CyberRes

ホワイトペーパー

ArcSight Intelligence の特長 (パート 3)

目次

はじめに 1
ArcSight Intelligence脅威検出プラットフォーム 1
より高速で、より正確な脅威検出のための機械学習 2
リスクを最大限に可視化するためのデータ取得 2
正確な結果を得るには正しいデータが重要 3
経験に基づく正しいデータの認識 4
学習による何百万もの個別ベースライン作成 4
教師なし機械学習での「ユニークノーマル」について 5
原理に基づく信頼できるシステム 5
自動的に変化に適応する自己学習システム 6
分析モデルによる異常の検出 6
確率密度関数による異常測定 7
ルールとしきい値のバイナリ性を含む確率的アプローチ 7
統計的証拠を使用した優先度の高い脅威リードの生成 8
統計を適用したエンティティリスクスコアの作成 8
誤検出の防止 8
断片化されたセキュリティエコシステムの統合 9
既存のセキュリティツールの強化と自動化 9
次のステップ 10

はじめに

サイバーセキュリティの世界では、AI、機械学習、脅威ハンティング、脅威検出の高速化、高度な分析、ビッグデータ、ディープラーニング、ニューラルネットワークといったバズワードを元に売り上げを競うベンダーが多いため、こうした言葉を見聞きするたびに懐疑的になっても不思議ではありません。

この ArcSight Intelligence ホワイトペーパーシリーズでは、多数のセキュリティ分析製品やユーザーエンティティ行動分析 (UEBA) 製品の評価結果を踏まえて、お客様から当社独自であると認められた重要な 3 つの特長について詳しくご紹介します。

当社の **AI** も、当社の**アプローチ**も、当社の**アーキテクチャ**も、他社とは異なる独自性を備えています。

このホワイトペーパーでは、**ArcSight Intelligence のアーキテクチャの特長**をご紹介します。ビッグデータを使用するセキュリティソリューションは多数ありますが、ほとんどの場合、使用されているのはビッグデータコンピューティングではなくビッグデータストレージです。この 2 つのテクノロジーの違いは、脳に記憶を保存することと、脳を使って問題を解決することの違いに似ています。適切なオープンソーステクノロジーを組み合わせて AI の頭脳を作りあげるのは容易なことではありません。先進的なデータサイエンスチームが当社の機械学習を評価すると、多くの場合「XYZ ソリューションで同じようなことを試したが、同じ結果は得られなかった」という反応が返ってきます。このホワイトペーパーでは、サイバーセキュリティを分析できる AI の頭脳を実現する ArcSight Intelligence 独自のビッグデータアーキテクチャと数学的アルゴリズムについて、そしてニューラルネットワークと確率密度が全体像に含まれる (または含まれていない) かについて詳しく説明します。

このホワイトペーパーで は、ArcSight Intelligence のアーキテクチャの特長をご紹介します。ビッグデータを使用するセキュリティソリューションは多数ありますが、ほとんどの場合、使用されているのはビッグデータコンピューティングではなくビッグデータストレージです。

ArcSight Intelligence 脅威検出プラットフォーム

ArcSight Intelligence の UEBA プラットフォームは、Apache Hadoop コンピューティングアーキテクチャを使用し、大規模な教師なし機械学習アルゴリズムを実行するためにゼロから構築されました。このアルゴリズムでは、ログファイルからエンティティ (個々のユーザー、マシン、IP アドレス、Web サーバー、プリンターなど) を抽出し、これらのエンティティに関するイベントを観察して、どのような行動が通常行われる、または予期されるかを判定します。分析プロセスで得られた新しい情報を収集して、以前に観察した行動や動的に測定された統計的ピアグループと比較して、潜在的リスクを評価します。

アーキテクチャはネイティブに Apache Hadoop 上に構築されているため、非常に拡張性の高いプラットフォームとなっています。セキュリティ情報とイベント管理 (SIEM) システム、エンドポイントソフトウェア、その他のセキュリティツールからデータを取得し、そのデータを、ビッグデータテクノロジーを使用した高効率のストレージと超高速のトランザクション機能で処理できます。ストレージの問題を避けるため、ArcSight Intelligence はメタデータを収集し、ログは再作成しません。ただし、セキュリティチームは時系列的フレンジック調査のために、時間枠を決めてメタデータを記録することができます。生データの代わりにメタデータを使用することは、一般データ保護規則 (GDPR) のプライバシー要件への対応にも役立ちます。

