VECTRA
SECURITY THAT THINKS

WHITEPAPER

Die KI hinter Vectra AI

INHALTSVERZEICHNIS

Vectra® schützt Unternehmen durch Erkennen und Stoppen von Cyber-Angriffen.

Vectra® ist ein führender Anbieter für Bedrohungserkennung und Response bei Hybrid- und Multi-Cloud-Unternehmen. Die Vectra-Plattform erkennt mithilfe von künstlicher Intelligenz (KI) innerhalb kürzester Zeit Bedrohungen in der Public Cloud sowie in Identitäten, SaaS-Anwendungen und Rechenzentren. Nur Vectra optimiert KI, um damit Angreifermethoden – die TTPs, die die Grundlage aller Angriffe bilden – zu erkennen, statt nur bei Abweichungen zu warnen. Das daraus entstehende zuverlässige Bedrohungssignal und der klare Kontext ermöglichen dem Security-Team, auf Bedrohungen früher zu reagieren und laufende Angriffe schneller zu stoppen. Als zuverlässiger Partner bietet Vectra Unternehmen weltweit Resilienz gegenüber gefährlichen Cyber-Bedrohungen und unterstützt sie bei der Abwehr von Ransomware, Kompromittierungen von Lieferketten, Identitätsübernahmen und anderen geschäftsschädigenden Cyber-Angriffen. Weitere Informationen finden Sie unter vectra.ai.

Einführung

Wir bei Vectra AI beschäftigen uns vor allem mit Data Science. Wir waren immer davon überzeugt, dass Data Science und künstliche Intelligenz (KI) bei richtiger Anwendung unseren Kampf gegen Cyber-Angriffe verändern und Verteidigern einen Vorteil geben kann. Doch KI ist nicht gleich KI. In diesem Whitepaper erläutern wir, was künstliche Intelligenz ist und welche Begriffe in Bezug auf KI-Lösungen wichtig sind. Zudem beschreiben wir die zwei wichtigsten Methoden bei der Anwendung von KI zur Bedrohungserkennung und gewähren schließlich einen Einblick in die Art und Weise, wie Vectra Bedrohungen mithilfe von künstlicher Intelligenz aufspürt.

Das Whitepaper richtet sich ebenso an KI-Skeptiker wie an solche, die in KI ein großes Potenzial sehen.

Was ist KI?

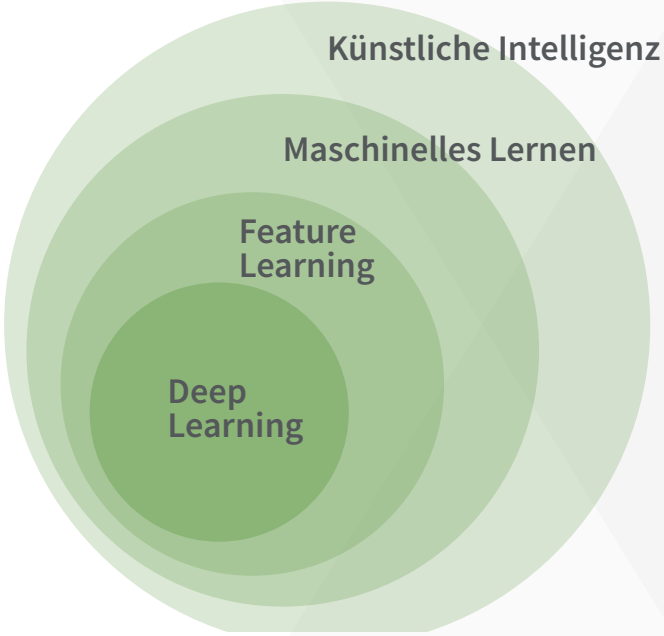
Definition von KI

Häufig werden die Begriffe künstliche Intelligenz, maschinelles Lernen (bzw. Machine Learning) und Deep Learning dahingehend missverstanden, dass sie zum selben Fachgebiet gehören oder es sich um verschiedene Qualitätsabstufungen handelt. Das ist jedoch falsch. Diese Begriffe sind zwar miteinander verwandt, doch jeder hat eine eigene Bedeutung. Wer versteht, wie sich diese Begriffe voneinander abgrenzen, kann KI-Tools besser verstehen.

Künstliche Intelligenz (KI): Künstliche Intelligenz wird definiert als ein System, das logisches Denken automatisieren und ähnlich wie der menschliche Verstand funktionieren kann. Es ist ein sehr allgemeiner Begriff, der die Teilgebiete des Machine Learning, Feature Learning und Deep Learning beinhaltet. Der Begriff KI gilt ebenfalls für Systeme, die explizit programmierte Regeln anwenden sowie für Systeme, die autonom Erkenntnisse aus einer großen Menge an Daten gewinnen. Die spätere Form der KI, die aus Daten lernt, bildet die Grundlage für Technologien wie selbstfahrende Autos sowie virtuelle Assistenten und fällt in das Teilgebiet des maschinellen Lernens.

Maschinelles Lernen (ML): Maschinelles Lernen ist ein Teilgebiet der KI, in dem die Aktionen eines Systems nicht explizit vom Menschen vorgegeben, sondern durch Lernen aus Daten bestimmt werden. Diese Systeme können Milliarden von Datenpunkten verarbeiten und auf diese Weise lernen, neue Instanzen von Daten optimal darzustellen und anschließend darauf zu reagieren.

Feature Learning (RL): Obwohl Feature Learning (im Engl. bekannt als „Representation Learning“, RL) kein allgemein diskutiertes Thema ist, ist es doch Hauptbestandteil vieler moderner KI-Technologien. In diesem Teilgebiet geht es um das Lernen neuer abstrakter Darstellungen aus Daten. Ein Beispiel für RL ist die Transformation von Bildern verschiedener Größen in eine Liste von Zahlen mit einheitlicher Länge, die eine Reduzierung der Originalbilder darstellt. Diese Abstraktion dient hauptsächlich dazu, nachgelagerten Systemen zu ermöglichen, besser auf neue Datentypen reagieren zu können.



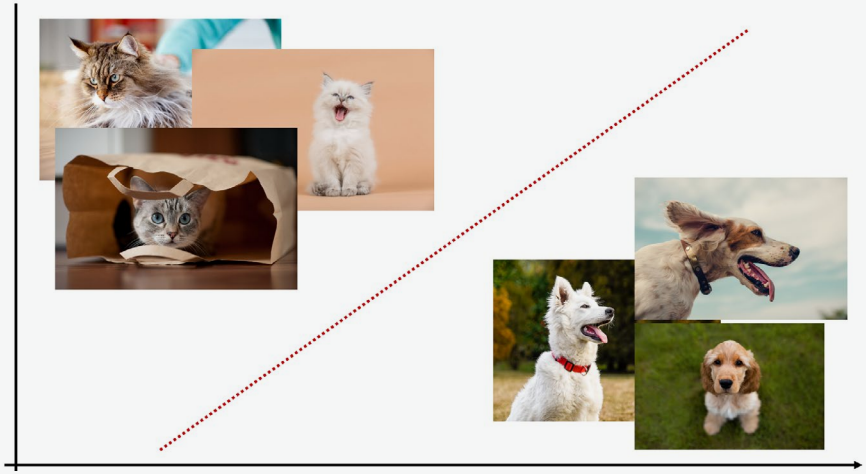
Die Beziehung zwischen den unterschiedlichen KI-Teilgebieten.
Quelle: „Deep Learning“, Goodfellow, Bengio & Courville (2016)

Deep Learning (DL): Deep Learning wird häufig mit neuronalen Netzen assoziiert und baut auf den allgemeineren Teilgebieten des ML und RL auf. Hier werden Abstraktionshierarchien aus Daten erkannt, die Eingaben in einer immer komplexer werdenden Weise darstellen. Die durch das menschliche Gehirn inspirierten DL-Modelle verwenden Schichten von Neuronen, deren Synapsengewicht sich bei der Reaktion auf Eingaben verändert. Tiefere Schichten im Netzwerk lernen neue abstrakte Muster, die helfen, Aufgaben wie das Kategorisieren von Bildern oder das Übersetzen eines Textes zu vereinfachen. Zwar lassen sich mit Deep Learning einige komplexe Probleme effektiv lösen, allerdings ist es kein Allheilmittel für die Automatisierung von Intelligenz.

Die verschiedenen Arten von Algorithmus-Lerntechniken

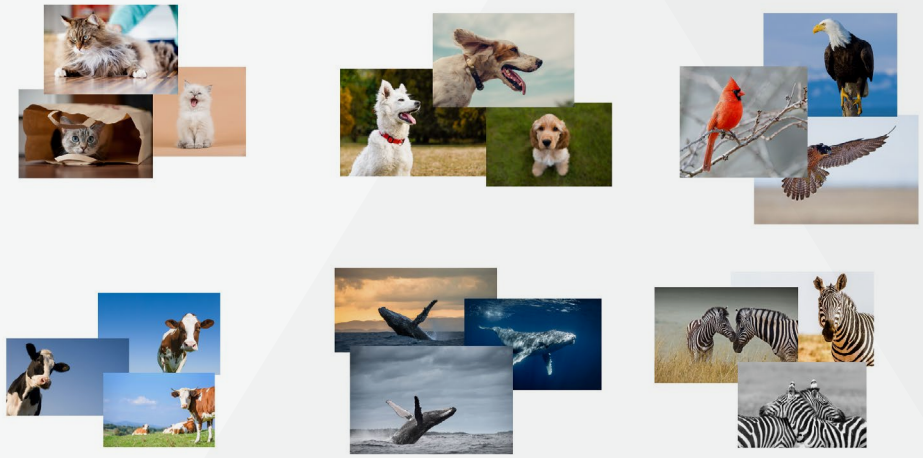
Eine Kernfunktion von ML-Algorithmen ist die Möglichkeit, Daten in verschiedene Klassen einzuordnen. Es gibt eine Handvoll weit gefasster Lernkategorien, die diese Funktion unterstützen, wobei am häufigsten die Kategorien **überwacht** und **unüberwacht** verwendet werden.

