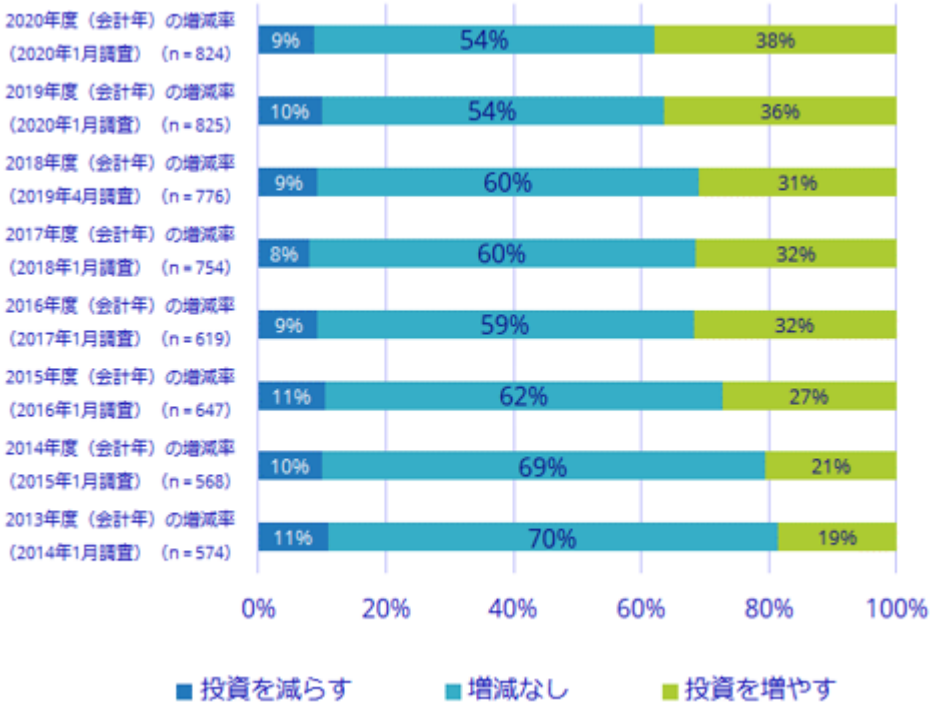
AIセキュリティシステムCloud Coffer
AI技術が実現する、攻撃者の一歩先を行くセキュリティシステム

法人会員　株式会社アリス　AIセキュリティ事業部　青木　登

１．情報セキュリティの概況

以下のIDCの調査結果に見られるように、大手企業の情報セキュリティへの投資は、ネットワークセキュリティとアイデンティティ／アクセス管理、クラウドセキュリティなどの投資重点項目を中心に依然として旺盛であるが、一方では6割近くの企業では、セキュリティ予算は確保されておらず、計画的なセキュリティ投資がなされていないという状況が続いています。



出典：IDC 調査　2013 年度(会計年)～2020 年度(会計年)の前年度に対する情報セキュリティ投資増減率

そのような中で、攻撃者たちの活動は相変わらず盛んで、近年では AI 技術を取り入れた多様な攻撃も多くみられ、攻撃対象も、最終的な目的である大企業への直接攻撃から、セキュリティ対策が十分でない企業や組織のシステムを侵害し、フィッシングメールなど様々な技術を駆使して堅牢なセキュリティ対策を施した企業を攻撃するというサプライチェーン攻撃に比重を移しているように思われます。また、コロナ禍においては、テレワーク、在宅勤

務が推奨され、VPN の脆弱性を狙った攻撃や、モバイルアプリへの攻撃なども話題になりました。

以下は、情報処理推進機構(IPA) が 2020 年の情報セキュリティ白書でまとめた、2019 年度の情報セキュリティの概況ですが、2019 年 7 月に米国の大手金融会社の 1 億人を超える

