サイバー攻撃の解析・検知に関する研究開発

R&D of detective and analytical technology against advanced cyber-attack

研究代表者

津田　宏　富士通株式会社

Hiroshi Tsuda　Fujitsu Limited

研究分担者

塩崎　哲夫[†]　津田　宏[†]　西田 助宏[††]　高倉 弘喜[†††]　嶋田 創[†††]

Tetsuo Shiozaki[†]　Hiroshi Tsuda[†]　Sukehiro Nishita[††]　Hiroki Takakura[†††]　Hajime Simada[†††]

[†]富士通株式会社　[††]NRI セキュアテクノロジーズ株式会社　[†††]国立大学法人名古屋大学

[†]Fujitsu Limited　[††]NRI SecureTechnologies Ltd.　[†††]Nagoya University

研究期間　平成 25 年度～平成 27 年度

概要

標的型攻撃をはじめとして高度化するサイバー攻撃に対し、既存の情報セキュリティ対策ではネットワークへの侵入、マルウェアの感染等の脅威を完全に防ぐことが困難となっている。本研究開発では、利用者の行動特性及び環境特性の活用に着目することにより、組織内ネットワークにおける不正な通信等を検知し、サイバー攻撃の被害状況を把握及びサイバー攻撃の影響を最小化し業務を継続するネットワーク制御技術を開発した。

１．まえがき

近年、標的型攻撃をはじめとしてサイバー攻撃の高度化・複雑化が進展し、既存の情報セキュリティ対策ではネットワークへの侵入、マルウェアの感染等の情報セキュリティ上の脅威を完全に防ぐことが困難となっている。

こうした状況においてサイバー攻撃の被害を最小化するには、攻撃を早期に検知し迅速に対処することが重要である。

そのため、本研究開発においては利用者の行動特性等に応じて不正な通信の痕跡を発見し、ネットワークへの侵入及びマルウェアの感染等のサイバー攻撃による被害の程度及び被害に至った経緯を明らかにする技術、及び当該情報に基づきサイバー攻撃への動的な防御を実現する技術の研究開発を実施した。

２．研究開発内容及び成果

２．１．利用者の行動特性に基づいた、サイバー攻撃を早期に検知する技術の研究開発（富士通）

本研究開発のシステム概要を図 2.1.1 に示す。利用者の心理・行動特性の分析により人や組織の ICT リスクを判定し、利用者の通信にまぎれた不正通信をネットワークで検知、さらにメールや Web などの行動シーケンスから標的型攻撃などをリアルタイムに検知する技術から成る。

これらの技術を組合せ、組織の静的・動的リスクをダッシュボードとして見える化するシステムを試作した(図 2.1.2)。行動特性分析により個人や組織の ICT 被害のあい易さを静的なリスクとして算出。メールや Web の履歴を元にリアルタイムで標的型攻撃を検知し、その対策としてネットワーク検知センサーのポリシーを強化したり、似たようなメールをやり取りしている人のメールポリシーを強化したりなど、先回りした防御を取ることが可能になる。

