R-6

脅迫文の特徴を利用したランサムウェア検知手法の研究

橋浦研究室　　1145221　小林 雄祐

１．はじめに

2016 年ごろからランサムウェアによる被害報告が増加し始め，2017 年 5 月には世界的な流行を見せた．現在，主に利用されている検知手法では日々変化し続けるランサムウェアに対応しきれていないという課題がある．

２．研究目的

ランサムウェアは身代金を要求するための脅迫文が必ず用意されている． 本研究ではこの脅迫文を利用し，ランサムウェアを検出する手法を提案する．

これにより最新のランサムウェアに対応することができ，なおかつ仮想環境のような特別な環境を必要としない環境が実現可能である．

３．提案手法

本研究ではランサムウェア検知のため，ランサムウェアの脅迫文の特徴を利用する．ランサムウェアの脅迫文を利用した手法としては，Amin ら[1]の研究がある．Amin らはランサムウェアを仮想環境で実行し，その際に画面に表示された脅迫文をスクリーンショットとして保存し，ユーザーへ提示することによって，検出を行うものである．

本研究では，ランサムウェアの脅迫文は一般的なテキストとは大きく異なるという点に着目し，分類器を用いて検出を行う．以下に本手法の手順を示す．

I. 実行ファイルをツールで開く
II. 実行ファイルから画像を見つける
III. 見つけた画像を外部へ保存する
IV. 保存した画像を OCR(光学的文字認識)にかける
V. OCR で出力したテキストを分類する
VI. 分類結果を通知する

４．実装

本研究で開発するツールは動作環境を Windows OS 環境と定め，Windows アプリケーションとして開発する．ツールに必要な 3 つの機能について，それぞれの概要を以下に示す．

ｉ．画像の抽出機能

Windows API を利用して実行ファイルにある画像を探索する．画像があった場合は，その画像を外部へ保存する．

ii．画像の分類機能

上記の抽出機能で保存された画像を OCR にかけ，画像からテキストを抽出する．さらに，そのテキストをナイーブベイズ分類器にかける．分類器にかけられたテキストはランサムウェアか，そうでないかに分類する．

iii．結果の通知機能

分類された結果を利用者に通知する．

５．実験と評価

開発したツールが，ランサムウェアかそうでないかを分類出来るかを確認するために実験を行った．データの内訳はランサムウェア 75 件，ランサムウェアでないもの 75 件の英文とした．実験で使用したデータはランサムウェアの脅迫文の場合，図 1 で示すような画像データになる．これらのデータを K 分割交差検証によって評価をする．K 分割交差検証とはデータ全体を K 個の項目に分割して，そのうち 1 個をテスト項目とし，残りを訓練項目にする．これを K 個に分割したすべての項目がテスト項目として検証されるまで続ける．これにより，すべての項目をテスト項目として扱う事が出来る．

図 1．ランサムウェアの脅迫文画像の例

今回の実験では，用意した全てのデータをランサムウェアとランサムウェアでないもの

に分け，さらにそれを 3 個ずつに分割する．

以下の表 1 に示す通り，25 件ごとに分割されたデータが合計 6 個出来る．データの判別がしやすいように A から F までのラベルを付与する．また，ランサムウェアを R，ランサムウェアでないものを NR と表す．

表 1．検証のために使用するデータ群

R	A 25 件	B 25 件	C 25 件
NR	D 25 件	E 25 件	F 25 件

分割されたデータを使って合計 3 回の検証を行う．その組み合わせを以下に示す．

・検証 1　訓練(A，B，D，E)　テスト(C，F)
・検証 2　訓練(A，C，D，F)　テスト(B，E)
・検証 3　訓練(B，C，E，F)　テスト(A，D)

6．実験結果と考察

前述した実験方法に従い実験を行った．検証 1 から検証 3 までの結果をまとめたものを表 2 に示す．

表 2．提案手法による分類結果

		正解	
		R	NR
分類器の出力	R	72(96%)	2(3%)
	NR	3(4%)	73(97%)

表 2 より，適合率と再現率の調和平均の F 値を算出すると約 96.7%である．また偽陽性に分類されてしまったデータが 2 件(以下①，②)，偽陰性に分類されてしまったデータが 3 件(以下③，④，⑤)あった．

偽陽性と偽陰性に分類された合計 5 件のデータについて考察する．①のデータは有料ソフトウェアの購入画面のテキストである．ソフトウェアはすべて Adobe 関連のものであった．ランサムウェアでない訓練データ内には Adobe に関連したテキストは 0 件であった一方で，ランサムウェア側の訓練データ内には Adobe Flash Player について記述されているものが 3 件見つかった．このことから，Adobe に関連したテキストによって分類がランサムウェア側に寄ってしまった原因のひとつと考えられる．②のデータは決済時のテキストである．その支払い方法は Bitcoin のみであった．ランサムウェアではない訓練データ内には Bitcoin で支払いができるものは 1 件しかなかった．一方で，ランサムウェア側の訓練データ内には Bitcoin で支払いできるものが 14 件あった．以上のことから，Bitcoin での支払いに関連するテキストが原因で偽陽性に分類されてしまったと考える．③，④のデータはテキストのほとんどがポルトガル語およびドイツ語で記述されていた．これらのデータを英語で書かれたデータを利用して分類しようとしたために正しく分類できなかったと考えられる．両方とも偽陰性に分類されたことから，まったくテキストが読み取れていない訳ではなく英単語として読める部分があったことが考えられる．⑤のデータはランサムウェアのテストデータとして定義されていたが，実際はランサムウェアでないテストデータとして定義すべき内容であったことが分類によって判明した．このデータは身代金を要求するような脅迫文ではなかった．

表 3．偽陽性と偽陰性の原因

番号	分類結果	原因
①	FP	Adobe 関連の情報
②	FP	Bitcoin 関連の情報
③	FN	英文ではないため
④	FN	英文ではないため
⑤	FN	テストデータの不備

7．まとめと今後の課題

今回行った実験より，F 値が約 96.7%であることがわかる．このことから，ランサムウェアの脅迫文を利用した検知手法は，効果が見込めるといえる．課題として考えられることとして，検知できないケースが存在することが挙げられる．提案手法では訓練データに使われていないテキストには対応できない．そのため，もしも脅迫文の文面が変化するなどしてテキストが変わってしまったら新たに訓練データを更新する必要がある．今後の発展として，画像を用いない脅迫文を表示するランサムウェアに対応出来るようにすることで，提案手法によるランサムウェアの検知はより効果を発揮することが期待される．

参考文献

[1] Amin Kharraz, Sajjad Arshad, Collin Mulliner, William Robertson, Engin Cerda, "UNVEIL: A Large-Scale, Automated Approach to Detecting Ransomware," Proceedings of the 25th USENIX Security Symposium, pp.757-772, Aug. 2016.