1W-09

コマンド入力特徴に着目したマルウェア不正活動の検知手法の提案

佐藤 至典† 稲村 浩† 中村 嘉隆†

公立はこだて未来大学†

1 はじめに

近年，スマート家電や家庭用ルータなどの IoT (Internet of Things) デバイスの普及にともなって，IoT デバイスを攻撃対象としたマルウェアが脅威となっている．これらマルウェアの多くは，IoT デバイスで適切に管理されずに動作している Telnet を利用して不正侵入や不正活動を行う．この脅威に対して早急な対策が望まれているが，IoT デバイスのマルウェア対策を行う際には，IoT デバイス特有の性質を考慮する必要がある．例えば，ライフサイクルが長いことや機能・性能が限られている場合があるという性質がある [1]．また，IoT デバイスを攻撃対象としたマルウェアには多種多様なものがある．そのため，特定のマルウェア以外にも対応できる汎用的な手法が必要である．したがって，IoT デバイスの性質を考慮した，ネットワークベースでアノマリ検知型のマルウェア対策手法を検討する．

本来，Telnet は人間がキーボードを用いてコマンド入力を行って対象デバイスを遠隔から操作することが目的である．それに対して，マルウェアはプログラムによってコマンド入力を行って不正活動をする．つまり，Telnet において人間とマルウェアはともにコマンド入力を行うという点は共通するが，コマンド入力の主体は異なることから，人間とマルウェアではコマンド入力特徴に違いがあるという仮説が立てられる．そこで，本研究ではコマンド入力特徴に着目して人間とマルウェアを判別することで，マルウェアの不正活動を検知する手法を提案する．

2 関連研究

コマンド入力特徴に関する研究として，入力ミスや入力速度，実行するコマンドなどに着目して，正規ユーザと非正規ユーザを判別する研究などがある[2][3]．これらの研究から，人間と人間の判別が可能であれば，人間とマルウェアの判別も可能であると考えられる．しかし，これらの研究はホストベースで行っていることに加えて，正規ユーザのコマンド入力特徴を事前に学習させる必要がある．そのため，IoT デバイスのマルウェア対策として実際に運用するのは困難である．

Detection of Malware Activity Based on Command Input Feature
Yoshifumi Sato†, Hiroshi INAMURA†, Yoshitaka NAKAMURA†
Future University Hakodate†

3 アプローチ

人間とマルウェアを判別するためには，それぞれのコマンド入力特徴を求める必要がある．そのため，人間とマルウェアによるコマンドの入力列を収集する実験を行う．その後，収集した入力列を元に，教師あり機械学習法の 1 つである，SVM (Support Vector Machine) を用いて人間とマルウェアの判別を行う．

4 実験

4.1 入力列の収集実験

4.1.1 マルウェアによる入力列

2017/09/21 から 2017/10/21 の期間でハニーポットを断続的に運用してマルウェアの不正活動ログの収集を行った．Linux などで標準インストールされているネットワーク調査ツールである Tcpdump を用いて，マルウェアが不正活動を行っている時のパケットダンプの収集も行った．その結果，マルウェアがハニーポットに侵入した後に 1 つ以上のコマンドを実行して不正活動を行った 620 件のデータを収集した．

4.1.2 人間によるマルウェアに近い入力列

人間がマルウェアに近い入力列を行う場合を想定した入力列の収集を行った．前項の実験と同じ環境を用いて，日常的にパソコンを使用している男性 11 名を対象に入力列の収集を行った．入力内容は，前項のハニーポットで収集したマルウェアの不正活動ログにおける入力列の一部を用いた．被験者に入力してもらったコマンドとその順序は表 1 のとおりである．

表 1 マルウェアを模した入力列

順序	コマンド
1	enable
2	system
3	shell
4	sh
5	/bin/busybox Mirai

4.1.3 人間による通常時の入力列

人間が通常のサーバ操作を行う場合を想定した入力列の収集を行った．4.1.1 項の実験と同じ環境を用いて，日常的にパソコンを使用している男性 5 名を対象に入力列の収集を行った．被験者に入力してもらったコマンドとその順序は表 2 のとおりである．

表 2 人間による通常操作時を模した入力列

順序	コマンド
1	dpkg -l vim
2	sudo apt-get install vim
3	mkdir test
4	cd ./test
5	vim test.txt
6	Hello World
7	:wq!

