jh171006-NWJ

Proactive and Reactive Cyber Security

関谷 勇司（東京大学）

概要　本研究は、Proactive(予測的)な攻撃解析手法と Reactive(対応的)な攻撃解析手法を組み合わせることで、確度の高いサイバー攻撃の予測を目指した。サイバー攻撃の予測を行うために必要な要素を「動機」「手法」「時間」と定義し、これらを明確にするための手法を明らかにすることを目指した。各種データセットを用いて攻撃の予兆をとらえることで、攻撃者の手法と挙動を明確化することを目指した。また、ネットワーク機器から得られる情報のみならず、ソーシャルメディアから得られる社会的情報を利用し、これらを機械学習の手法を用いて判定させることで、これから発生し得るサイバー攻撃を予測することを目指した。本研究によって、Proactive なサイバー攻撃の予知ができる可能性を示すことができた。

1. 共同研究に関する情報

(1) 共同研究を実施した拠点名

東京大学情報基盤センター

(2) 共同研究分野

□ 超大規模数値計算系応用分野

□ 超大規模データ処理系応用分野

■ 超大容量ネットワーク技術分野

□ 超大規模情報システム関連研究分野

(3) 参加研究者の役割分担

関谷 勇司

本研究プロジェクトのまとめ。研究の進捗管理と方向性の決定、ならびにサイバーセキュリティインシデントの予測手法に関する研究を担当する。

Paul Pang

本研究プロジェクトの副代表として、Unitec 工科大学側の研究者をまとめる。Unitec 工科大側は、既に有している AI に関する知見を活かし、解析アルゴリズムに関する提案とデータを用いた一次解析結果の評価を担当する。

宮本 大輔

主にネットワークインフラ機器から得られるデータの収集と解析、ならびに詐称された Web サイトに関連するデータの収集と解析を担当する。

荒巻 英治

サイバーセキュリティインシデントの予測に利用可能なソーシャルデータの収集と解析を担当する。

Denis Lavrov

解析に用いるソフトウェアの準備等を担当する

Veronique Blanchet

解析データの事前処理スクリプト作成を担当する。

Muyang He

解析データの事前処理スクリプト作成を担当する。

2. 研究の目的と意義

日々巧妙化、複雑化するサイバー脅威に対して、データを用いた攻撃の解析のみならず、攻撃の予測を行うことで、より先見的な対策を行う手法を確立することを目標とする。そのために、ネットワークインフラから得られる通信に関するデータと、ソーシャルメディアから得られる社会的データを用いることで、攻撃の予測に必要となる「動機」「手法」「時間」を明らかにすることを目指す。

従来のサイバーセキュリティの研究は、発生した攻撃に対して、その痕跡を示すデータを発見するものや、攻撃の兆候をとらえ実時間に近い形で攻撃を防御することを目指したものが多い。これら先行研究の多くは経験

則を用いたヒューリスティックな手法で攻撃を特定しているものが多く、攻撃の発生自体を予測することは難しい。そこで本研究では、機械学習を用いて攻撃の兆候を学習させることで、攻撃の発生自体を高い確度で予測させることを目指す。これがProactive(予測的)な対策手法であり、このProactiveな対策手法を実現するためには、Reactive(対策的)な手法による解析結果が必要となる。そのため本研究では、このProactiveとReactiveな手法を連携させることで、より確度の高い攻撃予測を行うことを目指す。

3. 当拠点公募型共同研究として実施した意義

本研究課題は、ネットワークにおけるセキュリティ脅威を解析、予測することを目指す。そのため、(1)ネットワークインフラ機器から生成されるログ情報を用いて攻撃を発見する手法を確立し、(2)ネットワークインフラ機器からのデータとソーシャルデータを組み合わせてサイバー攻撃を予測することが必要となる。これらを満たすために、ネットワークセキュリティを専門とする研究者とAI技術を専門とする研究者が協力する必要がある。また、実際のキャンパスネットワークの運用に精通している人物を加えることにより、より実運用に役立つ解析と予測の手法を確立できる。さらに、東京大学情報基盤センターが提供するFENNEL資源は大容量ネットワーク型の資源であり、リアルタイムにストリーミングされるデータを蓄積しながら解析を行うことができる。ネットワークインフラからのデータならびにtwitterに代表されるソーシャルネットワークのデータはリアルタイムに生成されるものであり、サイバー攻撃を防ぐ観点からは、より実時間に近い形で解析と予測が行える必要がある。これらの理由から、東京大学情報基盤センター、ならびに奈良先端科学技術大学院大学、Unitec工科大学の研究者が協力し、東京大学情報基盤センターが提供するFENNEL資源を用いた専有利用型による解析が必要であった。

