hora de generalizar y etiquetar correctamente nuevas instancias de datos.



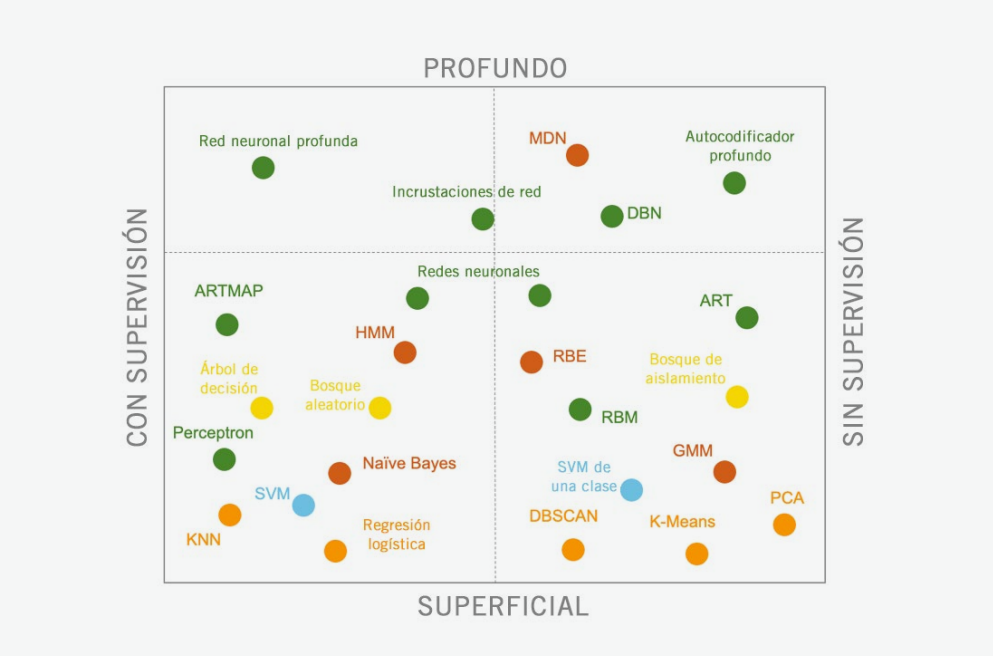
El aprendizaje con supervisión utiliza datos etiquetados para identificar los factores que distinguen las etiquetas. Los modelos que usan bien este tipo de aprendizaje pueden etiquetar datos nuevos.

El aprendizaje **sin supervisión** es aquel en el que el modelo aprende de un conjunto de datos sin etiquetar. Estos modelos aprenden estructuras a partir de los datos que se les suministran, y posteriormente pueden determinar si un dato nuevo encaja en la estructura aprendida y de qué manera. La ventaja de los modelos de aprendizaje sin supervisión es que no requieren entrenamiento inicial. Este modelo destaca por su capacidad para discriminar puntos de datos, sin embargo, no es capaz de asignar fácilmente etiquetas a las anomalías o los valores atípicos.



El aprendizaje sin supervisión aprende la estructura subyacente de los datos sin etiquetar. Los modelos que recurren a este tipo de aprendizaje pueden determinar en qué medida encajan los nuevos datos en una estructura aprendida.

Dentro de estos dos amplios modelos hay numerosos algoritmos de aprendizaje diferentes (en la imagen de abajo), y los investigadores siguen creando algoritmos nuevos. Por si fuera poco, los algoritmos pueden combinarse para formar sistemas aún más complejos. La pregunta entonces es: ¿cómo elige un científico de datos el algoritmo o los algoritmos idóneos para resolver un problema concreto? ¿Acaso hay un algoritmo mejor a todos los demás para cualquier tipo de problema?



Existen numerosos algoritmos de aprendizaje automático, y cada uno presenta puntos fuertes y débiles según el tipo de problema al que se enfrente.

Teorema No free lunch

Resulta que no hay ningún algoritmo que sea mejor que todos los demás ante cualquier tipo de problema. Es lo que se conoce como teorema "No free lunch". Dicho de otro modo: ante un determinado problema, siempre habrá un algoritmo especializado que logre mejores resultados que un algoritmo genérico. La diversidad de algoritmos es cada vez mayor, ya que para abordar problemas concretos hacen falta algoritmos específicos. Hay problemas en los que se necesitará una red neuronal supervisada y otros en los que la mejor opción será una agrupación jerárquica no supervisada.

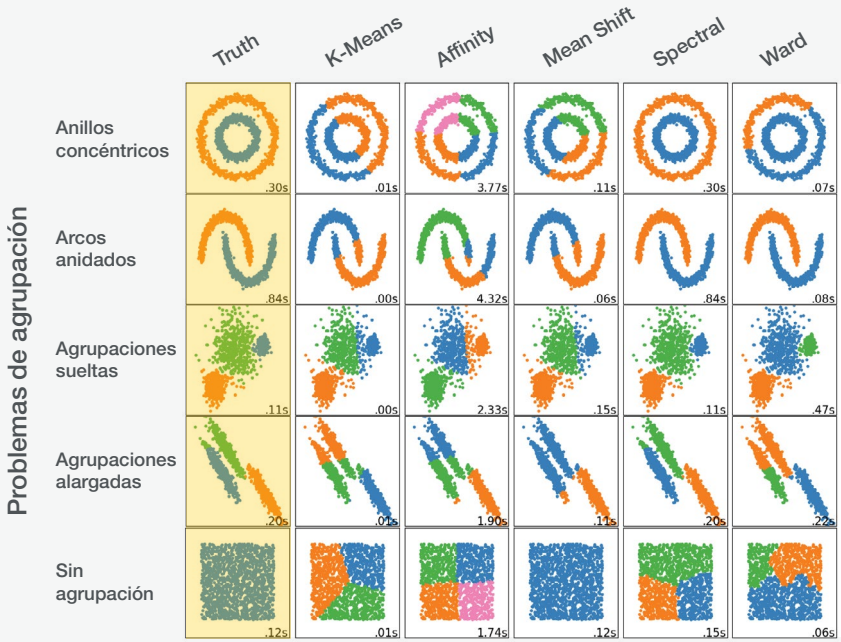
Por ejemplo, el algoritmo que se emplea para el reconocimiento de imágenes en los vehículos autónomos no puede aplicarse a la traducción de idiomas. Cada algoritmo es único y está optimizado para resolver un problema concreto con los datos en los que se basa el modelo.

Teorema No free lunch: no existe ningún algoritmo capaz de solucionar todo tipo de problema.