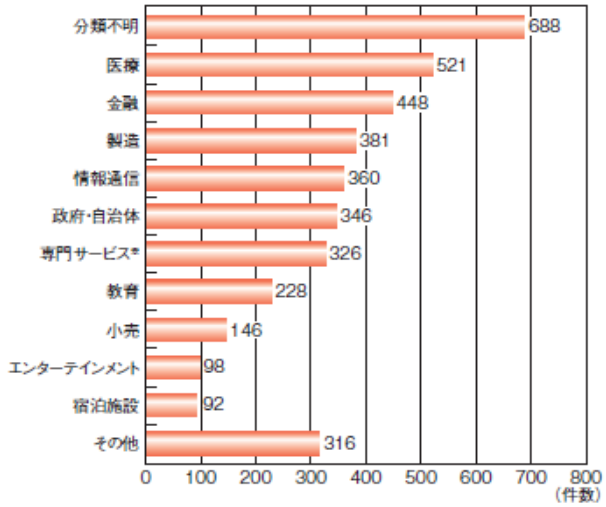
2019 年度の情報セキュリティの概況

	主な情報セキュリティインシデント・事件	主な情報セキュリティ政策・イベント
2019 年　4 月		経済産業省、「サイバー・フィジカル・セキュリティ対策フレームワーク　Version1.0」を策定(2.1.1) NISC「小さな中小企業と NPO 向け情報セキュリティハンドブック」公開(2.4.2)
5 月	EC サイトのアカウント 46 万 1,000 件に不正アクセス(1.2.7) アンケートモニターサービスの登録アカウント 77 万 74 件に不正アクセス(1.2.7)	NISC「サイバーセキュリティ 2019」公開(2.1.1) 米国で中国ベンダほか関連企業が輸出規制対象に(2.2.2)
6 月		G20 大阪サミット開催、信頼性のあるデータの自由な流通の概念を提唱(2.2.1) 経済産業省「サイバーセキュリティお助け隊」開始(2.4.2) 総務省・NICT「NOTICE」における注意喚起事業を開始(2.1.1、3.2.2)
7 月	米国の大手金融会社のクラウドから大量の個人情報漏えい(1.1.1、3.4.1) 福岡県警察、警視庁等、海賊版サイト運営者らを著作権法違反で検挙(2.1.4)	英国 ICO が航空会社及び宿泊事業者に GDPR 違反で巨額の制裁金(2.2.3)
8 月	スマホ決済サービスが不正アクセス被害を受けサービス廃止を発表(1.1.2) 就職情報サイト運営会社が「内定辞退率」データを販売(1.2.7) クラウドプラットフォームサービス大手が大規模障害で多数のサービスに影響(3.4.1)	米国で国防権限法 2019 が発効、中国の IT ベンダ・通信機器ベンダ 5 社の政府調達が禁止に(2.2.2) 東京 2020 組織委員会が AI を活用した顔認証技術導入を発表(3.3.3)
9 月	エクアドル国民約 2,000 万人分の個人情報流出(1.1.1) 大手新聞社米子会社、香港に 32 億円流出の詐欺被害(1.2.2)	経産省と IPA、インド太平洋地域向け日米サイバー演習を実施(2.1.1、2.2.1) ラグビーワールドカップ開催(1.2.3)
10 月	フィッシングの月間報告が 8,000 件を超え過去最多に(1.1.2、1.2.6)	EU 加盟国、5G セキュリティのリスク評価結果を報告(2.2.3) 重要インフラ専門調査会「「重要インフラの情報セキュリティ対策に係る第 4 次行動計画」に基づく情報共有の手引書(試行版)」策定(2.1.1)
11 月	JPCERT/CC、Emotet の感染に関する注意喚起(1.2.5)	NISC が東京 2020 オリンピック・パラリンピック競技大会を想定した「分野横断的演習」を実施(2.1.1)
12 月	情報機器リユース会社において廃棄予定 HDD の流出発覚(1.2.7) 自治体向けクラウドにおけるシステム障害でサービス停止等の影響(3.4.1) 日本への Emotet のばらまき型メールによる攻撃急増(1.2.5)	ドイツのモバイル通信ネットワーク構築で Huawei 社との契約が確定(2.2.3)
2020 年　1 月	国内防衛関連企業が不正アクセスによる情報流出を公表(1.2.1、1.2.7)	米国国防総省、サイバーセキュリティ成熟度モデル認証(CMMC)の初版を公開(2.2.2)
2 月	新型コロナウイルスに関連した内容の SMS からフィッシングサイトに誘導する手口発生(1.2.6)	英国、正式に EU を離脱、新しい自由貿易交渉開始(2.2.3)
3 月		個人情報保護法改正案閣議決定(1.2.7、2.7.4) 内閣府・経済産業省・総務省「政府調達のためのセキュリティ評価制度(ISMAP)」パブコメ開始(2.1.2、3.4.2) 米国国土安全保障省、新型コロナウイルス関連詐欺メール、詐欺サイトに注意喚起(2.2.2)

