2E-05

組織内ネットワーク攻撃進行度の自動推定技術の評価検証

矢野　翔太郎† 西野　琢也† 菊地　亮太† 丸橋　弘治† 福田　大輔†
齊藤　聡美‡ 鳥居　悟‡ 伊豆　哲也‡
株式会社富士通研究所　人工知能研究所†　セキュリティ研究所‡

1. 背景

近年，特定組織への侵害活動を目的とする標的型攻撃は活動の巧妙化，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判定するのが難しくなっている．そこで文献[1][2]では，組織内ネットワークを模擬した環境を構築し，攻撃者を誘引することで，実際の攻撃者の振る舞いの特徴を収集・分析している．その成果は攻撃進行度，と呼ばれる 8 段階のラベル付けを行った上で，マルウェア種類毎に異なる研究用データセット（以下 BOS dataset）として提供されている．前報[3]ではそのログ情報を用いて高リスク攻撃と，低リスク攻撃を自動判別する攻撃活動検知技術を開発した．この報告では，通信ログやプロセスログ等，複数の異なるログを統合し，攻撃者の一連の侵害行動をグラフ構造データとみなすことで，攻撃が進み C&C サーバとの通信を開始した高リスクな時間帯と，攻撃が進行せず組織内ネットワークのスキャンに留まった低リスクな時間帯の自動分類を行う新たな手法を提案した．

提案手法はテンソル分解と Deep Neural Network (DNN)を用いて，既存のグラフ分類手法に比較して高い分類精度が達成可能である事を報告した．しかし，前報ではモデルの解釈性に乏しい DNN をベースとした手法を用いており，実用上重要な評価結果の判断根拠の提示やグラフ構造データ以外のより一般的な手法を用いた場合の評価が不十分であるという課題があった．

そこで本稿では，既存の C&C サーバの検知技術の手法[4]による評価と，前報提案手法の結果を，局所的に解釈可能な線形モデルによる変換[5]を行い，高重要度の活動を抽出する新たな手法を提案し，その結果を考察する．

2. 適用手法

前報[3]は BOS dataset のログを 10 分間毎に区切り，ログ中の表 1 に記載したような IP アドレスやポート No 等の属性情報を各特徴量の 2 部グラフとみなし，可能な組み合わせ全てのパターンを網羅した入力データをテンソルへと変換して分類を行った．これにより攻撃者の 10 分間の活動履歴全体でのリスクを評価している．

表 1 BOS dataset より抽出した属性情報の詳細

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	実行したプロセスや属性
Target Command	コマンド，読取ファイル名等
Communication Flags	ビット列化した通信フラグ
Communication Types	通信種別（TCP/IP など）
Packet Numbers	通信パケット長

しかし，作成した DNN モデルは数層～数十層のニューロンで構成される非線形性の複雑なネットワークであり，その予測結果と入力データの関係性を明示するのは困難である．そこで，人間が解釈可能となるように，丸橋ら[5]は評価データの DNN への入力コアテンソルの要素ベクトルと，周辺のデータ集合のコアテンソルの要素ベクトルとの類似度が高く，DNN モデルと線形分類器の予測結果の差がより小さくなるように線形分類器の再学習を行うことで，DNN モデルを最も良く近似できる線形分類器を再計算するアルゴリズムを提案している．

本報ではこの再学習モデルを構築するため，学習時のデータとモデルを基準とし，精度検証に用いた評価データにおける線形分類器を再計算した．そして，得られた線形分類器の係数(寄与度)をコアテンソルの要素と対応付けし，テンソル分解の要素行列を用いて，元のテンソル(エッジごとの情報)に逆変換することで，分類に寄与したエッジを特定している．さらに，その特定したエッジが，一連の攻撃活動中のどの部分に該当するかを確認するためにログと対応付けした．併せて精度評価のため，最高精度のハイパーパラメータ条件での学習モデル構築を行い，前報と同様の評価データを用いて評価した．

Evaluation and verification of automatic estimation techniques of organization within the network attack progress
†Artificial Intelligent Research Laboratory Fujitsu Laboratories Ltd
‡Security Research Laboratory Fujitsu Laboratories Ltd

