

ネットワークログと社会情勢の関係分析

沖 龍磨¹⁾, 鈴木 大雅¹⁾, 小野田 誠也¹⁾, 松下 慶大¹⁾, 梅原 孝仁^{1),2)}, 塩崎 雅基^{1),4)},
阿部 祐輔^{3),4)}, 永田 正樹⁴⁾

1) 静岡産業技術専門学校 みらい情報科

2) 静岡大学大学院 総合科学技術研究科

3) 株式会社アバンセシステム

4) 静岡大学 情報基盤センター

san-j21009@sist.ac.jp

Analysis of the Relationship Between Network Logs and Societal Trends

Oki Ryouma¹⁾, Suzuki Taiga¹⁾, Onoda Seiya¹⁾, Matsushita Keidai¹⁾, Umehara Takahito^{1),2)},
Shiozaki Masaki^{1),4)}, Abe Yusuke^{3),4)}, Nagata Masaki⁴⁾

1) Shizuoka Professional Training College of I.T.

2) Graduate School of Integrated Science and Technology, Shizuoka University

3) AvanceSystem Corporation

4) Center for Information Infrastructure, Shizuoka University

概要

サイバー攻撃の多様化とともに、市場のさまざまな事象がネットワーク通信に与える影響が注目されている。たとえば、紛争情報や大規模な経済的変動などの情報周知の前後でサイバー攻撃の発生頻度が高まることが指摘されている。本研究では、社会情勢とネットワーク通信情報の関係を分析するために、統合脅威管理機器（UTM）のログデータを活用し、k-means クラスタリングや主成分分析を用いて通信パターンの特徴を抽出する。特定の社会情勢がネットワーク通信にどのように影響するかを明らかにし、将来的なセキュリティ対策に向けた知見を模索する。

1 はじめに

デジタル化の進展とともにサイバー攻撃のリスクが急速に増大している。これに対処するために、従来の対策を超えたより包括的なセキュリティフレームワークや、AI 技術を活用した新しい防御手法などが注目されている。このような社会環境において、社会情勢がサイバー攻撃にどのような影響を与えるかという視点から見たとき、経済的・政治的な不安定さがサイバー攻撃の要因となりうることが指摘されている[1]。

この背景を基に、本研究では統合脅威管理機器（UTM）の通信ログ分析を通じて、社会情勢がどのように組織内ネットワークの通信内容に影響を与えるかについて、両者の関係性を分析する。そのために、サイバー攻撃の動向と関連する社会情勢の情報収集手法および、ネットワークログデータの分析手法を検討する。

2 先行研究

サイバー攻撃の兆候を「プレッシャー」という指標で捉え、社会的・経済的状況がサイバー攻撃に影響を与える可能性を示した調査では、現実世界の緊張の高まりが攻撃の引き金になることを指摘している[2]。

筆者らの先行研究では、UTM ログを用いて、社会情勢と通信の関係性を試みた。当該研究では、ウクライナ紛争などの国際的な緊張が高まる局面で、ネットワーク上の脅威ログや通信量が増加することが確認され、社会情勢がサイバー攻撃の動機形成に寄与する可能性が示唆された。また、サイバーキルチェーンの初期段階である「偵察」における通信パターンを把握することで、攻撃の予兆を捉え、早期に対応する必要性を指摘した[3]。

3 研究方法

社会情勢に大きな影響を与えた事象とネットワークログの動きから関係性を考察し、通信内容がどのような社会的、政治的、または経済的な要因に影響されるかを調査する。これらの調査結果を用いて、将来的なリスク管理や予防策の改善を図ることを目指す。本研究では、ネットワークログに UTM データを用いるが、UTM データは多種多様な情報を扱うため、本研究では必要箇所のみを抜粋して用いる。

3.1 クラスタリング概要

はじめに、クラスタリングを行うことで通信の特徴をつかむ。クラスタリングのためにログデータの中から以下のカラム情報を使用する。

- ・ サブタイプ
- ・ 送信元 IP アドレス
- ・ ルール名
- ・ アプリケーション
- ・ 送信元ポート番号
- ・ 宛先ポート番号
- ・ 送信元国

本研究では、2023 年 10 月 5 日から 10 月 8 日までの UTM ログデータを用いて、k-means クラスタリングおよび、主成分分析 (PCA) による通信パターンの分析を行う。k-means クラスタリングとは、まずデータを適当なクラスタに分けた後、平均を用いてうまくデータが分かれるように調整させていくアルゴリズムである。任意に指定の k 個のクラスタを作成するアルゴリズムであることからこのように呼ばれる。主成分分析 (PCA) は、もとの特徴量から新たな特徴量（主成分）を作り出し、もとの特徴量よりも少ない数の変数データを説明する手法である。主に多次元データの次元削除や可視化に使用される。

