テンソル分解に基づくグラフ分類による組織内ネットワーク攻撃活動検知

西野 琢也[†1] 菊地 亮太[†1] 丸橋 弘治[†1] 福田 大輔[†1]
齊藤 聡美[†2] 鳥居 悟[†2] 伊豆 哲也[†2]

概要：組織内ネットワークへの侵害活動を目的とするマルウェアによる標的型攻撃は，近年その活動の巧妙化，複雑化が進んでおり，特定の操作の有無による判定だけでは攻撃者による操作の有無を特定するのが難しくなっている．そこで，本報告ではネットワーク内への通信試行からデータ搾取，などの攻撃者の一連の侵害行動をグラフ構造とみなし，攻撃が進み重篤な被害が生じた高リスクな時間帯と，攻撃が進行せず管理操作や軽微な被害のみ生じた低リスクな時間帯の分類を行う問題を解くことで，重篤な被害をもたらす攻撃の有無をその活動痕跡から検出することを可能とする，攻撃活動検知技術を開発した．

キーワード：動的活動観測，グラフデータ，マルウェア，テンソル分解，機械学習

Detections of attacker’s behavior using graph classification based on tensor decomposition

Takuya Nishino[†1] Ryota Kikuchi[†1] Koji Maruhashi[†1] Daisuke Fukuda[†1]
Satomi Saito[†2] Satoru Torii[†2] Tetsuya Izu[†2]

Abstract: The targeted attack using malwares for large-scale networks are proceeded and diversified recently. Therefore, it is difficult to specify the presence or absence of operation whether attacker’s behavior or not. In this report, a series of infringing behaviors of attackers such as communication trial or unauthorized data leak regarded as a graph data, and classified between the time zone of attacker’s activity and normal operation. We developed activity detection algorithm that enables to detect the presence or absence of an attacker from its behavior using integration logs of communication and process.

Keywords: Behavior Observable System(BOS), Graph analysis, Malware, Tensor decomposition, Machine Learning

1. はじめに

マルウェアを用いたサイバー攻撃は近年その活動の巧妙化，複雑化が進んでいる．特に侵入先の組織に合わせて作成した遠隔操作ツール（Remote Access Trojan, RAT）を使い，組織内ネットワークへの侵害活動を行う標的型攻撃が盛んに引き起こされるようになっている．現在の多くの標的型攻撃検知・対策技術では，C&C サーバの接続状況の把握や特定，情報搾取，などの特定挙動の存在を検知することや，各 RAT に特有の挙動を解析することに重点が置かれている．そのため，一度それらの検知をすり抜けて組織内ネットワークへの侵入が成功した場合，組織内外で発生する通常操作や管理操作と見分けることが難しい．そのため，攻撃者かどうかの判断が遅れてしまうことで，被害の拡大を防ぐ措置が遅れてしまうことも生じている．そこで，リスクの高い行動の履歴等から通常の振る舞いと違う活動を専門家が見出すことで，攻撃者の有無を検知することが行われている．しかし，巧妙化が進み，一見すると見分けのつきにくい履歴の中から判断するには専門家の高度な判断を要するため，人材確保，及び専門家の育成の難しさも課題となっている．

寺田ら[1][2][3]では組織内ネットワークでの特定の PC，サーバに対する攻撃者の行動観測を行うため，組織内ネットワークを模擬した環境を構築し，攻撃者の実際の活動を通して，攻撃時の振る舞いの特徴を収集・分析している．観測結果はデータセットである動的活動観測 BOS(Behavior Observable System)，以下 BOS dataset として提供されており，人材育成等の面で活用されている．BOS dataset は攻撃者の活動の進み具合に着目し，インターネットへの接続が可能な模擬環境におけるマルウェアの振る舞いの段階を進行度，と呼ばれる指標で区分し，標的型攻撃の段階に応じた分類を行っている．BOS dataset では，マルウェアが実行されていない段階を基点として，マルウェアによる通信の発生，C&C サーバとの通信確立，マルウェアの操作が継続的に実行されている段階までが 8 種類（進行度 1～進行度 8）に分類されている．

そこで，本研究では進行度を教師ラベルとみなし，攻撃の進行が進み，RAT が C&C サーバとの通信を開始した検体における攻撃者の振る舞い（進行度 6 以上）を高リスク攻撃，組織内ネットワークのスキャンに留まった振る舞い（進行度 5 以下）を低リスク攻撃，として攻撃者の振る舞

†1 株式会社富士通研究所　人工知能研究所
Artificial Intelligent Research Laboratory Fujitsu Laboratories Ltd.
†2 株式会社富士通研究所　セキュリティ研究所
Security Research Laboratory, Fujitsu Laboratories Ltd.