4.2 SVM を用いた判別

4.2.1 評価方法

実験で収集したデータを教師データとテストデータの 2 つに分割するホールド・アウト検定を用いて評価を行う．人間を良性としてマルウェアを悪性とする．特徴ベクトルは，収集した入力列から表 1 に示す 8 つの値を元に生成して，標準化を行う．そして，人間とマルウェアそれぞれ 7 割のデータを教師データとして，残り 3 割をテストデータとする．

特徴ベクトルの 1 つである WPS とは，1 秒あたりの入力単語数を示したものである．次に，コマンド入力の間隔とは，あるコマンドを入力してサーバがそのコマンドに対して応答を行った後から，次のコマンドを入力するまでに要した秒数のことを示す．そして，パケットサイズとは，クライアントからサーバに対して送信されたパケットのバイトサイズのことを示す．最後に，6 から 8 つ目はそれぞれのエスケープシーケンスを行った回数を示している．

表 3 特徴ベクトル

1	Word Per Second (WPS)
2	コマンド入力の間隔の平均
3	コマンド入力の間隔の標準偏差
4	パケットサイズの平均
5	パケットサイズの標準偏差
6	コマンド入力の補完回数 (TAB)
7	コマンド入力の修正回数 (Back Space)
8	コマンド入力の中断などの回数 (Ctrl+C)

4.2.1 実験結果

ホールド・アウト検定を用いた SVM による人間とマルウェアの判別結果における精度評価を表 4 に示す．また，混同行列を表 5 に示す．

表 4 判別結果における精度評価

	精度	再現率	F 値
人間	1.00	0.50	0.67
マルウェア	0.98	1.00	0.99

表 5 判別結果における混同行列

		真の結果	
		人間	マルウェア
予測結果	人間	3	0
	マルウェア	3	124

5 考察

評価実験より，マルウェアは高い確率で判別できるが，人間をマルウェアと誤判別することが多いという結果になった．原因としては，人間による入力列のデータ数がマルウェアに対して少ないということが考えられる．また，それぞれのコマンド入力特徴ごとの精度評価を行っていないため，判別に適さないコマンド入力特徴が含まれていることも考えられる．

6 まとめ

我々は，人間とマルウェアそれぞれのコマンド入力特徴に着目することで，マルウェアによる不正活動を検知する手法を提案するため実験を行った．その結果，コマンド入力特徴を元にするとマルウェアの判別を高い精度で行えるが，人間をマルウェアと誤判別することがあるという結果になった．今後の課題として，特徴量の寄与度を算出することで，特徴ベクトルの再検討を行う．また，人間が LINE モードを用いて入力を行った場合のコマンド入力特徴の変化を考慮して，追加実験を行い入力列の収集と分析を行う．

参考文献

[1] 総務省，経済産業省：IoTセキュリティガイドラインver1.0，(オンライン) 入手先 〈http://www.soumu.go.jp/main_content/000428393.pdf〉(参照 2017-01-12).

[2] 中國真教，堂薗浩，野口義夫：キーボード入力の監視による不正利用者の判別方法，情報処理学会論文誌，Vol. 41，No. 12，pp. 3276-3284，2000.

[3] 森裕子，小松賢嗣，赤池英夫，粕川正充，角田博保：打鍵データに基づく個人認証システムの作成と評価，情報処理学会研究報告，Vol. 1991，No. 5(1990-HI-034)，pp. 1-10，1991.