実装には Python の主要なライブラリである scikit-learn, Pandas, Plotly, Matplotlib を用い、非数値データの変換, 標準化, クラスタリング, 次元圧縮を行うことで可視化を実施した。ライブラリ／モジュールを表 1, 開発環境を表 2 に示す。

表 1 ライブラリ／モジュール

ライブラリ／モジュール	用途
Pandas	CSVファイルの読み込み
Scikit-learn	主要な機械学習モジュール
LabelEncoder	非数値データの数値変換
StandardScaler	特徴量の標準化
KMeans	クラスタリングアルゴリズムの適用
PCA	次元圧縮と累積分散説明率の計算
Matplotlib	エルボー法のプロットによるクラスタ数の最適化
PlotlyExpress	PCAによる次元削除後の結果をインタラクティブに可視化

表 2 開発環境

パーツ	内容
CPU	Intel Core i9-10980XE
GPU	NVIDIA GeForce RTX3090 x 4
RAM	128GB
STORAGE	SSD 1TB

3.2 実装内容

3.2.1 クラスタリング

クラスタリングでは、まずログデータを変数に格納し、使用するカラムのみを選択する。サブタイプ, 送信元 IP アドレス, ルール名, アプリケーション, 送信元ポート番号, 宛先ポート番号, 送信元国に対応するカラムのみを用いる。また、文字データとなっているカラムについて、LabelEncoder を用いて、配列の文字データを数値データに変換する。次に、StandardScaler を用いて、データの標準化を行う。次に、エルボー法を用いて最適なクラスタ数を選定する。今回はクラス数を 3 とした。0 から 2 までのクラス数に対して k-means を適応する。各クラスタ数の慣性をリストに保存し、k-means を用いてクラ

