VECTRA
SECURITY THAT THINKS

ホワイトペーパー

Vectraを裏側で支えるAI

DATA SCIENCE
SECURITY RESEARCH

CLOUD NATIVE

AUTOMATED

目次

サイバー攻撃を検知・阻止するVectra®が、ビジネスを保護します。

Vectra®は、ハイブリッドおよびマルチクラウドエンタープライズ環境の脅威検知およびレスポンスにおけるリーディングカンパニーです。AIを駆使したVectraプラットフォームが、パブリッククラウド、アイデンティティ（ID）、SaaSアプリケーション、データセンター全体にわたる脅威を迅速に検知します。単に「通常と異なる」という理由だけでアラートを配信するのではなく、攻撃者の手法（あらゆる攻撃の中核となるTTP）を検知するためにAIを最適化しているベンダーは当社のみです。信頼度の高い脅威シグナルおよび明確なコンテキストをお客様のセキュリティチームに提供するため、早い段階で脅威に対処し、進行中の攻撃をいち早く食い止めていただけます。危険なサイバー脅威からの回復力を確保し、ランサムウェア、サプライチェーンの侵害、IDの乗っ取りをはじめとするサイバー攻撃による業務への影響を回避するために、世界中の組織がVectraを利用しています。詳細は当社ウェブサイト（vectra.ai/jp）をご参照ください。

はじめに

Vectraにとって「データサイエンス」は、北極星とも言うべき究極の指標です。当社は常に、データサイエンスとAIを適切に活用すればサイバー攻撃との戦い方を抜本的に変え、防御における優位性を確立できると考えてきました。しかしAIは各種各様です。本書ではまず、AIの定義やAIソリューションに関する主な用語の全体像を説明し、AIを適用した脅威検知における2つの代表的な手法の特性について検討します。最後に、AIを駆使して脅威をあぶり出すVectraの仕組みについて掘り下げて解説します。

AIに懐疑的な皆様も、AIの可能性に関心をお持ちの皆様も、ぜひ本書をご一読ください。

そもそもAIとは？

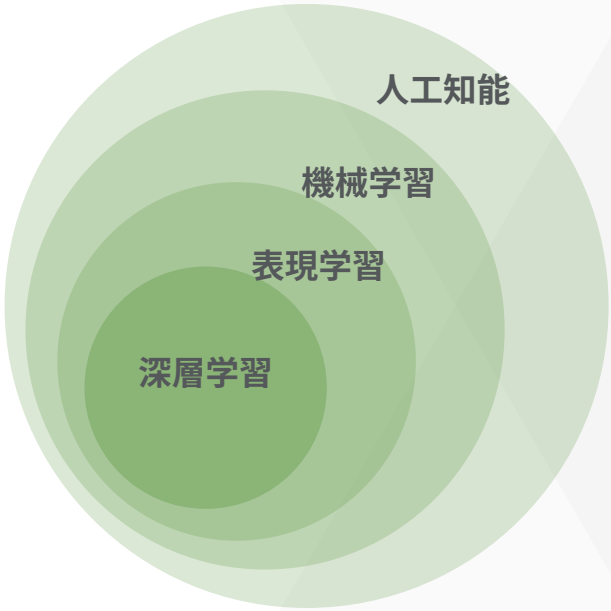
AIの定義

人工知能、機械学習、深層学習という用語はすべて同じ領域を指している、または同じような品質水準のもとで成り立っている、と誤解されがちですが、そうではありません。それぞれの用語に関連性はあるものの、具体的な意味は明確に異なります。各用語の範囲を理解することで、AIを活用するツールがどのようなものであるかをより深く理解できます。

人工知能 (AI)：「推論を自動化し、人間の心を近似できるシステム」をAIと呼びます。AIという広範で包括的な用語には、その下位領域である機械学習、表現学習、深層学習が含まれています。AIという用語は、明示的にプログラミングされたルールを利用するシステム、および膨大なデータを自律的に理解するシステムにも等しく当てはまります。後者（データから学習するシステム）は、自動運転車やバーチャルアシスタントなどの技術を支える基盤であり、AIの下位領域である「機械学習」に相当します。

機械学習 (ML)：機械学習は、AIの下位領域であり、人間による明示的な指示ではなく、データからの学習内容をもとにシステムが動作します。数十～数十億ものデータポイントを処理し、新たなデータインスタンスの最適な表現方法および応答方法を学習できます。

表現学習 (RL)：あまり馴染みがないものの、今日使われている様々なAI技術の中核を支えているのが、表現学習です。表現学習はAIの下位領域であり、データから学習した内容を抽象化して表現することが主な目的です。たとえば、大きさの異なる画像を固定長の数値リストに変換し、元の画像を蒸留して表現する、などの活用例があります。抽象化することで、主に下流システムが新種のデータを効率良く処理できるようになります。



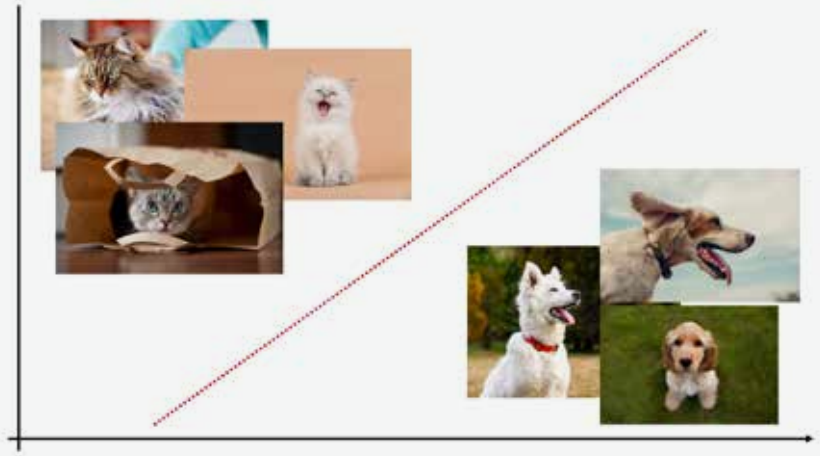
AIの下位領域の相関図
参照元：Goodfellow、Bengio、Courville共著「Deep Learning」（2016年出版）

深層学習 (DL)：ニューラルネットワークと関連づけられることが多い深層学習は、入力されたデータを、より複雑な形で表現する抽象化階層を構築します。このため、機械学習および表現学習の両方をカバーしています。人間の脳をもとに考案された深層学習モデルでは、入力内容に合わせてシナプスの重みを適応させるニューロン層を利用し、ネットワークの深層部分で、画像の分類や文章の翻訳などのタスクを簡素化する新たな抽象表現を学習します。深層学習は、特定の複雑な解析問題に対しては大きな効果を発揮しますが、インテリジェンスを自動化するための万能薬ではありません。

アルゴリズムによる学習手法の種類

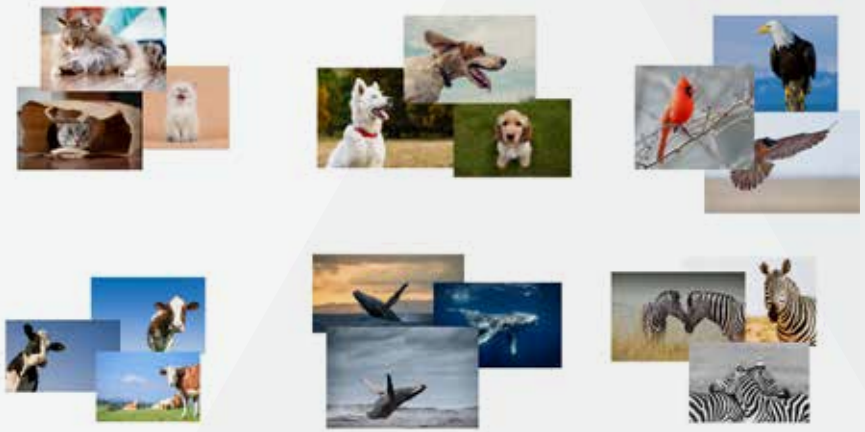
機械学習アルゴリズムの中核機能のひとつが、データインスタンスを異なるクラスに分類することです。そのための主な学習方法は数種類ありますが、主流となっている手法は、**教師あり学習**および**教師なし学習**の2つです。