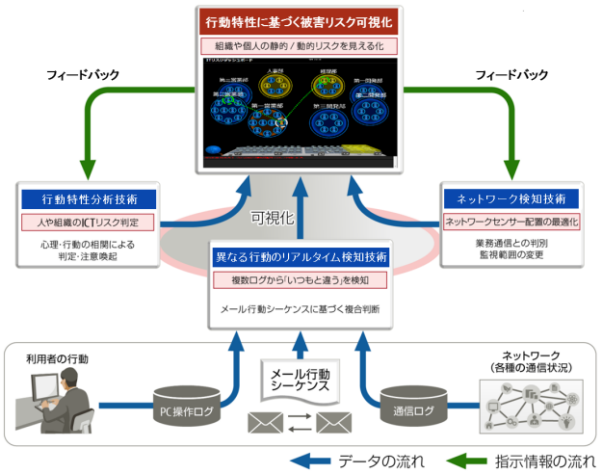

図 2.1.1 行動特性に基づく被害リスク検知システム概要

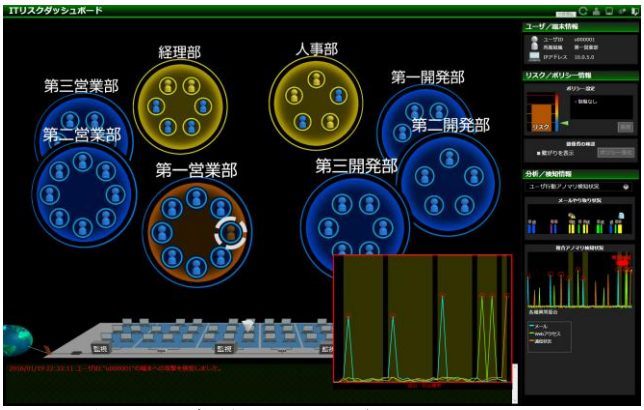

図 2.1.2 各種リスクの見える化システム

２．１．１．利用者の行動特性分析技術

詐欺・情報漏洩・ウィルス感染といった ICT 被害に遭いやすい人の行動特性を、2,000 名以上の Web アンケートと社会心理学の知見も生かした 200 項目以上の PC 操作ログの相関分析に基づき分析した。例えば、ベネフィット認知の高い人(赤信号でも道路を渡るなど、リスクが高くて

もメリットを優先する人)はウィルス・詐欺のリスクが高い、プライバシーポリシーをあまり読まない人はウィルス感染リスクが高いなどの傾向が得られた。

この分析結果に基づくリスク判定ツールを開発し、国内外の展示会にて 500 名以上に対して試行を行った(図 2.1.3)。その結果、職種においては営業、業種については金融や教育のリスクが高いなど、クラスによってリスクが異なることが分かった。組織や人のリスクに応じた、きめ細かいセキュリティ対応が必要と言える。

図 2.1.3 ICT リスク分析結果の表示例

また、リスクに応じた対策についてアンケート調査を実施し、周りの人の行動を適切に共有することがセキュリティ対策としても有効であることが分かった。サイバー攻撃を受けた際に、周囲の関連する人とのつながりから先回りして防御するなどの対策が考えられる。

なお、行動特性をセキュリティ対策に利用するにあたって、プライバシーの配慮も必要となる。上記リスク判定ツールの試行において受容性に関しても海外も含め調査を行ったところ、取得データの匿名化や、ローカル PC でのみ処理するなど共有範囲を制限することで受け入れられる可能性を確認した。

２．１．２．利用者の行動特性分析に基づく不正意図の検知技術及び通信の制御技術

組織内ネットワークを監視・解析し、その情報を統合サーバに収集・統合することで、利用者の通常通信に紛れた不正意図を抽出する技術を開発した。

単独のセンサー、もしくは複数のセンサー装置が連携してマルウェアの諜報活動を解析・検知し、踏み台を特定する技術を開発した(図 2.1.4)。実験環境にて、入手したマルウェア(9 種類)を動作させた場合、すべて検知することができた。この中には、既存のアンチウィルスソフトや IDS では検知できないマルウェア活動(諜報活動)も含まれる。

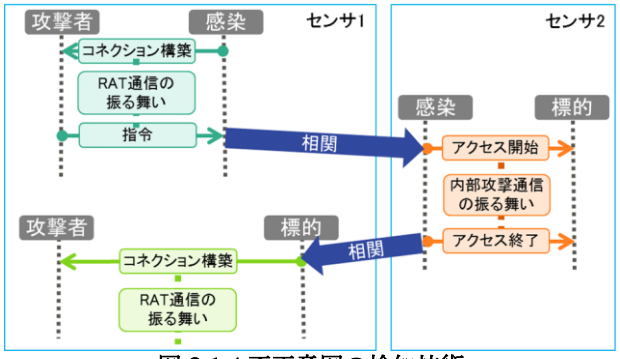

図 2.1.4 不正意図の検知技術

また、検知技術の最適配置技術においては、本技術を搭載した小型ネットワークセンサー装置(Intel Integrated Processor 600MHz/1 コア, メモリ 4GB、ストレージ 1GByte)試作により 10～20 台程度の小規模ネットワークにおけるマルウェア検知、また大型センサー(CPU: Xeon 1.7GHz / 4 コア、メモリ 4GB)により数百台程度の PC によるマルウェア検知が可能であることを確認した。