いを高リスク攻撃と低リスク攻撃に自動分類する，新たな機械学習手法を提案する．具体的には通信履歴から生成される通信挙動と Windows 端末上で実行されているプロセスの挙動の両方を加味した一連の活動を統合し, C&C サーバとの通信確立行動からプロセスの挙動に至るまでの一連の流れに攻撃者の行動の特徴が現れると仮定し，その履歴の特徴を総合的に判断する方式を提案する．

本提案手法では独立に生起された通信，プロセスの履歴を関連付けしたデータを一つのグラフデータとみなし，一定時間内で行われた履歴中に高進行度の契機となる遠隔操作を含む時間帯と，遠隔操作までは至らなかった低進行度の時間帯との教師あり分類問題を解くことで，高進行度に至った活動であるか否か，を自動で早期発見することを目的とする．

本論文ではその分類を行うためにテンソル分解と Deep Neural Network（以下 DNN）を複合したグラフ分類手法が有効であることを示し，既存のグラフ分類手法との比較を通して分類精度, 及びグラフの特徴を考察した結果を示す．

2. 関連研究

本研究の目的は，通信を含むネットワーク構成が大規模な組織をねらった標的型攻撃の早期発見であるが，標的型攻撃対策技術としては様々な検知手法が提案されている．

代表的なのは遠隔操作の攻撃送信元となる C&C サーバを直接特定することを目的とした研究である．特にサーバアクセスに関連する通信やプロセスの振る舞いを利用して攻撃者の存在を検知する技術が数多く知られている．

通信履歴の振る舞いを利用した技術として，文献[4]ではマルウェア検体とその通信履歴を用いて，その振る舞いから C&C サーバの存在を特定する技術が紹介されている．文献[5]では RAT 自体の DNS の特徴的な通信のパターンを観測して不正なドメインを検知する手法が知られている．文献[6]では，通信先 IP の異なり数やドメインの評価結果等の特徴をネットワーク履歴から抽出し，クラスタリングする事で正常通信が混じった場合でも悪意ある挙動を検知する技術を提案している．文献[7]では，ボットの感染モデルと IDS(Intrusion Detection System)ログを突き合わせることで，ボットの感染を検知する手法を提案している．

通信以外にも Windows 端末などの起動アプリケーションの特異性やそれらと通信の振る舞いを複合した行動に注目した攻撃検知技術も提案されている．文献[8]では，端末上のプロセスを定期的に取得し分析することで利用頻度の低いプロセスを特定し，不審なプロセスを検知する技術を提案している．文献[9]では，ネットワークイベントを対象として，発生したイベントのトリガーとなったイベント同士の関係性を分析することで，顕著に特徴が現れないマルウェアをも検知する手法を提案している．

これまでの研究では通信やプロセスの振る舞い方から検知したい C&C サーバの特定を行うことに焦点が置かれている．しかし，RAT の振る舞い等からの検知以外にも，攻撃者視点の通信試行や起動アプリケーションの特異性を元に，攻撃進行の程度を見積し，対処の優先順位をつけることも重要と思われる．実際に遠隔操作を行う攻撃者の様々な活動は，通信以外にも起動アプリケーションの選定，窃取する対象の検索，といった複数の活動を通して一つの振る舞いを形成すると考えられる．そのような活動の自動検知手法の構築には，活動履歴の繋がり方の特異性を判断する技術が必要である．

そこで筆者らはそれらの複合した活動進行度の推定を行うために，関係性に基づくグラフデータを用いた機械学習による分類手法に注目した．グラフデータは離散値で表されるノードとその接続関係をエッジで表して繋ぎ合わせ，複雑な関係を記述する数学表現の一つである．文献[10]では通信履歴が通信元と通信先の IP アドレスの関係に特異性が現れる事を利用し, 通信元と通信先の IP アドレスの履歴の関係をグラフデータとみなし, 通常と異なる IP アドレス間のエッジ数の増大を元に検知する事を提案している．文献[11]では大規模ネットワークにおけるマルウェアの感染拡大をグラフデータとみなし，IP アドレスだけでなくアクセスした URL やダウンロードしたファイル名等を複数ノードで表現し，マルウェア活動の感染拡大を検知する技術が提案されている．

2.1 グラフ分類手法の適用における課題

一般に機械学習では，入力データから分類に寄与する抽象化された特徴量を専門家が設計し，その特徴量に基づいた様々な分類手法，例えばサポートベクターマシン（以下 SVM）やランダムフォレスト（以下 RF）といった分類手法を用いて，データの学習と評価を行う．グラフデータの分類においても専門家が設計した部分グラフを使う方式[12]や，Random walk[13]や Shortest path[14]等の経路長を特徴量とする方式や，隣接するラベルのリラベリングとハッシュ配列化を行い，カーネルにより特徴量化する Weisfeiler-Lehman Graph Kernels[15]等で, グラフデータの特徴量を設計し，それを SVM や RF，もしくは DNN 等で分類する手法が提案されている．

しかし，本研究の対象である，通信とプロセスの複合関係を記述するグラフデータにおいては，通信とプロセスが入り混じった活動の中から特異な行動を特定，抽出することになる．そのため，複数の異なる要素を持つ多次元，かつ多様な活動形態の全ての要素を記述する多項関係を有したグラフデータになることが予想される．そのため，可能な組み合わせの数が膨大になってしまい，データから事前に特徴量の設計を行う事が困難になり，上述した事前の特徴量設計では高精度な分類を実現するのは限界がある．筆者らはグラフデータから分類に寄与する部分構造を自動抽出し，高精度な分類が可能な技術を開発した[16].