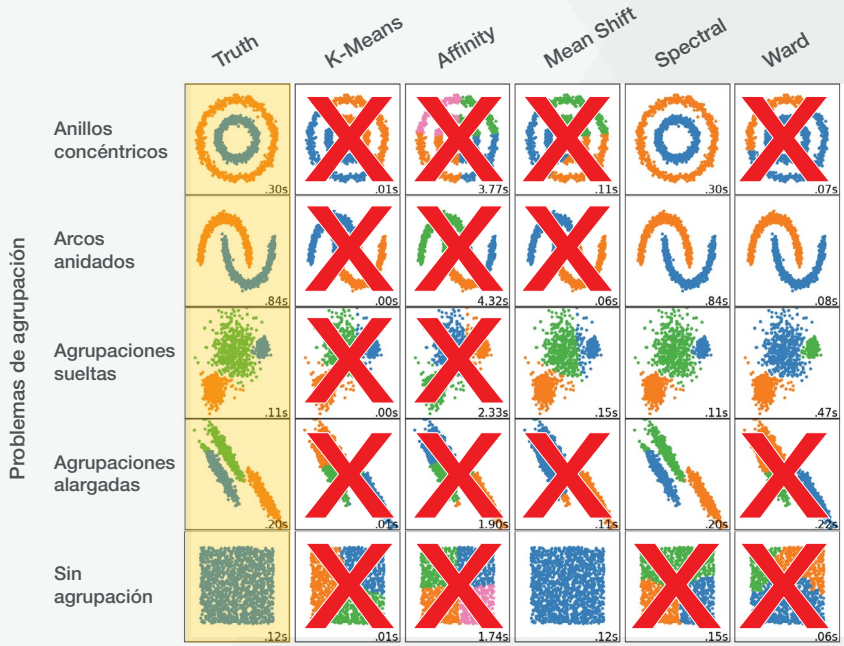
Encontrar la herramienta adecuada para cada tarea

Entonces, ¿cómo elige un científico de datos el algoritmo adecuado? Esta cuestión tiene tanto de arte como de ciencia. Para que un científico de datos se encamine en la dirección correcta, debe comprender perfectamente tanto el planteamiento del problema como los datos. Es importante tener en cuenta que si no se hace bien, los resultados no serán óptimos o peor aún: pueden ser completamente erróneos.

Pensemos en el ejemplo de abajo. Según el algoritmo que elijamos para cada conjunto de datos, obtendremos resultados muy diferentes. Para cada problema hay un algoritmo ideal, pero lo más importante es tener en cuenta que algunas opciones pueden arrojar resultados muy poco deseables. Por eso es tan importante elegir bien el enfoque correcto para cada problema.



Comparación de los resultados de los algoritmos de aprendizaje automático (eje X) con diferentes conjuntos de datos (eje Y). Las etiquetas reales se destacan en amarillo. Adaptado de scikit-learn.org.



Comparación de resultados. Se tachan con una X las predicciones erróneas que podrían conllevar resultados no deseados. Ningún algoritmo es eficaz con todos los conjuntos de datos. Adaptado de scikit-learn.org.

¿Cómo medir si un modelo funciona *bien*?

Para elegir el modelo correcto, los científicos deben determinar cómo medir si efectivamente funciona bien. Cuando se habla del rendimiento de un modelo, se suele hacer referencia a su *exactitud*.

$$Exactitud = \frac{(Positivos\ verdaderos + Negativos\ verdaderos)}{(Positivos\ verdaderos + Negativos\ verdaderos + Positivos\ falsos + Negativos\ falsoss)}$$

Como métrica, la exactitud aporta cierto valor, sin embargo, aun cuando sea aparentemente buena puede ocultar otros datos respecto al rendimiento real de un modelo. Imaginemos que nos encontramos ante un problema de clasificación en el que el objetivo es asignar a los datos una etiqueta A o una etiqueta B. Si la probabilidad de que se asigne la etiqueta A es mil veces mayor que de que se asigne la etiqueta B y etiquetamos los datos siempre con la etiqueta A, alcanzaremos fácilmente un nivel de exactitud del 99,9 %. Aunque el dato sea excelente, nunca podremos etiquetar nada correctamente como B. Está claro que la exactitud no es la métrica adecuada para detectar los casos de la etiqueta B. Afortunadamente, los científicos de datos disponen de otras métricas con las que optimizar y medir la eficacia de un modelo en los casos que les ocupan.

La precisión es otra de esas métricas. Mide el grado de acierto a la hora de asignar una etiqueta concreta en relación con el número total de conjeturas que hace el modelo de esa etiqueta.

$$Precisión = \frac{Positivos\ verdaderos}{(Positivos\ verdaderos + Positivos\ falsos)}$$

Los científicos de datos que quieran obtener un alto nivel de precisión crearán modelos que asignen las etiquetas sin generar demasiadas falsas alarmas. Lo que la precisión no nos dice es si el modelo falló al etiquetar los casos que nos interesan. Por suerte, hay otra métrica que aporta perspectiva a toda esta cuestión: la exhaustividad.

La exhaustividad mide la frecuencia con la que un modelo acierta a la hora de asignar una etiqueta concreta en relación con todas las instancias de esa etiqueta.

$$Exhaustividad = \frac{Positivos\ verdaderos}{(Positivos\ verdaderos + Negativos\ falsos)}$$

Los científicos de datos que quieran obtener un alto nivel de exhaustividad crearán modelos que no fallen al alertar de las instancias que les ocupan.

Controlar y equilibrar la precisión y la exhaustividad permite a los científicos de datos medir y optimizar eficazmente sus modelos para obtener buenos resultados.

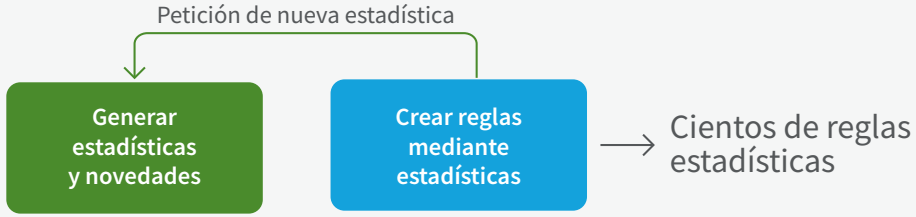
Para elegir el modelo correcto, los científicos deben determinar cómo medir si efectivamente funciona bien.

Aplicación de la IA a la detección de amenazas

La IA y todas sus disciplinas desempeñan una función importante en las empresas actuales a la hora de detectar y detener a los atacantes. Actualmente existen dos paradigmas de identificación activa de las ciberamenazas: uno basado en las matemáticas y otro basado en la seguridad. En esta sección, estudiaremos las diferencias entre ambos paradigmas y explicaremos por qué la IA basada en la seguridad es la mejor opción para los equipos de seguridad.

