Detecting Drive-by Download Attacks from Proxy Log Information using Convolutional Neural Network

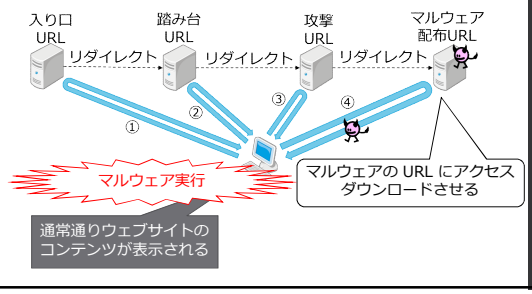
畳み込みニューラルネットワークを用いたプロキシログ情報からのドライブバイダウンロード攻撃検知

大阪大学大学院情報科学研究科
情報ネットワーク学専攻　村田研究室
山西宏平

1

ドライブバイダウンロード攻撃

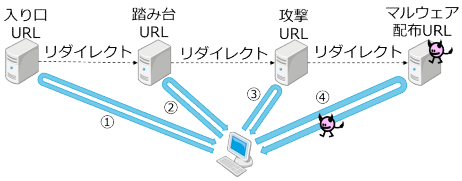
- 水面下で悪意のあるコードを実行させる攻撃　(DbD 攻撃)
- 改ざんしたウェブサイトを利用



2

ドライブバイダウンロード攻撃

- 水面下で悪意のあるコードを実行させる攻撃　(DbD 攻撃)
- 改ざんしたウェブサイトを利用



DbD 攻撃による被害を抑えるため、近年は攻撃発生時の対策に加えて
発生後のログ解析による検知も重要（本研究の対象領域）

3

研究の背景と目的

- コンテンツの解析
  - リソースが必要なため限られた環境でしか不可
- プロキシログ等のアクセス履歴の解析
  - 一般企業等でも行われており早期検知に利用可能

↓

- 本研究の目的
  - アクセス履歴情報を用いて DbD 攻撃を早期に検知

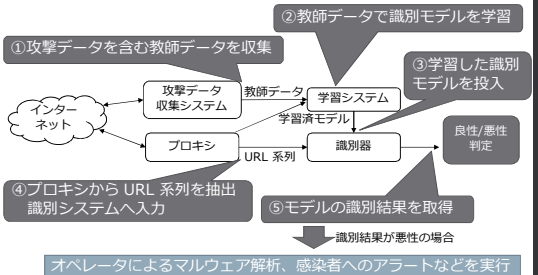
アクセス履歴情報: ユーザごとのアクセス先 URL の時系列

時刻	アクセス先 URL	ソース IP アドレス
13:14:05	aaa.asia/index.html	192.168.130.xxx
13:14:05	aaa.asia/frame.html	192.168.130.xxx
13:14:07	bbb.biz/index.html	192.168.129.yyy
13:14:08	aaa.asia/img.png	192.168.130.xxx
⋮	⋮	⋮

4

URL 系列に基づく攻撃検知

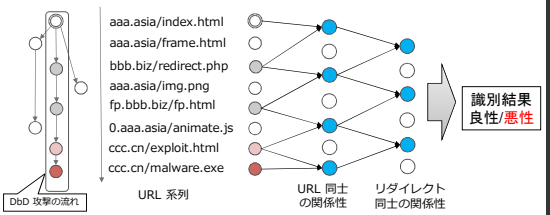
- DbD 攻撃発生時の URL 系列の特徴を学習し、攻撃発生を検知



5

URL 系列の識別手法

- 単一のウェブサイトにアクセスした際に生じる URL 系列を識別する手法[1]
  - Event De-noising CNN (EDCNN)を用いて学習・識別
    - URL の組を畳み込み、リダイレクトの情報を伝搬するニューラルネットワーク
    - URL の特徴量の系列を入力、良性・悪性の判定結果を出力

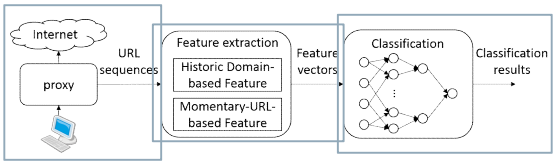


[1] Toshiki Shibahara et al., "Malicious URL Sequence Detection using Event De-noising Convolutional Neural Network," to be presented at IEEE International Conference on Communications, 2017.

6

悪性 URL 系列検出の流れ

1. プロキシログから URL 系列を取得
2. 系列中の各 URL を特徴抽出してベクトル化
   - URL の文字列や関連ドメイン / IP アドレスの特徴
3. ベクトル列となった URL 系列を識別器に入力



7

URL 系列の抽出

- プロキシはアクセス時刻とアクセス先を記録
  - アクセスしたウェブサイトの URL と css や画像などは近い時刻に連続的に記録される

⇒ アクセス時刻をもとに URL 系列を切り出す

時刻	アクセス先 URL
13:14:05	ウェブサイト A の URL
13:14:05	A を構成するコンテンツの URLs
⋮	
13:14:08	
13:15:22	ウェブサイト B の URL
13:15:23	B を構成するコンテンツの URLs
⋮	
13:15:27	
13:15:55	…