Beim **überwachten** Lernen lernt das Modell an einem gekennzeichneten Datensatz. Anschließend kann das Modell die Kennzeichnungen (Label) neuer Daten vorhersagen. Betrachten wir dazu das untenstehende Beispiel: Wird ein überwachtes Lernmodell mit sehr vielen Bildern von Katzen und Hunden gefüttert, kann es bei einem neuen Bild voraussagen, ob es sich um eine Katze oder einen Hund handelt. Für das überwachte Lernen ist ein Satz gekennzeichneter Trainingsdaten nötig, aus denen das Modell lernen kann. Nachdem sie trainiert wurden, können diese Modelle äußerst effektiv Muster in neuen Daten erkennen und diese kennzeichnen.



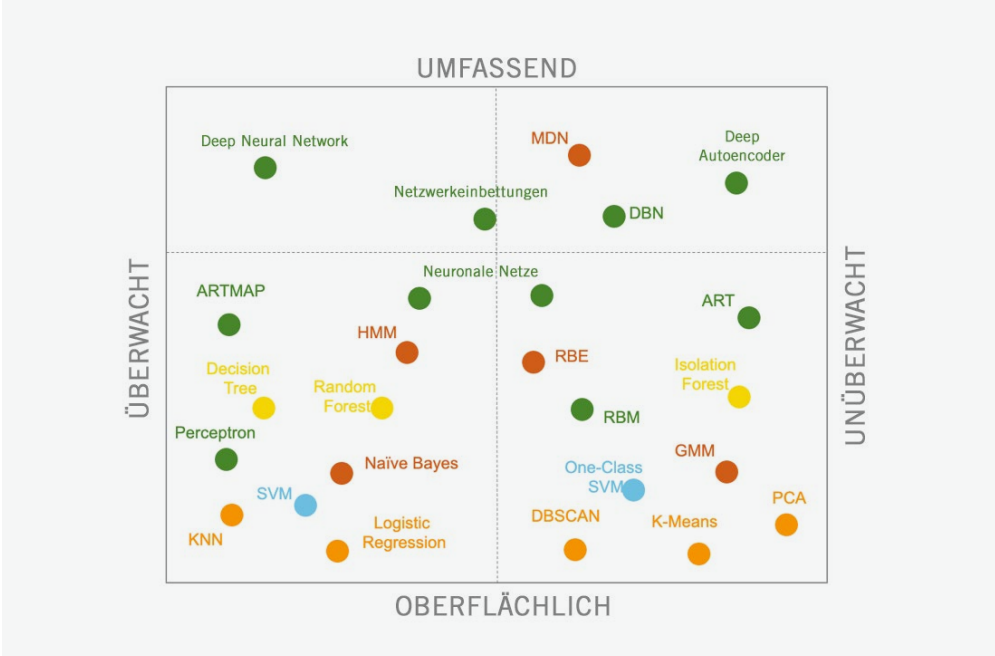
Überwachtes Lernen erkennt mittels gekennzeichneter Daten Faktoren, mit denen verschiedene Kennzeichnungen unterschieden werden können. Modelle mit diesem Lerntyp sind in der Lage, neue Daten zuverlässig zu kennzeichnen.

Beim **unüberwachten** Lernen lernt das Modell an einem ungekennzeichneten Datensatz. Diese Modellen lernen Strukturen aus vorgegebenen Daten und können dann beurteilen, ob und wie neue Daten zu den gelernten Strukturen passen. Unüberwachte Lernmodelle haben den Vorteil, dass sie im Vorfeld kein Training benötigen. Mit diesem Ansatz lassen sich besonders gut Datenpunkte finden, die sich von den anderen unterscheiden, allerdings können diese Anomalien oder Ausreißer dabei nicht gekennzeichnet werden.



Beim unüberwachten Lernen lernt das Modell die grundlegende Struktur nicht gekennzeichneter Daten. Modelle, die diesen Lerntyp nutzen, können beurteilen, wie gut neue Daten in eine gelernte Struktur passen.

Wie unten zu sehen ist, verbirgt sich hinter diesen allgemeinen Ansätzen eine Reihe verschiedener Lernalgorithmen, zu denen immer wieder neu entwickelte hinzukommen. Noch komplizierter wird es dadurch, dass Algorithmen zu noch komplexeren Systemen kombiniert werden können. Dann stellt sich die Frage, wie ein Datenwissenschaftler den richtigen Algorithmus zur Lösung eines bestimmten Problems findet. Oder könnte ein einzelner Algorithmus unabhängig vom jeweiligen Problem besser als alle anderen sein?

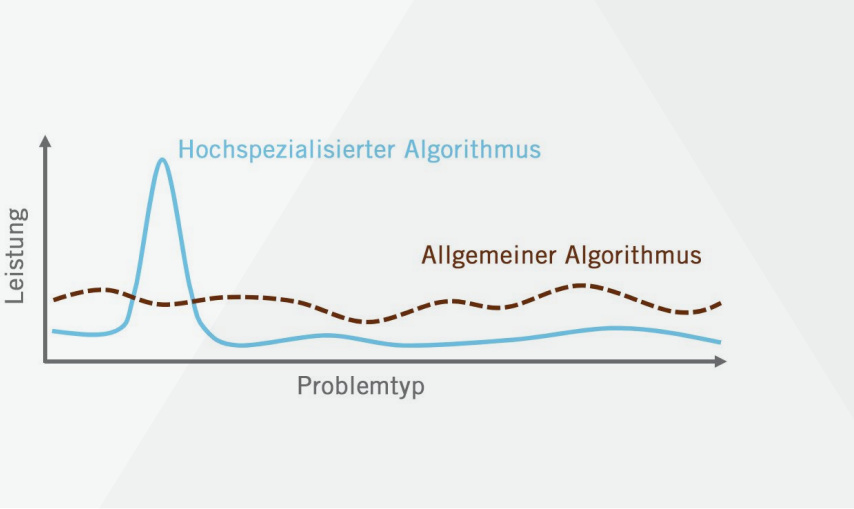


Es gibt eine große Zahl von ML-Algorithmen, die abhängig vom zu lösenden Problemtyp alle unterschiedliche Stärken und Schwächen aufweisen.

Das No-Free-Lunch-Theorem

Es hat sich herausgestellt, dass es keinen universellen Algorithmus gibt, der für alle möglichen Problemstellungen bessere Ergebnisse als alle anderen liefert. Dies wird als das „No-Free-Lunch-Theorem“ (engl. „kein kostenloses Mittagessen“ bzw. sinngemäß „nichts ist umsonst“) bezeichnet. Vereinfacht gesagt gibt es für ein Problem immer einen spezialisierten Algorithmus, der besser geeignet ist als ein allgemeiner Algorithmus. Der Bedarf an speziellen Algorithmen für die Lösung spezieller Probleme macht deutlich, warum (wie oben erwähnt) ein immer größerer Bedarf an Algorithmen besteht. Es gibt Probleme, für die ein überwachtes neuronales Netz am besten geeignet ist, und andere, bei denen unüberwachte hierarchische Cluster optimale Ergebnisse liefern.

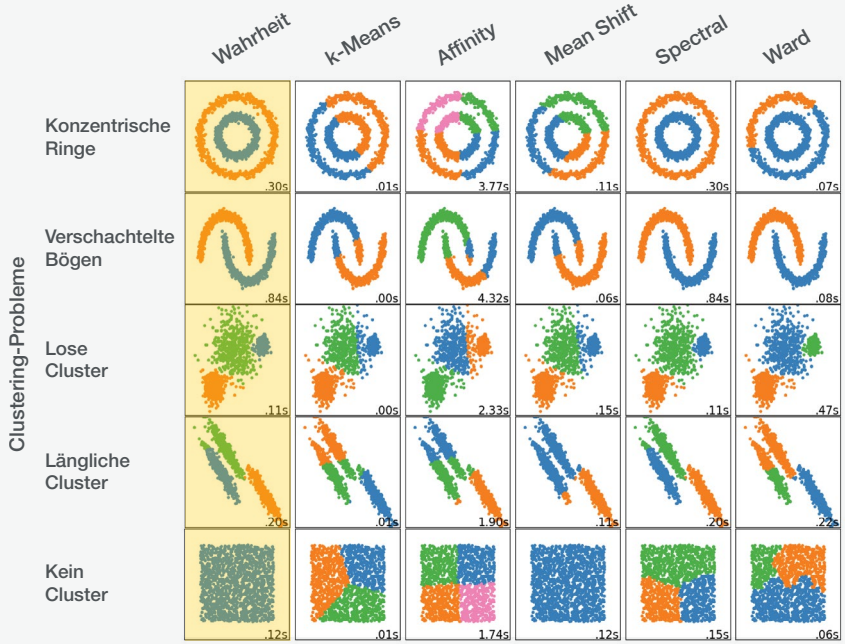
Der Algorithmus zur Bilderkennung in selbstfahrenden Autos lässt sich nicht zur Übersetzung einer Sprache in eine andere anwenden. Jeder Algorithmus ist speziell ausgewählt und optimal auf das zu lösende Problem und die Daten, mit denen das Modell arbeitet, zugeschnitten.



