企業内ネットワークの通信ログを用いたサイバー攻撃検知システム

大谷 尚通 †　　北野 美紗 †　　重田 真義 †

† 株式会社 NTT データ
135-8671　東京都江東区豊洲 3-3-9
{ootanihs,kitanoms,shigetam}@nttdata.co.jp

あらまし　最新のサイバー攻撃は, ウイルス対策ソフトで検知できるまでに時間が掛かり, 感染の未然防止が難しくなった. ネットワークセキュリティ機器も, 短時間の通信の情報だけでは攻撃の判別が難しくなった. そこで, ウイルス本体などの攻撃の主体を検知するのではなく, 攻撃に関係する通信のログや攻撃を受けたコンピュータのログから複数の特徴を探索し, 同攻撃によるインシデント発生を検知する方式を提案する. 既設のセキュリティ機器やネットワーク機器等のログを用いて, インシデントの早期検知と影響を把握する試作システムを構築し, 有効性を確認した. 本稿では, 同システムの攻撃検知方式の設計思想や運用の実績, および判明した課題を述べる.

Cyber Attack Detection System Using A Communication Log of The Enterprise Network

Hisamichi Ohtani†　　Misa Kitano†　　Masayoshi Shigeta

†NTT DATA Corporation
Toyosu 3-3-9, Koto-ku, Tokyo 135-8671, JAPAN
{ootanihs,kitanoms,shigetam}@nttdata.co.jp

Abstract　In the latest cyber attack it is difficult to prevent infection with computer-virus, since attack detection by anti-virus software often delays. It is also difficult for security devices to identify an attack only with information obtained from its short communications. Therefore, we propose a method to detect the incident due to cyber attacks, using multiple features from victim computers' logs and communication log. We confirmed the effectiveness by building a prototype system to understand the impact and early detection of the incident. In this paper, we describe design concept of attack detection method, operational performance, and problems of the prototype system.

1　はじめに

今や企業は, インターネット接続なしでビジネスができない. ほとんどの企業の社内ネットワークは, インターネットと接続されている. 社員は, 毎日のようにオフィスの PC を使って, お客様と電子メールをやり取りしたり, 検索エンジンで調べ物をしたり, クラウドサービスを利用したりする. 企業の情報セキュリティ管理部門は, この企業の社内ネットワークや社員のオフィス PC をサイバー攻撃から守らなければならない.

しかし, 最新のサイバー攻撃はセキュリティ対策が難しくなってきている. 侵入されて半年以上経って被害が大きくなってから, サイバー

攻撃が発見された事件が幾つもある. なぜなら, ブラックリストや定義ファイルの情報更新が追いつかず, ウイルス対策ソフトや URL フィルタ,IPS/IDS だけでは, 最新のサイバー攻撃の未然防止が難しくなってきているためである. そのため, サイバー攻撃の未然防止だけでなく, インシデントの発生を早期に検知して被害を最小化できる対策も求められている.

そこで筆者らは, 最新のサイバー攻撃によりマルウェアが感染したとしても, それを早期検知できるサイバー攻撃検知システムを開発した.

2 サイバー攻撃対策の問題点

最新のサイバー攻撃への対策の問題点は, 以下の 4 つの要因へ整理される.

1. **定義ファイルの配布タイムラグ**
   最近は, コンピュータウイルスやトロイの木馬プログラム等の多種の不正プログラムが次々と作成されている. そのため, セキュリティ企業によるブラックリストや定義ファイルの情報更新が追いつかず, 最新のサイバー攻撃を検知できるようになるまでにタイムラグが発生する. このタイムラグにより, 最新のサイバー攻撃を検知できずにマルウェアに感染されてしまう. さらにマルウェアにウイルス対策ソフトを無効化されてしまうと, より検知が遅れて被害が拡大してしまう.

2. **定義ファイル未対応による検知漏れ**
   狙っている組織やシステムに合わせて, 攻撃コードやマルウェアのカスタムメード化が進んでいる. 標的型攻撃や水飲み場型攻撃と呼ばれる, 特定の人物や組織を狙って特化した攻撃は, 情報が少なく, セキュリティ製品ベンダもウイルス対策ソフトの定義ファイルが提供できない. このようにマルウェアの検知や駆除ができない場合も増えてきた.