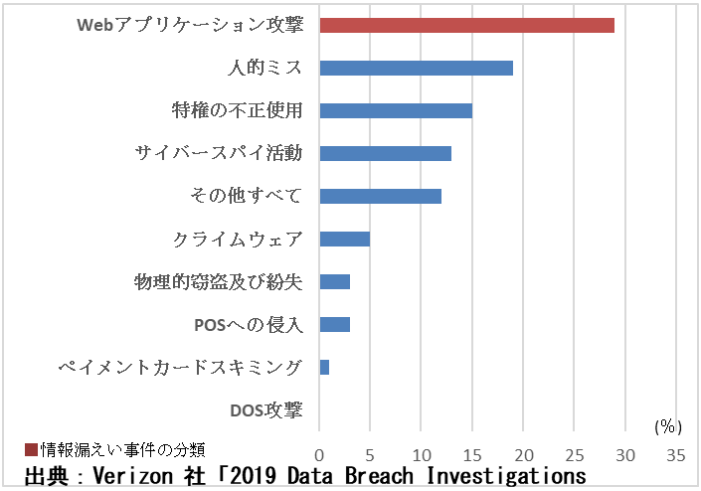
出典：独立行政法人情報処理推進機構「情報セキュリティ白書2020」

顧客情報の流出、スマホ決済サービスの不正利用、防衛関連企業からの情報流出、フィッシングメールによる不正送金、Emotet などのマルウェアによる情報窃取などが見られ、脆弱性や管理体制をついたものだけではなく、ゼロデイ攻撃や新種・亜種のマルウェアなども多く見られるようになってきました。

情報漏洩に限ってみても、以下のグラフの示すように、様々な分野でインシデントは発生しておりますが、2019年度には医療分野、金融、製造業が上位を占めております。しかもビジネスの利便性などのために用意されたWebアプリケーションを介したケースが全体の29%もの比率を占めています。



■図 1-1-4　業種別の情報漏えいの件数
(出典)Verizon 社「2020 Data Breach Investigations Report」を基に IPA が作成



出典：Verizon 社「2019 Data Breach Investigations

近年は、利便性もあって、Webアプリケーションを使ったサービスが多様に展開されており、個人向けのオンラインショッピングやオンラインバンキングなどにとどまらず、企業間の取引、公共機関や病院などによる各種サービスの予約や申請など、Webアプリケーションの活用の広がりはとどまるところを知りません。

攻撃者たちはそれらシステムに対して、各種自動化技術やAI技術を駆使し、伝統的な方式や手続きなどに縛られず攻撃をしかけています。既知のエクスプロイト、マルウェア、ファイルレス等の攻撃だけでも既に数10億件を超えていますが、誰でも容易に最新の攻撃技術にアクセスして利用できる環境とあいまって、攻撃の種類や数は加速度をつけて日々増え続けています。専門家たちは毎年1億以上の新たな脅威を発見し続けていますが、このギャップはますます広がるばかりで追いつける兆しがありません。更にモバイルやクラウド技術の広がりによってインターネットに接続される機器が増大する中で、仮想空間と物理空間の境目がますます曖昧になってきており、今までの考え方では簡単に対処できなくなってきています。

２．従来型の境界防御型ソリューションの限界ーゼロトラストという考え方の台頭

境界防御型のソリューションとして、Firewall、WAF（Web Application Firewall）やIDS（Intrusion Detection System）、IPS（Intrusion Protection System）などがありますが、Firewallを除くソリューションの多くは、シグネチャマッチング技術を使っております。これらは、通常のトラフィックや不正なトラフィックを見分ける上で、ある意味有効なのですが、既知の脅威にしか対応できません。シグネチャマッチングでは、ゼロデイやファイルレス攻撃、新種あるは亜種のマルウェアなどの攻撃に対しては効果がありません。また、有効に機能させるためには、シグネチャの更新や、カスタムシグネチャなどを必要としますが、未知もしくは亜種の攻撃に対しては効果がなく、対応するために数か月の脆弱期間ができてしまいます。これらの境界防御型のセキュリティ製品は、侵入を防ぐという目的で使われることが多いのですが、一度でも侵入を許したら存在意味が薄れてしまいます。そのために最近はそれらで防御できるものには限界があるということから、半ばあきらめ的な気持ちを込めてゼロトラストいうことが言われるようになってきました。

攻撃側は、既に述べたように高度の自動化ツールやAI技術を使った各種攻撃ツールを容易に入手できるようになりました。一方で防御側が、多くの労力をかけてシグネチャの更新やカスタムシグネチャなどを作っても攻撃側のスピードについてゆくことはできません。ゼロトラストの意識を持ってシステム環境を作り直すことも重要ですが、ビジネスの利便性などを考えると境界防御を全くしないで良いということにはなりません。また、攻撃側ができることには、必ず何らかの防御の方法があるはずです。

