7ZC-04

ネットワークトラフィックを用いた 分散表現に基づく亜種マルウェアの侵入活動検知

堀江　晟矢†　　梅澤　猛‡　　大澤　範高‡

所千葉大学工学部総合工学科情報工学コース†　　千葉大学大学院工学研究院‡

1. はじめに

ネットワークトラフィックの解析によるマルウェア検知は，IoT 機器などウイルス対策ソフトの導入・更新が困難な対象のセキュリティ対策に有効であるが，膨大なトラフィックの中から悪性なものを判別する検知モデル構築の自動化と効率化が課題となる．

マルウェア検知にあたっては，既存のマルウェアを改変した亜種マルウェアを高い精度で検出することが必要となる．セキュリティソフトで幅広く使用されているパターンマッチング方式では，亜種の検知精度は充分とは言えない．水野らはマルウェアの活動検知において使用されることが多い HTTP 通信に注目し，機械学習を用いてそのヘッダ情報から特徴量を抽出して悪性通信と良性通信の分類を行った[1]．実データを利用した評価実験では，正解率 0.905 で分類が可能であることを示した．しかし，亜種の検出性能は明らかにされていない．そこで本研究では，分散表現を利用することで，マルウェア活動の検知モデル構築の効率化を目指す．

分散表現は，自然言語処理の分野などで使われる，単語や文書を多次元の実数ベクトルで表現する技術である．代表的なツールとして，単語の分散表現を獲得する word2vec や文書や文章の分散表現を獲得する doc2vec が挙げられる．これらのツールによって得られる分散表現は加法構成性を有し，似た意味を持つ単語や文章の分散表現ベクトルは分散表現が成すベクトル空間で近くに分布するという特徴を持つ．

通信トラフィックを分散表現することで，亜種マルウェアが次元圧縮された類似性の高い分散表現ベクトルとして表現され，従来は検出が難しかった学習データに含まれない亜種マルウェアの検出が可能になることが期待できる．

Malware variant detection based on distributed semantic representation of network traffic

†Seiya Horie, Department of Information Engineering, Faculty of Engineering, Chiba University

‡Takeshi Umezawa, Noritaka Osawa, Graduate School of Engineering, Chiba University

2. 提案手法

提案手法によるマルウェアの検出手順を図 1 に示す．ネットワークトラフィックを分散表現学習に基づいてベクトル化し，ベクトル化したネットワークトラフィックを判別するマルウェア検知モデルを構築する．本研究では，モデル構築にパケットのヘッダ情報を利用する．ペイロードは，現代の情報通信において暗号化されていることが多く，適切な情報抽出が困難であるため利用しない．

パケットヘッダのフィールド名と，そのデータの組を「単語」と定義し，その分散表現への変換モデルを作成する．分散表現ベクトルとして表現されたネットワークトラフィックを時系列解析する分類モデルを教師あり学習によって構築する．

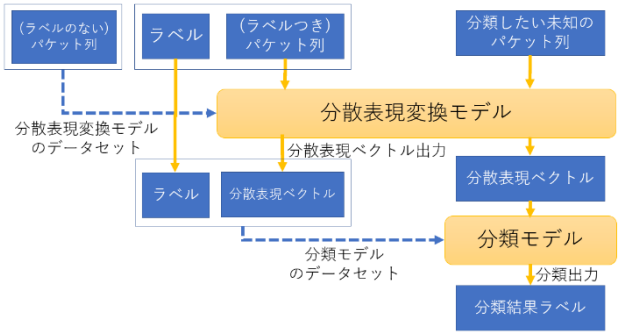

図 1　提案手法のフローチャート

3. 実験

提案手法による良性通信と悪性通信の分類精度を評価する．ネットワークトラフィックに tokenizer を用いて整数列に変換し，それを入力として分散表現変換モデルを作成する．その後，分散表現変換モデルで変換した分散表現ベクトルを用いて分類モデルを作成し，そのモデルの分類精度を評価する．

3.1 データセット

悪性通信データセットとして，The Malrec Dataset [2]の Packet Capture (PCAP)ファイルを利用する．本データセットは，2014/12/07-2016/12/03 の期間に Georgia Tech Information Security Center