より高速で、より正確な脅威検出のための機械学習

以下のセクションでは、Hadoop エコシステムを使用して ArcSight Intelligence が実装した、大規模にデプロイできる独自のコンビネーションの数学的アルゴリズムについて説明します。ArcSight Intelligence の脅威検出機能には、データ取得、ベースラインの作成、異常検出、脅威リードの 4 つの段階があります。

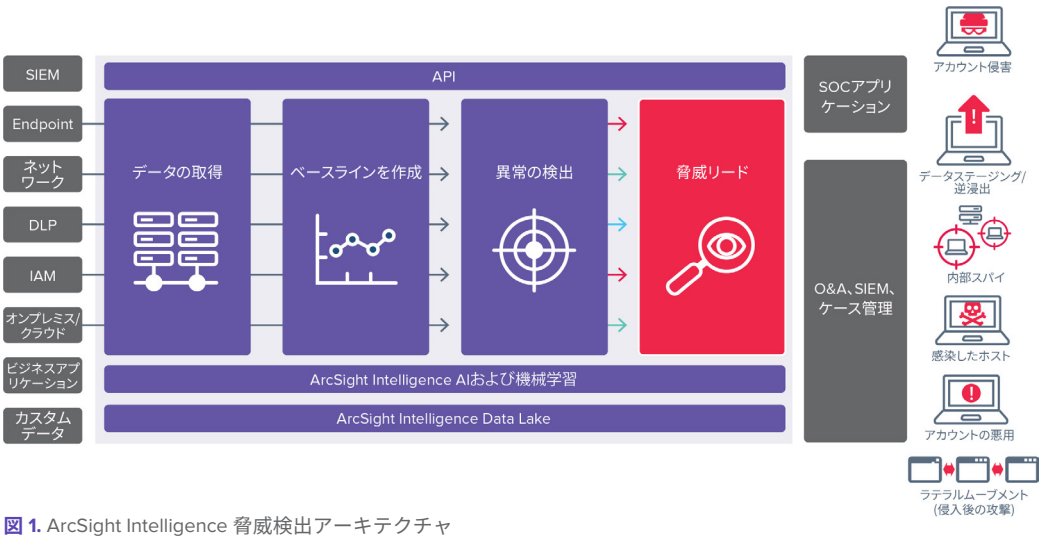

図 1. ArcSight Intelligence 脅威検出アーキテクチャ

リスクを最大限に可視化するためのデータ取得

最初の段階であるデータ取得では、分析に使用できるように、データを転送して変換します。この段階の目的は、分析用のデータセット数を最大限に増やし、リスクの可視化を行うことです。データソースがより広範で多岐にわたるほど、より総体的な企業の図を作成でき、より正確なリスク評価が可能になります。ただし、機械学習は、既存のルールとしきい値ベースのセキュリティツールでは不可能な、意味のあるインサイトをもたらすため、データセットが 1 つであっても貴重です。

脅威検出のユースケースによって、それぞれ異なるデータソースのセットを使用できます。以下の表は、6 種類の脅威の検出の分析モデルに使用できる多数のオプションを示したものです。6 種類の脅威とは、アカウントの悪用、データステージング / 盗難、感染したホスト、内部スパイ、ラテラルムーブメント (侵入後の攻撃) です。

アカウントの悪用	セキュリティ侵害を受けたアカウント	データステージング/盗難	感染したホスト	内部スパイ	ラテラルムーブメント(侵入後の攻撃)
認証ログ	認証ログ	エンドポイントログ	Webプロキシログ	認証ログ	認証ログ
ディレクトリサービスログ	ディレクトリサービスログ	ディレクトリサービスログ	ディレクトリサービスログ	IPリポジトリログ	IPリポジトリログ
エンドポイントログ	オペレーティングシステムログ	IPリポジトリログ	エンドポイントログ	ファイル共有ログ	ファイル共有ログ
オペレーティングシステムログ	ファイル共有ログ	プリンターログ	NetFlowログ	オペレーティングシステムログ	オペレーティングシステムログ
ファイル共有ログ	VPNログ	Webプロキシログ		リソースアクセスログ	リソースアクセスログ
VPNログ	リソースアクセスログ	プリンターログ		エンドポイントログ	エンドポイントログ
リソースログ	IPリポジトリログ	NetFlowログ		NetFlowログ	NetFlowログ
IPリポジトリログ		Eメールログ			
プリンターログ					