３．守る側にもAI技術を取り入れよう

境界防御の攻防戦においては、残念ながら常に攻撃側にリードされ続けてきたと言わざるを得ません。セキュリティ部門は、常に新しい攻撃に対応するために、OS、アプリケーション、上記各セキュリティソリューションのアップデートに追われ、SOCなどのインシデント対応チームも常に最新の情報に精通し、既存のシステムを迂回する攻撃がないかに目を光らせていなければなりません。この無尽蔵に人的リソースを食いつぶす状況を回避するため、AI技術を取り入れた製品が、徐々にセキュリティ分野にも現れ始めました。

ガートナー社 が、“AIは組織の在り方を変える可能性があり、ディジタルビジネスの中心的存在になる”と言うように、AI技術を使った新たな手法は、人手による人海戦術的な方法をやめ、効率的に未知の攻撃、攻撃の脅威の検出精度を大いに向上させ、ネットワークの可視化にも大いに役立つものです。

以下に、AIがセキュリティ問題について貢献できる点についてまとめてみました。

脅威の予測	次はどこから攻撃がやってくるかなどの予測に役立つもので、同様の技術を自動攻撃などに悪用される可能性があるものの、AI技術を活用することで、少なくとも攻撃者と対等に渡り合い、あるいは一歩先に行くことが可能となる。
アプリケーションのセキュリティ確保	シグネチャマッチングに依存した従来型のセキュリティ製品とは異なり、トラフィックデータそのものを、熟練のエンジニアと同じように解析し、不正な動きを検知できる可能性を持っている。人が行うのとは異なり、疲れによる見逃しなどをなくすことが期待されます。
セキュリティチームの効率的な行動	AI技術の活用によって、セキュリティ担当は、バグフィックスやアップデートの実装などに時間を費やすことなく、本来の防御対策により力を注ぐことができるようになる。脅威を特定するための時間を短縮するだけでなく、より効率的に行動できるようになる。
脆弱期間が不要	AIによって未知の脅威や既知の攻撃の亜種、難読化などを検知・遮断することができるようになると、脆弱期間（対策が提供されるまでの空白期間）を気にしなくて済むようになる。

４．CloudCoffer のAI技術とはどんなものか

AI技術は、人間が行うような複雑な問題の解決策を探すための科学技術分野であり、人間の持つ知的活動をシミュレーションするものです。人間が実際に行う意思決定メカニズムと同じような動きは、何らかのアルゴリズムでモデル化することができます。マシンラーニングは、意味のあるデータを人間が検知して学習するのと同じような能力を、コンピュータに持たせる技術で、データから情報を抽出するための数学的、統計的手法を使って行われ、これらの抽出された情報から未知の事象を予測します。それは、コンピュータが、人間と同じように、意味のあるデータを見つけてそれを学習する能力をコンピュータに持たせるための、最先端の技術です。人間が経験に応じて成長するように、AIも判断能力と解析能力が強化されます。

人工ニューラルネットワークは、脳内のニューラルネットワーク構造に触発されて作られた計算モデルであり、複雑な通信網によって互いに接続された基本の計算装置（ニューロン）が大規模に集積されたもので、それによって脳は高度に複雑な計算を行っています。ニューラルネットワークの理論は、多くのニューロンが通信網を経由して接合されて複雑な計算を実行しているという考え方が背景にあります。

CloudCoffer は、ニューラルネットワーク予測からなる推論クラス分けを定義しており、そこでは全ての推論はフィードフォワードネットワークにつながっていてエッジの重みづけで分けられています。単一のニューロンは図１、図２にあるように、シグモイド関数 $\sigma(z)=1/(1+\exp(-z))$ の活性化機能としてモデリングされています。グラフにあるそれぞれの

エッジは、あるニューロンの出力を別のニューロンの入力として関連付けられます。ニューロンへの入力は、それに接続されたすべてのニューロンからの出力に重みづけをした和から得られます。図3にあるように、CloudCoffer は悪意のあるトラフィックから特徴を抽出し、

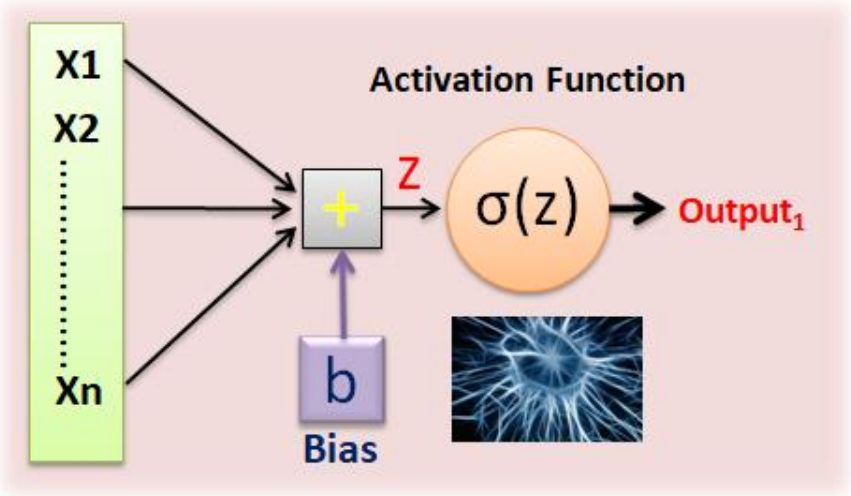


図１． A.I.の活性化関数.