3. **URL フィルタ遮断の限界**
   これまでは, 不審な Web サイトへアクセスすると,Drive-by-Download 攻撃によってマルウェアへ感染させる方法が主流だった. しかし最近は, 信頼できる組織が運営する正規の Web サイトを改ざんしておとりサーバに仕立てて, 全く気付かれずにマルウェアへ感染させる手法が用いられている. 正規の Web サイトの改ざんに気付いて URL フィルタで遮断するまでにタイムラグがあるため, その間にユーザがアクセスして感染してしまう. 改ざんされた Web ページも頻繁に変更されるため, すぐに URL ブラックリストが陳腐化してしまう.

4. **判断に必要な情報量の不足**
   高度なサイバー攻撃で使用される通信は, 通常の通信との区別が難しくなってきている. IPS 等のネットワークセキュリティ機器が, 短時間の通信から得られる情報だけでは, 検知できなかったり, 誤検知が多くなったりする.

このように, 検知漏れ（False Negative）の問題により, 定義ファイルを用いてマルウェア本体を検知する方式や URL フィルタを用いて危ない Web サイトへのアクセスを遮断する方式では, 感染の未然防止が困難になってきた. つまり予防対策だけでは, 最新のサイバー攻撃による情報セキュリティインシデントの発生を完全に止めることができない.

3 サイバー攻撃の分析

最新のサイバー攻撃は,SEO ポイズニングやフィッシングサイト, クリックジャッキングのように人間の視覚を騙す方法やソーシャルエンジニアリングのように心理的な落とし穴を悪用する方法, ドライブバイダウンロードのようにソフトウェアの脆弱性を悪用して気付かれずに自動的に感染する方法など, 攻撃を成功させるために複数の段階を経る複雑で高度な方法が用いられている. このように高度化されたサイバー攻撃は, 企業において大きな問題となっている.

そこで, この大きな脅威になっているサイバー攻撃「標的型攻撃」と「Web 待ち伏せ攻撃」の手法を分析した. 分析結果をもとに効果的な早期検知の方法を検討する.

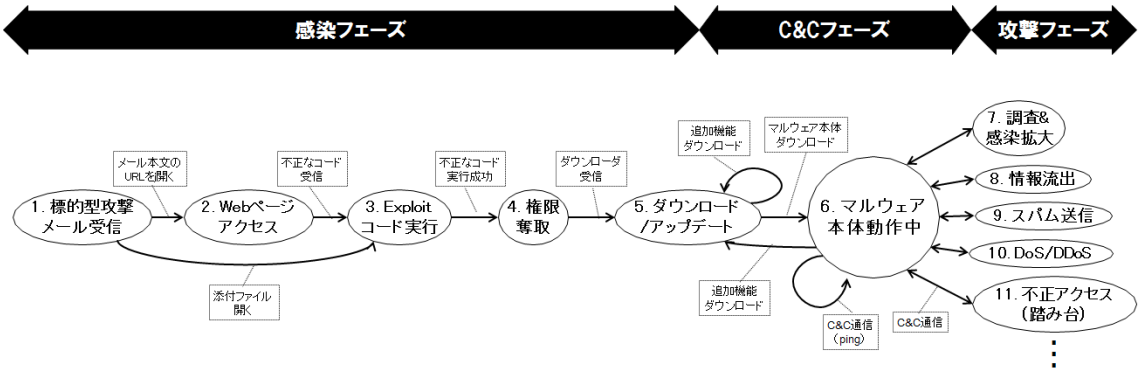

図 1: 標的型攻撃の状態遷移

3.1 サイバー攻撃のモデル化

この 2 つの高度化されたサイバー攻撃の感染から攻撃までの攻撃動作を分析してモデル化したものを説明する.

3.1.1 標的型攻撃

図 1 は, 標的型攻撃の攻撃動作の状態遷移図である. 標的型攻撃は, 主に電子メールを用いて, マルウェアを標的のユーザの PC へ感染させる. まずは, 電子メールを足掛かりにダウンローダーと呼ばれる補助的なマルウェアを標的のユーザの PC へ感染させてから, マルウェア本体をダウンロードして感染させる. マルウェア本体が起動してボット化すると,Command and Control（以下,「C&C」という.）サーバと通信したり,C&C サーバから命令を受けて次のサイバー攻撃を行ったりする. この攻撃の状態遷移を大きく 3 つに分け, それぞれ感染フェーズ,C&C フェーズ, 攻撃フェーズと呼ぶこととする.