表 1. データセットのオプション

ArcSight Intelligence のビッグデータアーキテクチャは、データにとらわれることなくデータを取り込み、ストリーミングし、解析し、分析モデルが実行できる 1 つ以上のサポートされた意味データタイプに正規化します。

正確な結果を得るには正しいデータが重要

データの形式やサイズは多岐に渡るため、データの取得は容易ではありません。ログファイルはさまざまなシステムから入ってきて、それぞれ異なる形でデータを生成するため、正確な分析を行うためには、異なるデータセットを互いに互換性のあるものにすることが重要です。たとえば、1:00 p.m のタイムスタンプが誤って 1:00 a.m と解釈されたらどうなるでしょうか。これらの 2 つのタイムスタンプに関連するユーザーの行動は大きく異なります (1:00 p.m は従業員の業務中である可能性が高いですが、1:00 a.m では状況が異なります)。

こういった非常に重要で時間のかかる問題が、機械学習やビッグデータプロジェクトの大きな課題の 1 つです。クリーンなデータは正確な分析にとって非常に重要なものであるため、データサイエンティストやデータエンジニアは、勤務時間の最大 80% をデータ分析用にデータをクリーニングするために費やします。クリーンでないデータを使用すれば、分析結果もクリーンなものではなくなります。

ArcSight Intelligence のビッグデータアーキテクチャは、データにとらわれることなくデータを取り込み、ストリーミングし、解析し、分析モデルが実行できる 1 つ以上のサポートされた意味データタイプに正規化します。これは、セキュリティデータを扱った豊富なデータ分析の経験があって初めて再現できるものです。ArcSight Intelligence のアーキテクチャにはこの経験が組み込まれています。このアーキテクチャは、Apache NiFi、Kafka、Flume でのビッグデータ取り込み、ストリーム、解析、データ正規化のために、主要なビッグデータコンポーネントを活用するために特別に開発されたものです。

経験に基づく正しいデータの認識

これらの 3 つのコンポーネントは、機械学習に必要なデータを提供するため、相互機能するように構築されています。これらの 3 つのオープンソースコンポーネントにより、必要な変換がすべて実行されます。ログファイルやデバイスの時系列データは、(SIEM、ビッグデータレイク、または生ログファイルなど、その保存場所に関わらず)ArcSight Intelligence がサポートする 1 つ以上のデータタイプに取り込まれます。データソース出力列とデータタイプ入力変数の間の関係を定義するこの「マッピング」により、一般的なデータソースと新しいデータソースを、正しく定義されたデータタイプ (ネットワークデータタイプ (NetFlow、VPN、Web プロキシなど)、認証データソース (IAM や Active Directory など)、リポジトリデータソース (ファイル共有、ソースコード、データベースなど)、エンドポイントおよび DLP データソース) のスキーマにマッピングすることが可能になります。データの取り込みまたはストリーム時に、入力列がデータソースから抽出されたモデルの機能にマッピングされます。

膨大な数のデータソースを特定数のデータタイプに変換することは、効果が低い 2 つの極端な例の間のバランスをとるものです。極端な例の 1 つ目は、1 つの汎用的な異常検出方法を使用して、あらゆる無数のデータタイプ設定をサポートしようとすることです。これは、膨大な数の誤検出とノイズに繋がります。このように一括で簡素化した場合、データソース間の重要な意味の違いが無視されます。たとえば、Active Directory で「悪い」ものは、NetFlow で「悪い」と解釈されるものとはほぼ確実に異なります。もう 1 つの極端な例は、マッピングに柔軟性を持たせずに、非常に限定された特定のデータソースのみをサポートすることです。この場合、複数のデータソースを追加するという重要な能力が制限され、柔軟性が失われます。直感的には両方をサポートする認証モデルがあるべきだと感じられるところです。しかし、実際問題として独自に構築した認証システムが Active Directory モデルでサポートされないということになります。マップされた、意味的に相当するデータタイプを処理する中間層を作成することにより、効果的なデータサイエンスモデルとデータソースの柔軟性を同時に実現できます。

