畳み込みニューラルネットワークを用いた URL 系列に基づくドライブバイダウンロード攻撃検知

山西宏平[1],芝原俊樹[2],高田雄太[2],千葉大紀[2],秋山満昭[2],八木毅[2],
大下裕一[1],村田正幸[1]

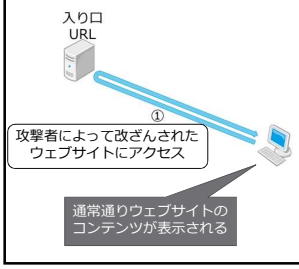
1 大阪大学大学院情報科学研究科

2 NTT セキュアプラットフォーム研究所

1

ドライブバイダウンロード攻撃

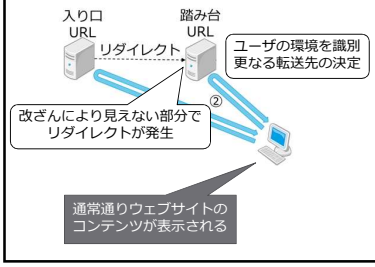
- 水面下で悪意のあるコードを実行させる攻撃 (DbD 攻撃)
- 改ざんしたウェブサイトを利用



2

ドライブバイダウンロード攻撃

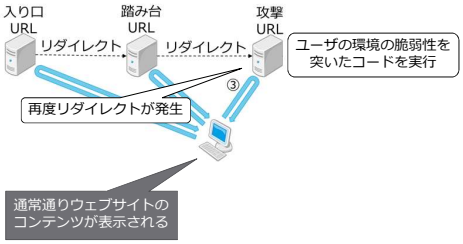
- 水面下で悪意のあるコードを実行させる攻撃 (DbD 攻撃)
- 改ざんしたウェブサイトを利用



3

ドライブバイダウンロード攻撃

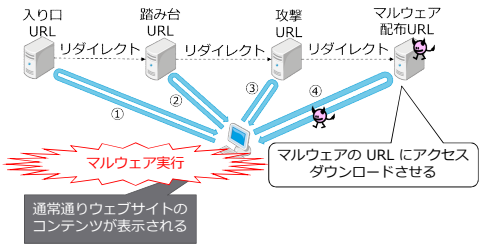
- 水面下で悪意のあるコードを実行させる攻撃 (DbD 攻撃)
- 改ざんしたウェブサイトを利用



4

ドライブバイダウンロード攻撃

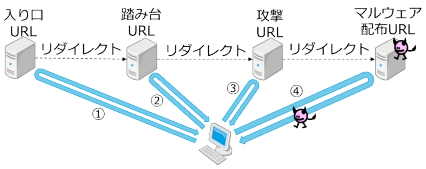
- 水面下で悪意のあるコードを実行させる攻撃 (DbD 攻撃)
- 改ざんしたウェブサイトを利用



5

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)

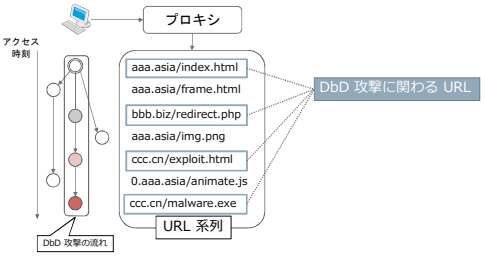
6

DbD 攻撃の検知

- 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るため早期検知も現実的
- 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
- アクセス履歴情報から得られるもの
  ⇒時系列的なアクセス先 URL の情報 (URL 系列)

