L-007

国別パケットのエントロピーの変化点検出に基づく異常検知

Anomaly detection based on detection of change points in entropy of packets by country

関 晃太郎 *　青木 茂樹 *　宮本 貴朗 *
Kotaro Seki　Shigeki Aoki　Takao Miyamoto

1　はじめに

近年, サイバー攻撃の増加に伴い, ネットワーク上の不正なトラフィックを検出する侵入検知システム (IDS) の研究が盛んにおこなわれている. 文献 [1] では, 正常な通信における国ごとのパケット数は一定の分布に従う傾向にあるが, 異常な通信が増加すると国毎のパケット数の分布が変化するという特徴を利用して, エントロピーを用いて単位時間毎の国単位でのパケット数の偏りを数値化することで, 異常な通信が増加した時間帯を特定する手法を提案している. この手法では, 時間帯の特定はできるが異常通信の内容を把握できないという点で課題がある. 本稿では, パケットデータからポート番号等の特徴を抽出し, 抽出した特徴の国毎の偏りをエントロピーで算出する. そして, エントロピーの時系列変化に対して ChangeFinder[2] により変化点検出することで異常を検出する. 本手法では, 国毎に分類することで, DoS 攻撃のようなパケット数に特徴のある攻撃を検出できる. また, ポート番号等の特徴を抽出しているため, 文献 [1] の手法で課題となっていた, 検出した異常の概要を把握することができる.

2　提案手法

本手法では, 組織のファイアウォールのインターネット側で収集したパケットデータからポート番号等の特徴を抽出し, 抽出した特徴の国毎の偏りに注目して異常を検出する. 以下では, 特徴としてポート番号に注目した場合の処理手順について述べる.

2.1　特徴量の抽出と国名の検索

まず, 収集したパケットデータからポート番号等の特徴を抽出する. ポート番号に注目する場合は, 送信元ポート番号と宛先ポート番号から値の小さい方を選択し, パケットデータを分類する. 次に, MaxMind 社が提供するパッケージである Geoip2 を使用して各 IP アドレスから送信元の国名を検索し, 最初のパケットからの経過時間とともに記録する. 22 番ポートに注目してパケット毎に国を調査して記録した例を表 1 に示す.

表 1. 22 番ポートのパケットの国の例

経過時間	送信元 IP アドレス	国名	選択ポート番号
0	11.11.111.11	A	22
1	22.22.22.22	B	22
…	…	…	…
600	33.33.33.33	C	22

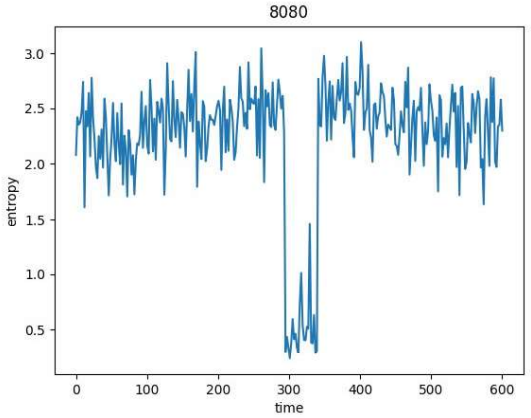

図 1. 8080 番ポートでのエントロピーの時系列変化の例

2.2　エントロピーの算出

抽出した特徴量毎に分類したパケットデータそれぞれで国毎のパケット数の偏りを表すエントロピーを式 (1) で計算する.

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{1}$$

ここで, n は単位時間内での通信相手国の総数であり, p_i は式 (2) で計算される単位時間内での国 i のパケットの割合である.

$$p_i = \frac{\text{単位時間内での国 } i \text{ のパケット数}}{\text{単位時間内での全パケット数}} \tag{2}$$

図 1 に, 8080 番ポートでのエントロピーの時系列変化の例を示す.

