講演番号： 207

機械学習を用いたサイバー攻撃のリアルタイム検知方式

D－19　Study on Real-time Cyber-Attack Detection System by using Machine Learning

狩俣　知希†　宮保　憲治†
Tomoki KARIMATA†　Noriharu MIYAHO†
† 東京電機大学大学院情報環境学研究科
† Graduate School of Information Environment, Tokyo Denki University

1. はじめに

近年，既存のマルウェアの挙動を改変した亜種が爆発的に増加し，従来のセキュリティ対策では検知できないマルウェアの亜種も多く出現している[1]．機械学習を用いることにより，既存マルウェアの亜種による攻撃を検知する研究も進みつつある．しかし，機械学習を用いた異常検知手法には正常通信を異常通信と誤検知する可能性も存在する．

本稿では，通信セッションの早期の段階において，異常通信（外部との通信を行うマルウェアに感染した PC の通信）の検知を支援する手法を提案し，評価結果を述べる．

2. 解析用通信データ

解析した正常通信データは，メールの送受信やブラウジングなどにより独自にキャプチャした pcap データを用いた．異常通信データとしては，D3M データセット[2]に含まれる pcap データを用いた．

3. 実験手法

本実験では，[3]を参考に Early Stage から特徴量を抽出し，複数の機械学習アルゴリズムを組み合わせることにより，リアルタイムに誤検知の防止を行う手法を適用した．

通信セッションの Early Stage とその他の通信を分割するための送受信間隔の閾値を 1 秒として Early Stage を定義した．この Early Stage から，正常通信と異常通信とを識別するための特徴量として，Early Stage 内におけるパケット総数，受信パケット数，送信パケット数，受信バイト数，送信バイト数，送受信パケット数，および平均送信パケット長を用いた．正常通信と異常通信のデータを識別するための機械学習アルゴリズムとして，RF(Random Forest)，DT(Decision Tree)，SVM(Support Vector Machine)，NB(Naive Bayes)，KNN(K-Nearest Neighbor)を用いた．

提案手法を用いて構築した処理システムを図 1 に示す．はじめに，抽出した特徴量を学習用データと評価用データに分割する．学習用データを用いて複数の機械学習アルゴリズムにより識別器を生成し，評価用データを逐次識別する．最後に，各識別器における識別結果に対して集計処理を行う．集計処理では，各識別器における識別結果が全て正常通信であれば正常通信，また，全て異常通信であれば異常通信，その他の場合は UNKNOWN と判定する．これは，複数の機械学習アルゴリズムにおける全ての識別器において同一の識別結果であれば，誤識別である可能性は非常に低いと考えられるためである．また，学習用データや評価用データにより異常通信の検知に有効な機械学習アルゴリズム異なると考えられる．したがって，多数決のような集計処理は行わず，1 つでも異常通信であると識別された場合には UNKNOWN の判定を行うようにした．

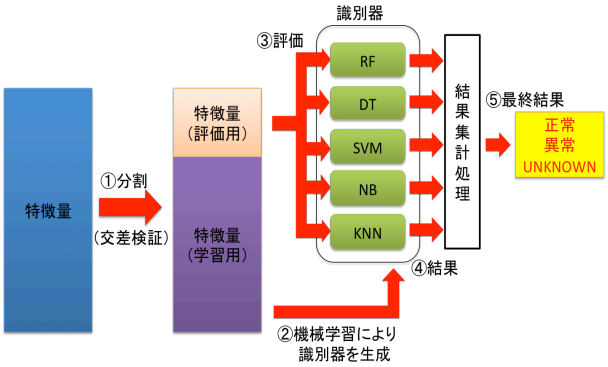

図. 1 提案手法を用いた処理システム

4. 実験結果

提案手法の判定結果により正常通信，もしくは異常通信と判定された通信は全体の約 63.3%であり，すべて正しい判定結果であった．したがって，複数の機械学習アルゴリズムにおける全ての識別器において同一の識別結果であれば，誤識別である可能性は非常に低いという考えは，正しいと考えられる．今回使用したデータセットによる識別では，UNKNOWN と判定された約 36.7%のみを詳しく解析すれば済むので，未知のマルウェアの通信の大半の解析を行う手間を省くことができると考えられる．

5. 今後の課題

通信によっては Early Stage が長くなってしまい，リアルタイム性が損なわれる可能性がある．今後は，リアルタイム性を保証できるサイバー攻撃の検知手法について、検討を進める予定である．

参考文献

[1] G DATA Software , ”G DATA SECURTYLABS REPORT”, 2014, https://public.gdatasoftware.com/Presse/Publikationen/Malware_Reports/GData_PCMWR_H2_2014_EN_v1.pdf（2016/01/19 参照）

[2] 秋山満昭，神薗雅紀，松木隆宏，畑田充弘:”マルウェア対策のための研究用データセット”，情報処理学会報告， Vol.2014-CSEC-66 No.19， pp.1-7, 2014

[3] Dan Jang and Kazumasa Omote, ”An Approach to Detect Remote Access Trojan in the Early Stage of Communication” , 2015 IEEE 29th Internet Conference on Advanced Information Networking and Applications, pp.706-713