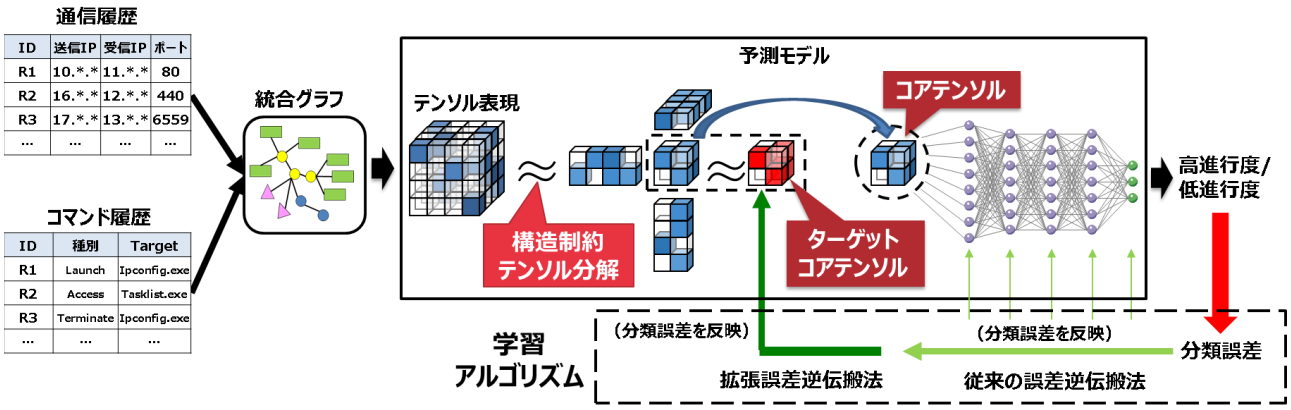

図 1 提案手法の概要

Figure 1 Overview of proposed method

そこで，本論文では履歴データの結合によって複雑な多項関係を有する多次元のグラフを接続テンソルとみなし，それらの中からテンソル分解と DNN を複合したアルゴリズムを用いることで，進行度を自動推定する方式を提案する．

3. 提案手法

図 1 に提案手法の概要を示す．提案手法は以下の 3 つの要素で構成される．まず通信とコマンドを統合した一つの配列データを作成し，グラフデータに変換する．次に，2 つ目として，それらのグラフをノード，ラベルの並べ方の組み合わせ全ての接続関係を表記する接続テンソルであるとみなし，テンソル分解によりコアテンソルと要素行列に分解し，作成したコアテンソルをニューラルネットワークに入力する．コアテンソル算出時に，分類に寄与する重要な特徴量が表現されたターゲットコアテンソルへの類似度が最も高いコアテンソルを選択する．更に 3 つ目としてターゲットコアテンソルが分類精度を高めるようにニューラルネットワークの学習で用いられる誤差逆伝搬法をターゲットコアテンソルへ拡張し，更新する．

3.1 履歴の結合とグラフデータ化

本手法では通信を記述するパケットキャプチャである Pcap，及び Windows 上の Firewall，Eventlog 等の通信履歴とコマンドを記述する Process，Eventlog を複合してグラフデータを構成する．特に RAT による遠隔操作による被害を想定し，通信を基点とした遠隔操作までの一連の流れをグラフデータとして表現する．そのため，送信・受信 IP アドレス及びポート番号の繋がりに加え，紐付けられた起動アプリケーションの履歴で表現する．それらが一定時間内に行われた履歴全てを一つの判別クラスとみなすことで，進行度を推定する．各履歴から生成されている時間を元に時系列順に全履歴を並び替え，一つの多次元時系列データの配列を構成し，配列を時間帯で区切ることで一つのエッジリストとして扱うことでグラフデータ化を行う．この配列中に存在するユニークノードの離散値全ての組み合わせが多次元配列であるテンソルとみなし，そこから各グラフデータとしての塊毎に分解されて DNN へと入力され，学習が行われる．

3.2 構造制約テンソル分解による特徴抽出

次に，入力データから，DNN への入力とする際の流れを示す．テンソル分解を用いたグラフデータの学習では入力データの接続関係を記述するテンソルを主要な部分テンソルへと分解する．その際，Tucker 分解[17]と呼ばれるテンソル分解を用いることが考えられるが，通常の Tucker 分解ではグラフの分類に寄与する特徴とは無関係にコアテンソルや要素行列の最適化が行われてしまう．そこで，筆者らが提案する構造制約テンソル分解は，分類に寄与する特徴が表現されたターゲットコアテンソルという概念を導入し，このターゲットコアテンソルにコアテンソルが最も類似するようにコアテンソルを算出する．

それらを実現するために 2 段階の最適化計算を行う．図 2 に示すように，まずステップ 1 として，分解時にターゲットコアテンソルをコアテンソルとみなし，不変とした時に要素行列のみを変更し，入力テンソルを最も良く近似する要素行列を求める．その後，ステップ 2 としてステップ 1 で求めた要素行列を不変として，同様に入力テンソルを最もよく近似するコアテンソルを求めることで，コアテンソルの最適化を行う．

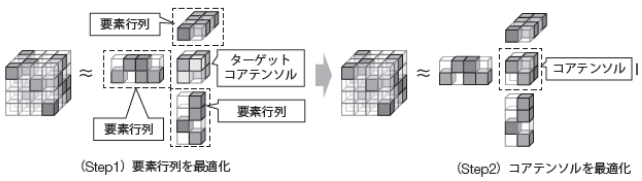

図 2 構造制約テンソル分解の概念 [16]

Figure 2 Restricted tensor decomposition [16].