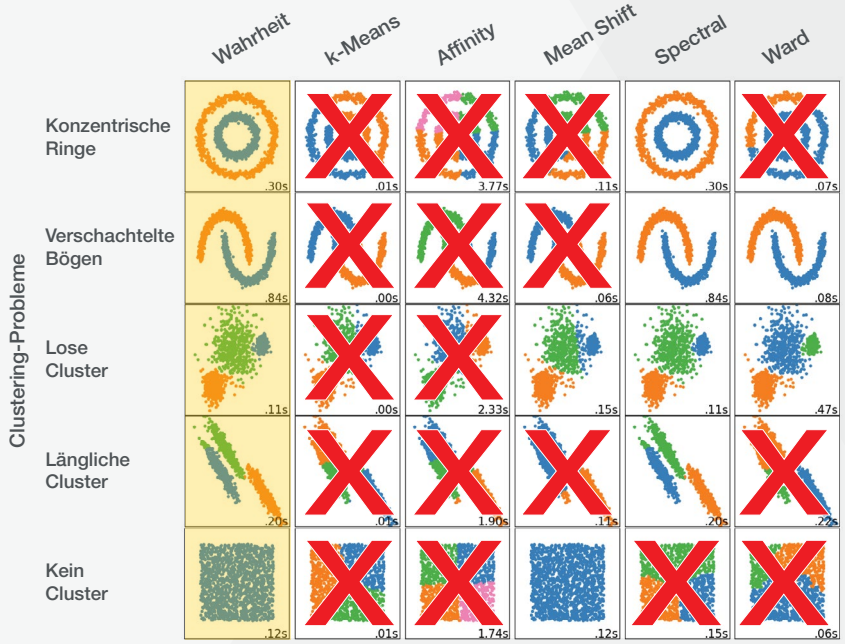
Das No-Free-Lunch-Theorem: Es gibt keinen universellen Algorithmus, der sich für jedes Problem eignet.

Das Suche nach dem richtigen Werkzeug für die Aufgabe

Wie findet also ein Datenwissenschaftler den richtigen Algorithmus? Das ist gleichermaßen eine Kunst wie eine Wissenschaft. Die richtige Problemstellung in Kombination mit einem tiefgehenden Verständnis der Daten kann den Datenwissenschaftler auf die richtige Spur bringen. Dabei sollte unbedingt beachtet werden, dass die falsche Vorgehensweise nicht nur zu suboptimalen, sondern auch zu völlig falschen Ergebnissen führen kann! Sehen wir uns dazu das untenstehende Beispiel an. Für jeden Datensatz führen die verschiedenen Algorithmen zu gänzlich unterschiedlichen Ergebnissen. Für jedes Problem gibt es einen optimalen Algorithmus. Noch wichtiger ist aber, dass bestimmte Algorithmen zu gänzlich falschen Ergebnissen führen. Es ist daher äußerst wichtig, den richtigen Ansatz für das richtige Problem auszuwählen.



Vergleich der Ergebnisse von ML-Algorithmen (x-Achse) bei unterschiedlichen Datensätzen (y-Achse). Tatsächliche Kennzeichnungen gelb hervorgehoben. Vorlage von scikit-learn.org.



Vergleich der Ergebnisse. Falsche Voraussagen, die zu falschen Ergebnissen führen würden, sind mit einem X gekennzeichnet. Es gibt keinen einzigen Algorithmus, der für jeden Datensatz geeignet ist. Vorlage von scikit-learn.org.

Wie misst man *gut*?

Ein wesentlicher Aspekt bei der Auswahl des richtigen Modells ist die Entscheidung darüber, wie der Erfolg eines Modells gemessen werden soll. Wenn es um die Leistung eines Modells geht, wird häufig auf die *Treffergenauigkeit* des Modells verwiesen.

$$Treffergenauigkeit = \frac{(True\ Positives + True\ Negatives)}{(True\ Positives + True\ Negatives + False\ Positives + False\ Negatives)}$$

Die Treffergenauigkeit ist eine wichtige Kennzahl, allerdings kann eine scheinbar gute Treffergenauigkeit auch die tatsächliche Leistung eines Modells verschleiern. Betrachten wir dazu ein Klassifizierungsproblem, bei dem das Ziel darin besteht, Daten entweder mit A oder B zu kennzeichnen. Wenn A mit einer 1.000 Mal höheren Wahrscheinlichkeit als B auftritt, können leicht 99,9 % erreicht werden, indem immer Daten mit A gekennzeichnet werden. Damit wird zwar eine sehr gute Treffergenauigkeit erreicht, allerdings wird dabei nie etwas korrekt als B gekennzeichnet werden. Dies zeigt eindeutig, dass die Treffergenauigkeit als Messwert nicht dazu geeignet ist, Fälle mit B zu finden. Zum Glück stehen Datenwissenschaftlern weitere Kennzahlen zur Verfügung, die ihnen dabei helfen, die Wirksamkeit des Modells für das Finden relevanter Fälle zu messen und zu optimieren.

Eine solche Kennzahl ist die *Wirksamkeit*. Mit ihr wird gemessen, wie gut ein Modell darin ist, eine bestimmte Kennzeichnung zu erraten, und zwar relativ zur gesamten Zahl der Vermutungen für diese Kennzeichnung, die das Modell vornimmt.

$$Wirksamkeit = \frac{True\ Positives}{(True\ Positives + False\ Positives)}$$

Datenwissenschaftler, die einen hohen Wirksamkeitswert erzielen wollen, entwickeln Modelle, die Kennzeichnungen vorhersagen und dabei nicht allzu viele Fehlalarme produzieren. Die Wirksamkeit verrät uns jedoch nicht, ob das Modell für uns relevante Fälle *nicht* gekennzeichnet hat. Eine weitere Kennzahl – die Sensitivität – verdeutlicht dies.

Die *Sensitivität* gibt an, wie oft ein Modell eine bestimmte Kennzeichnung bezogen auf alle Instanzen dieser Kennzeichnung korrekterweise gefunden hat.

$$Sensitivität = \frac{True\ Positives}{(True\ Positives + False\ Negatives)}$$

Datenwissenschaftler, die einen hohen Sensitivitätswert erreichen wollen, entwickeln Modelle, die auch wirklich alle relevanten Instanzen melden.

Datenwissenschaftler können den Erfolg ihrer Modelle messen und optimieren, indem sie die Treffergenauigkeit und Sensitivität nachverfolgen und ins Gleichgewicht bringen.

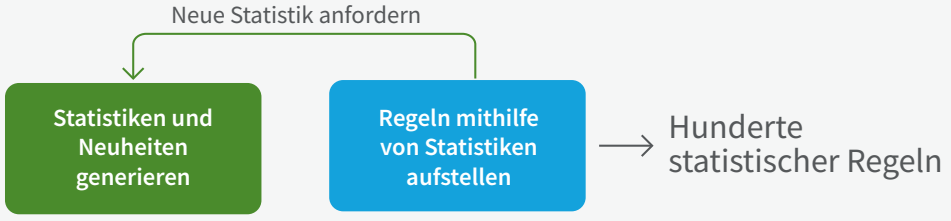
Ein wesentlicher Aspekt bei der Auswahl des richtigen Modells ist die Entscheidung darüber, wie der Erfolg eines Modells gemessen werden soll.

Anwendung der KI zur Bedrohungserkennung

Die künstliche Intelligenz und ihre vielen Fachgebiete spielen eine wichtige Rolle, wenn es darum geht, Angreifer in modernen Unternehmen aufzuspüren und zu stoppen. Für die aktive Identifizierung von Cyber-Bedrohungen haben sich zwei Herangehensweisen herauskristallisiert: eine mathematikorientierte und eine sicherheitsorientierte. In diesem Abschnitt untersuchen wir die Unterschiede zwischen den beiden Methoden und erläutern, warum sicherheitsorientierte KI den Security-Teams optimale Ergebnisse bietet.

Mathematikorientierte KI: ein fehlerhafter Ansatz zur Bedrohungserkennung

Bei der mathematikorientierten Herangehensweise generieren Datenwissenschaftler einfache Statistikdatensätze durch eine begrenzte Anzahl generischer Algorithmen, die auf die Erkennung von Ausreißern und neuen Merkmalen ausgerichtet sind. Sicherheitsforscher kombinieren diese Statistiken, um hunderte von statistischen Regeln aufzustellen. Wird eine neue Statistik benötigt, wird diese mithilfe desselben generischen Ansatzes generiert. Die statistischen Regeln werden im Zuge der Nachbearbeitung häufig mit eindeutigen Unterdrückungsfiltern ergänzt, um zusätzliche Erkennungsmengen auszublenden, die mit diesem generischen Ansatz erzeugt werden (hier kommen wir zurück auf das No-Free-Lunch-Theorem – generische Algorithmen führen zu suboptimaler Leistung).



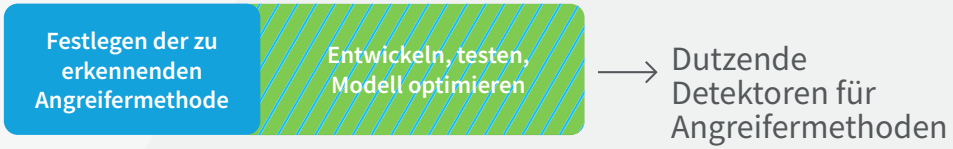
Der mathematikorientierte Ansatz zur Entwicklung von Erkennungen.