* 大阪公立大学大学院情報学研究科 Graduate School of Informatics, Osaka Metropolitan University

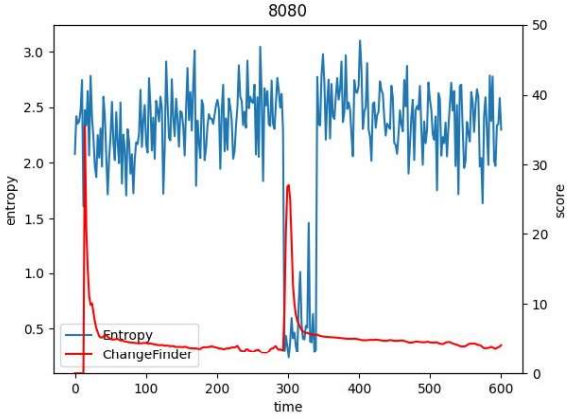

図 2. 8080 番ポートで変化点スコアを算出した例

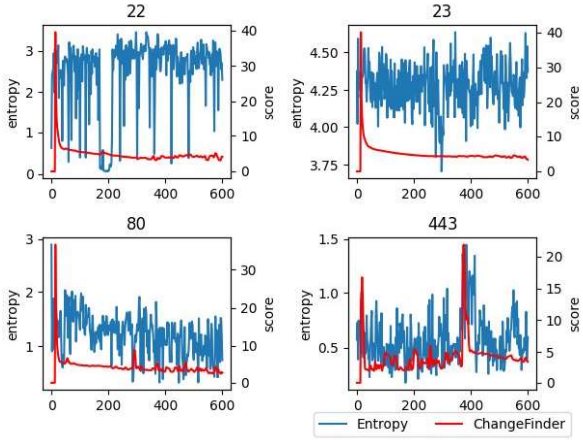

図 3. 4 つのポート番号での実験結果

表 2. 4 つのポート番号とそのパケット数

ポート番号	22	23	80	443
パケット数	239,820	149,277	508,442	3,103,937

2.3 変化点検出

算出したエントロピーの時系列データの変化点を, ChangeFinder を用いて検出する. ChangeFinder は SDAR モデルを使用して二段階学習し, 忘却パラメータ, モデルの次数, 平滑化の範囲の 3 つのパラメータ設定を行うことで, 時系列データの変化点をリアルタイムに検出する手法である. SDAR モデルは AR(自己回帰) モデルの学習に逐次学習と忘却機能を追加したモデルである. 逐次学習を行うことで, 時系列データの変化点をリアルタイムに検出できる. また, 忘却機能は過去のデータの影響を少なくすることができるため, 非定常なデータにも対応でき計算量を減らすことができる. ChangeFinder で変化点スコアを算出した例を図 2 に示す. 300 秒付近のエントロピーの値が減少している箇所で, 変化点スコアの大幅な増加を確認できる.

3 実験

3.1 実験環境

本手法の有効性を確認するために, 大阪府立大学のファイアウォールのインターネット側で取得したパケットデータを用いて実験を行った. 実験データは 2019 年 7 月 8 日 7 時 35 分から 7 時 45 分の 10 分間の, 約 1000 万パケットで構成されている. ここで, UTM(Unified Threat Management) で 7 時 41 分に TCP Flood 攻撃が検出されている. 実験条件として, エントロピーを算出するための単位時間を 1 秒に設定した. 表 2 にパケット数の多かった上位 4 つのポート番号とそれぞれのポート番号でのパケット数を示す.

3.2 実験結果

パケット数の多かった 22, 23, 80, 443 ポートのパケットを利用して異常検知を行った. 実験の結果を図 3 に示す. 実験結果のグラフから, 22, 23, 80 番ポートでは ChangeFinder による変化点スコアの変化は確認できなかったが, 443 番ポートで変化点の値が高くなった箇所を確認できた.

3.3 考察

図 3 に示す実験結果より, 443 番ポートの 400 秒付近で変化点スコアが高くなっていることを確認できた. このことから, 少数の国から https の大量のパケットを受信したことを把握できる. これは, ポート番号ごとに分類して変化点を検出したために把握できた攻撃の概要である. また, 今回の実験で異常を検出した時刻が, UTM で TCPflood 攻撃を検出した時刻と一致していることを確認した. この時刻でのパケットを詳細に調査した結果, ある国から正常時の約 3000 倍のパケットを受信していたことを確認した. このことから, https に対する DoS 攻撃を受けていたことが分かり, 本手法で把握した攻撃の概要が正しいことを確認できた.

まとめ

本稿では, エントロピーを用いて単位時間ごとに国単位でのパケット数の偏りを計算し, ChangeFinder を使って変化点を検出することで異常検知を行う手法を提案した. パケットデータからポート番号等の特徴を抽出しエントロピーを計算することで, 攻撃元の国の特定と攻撃の概要把握が可能であることを確認した. 今後の課題としては, DoS 攻撃以外の攻撃を検知する方法や検出できる攻撃の種類を増やすこと等が挙げられる.

4 参考文献

[1] Yuki Uemoto, Koji Okamura: Detection of increase in number of communication countries using entropy, Computer Security Symposium 2019, 3C1-1, 2019.

[2] J. Takeuchi and K. Yamanishi: A Unifying Framework for Detecting Outliers and Change Points from Time Series. IEEE Transactions on Knowledge and Data Engineering, Vol.18, No.4, pp.482–492(2006).