２．１．３．利用者の行動特性分析に基づくリアルタイム・アノマリ分析技術

通信データ、メール操作、Web アクセスなどの個人および組織の行動ログに対して複数の検知処理を適用し、それらの結果を組み合わせてパターン化したものを普段の行動として学習し、その学習結果と現状の行動の違いをアノマリとして検知する複合判断機能を開発した。本技術は、分散並列型 CEP(Complex Event Processing)技術を機能拡張により実現した。

また、一つひとつの時系列データ（イベント）を高速に分析するとともに、長期間に及ぶデータをコンパクトに効率良く処理する高速化機能を開発した。これにより、マルウェアのような短時間でのアノマリに加え、長期間に渡るアノマリのリアルタイム検知も可能となった。

やり取り型の模擬標的メール攻撃に対して、本技術を適用し、社内の実証環境にて検知実験に成功した。やり取り型標的型メール攻撃では、攻撃者から複数回のメールのやり取りを通して、利用者に攻撃者を信頼させたり、利用者の心理的な隙を突いたりして、利用者を攻撃サイトへ誘導する攻撃手法が取られる。単純な時系列のアノマリ検知では、関係ない利用者のメールや Web 行動により検知にゴミが多くなる。そこで、メールを起点とする Web やファイルダウンロードなど一連の操作を関連付ける(図 2.1.5)ことにより、従来技術(各ログ毎の単体検知)に比べて、検知のためのデータ量を 1/10 以下、また過剰検知も 1/10 以下に抑制できた。

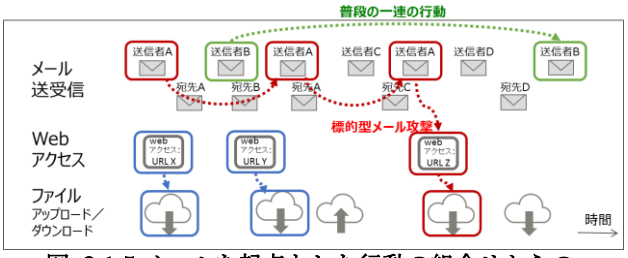

図 2.1.5 メールを起点とした行動の組合せからのアノマリ検知

２．２．端末および、ゲートウェイにおける「グレーアクティビティ」の蓄積による攻撃経路の解析と被害範囲の特定技術の研究開発(NRI セキュア)

本研究開発のシステム概要を図 2.2.1 に示す。

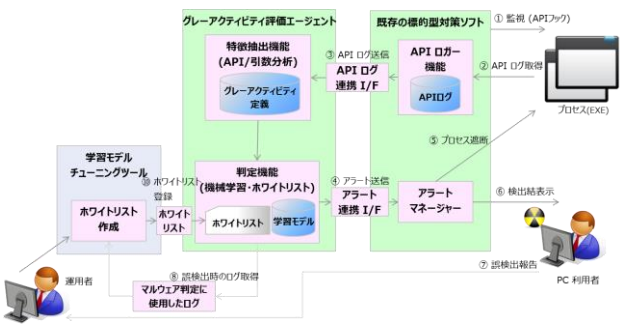

図 2.2.1 システム概要図

既存の振る舞い検知型の標的型攻撃対策ソフトウェアと、本研究開発技術であるグレーアクティビティ評価エージェントを連携させる。ここで、誤検知抑制対策のツールとして、学習モデルチューニングツールをグレーアクティビティ評価エージェントに組み込んでいる。

2．2．1．グレーアクティビティ評価エージェント

グレーアクティビティ評価エージェントは、SVM(Support Vector Machine)を採用したマルウェア検出エンジンである。グレーアクティビティ評価エージェント内で処理される特徴抽出から学習モデル作成までの流れを図 2.2.2 に示す。

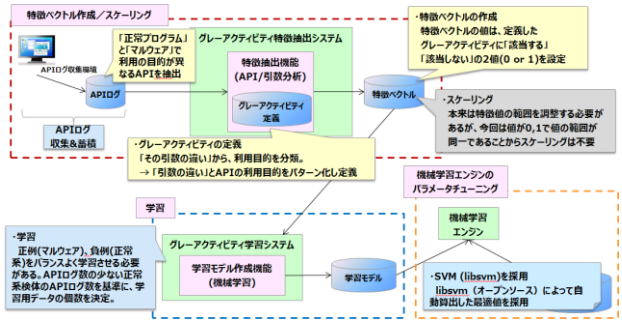

図 2.2.2 特徴抽出から学習モデル作成までの流れ

グレーアクティビティ評価エージェントでは、正常プログラムまたはマルウェアから呼び出される Windows API の引数を分析し、Windows API の利用有無と引数をルール化したグレーアクティビティの有無を特徴ベクトルとして利用している。今回利用したグレーアクティビティは、5749 個のマルウェア検体と 116 個の正常ソフトウェアから API ログを取得して分析したものである。