やプロバイダ，アンチウイルスベンダーから提供された総数 66,301 件のマルウェア検体に対して，サンドボックス解析システムを用いて挙動を解析・記録したデータ群である．

良性通信のデータセットとしては，NCD in MWS Cup 2014 [3]の 5 つの PCAP ファイル，ならびに Stratosphere Lab が提供する Normal Datasets における CTU-normal-18，CTU-normal-20, CTU-normal-33 の PCAP ファイルを利用した．

3.1.1 分散表現変換モデルのデータセット

悪性と良性の通信ログを記録した PCAP ファイルを訓練データとして分散表現変換モデルを構築する．

3.1.2 分類モデルのデータセット

良性通信と悪性通信の分類実験に使用した学習データの内訳を表 1 に示す．

悪性通信には，次の 5 種類のファミリが含まれる．

(a) P2P-Worm.Win32.Sytro,
(b) Trojan.Win32.Reconye,
(c) Virus.Win32.Expiro
(d) Virus.Win32.PolyRansom
(e) Worm.Win32.WBNA

検証データの悪性通信データには訓練データには含まれない亜種を使用する．

表 1　良性・悪性の分類のためのデータの内訳

	良性	悪性				
		(a)	(b)	(c)	(d)	(e)
訓練	31	220				
		30	43	78	39	30
検証	12	50				
		10	10	10	10	10

3.2 モデル構築

3.2.1 分散表現変換モデル

悪性と良性の通信ログを記録した PCAP ファイルのパケットの「単語」を数列に変換し，それを基に分散表現変換モデルを作成する．語彙数は 551,981 であった．分散表現変換モデルは Skip-gram モデルを用いて作成した． negative sample 数は 4，次元数は 50，ウィンドウサイズは 3，エポック数は 10 とした．

3.2.2 分類モデル

パケットの「単語」を整数に変換済みのデータを分散表現ベクトルに変換し，それを用いて図 2 の手順にしたがって分類モデルを作成する．

分類モデルは Python のニューラルネットワークライブラリ Keras を用いて構築した．この分類モデルは，ネットワークトラフィックを変換して得た分散表現ベクトルの系列データを入力とし，良性・悪性の二値分類を行う．時系列データを用いるため，学習モデルには LSTM を用いた．ユニット数 50 の LSTM 層，過学習防止のための Dropout 層，出力ユニット数 1，活性化関数をシグモイド関数とした Dense 層で構成される．エポック数は 20 とした．

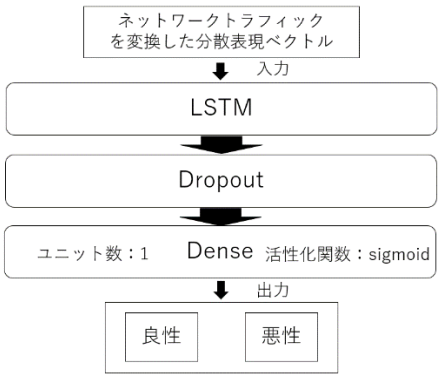

図 2　分類モデル

4. 評価

訓練データで学習したモデルに対して検証用のデータを入力し，悪性通信の検出性能を評価する．

評価指標として，正解率，適合率，再現率，F-値の 4 つを利用する．表 2 に，分類結果を示す．今回構築したモデルで悪性の亜種も検出可能であることが示唆された．ただし，適切な評価を行うためには，今後，より多くのデータでの検証が必要である．

表 2　良性・悪性の分類結果

		訓練		検証	
		良	悪	良	悪
正解	良	31	0	11	1
	悪	1	219	0	50
正解率		0.9960		0.9837	
適合率		0.9687		1.0000	
再現率		1.0000		0.9166	
F1 スコア		0.9841		0.9564	

参考文献

[1] 水野翔，畑田充弘，森達哉，後藤滋樹．“HTTP ヘッダフィールドの可変性に基づくマルウェア感染端末の特定”，コンピュータセキュリティシンポジウム 2016 論文集，2016(2)，pp.632-639.
[2] The Malrec Dataset, https://giantpanda.gtisc.gatech.edu/malrec/dataset/
[3] NCD in MWS Cup 2014, https://www.iwsec.org/mws/datasets.html