教師あり学習：ラベル付きデータセットをもとに学習する機械学習モデルです。一度学習したモデルに新たなデータを与えると、ラベルを推測できます。以下の例をご覧ください。教師あり学習モデルに大量の犬や猫の画像を提供した後に新たな画像を与えると、ラベル（犬、猫）を予測して識別できるようになります。膨大な数のラベル付きトレーニングデータを集約した大規模コーパスを使ってモデルに学習させる必要がありますが、いったんトレーニングが完了すると、新たなデータインスタンスの汎化および適切なラベル付けをきわめて効果的に実行できます。



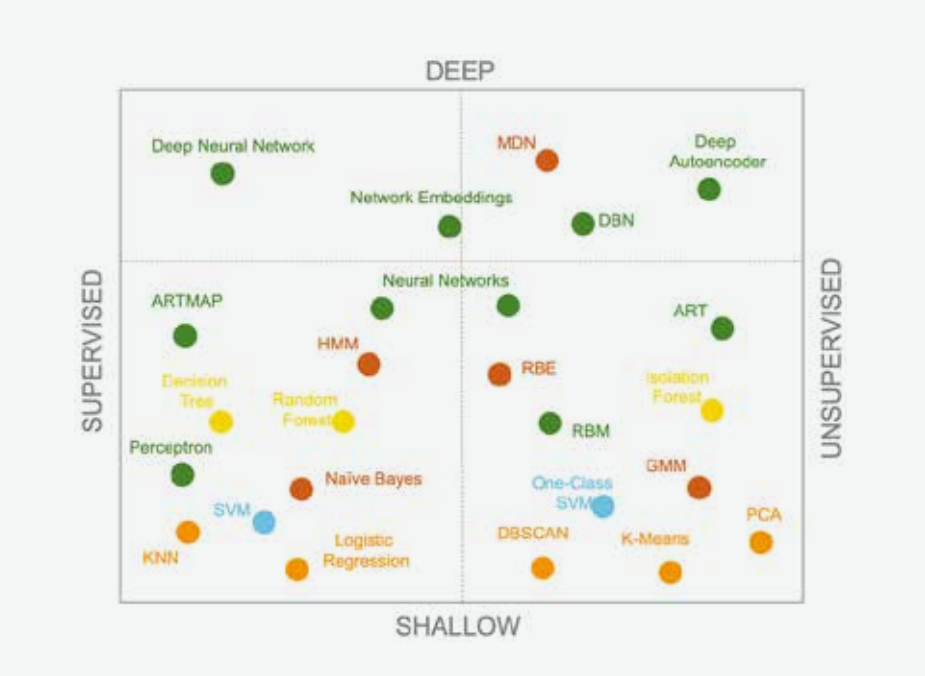
教師あり学習では、ラベル付きデータを使って各ラベルを特徴づける要素を識別します。学習を終えたモデルは、新たなデータをラベル付けすることができます。

教師なし学習：ラベルが付いていないデータセットをもとに学習する機械学習モデルです。与えられたデータの構造をモデルに学習させ、新たなデータが当該構造に当てはまるのか、当てはまる場合はどの箇所に該当するのかを判定させることができます。教師なし学習モデルのメリットは、先験的なトレーニングが不要なことです。このアプローチは、他とは異なるデータポイントを特定するための手法としては効果的ですが、特定した異常値や外れ値をすぐにラベル付けすることはできません。



教師なし学習では、ラベルなしデータの根底にある構造を学習します。学習を終えたモデルは、新たなデータが当該構造にどの程度適合しているのかを判定できます。

こうした広義のアプローチ（教師あり、教師なし）の中には、様々な学習アルゴリズムが存在します（下図参照）。研究者によって新たなアルゴリズムも次々と生み出されています。もっと面倒なことに、これらのアルゴリズムを組み合わせて、ますます複雑なシステムが形成される場合もあります。そこで疑問となるのが、「個々の問題を解決する際、データサイエンティストはどのようにして適切なアルゴリズムを選択しているのか」、もしくは「問題の種類にかかわらず、常にどのアルゴリズムよりも優れたアルゴリズムが存在するのか」という点です。



数多くの機械学習アルゴリズムが存在しますが、それぞれの長所・短所は対処すべき問題の種類によって異なります。

ノーフリーランチ定理

結局のところ、考えられるすべてのプロブレムステートメント（定義された問題）に対し、どのアルゴリズムよりも優れた成果を出せるアルゴリズムなど存在しません。これが「ノーフリーランチ（＝只飯はない）」と呼ばれる定理です。端的に言うと、個々の問題に特化した専用アルゴリズムの性能は、常に汎用アルゴリズムの性能を上回る、という意味です。問題の種類に合わせて異なるアルゴリズムが用いられるため、必要とされるアルゴリズムの数（前述）も増加の一途をたどっています。教師ありのニューラルネットワークが最大の成果を発揮する問題もあれば、教師なしの階層型クラスタモデルを用いるのが最適な問題もあります。

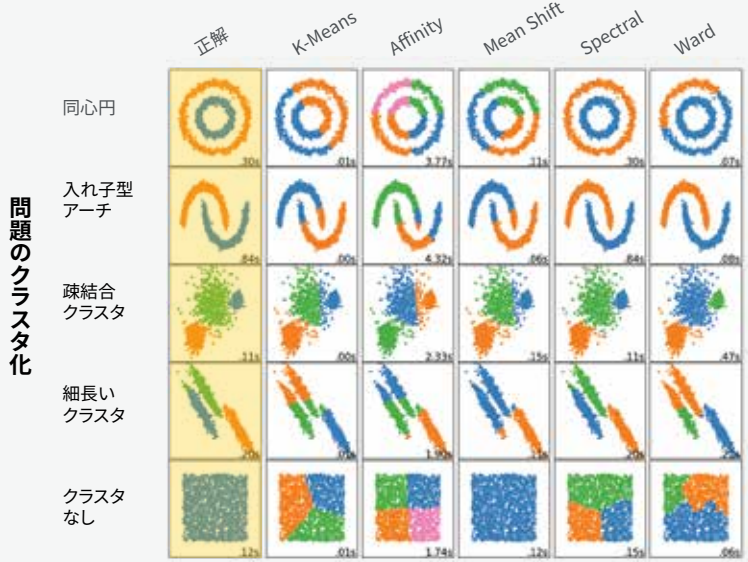
自動運転車の画像認識アルゴリズムを翻訳システムに適用することはできません。これらはいずれも目的特化型の選択肢であり、解決すべき問題やモデルの運用対象データに合わせて最適化されています。