2．2．2．学習モデルチューニングツール

グレーアクティビティ学習システムは、既存の振る舞い検知型の標的型攻撃対策ソフトウェアでは検知できない検体に対して高い効果を発揮できることを評価した。しかし、一方で誤検知率は約 5%と製品化を考えると高い値であり、誤検知の抑制が課題として挙がった。

学習モデルチューニングツールは、グレーアクティビティ評価エージェントの製品化に向けて課題となる機械学習エンジンの誤検知を抑制するためのツールである。機械学習エンジンが誤検知したときに出力するログファイルを読み込み、どの特徴が検知結果にどの程度の影響を与えているのかを調べ、その結果を基に特徴の有無を再調整し、機械学習エンジンのホワイトリストへ登録する機能を持つ。

学習モデルチューニングツールによって作成されたホワイトリストは、利用者の PC のレジストリ上に書き込まれ、以降、グレーアクティビティ評価エージェントでは、マルウェア判定の際、レジストリに書き込まれたホワイトリストを読み出し、判定する特徴と一致するかを確認する。そして、確認の結果、ホワイトリストと一致した場合には判定を終了し、ホワイトリストと一致しなかった場合には、通常通り抽出した特徴を識別機に入力して判定を行う。

2．2．3．評価とまとめ

グレーアクティビティ評価エージェントは、既存の振る舞い検知型の標的型攻撃対策ソフトウェアでは検知できない検体に対して、高い効果があることを検証できたが、同時にグレーアクティビティを含まないマルウェアに対しては効果がなく、既存の振る舞い検知型の標的型攻撃対策ソフトと連携させることによって相互に特徴が異なるエンジンを組み合わせて補完しあうことが望ましいと考えた。

従って、試作したグレーアクティビティ評価エージェントを改修し、既存の振る舞い検知型の標的型攻撃対策ソフトウェアと連携させることで、よりセキュアなシステムを開発した。そして、評価の結果、これまでに収集した既存の振る舞い検知型の標的型攻撃対策ソフトでは検知できない検体に対して、グレーアクティビティ評価エージェント単体では検知率 63.4%、誤検知率 4.9%いう評価結果を得ることができた。また、既存の振る舞い検知型の標的型攻撃対策ソフトウェアと連携させることにより、最終的に検知率 88.7%、誤検知率 1.7% という高い評価結果を得ることができた。

一方で、グレーアクティビティ評価エージェントで使用しているグレーアクティビティおよび学習モデルは、時間とともにマルウェアの特徴が変化することによって性能が劣化する場合があるため、グレーアクティビティや再学習などの定期的なメンテナンスが必要であることがわかった。どのようなタイミングで学習モデルを作り直すのか、作り直した学習モデルをどのように配布するかについては、運用上の課題となりえる問題であり、今後検討していく必要がある。

誤検知に関しては、学習モデルチューニングツールを用いた誤検知率の低減を実施したが、ツールを用いた誤検知の抑制を実施しすぎると検知率までも低下させてしまうことがわかった。検知率への影響を抑えつつ、誤検知率を下げて運用の負荷を減らす工夫としては、誤検知の抑制においても既存の振る舞い検知型の標的型攻撃対策ソフトウェアの当該機能を活用する方法が考えられる。その場合には、既存の振る舞い検知型のソフトウェアとグレーアクティビティ評価エージェントそれぞれのホワイトリスト機能の特徴を生かし、下記の通り設定を行うことが望ましい。

グレーアクティビティ評価エージェントのホワイトリストでは、自己解凍書庫のように、同じグレーアクティビティの組合せを持つが、多様な種類が存在するソフトウェアを対象とする。