3.1.2 Web 待ち伏せ攻撃

図 2 は,Web 待ち伏せ攻撃の攻撃動作の状態遷移図である. Web 待ち伏せ攻撃は, 有名な Web ページや有用な情報が掲載された Web ページなど, ユーザがアクセスする可能性の高い Web ページへ罠を仕掛けておき, 当該 Web ページをアクセスしたユーザの PC へマルウェアを感染させる Drive-by-Download 攻撃の一種である. 特定の人物や組織を狙った同様の攻撃は, 水飲み場型攻撃 (Watering Hole Attack) と呼ばれている. この攻撃方法は, 水飲み場型攻撃と少し異なり, 不特定多数の Web 閲覧者を狙うため,Web 待ち伏せ攻撃と呼ぶこととする. 特に「おとり」「リレー」「Pre-Exploit」「Exploit」を使って段階的に感染する方式は, マルネット 4 層モデル [1] と呼ばれている. このマルネット 4 層モデルを参考に分析した同攻撃の全体像を図 2 に示す. 図 2 の破線で囲まれた C&C フェーズと攻撃フェーズの状態遷移の構造は, 図 1 の標的型攻撃と同じである.

3.2 攻撃動作の特徴

標的型攻撃も Web 待ち伏せ攻撃も, どちらも感染フェーズ,C&C フェーズ, 攻撃フェーズの 3 つのフェーズから構成される. 各フェーズの攻撃動作とその特徴を以下にまとめた.

- 標的型攻撃の感染フェーズの攻撃動作は, 電子メールへ添付したマルウェアを実行させて感染させたり, 電子メール本文に記載された URL へアクセスさせてマルウェアを感染させたりする動作である. 通常の電子メールの開封動作とあまり違いがなく, 特徴がない.
- Web 待ち伏せ攻撃の感染フェーズは,Drive-by-Download 攻撃を用いる. 特に Exploit Kit と呼ばれるソフトウェアパッケージを使ったマルネット 4 層モデルによる感染のしくみには, 特徴がある. 正規の Web ページに埋め込まれた iframe や JavaScript によってリレー用の Web ページへ転送され

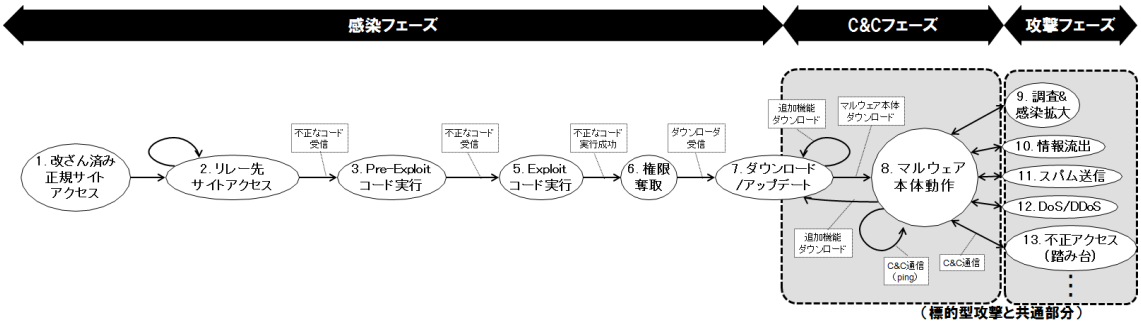

図 2: Web 待ち伏せ攻撃の状態遷移

たり,Pre-Exploit コードの動作結果によって Exploit コードをダウンロードするアクセス先 URL が変化する．感染フェーズの通信に現れる特徴は,Exploit Kit ごとに異なる [2].