ノーフリーランチ定理：すべての問題に対して優れた性能を発揮できる「万能の」アルゴリズムは存在しません。

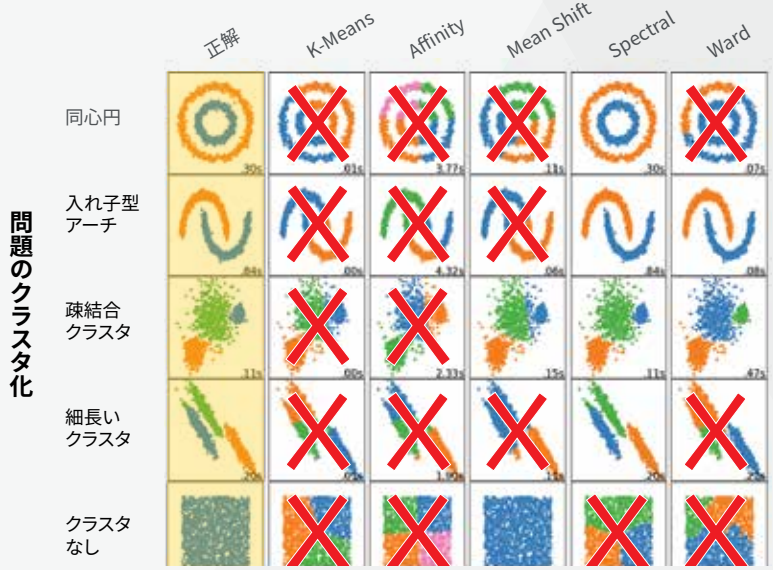
ジョブの内容に適したツールの選び方

では、データサイエンティストは適切なアルゴリズムをどのようにして選択しているのでしょうか？アルゴリズム選びは、サイエンスだけでなく芸術的な側面もあります。データを深く理解し、プロブレムステートメントに当てはめて考えることが、正しいアルゴリズムの選択につながります。忘れてはならないのが、「選択を誤ると、最適な結果が出ないだけでなく、完全に間違った結果が出る可能性がある」という点です。以下の例をご覧ください。

各データセットに対して選択されたアルゴリズムによって、出力結果に大きなバラつきが生じています。個々の問題に対して最適なアルゴリズムが存在しますが、何よりも「アルゴリズムの選択次第では、全く望まない結果が生じる」という点を覚えておいてください。そのため、問題の内容に最も適したアプローチを選択することが不可欠となります。



異なるデータセット（Y軸）に対する機械学習アルゴリズムの出力結果（X軸）の比較：黄色くハイライトされた部分が正解ラベルです。（scikit-learn.orgの公開資料を加工）



出力結果の比較：X印が付いている部分は、望ましくない結果につながる誤った予測。すべてのデータセットに対して有効なアルゴリズムはありません。（scikit-learn.orgの公開資料を加工）

出力結果の評価方法（何をもって「良し」とするのか）

データサイエンティストが適切なモデルを選択するためには、「モデルの成果測定方法」を決めておく必要があります。モデルの性能を検討する際は、**正解率 (Accuracy)** という言葉がよく用いられます。

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

（True Positives: 真陽性　True Negatives: 真陰性　False Positives: 偽陽性　False Negatives: 偽陰性）

正解率は有用な測定指標ではありますが、表面上の正解率が良好であってもモデルの本質的な性能を判断できない可能性もあります。「2つのラベル（AとB）をもとにデータを分類する」という問題を考えてみてください。ラベルAの発生確率がラベルBの1,000倍である場合、常に「ラベルA」と回答すれば99.9%の正解率をたやすく達成できます。つまり正解率は非常に高いものの、ラベルBに該当するものを正しく分類することは決してできません。つまり、「Bに該当する事例の特定」を重視する場合は、この指標では適切に測定できません。データサイエンティストにとって幸運なことに、重視する事例に対するモデルの成果を最適化し、測定するための指標は他にもあります。

そのひとつが**適合率 (Precision)** です。適合率とは、モデルが「陽性」と予測したラベル全体のうち、正しく予測できているものの割合を測定するための指標です。

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

「適合率を高めること」が目的のデータサイエンティストは、不正解（偽陽性）を大量に出さずにラベルを予測可能なモデルを構築するはずです。しかし適合率を使っても、対象事例におけるモデル予測の正誤判定はできません。この点で役立つ指標が、**再現率 (Recall)** です。

再現率は、対象となるラベルのインスタンス全体に対して、モデルが正しく「陽性」と予測できた割合を測定する指標です。

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

「再現率を高めること」が目的のデータサイエンティストは、重視するインスタンスにおける「陽性」を取りこぼさないようなモデルを構築するはずです。

適合率と再現率の両方をバランスよく組み合わせて追跡管理することで、モデルの出力結果を効果的に測定し、正しい運用に向けて最適化できます。

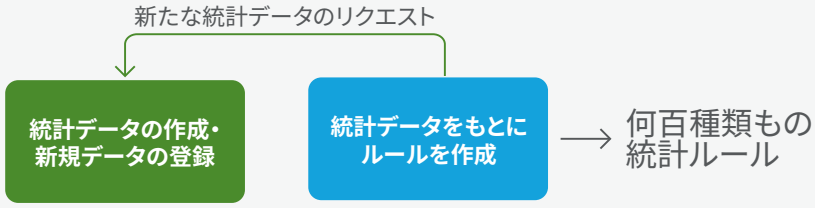
データサイエンティストが適切なモデルを選択するためには、「モデルの成果測定方法」を決めておく必要があります。

AIを適用した脅威検知

AIおよび様々なAI関連領域は、今日のエンタープライズ環境を狙う攻撃者の特定・阻止において重要な役割を担っています。AIを駆使してサイバーセキュリティ脅威を機動的に特定する2つの枠組み（数学ベースのアプローチ、およびセキュリティ主導型アプローチ）も登場しました。このセクションでは両者の違いについて掘り下げて解説すると共に、セキュリティチームにとって最適な結果を導き出せる「セキュリティ主導型AI」の仕組みについて説明します。

数学ベースのAI：脅威検知のアプローチとしては不十分

この枠組みでは、データサイエンティストが、外れ値や新規性の検知に特化したごく小数の汎用アルゴリズムを使って、シンプルな統計データのセットを作成します。続いてセキュリティリサーチャーが統計データを合算し、何百もの統計ルールを作成します。新たな統計が必要になると、前述の普遍的アプローチを使って統計データを追加作成します。このように検知対象データが次々と追加されるため、統計ルールの後処理として、明示的な除外フィルターを使った補完処理が頻繁に発生します（＝汎用アルゴリズムでは、最適な結果が得られない、というノーフリーランチ定理に帰結）。



数学ベースのアプローチを用いた検知モデルの開発。

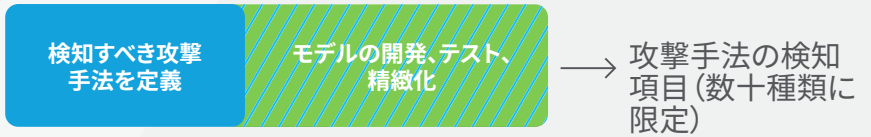
たとえばこれを「コマンド&コントロール（C2）チャネルの検知」に適用するとどうなるでしょうか。まずはデータサイエンスチームが、外部ドメイン全体のレアドメイン（希少性）に関する統計データを作成します。次にセキュリティリサーチチームにて、C2チャネル検知のきっかけとなる希少性判定の閾値を決定します。IoTデバイスが使用するドメインの多くが希少性の閾値を上回っている場合は、「IoTデバイスをすべて無視する」という除外フィルターを適用する必要があります。