Schauen uns wir nun ein Beispiel an, bei dem ein Command & Control-Kanal erkannt werden soll: Zunächst generiert das Team der Datenwissenschaftler eine Statistik zur Seltenheit aller externen Domains. Die Sicherheitsforscher müssen dann den Schwellenwert für die Seltenheit festlegen, um C&C-Kanäle erkennen zu können. Tauchen viele von IoT-Geräten genutzte Domains über dem Schwellenwert auf, müssten mit einem Unterdrückungsfilter IoT-Geräte ignoriert werden. Anschließend werden weitere Unterdrückungsfilter für Anwenderagenten, Subnetze sowie andere Attribute angewendet, bis die Warnungen auf eine überschaubare Menge reduziert wurden. Aufgrund dieses allgemeinen Ansatzes müssen Unterdrückungsregeln angewendet werden, obwohl dadurch die Gefahr besteht, von Angreifern verwendete Umgehungstechniken zu übersehen.

Sicherheitsorientierte KI: ein Bedrohungserkennungsansatz mit maximaler Abdeckung und minimalem Rauschen

Bei der sicherheitsorientierten Herangehensweise sind die Definition des Problems (die Angreifermethode) und die Suche nach dem richtigen Modell eng miteinander verknüpft. Die Sicherheitsforscher definieren das Problem, indem sie eine breit angelegte Angriffsmethode identifizieren und nicht nur ein Tool oder einen einzelnen Exploit. Die Datenwissenschaftler suchen dann nach dem richtigen Algorithmus, der diese Methode erkennen kann, und arbeiten dabei eng mit den Sicherheitsforschern zusammen, um sich der Lösung schrittweise zu nähern. Mit diesem Ansatz wird die Angreifermethode direkt erkannt, wobei nicht nur oberflächliche Anomalien identifiziert werden, die häufig bei mathematikorientierten Ansätzen vorkommen.

Der sicherheitsorientierte Ansatz führt – gemessen an der Sensitivität und Wirksamkeit – zu besseren Ergebnissen. Außerdem ist dieser Ansatz weniger anfällig gegenüber Änderungen der Angreifer-Tools und erfordert weniger Erkennungstypen, sodass er für Security-Teams leichter zu handhaben ist. Tritt eine neue Angreifermethode immer häufiger auf, wird der sicherheitsorientierte Prozess in Gang gesetzt und eine neue Erkennung entwickelt. Die Verfeinerung des Ansatzes kann zusätzliche Entwicklungszeit in Anspruch nehmen. Allerdings verändern sich die Angreifermethoden nur sehr langsam und tauchen immer wieder zusammen mit älteren Methoden auf, die bereits abgedeckt sind.



Der sicherheitsorientierte Ansatz zur Entwicklung von Erkennungen.

So arbeitet Vectra

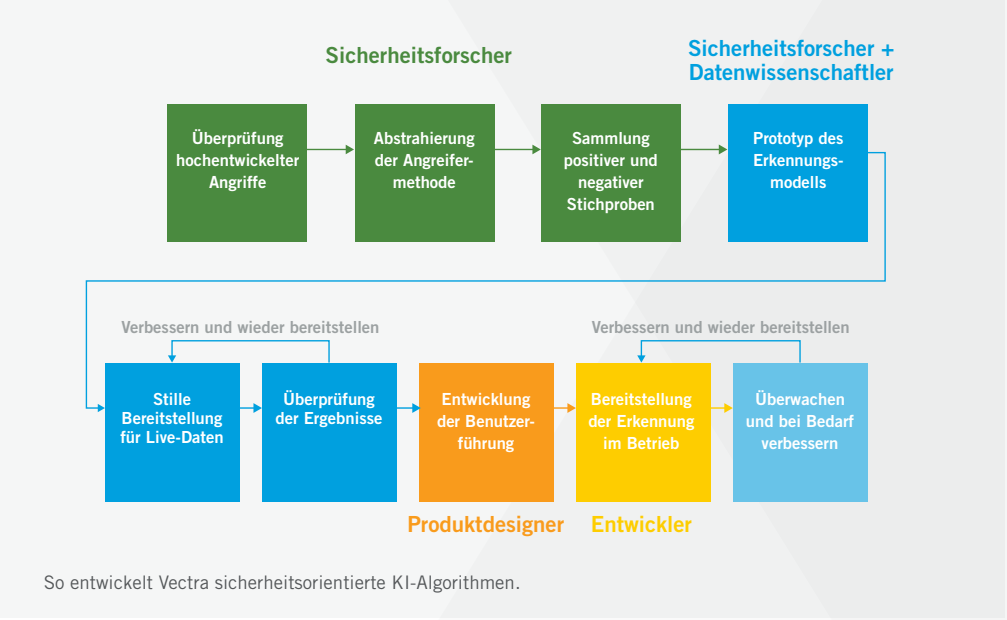
Vectra ist ein Wegbereiter für den sicherheitsorientierten Ansatz, mit dem sich Angreifermethoden im Netzwerk, in der Public Cloud sowie in SaaS-Anwendungen und Identitäten finden lassen. In den nächsten Abschnitten erläutern wir, wie groß der Umfang der Abdeckung von Vectra ist und schauen uns den Entwicklungsprozess sowie das Modul an, das Erkennungen erfasst und generiert. Zudem erklären wir, wie einzelne Ereignisse zu verwertbaren Sicherheitszwischenfällen korreliert werden, und untersuchen die inneren Abläufe zweier Vectra-Erkennungen.

Erkennungsentwicklung bei Vectra

Vectra konzentriert sich bei Erkennungen ausdrücklich auf das Aufspüren von Angreifern und die Identifizierung von laufenden Angreifermethoden – wobei nicht nur außergewöhnliche Anomalien erkannt werden. Die abgedeckten Erkennungen werden von Sicherheitsforschern mit unterschiedlichen Werdegängen sowie von Datenwissenschaftlern entwickelt, die genau wissen, wie sich ein optimaler Mehrwert aus massiven komplexen Datensätzen schöpfen lässt. Während der letzten zehn Jahre haben diese beiden Gruppen in enger Zusammenarbeit einen Ansatz zur Entwicklung von Bedrohungserkennungen erarbeitet, der für alle Sicherheitsbereiche und Datentypen skaliert werden kann und Angreiferverhalten mit minimalen False-Positives effektiv aufspürt.

Die Vectra-Sicherheitsforscher sind während des gesamten Entwicklungsprozesses einer Erkennung anwesend. Sie leiten mit ihrer Arbeit den Prozess und überwachen sowie überprüfen die in der Praxis aufgespürten Angreifermethoden permanent. Die Forschung konzentriert sich nicht auf bestimmte Tools oder Angriffsgruppen, sondern auf die von Angreifern eingesetzten allgemeinen Methoden.

> Vectra ist ein Wegbereiter für den sicherheits-orientierten Ansatz, mit dem sich Angreifer-methoden im Netzwerk, in der Public Cloud sowie in SaaS-Anwendungen und Identitäten finden lassen.



So entwickelt Vectra sicherheitsorientierte KI-Algorithmen.

Schauen wir uns das an einem Beispiel an: Die Sicherheitsforscher stellen fest, dass die Cobalt Strike-Beacons bei Ransomware-Angriffen verwendet werden. Statt nun einzig und allein die Beacons zu untersuchen, abstrahieren sie die Aktionen der Technologie und schauen sich die *Kontrollmethoden* der Angreifer näher an. Durch die Fokussierung auf die abstrahierte Methode kann Vectra Erkennungen entwickeln, mit denen sowohl Tools abgedeckt werden, die diese Methode aktuell verwenden, als auch solche, die erst noch entwickelt werden.

Nachdem die Angreifermethode identifiziert wurde, stellen die Sicherheitsforscher eine Sammlung schädlicher und harmloser Beispiele zusammen. Stichproben von Schaddaten stammen von verschiedenen Stellen – zum Beispiel von Kunden, die freiwillig anonymisierte Metadaten zur Verfügung stellen, sowie von synthetischen Algorithmen zur Datenerzeugung, öffentlich dokumentierten Cyber-Zwischenfällen und Angriffen in unseren internen Laboren. Harmlose Stichproben stammen aus dem großen Datensatz anonymisierter Metadaten der Kunden von Vectra.

Anhand der Angreifermethode und den Begleitdaten entwickeln die Sicherheitsforscher zusammen mit den Datenwissenschaftlern einen Modellprototyp mit dem optimalen Schwellenwert zur Erkennung der Angreifermethode. Der Prototyp wird im stillen Betamodus bereitgestellt und läuft hinter den Kulissen. Dabei schickt er zusammenfassende Berichte von einem großen Stamm von Kunden zurück, die dem zugestimmt haben. Um die Effektivität des endgültigen Modells noch weiter zu steigern, schickt der Prototyp jedes Mal einen Bericht, wenn eine Angreifermethode beobachtet wurde und wenn etwas auftrat, das dieser ähnlich sah – d. h. Ereignisse, die direkt unterhalb des Schwellenwerts liegen. Die Meldung der Ereignisse direkt unter dem Schwellenwert ermöglicht den Datenwissenschaftlern, ihre Modelle weiter zu verfeinern und sicherzustellen, dass kein Verhalten übersehen wird. Die Modelle werden schnell iteriert, bis die strengen Qualitätsanforderungen für die Erkennung von Angreifermethoden in der Praxis erfüllt sind.

Im letzten Schritt der Erkennungsentwicklung geht es um den Aufbau einer angepassten Benutzeroberfläche, die alle Kontextinformationen zur identifizierten Angreifermethode zeigt und bei Bedarf weitere Informationen dazu gibt, was für die betroffenen Systeme als normal gilt. Danach werden die Modelle für den Produktivbetrieb bereitgestellt und melden Kunden Zwischenfälle. Die gleiche Prototyp-Pipeline, mit der Daten erfasst werden, wird nun dazu genutzt, die Wirksamkeit des Modells in der Praxis zu überwachen und bei Bedarf die Erkennung weiter zu verbessern.

