IoTマルウェア自動検知のための 悪性コマンド列の特徴に対する概念ドリフト検出

Concept drift detection for malicious command sequence for automatic identification of IoT malware

小川真聖・ネットワーク分科会・情報セキュリティ大学院大学

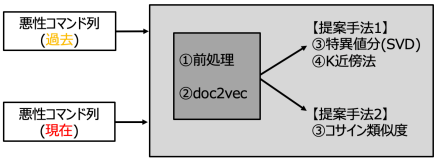
Abstract

In recent years, machine learning has been used in the field of cyber security, and security research using machine learning has achieved remarkable results in attack detection and attack classification. However, the deterioration of machine learning models due to conceptual drift, in which the tendency of data changes. In this research, we focus on the time series of malignant command sequences executed by IoT malware, and propose two conceptual drift detection methods for the characteristics ofmalignant command sequences. The first method is a conceptual drift detection method that treats a malicious command sequence as time-series data and uses a natural language processing method and a K-nearest neighbor method to detect changes. The second method proposes a conceptual drift detection method based on natural language processing and changes in cosine similarity. In the experiment, it was shown that the conceptual drift can be detected by applying the above two proposed methods

背景

近年、IoTマルウェアの増加が深刻な脅威になっている。
それに対し、IoTマルウェアの検知には、機械学習が活用されている。しかし、概念ドリフトに起因する機械学習モデルの劣化により、新種のIoTマルウェアや亜種が検知できなくなるといった問題が存在する。そこで、本研究では、IoTマルウェアの悪性コマンド列における概念ドリフト検出手法を提案する

提案手法



提案手法1：特異値分解とK近傍法による概念ドリフト検出

提案手法2：コサイン類似度による概念ドリフト検出

実験

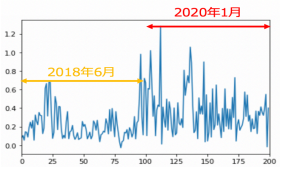
■データセット
- ●2018年6月に観測された悪性コマンド列
- ●2020年1月に観測された悪性コマンド列

■収集方法
- ●ハニーポットCowrieを用いて収集

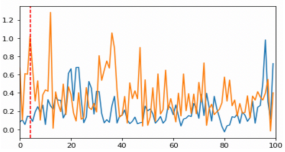
ハニーポット構築環境

種別	環境名
プラットフォーム	AWS EC2
インスタンスタイプ	T2.Micro
CPU	Intel Xeon E5-2670 v2 2.50GHz
OS	Debian
ハニーポット	Cowrie

実験結果



2つの期間で収集された悪性コマンド列データの比較



K近傍法による概念ドリフト検出

コサイン類似度による概念ドリフト検出

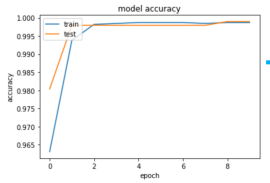
比較	コサイン類似度
2018年5月と2018年6月の比較	0.89
2020年1月と2020年2月の比較	0.87
2018年6月と2020年1月の比較	0.48

評価実験

実際に概念ドリフトが発生しているか検知器を用いて評価

■学習用データ
- ●悪性コマンド列 2003個
- ●良性コマンド列 2840個

■機械学習アルゴリズム
- ●LSTMを使用



→ 約99%の精度が約61%に低下