IA basada en las matemáticas: un mal enfoque para la detección de amenazas

En el paradigma basado en las matemáticas, los científicos de datos generan conjuntos simples de estadísticas usando pocos algoritmos genéricos que se centran en la detección de valores nuevos o atípicos. Más tarde, los investigadores de seguridad combinan esas estadísticas para crear cientos de reglas. Si hace falta una nueva estadística, se sigue el mismo enfoque genérico para crearla. Estas reglas estadísticas suelen mejorarse con filtros de supresión explícitos en la fase de posprocesamiento para abordar los volúmenes de detección adicionales que se originan con este tipo de enfoque genérico (recordemos que según el teorema No free lunch, la aplicación de algoritmos genéricos ofrece un rendimiento que no alcanza el nivel óptimo).



Paradigma para desarrollar métodos de detección, basado en las matemáticas.

A modo de ejemplo, vamos a tratar de detectar un canal de mando y control (C2). El equipo de científicos de datos empieza por generar una estadística relativa a la prevalencia de todos los dominios externos. El equipo de investigadores de seguridad deberá determinar entonces qué umbral de anomalía implicará que se ha detectado un canal C2. Si muchos de los dominios que utilizan los dispositivos IoT superan este umbral de anomalía, habría que aplicar un filtro de supresión para ignorar todos esos dispositivos. Se aplicarían filtros de supresión adicionales a agentes de usuario, subredes y otros atributos hasta tener un volumen manejable de alertas. Estas reglas de supresión son necesarias debido a la naturaleza genérica de este enfoque, aunque es cierto que plantean el riesgo de pasar por alto una técnica de evasión utilizada por un atacante.

IA basada en la seguridad: un enfoque con la máxima cobertura y el mínimo ruido para la detección de amenazas

El paradigma basado en la seguridad aúna la definición del problema (método del atacante) y la búsqueda del modelo adecuado. Los investigadores de seguridad definen el planteamiento del problema identificando un método amplio de ataque (no solo una herramienta o una única vulnerabilidad). Por otro lado, los científicos de datos buscan el algoritmo adecuado para identificar ese método, y finalmente ambos equipos colaboran estrechamente para dar con la solución mediante iteración. Este enfoque detecta directamente el método del atacante, y no se limita a identificar las anomalías superficiales que a menudo se advierten en los enfoques basados en las matemáticas.

El enfoque basado en la seguridad ofrece un mejor rendimiento, medido por la exhaustividad y la precisión. Aparte de estas métricas, resiste a los cambios en las herramientas de ataque y emplea menos tipos de detección, lo que facilita las operaciones para los equipos de seguridad. Cuando un nuevo método de ataque comienza a generalizarse, se pone en marcha el proceso basado en la seguridad y se desarrolla un nuevo método de detección. Aunque, debido a lo sofisticado del enfoque, es posible que se requiera más tiempo de desarrollo, las técnicas de los atacantes no cambian muy rápido y aparecen normalmente junto a otras más antiguas para las que ya existen soluciones.

Paradigma para desarrollar métodos de detección, basado en la seguridad.

Cómo funciona Vectra

Vectra ha sido pionera en el enfoque basado en la seguridad para detectar métodos de ataque en la red, la nube pública, las aplicaciones SaaS y las identidades. En las secciones siguientes, veremos hasta dónde llegan el proceso de desarrollo y la cobertura de Vectra, conoceremos el motor que recopila y genera detecciones, descubriremos cómo se correlacionan eventos concretos con incidentes de seguridad de utilidad práctica y descubriremos los entresijos de dos detecciones de Vectra.

Desarrollo de las detecciones en Vectra

En Vectra, las detecciones se centran en localizar a los atacantes e identificar sus métodos de ataque, no solo las anomalías. La cobertura la desarrollan investigadores de seguridad con muy variada experiencia y científicos de datos con profundos conocimientos acerca de cómo extraer valor de conjuntos de datos enormes y complejos. En los últimos 10 años o más, ambos grupos han ideado un enfoque de estrecha colaboración con el que desarrollar métodos de detección de amenazas comunes a todos los dominios de seguridad y tipos de datos para identificar con eficacia los comportamientos de los atacantes con el menor ruido posible.

El equipo de investigación de seguridad de Vectra lidera todo el proceso de desarrollo de detecciones, en el que constantemente supervisa y revisa las técnicas de ataque que hay en circulación. La investigación no se centra en herramientas o grupos de ataque concretos, sino en los métodos generales que emplean los atacantes. Por ejemplo, si los investigadores de seguridad observan que la funcionalidad de balizas de Cobalt Strike se utiliza en ataques de ransomware.

> Vectra ha sido pionera en el enfoque basado en la seguridad para detectar métodos de ataque en la red, la nube pública, las aplicaciones SaaS y las identidades.

¿Cómo desarrolla Vectra los algoritmos de IA basados en la seguridad?

En lugar de fijarse solo en las balizas de Cobalt Strike, abstraen las acciones de esta tecnología y estudian el método de *control* del atacante. Gracias a la técnica de abstracción, Vectra puede proteger a cualquier organización frente a las herramientas que se sabe que emplean este método actualmente, así como a las que se desarrollarán en el futuro.

Una vez que los investigadores de seguridad han identificado el método de ataque, pasan a crear un corpus de muestras maliciosas y benignas. Las muestras maliciosas proceden de diversas fuentes: clientes que comparten voluntariamente metadatos anonimizados, algoritmos de creación de datos sintéticos, ciberincidentes en la red documentados públicamente o ataques en nuestros laboratorios internos. Las muestras benignas se obtienen del enorme conjunto de datos que Vectra mantiene con metadatos de clientes anonimizados.

Una vez que los investigadores de seguridad conocen el método de ataque y los datos correspondientes, trabajan con el equipo de científicos de datos para desarrollar un modelo prototipo con un umbral óptimo para detectar el método del atacante. El prototipo se implementa en un modo beta sigiloso en el que se ejecuta en segundo plano y genera resúmenes de una base de clientes más amplia que participa voluntariamente en la prueba. Para que el modelo final sea lo más eficaz posible, el prototipo informa de todas las instancias relativas al método de ataque y de cualquier otra instancia que se asemeje a ese método; es decir, de los eventos que estén justo por debajo del umbral. Estos últimos eventos permiten a los científicos de datos perfeccionar sus modelos para no pasar por alto ningún tipo de comportamiento. Se iteran modelos rápidamente hasta que se cumplen los estrictos estándares de calidad en cuanto a detectar métodos de ataque en el mundo real.