アラートの件数が管理可能な水準に落ち着くまで、ユーザーエージェント、サブネットなどの属性に対して除外フィルターを追加適用します。除外ルールの適用によって、攻撃者が検知回避のために用いる手法までもがブロックされてしまうリスクがありますが、このような普遍的アプローチでは、除外フィルターを使わざるを得ません。

セキュリティ主導型AI：最小限の誤検知で最大限のカバレッジを実現

この枠組みは、「問題の定義（攻撃者の手法）」と「適切なモデルの選択」を密接に連動させたアプローチです。セキュリティリサーチャーは、単一のツールやエクスプロイトだけでなく、広義の攻撃手法を見極めたうえでプロブレムステートメントを策定します。次に、データサイエンティストが当該攻撃手法を特定するために最適なアルゴリズムを見極めます。一連の過程において、両者が密に協力しながらソリューションを共同開発します。これにより、数学ベースのAIで頻繁に検知される「表面上の異常値」にとどまらず、攻撃者が用いる手法を直接検知することができます。

セキュリティ主導型アプローチによる検知結果（再現率、適合率）は、数学ベースのアプローチによる検知結果を全般的に上回っています。また、攻撃ツールの変化に合わせてアラートの内容を柔軟に変え、検知項目の種類を絞り込むことができています。新たな攻撃手法が普及し始めると、セキュリティ主導型のAI検知プロセスが発動し、新たな検知項目が作成されます。高度なアプローチのため開発期間が多少長引く可能性はありますが、攻撃者の手法はそれほど頻繁には入れ替わりません。また、新たな攻撃手法は常に、検知モデルがフルカバー済の従来型手法と並んで表示されます。



セキュリティ主導型アプローチを用いた検知モデルの開発。

Vectra AIの仕組み

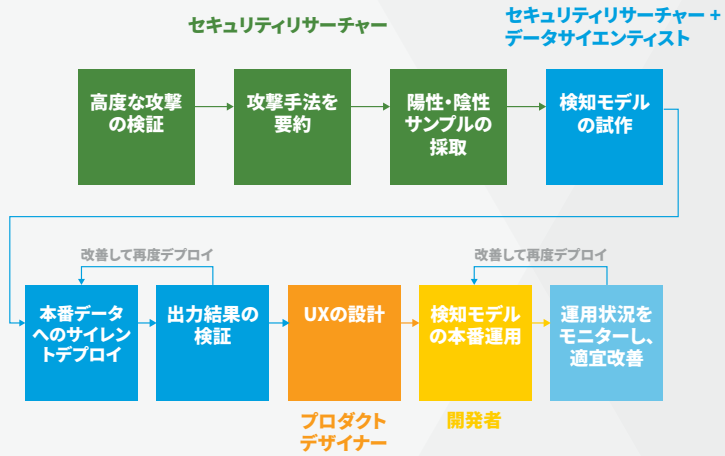
当社は、ネットワーク、パブリッククラウド、SaaSアプリ、IDに潜む攻撃者の手法をあぶり出すセキュリティ主導型アプローチを他社に先駆けて推進してきました。ここからは、当社独自の検知モデルのカバレッジ、開発プロセス、検知データ収集・アラート生成エンジンおよび、個々のイベントを結び、対処可能な「セキュリティインシデント」として相関付けする方法について詳しく解説します。また、2種類の検知事例をもとにモデルの内部処理の仕組みについても詳述します。

当社独自の検知モデル開発プロセス

Vectraの検知モデルは、特異な異常値を検知するだけでなく、攻撃者および現在進行形の攻撃に用いられている手法を特定することにも明確な重点を置いています。様々な経歴をもつセキュリティリサーチャーおよび、複雑で膨大なデータセットから価値を見い出すための方法を熟知したデータサイエンティストによって、検知モデルのカバレッジが拡充されています。この両者は10年以上にわたり、脅威検知モデル開発における密接な協力体制を築き上げてきました。様々なセキュリティ領域や、データタイプ全体に拡張可能な協業型アプローチを通じ、攻撃者の振る舞いを最小限の誤検知で、効果的に特定することが可能な検知モデルが出来上がります。

当社の検知モデル開発プロセスでは、初期段階から最後までセキュリティリサーチチームが関与しています。同チームの役割は、開発プロセスを先導し、実際に出回っている攻撃手法を常に監視・検証することであり、特定のツールや攻撃グループではなく、攻撃者が用いている手法全般に焦点を当てたリサーチを行うことです。

Vectraは、ネットワーク、パブリッククラウド、SaaSアプリ、アIDに潜む攻撃者の手法をあぶり出すセキュリティ主導型アプローチを他社に先駆けて推進してきました。



当社独自のセキュリティ主導型AIアルゴリズムの開発プロセス。

仮に、セキュリティリサーチャーが Cobalt Strikeのビーコン機能を使ったランサムウェアインシデントを発見したとしましょう。Cobalt Strikeのビーコン機能だけに着目するのではなく、当該テクノロジーを使って実行されるアクションを要約し、攻撃者による標的システムのコントロール方法を調査します。攻撃手法を俯瞰することで、当該手法を実行するために現在用いられているツールだけでなく、今後開発されるであろうツールをもカバーすることができます。

攻撃方法を特定したセキュリティリサーチャーが次に行う作業は、悪性データおよび良性データのサンプルを取り揃えたコーパスの収集です。悪性データのサンプルは、匿名加工したメタデータをお客様から任意で提供してもらう、合成データの生成アルゴリズムを使う、サイバーインシデントの公開資料を参照する、社内ラボ環境でサイバー攻撃を再現する、などの方法で収集します。良性データのサンプルは、お客様のメタデータを匿名加工した大規模データセットから採取します。

セキュリティリサーチャー側で攻撃手法と補完データの準備ができたら、攻撃手法を検知するために最適な閾値を設定した試作モデルを、データサイエンスチームと一緒に開発します。この試作モデルをサイレントベータ版としてデプロイして非公開環境で運用し、分母を拡大したお客様データ（任意提出）に対する運用結果のサマリーレポートを出力します。完成版の検知効率を最適化するために、試作モデルのレポートには観測されたすべての攻撃手法に加えて、攻撃者の手法と見られる事例（閾値をわずかに下回る事象など）もすべて記載します。疑わしい事例もあわせて捕捉することで、データサイエンティスト側でモデルのチューニングを重ね、いかなる振る舞いも逃さず検知するモデルを実現できます。その後、厳格な品質基準（現実世界での攻撃手法検知における性能基準）をクリアするまでモデルを短期集中的に反復運用します。

検知モデル開発の最終ステップは、特定された攻撃手法のフルコンテキストおよび、対象システムの通常状態に関する補足情報（適宜）を表示可能な専用UIの開発です。その後、モデルを本番環境にデプロイして運用を開始し、お客様向けのインシデントレポートを出力します。データ収集時に使った試作開発パイプラインを再利用して、現実世界におけるモデルの効率をモニタリングし、必要に応じて検知機能のさらなる改善を図ります。

このように様々な努力を重ねた結果、頻繁なチューニングを必要とせず、現行世代のみならず将来的な攻撃ツールに対しても有効なモデルが実現します。Vectraのセキュリティ主導型アプローチは、通常と異なるイベントだけでなく攻撃者のアクションを検知する際にも強みを発揮します。

実用的な結果を導き出すためのリアルタイムストリーミングエンジン

検知はスピードが命です。アラートの発動が遅れると、攻撃者にさらなる侵入を許してしまいます。Vectraのアルゴリズムは定期バッチではなく、ストリーミングデータを対象に運用されています。これにより攻撃者を遅滞なく特定し、充分な余裕をもって攻撃の進行を阻止することができます。