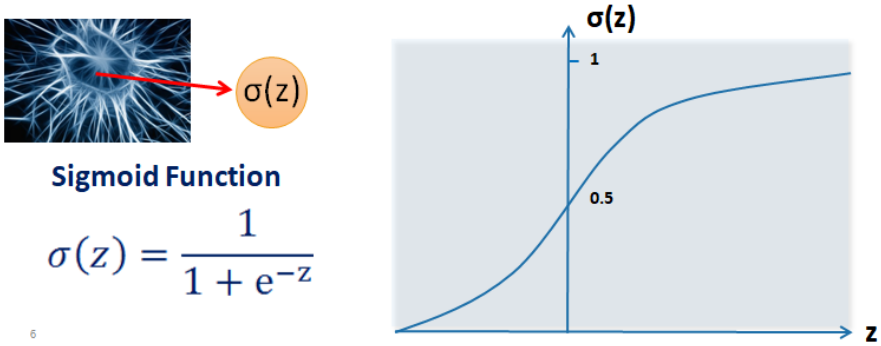


図２． A.I.におけるSigmoid 関数

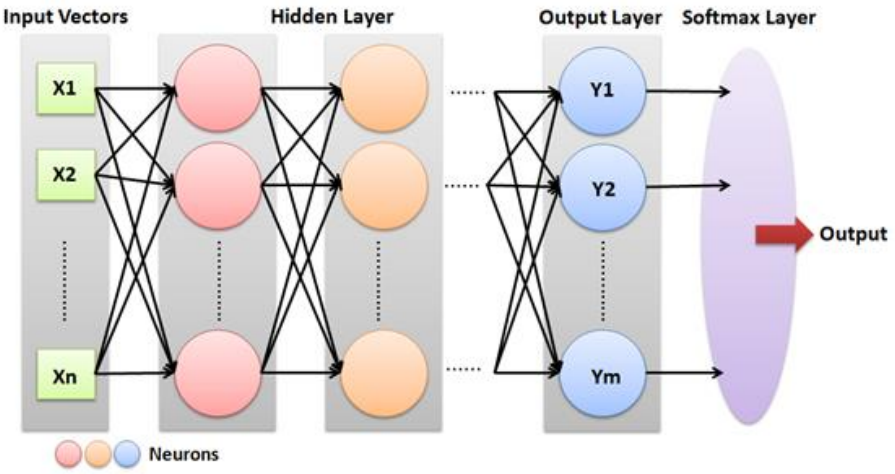


図３． AIのニューロン構造

これらの特徴をベクトルに変換します。入力ベクトルは、次に、学習のために、CloudCoffer のアルゴリズムに放り込まれます。

最も複雑なオペレーションは適切な解を求めて行うエラー機能の最急降下法の計算です。一般的に、ニューラルネットワークの学習に広く使われる発見教授法は、SGD（Stochastic gradient descent－確率的勾配降下法，バッチ学習で使われる最急降下法をオンライン学習に改良したもの）フレームワークに基づくものです。CloudCoffer のAIエンジンも、不正トラフィックの検知率を飛躍的に高め、効率的に行うためにカスタマイズモデルを使用しています。脆弱性を発見するAI技術を使ったアプリケーションの開発にあたって、最初に行ったのは、何を差別化機能にするかを決定することでした。全ての要素や機能は、当初からAIに盛り込みました。CloudCoffer では、不正データ（トラフィック）からその特徴を抽出してそれらをベクトル化しています。私たちのエキスパートシステムはAIにコンピュータ言語の持つ意味を理解するよう学習させるため、AIエンジンは、トラフィックの持つ意味を理解し自身で進化してゆきます。トラフィックの持つ特徴は、記述、分類、数式化などによって意味のある表現にされます。次に、ビッグデータ解析により、AIエンジンは学習を始め、独自のモデルを構築しますが、それは人間が新しいことや考えを学習してゆくのに似た動きです。データは、エクスプロイトデータベース、NVD（National Vulnerability Database－ 脆弱性情報データベース)、Web Crawler（ウェブクローラー）、CloudCoffer 自身が世界中に配置した16万か所以上のハニーポットなどから収集されます。これらのAIエンジンの学習以外に、David Brumley教授の論文が、エクスプロイトの提供などで、CloudCoffer のAIエンジンの学習に大きな役割を果たしています[参照1]。図4と図5にあるように、CloudCoffer 及び関連会社のRayAegis 社では、自動的にエクスプロイトを生成する特製のアルゴリズムによる高度なファジングシステムを開発しました。エクスプロイトとファジング技術[参照2]、特製の生成アルゴリズム[参照3]とを統合することで、既存のエクスプロイトをもとに、全く新たなエクスプロイトを生成することができるようになりました。これらのエクスプロイトだけでなく、両社は、OpenAI フレームワークを使った、AIベースのハッキングツールを開発しました。このツールはCloudCoffer システムに搭載された防衛型のAIエンジン（MatrixShield と呼んでいます）を攻撃するツールとして使われています。