3.3 誤差逆伝搬によるターゲットコアテンソル最適化

3.2 にてターゲットコアテンソルを元にコアテンソルの最適化を行う手法を示した．ターゲットコアテンソルは分類に重要な情報をそのテンソル中に含むコアテンソルである事を仮定している．その一方で，ターゲットコアテンソルが何かしら最初に定められるのであれば問題ないが，通常そのような情報は存在していない．そこで，入力テンソルからの要素行列とコアテンソルの最適化と同時に，DNN からの誤差逆伝搬を活用して，同時にターゲットコアテンソルの最適化を行う．誤差逆伝搬法は DNN において，分類誤差を下層に伝搬させる形で，分類誤差を小さくするように勾配を求め，パラメータの修正方向を算出するアルゴリズムであり，現在の多くの DNN をベースとする手法で搭載されている．入力として使われたコアテンソルによる生成された第一層が，逆伝搬時にはターゲットコアテンソルに戻るように拡張することで，ターゲットコアテンソルと DNN のパラメータを更新する．得られたターゲットコアテンソルが分類に寄与するように最適化されることにより，コアテンソルも類似度を最大にするように計算されることで，ターゲットコアテンソルが分類に重要な構造を保持すると同時に，コアテンソルも重要な構造に類似した形へ変換される．

4. 評価実験

4.1 動的活動観測データセットと活用データ詳細

今回の実験では BOS dataset 2015 及び 2016 に含まれる各マルウェア検体の中で BKDR 系の検体の実行記録を用いた．これらは組織内ネットワークを模擬したネットワーク環境下で実際に検体を特定のクライアント PC 上で実行し，C&C サーバからの通信や遠隔操作が実際に観測できた時間帯の履歴を抽出したデータセットである．各履歴はネットワーク内の通信キャプチャを記録した Pcap，検体を実行しているクライアント PC で発生したプロセスを記述した process.log，クライアント PC 内で発生したイベントを記述した Eventlog.evtx，クライアント PC と外部の間で発生する通信記録である Firewall.log，ネットワークプロキシの実行記録である proxy.log 等が提供されている．これらの履歴は各々が独立に生起されており，履歴中から遠隔操作の証拠となる事象を抽出，組み合わせることで遠隔操作の有無や攻撃進行の程度を評価し，進行度，と呼ばれるラベルで各検体の特徴を評価している．

4.2 問題設定

本論文では BOS dataset を用いて，論文中に記述されている進行度を教師ラベルとして用い，進行度 6 以上か，それ以下なのか，を分類する問題を設定した．進行度は 1～8 までの 8 段階設定されており，進行度 7 で遠隔操作に至り，実際のファイル窃取等に至ったケースに該当する．そのため，この問題設定は遠隔操作に至り，重篤な被害に至った場合の攻撃者の振る舞いと，それに至らず，重篤な被害には至らなかったケースを分類する問題設定と同義である．

表 1 に BOS dataset 中から実験に使用した各検体の種類，進行度を示す．検体 d18～e04 までは進行度 7 で遠隔操作に至った検体の実行記録であり，それ以外は進行度 5 以下で遠隔操作には至らなかった検体の実行記録である．BOS データではこれ以外にも TROJ 系の検体動作も存在するが，検体種類が大きく異なるため，本論文では対象外として，BKDR 系の検体のみに限定して実験を行った．

表 1　BOS dataset より抽出した入力データの詳細[2][3]
Table 1　Details of dataset from BOS dataset

識別名	マルウェア検体名	進行度	遠隔操作時間 [minutes]
d18	BKDR_EMDIVI.I	7	201
d19	BKDR_EMDIVI.F	7	28
d33	BKDR_PLUGX.DUKLR	7	355
d37	BKDR_EMDIVI.AB	7	181
e04	BKDR_EMDIVI.MSB	7	216
e12	BKDR_EMDIVI.L	5	--
e70	BKDR_PLUGX.DUKOA	4	--
e435	BKDR_PLUGX.DUKOQ	4	--

更に，この問題設定中で 2 種類の問題を設定した．図 3 にその詳細を示す．一つ目が遠隔操作に至った各検体の時間帯を 2 分割して学習用と評価用にすることで，自検体の振る舞いの特徴からその検体の振る舞いの特徴を抽出可能か検証する問題設定である（問題設定①）．二つ目は遠隔操作に至った異なる検体のふるまいの学習結果から，他の亜種検体の振る舞いを正しく予測可能か検証する問題設定である（問題設定②）．

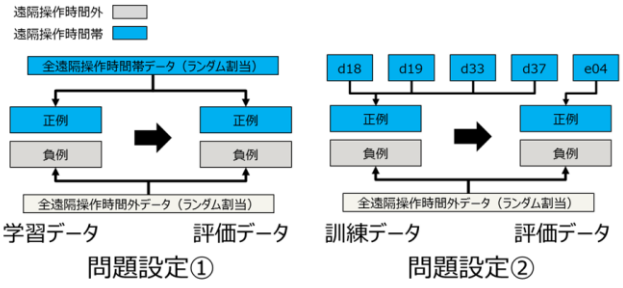

図 3　実験の問題設定
Figure 3　Problem settings of experiments.