振る舞い検知型のソフトウェアのホワイトリストでは、業務アプリケーション等、特定用途に使用するソフトウェアを対象とする。

このように、正常系ソフトウェアにのみ広く認められる特徴に対しては、グレーアクティビティ評価エージェントのホワイトリスト機能は有効であるが、正常系ソフトウェアとマルウェア双方が持つ特徴に対しては、誤検知を抑制することにより検知率が低下してしまう場合がある。このような場合には、振る舞い検知型のソフトウェアのホワイトリスト機能の活用を有効な対策として挙げることができる。また、更なる誤検知抑制の方法としては、グレーアクティビティ以外の特徴（検知に使用しない特徴）を定義し、それをホワイトリストに使用することにより、誤検知を抑制できる可能性もある。

最後に、グレーアクティビティ評価エージェントと振る舞い検知型のソフトウェアを連携させたシステムを利用者の端末に導入することによって、端末上の CPU やメモリ等のパフォーマンスにどの程度の影響を及ぼすのか未

検証であるため、今後それらの性能評価を行っていく必要がある。

このように、今後更なる改善を行っていく必要があるが、従来利用されていなかったグレーアクティビティを活用し、新たな標的型攻撃検知技術を確立するという本テーマの目的については、一定の成果を達成できたと考えられる。

２．３．インシデントを考慮した組織内ネットワーク制御(名古屋大学)

近年の高度化し洗練されたサイバー攻撃では、侵入検知装置などの入口対策をすり抜けてくる攻撃が多くなってきており、組織内ネットワークに被害端末が発生した後の対策が重要となっている。これは、事前に業務定義やそれに基づくネットワーク分割やアクセス制限などによる被害範囲の削減、および、被害発生後の被害範囲封じ込めと通常業務への影響を最小化することが重要になる。そこで、本課題において、被害発生前に被害範囲を最小化することを目的とした「(1)管理ポリシーに基づく自動ネットワーク構成」、被害発生時に通常業務継続と被害範囲封じ込めを目的とした「(2)緊急時における自動ネットワーク構成変更」、インシデント対応の起点となるサイバー攻撃候補の同定を行う「(3)動的管理ポリシー生成技術」について研究開発を実施した。

図 2.3.1 に本研究開発で実現する、インシデント発生を考慮した組織内ネットワーク構築・監視・制御の概要を示す。図中の括弧書きの数字は、前段落の各研究開発項目に対応している。本システムでは、ネットワーク管理者が業務情報や重要度をもとにした監視ポリシーを戦略的運用管理システムに入力し、これをもとに、ネットワーク自動設計システムが初期ネットワークの生成とサイバー攻撃監視を実施する。通信解析システムによって発見されたサイバー攻撃やマルウェア解析システムによって発見されたマルウェアの情報は対攻撃情報セキュリティ設計システムに入力された後、新たなポリシーとして戦略的運用管理システムに入力され、ネットワーク自動設計システムによって業務影響度を抑えた対策ネットワークがネットワーク管理者に提示される。このシステムは単に攻撃対策を行うのみならず、攻撃対策実施後の業務に関連するトラフィックの解析に用いることにより、攻撃対策実施によって発生した副作用の検知も可能である。

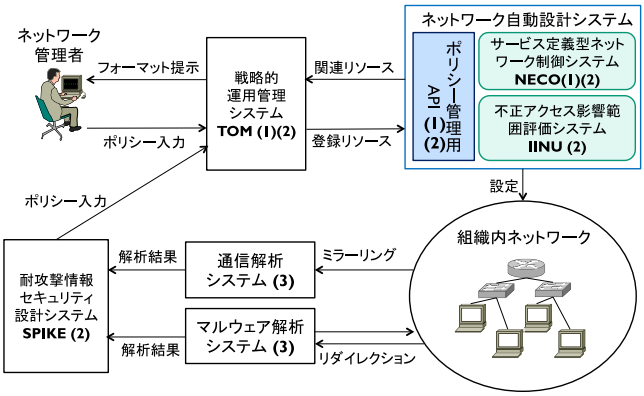

図 2.3.1 インシデント発生を考慮した
組織内ネットワーク構築・監視・制御システム

２．３．１．管理ポリシーに基づく自動ネットワーク構成技術

本研究項目では、100 台規模の情報機器が接続された組織内ネットワークに対し、(1)ネットワーク構成情報、組織情報、サービス定義情報などのリソースを関連付けた管理ポリシーを記述する手法、(2)インシデント対応による通常業務通信の遮断などの副作用検知方法の 2 点について研究を行った。