エクスプロイトやその進化系を収集した後、AIは攻撃を構成する重要な要素について学習を行います。脆弱性を判断するうえで、以下のような項目を確認します。

- Special Symbol
- Malicious IP Addresses
- Length
- System Commands
- Customized Images
- その他

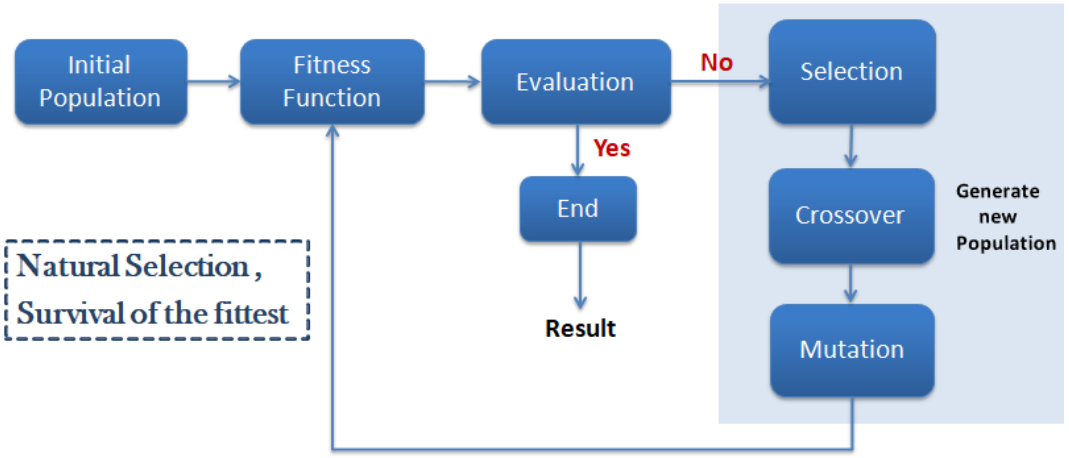


図４．既知のエクスプロイトからプログラムによって自動的により高度なエクスプロイトを生成
AIは、ハッカーたちがどのように攻撃を進化させるかのプロセスを理解

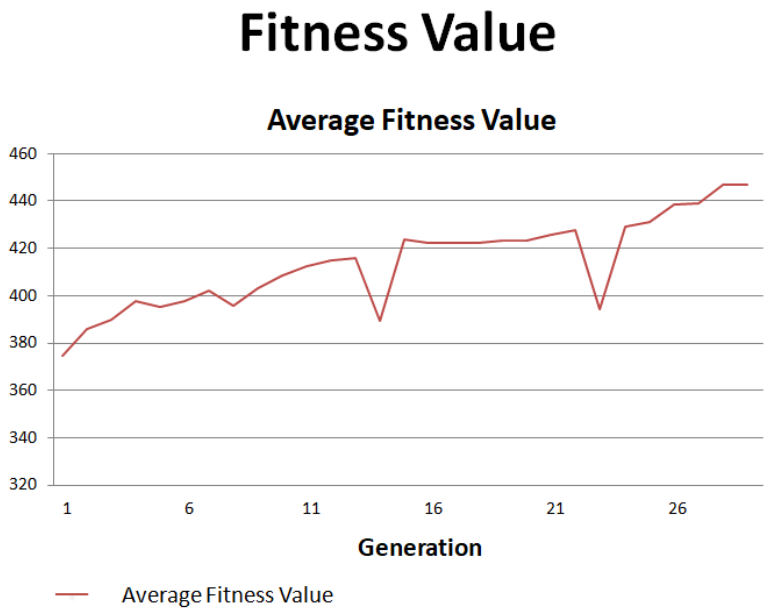


図５．エクスプロイト各世代の変異における裁量適合

学習プロセスを経た後、AIは不正データと正常なものとを正確に区別するモデルを、図6にあるような形で作り始めます。最初の段階では、AIエンジンは正常か不正かを決定づける因子を多数受け取りますが、学習が進むにつれ、AIはそれらを大幅に減らして450の重要な重みづけをもったベクトルのみでトラフィックをフィルターするようになりました。誤検