- 標的型攻撃と Web 待ち伏せ攻撃の C&C フェーズと攻撃フェーズの攻撃動作は, 共通している. この部分には, 以下の 2 箇所に特徴がある. まず, ダウンローダーと呼ばれる補助的なマルウェアを標的のユーザの PC へ感染させてから, マルウェア本体をダウンロードして感染させる．マルウェア本体の感染後の C&C 通信は,IRC プロトコルや RAT (Remote Access Trojan) 通信, 特徴のある HTTP 通信が使用される [3].

3.3 新しく開発されるサイバー攻撃の特徴

標的型攻撃と Web 待ち伏せ攻撃の攻撃動作には, 共通した特徴があった. このように, 高度化されたサイバー攻撃は一部に共通した動作やしくみを持つ場合が多い．高度化されたサイバー攻撃は動作が複雑なため, 全て新しい手法で構成された攻撃は, 開発にコストと期間が必要であり, 非常に稀である. したがって, 新しい攻撃手法は攻撃動作の一部だけで, 既存や類似の攻撃手法を組み合わせたサイバー攻撃を開発することが多い. すなわち, もし別の新しいサイバー攻撃が出現しても, その中に含まれている既存の攻撃手法の特徴を捉えれば, この攻撃を検知できる可能性がある.

4 提案方式

2 章の問題点と 3 章の分析結果をもとに検討を進め, 以下のサイバー攻撃検知システムのコンセプトとアーキテクチャを決定した.

4.1 基本方針

サイバー攻撃検知システムは, 最新の高度化されたサイバー攻撃が検知可能で, かつ当社社内への導入できる方式でなければならない. プロトタイプシステムを構築して, 当社社内の実環境で運用し, 評価することを目標に定めてアーキテクチャを検討した．また, 同システムの効果が認められた場合は, 社内への本格導入も考慮する. そのため, 当社社内へ同システムを導入する場合を想定し,「既存の社内システムへの影響が少ないこと」「新規投資する対策コストを押さえること」という制約条件を追加した．4 つの問題点と上記の 2 つの制約条件に対して, 以下の解決方針を採用した.

- 既存の検知方式へ, 定性的な特徴やその遷移のパターンを用いた検知方式を追加
- 独自のブラックリストや検知パターンを自社開発
- 既存リソース (設置済みセキュリティ機器) の利活用
- 定常監視, および高度分析を合わせた継続的な運用

4.2　システムアーキテクチャ

本研究では, ネットワーク機器やネットワークセキュリティ機器のログから, サイバー攻撃の攻撃動作の特徴を検知する方式を提案する. 以下にその理由を説明する.

4.2.1　検知方式

3 章において, 標的型攻撃や Web 待ち伏せ攻撃は,3 つのフェーズに分割でき, それぞれに特徴的な攻撃動作を持つことを説明した. 2 つの攻撃の特徴的な攻撃動作のうち, 定性的な特徴やその遷移に注目することにより, 汎用性の高い検知方式 [2] を実現できる. 定性的な特徴やその遷移とは,Pre-Exploit 攻撃や Exploit 攻撃時に攻撃者サーバから標的のブラウザへ送り込まれるファイルの種別や,Exploit 攻撃を受けた標的のプログラムが攻撃者サーバへ渡す UserAgent 情報, 引数の有無, 受信データサイズの多寡など, 標的側のシステムの処理に依存するために, 攻撃者が詐称できない情報や状態遷移レベルの情報である.

4.2.2　検索対象データ

本試作システムは, ネットワークから通信データを収集して, 特徴的な攻撃行動を検知する方法を採用した. オフィス PC から情報を取得する場合は, 情報収集用のエージェントプログラムを全オフィス PC へ設置して, 定期的に情報を収集しなければならず, 全オフィス PC の設定変更と情報を収集するための通信が社内ネットワークへ与える影響が大きい. また, 既にネットワーク通信から標的型攻撃を検知する方式 [3] が提案されている. 社内ネットワークとインターネットの境界など, 少ない箇所から効率的に通信情報を収集し, サイバー攻撃を検知できる. さらに, 通信をキャプチャする方式ではなく, ネットワーク機器やネットワークセキュリティ機器のログなどから通信情報を取得する方法を採用した. この方法ならば, ネットワークやシステムの構成変更が不要で, かつ既存のリソースを利活用できる.