学習による何百万もの個別ベースライン作成

ArcSight Intelligence アーキテクテクチャの 2 番目の段階は、ベースラインの測定と作成です。前述のように ArcSight Intelligence は、各エンティティと、そのエンティティと他のエンティティとの関係に対し、固有の行動のベースラインを作成します。各ユーザー、マシン、ファイル、IP アドレス、プロジェクト、リソース、サービス、共有、ウェブサイト、ドライブ (ボリューム)、プリンタの「ユニークノーマル」と、これらのエンティティの組み合わせ間のインタラクションの「ユニークノーマル」が測定されます。

図 2. エンティティのベースライン

教師なし機械学習での「ユニークノーマル」について

ArcSight Intelligence は、教師なし機械学習でベースラインを作成し、数学的にパターンを発見します。今日サイバーセキュリティで使用されている AI や機械学習の技術は他にも多数ありますが、ラベル付けされていないデータセットで関連パターンを発見できるのは、教師なし機械学習のみです。これは多量のラベル付けされたデータを必要とする、ディープラーニングなどの教師あり機械学習とは異なります。マルウェア検出では、ラベル付けされたデータセットが一般的なため、ディープラーニングはマルウェア検出に適しています。

教師なし機械学習をサポートするため、ArcSight Intelligence は、エンティティベースライン作成に使用される集計と係数の保存とクエリに Hadoop テクノロジーの HDFS、HBase、そして Phoenix を使用します。機械学習には、トレーニング (新しいデータに基づいてモデルの精度を高める段階) と、スコアリング (モデルを使用して脅威を検出し、リスクを計算する段階) の 2 つの段階があります。一般的に、HDFS 上の Phoenix または HBase を使用する形は、性質的にマッピングが少ない傾向にあるため、トレーニングに適しています。一方、脅威をできるだけ迅速にリアルタイムで検出できる Spark は、スコアリングに適しています。

この目的は、母集団 (これには、何千ものユーザー、何百万ものファイル、何百ものサーバー、何千ものマシンが含まれる可能性があります) の各エンティティの正常な行動と、それぞれがどのようにインタラクトしているかをデータから学ぶことです。この観察の結果を、「ユニークノーマル」と呼びます。

ArcSight Intelligence はこれを数学の力で達成します。さまざまなデータタイプにわたる何百ものアルゴリズムが、各エンティティの正常な行動を自動的に学習します。このベースライン学習により、各エンティティに対して、「オンライン」またはお客様の環境内で自動的に学習された、統計に基づく包括的な「指紋」が作成されます。オンライン学習は、データサイエンティストによってサイエンスラボで機械学習が手動で行われる「オフライン」学習と対比することができます。ArcSight Intelligence のすべてのアルゴリズムは、教師なし機械学習アプローチを採用しています。このアルゴリズムは、各エンティティの正常な行動に関連付けられた確率密度関数を学習します。そのほとんどは、シンプルな一変量または二変数モデルです。

教師なし機械学習をサポートするため、ArcSight Intelligence は、エンティティベースライン作成に使用される集計と係数の保存とクエリに Hadoop テクノロジーの HDFS、HBase、そして Phoenix を使用します。

原理に基づく信頼できるシステム

シンプルな多数のモデルを並列して実行することには、いくつかの利点があります。第一に、ユニークノーマルを学習するために ArcSight Intelligence が選択したアルゴリズムはすべて、何百年にもわたって研究されてきた、十分に理解され証明された統計的な学習アルゴリズムである最尤推定とカーネル密度推定です。これらは、検証を受け、原理に基づいており、理論的に適正なアルゴリズムです。

第二に、これらのアルゴリズムは、脅威の検出に繋がる特定の行動の異常について、人間が読める形式で説明を提供できます。たとえば、ディープラーニングニューラルネットワークから、人間が理解できる説明を生成することは非常に困難です。しかし、監査性、追跡性、責任性を担保したデータ使用の透明性が求められる業界では、このような説明が不可欠です。

分析モデルの出力は、脅威リーダーボードに使用される相対的リスクスコアです。セキュリティ担当者は、リスクスコアを使用して、より効果的な脅威検出、脅威探索、そして脅威の調査に集中することができます。