(1)については、JSON で個別に記述されたネットワーク構成情報、組織情報、業務定義情報を関連付けるシステムである。このシステムにより、新たな端末の追加などの操作時に記述すべき内容が削減されるため、管理コストが削減できることを確認した。

(2)については、インシデント対応前後の組織内ネットワークに対して実施した流量監視結果と、(1)によって記述された業務定義データベースからの情報を用いてインシデント対応によって滞っている業務の発見を補助するシステムである。これは、想定内業務変動率と想定外業務変動率を算出し、業務変動率のランキング形式の表とそれに対応する通信量変化のグラフによる可視化を行うシステムとなっている。これらのシステムによりネットワーク管理者の組織内ネットワーク管理ポリシー再設計を容易にする。

２．３．２．緊急時における自動ネットワーク構成変更技術

本研究項目では、100 台規模の情報機器が接続された組織内ネットワークに対し、サイバー攻撃を受けて不審な通信が観測される状況に対応するネットワーク構成技術について、 (1)ディレクトリ・サービスを利用した自動ネットワーク構成変更検討用情報の自動収集、(2)標的型攻撃に対するインシデント対応支援システムの 2 点について研究を行った。

(1)については、2.3.1(1)の業務情報の生成をさらに自動化するため、アクティブ・ディレクトリ等のディレクトリ・サービスから人事情報やアクセス権限を取得すると同時に、初期ネットワークの分割案とアクセス制御案を構築するシステムである。被験者評価の結果、提案システムの併用により、過剰なアクセス制限や冗長なアクセス制限を実施しがちな人間のみによるアクセス制限案よりも、最低限かつ効果的なアクセス制限を選択できることを確認した。

(2)については、初期ネットワーク情報と 2.3.1(1)から得られる組織のワークフローをもとに、分割済みネットワークとそれに対するアクセス制限実施、不正通信報告をもととした通信遮断の業務影響度評価、業務影響度と被害封じ込めのトレードオフ評価をもととした図 2.3.2 に示すようなインシデント対応案の提示を行うシステムである。これにより、ネットワーク管理者は視覚情報と文字情報を複合させて提示された複数インシデント対応案から一案を選択することによりインシデント対策を適用できるため、インシデント対策実施までの時間を短縮できる。被験者評価の結果、このシステムはより短い時間で何らかの対応案を選択させることができることを確認できた。

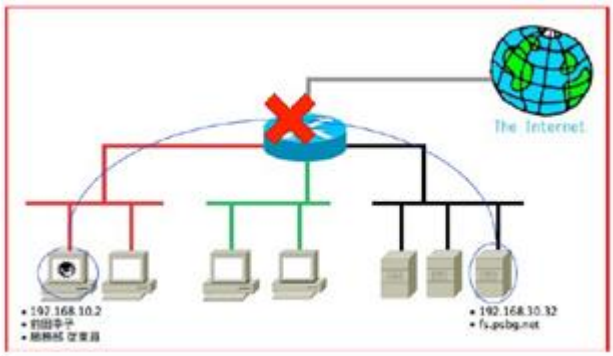

図 2.3.2 インシデント対応案の提示

２．３．３．動的管理ポリシー生成技術

本研究項目では、緊急対応のトリガとなるサイバー攻撃を起因とする不審な通信やマルウェア活動を検知するため、(1)ネットワークトラフィックから抽出した通信特徴量からの攻撃検知、(2)マルウェアファミリや新規マルウェアの分類精度向上による脅威度判定の高速・高精度化、(3)FPGA によるサイバー攻撃検出用通信特徴量抽出を実施した。

(1)については、新規のサイバー攻撃が続出している現状を踏まえ、通常通信を定義してそれから外れたものを攻撃と判断する、アノマリ検知について複数の研究を実施した。ある研究においては、時系列で分割された通信特徴量に対し、先行する時系列の通信特徴量から生成した OC-SVM による識別器に対し、後続の時系列における通信を評価させる方式の研究を実施した。この方式により、通常通信のみから作成した識別器による攻撃の検知、および、既知攻撃と通常通信から作成した識別器による既存攻撃と未知攻撃の分類の 2 種類の検知を実現した。また、通信の TCP セッションをグリッド分割で複数のクラスタに分割し、不自然なセッション遷移に対して高スコアをつける遷移スコアがアルゴリズムを開発した。いずれの方法も、誤検知率を押さえつつ高い検知率を達成した。