図 3: 標的型攻撃 (感染フェーズ)

図 4: Web 待ち伏せ攻撃 (感染フェーズ)

自社への試験的な導入を前提に, 同サイバー攻撃の検知に効果的なログ (図 3〜5 参照) を検討した. 例えば, セキュリティを考慮している企業の社内ネットワークは, オフィス PC からインターネットへの直接通信を Firewall によって遮断し,Proxy サーバを経由した HTTP 通信のみを許可している場合が多い. そのため, 近年は HTTP 通信を使うマルウェアが多い. HTTP ステータス,HTTP ヘッダの情報から, マルウェアの特徴的な通信を検知する方法 [6] も提案されていることから, 本方式も Proxy サーバのログを利用してサイバー攻撃を検知する.

特徴的な攻撃行動に応じて,Proxy サーバ以外のログも利用できる. 例えば, ダイナミック DNS を使用したリレーサーバや C&C サーバと, 感染 PC の間の通信は,DNS ログから検知しやすい. Proxy サーバに対応していないマルウェアの通信は,Firewall の遮断ログに通信を試行した痕跡が残る. 複数のネットワーク機器やセキュリティ

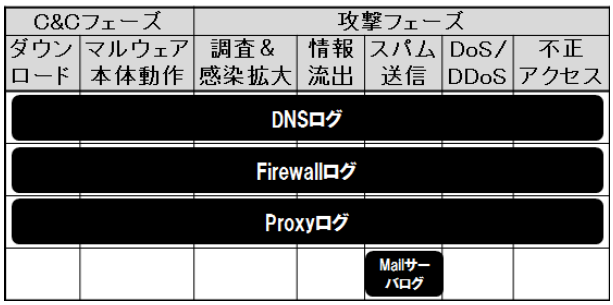

図 5: 2 つの攻撃共通 (C&C〜攻撃フェーズ)

機器のログを組み合わせることによって, サイバー攻撃を検知しやすくなる.

4.2.3 データ処理方式

高度化されたサイバー攻撃は HTTP 通信を使うため, 短時間の通信から得られる情報で, サイバー攻撃の通信と正常な通信を見極めて, 攻撃を検知することが難しい. そのため, 監視する時間や取得する情報の種類を増やして, より多くの情報からサイバー攻撃を検知する. 以前から, 複数台のセキュリティ機器や異なる種類のセキュリティ機器の情報を組み合わせて, それらの相関性を分析してサイバー攻撃の有無や影響を判断する方式 [7] が提案されている. また, 感染フェーズの特徴からサイバー攻撃を検知できなかった場合は,C&C フェーズと攻撃フェーズの特徴を用いて検知しなければならない. 感染フェーズは, 数分から数時間で行われると言われているが,C&C フェーズと攻撃フェーズは, 数か月から年単位におよぶ場合がある. したがって, サイバー攻撃の検知には, 長期間のログが必要になる. 本システムには, 大量のデータを蓄積でき, 高速に検索できることが求められる.

複数台数の機器のログや異なる種類のログを統合して, セキュリティインシデントを分析するシステムには, 情報の集約 (Aggregation), 正規化 (Normalization), 相関分析 (Correlation) の 3 つの基本処理が必要といわれている [8]. 本方式も, この 3 つの基本処理に基づいて設計した.

4.3 運用方式

ログサーバからログを自動的に収集して, サイバー攻撃を検知するシステムの構成と処理の流れを図 6 に示す. サイバー攻撃の検知処理は, サイバー攻撃検知システムを中心とした 3 段構成で運用する. 以下にそれぞれの処理内容を示す.

1. **簡易チェック** 従来のブラックリストを用いた単純な検索処理によって不審な通信を自動的に検知
2. **検知パターンによる検索** URL 文字列の特徴だけでなく, 攻撃動作のような定性的な特徴やその遷移も検索し, 自動的に検知する
3. **専門家による高度な分析** セキュリティ専門家が, 定期的に手動分析する. 検索条件を変化させたり, 新しい検知パターンを使ったりして, 自動検知から抜け漏れ (False Negative) たサイバー攻撃を検知する

標的型攻撃や水飲み場型攻撃は, 狙った組織や人物専用に作成したメール文面や攻撃コード, マルウェアを使用するため, ブラックリストやウイルス対策ソフトの定義ファイルでは検知できない恐れがある. このような自組織に特化した新しい攻撃に対応するために, 独自のブラックリストや検知パターンを自社で開発する. ニュースやコミュニティから新しい攻撃情報を積極的に収集し, その情報をもとに新たな検索パターンを開発する.