具体的には、FENNELの資源を利用してリアルタイムに送信されてくるネットワーク機器からの通信情報を蓄積し、解析するシステムを構築した。syslogというプロトコルにて送信されてくるデータを、ストレージノードにて受信し、後述するhayabusaというシステムにて利用するバイナリ形式に変換してデータの蓄積を行った。このsyslogプロトコルにて送付されてくるメッセージは複数のネットワーク機器やセキュリティ機器から送付されてくるものであり、このログメッセージを直接受信してリアルタイムにストレージノードに書き込むために、FENNELの特徴であるインターネットに直接接続したストレージノードを有効に活用した。また、ストレージノードに蓄積されたデータは、iSCSIによって計算ノードと共有され、計算ノード側においてhayabusaシステムを用いた高速リアルタイム検索と、TensorFlowを用いた解析に利用された。これらもFENNELの特徴である、ログインノードからのジョブ制御に依らない自由な計算ノード構成を活用した。

4. 前年度までに得られた研究成果の概要

該当無し

5. 今年度の研究成果の詳細

本年度に得られた成果は、(1)データ収集と高速検索の基盤となるHayabusaシステムの開発、(2)攻撃者のスキャニング挙動を利用した攻撃者の特定に関する研究、(3)従来のシグネチャ手法に依らない攻撃挙動の発見に関する研究、(4)DNSのログデータを用いた特異端末の発見手法に関する研究、(5)Twitter情報を用いた攻撃余地の可能性に関する研究、(6)複数データセットのリン

クによる攻撃挙動の発見手法に関する研究、である。

まず(1)に関しては、リアルタイムにストリーミング転送されてくるネットワークインフラ機器からのデータを蓄積し、高速な検索が行えるシステムを構築した。目標値は 20,000 ログ/秒であり、この速度での蓄積を実現するシステムを構築した。これは本研究の解析基盤となるシステムであり、この基板上に蓄積されたリアルタイムデータを、そのまま各種解析ソフトウェアやアルゴリズムに展開できるようシステムを構築する。現時点では複数ノードにまたがった分散ファイル蓄積と、ファイルからの並列型高速検索を実現した。

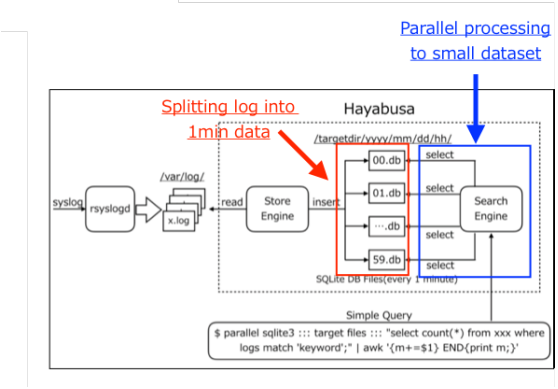

図 1 ： Hayabusa システムアーキテクチャ

図 1 に Hayabusa システムの基本アーキテクチャを示す。ネットワークインフラ機器からのデータを直接 UDP 等で受信し、受信した時間を基準としてファイル単位に分割し、単一のストレージもしくは複数のノードに付属のストレージに分散して蓄積する。蓄積された情報は、検索範囲となる時間を指定することで該当ファイルが特定できるため、該当するファイル数と検索に利用できるノード数を考慮した分散処理を行う。基本的な思想は Hadoop で実現される Map & Reduce の概念であるが、ネットワークインフラ機器からリアルタイムに送出されるデータを蓄積し、検索するのに適したアーキテクチャとなっている。キーワード検索に関する性能を、Apache Spark と Hayabusa を比較した結果を図 2 に示す。