Dank dieser ganzen Arbeit müssen die Modelle nicht regelmäßig optimiert werden und erkennen effektiv aktuelle sowie zukünftige Generationen der Angreifertools. Der sicherheitsorientierte Ansatz von Vectra überzeugt durch die Erkennung von Angreiferaktionen und nicht nur ungewöhnlichen Ereignissen.

Echtzeit-Streaming-Modul für verwertbare Ergebnisse

Bei der Erkennung spielt vor allem Zeit eine Rolle. Kommt es bei Warnmeldungen zu Verzögerungen, gibt dies den Angreifern Gelegenheit, ihren Angriff weiter auszubauen. Vectra-Algorithmen arbeiten mit gestreamten Daten und nicht mit regelmäßig erstellten Chargen von Daten. Dadurch ist es den Vectra-Erkennungen möglich, Angreifer ohne Verzögerungen aufzuspüren, sodass genügend Zeit zur Abwehr bleibt.

Auch die Größe der Umgebung spielt dabei eine Rolle, denn der Umfang der Unternehmensnetzwerke, Cloud-Bereitstellungen und SaaS-Dienste nimmt stetig zu, weshalb die Erkennungen von Vectra immer mehr Daten verarbeiten müssen. Das Echtzeit-Streaming-Modul von Vectra ist auch für internationale Großunternehmen geeignet, da es die nötigen Daten extrahiert und somit langfristiges Lernen unabhängig vom Datenumfang ermöglicht.

Die Effektivität der Algorithmen – insbesondere derer, die unüberwacht lernen – hängt erheblich davon ab, wie viele historische Daten ihnen zur Verfügung stehen. Werden die Erkennungen chargenweise mit Daten versorgt, können sie nur eine bestimmte Menge an Daten in einer bestimmten Zeit verarbeiten. Beim Streaming-Ansatz von Vectra extrahieren die Algorithmen die relevanten Informationen aus einem Ereignis und beziehen diese in neue Basislinien für Modelle ein. Durch das Lernen aus gestreamten Daten können Basislinien aus Daten und Millionen Ereignissen gewonnen werden, die über Monate gesammelt wurden, sodass die Zuverlässigkeit der Warnungen äußerst hoch ist.

Künstliche Intelligenz zur Bedrohungskorrelation

Die Vectra-KI identifiziert nicht nur einzelne Angreifermethoden, sondern korreliert diese auch und kann dadurch laufende Angriffe identifizieren, kategorisieren und priorisieren. Die Korrelation ist erforderlich, da Cyber-Angreifer für das Erreichen eines Endziels mehrere Aktionen in verschiedenen Bereichen ausführen. Ein spezieller Korrelationsalgorithmus analysiert das Verhalten in allen Konten, Hosts, Netzwerken sowie in der Cloud und kann dadurch einen klaren Hinweis auf einen Sicherheitszwischenfall geben.

Der Korrelationsalgorithmus ordnet den Verhaltensweisen anschließend feste Anker in Form von Konten oder Host-Rechnern zu.

So werden zum Beispiel festen Host-Rechnern in Netzwerk- oder hybriden Cloud-Umgebungen vorübergehende IP-Adressen anhand von Artefakten zugeordnet, die ein Algorithmus namens „host-id“ erkannt hat. Die Artefakte werden aus Netzwerk-Metadaten entnommen, die unter anderem Informationen wie Kerberos-Host-Principals, DHCP-MAC-Adressen und -Cookies sowie Daten aus API-Integrationen wie EDR, vCenter, Azure und AWS enthalten. Sobald Artefakte einem bestimmten Host-Rechner zugeordnet wurden, fließen diese Metadaten jedes Mal, wenn eine IP-Adresse mit einem bestimmten Artefakt registriert wird. Alle damit verbundenen Verhaltensweisen der Angreifer können dann dem benannten Host-Rechner und nicht nur der IP-Adresse zugeordnet werden.

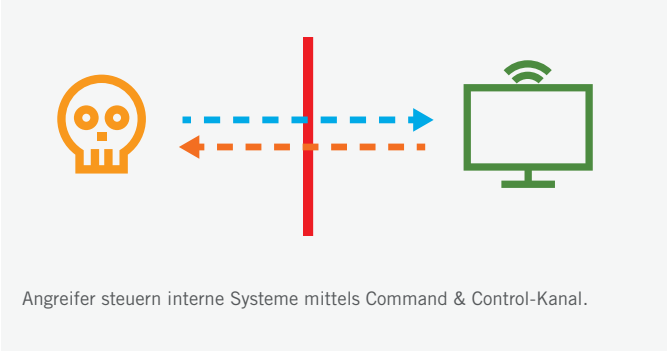
> Die Vectra-KI identifiziert nicht nur einzelne Angreifermethoden, sondern korreliert diese auch und kann dadurch laufende Angriffe identifizieren, kategorisieren und priorisieren.

Bei AWS gestaltet sich wiederum die Zuordnung schwierig, da Ereignisse auf der AWS-Steuerebene in Verbindung mit übernommenen Rollen und nicht den eigentlichen Anwenderkonten gespeichert werden. Eine Rolle kann von beliebig vielen Konten übernommen werden, allerdings ist es zur Reaktion auf einen Angriff wichtig zu wissen, welcher IAM- oder SAML-Anwender die Rolle tatsächlich übernommen hat.

Raffinierte Angreifer können es den Verteidigern dabei noch schwerer machen, indem sie mehrere Rollen hintereinander übernehmen und damit den Angriffsursprung verschleiern. Vectra kann die Rollen mithilfe einer speziell entwickelten Technologie namens Kingpin zurückverfolgen und die beobachteten Angriffe dem eigentlichen Anwender und nicht einer undurchsichtigen Rolle zuordnen.

Sobald Angreiferverhaltensweisen einem stabilen Indikator zugeordnet wurden, werden sie zusammen korreliert, um das zugrundeliegende Verhaltensprofil des Systems zu identifizieren, das anschließend die laufenden Bedrohungen kennzeichnet und priorisiert. Der Korrelationsalgorithmus dient dazu, die ausgeführten Aktionen der Vectra-Analysten und -Sicherheitsforscher bei der Untersuchung von Bedrohungen zu replizieren. Er ermöglicht zudem die Klassifizierung komplexer Angreiferszenarien wie „externe Bedrohungsakteure“ oder „Insiderbedrohungen auf Administratorebene“, die sofort untersucht werden sollten.

Reales Beispiel für KI-Erkennung: Verschlüsselte Command & Control-Kanäle



Angreifer steuern interne Systeme mittels Command & Control-Kanal.

Angreifermethode

Das Herzstück aller netzwerkbasierten Angriffe bildet ein Command & Control-Kanal (C&C). Angreifer mit Zugriff auf einen Host-Rechner laden Schadsoftware hoch, die sich mit einem externen Server verbindet. Obwohl die Verbindung anfangs durch den internen Rechner aufgebaut wird, enthalten die Antworten des externen Servers Anweisungen, die der infizierte Host-Rechner ausführt und den Angreifern somit ermöglicht, ihren Angriff fortzuführen.

Command & Control-Tools finden sich in käuflich erwerbbaren Angriffs-Frameworks wie Cobalt Strike und Metasploit sowie anderen, die Angreifergruppen selbst entwickeln. Diese Frameworks unterstützen alle die Verschlüsselung des Kanals sowie andere Techniken wie „Domain Fronting“ oder „Session Jitter“, die Angreifern helfen, der Erkennung zu entgehen.

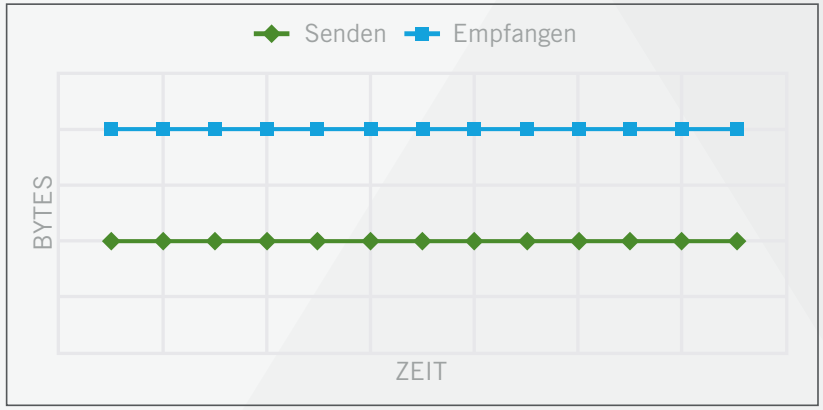
Vectra erkennt Command & Control-Kanäle unabhängig von der Verschlüsselung und anderen Umgehungstechniken.

Erkennungsmethodik

Vectra erkennt Command & Control-Kanäle unabhängig von der Verschlüsselung und anderen Umgehungstechniken. Ermöglicht wird dies durch den oben erwähnten sicherheitsorientierten Ansatz, der viele Probleme des mathematikorientiertes Ansatzes löst.

Als die Vectra-Sicherheitsforscher das Verhalten des Command & Control-Kanals abstrahierten, stellten sie fest, dass die eindeutigsten Anzeichen der Methode keine nebensächlichen Traffic-Elemente wie selten vorkommende Domains oder Benutzer-Agenten waren, sondern die Struktur des Netzwerk-Traffics über einen bestimmten Zeitraum.

Betrachten wir das untenstehende Beispiel eines externen Systems mit harmlosem Traffic.