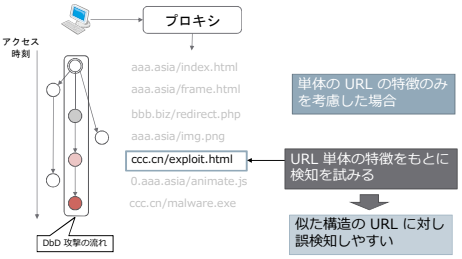
7

URL 系列への着目



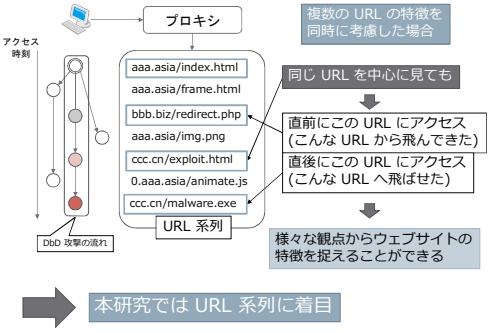
8

URL 系列への着目



9

URL 系列への着目



本研究では URL 系列に着目

10

URL から抽出される特徴量

ID	特徴量
1	ドメインの長さ
2	ファイル名の長さ
3	URL の長さ
4	パスの長さ
5	公開ブラックリスト掲載の有無
6	ファイル名内に知られた悪性パターンの登場の有無
7	サブドメインの有無
8	URL 中の IP アドレスの有無
9	URL 中のポート番号の有無
10	ドメインに対応する IP アドレスの総数
11	AS 番号
12	IP アドレスに対応した経度
13	TLD
14	IP アドレスに対応した国名
15	IP アドレスに対応した都市名

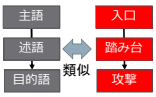
ID	RHP の特徴量
1	BGP prefix 数 (FQDN)
2	BGP prefix 数 (SLD)
3	BGP prefix 数 (2LD)
4	登録国の総数 (FQDN)
5	登録国の総数 (SLD)
6	登録国の総数 (2LD)
7	IP アドレス数 (SLD)
8	IP アドレス数 (2LD)
9	経路数 (FQDN)
10	AS 番号の総数 (FQDN)
11	AS 番号の総数 (SLD)
12	AS 番号の総数 (2LD)
13	レジストリ数 (FQDN)
14	レジストリ数 (SLD)
15	レジストリ数 (2LD)
16	登録後の日数 (FQDN)
17	登録後の日数 (SLD)
18	登録後の日数 (2LD)

ID	RHDN の特徴量
19	FQDN 数
20	長さの平均
21	長さの標準偏差
22	出現確度の平均 (1-gram)
23	出現確度の中央値 (1-gram)
24	標準偏差 (1-gram)
25	出現確度の平均 (2-grams)
26	出現確度の中央値 (2-grams)
27	出現確度の標準偏差 (2-grams)
28	出現確度の平均 (3-grams)
29	出現確度の中央値 (3-grams)
30	出現確度の標準偏差 (3-grams)
31	TLD 数
32	.com の割合
33	出現確度の平均 (TLD)
34	出現確度の中央値 (TLD)
35	出現確度の標準偏差 (TLD)

11

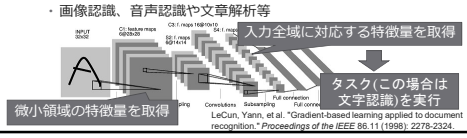
URL 系列の分析

- URL 系列の分析方法
  - DbD 攻撃の一連のリダイレクト関係は文章の構文関係と類似
  - 文章解析は畳み込みニューラルネットワーク (CNN) の得意分野の一つ



⇒**URL 系列の分類にも CNN が有効な可能性が高い**

- CNN
  - 局所領域の特徴抽出 (畳み込み) を繰り返すネットワーク
  - 教師有り学習を行うことで様々なタスクをこなすモデルが作成可能
    - 画像認識、音声認識や文章解析等

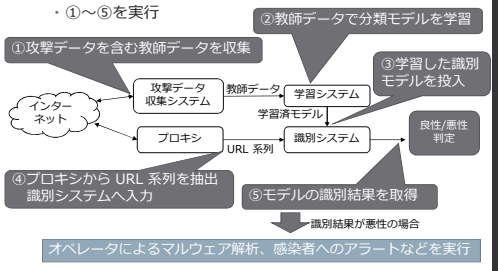


LeCun, Yann, et al. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86.11 (1998): 2278-2324.