En las últimas fases del desarrollo de métodos de detección se crea una interfaz de usuario *ad hoc* que brinda todo el contexto del método de ataque y, si procede, información adicional acerca de lo que se considera normal en los sistemas que se van a proteger. Los modelos pasan a producción, donde comienzan a funcionar y a informar de los incidentes a los clientes. El mismo procedimiento con el que se recopilan los datos para crear prototipos se utiliza para supervisar la eficacia del modelo en el mundo real y, si es necesario, para hacer mejoras en la detección.

Todo este esfuerzo vale la pena porque de este modo no hay que ajustar frecuentemente los modelos que, además, son eficaces tanto frente a las herramientas de ataque actuales como frente a las que puedan aparecer en el futuro. El enfoque de Vectra basado en la seguridad destaca por su capacidad para detectar la actividad de los atacantes, no solo eventos atípicos.

Motor de streaming en tiempo real para obtener resultados prácticos

La detección debe realizarse lo antes posible. Si las alertas tardan en dispararse, los atacantes aprovecharán para llegar más lejos. Los algoritmos de Vectra se ejecutan en datos en streaming, en lugar de hacerlo con lotes periódicos, de modo que las detecciones de Vectra permiten hallar a los atacantes en poco tiempo y así tener margen suficiente para detener su avance.

Es importante tener en cuenta a qué escala se trabaja porque las redes empresariales, las implementaciones en la nube y los servicios SaaS crecen constantemente, lo que para las detecciones de Vectra supone tener que procesar cada vez más datos. El motor de streaming en tiempo real de Vectra da cobertura a grandes multinacionales extrayendo los datos que necesita para generar conocimiento a largo plazo, sin que el volumen de datos suponga un problema.

La eficacia de los algoritmos, sobre todo la de aquellos que siguen un método de aprendizaje sin supervisión, queda considerablemente lastrada por la cantidad de datos históricos disponibles. Cuando se ejecutan detecciones en lotes, se limita el volumen de datos que se puede procesar en un plazo razonable. En el modelo de streaming de Vectra, los algoritmos extraen solo los datos que necesitan de un evento y los convierten en referencias nuevas para los modelos. Como aprenden de datos en streaming, esas referencias se crean a partir de meses de datos y millones de eventos, lo que garantiza una máxima calidad de las alertas.

Inteligencia artificial para correlacionar amenazas

La IA de Vectra no solo se emplea para identificar métodos de ataque concretos, sino también para correlacionar esa actividad y poder identificar, categorizar y priorizar los ataques que progresan activamente. Es necesario hacer esta correlación porque los ciberdelincuentes llevarán a cabo varias acciones en distintos dominios para alcanzar su objetivo final. Un algoritmo de correlación especializado analiza comportamientos en cuentas, hosts, la red y la nube, para indicar cuándo se produce un incidente de seguridad real.

A continuación, este algoritmo de correlación atribuye los comportamientos a cuentas o hosts como anclajes estables.

Por ejemplo, en los entornos de red y de nube híbrida, las IP transitorias se atribuyen a hosts estables en función de los artefactos que se identifican mediante un algoritmo conocido como host-id. Esos artefactos se extraen de los metadatos de red que incluyen, por ejemplo, información acerca de los host principales de Kerberos, direcciones MAC DHCP y cookies, e información de integraciones de API como EDR, vCenter, Azure y AWS. Una vez que los artefactos se han atribuido a un host concreto, cada vez que se observa una IP con un artefacto determinado, ese flujo de metadatos y el comportamiento de ataque asociado pueden atribuirse a un host concreto; no solo a la IP.

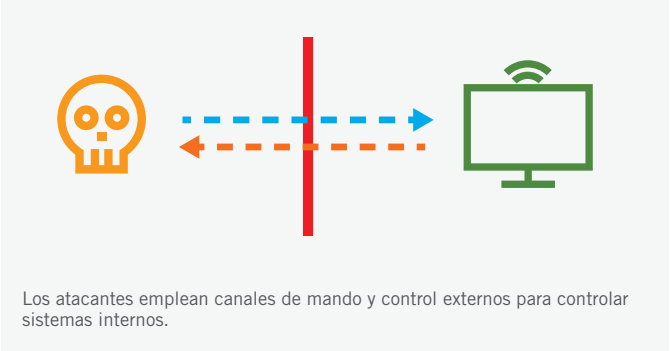
> La IA de Vectra no solo se emplea para identificar métodos de ataque concretos, sino también para correlacionar esa actividad y poder identificar, categorizar y priorizar los ataques que progresan activamente.

En AWS se presenta otro tipo de dificultad relacionada con la atribución: en el plano de control de AWS los eventos se registran como asociados a roles asumidos, no a las cuentas de usuario subyacentes. Puede haber muchas cuentas que asuman un rol determinado, pero para poder responder a un ataque, es fundamental conocer al usuario de IAM o SAML real que hay detrás de este rol.

Los atacantes expertos pueden ponérselo aún más difícil a los responsables de seguridad encadenando roles para tratar de ocultar el origen de un ataque. Gracias a una tecnología creada a medida, que se conoce como Kingpin, Vectra puede seguir el rastro por la cadena de roles para atribuir un ataque a un usuario subyacente, en lugar de a un rol ambiguo.

Una vez que se logran atribuir los comportamientos de ataque a un indicador estable, se correlacionan para identificar el perfil de comportamiento subyacente del sistema, que posteriormente se sirve para etiquetar y priorizar las amenazas en curso. El algoritmo de correlación se diseñó para reproducir las acciones que los analistas y los investigadores de seguridad de Vectra llevan a cabo cuando investigan amenazas, para ofrecer la posibilidad de clasificar situaciones de ataque complejas (como las amenazas externas o las amenazas internas a nivel de administrador) para que se analicen al instante.

Estudio de caso de detección con IA: canales de mando y control cifrados



Los atacantes emplean canales de mando y control externos para controlar sistemas internos.

Método de ataque

Cualquier ataque que se lanza a través de una red necesita un canal de mando y control (C2). Los atacantes con acceso a un host desplegarán un software malicioso que se pone en contacto con un servidor externo. Aunque es la máquina interna la que inicia la conexión, las respuestas del servidor externo contienen instrucciones que el host infectado ejecuta y permiten al atacante avanzar hacia su objetivo.

Las herramientas de mando y control se incluyen en marcos de ataque comerciales como Cobalt Strike y Metasploit, y también las desarrollan internamente los grupos de atacantes expertos. Todos estos marcos admiten el cifrado del canal, además de otras técnicas (como domain fronting o session jitter) que ayudan a los atacantes a evitar la detección.

Vectra detecta los canales de mando y control a pesar del cifrado u otras técnicas de evasión.

Metodología de detección

Vectra detecta los canales de mando y control incluso si usan cifrado u otras técnicas de evasión. Este grado de protección se consigue gracias al empleo del enfoque basado en la seguridad que mencionamos antes y que, como vemos, ataja muchos de los problemas que surgen al abordar la cuestión con un enfoque basado en las matemáticas.