知を最小化し、性能を最大化するなかでの最適値を求めた結果ベクトル数は450まで削減されたのです。また、AIは、人間の目には簡単に理解できない隠された要素まで発見することができます。

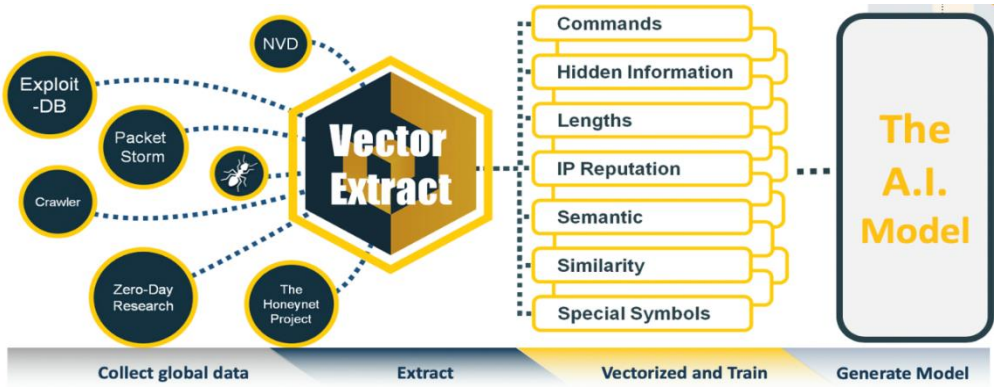


図6．大量のエクスプロイトから特徴を抽出し、エクスプロイトを構成する要素を学習、推論

新たなデータをAIエンジンに与えると、AIエンジンはデータを数値化された配列に変換し、入力データを表します。入力ベクトルは判定アルゴリズムに投入されます。

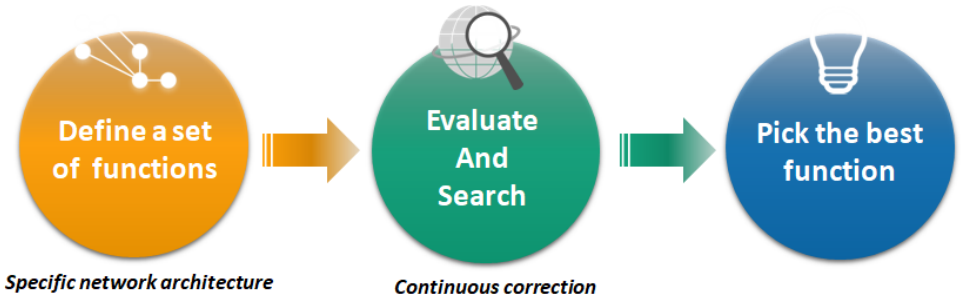


図7．補正・成長プロセス

学習においては、効率的に実行可能予測を行う仮設分類を行う上で、近似誤差があります。CloudCoffer のAIエンジンは、新しいモデルへのアップグレードが簡単にできる柔軟性とマシンラーニングモデルが十分に性能を発揮し堅牢であるという性格を兼ね備えています。ニューラルネットワークをより大型のシステムの導入に耐えられるようにするために、AIエンジンは、多くの計算装置が共有のパラメータによって、様々なレプリケーションやコアモデルのデータフローグラフの並列処理を表現できるようにしてあります。

５．CloudCoffer のAIエンジンによるマルウェア検知の仕組み

高度な技術をもったハッカーは、発見が極めて困難なエクスプロイトやマルウェアを創ることができます。そのため、このような者たちの攻撃は解析がますます困難になってきています。この困難に打ち勝つために、ファイルやコマンドを単純に解析するだけではなく、CloudCoffer ではファイルやコマンドを隔離されたマシン上で実行してその動きを解析するSandboxを開発しました。ファイルやコマンドを動的に解析することで、ファイルの持つ意図を見つけることが可能となります。ファイルやコマンドがバックドアを開けようとしたり、機密情報を盗み出したり、あるいは他の不正を働こうと意図したものである場合、CloudCoffer はそれらのファイルが不正であるとみなします。CloudCoffer のSandboxでは、定義済みの不正な動きをトラップと呼びますが、マシンラーニングの結果与えられるそれぞれのトラップの重みづけを用いて、解析したファイルやコマンドにリスクレベルのスコアが与えられます。