12

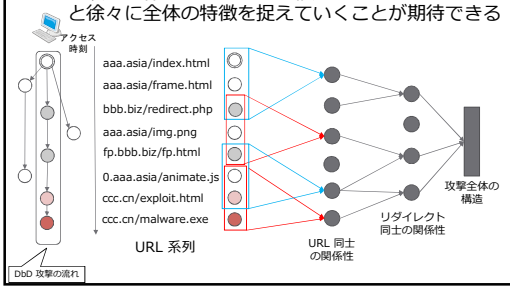
アプローチの全体像

- DbD 攻撃発生時の URL 系列の特徴を学習し、攻撃発生を検知
  - ①～⑤を実行



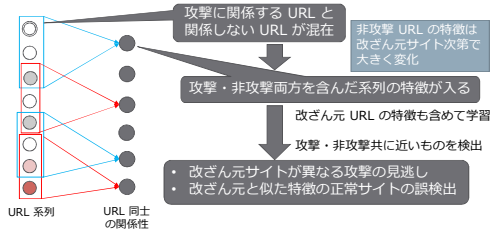
攻撃の一連の URL の畳み込み

URL 同士のリダイレクト関係
⇒リダイレクトとリダイレクトの関係
⇒攻撃全体のリダイレクト構造
と徐々に全体の特徴を捉えていくことが期待できる



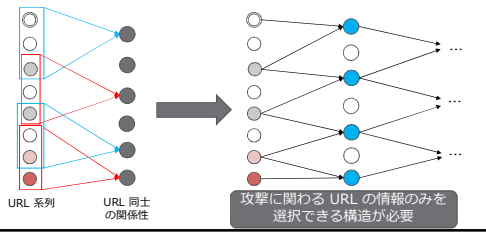
CNN 適用にあたる問題点

- DbD 攻撃発生時の URL 系列の特性
  - 攻撃とは関係のない URL が混在している
- CNN の特性
  - 入力中の一定範囲の情報を畳み込み局所的な特徴量を得る



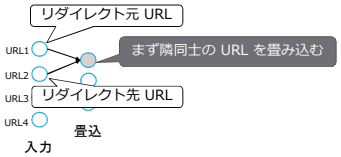
CNN 適用にあたる問題点

- DbD 攻撃発生時の URL 系列の特性
  - 攻撃とは関係のない URL が混在している
- CNN の特性
  - 入力中の一定範囲の情報を畳み込み局所的な特徴量を得る



Event-Denoising CNN

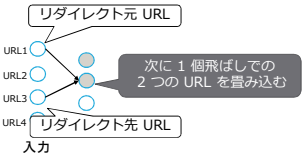
- 攻撃に関係しない URL の影響を削減するよう CNN を改良した **Event-Denoising CNN** (**EDCNN**) を提案
- EDCNN における畳み込み
  - 系列中の近くにある 2 つの URL の組を畳み込み
    - 攻撃に関わるリダイレクトの可能性に相当する値を出力
  - 起点のリダイレクト元 URL 毎に最高値の組を抽出 (プーリング)
    - 各 URL から最も攻撃に関わっている可能性が高い組を探索



URL1 を起点とした URL 組の畳み込みとプーリング

Event-Denoising CNN

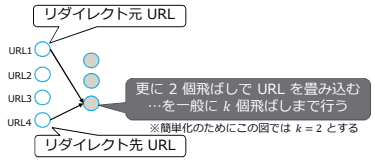
- 攻撃に関係しない URL の影響を削減するよう CNN を改良した **Event-Denoising CNN** (**EDCNN**) を提案
- EDCNN における畳み込み
  - 系列中の近くにある 2 つの URL の組を畳み込み
    - 攻撃に関わるリダイレクトの可能性に相当する値を出力
  - 起点のリダイレクト元 URL 毎に最高値の組を抽出 (プーリング)
    - 各 URL から最も攻撃に関わっている可能性が高い組を探索



URL1 を起点とした URL 組の畳み込みとプーリング

Event-Denoising CNN

- 攻撃に関係しない URL の影響を削減するよう CNN を改良した **Event-Denoising CNN** (**EDCNN**) を提案
- EDCNN における畳み込み
  - 系列中の近くにある 2 つの URL の組を畳み込み
    - 攻撃に関わるリダイレクトの可能性に相当する値を出力
  - 起点のリダイレクト元 URL 毎に最高値の組を抽出 (プーリング)
    - 各 URL から最も攻撃に関わっている可能性が高い組を探索

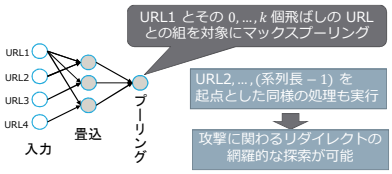


URL1 を起点とした URL 組の畳み込みとプーリング

19

Event-Denoising CNN

- 攻撃に関係しない URL の影響を削減するよう CNN を改良した **Event-Denoising CNN** (**EDCNN**) を提案
- EDCNN における畳み込み
  - 系列中の近くにある 2 つの URL の組を畳み込み
    - 攻撃に関わるリダイレクトの可能性に相当する値を出力
  - 起点のリダイレクト元 URL 毎に最高値の組を抽出 (プーリング)
    - 各 URL から最も攻撃に関わっている可能性が高い組を探索