自動的に変化に適応する自己学習システム

何百ものアルゴリズムを使用してさまざまな行動を追跡するため、実際問題として、脅威が検出から逃れる唯一の方法は、何百ものさまざまな形で正常に行動することです。しかしシステム内には多数のデータソースとモデルがアクティブに追加されるため、それは次第に難しくなります。これらの行動は、ラボの「オフライン」ではなく「オンライン」で学習されるため、学習プロセスがより迅速に行われるだけでなく、お客様間の違いまたはお客様の母集団内のエンティティの間の違いが、各エンティティのオンライン学習の「ユニークノーマル」によって自動的に把握されます。

このため、自己学習して各顧客の環境の特定の性質や傾向に合わせてカスタマイズすることによって正確さを高めるだけでなく、環境内の変化に反応し、適応し、歩調を合わせるシステムが実現できます。この自己学習という要素により、従来のルールとしきい値に基づくシステムで必要だった手作業は、最低限または皆無になります。従来のシステムでは、環境に変化があると手動で継続的に調整する必要がありました。教師なしアプローチを使用することで、効果的な結果を得るために、ラベル付けされたデータの巨大でクリーンなデータセット (サイバーセキュリティにはほとんど存在しない) を用意する必要がなくなります。最後に、シンプルな一変量または二変数モデルを使用することにより、あるお客様から他のお客様向けに迅速に一般化できる低い次元の学習が可能になります。

分析モデルによる異常の検出

ArcSight Intelligence アーキテクチャの第三の段階は、異常の検出です。ここで ArcSight Intelligence は、何百もの分析モデルを適用してエンティティの正常な行動からの逸脱を測定し、「ユニークノーマル」からの違いのどの組み合わせが優先脅威リードとなるかを統計的に決定します。分析モデルには、尤度推定、確率密度推定、期待値最大化、クラスタリング法 (K 平均法、ガウス混合モデル、ベキ乗クラスタリングなど)、そして PCA のような次元縮小法など、多くのさまざまなタイプの統計法および機械学習アルゴリズムが含まれます。これらの手法は、Apache Hadoop、HBase、Kafka、Spark、Elasticsearch などのテクノロジーを使用し、世界最大規模の組織に合わせてスケーリングし、オープンなビッグデータインフラストラクチャで実行するように最適化されています。

分析モデルの出力は、脅威リーダーボードに使用される相対的リスクスコアです。セキュリティ担当者は、リスクスコアを使用して、より効果的な脅威検出、脅威探索、そして脅威の調査に集中することができます。

モデルは、正常、期待される行動、そして観察された現在の行動の間の違いを検出し、数量化して脅威のリスクを予測します。分析モデルにより、ArcSight Intelligence 脅威検出プラットフォームは、コードを記述してモデルを作成するサイバーセキュリティデータサイエンティストのチームに頼らずに脅威を検出でき、ビッグデータサイバーセキュリティ分析プラットフォームの価値実現を大きく加速します。ビッグデータのサイバーセキュリティプロジェクトでは、深い専門知識を持つセキュリティチームと、コード記述と統計モデリングの技術を持つデータサイエンスチームからなる、多量のリソースを必要とするチームが結成されることがよくあります。さらに、モデルを構築するためのカスタムコーディングには多大な労力が必要であるため、これらのモデルを脅威検出に適用できるようになるまでには、一般的に何週間、何か月という時間がかかります。ArcSight Intelligence のすぐに使用できるモデルは、この何か月もの労力を節約します。

確率密度関数による異常測定

正常を説明する確率密度関数 (PDF)、観察された行動 (データから) と期待される行動 (PDF から) の間の距離関数を構築することにより、異常と危険な可能性のある行動を検出し、数量化することが可能になります。特定の入力機能、PDF と距離測定基準 (これらをまとめて異常モデルと呼びます) は、データタイプと対象のユースケースによって異なります。モデルの重み付けにより、さまざまな行動に対する異常の検知度を上げることができます。エンティティの重み付けでは、さまざまなエンティティに対する距離の検知度を上げることができます。