Cuando el equipo de investigación de seguridad de Vectra abstrajo el comportamiento de un canal de mando y control, observó que los indicadores más claros que el método detectaba no eran aspectos circunstanciales del tráfico, como dominios inusuales o agentes de usuario, sino la forma real que el tráfico de la red adoptaba a la larga.

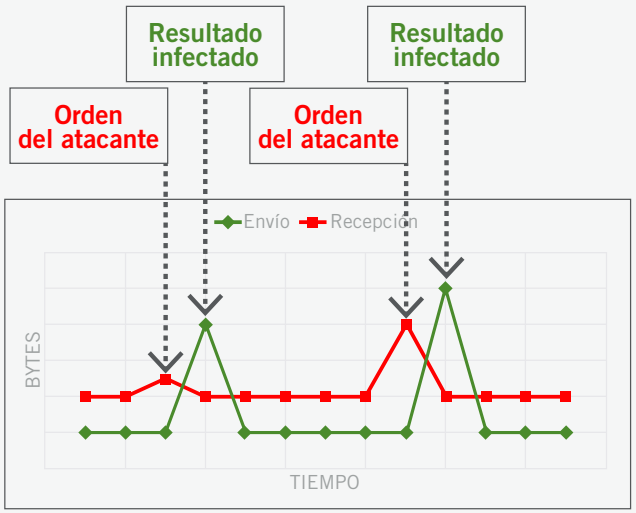
Observemos el siguiente ejemplo de tráfico benigno de un sistema externo:

Tráfico benigno

Ejemplo de tráfico benigno de transferencia de datos de balizas.

Este ejemplo de tráfico representa un host que emite balizas con un servidor externo. Las balizas (o *beacons*) son una función de red muy habitual que se emplea en servicios como teletipos bursátiles, aplicaciones de chat y rastreadores de anuncios, y permiten que sistemas locales y remotos se sincronicen y se comuniquen. Esta misma función la utilizan los canales de mando y control maliciosos.

Sin embargo, hay una pequeña diferencia en la forma en la que aparece una baliza cuando se utiliza en un teletipo de bolsa y en un canal malicioso. Veamos los siguientes datos que representan un túnel cifrado malicioso:

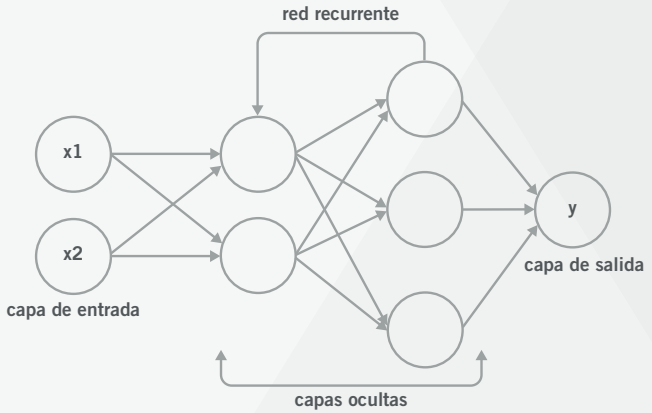


Ejemplo de tráfico malicioso de transferencia de datos de mando y control.

¿Observa los picos? Se producen cuando el atacante envía una orden y el sistema infectado devuelve un resultado. El primer pico de datos llega inadvertido en los bytes que se reciben y rápidamente le sigue la respuesta del equipo infectado.

Los científicos de datos de Vectra analizaron esos patrones y consiguieron desarrollar el enfoque óptimo para detectar este tipo de comportamiento. Los datos de las series temporales que caracterizan el comportamiento del canal de mando y control son muy parecidos a los que se emplean en el reconocimiento del habla y el procesamiento del lenguaje natural, por lo que el equipo decidió recurrir a un modelo de aprendizaje profundo.

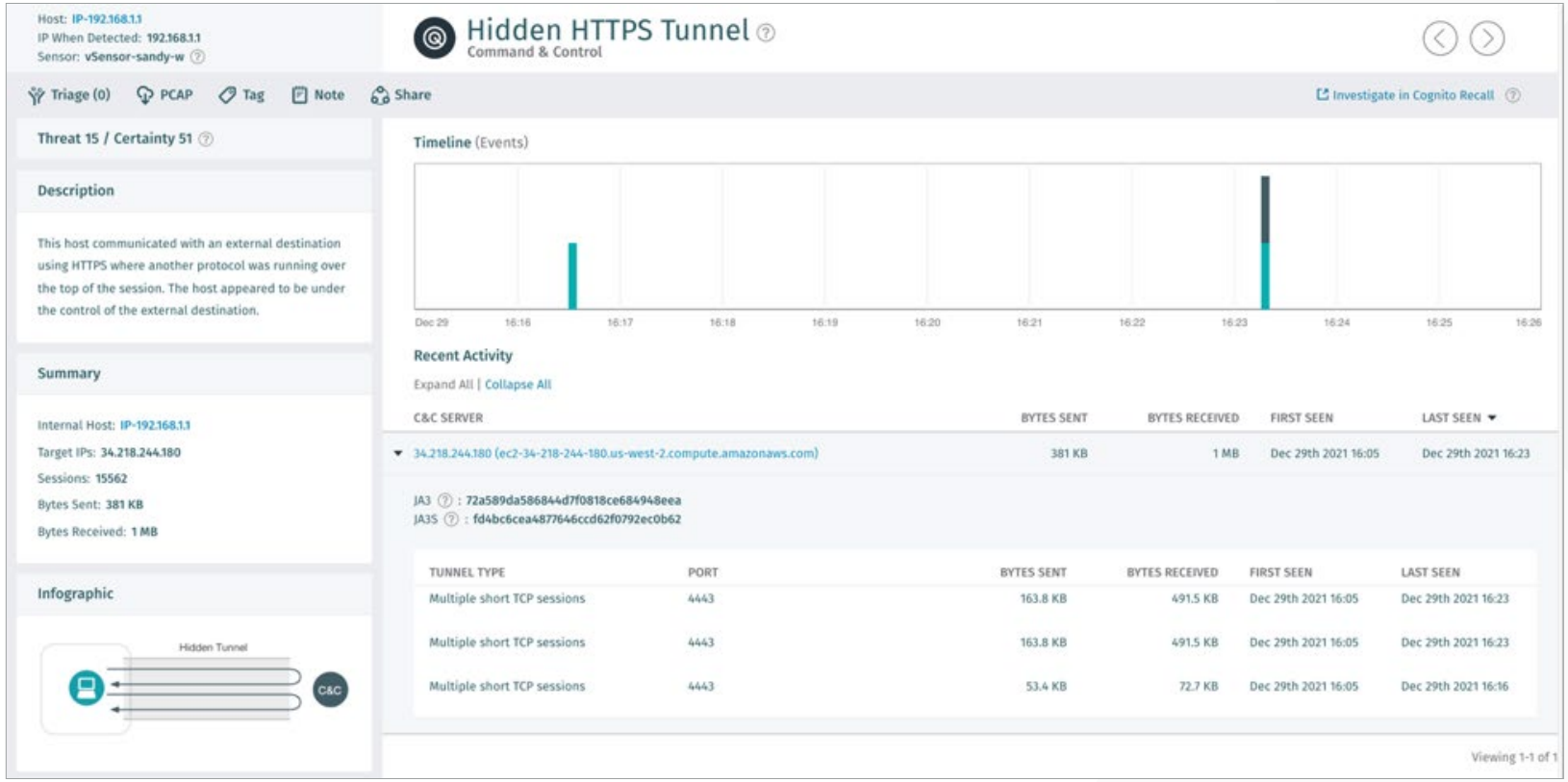
Vectra emplea una arquitectura de red neuronal recurrente específica conocida como "memoria a corto y largo plazo" (LSTM) para identificar el comportamiento de los ataques. Este tipo de algoritmo destaca en la comprensión de los eventos en varias escalas de tiempo diferentes, lo que resulta clave para comprender bien la naturaleza de los datos de la comunicación de mando y control. La LSTM se entrena con muestras reales y generadas mediante algoritmos. El conjunto de datos abarca situaciones, herramientas, configuraciones y entornos muy distintos que permiten al modelo identificar la señal generalizable de un canal de control con independencia de la herramienta que se utilice.