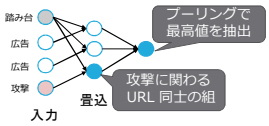


URL1 を起点とした URL 組の畳み込みとプーリング

20

Event-Denoising CNN

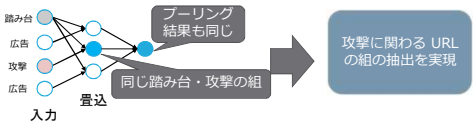
- 攻撃に関係しない URL の影響を削減するよう CNN を改良した **Event-Denoising CNN** (**EDCNN**) を提案
- EDCNN における畳み込み
  - 系列中の近くにある 2 つの URL の組を畳み込み
    - 攻撃に関わるリダイレクトの可能性に相当する値を出力
  - 起点のリダイレクト元 URL 毎に最高値の組を抽出 (プーリング)
    - 各 URL から最も攻撃に関わっている可能性が高い組を探索



21

Event-Denoising CNN

- 攻撃に関係しない URL の影響を削減するよう CNN を改良した **Event-Denoising CNN** (**EDCNN**) を提案
- EDCNN における畳み込み
  - 系列中の近くにある 2 つの URL の組を畳み込み
    - 攻撃に関わるリダイレクトの可能性に相当する値を出力
  - 起点のリダイレクト元 URL 毎に最高値の組を抽出 (プーリング)
    - 各 URL から最も攻撃に関わっている可能性が高い組を探索

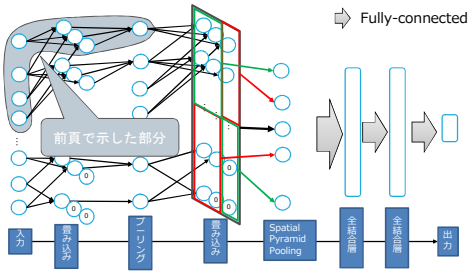


後半の URL の順序が入れ替わった場合

22

EDCNN の構造

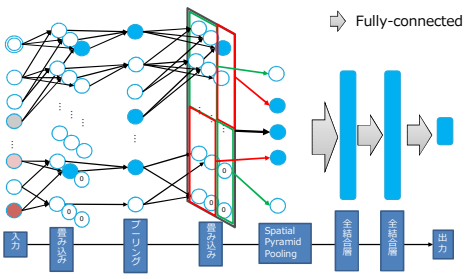
- 全体構成としては下図の通り
  - ○: N 次元ベクトル (URL の特徴量や畳み込み結果)



23

EDCNN の動作

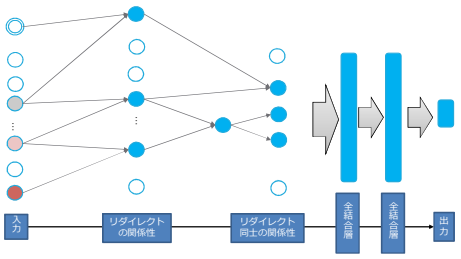
- ●: 攻撃に関する遷移の情報



24

EDCNN の動作

簡略化するとこのように攻撃に関わるリダイレクトの情報を選択して伝播していることになる



25

評価実験

- データセット
  - 収集元
    - 有名サイトのリストや公開ブラックリストに掲載されたウェブサイト
  - 収集期間
    - 2015 年 2 月～ 2016 年 1 月
  - 事前にコンテンツ解析を行いラベル付け (良性/悪性) 済
  - **収集元の各ウェブサイト一つ一つにアクセスしたときに発生する URL 系列のみを使用** (一度に複数サイトにアクセスしない)
- 実験内容
  - 以下に示す教師データでモデルを学習させ、テストデータを分類
  - テストデータ
    - 収集元から 1 か月間に渡り集めたデータ (重複データは削除)
  - 教師データ
    - テストデータの直近 6 か月間で集めたデータ (重複データは削除)

	教師データ			テストデータ		
	期間	悪性	良性	期間	悪性	良性
1	2015/2～7	1,039	5,905	2015/8	399	967
2	2015/3～8	1,182	5,905	2015/9	352	957
3	2015/4～9	1,126	5,892	2015/10	158	956
4	2015/5～10	2,785	5,895	2015/11	123	956
5	2015/6～11	2,515	5,880	2015/12	155	956
6	2015/7～12	1,906	5,891	2016/1	119	951

