1W-04

決定木と自己組織化マップを用いた標的型攻撃に対するハイブリッド型通信検知手法

中川 雄太郎 †　　安達 直世 †　　滝沢 泰久 †
† 関西大学環境都市工学部都市システム工学科

1 はじめに

近年，標的型攻撃における情報漏えいなどの被害が問題となっている．標的型攻撃では Remote Access Trojan(RAT) が用いられることが多い．RAT は諜報活動を行なうためのマルウェアであり，感染端末の遠隔操作を可能にするため，早急に検知する必要がある．そのため RAT が行なう通信と通常通信の区別することで，RAT 検知を行なう手法の研究が進められている．現在の多くの IDS では，既存 RAT の解析から得られる検知可能な特徴に基づいたシグネチャベースによる検知が行われている．シグネチャ検知では既存 RAT にない特徴を持つ RAT を検知できず，またシグネチャの登録にかかるコストが大きい．そのため機械学習を用いて通信特徴の区別を行なう手法の研究が重要となる．

機械学習を用いた検知は，ミスユース検知とアノマリ検知に大別される．ミスユース検知では既存 RAT の通信データを参考に検知を行なうため，既存攻撃に対して有効だが未知攻撃の検知は困難である．アノマリ検知では通常通信を参考に異常値の検知を行なう．そのため未知攻撃に対して有効だが，既存攻撃に対する検知精度が低い．そこでミスユース検知とアノマリ検知を組み合わせることで，それぞれ単一での検知と比較し，検知精度を向上させる手法が提案されている．

文献 [1] では，ミスユース検知である決定木とアノマリ検知である one-class SVM を階層的に組み合わせ，決定木から得られる情報をアノマリ検知で用いることで通常通信のプロファイルを構築する能力を向上させている．標的型攻撃では，対象に合わせて既存の RAT をカスタマイズした亜種 RAT を用いることが多い．そのため検知対象となる通信では，既存 RAT が持つ特徴の一部が変化している可能性がある．このような標的型攻撃の検知を行なうためには，既存 RAT の通信からより多くの特徴を学習する必要がある．しかし，決定木のみでは既存 RAT 通信の特徴を幅広く学習することは難しい．

そこで本研究では既存 RAT のより多くの特徴を学習するために，決定木と自己組織化マップ (SOM) の学習方法の違いに着目し，ミスユース検知の精度を向上させる．さらに SOM によりミスユース検知とアノマリ検知を並行して行なうことで，未知攻撃に対しても有効な検知手法を提案する．

2 提案手法

提案手法の概要を以下の図 1 に示す．提案手法では，決定木における検知と SOM による検知の 2 つの段階をもって，RAT 通信の検知を行なう．入力段階の処理として，通信データから特徴量 (後述) を抽出し，特徴ベクトルを生成する．

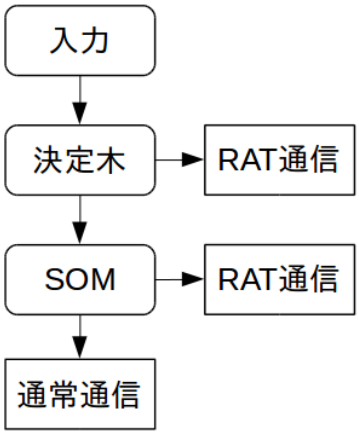

図 1　提案手法の概要

2.1 決定木を用いたミスユース検知

すべての学習データを用いて分類器を生成する．分類器の入力に対して「RAT 通信」と判定される場合は，システム全体の判定として RAT 通信とする.「通常通信」と判定される場合には，入力された特徴量を SOM で判別を行なう．

2.2 SOM を用いたミスユース・アノマリ検知

SOM は決定木で生成される分類器において，正常と判定される終端ノードごとに生成する．各 SOM が学習するデータは対応する終端ノードごとに分割された学習データを用いる．SOM の競合層を 2 つに分割し，片側でミスユース検知，もう片側でアノマリ検知を行なう．ミスユース検知では RAT 通信のみを学習させ，参照ベクトルと類似する入力を検知する．アノマリ検知では通常通