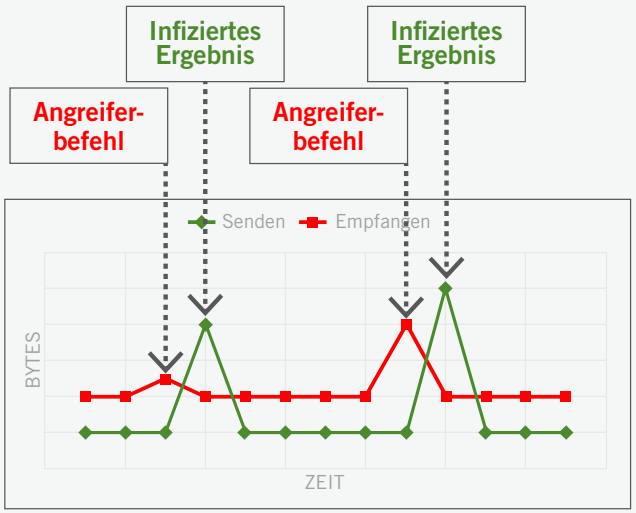


Beispiel für harmlosen Beacon-Traffic.

Harmloser Traffic

Das Beispiel zeigt den Traffic eines Host-Rechners, der über die Beacon-Funktion mit einem externen Server kommuniziert. Beacons sind eine häufig genutzte Netzwerkfunktion, die von Diensten wie Börsentickern, Chat-Apps und Werbetrackern verwendet wird. Sie ermöglichen lokalen und Remote-Systemen, synchron zu bleiben und miteinander zu kommunizieren. Die gleiche Funktionalität wird auch von schädlichen Command & Control-Kanälen genutzt.

Allerdings gibt es einen feinen Unterschied in der Art und Weise, wie ein Beacon erscheint, je nachdem, ob er von einem Börsenticker oder einem schädlichen Kanal verwendet wird. Die folgenden Daten stellen den Einsatz eines schädlichen verschlüsselten Tunnels dar.

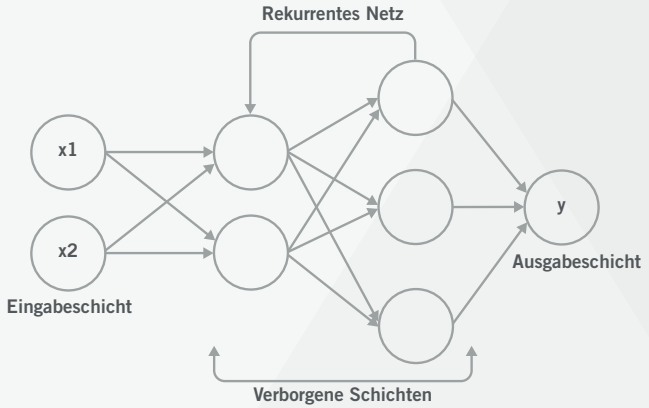


Beispiel für schädlichen Command & Control-Traffic.

Die Spitzen treten auf, wenn die Angreifer einen Befehl senden und das infizierte System darauf antwortet. Die erste Spitze tritt spontan in den empfangenen Bytes auf. Dahinter folgt umgehend die Antwort des infizierten Rechners.

Vectra-Datenwissenschaftler untersuchten diese Muster und konnten einen Ansatz erarbeiten, mit dem dieses Verhalten optimal identifiziert werden kann. Die Zeitreihendaten, die dieses Verhalten des Command & Control-Kanals beschreiben, weisen viele Ähnlichkeiten zu Daten auf, die in der Spracherkennung und bei der Verarbeitung natürlicher Sprache verwendet werden. Deshalb entschied sich das Team für den Einsatz eines Deep Learning-Modells.

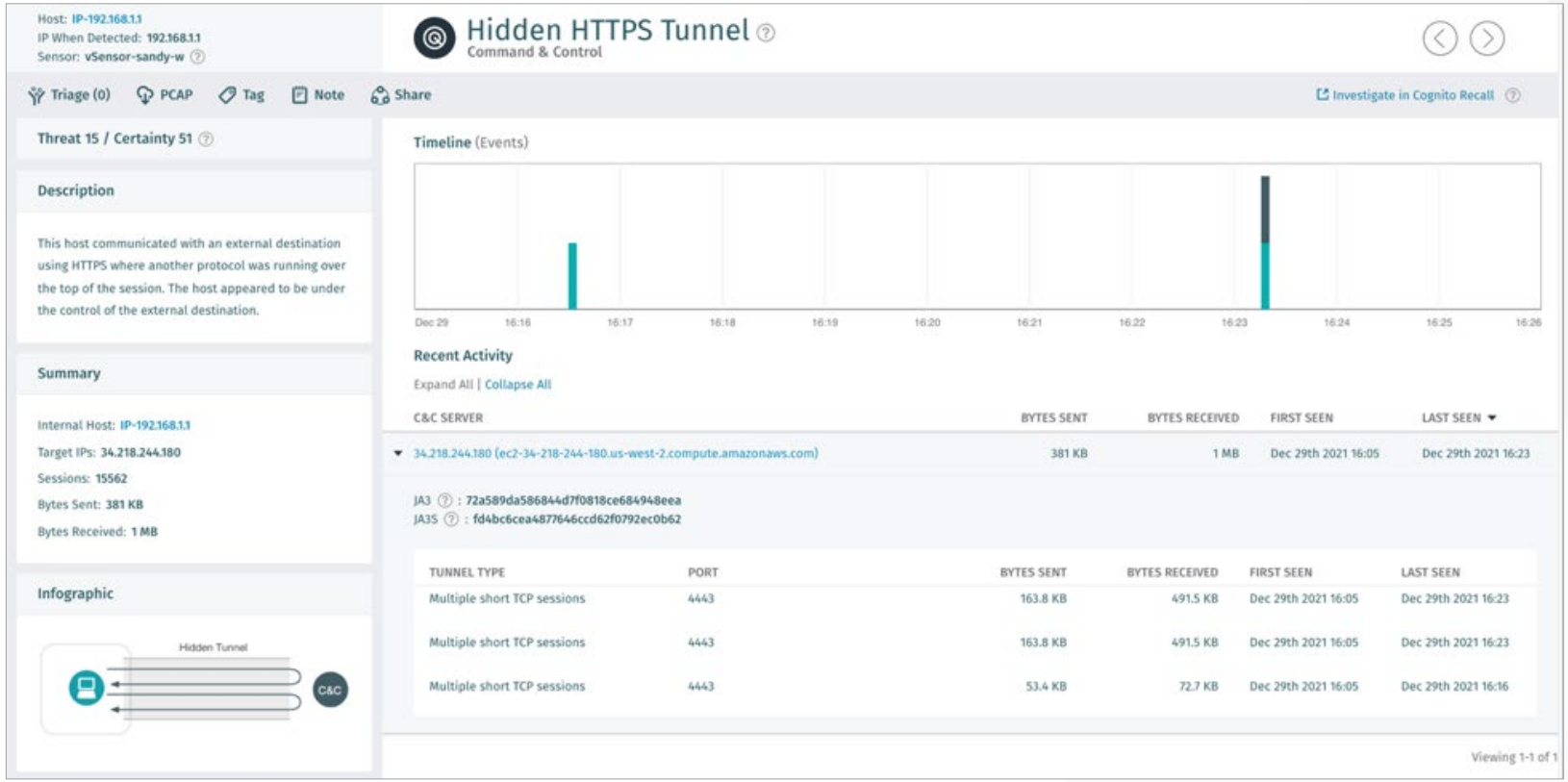
Zur Identifizierung des Angriffsverhaltens nutzt Vectra die spezielle Architektur eines rekurrenten neuronalen Netzes namens LSTM (Long Short-Term Memory, dt. langes Kurzzeitgedächtnis). Dieser Algorithmustyp kann besonders gut Ereignisse in mehreren verschiedenen Zeiträumen verstehen. Dies ist besonders wichtig, wenn man die Beschaffenheit der Gesprächsdaten des Command & Control-Kanals verstehen will. Das LSTM wird mit realen und algorithmisch erzeugten Beispielen trainiert. Der Datensatz umfasst dabei eine breite Palette an Szenarien, Tools, Konfigurationen und Umgebungen, die es dem Modell ermöglichen, das verallgemeinerbare Signal eines C&C-Kanals unabhängig vom eingesetzten Tool zu identifizieren.



Vectra nutzt rekurrente neuronale Netze, um schädliche Command & Control-Kommunikation von harmlosen Beacons zu unterscheiden.

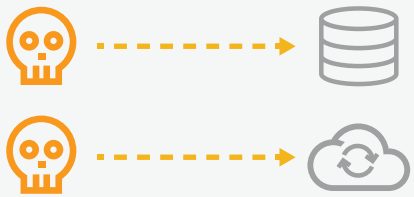
Es sollte zudem erwähnt werden, dass dieser algorithmische Ansatz durch die Art und Weise ermöglicht wurde, wie Vectra die Daten von Netzwerksitzungen formatiert. Vectra ist zwar in der Lage, Metadaten in einem Zeek-ähnlichen Format auszugeben, allerdings bietet der von Vectra eigens entwickelte Parser eine höhere Genauigkeit der Metadaten als das normale Zeek-Format und analysiert Netzwerk-Traffic innerhalb von Sekundenbruchteilen. Diese detaillierte Ansicht gibt einen klaren Überblick über alle Arten harmloser und schädlicher Kommunikation und ermöglicht den Vectra-Datenwissenschaftlern, Algorithmen zu nutzen, die eine breite Palette an Problemen bestmöglich abdecken können.