本研究では上述した問題設定の中で，各履歴の中から可能な限り統一した情報を抽出するため，ほぼ全ての検体実行記録に含まれている Pcap, Process, Eventlog, 及び Firewall の 4 つを用いた．

4.3 サンプリングしたデータセット

本研究では独立に生起している履歴を結合し，一つのグ

ラフデータによる複合的な事象の結果として，進行度を分類する．しかし，検体毎に攻撃元が異なるため，これらの履歴の中から攻撃者の振る舞いの特異性を学習するには，特定の IP アドレスや通信の有無に依存しないための抽象化が必要である．そのため，本研究では，データセットの中から他の履歴と情報の粒度，抽象度を揃えるために，表 2 の 9 つの異なる属性を用いてデータの抽象化を行った．

表 2 のように，まず，通信に関連する属性として，通信元と通信先の IP アドレス，ポート No，及び各組合せに紐付いた通信種別やパケット長，通信フラグを履歴から抽出した．次に，プロセスに関連する属性として，ターゲットとなる exe ファイルを起動するために使用したアプリケーションの種類やコマンドの属性（Launch や Terminate 等），及びターゲットのアプリケーション(exe ファイル)を履歴中から抽出した．

これらは検体毎に使用されている IP アドレスが異なっていることや, ポートスキャン等のように固有のポート No を使った事自体よりもポート No が短期間で大量に変更された通信が存在すること，自体に特徴があると考えた．そこで，グラフ単位で新たにユニークなラベルを割り当てなおし，接続関係の広がりに注目した抽象化を行った．その一方で通信フラグ，プロセスコマンドやコマンド属性，パケット長等は検体毎に固有の情報を用いるのではなく，通信の種別や検体動作端末中に存在する操作コマンドの情報であるため，その特異性を加味するよう，ラベルの振り直しは行っていない．これは，攻撃者がどのような素性や思惑であっても特定の動作が行われる可能性があるからである．

表 2　入力データセットより抽出した属性情報の詳細
Table 2　Details of dataset from BOS dataset 2015 and 2016

抽出した離散値	内容
Source IP address	通信元 IP アドレス
Destination IP address	通信先 IP アドレス
Source Port Number	通信元ポート No
Destination Port Number	通信先ポート No
Command Attribute	処理を実行したプロセスや属性
Target Command	起動先コマンド，読取ファイル名等
Communication Flags	ビット文字列化した通信フラグ
Communication Type	通信種別（TCP/IP など）
Packet Numbers	通信パケット長

グラフデータは繋がりの起点を表すノード，及びそのノードに紐付いた識別記号であるラベル，そして各ノード間を接続する情報を含むエッジの 3 つで構成されている．表 2 で抽出された属性をその履歴の記載された時系列順に並べ，図 4 にそのグラフのノードの関係を表した可視化の一例を示す．色毎に異なる属性の通信とコマンドを示し，それらの繋がりをグラフとして表現している．この例では，赤色が通信元 IP アドレス，青色が通信先 IP アドレス，緑とオレンジがそれぞれポート No を表し，紫色がパケット数を表し，黄色と桃色がコマンドに関係する．この活動においては数種類の少ない IP アドレスやコマンド種別に対し，通信先及び通信元ポート No やパケット数の組み合わせにユニークな活動が多く紐付いていることから，特定の IP 間での通信頻度やポートの走査活動の多さで，特徴づけられたグラフである．

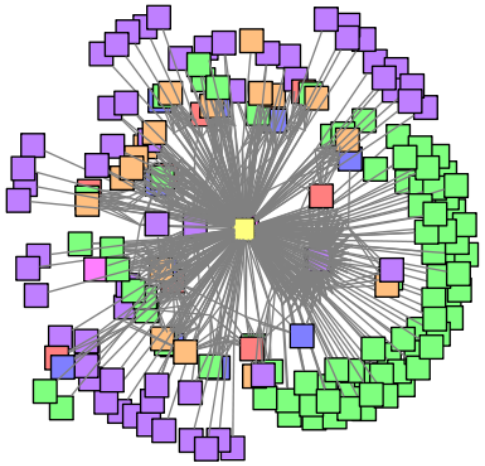

図 4　グラフデータによる攻撃者の活動表現の一例
Figure 4　Graph representation from an attacker's behavior

4.4 履歴の結合方式とグラフ化

本報では上述で各抽象化された属性を履歴から抽出し，時刻情報を元に各履歴を結合して一つのグラフを構成した．しかし，履歴毎に内包している情報の質と量が異なっているため，適切な結合方式が必要である．例えば Pcap 等の通信キャプチャにはコマンド情報は記載されておらず，逆に Process には通信の情報は含まれていない．そのため，各属性を結合するためには欠損が生じる．しかし，本来は遠隔操作によってつながりを持っているはずである．そこで，ある時刻におけるコマンドの直近時刻には関連する通信が行われているはず，という仮定の元に結合処理を行っている．表 3 にそれらの結合後の検体毎の履歴数（エッジ数）と，教師クラス数を示す．各時系列の情報から，10 分毎に時間帯で区切り，その時間帯で行われた活動全てを一つの教師クラスとして扱い，グラフデータとしている．検体毎に混在することを避けるため，10 分間で区切るまでは検体毎に処理を行い，最後に学習する際にのみ検体を混在させたデータを入力としている．

遠隔操作時間帯として定義されている時間はそれ以外の時間帯に比べると短く，教師クラス数は遠隔操作時間帯を正例，それ以外の時間帯を負例とすると，正例 : 負例比は 108 : 45021 となり，著しく偏りが生じている．検体毎に遠隔操作時間は異なり，かつ各履歴からの情報量もさまざまであるが, 遠隔操作時間のエッジ数平均は 1307 なのに対

し，遠隔操作時間外でのエッジ数平均は 243 と少ない傾向にある．

表 3　履歴結合データのエッジ数と教師クラス数

Table 3　The number of graph edges and classes

識別名	遠隔操作時間行数	遠隔操作時間外行数	教師数正例	教師数負例
d18	56915	1149544	25	8422
d19	1952	153317	6	10224
d33	33970	163539	36	5835
d37	46757	2583904	19	9791
e04	261	32033	22	6118
e12		4770699		2580
e435		1995506		1596
e70		381061		455
総計	139855	11229603	108	45021