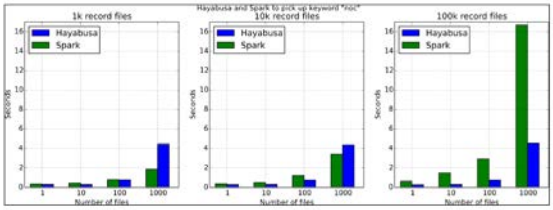

図 2 ： 検索性能比較

特定のキーワードを Hayabusa システムにて検索した場合に要した時間を青軸、Apache Spark システムの場合を緑軸にて示した。また、3 つのグラフはそれぞれ単一のファイルサイズが 1KB、10KB、100KB の場合について結果を示した。1 つのファイルサイズが大きく、かつファイル数が大きくなるほど Hayabusa システムの方が高速に検索できることを示している。この研究成果は論文[6]にて発表した。

さらに、データ量の増大に伴い、検索サーバを分散するための仕組みに関して提案を行い、現在検証中である。比較対象としては、Google Cloud Platform の BigQuery や、Elastic 社の Elasticsearch などがあげられる。どちらもリアルタイムデータを格納しながら、高速な全文検索を実現している。今後は、単なるキーワード検索のみならず、検索した結果を各種解析ソフトウェアに展開し自動的に解析ができる基盤の構築を目指す。

次に、(2)に関する成果を述べる。攻撃者が行うスキャニングという挙動をとらえ、攻撃者のスキャニング行為自体を防御する手法を検証した。既存研究においては、同一の送信元アドレスから送られてくる TCP 接続要求の数をカウントし、一定数を超えたものをブロックするという、経験則に基づいた手法にて防御が行われていた。一方本研究では、攻撃者が行うスキャニング挙動自体をパターンとしてとらえ、送信元アドレスが複数であったとしてもその挙動を攻撃者による一連のスキャニングとしてとらえ、スキャニング

早期段階にて防御することを可能にした。

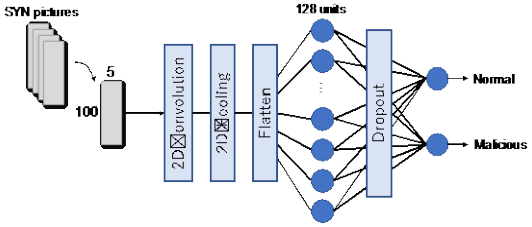

図 3: CNN を用いた SYN パケット解析

図 3 に示す通り、SYN パケットの挙動をプロットし画像として生成することで、既存の CNN を用いた画像判別の手法を適用した。従来のネットワークインフラからのデータを用いた機械学習では、IP アドレスやポート番号をそのまま数値データとして用いてベクトル化することで行列を生成する事例が多く見受けられたが、本研究ではまず SYN パケットの挙動を画像化することにより、IP アドレスやポート番号といった情報を数値化することなく、画像として認識することによって CNN による深層学習アルゴリズムを適用した。手法の検証結果を図 4 に示す。

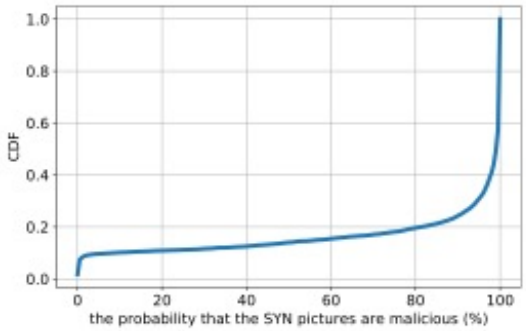

図 4 ： 手法の検証結果

画像に基づいた SYN パケットの挙動が malicious 判定である場合、実際にその通信が malicious であった割合を、縦軸の CDF グラフとして示した。結果、IP アドレスにとらわれずに、SYN パケットの送信挙動に基づいた攻撃者の特定が可能であることを示した。この成果の詳細については、論文[8]にて発表した。