Dank der einzigartigen Herangehensweise an Metadaten sowie dem hochentwickelten Algorithmus kann Vectra Angreifer äußerst effektiv aufspüren. Da der Fokus auf den Kommunikationsdaten selbst und nicht auf oberflächlichen Anzeichen liegt, ist diese Methode unempfindlich gegenüber Änderungen an Tools sowie verschlüsseltem Traffic. Zudem sind durch das klare Verhaltenssignal keine Unterdrückungsfilter mehr nötig, die möglicherweise Domain Fronting oder gut getarnte Angriffsaktionen herausfiltern würden.



Vectra-Erkennung für einen verschlüsselten Command & Control-Kanal.

Reales Beispiel für KI-Erkennung: Missbrauch von Anmeldedaten im Netzwerk und der Cloud



Angreifer nutzen Anmeldedaten mit hohen Berechtigungen, um sich lateral zu bewegen und andere Taktiken im Netzwerk und der Cloud umzusetzen.

Angreifermethode

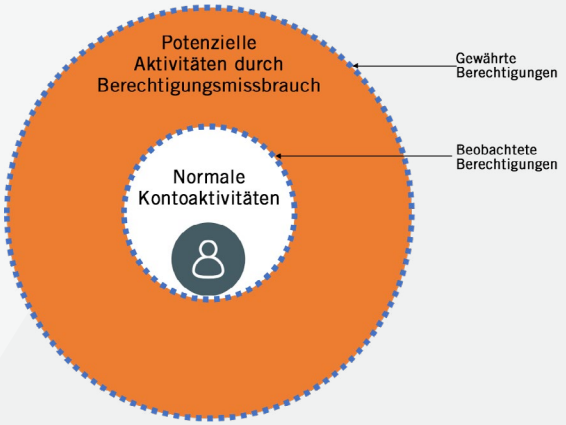
Durch Anmeldedaten mit hohen Berechtigungen erhalten Angreifer weitreichenden Zugriff auf Netzwerk- und Cloud-Ressourcen. Zudem bekommen sie damit Zugriffsrechte, ohne dafür Malware oder Exploit-Schaddaten hochladen zu müssen, die Spuren hinterlassen oder Schutzmaßnahmen auslösen könnten. Die Durchsetzung von Zugriffsanforderungen nach dem Least-Privilege-Prinzip kann die Folgen von Angriffen minimieren, doch die jüngsten Attacken zeigen, dass dies immer noch schwierig ist.

Um den Missbrauch gestohlener Zugangsdaten zu verhindern, ist es notwendig, solche Bedrohungen auch zu erkennen. Es ist eine besondere Herausforderung, zu erkennen, wann ein Konto von Angreifern gestohlen und missbraucht wird. Jede vom Angreifer ausgeführte Aktion ist durch die festgelegten Berechtigungen ausdrücklich erlaubt. Effektive Warnungen aufgrund neuer oder neuartiger Aktionen sind kaum möglich, da Anwender sich in dynamischen Umgebungen bewegen, in denen der Zugriff auf neue Ressourcen Teil ihrer täglichen Arbeit ist. Wenn ein Angreifer die Umgebung kennt, wird er versuchen, sich daran anzupassen, und keine Aufmerksamkeit auf sich ziehen, wenn er Aktionen ausführt, die für das jeweilige Konto nicht ungewöhnlich sind. Um den Missbrauch von Anmeldedaten effektiv zu erkennen, ist ein sicherheitsorientierter Ansatz nötig, der berücksichtigt, welche Absicht ein Angreifer mit gestohlenen Anmeldedaten verfolgt.

Erkennungsmethodik

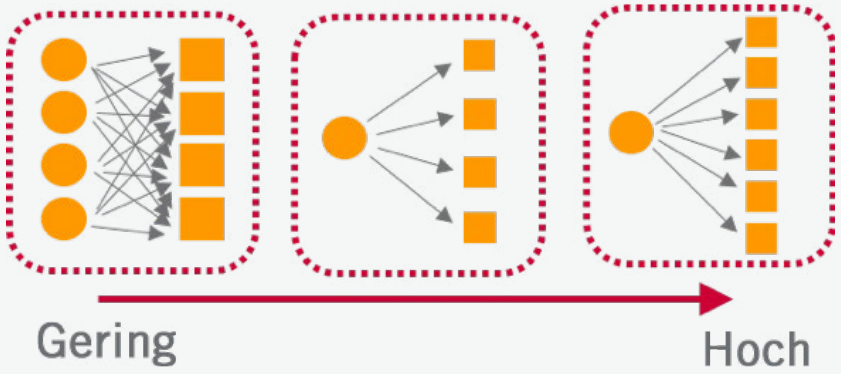
Vectra kann den Missbrauch gestohlener Anmeldedaten sowohl in Netzwerk- als auch in Cloud-Umgebungen erkennen. Im Kern des sicherheitsorientierten Erkennungsansatzes geht es darum, zu verstehen, wie Angreifer die gestohlenen Anmeldedaten nutzen. Für sie besteht der Wert von Anmeldedaten darin, auf Services und Funktionen zugreifen zu können, die in der Umgebung einen hohen Wert haben und hohe Berechtigungen erfordern.

Vectra-Sicherheitsforscher fanden heraus: Würde man alle tatsächlichen Berechtigungen jedes Kontos, Host-Rechners, Service und Cloud-Prozesses kennen, hätte man eine Übersicht aller vorhandenen wertvollen Ressourcen. Auch wenn die Grundsätze von *gewährten Berechtigungen* allgemein bekannt sind, veranschaulicht eine solche Darstellung die obere Grenze der tatsächlichen Berechtigungen eines Kontos im Vergleich zu den minimal erforderlichen Berechtigungen. Das Team der Sicherheitsforscher und Datenwissenschaftler von Vectra hat daher eine neue Methode entwickelt, um den Wert von Systemen in einer Umgebung basierend darauf darzustellen, was über einen bestimmten Zeitraum hinweg beobachtet wurde. Diese dynamische und grundlegende Betrachtung der Werte bezeichnen wir als *beobachtete Berechtigungen*. Der datenbasierte Überblick über Berechtigungen ermöglicht einen effektiven Zero-Trust-Ansatz zur Nutzung von Anmeldedaten ohne manuelle Konfiguration.



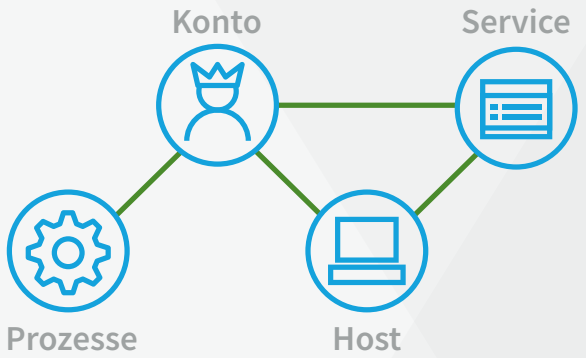
Beobachtete Berechtigungen ermöglichen einen Zero-Trust-Überblick über die normalen Berechtigungen, die Anwender für ihre Arbeit benötigen. Die Nutzung von Berechtigungen über das normale Maß hinaus erfordert eine genauere Prüfung.

Die Vectra-KI berechnet die *beobachteten Berechtigungen*, indem sie nicht die vom IT-Administrator festgelegten Berechtigungen, sondern die historischen Interaktionen zwischen den überwachten Objekten berücksichtigt. Der Umfang und die Spezifität der Zugangsrechte sowie der Nutzung tragen erheblich zu den errechneten Werten bei. Ein System, das auf mehrere Systeme zugreift, auf die üblicherweise von anderen Systemen zugegriffen wird, wird niedrige Berechtigungen haben. Im Gegensatz dazu hat ein System, das auf sehr viele Systeme zugreift, auf die aber von anderen Systemen nicht zugegriffen wird, einen hohen Berechtigungswert. Durch diesen Ansatz ist Vectra in der Lage, zwischen Domain-Administratorkonten und normalen Anwenderkonten zu unterscheiden.



Vectra lernt beobachtete Berechtigungen anhand des Anwenderverhaltens. Ein Konto, das auf viele häufig genutzte Services zugreift, hat geringere Berechtigungen als eines, das auf Services zugreift, auf die nur wenige andere Konten zugreifen.