4.5 実験の方法

問題設定①では負例データから検体毎の偏りが出ないよう，各 15 クラスをランダム選択し，計 120 クラスを学習用データとして抽出，各 10 クラスを評価用データとしてランダム抽出した．正例は評価用データを各 6 クラスランダム抽出して 30 クラスを評価用とし，残りの 78 クラスを学習用データとした．問題設定②では検体毎に学習データは対象となる検体以外の正例全てを用いた．問題設定②でも負例データは問題設定①と同様の設定を用いた．問題設定①，②共通で評価時は学習時とは異なる時間帯の負例，正例を割当し，各検体の実行時間帯で重複して混じることの無いように抽出した．負例は正例に比較してクラス数が大きいため，学習時の悪影響を避けるため，アンダーサンプリングで極端な不均衡データとならないようにした．その際負例の抽出データの偏りによる学習と評価への影響を取り込むため，問題設定毎に学習データの負例のランダム抽出を 5 回行い，平均する事で抽出による誤差の評価も行っている．

問題設定①，②共通で学習時に最低一クラスは割当可能な 5 分割交差検証を行って，Neural Network 及びテンソル分解のパラメータの最適化を行った．交差検証後，評価用データの評価のための学習モデル構築の際には交差検証で求められた際の最良のパラメータを用いて学習データ全てを用いてモデルを生成した．

4.6 評価指標

精度評価の指標として，分類問題における各種指標を複合的に用いた．負例，正例問わず正解教師データに対する予測結果の正解率，及び正例側から見た TP（True Positive）と TN（False Negative），負例に対する正解数である FP（False Positve），負例に対する誤り数である FN（False Negative）を元に算出した適合率，検出率，Geometric mean [18]（以下 GM）を用いた．GM は今回のような正例と負例のバランスの悪い不均衡データの評価に用いられる指標の一つであり，式(1)のように評価データ数に偏りのある場合にその正解率が良いバランスで最大化するよう設計された指標である．本報では遠隔操作時間帯の正解率が高いほど精度が相対的に良くなる傾向にある．

$$GM = \sqrt{\frac{TP}{TP+FN} \cdot \frac{TN}{FP+TN}} \tag{1}$$

実験結果はこれらの評価結果を元に複数の観点から行って傾向を評価した．精度評価結果は，抽出時と初期値の設定によってバラつきが生じてしまうため，5 回平均値と標準偏差を記載した．

4.7 実験結果

表 4 に実験結果を示す．問題設定毎に正解率，適合率，網羅率，及び GM の値を示し，更に既存手法として Weisfeiler-Lehman Graph Kernel による特徴量抽出を元にした SVM による分類結果（WL）を示す．WL は文献[14]に

表 4 実験結果

Table 4　Experimental Results

問題設定	評価データ比率 負例数:正例数	提案手法 正解率	提案手法 適合率	提案手法 検出率	提案手法 GM	WL 正解率	WL 適合率	WL 検出率	WL GM
①Random	80 : 30	0.93±0.01	0.89±0.01	0.83±0.01	0.90±0.01	0.76±0.02	0.55±0.04	0.79±0.07	0.77±0.02
②d18	80 : 25	0.89±0.01	0.71±0.01	0.88±0.01	0.88±0.01	0.76±0.04	0.49±0.09	0.44±0.22	0.60±0.14
②d19	80 : 06	0.85±0.01	0.30±0.01	0.83±0.01	0.84±0.01	0.76±0.02	0.17±0.04	0.63±0.25	0.69±0.12
②d33	80 : 36	0.92±0.01	0.84±0.01	0.92±0.01	0.92±0.01	0.61±0.01	0.24±0.03	0.12±0.02	0.31±0.03
②d37	80 : 19	0.84±0.01	0.56±0.01	0.67±0.02	0.77±0.01	0.81±0.03	0.50±0.04	1.00	0.87±0.02
②e04	80 : 22	0.76±0.00	0.00	0.00	0.00	0.61±0.00	0.03±0.04	0.04±0.04	0.09±0.13

て公開されている Matlab 版プログラムと LIBSVM[19]を用いて提案手法と同一のデータを用いて行った．本研究では他にも，Graphlet Kernel，Shortest Path，Random Walk に基づくカーネル法による特徴量抽出を行い，同様に SVM による評価を行ったが，その中で最も正解率の高かった WL のみ記載している．

各検体から 2 分割する問題設定①では既存手法に比較して検出率を除いて 0.1〜0.2 程高精度である．また，標準偏差も既存手法よりも小さく，ランダム性による評価データのバラつきも小さい．既存手法の場合は問題設定①，②共に検知率が高い傾向にあるが，適合率が低く，負例を正例と誤認している割合が高い．提案手法では逆に，検知率は既存手法と同程度かわずかに高い程度であるが，適合率が高く，負例を誤認する割合が小さい．特定検体を他検体の学習から予測する問題設定②では検体毎に顕著な差がある．特に d18，d19，d33 ではいずれも提案手法で GM が 0.84 以上の高い検知精度を達成しており，他検体から共通する特徴を学習できている．その反面，d37 では提案手法の適合率，検知率が共に 0.53，0.62 と低く，ランダムに選択する場合の 0.5 を僅かに上回る程度であり，学習が困難であることがわかる．更に e04 を評価する場合は適合率が 0，つまり全て負例と予測しており，他検体から e04 の特徴の学習をすることができていない．既存手法では問題設定②では d37 を除き低精度であり，特徴を学習することが困難であることがわかる．更に，d37 の検出率 1.0 と高精度であるが適合率が低く，ランダム選択した場合と同じであり，e04 についても提案手法と同じく低精度である．