5 試作システムの実装

サイバー攻撃やそれによって引き起こされたインシデントを早期に検知し, 影響を把握できるサイバー攻撃検知システムを構築した. 以下に本試作システムの構成を説明する.

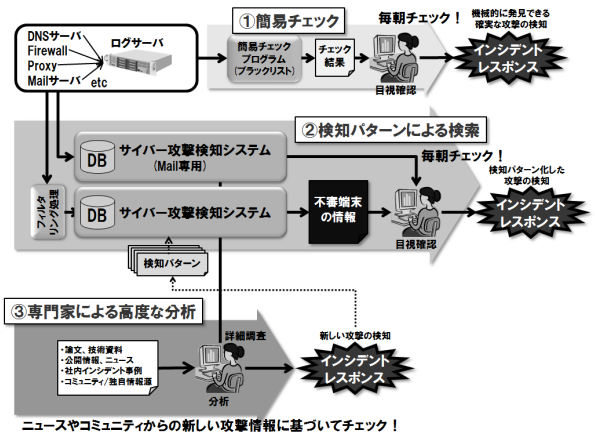

図 6: 3 段構成の攻撃検知

5.1 データベース基盤（DBMS）

データベース処理の基本能力に加え, 情報の集約, 正規化, 相関分析の 3 つの基本処理に対応できるソフトウェア Splunk [1] を選択して, データベース基盤を構築した. 検索性能や対話的な検索機能, 追加のプラグイン開発による機能の拡張性を重視し, セキュリティ用に完成されたセキュリティ情報イベント管理用のソフトウェア (SIEM:Security Information and Event Management) ではなく,Splunk を採用した.

5.2 プラグインの追加開発

Splunk の基本コマンドで不足している以下の機能は, 独自のプラグインを開発して Splunk へ導入した. 以下に追加した機能を示す.

- ドメイン/サブドメイン/FQDN の抽出処理
- 繰り返し発生する通信の検知
- スコア処理

本試作システムは, 定性的な特徴とその遷移を用いた検知方式を採用し, 検知漏れ (False Negative) を減らすことを優先して, 検知パターンのチューニングを行っている. そのため, 誤検知 (False Positive) が多くなってしまう. そこで本試作システムは, スコア処理のプラグインを開発した. スコア処理のプラグインは, 検知結果を用いてオフィス PC ごとにスコアリングを行い, 数値を用いてマルウェアへの感染の恐れや被害の深刻度を表現する. スコアの高い不審なオフィス PC から優先的に証拠保全などのインシデント対応を行う.

6 運用状況の報告

以下に, 試作したサイバー攻撃検知システムの運用実績を示す.

[1]Splunk 社のログ分析エンジン. http://ja.splunk.com

6.1 稼働中の検知パターン数

独自に開発して運用している検知パターンとその数を表 1 に示す. 70 個の検知パターンを用いてログを日次検索している. 特に感染フェーズの検知パターンは, 検知率が高く, 誤検知率が低いことが確認された [2].

表 1: 稼働中の検知パターン 一覧

感染フェーズ		個数
汎用 Exploit		3
Exploit Kit 用	BlackHole	9
	RedKit	6
	Neutrino	1
	Glazunov	2
	Sakura	1
C&C フェーズ		個数
汎用 C&C 通信		15
汎用ダウンロード		2
攻撃フェーズ		個数
汎用情報漏えい		7
その他		個数
マルウェア通信		23
合計		70

6.2 処理性能

サイバー攻撃検知システムにおいて, 平日に検索処理しているログ量, 処理時間を以下に示す. どちらも, 検索パターンの 1 個あたりに要した処理時間と処理ログ行数である. クアッドコア CPU を活用し,70 個の検知パターンは, 並列処理している.