3. 結果と考察

最初に精度評価結果を述べる．本報では前報の提案手法をより精緻にパラメータチューニングを行った方式に加え，文献[4]中から使用可能な 2 種類の特徴量を算出した．1 つは 10 分間に転送されたバイト数の平均値，標準偏差，自己相関の平均値，自己相関の標準偏差，及び 2 つ目は通信元 IP アドレスと通信先 IP アドレスの組み合わせ数の平均値，標準偏差である．各特徴量の単独の場合と，全てを使用する場合の 3 つのパターンをランダムフォレストによって評価した．

正確度の評価結果において，提案手法 0.97，残り 3 つの手法においては 0.89，0.81，0.91 と最も提案手法が高精度の結果となった．

この結果は，IP アドレスの組み合わせに基づく情報は，攻撃活動を識別するうえで支配的な特徴量である事を示唆しており，大半の評価データは少数の支配的な特徴量で分類が可能な事を示している．

表 2 に d33 における高進行度時間帯の寄与度評価結果を示す．d33 の事例は PlugX と呼ばれるマルウェアの感染事例である．感染後，特定の実行ファイルが実行され，攻撃者からの指示をやり取りする C&C 通信をはじめとする攻撃活動が行われている．我々の再学習モデルから得られた結果では，従来のウィルス対策で着目されている実行ファイルの実行ログではなく，C&C 通信に関するログが識別の寄与度が高いとされている．すなわち，攻撃活動時に定常的に行われている C&C 通信に着目することが，攻撃活動の進行度を見極めるうえで有効であると示唆するものである．この観点は，仮に実行ファイルの実行ログを見逃しても攻撃活動の捕捉が可能であり，実際の運用監視業務においても有効であると考えられる．

4. 参考文献

[1] 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), pp.1387-1393(2015)

[2] 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), pp.892-895(2016)

[3] 西野 琢也，菊地 亮太，丸橋弘治，福田 大輔, 齊藤 聡美，鳥居 悟，伊豆 哲也.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. コンピュータセキュリティシンポジウム 2017 (CSS 2017) ,pp.7-14(2017)

[4] Leyla Bilge, Davide Balzarotti, William Robertson Engin Kirda Christopher Kruegel:DISCLOSURE:Detecting Botnet Command and Control Servers Through Large-Scale NetFlow Analysis,ACSAC’12 Proceeding of the 28th Annual Computer Security Applications Conference,pp.129-138(2012)

[5] K. Maruhashi, M. Todoriki, T. Ohwa, K. Goto, Y. Hasegawa, H. Inakoshi, and H. Anai: Learning Multi-way Relations via Tensor Decomposition with Neural Networks, In AAAI, to appear. (2018)

表 2　BKDR PlugX による高進行度時間帯の寄与度評価結果

TIME	送信元 IP	送信先 IP	送信元 ポート No	送信先 ポート No	通信種別	パケット長	実行プロセス	コマンド	寄与度
10:17:31	a.a.a.a	b.b.b.b	61272	80	TCP/IP	101	-	-	0.0069
10:14:47	a.a.a.a	b.b.b.b	61235	80	TCP/IP	56	-	-	0.0060
10:15:27	a.a.a.a	b.b.b.b	61245	80	TCP/IP	101	-	-	0.0047
10:10:01	a.a.a.a	b.b.b.b	61076	80	TCP/IP	56	-	-	0.0044
10:14:47	b.b.b.b	a.a.a.a	80	61235	TCP/IP	56	-	-	0.0043
10:11:40	a.a.a.a	c.c.c.c	61185	445	TCP/IP	0	System	-	9E-28
10:18:48	a.a.a.a	c.c.c.c	61298	445	TCP/IP	0	System	-	6E-28
10:13:16	a.a.a.a	c.c.c.c	61191	445	TCP/IP	0	Access_conhost.exe	Conhost.exe	6E-28
10:11:50	a.a.a.a	c.c.c.c	61185	445	TCP/IP	0	Access_PING.EXE	Ping	6E-28
10:10:41	a.a.a.a	d.d.d.d	61165	139	TCP/IP	0	System	-	4E-31