(2)については、新規マルウェア、および、標的型攻撃などの時間をかける攻撃で用いられるシーケンシャルマルウェア解析など、緊急度の高いマルウェアの同定に注力した。ある研究においては、マルウェアが生成する通信の TCP セッションに対してクラスタリングを行った後、クラスタシーケンスをもとにマルウェアを分類することにより、既存のマルウェアファミリのマルウェアもしくは新規マルウェアと同定する方法を実施した。さらに、判定精度と判定処理にかかる計算量を削減するため、クラスタシーケンスの解析に対する fuzzy hashing の利用と、マルウェアバイナリに対する fuzzy hashing の併用を試み、解析精度の向上を確認した。また、シーケンシャルマルウェアのダウンロードフローを安全に解析するため、通信からのバイナリ抽出と仮想マシンによる実行、感染拡大通信を遮断しつつ C&C サーバや新規マルウェアダウンロードは許可するネットワーク制御を行う、悪性通信解析システムを実現した。

(3)については、10Gbps オーバの通信ペイロードからリアルタイムに(1)で利用する通信特徴量を抽出することを目的とすると同時に、高スループット対応が可能な FPGA ならではの新規特徴量の抽出を試みた。この研究においては、パケットからの TCP セッション構築とペイロードからの 1-gram 特徴量抽出を 10Gbps クラスで実現できる回路の実現、および、オフチップメモリを利用した大量のセッションの並行構築の実現を達成した。

３．今後の研究開発成果の展開及び波及効果創出への取り組み

本研究開発の成果が社会的に活用されることで、サイバー攻撃下におけるシステムの可用性の向上、損失額の減少、利用者が要するサイバー攻撃対応の時間の減少といったアウトカムの達成が見込まれるだけでなく、その派生効果として、こうした本研究開発成果を活用した ICT 環境による高いレジリエンスを備えた社会の実現が期待される。

４．むすび

本研究の実施に当たり、研究開発全体の方針について幅広い観点から多くの助言を頂くとともに、研究開発の進め方についてご指導いただいた、神戸大学の森井 昌克教授、アルテア・セキュリティ・コンサルティング 代表の二木 真明氏、情報セキュリティ大学院大学の後藤 厚宏 教授、東京大学の高木 大資 講師、東京工科大学の村上 康二郎 准教授に深く感謝いたします。

【査読付発表論文リスト】

[1]長谷川皓一、 山口由紀子、 嶋田創、 高倉弘喜、 "標的型攻撃に対するインシデント対応支援システム"、 情報処理学会論文誌 Journal of Information Processing Vol.57 No.3 pp. 836-848 2016 年 3 月

[2]Takeaki Terada, Yoshinori Katayama, Satoru Torii, Hiroshi Tsuda, "Preliminary Investigation on Psychological Traits of Users Prone to be damaged by Cyber-attack", Poster The 11th Symposium on Usable Privacy and Security (SOUPS) 2015.7

[3]Masahiro Yamada, Masanobu Morinaga, Yuki Unno, Satoru Torii, Masahiko Takenaka, “RAT-based Malicious Activities Detection on Enterprise Internal Networks”, ICITST-2015(The 10th International Conference for Internet Technology and Secured Transactions) 2015.12

【受賞リスト】

[1]片山佳則、DICOMO2014 シニアリサーチャー賞、“IT 被害に遭いやすい心理的・行動的特性に関する調査”、2015 年 7 月 11 日

[2]富士通株式会社、Interop Tokyo 2015 “Best of Show Award”、iNetSec Intra Wall 、2015 年 6 月 10 日

【報道掲載リスト】

[1]"業界初！心理・行動特性でサイバー攻撃に遭うリスクを判定する技術を開発"、 日刊工業新聞、2015 年 1 月 19 日

[2]"攻撃リスク個人別評価 サイバー被害 PC 操作の癖分析"、 日経産業新聞、 2015 年 1 月 20 日

[3]"標的型メール攻撃 先回り警告発信"、 日刊工業新聞、2016 年 1 月 22 日