Vectra utiliza redes neuronales recurrentes para distinguir las comunicaciones de mando y control maliciosas de las balizas benignas.

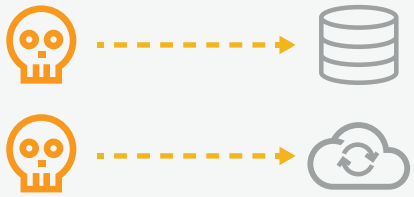
También es importante saber que este enfoque algorítmico existe gracias a la forma en que Vectra da formato a los datos de las sesiones de red. Aunque Vectra puede generar metadatos parecidos a los de Zeek, los metadatos que extrae el analizador propio de Vectra son más fiables que los del estándar de Zeek porque puede analizar las comunicaciones de red a intervalos de menos de un segundo. Esta perspectiva detallada permite observar claramente todo tipo de comunicaciones benignas y maliciosas, y permite a los científicos de datos de Vectra emplear los algoritmos que ofrecen la mejor cobertura posible ante situaciones muy distintas.

El resultado de estos metadatos exclusivos y el sofisticado enfoque algorítmico hacen posible conjuntamente detectar a los atacantes. La decisión de centrarse en los propios datos de comunicación, en vez de hacerlo en las señales superficiales, blinda la seguridad ante los cambios de herramientas y el cifrado del tráfico. Gracias a la detección de señales de comportamiento inequívocas, evitamos usar filtros de supresión que podrían evitar la detección de canales ocultos o las acciones de atacantes furtivos.



Método de detección de Vectra para un canal de mando y control cifrado.

Estudio de caso de detección con IA: uso fraudulento de credenciales con privilegios en la red y la nube



Los atacantes utilizan las credenciales con privilegios para desplazarse lateralmente y ejecutar otras tácticas en la red y la nube.

Método de ataque

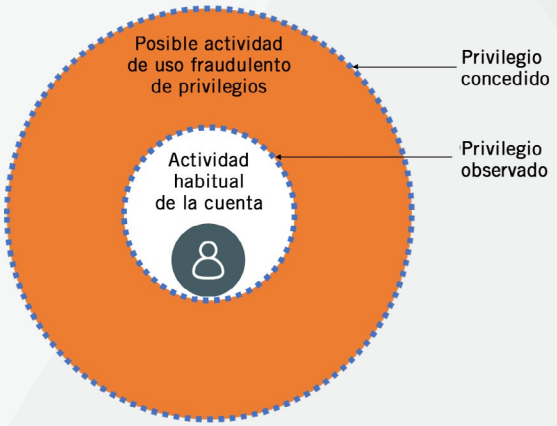
Los atacantes que logran hacerse con credenciales con privilegios pueden acceder a más recursos de red y de nube. Con las credenciales, los atacantes pueden acceder a una red sin tener que usar malware ni cargas útiles de exploits, que pueden dejar rastro o hacer saltar las alarmas. Si se implantasen requisitos en cuanto al establecimiento del mínimo de privilegios de acceso para los usuarios, se podrían mitigar ciertas amenazas. Sin embargo, como se demuestra en ataques recientes, esto sigue siendo muy difícil.

Para evitar el acceso fraudulento con credenciales robadas se debe detectar el incidente de vulneración de las credenciales, pero descubrir a un ciberdelincuente cuando roba una cuenta y la aprovecha para lanzar un ataque presenta una serie de retos únicos. En estos casos, los permisos de la cuenta autorizan expresamente todas las acciones del atacante. Tampoco saltarán las alarmas que se basen en conceptos como las interacciones nuevas porque los usuarios se mueven en entornos dinámicos en los que acceder a nuevos recursos es indispensable en su labor diaria. Un atacante que conozca bien el entorno intentará realizar acciones que no sean nuevas en una cuenta concreta, para evitar levantar sospechas. Para poder identificar un uso fraudulento de credenciales con privilegios, se requiere un enfoque basado en la seguridad que tenga presente qué pretende lograr el atacante con las credenciales que ha robado.

Metodología de detección

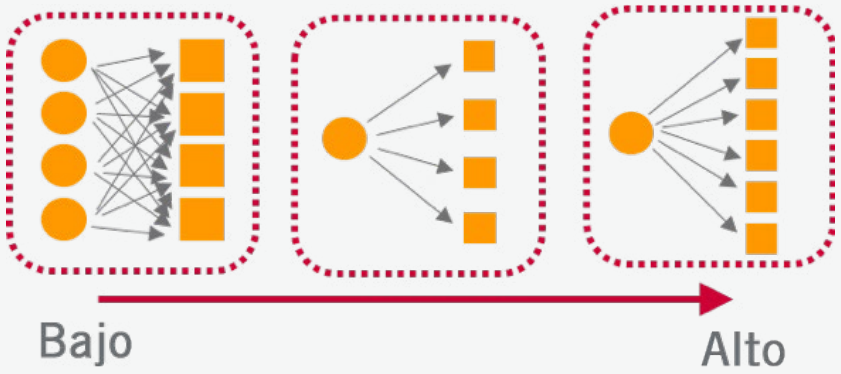
Vectra puede identificar el uso fraudulento de credenciales con privilegios en entornos de red y de nube. En este enfoque de detección basado en la seguridad es fundamental saber qué pretenden hacer los atacantes con las credenciales robadas. Con unas credenciales con privilegios, un atacante puede acceder a servicios y funciones considerados valiosos y privilegiados.