⇒ ウェブサイト A の URL 系列

⇒ ウェブサイト B の URL 系列

特定ユーザに対するアクセス履歴情報

8

URL 系列抽出時に発生する問題

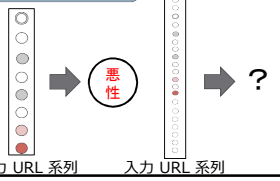
- プロキシログから**時刻ベース**で URL 系列を抽出

時刻	アクセス先 URL
13:14:05	ウェブサイト A の URL
13:14:05	A の css
13:14:06	A の画像
13:14:07	ウェブサイト B の URL
13:14:08	A の Javascript
13:14:08	B の画像
⋮	⋮

ウェブサイト A の読み込み途中にウェブサイト B にもアクセス

複数のウェブサイトに連続してアクセスするとそれらが同じ URL 系列に切り出される

そのままでは EDCNN で攻撃を検知できない



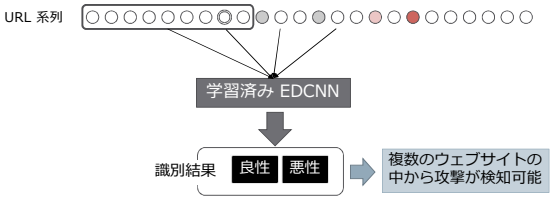
入力 URL 系列　　入力 URL 系列

9

複数のウェブサイトのアクセスが混在した場合への対処

- 同じウェブサイトから辿られた URL 同士は連続している割合が高い
  - ユーザの各ウェブサイトへのアクセスは完全に同時ではないため

⇒ URL 系列の一部分を随時判定することで対応



10

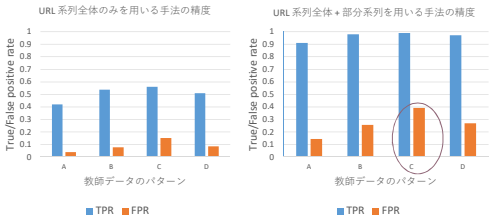
評価環境

- データセット
  - 良性 URL 系列
    - ① 有名サイトのリストに掲載されたウェブサイトにアクセスしたときの URL の系列
    - ② 研究室内でのトラヒックをキャプチャした URL の系列
  - 悪性 URL 系列
    - ① 公開ブラックリストに掲載されたウェブサイトの URL の系列
    - ② 研究室内でキャプチャした URL と上記の URL を混在させた系列 (テストにのみ使用)
- 教師データ
  - 総悪性サンプル数 ≪ 総良性サンプル数 のため、比率を調整
  - 本評価では 4 パターンの比率を用意
    - パターン A …　良性②：悪性① = 1：1
    - パターン B …良性①：良性②：悪性① = 1：1：1
    - パターン C …良性①：良性②：悪性① = 1：1：2
    - パターン D …良性①：良性②：悪性① = 1：1：3

11

複数のウェブサイトを含む URL 系列の検知性能

- 2016 年 10 月 25 日以前のデータで学習したモデルを使用
- 10 月 26 日以降 1 か月間に収集した URL 系列を識別



他のウェブサイトの URL が混ざることで約半数の場合で攻撃が検知できなくなる

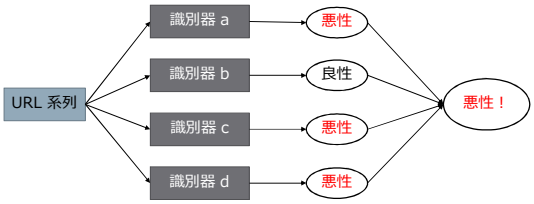
系列の各部分を入力することで 9 割以上の場合で攻撃検知可能になる
ただし、FPR も最大で 4 割弱まで増加

※評価指標: $TPR(True\ Positive\ Rate) = \frac{\text{悪性と識別した数}}{\text{悪性テストデータ数}}$, $FPR(False\ Positive\ Rate) = \frac{\text{悪性と識別した数}}{\text{良性テストデータ数}}$

12

FPR の軽減方法

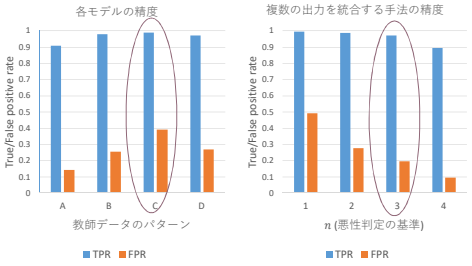
- 複数のモデルの出力を統合
  - n 個以上のモデルが悪性と識別したら悪性と判定
    - 多少、モデルが誤検知や見逃しをしても他が正しければ その影響を吸収できる



13

出力の統合の結果

- パターン A~D の教師データでそれぞれ学習した 4 つの EDCNN の出力を利用した場合



$n = 3$ の場合、FPR が最悪の場合 (学習パターンC) と比較して検出率を大きく落とさずに FPR を半減することができた

14

まとめと今後の課題

- プロキシログの情報から DbD 攻撃を検知
  - ログから抽出した URL 系列に CNN を適用
    - リダイレクトの特徴と文章の構文関係の類似性に着目
  - URL 系列の特性に合わせ CNN を拡張した EDCNN を提案
    - 悪性な URL 同士の組を畳み込んだ結果を伝搬
  - 部分系列を考慮することで同時刻帯に 複数のウェブサイトの URL が混在した場合に対応
  - 複数のモデルの出力を統合して誤検知を軽減
- 今後の課題
  - FPR の更なる軽減
    - 現状では URL 系列が長いものほど誤検知しやすい
  - 適切な教師データの期間・量の調査
  - より実用を想定した実験

15