ルールとしきい値のバイナリ性を含む確率的アプローチ

確率的アプローチとは、任意のしきい値を下回る行動を無視したり (ゆっくりとした小さな攻撃が見逃されるため再現率に悪影響を与える)、または任意のしきい値を超えるすべての行動を悪いものとみなす (ノイズが増加するため精度に悪影響を与える) 必要がないことを意味します。つまり、誤検出が生じがちなしきい値ベースのアラートのバイナリ性を回避できます。個々のモデルが性質上「ファジーな」継続的数値を出力するため、より確率的で統計的なアプローチを取ることができます。最後に、モデルとエンティティの重み付けにより業務上のコンテキストを提供することで、これをリスクスコアに統合することができます。たとえば、会社で特にプリンター関連の異常に興味がある場合にプリンターモデルの重み付けを増やしたり、または契約期間最後の 2 週間に入った臨時雇用者を追跡するために関連エンティティの重み付けを増やしたりすることができます。

統計的証拠を使用した優先度の高い脅威リードの生成

ArcSight Intelligence アーキテクチャの第 4 段階では、すべてのモデルとデータセット全体ですべての統計的証拠が取集され、各エンティティに対して測定リスクスコアが計算されます。ここですべての個々のリスクスコアが相互に重み付けされ、優先度が付けられた脅威リードのリストが作成されます。セキュリティ担当者は、ここから対応を開始することができます。この段階では、Elasticsearch を使用して分析結果およびリスクスコアを保存し、ArcSight Intelligence の UI を使用して可視化を行い、ArcSight Intelligence の UI またはオプションとして Kibana を介して即座にデータの検討を行います。

ArcSight Intelligence アーキテクチャの第 4 段階では、すべてのモデルとデータセット全体ですべての統計的証拠が取集され、各エンティティに対して測定リスクスコアが計算されます。

統計を適用したエンティティリスクスコアの作成

すべてのモデルから出力された確率は、共通の期間について、エンティティごとに統合されます。本当に危険な何かが起こっていることを示す証拠が増え、積み重なっていくイメージを直感的に掴もうとします。統合されたデータは次に、ロジスティックまたは Pareto スカッシング関数に送られ、最終的な値の範囲は [0,1] に効率的に統合されます。エンティティリスクスコアと呼ばれるこの最終値が、0 から 100 の間の値で UI に表示されます。スカッシング関数のシェイプパラメータを調整することで、すべてのエンティティリスクスコアにターゲットの分布を作成することができます。

誤検出の防止

複数の異常モデルを統合することで、誤検出のインシデントを大幅に削減できます。これは、大きな異常リスク値は、多数の異常 (複数ソースからの大量の証拠) がある場合、または少数の甚だしい異常 (少数の特にリスクの高い異常) があるのみに発生するためです。[0,1] (または UI 内で [0,100]) の範囲の出力が保証されるということは、出力リスクスコアが安定しており非有界ではないことを示し、戦略ガイドのオーケストレーションと開発がより実践的になります。リスクスコアに上限がなかったり、新しいデータソースや追加モデルが追加されるたびに上限が変化すると、戦略ガイドの定義は困難です。

最後に、エンティティリスクスコアの特定の分布を対象とすることができるため、業務上、統計上、および見やすさの要件を満たすように低から高までのエンティティリスクスコアを分布させることができます (たとえば、誤検出を減らす、対象の精度 / 再現率、赤・黄色・灰色の「見やすい」分布など)。