Los investigadores de seguridad de Vectra descubrieron que conocer cuáles son los privilegios para cada cuenta, host, servicio y operación en la nube permite elaborar un mapa de todos los recursos de gran valor que existen en un entorno. Si bien el concepto de *privilegio concedido* está claro, presenta un límite máximo con el que comparar el verdadero privilegio para una acción. Los equipos de Vectra de ciencia de datos y de investigación de la seguridad identificaron una forma nueva de representar el valor de los sistemas de un entorno basándose en lo que habían observado a lo largo del tiempo. Esta perspectiva dinámica y real del valor se denomina *privilegio observado*. Esta concepción basada en datos de los privilegios ofrece un enfoque eficaz de confianza cero (o Zero Trust) para el uso de credenciales sin configuraciones manuales.



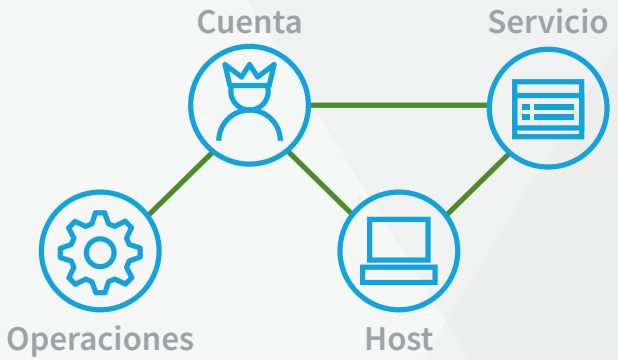
El nivel de privilegio observado otorga una perspectiva Zero Trust del privilegio normal que un usuario necesita para hacer su trabajo. Si el privilegio se usa al margen de lo que normalmente es necesario, se somete a un escrutinio extra.

La IA de Vectra calcula el nivel de *privilegio observado* teniendo en cuenta el historial de interacciones entre las entidades a las que se les ha hecho un seguimiento, en lugar de tener en cuenta los privilegios que define un administrador de TI. La amplitud y la especificidad del acceso, así como su uso influyen enormemente en la puntuación. El nivel de privilegios será bajo en el caso de un sistema que accede a varios sistemas a los que normalmente acceden otros, mientras que será alto si un sistema accede a muchos sistemas a los que no acceden los demás. De este modo, Vectra distingue entre cuentas de administrador de dominio y cuentas de usuario normales.



Vectra aprende cuáles son los niveles de privilegio observado basándose en el comportamiento de los usuarios. El privilegio de una cuenta que accede a muchos servicios comunes es menor que el de otra cuenta que accede a servicios a los que pocos acceden.

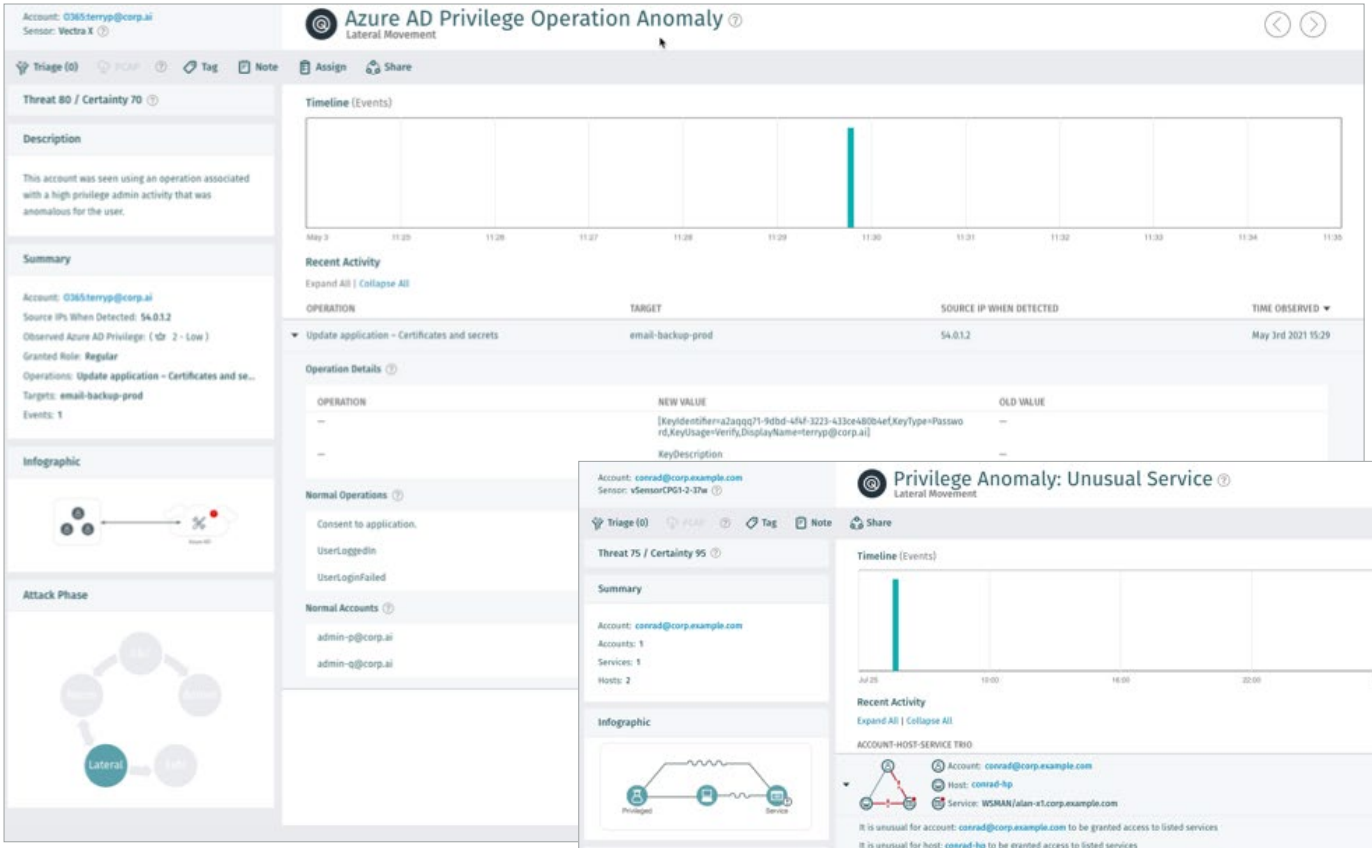
Una vez que se han calculado las puntuaciones de privilegios observados, se elabora un mapa de todas las interacciones entre cuentas, servicios, hosts y operaciones en la nube para conocer cuáles son las que normalmente se dan entre sistemas. A continuación, un conjunto de algoritmos de aprendizaje sin supervisión, que incorporan la puntuación de privilegios, identifican los casos anómalos de uso fraudulento de privilegios, mediante el empleo de algoritmos de detección de anomalías personalizados e implementaciones de HDBSCAN (o la agrupación jerárquica espacial de aplicaciones con ruido, basada en la densidad).