5. 考察

この章では提案手法の優位性について既存手法との対比を通して考察する．実験結果から提案手法は既存手法に比較して適合率が高く，検知率が同程度だがバラつきも小さいことが分かった．本論文では問題設定①において，既存手法と提案手法の間で対比が可能な時間帯において，グラフの頻度分布の対比によって考察を行う．問題設定①では全ての検体の低進行度のデータと高進行度のデータが一定数以上教師クラスとして含まれるよう学習を行う．そのため，必ず自検体の負例も正例も含んだ状態で学習と評価を行っている．提案手法では学習データ中に含まれる何らかの特徴をターゲットコアテンソルの形で学習データから入力データのテンソルを最適化する．それに対し，既存手法では特定のノードを基準として近傍のノードとの関係を更新して学習するものであり，どうしても近傍のノードの関係に引きずられてしまうと思われる．これらの仮説の検証の一つとして，本論文ではノードユニーク数を可視化し，提案手法が得意なグラフデータと，既存手法が不得意なグラフデータとを比較し，その結果から考察することを試みた．

5.1 既存手法との対比による提案手法の優位性考察

図 5 に対比の一例を示す．この図は一教師クラス単位での表記であり，横軸に表 2 で示したノードの属性，縦軸にユニークなノードのカウント数を示し，各属性同士のユニークな結合を図中に表したものである．図 5(a)は e04 のある時間帯に取得された履歴から作成したグラフであり，正例と設定したもの，図 5(b)は e70 のある時間帯に取得された履歴からの負例と設定したものである．この二つはどちらも通信元のソースポート No と通信パケット数の種類が多く，ついで通信先ポート No の数が多く，一見すると似通ったグラフである．この例では既存手法はどちらも正例と誤判定しており，全く同じものと判定しているが，提案手法では負例と正例で正しく予測できている．既存手法では特にこのようなユニークカウント数が多く，学習データと似通った可視化結果となるものは同様に誤判定している．既存手法は特定のノードを基準としてその隣接するノードとの関係をカーネルとして重ね合わせるものであり，近傍のノードの影響が大きいが，図 5 のようにグラフデータとしての各属性に一定のユニーク数が存在しており，多岐にわたる属性の操作が入っている場合では，近傍同士の関係がそのグラフの特徴をカバーできなくなるため，特徴量が似通ったものになり，分類が難しくなると考えられる．それに対し，提案手法では分類に寄与する観点のみから高次元のテンソル中の重要な特徴を自動選択するため，そのような近傍に対する依存性は小さく，図 5 のような複数の属性に跨る活動を記述するグラフデータであっても正しく学習できている．

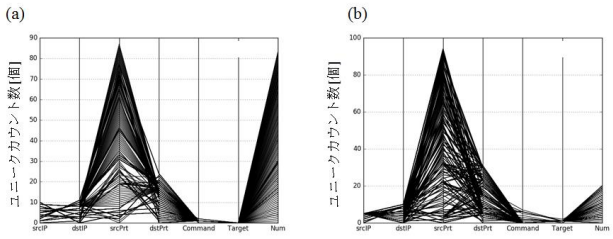

図 5　ノードユニーク数と関連付けた可視化
(a): 正例における対比，(b): 負例における対比
Figure 5　Distribution graph for node unique counts

最後に提案手法，既存手法共に不得手なグラフデータについて述べる．正解率は最大でも 0.9 程度であり，一定数の誤判定を内包している．図 6(a)は d18 における特定の時間帯に取得された履歴から作成したグラフであり，正例と設定したもの，図 5(b)は d37 のある時間帯に取得された履歴からの負例と設定したものである．図 6(a),(b)いずれの場合でも提案手法，既存手法共に誤判定している．これらのグラフデータの特徴としては，全体として図 5 に比べユニークカウント数が小さく，単一の突出したカウント数が存

在しているものの，全体に対しての比率の差は小さく，特定の起動アプリケーションや IP アドレスの関係によって特徴付けられたものであると推測される．このような場合では人手による目視で確認可能な範囲であることもあり，グラフデータによる分類と人手による分類とで役割分担する措置が必要であると考えられる．

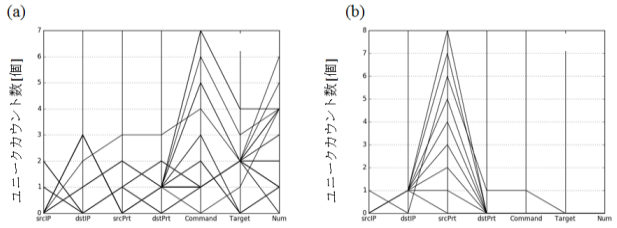

図 6　ノードユニーク数分布
(a): 正例における誤判定，(b): 負例における誤判定

Figure 6　Distribution graph for node unique counts

6. まとめ

RAT を利用した攻撃者の通信と起動アプリケーション，両方の振る舞いの特異性に着目し，履歴データの結合による通信挙動とプロセス挙動両方を加味した一連のグラフを用いて，重篤な遠隔操作に至る過程を自動推定する新たな攻撃検知手法を提案した．提案手法は既存のグラフ分類手法に比べ高い検知精度を示した．特に複数の属性に跨った活発な活動痕跡を示すデータの場合，効果の高いことが分かった．これらの結果は今後の複雑化，巧妙化していく挙動を分析するための強力なツールとなることが期待される．その一方で活動量が小さく，特定の内容に依存する判断が必要な場合ではこの方式だけでは限界があることも分かった．今後の課題は，それらの取得データの拡充による効果の確認と見極め，そして可視化による専門家へのフィードバック機能の追加等の説明力の向上である．