検知モデルの運用規模も重要です。拡大を続けるエンタープライズネットワーク、クラウドのデプロイ範囲、SaaSサービスのフットプリントに比例して、Vectraの検知モデルが処理するデータ量もますます増えているためです。Vectraのリアルタイムストリーミングエンジンは、長期学習モデルの構築に必要なデータをあらゆる規模で抽出可能なため、大規模グローバル企業でもご利用いただけます。

アルゴリズム（とくに教師なし学習を使用する場合）の有効性は、利用可能な過去データの量によって大きく影響されます。検知モデルをバッチ処理で実行すると、妥当な時間内で処理できるデータの量が制限されてしまいます。Vectraのストリーミング型アプローチでは、個々のイベントにおいて重要性の高い部分を、アルゴリズムが抽出し、検知モデルの新たなベースラインとして取り込みます。学習対象がストリーミングデータのため、数ヵ月単位のデータ、数百万件のイベントをもとにベースラインを構築し、最高品質のアラートを生成することができます。

人工知能を活用した脅威の相関分析

Vectraでは、個々の攻撃手法を特定する際だけでなく、攻撃アクションを相関付けする際にもAIを適用し、現在進行形の攻撃の特定、分類、優先度判定に役立てています。サイバー攻撃者は、異なる領域で複数のアクションを実行しながら最終目標に向かって侵攻を続けるため、こうした相関付けが必要となります。相関付けに特化したアルゴリズムが、ネットワークとクラウド環境全体にまたがるアカウントやホストすべてを対象に振る舞い分析を実行し、セキュリティインシデントのシグナルを明確に表示します。このアルゴリズムでは、基点となるアンカー（アカウントやホストマシン）に振る舞いを結びつけることで相関性を分析します。

たとえば「host-id」と呼ばれるアルゴリズムが観測したアーティファクトをもとに、ネットワークやハイブリッドクラウド環境における過渡的な (transient) 状態のIPを固定ホストマシンに結びつける、などがその一例です。アーティファクトは、ネットワークメタデータ（Kerberosホストプリンシパル、DHCP MACアドレス、クッキーなどの情報）や、API連携先（EDR、vCenter、Azure、AWSなど）から収集します。該当するホストマシンとアーティファクトの相関性を確立しておけば、メタデータが流入するアーティファクトおよび関連する攻撃者の振る舞いと共にIPが発現する度に、IP単体ではなく該当するホストマシンとの相関性を裏付けることができます。

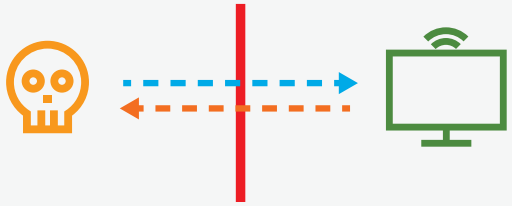
AWS環境では、本来のユーザーアカウントではなく、別のアカウントが代理で引き受けたロール (Assumed Role) にひもづけられたイベントがAWSコントロールプレーンに記録されるため、相関付けに関するAWS固有の課題が発生します。代理アカウントは、無数に存在する可能性がありますが、実際にロールを引き受けたIAMユーザーやSAMLユーザーを特定しなければ攻撃に対処できません。

Vectraでは、個々の攻撃手法を特定する際だけでなく、攻撃アクションを相関付けする際にもAIを適用し、現在進行形の攻撃の特定、分類、優先度判定に役立てています。

ロールを連鎖させて身元を隠蔽しようとする高度な攻撃者もいるため、防御側の対応がますます困難になる可能性もあります。Vectraでは、ロールの連鎖を遡って調査可能なテクノロジー「Kingpin」を独自開発しました。これにより、観測された攻撃を、属性不明のロールに対してではなく、本来のユーザーと結びつけることができます。

攻撃者の振る舞いを定点指標に結び付けたら、全体の相関関係を確立し、システムでプロファイル分析するための基本的な振る舞いの種類を特定します。続いて、進行中の脅威をシステム側でラベル付けして、優先度を振り分けます。相関アルゴリズムは、当社のアナリストおよびセキュリティリサーチャーが、脅威を調査する際のアクションを再現する目的で設計されました。また、相関アルゴリズムを使って、外部の脅威アクターや管理者権限をもつ内部脅威をはじめとする高度な攻撃シナリオを分類し、その場で検証することもできます。

AIを活用した検知事例：暗号化されたコマンド&コントロールチャネルの検知



標的組織内のシステムを掌握するために、外部のコマンド&コントロールチャネルを使って攻撃。

攻撃手法

コマンド&コントロールチャネル (C2) は、ネットワークベースのあらゆる攻撃の中核的役割を果たします。ホストマシンにアクセスした攻撃者は、外部サーバーへの接続手段となる悪質ソフトを実装します。接続要求を出すのは内部マシンですが、外部サーバーからは「感染したホストマシンに実行させる命令」を含む応答が返されるため、攻撃をさらに進めることができます。

コマンド&コントロール攻撃のツールは、市販の攻撃フレームワーク（Cobalt StrikeやMetasploitなど）に搭載されていますが、高度な攻撃グループが独自に開発することもあります。これらのフレームワークはいずれも、C2チャネルの暗号化およびドメインフロンティング、セッションジッタなどの攻撃手法をサポートしているため、検知の網をすり抜けることができます。

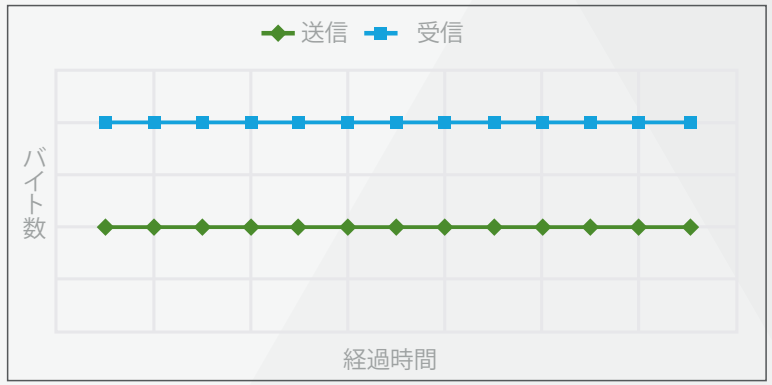
Vectraは、暗号化などの検知回避手法の影響を受けることなくコマンド&コントロールチャネルを検出します。

検知手法

Vectraは、暗号化などの検知回避手法の影響を受けることなくコマンド&コントロールチャネルを検出します。このようなカバレッジを実現できる理由は、前述のセキュリティ主導型アプローチによるものです。また、ご覧のとおり、数学ベースのアプローチによる問題解析で見られる多くの課題をクリアすることもできます。

当社のセキュリティリサーチチームがC2チャネルの振る舞いを要約する過程で、攻撃手法を最も端的に示す指標は、トラフィックの状況的側面（レアドメイン、ユーザーエージェントなど）ではなく、ネットワークトラフィックの経時的変化であることが判明しました。