さらに、(3)に関する研究成果について述べる。ユーザがブラウザ等を用いて Web ページ等のインターネット上のリソースにアクセスするためには URL(URI)とよばれる識別子が用いられる。この識別子にはドメイン名と呼ばれる、人間が認識しやすいよう命名された木構造の名前空間が構築されている。この URL に含まれる名前や、指し示している Web ページの名前、もしくはファイルの名前等から、ユーザにとって危険なコンテンツを特定する手法を提案し、検証した。現在のセキュリティ対策では、ブラックリストと呼ばれる URL や IP アドレスの膨大なリストを保持し、そのブラックリストに適合する URL にユーザがアクセスしようとした場合、ユーザの通信を遮断するという手法が用いられている。このブラックリストはセキュリティ会社や有志のメンバーによってメンテナンスされ続けており、Web サイトをクローリングし続けて発見した危険なサイトや、ユーザが実際にアクセスしてマルウェア等に感染してしまった危険なサイト等の情報が集められ、構築されている。

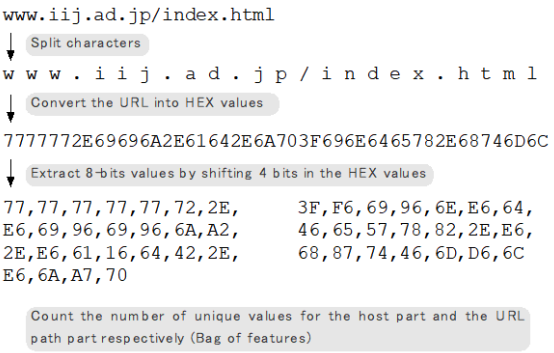

図 5 ： Bag of Bytes を用いた URL 分類

しかし本研究では、このようないわば力技に頼った手法ではなく、URL の文字列を単なるビット列として解釈し、Bag of Bytes の手法を用いて危険な URL と安全な URL を判定することを試みた。具体的には、図 5 に示す通り URL を 16 進数として表現した後にビットシフトを行うことで Bag of Bytes を実現し、さらに URL のホスト部とファイル部を分割して特徴量として扱うことで判定を行った。学

習データとしては、既存のブラックリスト URL を用い、ホワイトリストには Alexa[1]にてアクセス上位に位置するサイトを用いた。その結果、既存の研究よりも高い確度にて悪性サイトを判定できることがわかった。この研究成果は論文[5]にて発表した。

(4)の研究では、DNS のキャッシュサーバから得られるログを用いて、各端末から問い合わせが行われた FQDN に対する関連性をクエリグラフとして構成し、特異な FQDN を共有するホストをクラスタリングによって発見する手法を試みた。図 6 にクエリグラフの事例を示す。

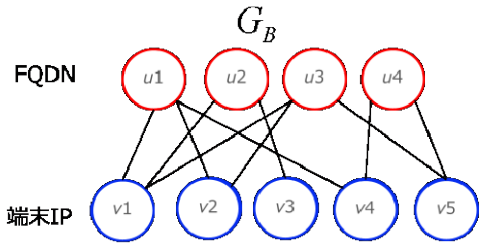

図 6 ： DNS クエリグラフ

このように、端末から問い合わせが行われた DNS クエリの FQDN をクエリグラフとして構成することで、多くの端末が問い合わせを行う共通の FQDN と、ごく少数の問い合わせしか行われない FQDN を分類することができる。分類後、ごく少数の問い合わせが行われる FQDN を基準としてその問い合わせをおこなったホスト群を調査し、そのホスト群の関連性から C&C サーバを発見する手法である。具体的には、図 7 に示すような中継点となっている FQDN とノードを発見することにより、その中継点にアクセスしているノード群の共通となる悪性ノードを発見するという手法である。

表 1 ： DNS クエリグラフによる判定結果

	データ収集月	悪性端末数	良性端末数
dataset2016	**2016 年 3 月**	**22**	**1703**
dataset2017	**2017 年 4 月**	**19**	**1734**

