L-015

ハニーポットによる攻撃ログの収集と機械学習による未知の攻撃の判別
The collection of attack logs using honeypots and the identification of unknown attacks through machine learning.

亀谷 広務† 八槇 博史†
Kametani Hiromu Yamaki Hirofumi

1. はじめに

近年サイバー空間での攻撃・探索活動は活発化しており，わが国でも能動的サイバー防御の導入に向けた議論が進むなど，先手を打ったセキュリティ対策が求められている．こうした中，企業を標的とした APT 攻撃も増加しており，企業単位で脅威解析を行い，攻撃トレンドに基づいた将来の攻撃予測を行う必要があると考えた．本研究では，Web サービスを標的とした攻撃に着目し，定点観測のログを機械学習した結果に基づいて，新たに受けた攻撃・探査活動が既存の攻撃トレンドに基づくものか，新規の攻撃パターンかを判別するモデルを作成した．

2. 先行研究

Web サービスへの攻撃の解析，すなわち HTTP リクエストのパスの特徴抽出について，Yu[1]らはリクエストパスやパラメータの長さ、大文字や小文字の出現頻度といった特徴量を抽出し，k-means による正常なアクセスのモデルを作成している．本研究では，パスを TF-IDF でベクトル化し，ハニーポットで収集した攻撃ログのクラスタリングを行った．

3. ハニーポット

サーバー上に設置し，サイバー空間で日々行われている攻撃・探査活動を待ち受けてログを収集するためのおとりのシステムである．ハニーポットは実際のサービスの挙動を模倣して待ち受けるため，攻撃者に本物のサービスと誤認させ，より踏み込んだ攻撃ログを収集できる．本実験では，複数のハニーポットを統合した T-Pot に組み込まれている，対 Web サービス攻撃に特化した tanner/snare を使用した．

4. T-Pot

種々の低対話型ハニーポットを統合し，Kibana を用いて集計できるシステムである．管理画面を図 1 に示す．

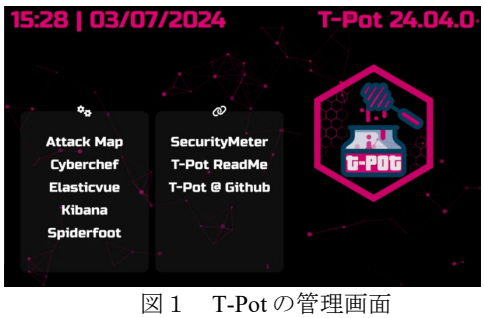

図 1　T-Pot の管理画面

† 東京電機大学システムデザイン工学研究科情報システム工学専攻　Tokyo Denki University Graduate School of System Design and Technology Design Engineering and Technology

5. tanner/snare

低対話型ハニーポットで，ポート 80 番で待ち受けを行い，対 Web サービスの攻撃・探査活動のログを収集する．snare が外部からのアクセスを待ち受け，tanner がログを集計する役割を持つ．

6. 実装

ハニーポットの構築にあたり，VPS サービスの Amazon Lightsail を使用し，OS として Debian 12.9 の環境を用意した．なおリージョンはカナダで実施している．ハニーポットの概要を図 2 に示す．

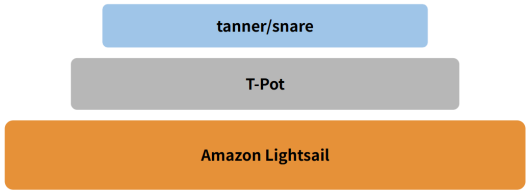

図 2　ハニーポットの概要

T-Pot の環境構築は公式リポジトリをクローンし，用意されているインストール用シェルスクリプトを実行してインストールを行った．ログの収集は 2024 年 7 月 4 日～2024 年 7 月 11 日までの 8 日間行い、図 3 のようなログが収集できた（IP アドレス部を加工）．

図 3　収集したログの一部

次に，収集したログの解析を行うプログラムを作成した．実装にあたり，収集されたログ情報（メソッド,宛先パス,ヘッダー,送信元 IP アドレス/ポート,ユーザーエージェント,送信先 IP アドレス/ポート,タイムスタンプ等）のうち，攻撃対象のサービスを把握可能な宛先パスの文字列のみを使用し，”/”のようにルートディレクトリのみを示すものを削除して機械学習を行った．学習は宛先パスを TF-IDF によりベクトル化し，k-means によるクラスタリングを行いモデルを作成した．攻撃トレンドの新規・既存の判別を行いたいパスについて，学習済みモデルの各クラスター中心とのコサイン類似度を計算し，閾値未満であれば「新規の攻撃」に，閾値以上であれば「既存の攻撃」に分類するとともに最も距離が近いクラスターに分類した．この処理を，新しい宛先パスのベクトルを x，クラスタ k の中心を μ k，閾値を θ （θ =0.5）として表したものを式(1)に示す．

$$\text{判定結果} = \begin{cases} \text{新規} & \dfrac{\boldsymbol{X}\cdot\boldsymbol{\mu}_k}{\|\boldsymbol{X}\|\,\|\boldsymbol{\mu}_k\|} < \theta, \\ \text{既存} & \dfrac{\boldsymbol{X}\cdot\boldsymbol{\mu}_k}{\|\boldsymbol{X}\|\,\|\boldsymbol{\mu}_k\|} \geq \theta. \end{cases} \quad (1)$$