謝辞　BOS dataset 2015, 2016 のデータ解釈に当たり，一部追加情報提供を株式会社日立製作所の寺田真敏様より行っていただきました．ここに感謝致します．

参考文献

[1] 寺田 真敏, 青木 翔, 楠美　淳弥, 重本　倫宏, 萩原 健太. 研究用データセット「動的活動観測 2014」の検討. 情報処理学会, マルウェア対策研究人材育成ワークショップ 2014 (MWS2014), 2014, p. 1121-1125.

[2] 寺田 真敏, 堀 健太郎, 成島 佳孝, 吉野龍平, 萩原 健太, “研究用データセット「動的活動観測 2015」の検討“, 情報処理学会, マルウェア対策研究人材育成ワークショップ 2015 (MWS2015), 2015, p. 1387-1393.

[3] 寺田 真敏, 佐藤 隆行, 堀 健太郎, 吉野龍平, 萩原 健太, “研究用データセット「動的活動観測 2016」の検討“, 情報処理学会, マルウェア対策研究人材育成ワークショップ 2016 (MWS2016), 2016, p. 892-895.

[4] J. Ma, L. K. Saul, S. Savage and G. M. Voelker.. Beyond Blacklists. Learning to Detect Malicious Web Sites from Suspicious URLs. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009, p. 1245–1254.

[5] M. Felegyhazi, C. Kreibich, and V. Paxson.. On the Potential of Proactive Domain Blacklisting. USENIX Conference on Large-scale Exploits and Emergent Threats, 2010, p. 6-6.

[6] Ting-Fang Yen, Alina Oprea, Kaan Onarlioglu, Todd Leetham, William Robertson, Ari Juels, and Engin Kirda.. Beehive: Large-scale log analysis for detecting suspicious activity in enterprise networks. Proceedings of the 29th Annual Computer Security Applications Conference, 2013, p. 199-208.

[7] Guofei Gu, Phillip Porras, Vinod Yegneswaran, Martin Fong, Wenke Lee, "BotHunter: Detecting Malware Infection Through IDS-Driven Dialog Correlation." USENIX Security Symposium, 2007, vol. 7, p. 167-182.

[8] 中里純二, 津田侑, 高木彌一郎, 衛藤将史, 井上大介, 中尾康二. ホスト型 IDS を用いた標的型攻撃対策. コンピュータセキュリティシンポジウム 2014 (CSS 2014), 2014, p. 466-473.

[9] Zhang Hao, Danfeng (Daphne) Yao, and Naren Ramakrishnan.. Detection of stealthy malware activities with traffic causality and scalable triggering relation discovery. Proceedings of the 9th ACM symposium on Information, computer and communications security, 2014, p. 39-50.

[10] Collins, M.P., Reiter, M.K.. Hit-list worm detection and bot identification in large networks using protocol graphs. RAID 2007, 2007, vol 4637, p. 276-295.

[11] Luca Invernizzi, Stanislav Miskovic, Ruben Torres, Sabyaschi Saha, Sung-Ju Lee, Christopher Kruegel, and Giovanni Vigna.. Nazca: Detecting malware distribution in large-scale networks. Proceedings of the ISOC Netwrk and Distributed Systems Security Symposium 2014, 2014, p. 1-16.

[12] Gartner, T., Flach, P., and Wrobel, S.. On graph kernels: Hardness results and efficient alternatives. Proceedings of 16th Annual Conference on Learning Theory and 7th Kernel Workshop, 2003, p. 129-143.

[13] Borgwardt, K. M., & Kriegel, H.-P.. Shortest path kernels on graphs. Proceedings of International Conference on Data Mining, 2005, p. 74-81.

[14] Nino Shervashidze, S.V.N. Vishwanathan, Tobias H. Petri, Kurt Mehlhorn, Karsten M. Borgwardt.. Efficient graphlet Kernels for large graph comparison. Proceedings of Machine Learning Research, 2009, 5, p. 488-495.

[15] Nino Shervashidze, Pascal Schweitzer, Erik Jan van Leeuwen, Kurt Mehlhorn, Karsten M. Borgwardt.. Weisfeiler-Lehman Graph Kernels. Journal of Machine Learning Research, 2011, 12, p. 2539-2561.

[16] 丸橋 弘治.. 人やモノのつながりを表すグラフデータから新たな知見を導く新技術Deep Tensor. 雑誌富士通, 2017, 8, 5, p. 1-7.

[17] Tamara G. Kolda, Brett W. Bader.. Tensor Decompositions and Applications. SIAM Review, 2009, 51(3), p. 455-500.

[18] R. Barandela, J.S. Sánchez, V. García, E. Rangel.. Strategies for learning in class imbalance problems, Pattern Recognition, 2003, 36 (3) p. 849–851.

[19] Chang, Chih-Chung, and Chih-Jen Lin.. LIBSVM: a library for support vector machines. ACM transactions on intelligent systems and technology (TIST), 2011, no. 2.vol. 3, p. 27:1-27.