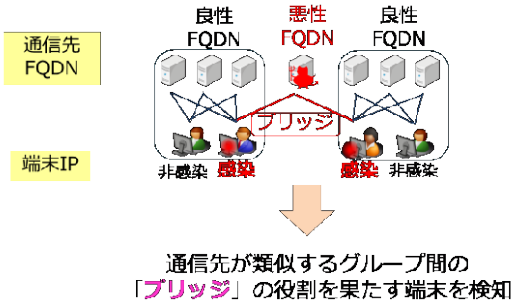

図 7 ： 中継 FQDN の発見

この手法を用いることで、著名な C&C ボットネット群である CryptoWall の感染端末を発見することが可能であることを示した。表 1 に、収集したテストデータによる悪性端末の判定結果を示す。

しかし、本研究では悪性と判定された端末が本当に感染端末であるかの真偽判定が行われていないため、手法の有用性が評価できていない。今後の課題となっている。

(5)においては、ソーシャルデータを利用したサイバー攻撃の検知可能性に関する検証を行った。本研究のソーシャルデータ収集システム概要を図 8 に示す。

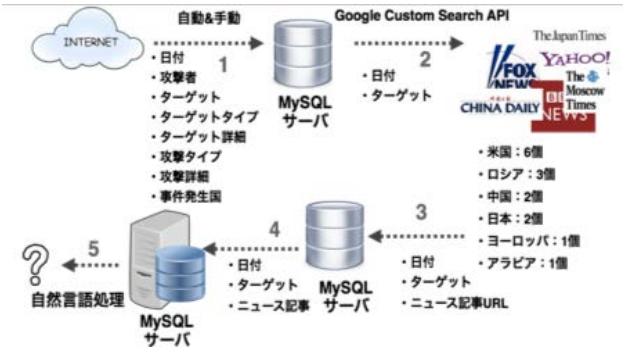

図 8 ： ソーシャルデータ収集システム

著名なニュースサイトやブログにおける記事をクローリングして収集するとともに、Twitter のフィードにて特定のユーザや特定のハッシュタグ、もしくはキーワードを含むツィートを収集して蓄積した。

このデータを用いて、実際に発生したサイバー攻撃に関して、その発生より前の時刻のニュース記事やブログ記事、ツィート等から

[1] https://www.alexa.com/

その攻撃を予測することが可能であったかどうかの検証を行った。蓄積されたデータからラベル付けされたデータを作成するために、実際に発生した攻撃に関連すると思われる記事やツィートを抜き出し、それをキーワードとして記録することで、活性化関数に ReLU を用いた K-Means による機械学習を行った。さらに、収集したデータと実際にあった過去の攻撃データを用いてクロス評価を行うことで、収集したキーワードの妥当性を評価した。その結果、92.3% の確度攻撃に関連する記事を特定することができた。本実験はまだ予備段階であり、攻撃予測に必要となる、「動機」「手法」「時間」のうち「動機」に関しての予測がソーシャルデータを用いて可能であるという可能性を示しただけに過ぎない。そのため、今後「手法」や「時間」を決定するための予測手法を明確にする必要がある。

最後に(6)の研究について述べる。本研究では、ネットワークインフラ機器やセキュリティ機器、ならびに社会的サイトから複数のデータセットを収集し、解析することで攻撃者の攻撃挙動を発見することを目指した。その攻撃者の攻撃挙動を発見するにあたって、データセット同士を連結するインジケータの存在が重要であると判断し、インジケータを定義することを行った。インジケータとして利用可能なものは、IP アドレス、発生時間、発生場所、利用者、行為者等があげられる。これらの複数のインジケータを活用し、ネットワークインフラから得られるデータのみならず、社会的データセットも含めた複数データセットの関連性を定義し、リンクグラフを作成した。この成果は論文[3]にて述べた。また、データセット間の関連を元に、関連するデータを高速に検索するためのプラットフォームとして、(1)にて開発した hayabusa と、Google Cloud Platform の BigQuery を活用することとした。全てのデータセットを共通の解析基盤に載せることで、事象が発生した際に関連するデータセットを全て抜きだすことが可能となった。

このデータセットの関連性を元に、実際に大学内で発生したインシデントからインジケータを抽出し、インジケータに基づいたデータ抽出を行い、その結果を学習データとして用いることで、Proactive なインシデントレスポンスの実現を目指した。この研究はまだ着手したばかりであり、今後のさらなるデータ蓄積と学習データの蓄積が必要である。