7. 結果

収集したログのうち，その日以前のログがない 7 月 4 日を除く，28,126 件のログについて，作成したモデルを使用して新規攻撃に分類された件数，および割合を表 1 に示す．

表 1　集計結果

年月日	全数（件）	新規攻撃件数（全件数からの割合）
2024/7/5	6801	1321(19.42 %)
2024/7/6	75	25(33.33 %)
2024/7/7	541	267(49.35 %)
2024/7/8	75	17(22.67 %)
2024/7/9	6867	1441(20.98 %)
2024/7/10	6916	1419(20.52 %)
2024/7/11	6851	1400(20.43 %)

次に，全 34,916 件のログを TF-IDF でベクトル化し，2 次元に次元削減したものを可視化した結果を図 4 に示す

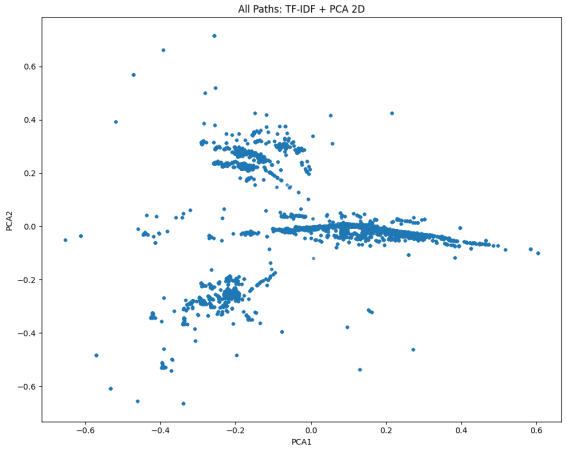

図 4　全パスの可視化

次に、新規攻撃に分類されたパスの傾向を調べた．2024 年 7 月 5 日と 2024 年 7 月 11 日について、パスの中に「.env」が含まれる新規攻撃を比較した結果を図 5 に示す．

図 5　2024/07/05 と 2024/07/11 の新規攻撃の比較

同じく環境変数を狙った攻撃であっても，7 月 5 日は Web アプリケーションフレームワークを狙ったものを中心に，7 月 11 日は Docker を狙ったものを中心に新規と判別された．学習済みの攻撃については既存と判別され，学習していない攻撃を新規として判別できていることが分かる．

次に，7 月 4 日以降で Remote Code Execution を狙った攻撃のパスと合計件数を表 2 に示す．

表 2　Remote Code Execution を狙った攻撃のパスと合計件数

パス	件数
/cgi-bin/luci/;stok=/locale?form=country&operation=write&country=$(id%3E%60for+proc_dir+in+%2Fproc%2F%5B0-9%5D%2A%3B+do+pid%3D%24%7Bproc_dir%23%23%2A%2F%7D%3B+buffer%3D%24%28cat+%22%2Fproc%2F%24pid%2Fmaps%22%29%3B+if+%5B+%22%24%7B%23buffer%7D%22+-gt+1+%5D%3B+then+if+%5B+%22%24%7Bbuffer%23%23%2A%22%2Flib%2F%22%7D%22+%3D+%22%24buffer%22+%5D+%26%26+%5B+%22%24%7Bbuffer%23%23%2A%22telnetdbot%22%7D%22+%3D+%22%24buffer%22+%5D%3B+then+kill+-9+%22%24pid%22%3B+fi%3B+fi%3B+done%60)	4
/cgi-bin/.%2e/.%2e/.%2e/.%2e/.%2e/.%2e/.%2e/.%2e/.%2e/.%2e/.%2e/bin/sh	3
/cgi-bin/%%32%65%%32%65/%%32%65%%32%65/%%32%65%%32%65/%%32%65%%32%65/%%32%65%%32%65/%%32%65%%32%65/%%32%65%%32%65/bin/sh	3
/index.php?lang=../../../../../../../usr/local/lib/php/pearcmd&+config-create+/&/<?echo(md5("hi"));?>+/tmp/index1.php	2
/index.php?lang=../../../../../../../../tmp/index1	2
/index.php?s=/index/\think\app/invokefunction&function=call_user_func_array&vars[0]=md5&vars[1][]=Hello	2

8. まとめ

8 日間の実験期間のうち、全数が極端に少ない 7 月 6 日～7 月 8 日を除くと，平均して 20.338 %が新規の攻撃に分類された．ログの件数が大幅に少ない日も存在し、攻撃・探査活動を行う Bot のクローリング活動にも波があることを示唆している．ログの収集期間を長くすることでログ件数の増減も含め、攻撃トレンドの正確な把握が可能になると考える．表 2 について，アクセスのみでサーバーが侵害されうる RCE 攻撃についても新規攻撃として抽出できているが，各日で新規攻撃として分類されていた．これは件数が少ないために k-means クラスタリングにおいて RCE 攻撃についてのクラスターが形成されなかったことが考えられる．

参考文献

[1] J. Yu, D. Tao and Z. Lin, "A hybrid web log based intrusion detection model," 2016 4th International Conference on Cloud Computing and Intelligence Systems (CCIS), Beijing, China, 2016, pp. 356-360, doi: 10.1109/CCIS.2016.7790283.