マシンラーニング技術によってSandboxを効果的に使うことで[参照4]、信用レベルを出力します。

６．CloudCoffer のAIエンジンの誤検知、過検知が少ない理由

CloudCoffer が誤検知や過検知を最小化できる理由の一つに、「カスタマイズされたイメージ」があります。マルウェアやエクスプロイトをカスタマイズされたイメージに変換し、このイメージから特徴を抽出して入力データに使います。データはベクトル化されてCloudCoffer のAIエンジンによって処理されます。

例えば、洗練されたマルウェアは、Sandbox内で動作を止め、特定の条件がトリガーとなって動き始めるように作られています。マルウェアによっては、機密情報を月の最初の日にだけ送信するようなものもあります。しかしながら、このタイプのマルウェアは、他の類似のマルウェアのカスタマイズされたイメージと比較され、CloudCoffer のAIエンジンが検知できるのです。

この技術を使って、AIエンジンは、最も最先端のマルウェアを、実行させなくとも、また事前にハッシュ値などの情報がなくとも検知することができます。AIエンジンは、偽装されたエクスプロイトも同様の技術で検出することができるのです。

７．フィードバックループを減らす

AIエンジンは、実際に使われる環境を使って、大量のデータによる学習が必要ですが、CloudCoffer はユーザサイトでの学習は、以下の3つの理由から行っておりません。

1. AIエンジンは、正しくカテゴリ毎に分類された、膨大な量の不正トラフィックのサンプルで学習を済ませており、またその学習結果も時間をかけて検証している

ため、CloudCoffer のAIエンジンが新たな脅威を発見するに十分な能力を備えているという自負があります。

2. AIエンジンの学習は、システムリソースを大量に消費します。ユーザ先での学習をするようなシステムでは、膨大なCPUとメモリを搭載したシステムを必要とし、ユーザにとって現実的な選択肢ではなくなります。
3. ハッカーによる攻撃によってAIエンジンを狂わせたり、偏向を持たせたりすることを避けるためです。CloudCoffer のAIエンジンは、既に学習を終え自己完結したものを使っているため、エンジンの整合性を保証し、外部からの影響を受けることはありません。

８．CloudCoffer の使い方

CloudCoffer は、使い方によってその機能や表情を変えてきますが、基本的な使い方としては、以下の3つが考えられます。

1. IPS・IDS・NDRとして
   （ア） CloudCoffer を防御したいシステムの手前に透過モードで配置し、通過するL3からL7のトラフィックを検査し、不正なトラフィックを検知すると遮断する。IPSとしての使い方。
   （イ） CloudCoffer を境界防御システムの後ろにミラーモードで配置し、既存の境界防御システムをすり抜けたトラフィックに疑わしいトラフィックがあるか、また、ネットワーク内部から疑わしいサイトへの通信や、システム間での疑わしいトラフィックがないかを検出すると警告を発する。IDSあるいはNDRとしての使い方。
2. WAFとして
   （ア） オンプレミスで配置し、L7のトラフィックをヘッダーからボディまで検査し、不正なトラフィックを遮断し、また疑わしいトラフィックについては警告を発する。WAFとしての使い方となるが、一般のWAFと異なりボディまで検査するため、従来よりも幅広く不正トラフィックを検出することができる。
   （イ） クラウド上にCloudCoffer のWAFエンジンをお客様毎に配置し、Webサーバを守るクラウド型WAFサービスとして提供。アリスからCCC（CloudCoffer on Cloud）の名称でサービス提供中。
3. Sandboxとして
   （ア） オンプレミスで、アプリケーションサーバ、ファイルサーバ、Webサーバに送られてくるファイルがマルウェアでないかどうかを、アンチウィルスソフトのようにハッシュ値で調べた上で、ハッシュ値では判別できないものを実

際に仮想空間で動作させてマルウェアの可能性を判別する。メールサーバの添付ファイルやメール文中のURLなどを検査するような構成も可能。

（イ）　クラウド上にSandboxを配置してマルウェア検知を行うクラウド型サービスをアリスから提供予定。

参照

[1] https://dl.acm.org/citation.cfm?doid=2556647.2560219
[2] https://arxiv.org/pdf/1808.09700.pdf
[3] http://www.cs.cmu.edu/~02317/slides/lec_8.pdf
[4] http://www.cs.cmu.edu/~mmaass/pdfs/dissertation.pdf