代表例として、外部システムによる良性トラフィック（下図）を考えてみましょう。

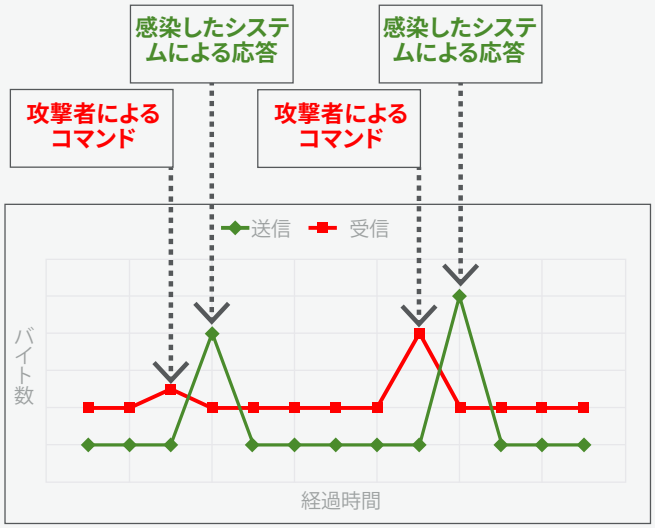


良性トラフィック

ビーコンデータ送受信における良性トラフィックの例。

この例は、ビーコンを介して外部サーバーと通信するホストマシンの特徴を示しています。ビーコンとは、株価情報の取得、チャットアプリ、追跡型広告などのサービスで、ローカルシステムとリモートシステムの同期や接続を維持するために、ごく一般的に使われているネットワーク機能です。このビーコン機能が悪意をもったコマンド&コントロールチャネルに使用されることもあります。

しかし、株価情報に適用される場合と、悪意のあるチャネルに適用される場合ではビーコンの発現形態が微妙に異なります。こちらの「悪質な暗号トンネル」のデータをご覧ください。

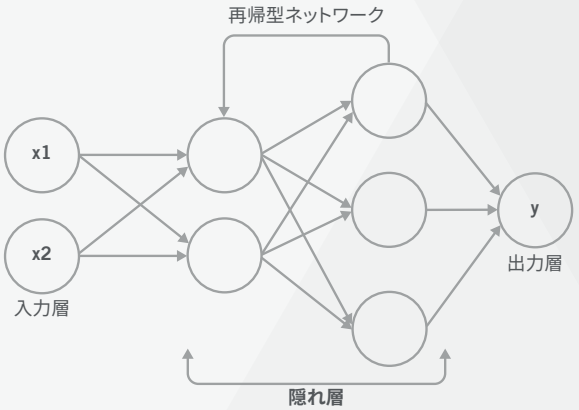


コマンド&コントロールを介した悪質なデータ送受信トラフィックの例。

突出箇所（スパイク）のパターンにご注目ください。攻撃者がコマンドを送信し、感染したシステムが応答した時に発生しています。最初のスパイクは、受信データのバイト数が自然に増えたものですが、その直後に感染したマシンが応答する際に次のスパイクが発生しています。

当社のデータサイエンティストがこれらのパターンを検証し、同様の振る舞いを特定する最適なアプローチを考案しました。コマンド&コントロールチャネルの振る舞いを特徴づける時系列データは、音声認識や自然言語処理に用いるデータとの類似点が多いことから、「深層学習モデルを応用できる」という判断に至りました。

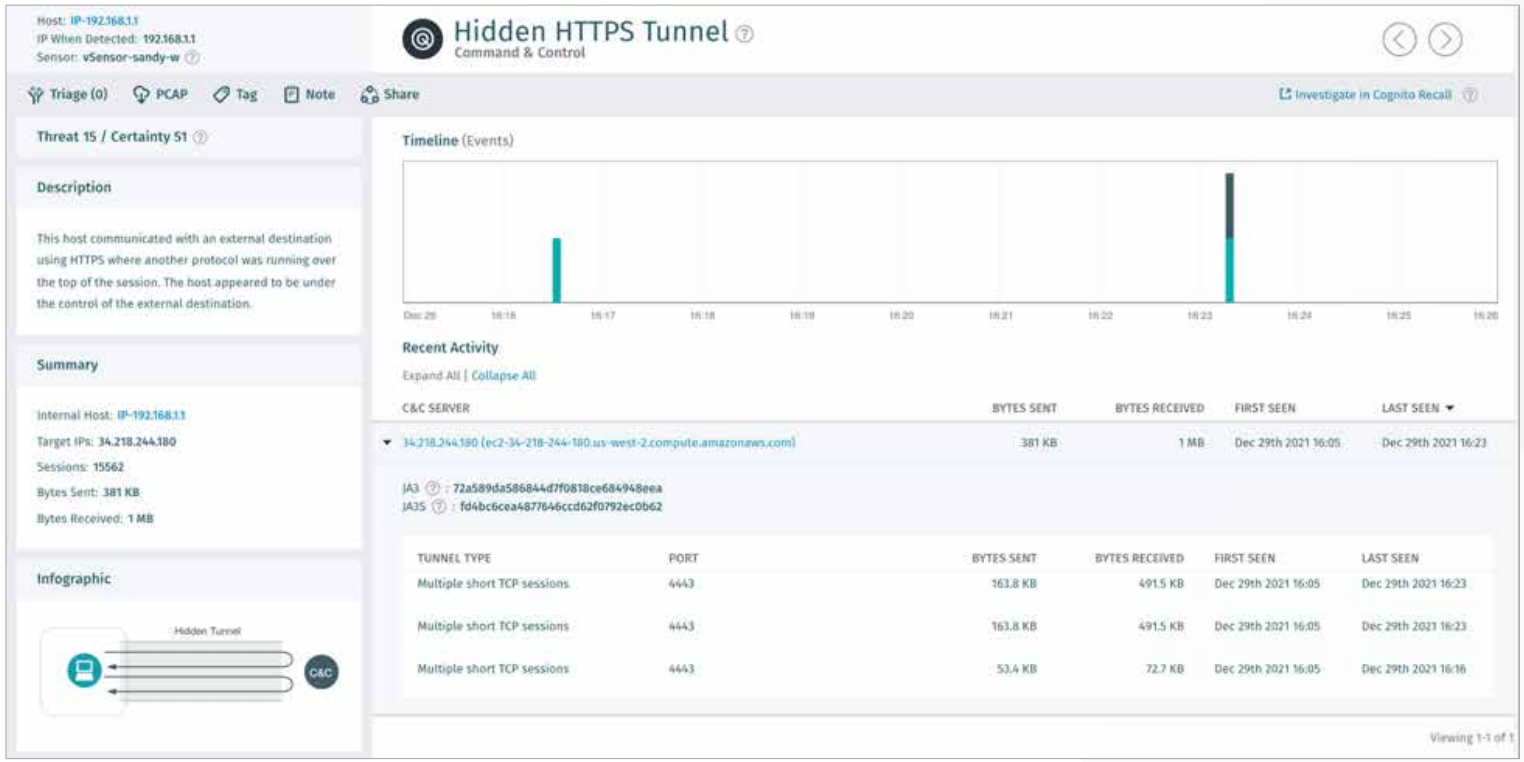
Vectraでは、再帰型ニューラルネットワークの一種である長・短期記憶（LSTM）アーキテクチャを用いて攻撃者の振る舞いを特定しています。この種のアルゴリズムは、コマンド&コントロールを介した通信データの特性をくまなく把握するための鍵となる「時間軸をまたいだイベント」を、高い精度で認識できます。LSTMのトレーニングデータには、実際のサンプルおよびアルゴリズムが生成したサンプルを使います。様々なシナリオやツール、構成設定、環境を幅広くカバーするデータセットを使ってモデルを訓練するため、コントロールチャネルで実際に使用されているツールの種類を問わず、汎化可能なシグナルを特定できます。