6. 今年度の進捗状況と今後の展望

今年度は、前述の(1)～(6)に関する研究を行った。その結果、データセットを収集し、そのデータセットの分析手法を変えることで、攻撃者の様々な攻撃挙動を把握できる可能性があることがわかった。まだ、複数のデータセットを組み合わせてその関連性を元に攻撃者の挙動を発見することで、実際のインシデントレスポンスに役立てることのできる、Proactive な攻撃予測の可能性を把握した。本年度に行うことができた研究は、全体像のまだ初歩的な部分であり、Proactive な攻撃予測を実現するためにさらなるデータセットの収集と解析手法の検討、ならびにデータセット間のリンクによる攻撃挙動の把握を明確化していきたいと考えている。

本研究にてデータセットの蓄積と解析による攻撃検知の可能性を見出すことができ、かつ Unitec 工科大学と共同研究を行うことによってインジケータを用いたデータセットの関連付けという手法を明確化することができた。この成果を元に、JST/CREST に応募することで競争的資金を獲得できたため、今後の研究は JST/CREST の枠組みにおいて研究を遂行する。

7. 研究成果リスト

(1) 学術論文

[1] Munkhdorj Baaatarsuren, and Yuji Sekiya, “Cyber attack prediction using social data analysis”,

IOS Press, Journal of High Speed Networks, vol. 23, no. 2, pp. 109-135, 2017, DOI : 10.3233/JHS-170560.

(2) 国際会議プロシーディングス

[2] Bo Hu1, Atsutoshi Kumagai, Kazunori Kamiya, Kenji Takahashi, Daniel Dalek, Ola Soderstrom, Kazuya Okada, and *Yuji Sekiya*, "Alchemist : Stochastic Feature Regeneration for Malicious Network Traffic Classification", ACM CCS2018 (submitted).

[3] Jing Zhao, Shaoning Pang, *Yuji Sekiya*, and Daisuke Miyamoto, "Task and Instance Quadratic Ordering for Active Online Multitask Learning", International Joint Conference on Neural Networks (IJCNN 2018), Rio, Brazil, July 2018. (Submitted)

[4] Ryo Nakamura, Yuji Sekiya, Daisuke Miyamoto, Kazuya Okada, Tomohiro Ishihara, "Malicious Host Detection by Imaging SYN Packets and A Neural Network", International Symposium on Networks, Computers and Communications (ISNCC 2018), Rome, Italy, June 2018. (Accepted)

[5] Keiichi Shima, Hiroshi Abe, Daisuke Miyamoto, Tomohiro Ishihara, Kazuya Okada, Yuji Sekiya, Hirochika Asai, and Yusuke Doi, "Classification of URL bitstreams using Bag of Bytes", Proceedings of the 21st Innovations in Clouds, Internet and Networks, February 2018

[6] Hiroshi Abe, Keiichi Shima, Yuji Sekiya, Daisuke Miyamoto, Tomohiro Ishihara, and Kazuya Okada, "Hayabusa: Simple and Fast Full-Text Search Engine for Massive System Log Data", Proceedings of the 12th International Conference on Future Internet Technologies, ACM, DOI: 10.1145/3095786.3095788, June 2017

[7] Keiichi Shima, Hiroshi Abe, Daisuke Miyamoto, Tomohiro Ishihara, Kazuya Okada, and Yuji Sekiya, "URL Classification using BoF of URL bitstream", Proceedings of the 12th International Conference on Future Internet Technologies, June 2017

[8] Ryo Nakamura, Yuji Sekiya, Daisuke Miyamoto, Kazuya Okada, and Tomohiro Ishihara, "Malicious Host Detection by Imaging SYN Packets and A Neural Network", Proceedings of 2018 International Symposium on Networks, Computers and Communications (ISNCC) (in press).

(3) 国際会議発表

(4) 国内会議発表

[8] 神谷 和憲 , 長谷川 彩子 , 関谷 勇司 , 岡田 和也 :「DNS キャッシュサーバログのグラフ分析による感染端末検知の検討」, 電子情報通信学会, 信学技報, vol. 117, no. 316, pp. 53-58, 2017 年 11 月

(5) その他（特許，プレス発表，著書等）