- 1 パターンあたりの処理時間 平均 約 20 分
- 処理ログ行数 約 12×10^6行

6.3 検知実績

本試作システムの検知実績と検知のポイントを以下に示す. 本試作システムは, メールゲートウェイ方式のウイルス対策ソフトをすり抜けた標的型攻撃メールや,URL フィルタやオフィス PC 上のウイルス対策ソフトのリアルタイム

スキャンで検知できなかった攻撃コードのダウンロードなどを検知できた.

- ウイルス対策ソフトをすり抜けた標的型攻撃メール/添付ファイルの検知 (件名, 差出人などの文字列)
- 標的型攻撃メールに感染したオフィス PC の検知（C&C 通信）
- Web 待ち伏せ攻撃に感染したオフィス PC の検知 (Pre-Exploit 通信,Exploit 通信)

7 まとめ

本稿では, 実環境への導入を考慮して, 既設のセキュリティ機器やネットワーク機器等のログを用いて, 最新のサイバー攻撃を検知する方式を提案し, その試作システムの実装および運用結果を述べた. 試作したサイバー攻撃検知システムは, 定性的な特徴とその遷移を用いた検知方式を用いて, ウイルス対策ソフトでは検知できなかった標的型攻撃と Web 待ち伏せ攻撃を検知したり, 市販のウイルス対策ソフトの定義ファイルの配信より早く, マルウェアが感染したオフィス PC による通信を検知したりできた. また, 独自に攻撃情報を入手して検知パターンを開発し, 日々, 監視を継続的に運用できること, およびその有効性を確認した.

7.1 課題と今後の予定

これまで試作システムを運用した結果, 以下に示す課題が明らかになっており, その解決を検討している.

検知精度の向上 現在は, 感染フェーズの検知パターンを重点的に開発している. しかし, 感染フェーズに特徴がない攻撃は検知できない. 感染フェーズで検知できなくても,C&C フェーズや攻撃フェーズで検知できるよう, 検知パターンを強化する. 今後は, 検知パターンの増強と IDS/IPS などの他のログの利用を進める.

スケールアウト構成 検知パターン数やログ量が増加すれば, 検索処理時間が延びて, 現在のシステムでは日々の検索処理と検索結果のチェックを時間内に完了できなくなる. データベースを NAS 上へ配置し, 複数台の PC から同時に検索できるスケールアウト構成を導入する.

スコア処理の高度化 C&C フェーズの検知パターンは誤検知が多いため, スコア処理を高度化し, 毎日のチェック作業の工数を削減する.

統計分析および機械学習 本システムの検知パターンもヒューリスティック手法に基づいており, セキュリティ専門家が開発しなければならない. 今後は, システムへ蓄積された大量のデータを有効利用し, 統計分析や機械学習を応用した検知方式を開発する.

参考文献

[1] Chris Larsen, マルネットのトラッキングと可視化, ガートナー セキュリティ＆リスク・マネジメント サミット 2012

[2] 北野 美紗, 大谷 尚通, 宮本 久仁男,Drive-by-Download 攻撃における通信の定性的特徴とその遷移を捉えた検知方式

[3] 山田 正弘, 森永 正信, 海野 由紀, 鳥居 悟, 武仲 正彦:組織内ネットワークにおける標的型攻撃の検知方式, 情報処理学会研究報告 コンピュータセキュリティ CSEC62 (2012)

[4] 北澤 繁樹, 祢宜 知孝, 河内 清人, 榊原 裕之, 藤井 誠司, 標的型攻撃検知システムの評価, MWS 2009

[5] 東角 芳樹, 鳥居 悟,TCP セッションの特徴に基づくボット制御通信の検知方式の検討,MWS 2009

[6] 大谷 尚通, 与那原 亨, 馬場 達也, 稲田 勉,HTTP 利用型スパイウェアの検知および遮断方式の検討, 情報処理学会研究報告 コンピュータセキュリティ CSEC31 (2005)

[7] 大谷 尚通, 桑田 喜隆, 小迫 明徳, 井上 潮, 岩田 恵一, 広域不正アクセスに対する侵入検出・状況把握システムに関する検討, 情報処理学会 全国大会 (2001)

[8] David Swift,A Practical Application of SIM/SEM/SIEM Automating Threat Identification,SANS Institute (2006)