再帰型ニューラルネットワークを用いて、悪意あるコマンド&コントロール通信と良性のビーコン通信を識別するVectraのモデル。

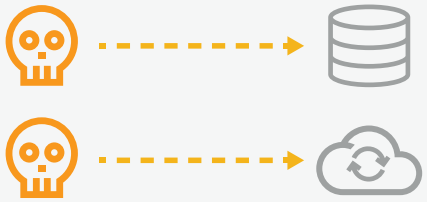
さらに、このアルゴリズムを使ったアプローチは、Vectra独自のネットワークセッションデータ形式の定義をもとに実現したことも注目に値します。Vectraでは、Zeekフォーマットに準じたメタデータの出力も可能ですが、Vectra独自の構文解析機能「カスタムパーサー」（ネットワーク通信の構文解析処理速度：1秒未満）を適用することで、標準的なZeekフォーマットよりも高い信頼度でメタデータを解析できます。データのビューが細かい粒度で表示されるため、あらゆる種類の良性通信および悪質な通信をクリアに可視化できます。当社のデータサイエンスチームは、このビューをもとに、多種多様な問題を最適な形でカバーするアルゴリズムを採用できます。

Vectra独自のメタデータおよび高度なアルゴリズムを活用したアプローチにより、攻撃者を効果的に特定可能なモデルが実現しました。表面的なシグナルではなく通信データそのものに焦点を当てる、という方針のため、攻撃ツールの入れ替わりやトラフィックの暗号化に対しても柔軟に対応できます。また、振る舞いに関する明確なシグナルを用いることで除外フィルターが不要となるため、攻撃用のフロントチャネルや身を潜めて行動する攻撃者までもブロックしてしまうというリスクを低減できます。



暗号化されたコマンド&コントロールチャネルをVectraが検知。

AIを活用した検知事例：ネットワークとクラウド環境における特権認証情報の乱用



特権認証情報を使って、ネットワークやクラウド環境内部を横移動する、もしくは別の攻撃戦術を実行する攻撃者。

攻撃手法

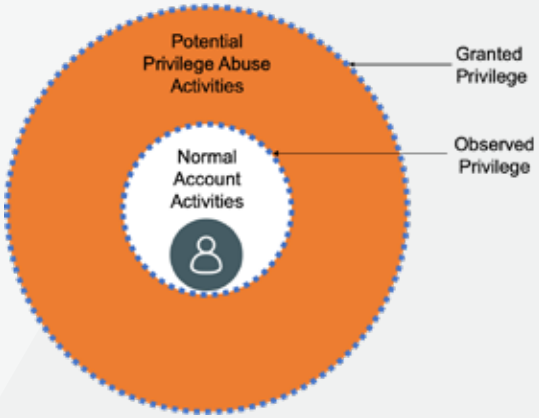
特権認証情報を取得した攻撃者は、ネットワークおよびクラウド環境内の様々なリソースにアクセスできるようになります。マルウェアやエクスプロイト用ペイロードを使う必要がないため、攻撃の証跡が残ったり侵入防止アラームが発動したりすることもありません。「最小権限にもとづくユーザーアクセス」のルールを実装すれば、一部の攻撃を回避できる可能性はありますが、最近の攻撃を見ると、このルールを厳密に運用することがいかに難しいかがよくわかります。

「盗難にあった認証情報の乱用による不正アクセス」を防止するには、「乱用」が発生した時点で、インシデントを検知する必要があります。しかし攻撃者によるアカウント情報の盗用は、一筋縄では検知できません。攻撃者が実行するすべてのアクションは、当該アカウントに付与された権限によって明示的に許可されているためです。新規またはこれまでにないパターンのインタラクションを検知する、などの概念にもとづくアラート配信では充分な効果が得られないでしょう。なぜならユーザーは動的な環境に存在し、日々の業務を遂行するために当たり前のように新たなリソースへのアクセスを繰り返しているからです。標的組織の内情を調べ上げた攻撃者は、内部ユーザーの中に紛れ込み、盗用したアカウントのユーザーが日常的に行うアクションを実行し、疑いの目が向けられないようにします。特権認証情報の盗用を効果的に特定するためには、「攻撃者がこれらを使って何を達成しようとしているのか」を考慮したうえで、当該インシデントを適切に検知するセキュリティ主導型アプローチが必要になります。

検知手法

Vectraのセキュリティ主導型アプローチを用いると、ネットワークおよびクラウド環境における特権認証情報の盗用を特定できます。このアプローチの中核となるのが「盗難にあった認証情報を手にした攻撃者が取るであろうアクション」の解明です。攻撃者にとっての特権認証情報は、標的環境で価値が高く特権的とみなされるサービスや機能への橋渡し手段となります。

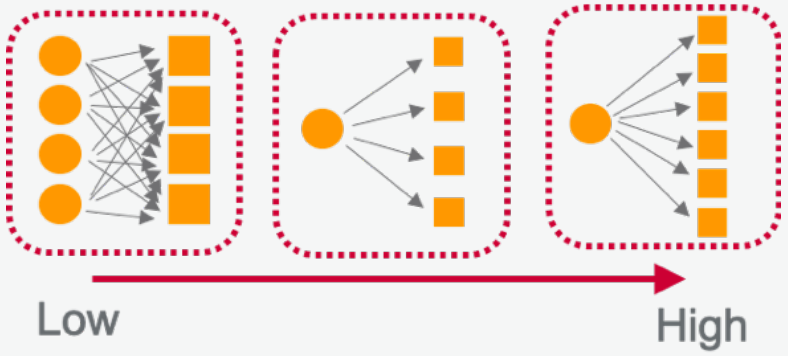
当社のセキュリティリサーチャーは、個々のアカウント、ホストマシン、サービス、クラウドオペレーションに付随する実際の権限を把握できれば、その中に存在する価値の高いリソースを俯瞰したマップ（全体像）ができあがるという点に着目しました。「Granted Privilege（付与された特権）」という概念は、すでに定着しています。しかし最小権限の原則と違い、この概念では、対象物にとって本当に最上位の特権を特定することができません。そこで当社のセキュリティリサーチチームとデータサイエンスチームは、環境内のシステムの価値を表現するために、「時系列的に観測された価値」に着目するという新たな方法を見い出しました。価値の内容を平面図形式で動的に表現するこの手法は「Observed Privilege（観測された特権）」と



Observed Privilege：ユーザーの業務遂行に必要な通常の権限をゼロトラストの観点で可視化します。これを超える権限が使用されている場合は、さらなる精査が必要です。

呼ばれます。観測したデータをもとに特権を可視化するため、手作業による設定なしで、認証情報の乱用を効果的に検知可能なゼロトラストアプローチを実践できます。

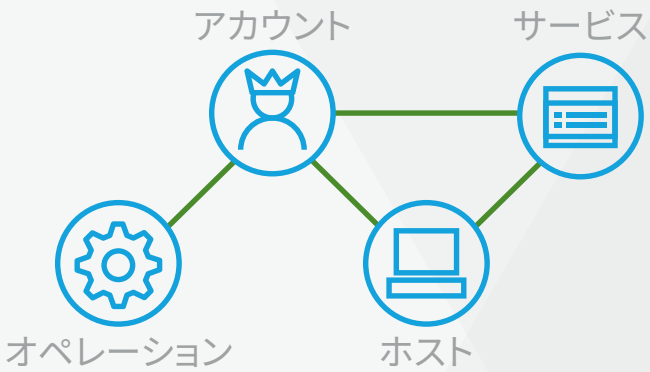
VectraのAIは、IT管理者が定義した権限そのものではなく、追跡対象のエンティティ間で発生したインタラクション履歴を考慮してObserved Privilegeのスコアを算出します。スコアは、アクセスおよびアカウント使用範囲の広さや特異性によって大きく左右されます。あるシステムが、他のシステムから日常的にアクセスされているシステムに接続する場合は、低い権限スコアを付与します。反対に、他のシステムから滅多にアクセスされないシステムに数多く接続している場合は、