A hybrid detection method against Targeted Attacks on Decision tree and Self-organizing map
†Yutaroh Nakagawa, Naotoshi Adachi and Yusuhisa Takizawa, Kansai University

信のみを学習させ，参照ベクトルから乖離する入力を検知する．

3 実験

提案手法の有効性を示すために，提案手法と先行手法 [1] の比較評価を行なう．

3.1 実験準備

本研究では，検証用データセットとして PRACTICE Dataset 2013[4] を用いた．データセットには 5 つのマルウェアの通信データが収録されている．そのうち 4 つのマルウェア通信データをセッション数で二等分し，半分を学習データ，もう半分をテストデータとした．1 つのマルウェア通信データは学習データに用いず，テストデータのみに含めることで未知攻撃とした．また正常データとして大学研究室のルータにおける 6 日分の通信を取得したデータを用いた．各データのセッション数を以下の表 1 に示す．

表 1　各データのセッション数

	学習	テスト
正常	44066	44066
RAT 通信 (既知)	176266	176266
RAT 通信 (未知)	0	6448

用いる特徴量として，文献 [2, 3] から RAT 通信の検知に有効とされるものを用いた．用いた特徴量を以下に示す．

1. 合計パケット数
2. 初期段階セッションの持続時間
3. OutBound のデータ量 (OB)
4. OutBound のパケット数 (OP)
5. InBound のデータ量 (IB)
6. InBound のパケット数 (IP)
7. OB/OP:OutBound のパケット平均データ量
8. IB/IP:InBound のパケット平均データ量
9. パケット到着間隔分布のエントロピー

3.2 評価指標

以下に混合行列 (表 2) と各指標の定義を示す．式 (1) に示す Detection Rate (DR) は RAT 通信の検知率であり，1 に近いほど高い精度であることを示す．式 (2) で示す False Rejection Rate (FRR) は正常通信を誤って RAT 通信と判定している割合を示し，0 に近いほど誤判定が少ないことを示す．

$$DR = \frac{TP}{TP + FN} \tag{1}$$

$$FRR = \frac{FP}{FP + FN} \tag{2}$$

表 2　混合行列

	予測 (正常)	予測 (RAT 通信)
正常通信	False Positive (FP)	True Positive (TP)
RAT 通信	True Negative (TN)	False Negative (FN)

3.3 実験結果

表 1 に示した学習データを用いて学習を実行し，すべてのテストデータに対して予測を行なうことで各指標を算出した．提案手法と先行手法 [1] の各 DR, FRR をまとめた表 3 を以下に示す．

表 3　各手法の評価値

	DR	FRR
提案手法	0.954	0.046
先行手法	0.631	0.369

検証の結果，提案手法の DR は 0.954，先行手法では 0.631 であり，これは RAT 通信の検出について，提案手法がより正確な検知が可能であることを示す．また提案手法の FRR は 0.046 であり，先行手法では 0.369 となっており，提案手法では通常通信に対して誤った判定を行なうことが少ないことがわかる．

4 まとめ

本研究では，既存 RAT のより多くの特徴を学習するために，決定木と自己組織化マップを組み合わせた検知手法の提案を行った．提案手法と先行手法の比較評価により，提案手法ではより高い精度での検知が可能であることがわかった．

参考文献

[1] Gisug Kim, Seungmin Lee and Sehun Kim: A novel hybrid intrusion detection method integrating anomaly detection with misuse detection, Expert Systems with Applications, Volume 41, Issue 4, pp.1690-1700 (2014).

[2] 蒋丹，面和成：初期段階における Remote Access Trojan の検知手法，Computer Security Symposium 2014，pp.22-24 (2014).

[3] 宇野真純，石井将大，猪俣敦夫，新井イスマイル，藤川和利：エントロピーを用いた初期侵入段階における Remote Access Trojan の通信検知，信学技 IEICE Technical Report SITE2016-68, IA2016-98 (2017).

[4] 神薗 雅紀，畑田 充弘，寺田 真敏，秋山 満昭，笠間 貴弘，村上 純一：マルウェア対策ための研究用データセット MWS Datasets 2013, Computer Security Symposium 2013 21-23 (2013).