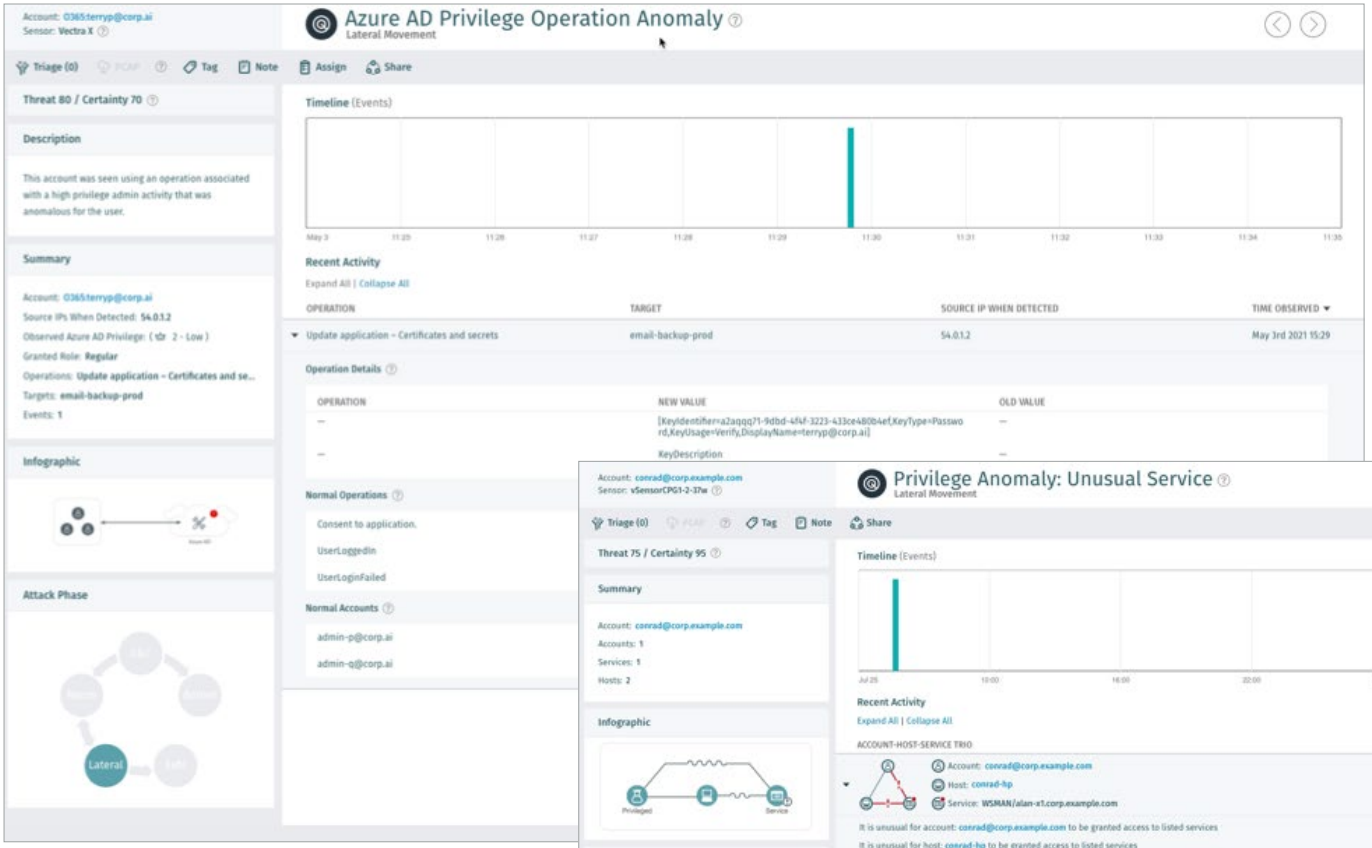
Nachdem die beobachteten Berechtigungswerte berechnet wurden, werden alle Interaktionen zwischen den Konten, Services, Hosts und Cloud-Prozessen auf eine Karte übertragen, um die regulären menschlichen Interaktionen zwischen den Systemen nachvollziehen zu können. Eine Reihe unüberwachter Algorithmen, die Berechtigungswerte berücksichtigen, identifiziert anschließend ungewöhnliche Fälle von Berechtigungsmissbrauch. Dabei werden eigens entwickelte Algorithmen zur Anomalieerkennung sowie HDBSCAN-Implementierungen (Hierarchical Density-Based Spatial Clustering of Applications with Noise) genutzt.

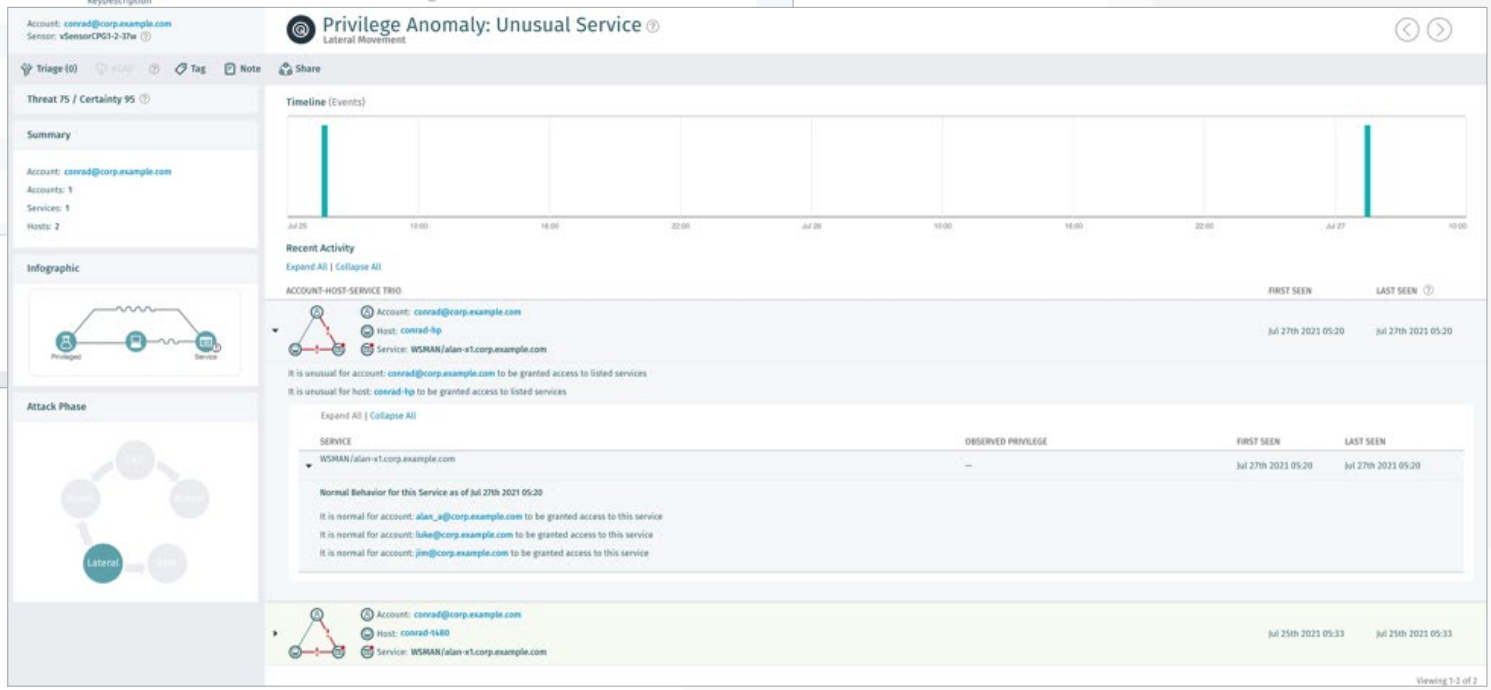


Um den Missbrauch von Anmeldedaten aufzuspüren, nutzt Vectra unüberwachtes Lernen, das beobachtete Berechtigungen sowie die Interaktionen zwischen Konten, Hosts, Services und Cloud-Prozessen berücksichtigt.

Der hochentwickelte sicherheitsorientierte Ansatz versetzt Vectra also in die Lage, den Missbrauch gestohlener Anmeldedaten sowohl in der Cloud als auch in lokalen Netzwerken zu erkennen. Die Kennzahl *beobachtete Berechtigungen* beschränkt die Erkennung auf anormale Aktionen, die relevant sind. Zudem ermöglicht sie eine höhere Wirksamkeit und Sensitivität als ein Ansatz, der diese kritische Betrachtung ignoriert.

Die Vectra-KI berechnet die *beobachteten Berechtigungen*, indem sie nicht die vom IT-Administrator festgelegten Berechtigungen, sondern die historischen Interaktionen zwischen den überwachten Objekten berücksichtigt.





Vectra-Erkennungen von Konten, die Berechtigungen missbrauchen.

Den Angreifern bei Innovation und Reaktion voraus

Die Angreifer erfinden immer neue Methoden – und die Verteidiger müssen mit ihnen mithalten. Vectra hat in den vergangenen Jahren kontinuierlich Innovationen hervorgebracht und damit eine der effektivsten Plattformen zur Bedrohungserkennung und Reaktion für lokale sowie Cloud-Assets entwickelt.

Dabei hat Vectra über 100 sicherheitsorientierte KI-Erkennungen entwickelt, die zahllose Bedrohungen in den Netzwerk- und Cloud-Umgebungen seiner Kunden identifiziert und somit Angreifer davon abgehalten haben, ihre Ziele zu erreichen. Jede Erkennung wurde mit einem tiefgehenden Verständnis der Vorgehensweise von Angreifern bei ihren Attacken sowie einer Reihe der fortschrittlichsten ML-Techniken auf dem Markt entwickelt. Insgesamt hält Vectra 33 Patente für die Technologien, die für diese Erkennungen eingesetzt werden.

Neben der Berichterstattung zu den patentierten Vectra-Technologien freuen wir uns, der meistzitierte Anbieter des von der NSA und MITRE entwickelten Frameworks MITRE D3FEND zu sein, das für Verteidiger Maßnahmen zum Schutz ihrer Umgebung definiert. Das D3FEND-Framework stellt schematisch dar, wie Verteidiger Angriffe stoppen und welche Maßnahmen sie gegen die im MITRE ATT&CK-Framework definierten Angreifertechniken ergreifen können. Insgesamt zitiert das D3FEND-Framework zwölf verschiedene Vectra-Patente, die im Rahmen von Gegenmaßnahmen für Verteidiger erwähnt werden.

Wir bei Vectra setzen uns dafür ein, die Welt sicherer und gerechter zu machen. Wir werden daher weiterhin sicherheitsorientierte KI-Technologie einsetzen, um Erkennungsfunktionen neu und weiter zu entwickeln, die Angreifer davon abhalten, ihre Ziele zu erreichen.

Weitere Informationen erhalten Sie unter info_dach@vectra.ai.

E-Mail: info_dach@vectra.ai vectra.ai/de