Vectraは、ユーザーの振る舞いをもとにObserved Privilegeのレベルを学習します。よく使われるサービスに数多くアクセスするアカウントは、あまり使われないサービスにアクセスするアカウントよりも権限レベルが低くなります。

高い権限スコアを付与します。このアプローチにより、Vectra側でドメイン管理者アカウントと一般的なユーザーアカウントを区別します。

Observed Privilegeのスコアを算出したら、アカウント、サービス、ホスト、クラウドオペレーション同士で発生する、すべてのインタラクションをマッピングし、システム間の定常的なインタラクションの実績を把握します。次に、教師なし学習のアルゴリズム群が権限スコアを考慮し、特権乱用を示す異常なケースを特定します。

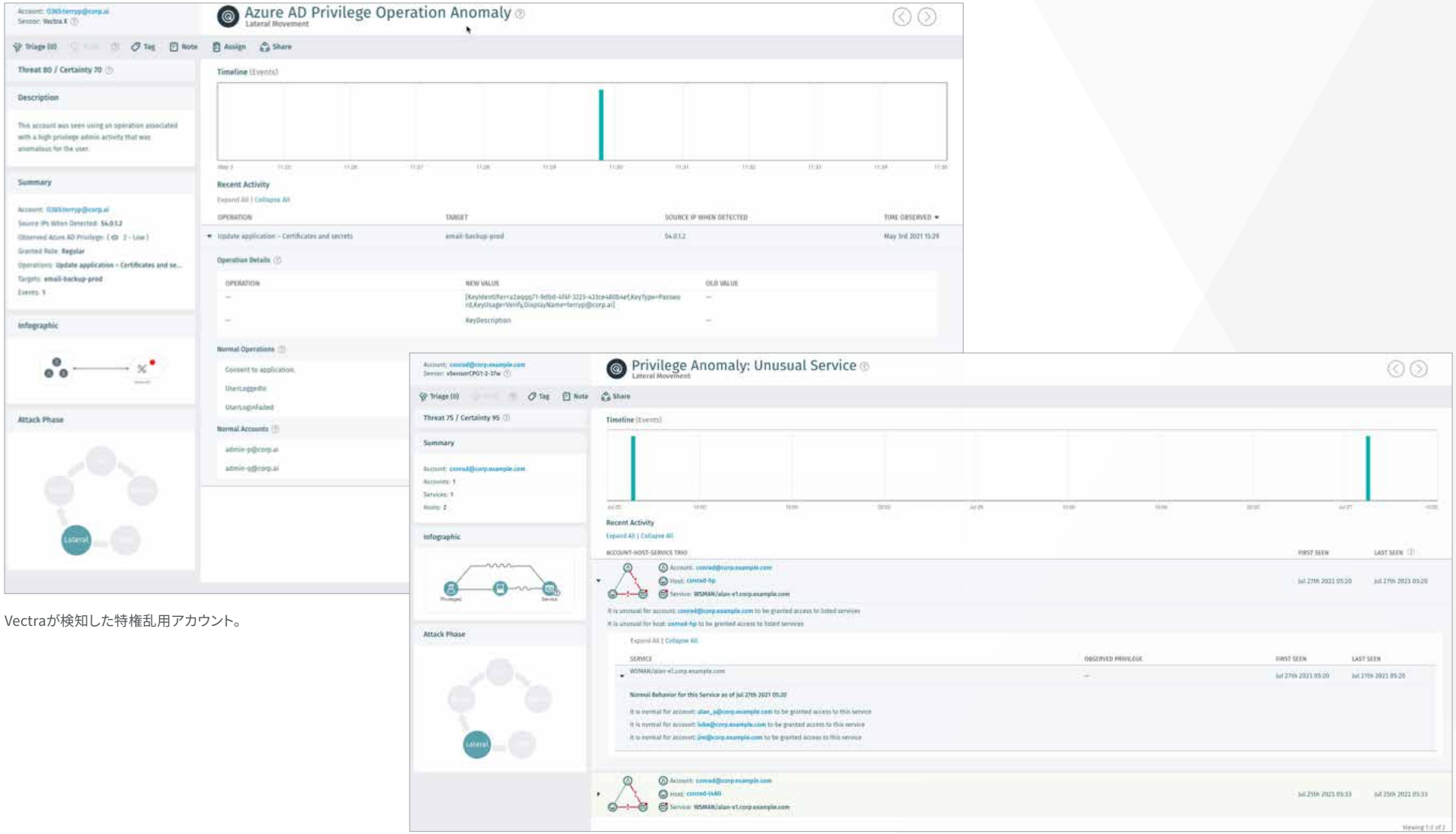


Vectra上で稼働する教師なし学習が、Observed Privilegeおよびアカウント、ホスト、サービス、クラウドオペレーション間で発生するインタラクションを考慮し、特権乱用のケースを特定します。

これらは、HDBSCAN（誤検知を伴うアプリケーションの密度ベースの階層型空間クラスタリング）を実装した異常検知のカスタムアルゴリズムを用いて行います。

こうした高度なセキュリティ主導型アプローチによって、クラウドおよびオンプレミスネットワーク環境における認証情報の盗用を特定することができます。Observed Privilegeは、「通常とは異なる重大なアクションの特定」を重視した指標です。そのため、この重要な視点を考慮しないアプローチよりも高い適合率および再現率を達成できます。

VectraのAIは、IT管理者が定義した権限そのものではなく、追跡対象のエンティティ間で発生したインタラクション履歴を考慮してObserved Privilegeのスコアを算出します。



Vectraが検知した特権乱用アカウント。

まとめ

革新的な手法を次々に生み出す攻撃者から組織を防御するには、攻撃者に後れを取らずに革新を続ける必要があります。Vectraは長年にわたってイノベーションを重ね、オンプレミスおよびクラウド環境の資産に対する脅威に対し、最も効果的に検知およびレスポンス可能なプラットフォームを開発してきました。

当社は、これまで100種類以上のセキュリティ主導型AI検知モデルを開発し、お客様のネットワークやクラウド環境で数々の脅威を特定して、攻撃者の目標達成を阻止するために役立ててきました。それぞれの検知モデルは、攻撃者の振る舞いに関する深い見識および、入手可能な最先端の機械学習の手法をベースに構築されています。当社は、AIを活用した検知技術に関して合計33件の特許を取得しています。

当社の強みは、特許取得済み技術による広範なカバレッジだけではありません。当社はアメリカ国家安全保障局 (NSA) およびMITRE社のフレームワーク（組織環境の防御に必要な対策を定めた、通称MITRE D3FEND）で最も頻繁に引用されるベンダーでもあります。D3FENDはMITRE ATT&CKフレームワークに明文化されている攻撃手法をベースにしたフレームワークで、それぞれの項目に該当する攻撃を阻止するための方法および攻撃手法への対抗手段が掲載されています。D3FENDには、当社が発案した合計12種類の特許技術が引用されており、防御対策の参照事例として紹介されています。

Vectraは、世界をより安心かつ安全な場所にするために全力を注いでいます。この理念をもとに、セキュリティ主導型AIを駆使したイノベーションおよび、攻撃者の目標達成を阻止するための検知機能の開発を続けていきます。

お問い合わせ：info-japan@vectra.ai　vectra.ai/jp