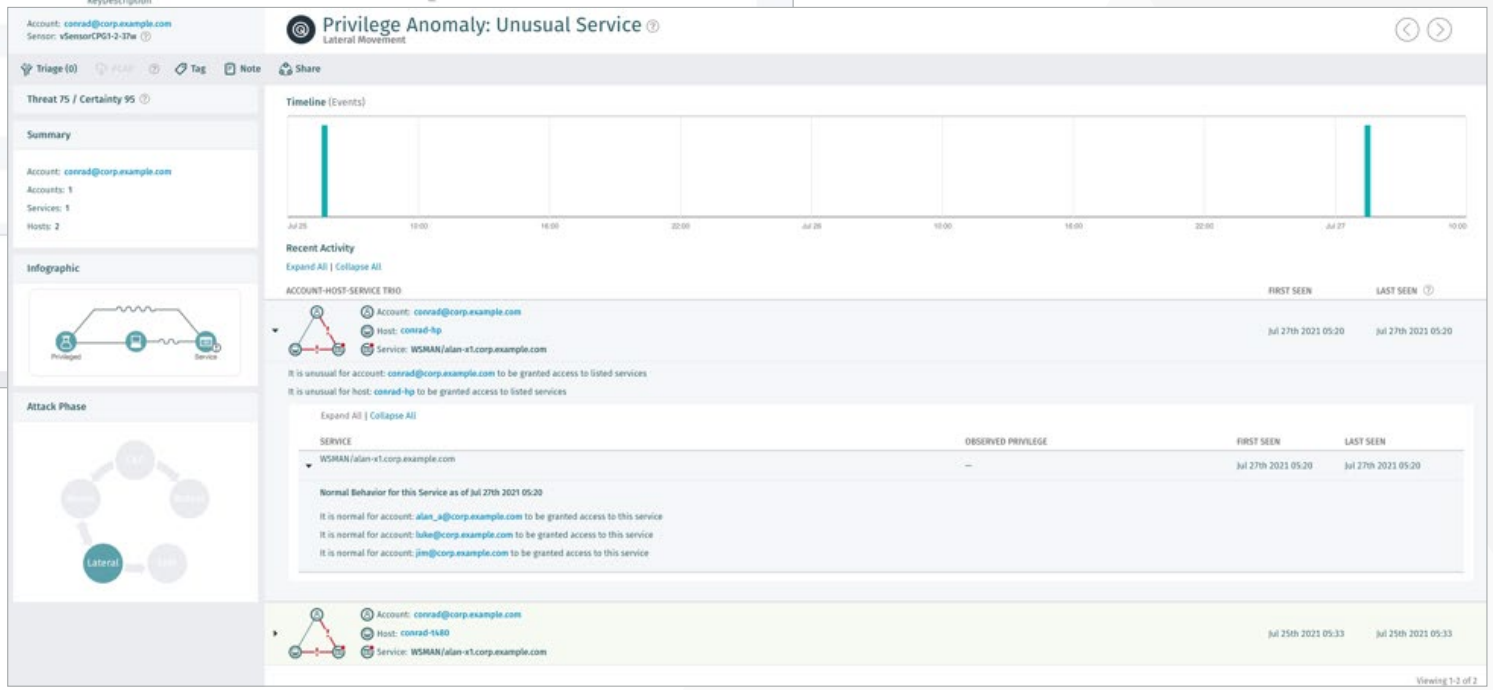


Para identificar el uso fraudulento de credenciales, Vectra aplica el aprendizaje sin supervisión, que tiene en cuenta el nivel de privilegio observado y las interacciones entre cuentas, hosts, servicios y operaciones en la nube.

Este sofisticado enfoque basado en la seguridad ofrece la capacidad de identificar credenciales robadas de las que se hace un uso fraudulento, tanto en redes locales como la nube. La métrica del nivel de *privilegio observado* centra la detección en las acciones anómalas más relevantes y brinda mayor precisión y exhaustividad que otro enfoque que no preste atención a esta perspectiva crucial.

> La IA de Vectra calcula el *privilegio observado* teniendo en cuenta el historial de interacciones entre las entidades a las que se les ha hecho un seguimiento, no como lo define un administrador de TI.





Métodos de detección de Vectra para cuentas que hacen un uso fraudulento de los privilegios.

Más innovación y mejor que los atacantes

Sin duda, los atacantes seguirán innovando, por lo que los responsables de la seguridad deberían hacer lo mismo. A lo largo de los años, Vectra ha innovado constantemente para desarrollar la plataforma de detección de amenazas y respuesta más eficaz posible para proteger recursos locales y en la nube.

Vectra ha desarrollado más de cien detecciones con IA basadas en la seguridad y ha identificado un sinfín de amenazas en la red de los clientes y en los entornos de nube, frustrando así las intenciones de los atacantes. Cada detección se desarrolló gracias a los profundos conocimientos de los equipos acerca de cómo se llevan a cabo los ataques y con el empleo de algunas de las técnicas de aprendizaje automático más avanzadas que existen. En total, Vectra tiene 33 patentes de tecnología para estas detecciones.

Además de la protección que ofrece la tecnología patentada de Vectra, nos sentimos orgullosos de ser el proveedor con más menciones en el marco de la NSA y MITRE (conocido como MITRE D3FEND), que define las contramedidas que deben adoptar los responsables de seguridad para proteger su entorno. D3FEND es un gráfico que indica a estos responsables cómo detener los ataques y evitar las técnicas empleadas por los ciberdelincuentes, que se definen en el marco ATT&CK de MITRE. En total, el marco D3FEND menciona 12 patentes de Vectra, que se utilizan como referencia para las contramedidas de seguridad.

En Vectra tenemos el compromiso de hacer del mundo un lugar más seguro y justo. Por eso, seguiremos aprovechando la IA basada en seguridad para innovar y crear métodos de detección con los que impedir que los atacantes consigan sus objetivos.

Para obtener más información, escríbanos a la dirección info@vectra.ai.

Correo electrónico info@vectra.ai vectra.ai