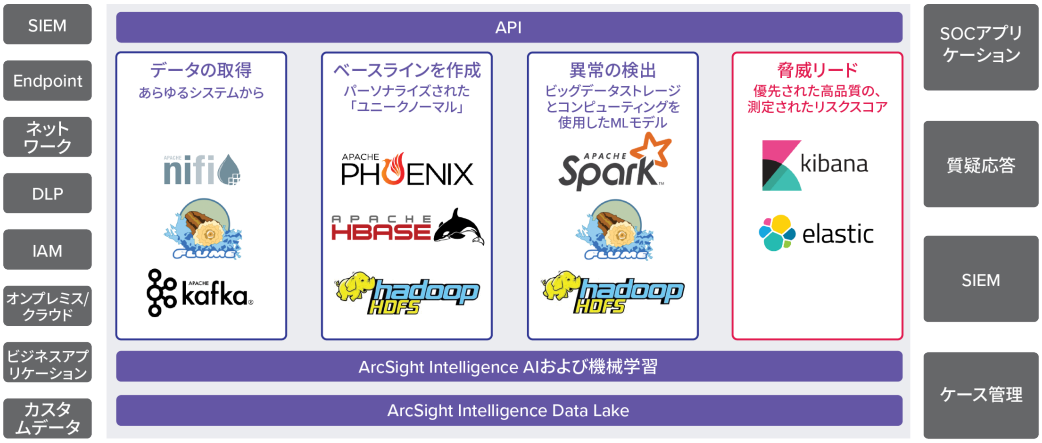

図 3. ArcSight Intelligence の AI 搭載脅威検出の概要

ArcSight Intelligence の脅威検出プラットフォームは、ビッグデータテクノロジーと数学的分析を使用してルールとしきい値に基づくシステムの制限を克服し、脅威リードの優先度付きリストを生成します。

断片化されたセキュリティエコシステムの統合

ArcSight Intelligence の脅威検出プラットフォームは、ビッグデータテクノロジーと数学的分析を使用してルールとしきい値に基づくシステムの制限を克服し、脅威リードの優先度付きリストを生成します。ネイティブのビッグデータ Hadoop アーキテクチャのため、本質的に拡張性が高く、デジタルワールドのサイバーセキュリティの需要に応じて拡大できます。実証済みの設計によいり、断片化されたセキュリティエコシステムを含む、広範なデータソースからデータを取り込むことができます。SIEM、エンドポイント、ネットワーク、DLP、IAM、ビジネスアプリ、セキュリティデータレイクなどからのデータをまとめ、統計的に重要なリスクについて分析できます。

既存のセキュリティツールの強化と自動化

ArcSight Intelligence 脅威検出プラットフォームがリスクを識別したら、それを外部システムと共有してその後の処理を行うことができます。一般的な統合では、リスクスコアを次を含む送信先に送信します。

- セキュリティ担当者が対策を取れるように SIEM に戻す
- 検出された脅威に対して人間の操作よりも迅速に対応するために Phantom、Swimlane、Demisto などの操作システムおよび自動化システムに送る
- IT サービスおよび操作管理のために、ServiceNow、BMC Remedy、TheHive、IBM Resilient などのケース管理ツールに送る。実際、E メールや SMS から新しいケースをトリガまたは生成可能な、REST API を備えたあらゆるプラットフォームに送信できます (ServiceNow、BMC Remedy、TheHive、IBM Resilient など)。

お問い合わせ先：**CyberRes.com**
この記事はいかがでしたか？
シェアはこちら

これらすべては、オープンソーステクノロジーを使用し、広範なシステムとの統合を受け入れるオープンプラットフォームに基づくものです。これにより、企業は特定のベンダーに囲い込まれることがなくなり、自由に、柔軟にセキュリティツールを選択して機械学習の力を活用することができます。

次のステップ

以下の表は、ユーザーおよびエンティティの行動分析に使用する、ArcSight Intelligence の数学的ビッグデータアーキテクチャの主要コンポーネントをまとめたものです。多くの状況に当てはまるように、企業に真の価値をもたらすのは、本ホワイトペーパーに説明されている概念やテクノロジーの実行です。リスクの可視化を高め、脅威検出の精度と速度を向上したいと希望する CISO、セキュリティアーキテクト、そしてその他のお客様は、**microfocus.com** で詳細をご覧ください。

	第 1 段階	第 2 段階	第 3 段階	第 4 段階
用途	リスク可視化の最大化	「ユニークノーマル」の測定	異常の検出	脅威リード
操作	データの取得	ベースラインを作成	数学的確率でリスクを評価	エンティティリスクスコアの正規化
データサイエンスのプロセス	データの取り込み、正規化、変換、充実化	特徴の抽出、機械学習モデルのオンライントレーニングと検証	機械学習スコアリング、重み付け、期待効用理論ベースの行動リスクスコアリング	重み付け、共通の期間にわたるエンティティとエンティティリスクスコアの統合、統計的正規化
ビッグデータテクノロジー	Apache NiFi Apache Kafka Apache Flume	Apache Phoenix Apache HBase Apache HDFS	Apache Spark Apache Flume Apache HDFS	Kibana Elastic

表 2. 主要コンポーネントの概要

マイクロフォーカスエンタープライズ株式会社
jp-info-enterprise@microfocus.com
www.microfocus-enterprise.co.jp

CyberRes
A Micro Focus Line of Business