各データの収集期間と系列の個数

26

評価方法

- 評価指標
  - 識別性能を表す指標 Precision, Recall, F 値 (F-measure)

Precision= $\frac{TP}{TP+FP}$ (誤検知の少なさの指標)

Recall = $\frac{TP}{TP+FN}$ (見逃しの少なさの指標)

出力＼正解	悪性	良性
悪性	TP	FP
良性	FN	TN

テストデータの各正解ラベルに対しそれぞれの出力を行った数

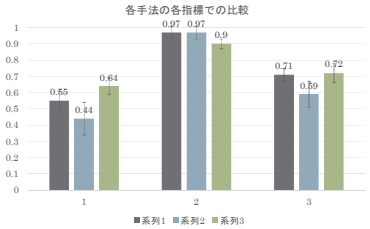
F-measure = $\frac{2\times Precision\times Recall}{Precision+Recall}$ (上記両方を評価する指標)

- 使用した学習・識別器
  - 系列中の URL を個別に分類する手法
    - 分類器は Random Forest (RF) を使用
    - URL 系列の中で悪性と判定した URL があれば系列全体を悪性と出力
  - CNN
  - EDCNN

27

識別性能

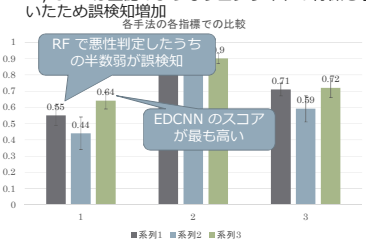
- 識別手法：RF (個別分類), CNN, EDCNN
- 評価指標：Precision, Recall, F 値



28

識別性能

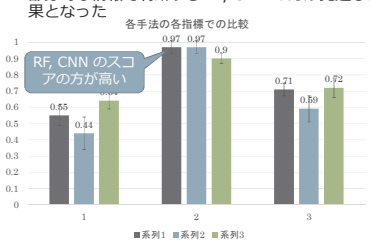
- Precision については EDCNN が最も高い
  - 攻撃に利用されやすい構造のウェブサイトが誤検知対象
  - RF, CNN は上記のようなウェブサイトの特徴も学習していたため誤検知増加



29

識別性能

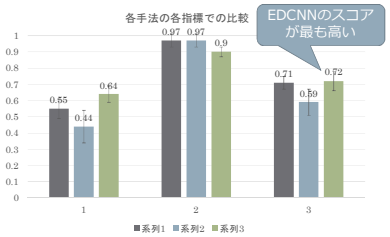
- Recall については EDCNN は比較的低い
  - 攻撃までの間に多量の広告があったウェブサイトを見逃し
  - 部分的な情報で判断する RF, CNN の方が見逃しにくい結果となった



30

識別性能

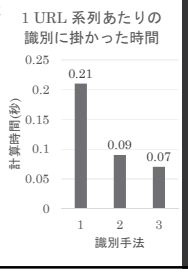
- F 値については EDCNN が最も高い
  - 総合的には URL 系列の分類には EDCNN が良いと言える
    - ただし Recall の低さを改善する工夫は必要



31

計算時間

- EDCNN, CNN, RF の順となった
- Random Forest による個別分類より EDCNN は平均 3 倍の速度での識別が可能
  - 個別に URL を識別していく手法は時間がかかる
- CNN より EDCNN の方が早い速度で識別することが可能
  - 畳み込み層に繋がるリンクの数が CNN より少ないためと考えられる
  - 計算時間の面でも EDCNN は CNN より性能向上できているといえる



32

まとめと今後の課題

- まとめ
  - プロキシログに含まれる宛先 URL の系列から，ドライブバイダウンロード攻撃に関する通信が含まれる URL 系列を検知する手法を検討
  - URL 系列の分類に CNN を適用
  - URL 系列の特性に合わせ CNN を拡張した EDCNN を提案
  - 1 年間に渡り収集したデータで EDCNN の識別性能を評価
- 今後の課題
  - **EDCNN の識別性能改善**
    - 見逃し・誤検知共に改善の余地あり
  - 同時に複数のウェブサイトへの通信が行われそれらの URL が混在した場合における識別の実施と評価
    - 本評価では単一のウェブサイトへのアクセス結果に対してのみ実施

33