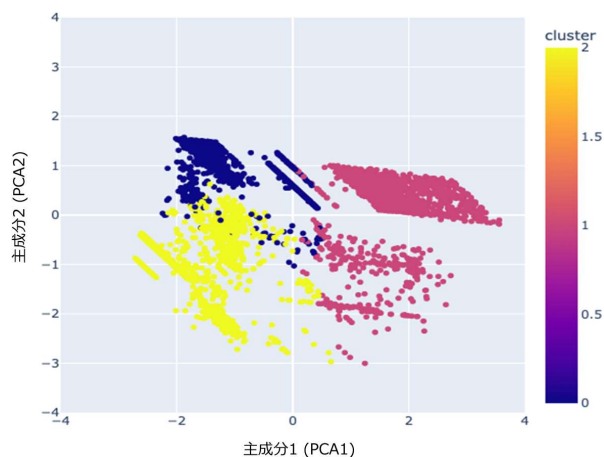


図1 2023年10月5日

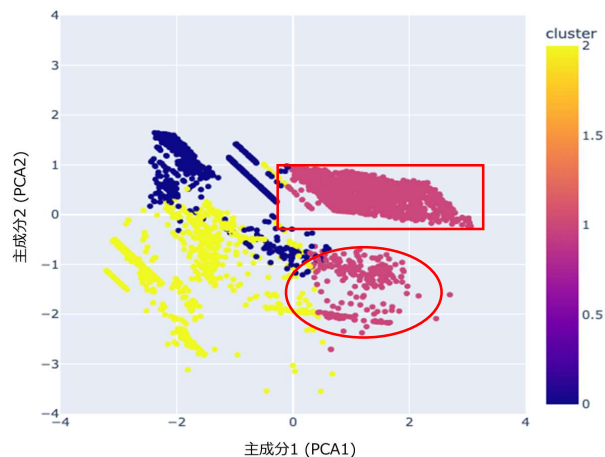


図3 2023年10月7日

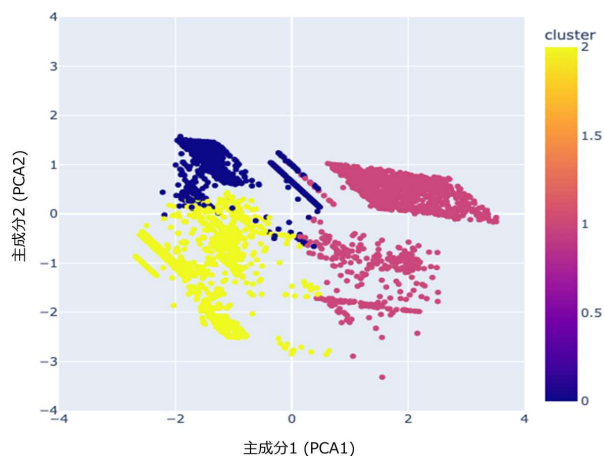


図2 2023年10月6日

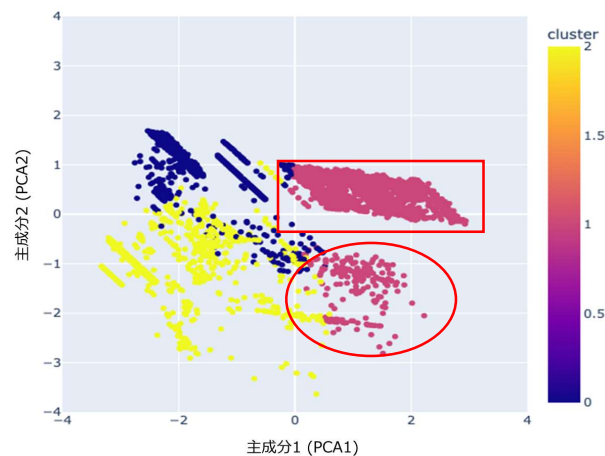


図4 2023年10月8日

スタリングを実行する。上記で求めた最適なクラスタの数だけ処理を繰り返す。

・3.2.2 次元圧縮

次元圧縮はPCAを用いて実施する。はじめに、Scikit-learnのPCAモジュールを用いて、元のデータセットの主要な分散を説明する主成分を計算し、データの次元を削除する。情報の損失を最小限に抑えつつ、次元を削除するために、データ全体の累積分散説明率を算出し、90%以上の分散を説明できる主成分を選定した。

・3.2.3 可視化

可視化は、Plotlyを用いてわかりやすい二次元散布図を作成し、クラスタ情報をマウスオーバーで詳細に表示する。各データポイントに関連する情報をホバー機能として実装した。Plotly Expressのpx.scatter関数を用いて、PCAで次元削除された2つの主成分を軸に、クラスタごとに色分けされたデータを視覚化した。

3.3 クラスタリングの結果

2023年10月5日から10月8日のクラスタリング結果を図1～4に示す。まず、 $x=0$ を境界として x がマイナスの場合は通信を許可、プラスの場合は通信が破棄、と別れている。正常に通信が終了した各クラスタの主要通信は、青色クラスタ0はポート53への通信、黄色クラスタ2はポート443への通信である。

破棄された通信である赤色クラスタ1はポート23への通信と一部不明ポートが含まれている。 y 軸方向に増加するほどポート番号の数値は小さい値となっている。軸赤色クラスタ1は $y=1$ を基準として、平行四辺形の塊と、散り散りになっている塊の2つで構成されている。そのうち散り散りになっている塊は様々なポートに向けた通信となっているが、平行四辺形の塊はポート23番に向けた通信である。10月7日の図3と10月8日の図4中の円および矩形箇所では、赤色クラスタ1の平行四辺形の塊の大

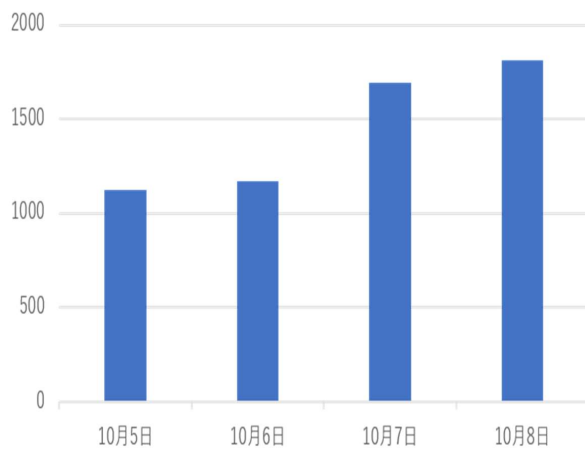


図5 ポート23番(Telnet)へ通信数
きさが、わずかではあるが増大している。

4 社会情勢との比較

2023年10月7日に発生したハマスによる大規模侵攻に関連し、ネットワークトラフィックの解析を行った結果、以下の知見を得た。特定の国からの通信が急増するといったことは確認できなかった。一方で、宛先ポート番号に関して、図5を見ると、侵攻日10月7日を境に大きな変化が見られた。ポート23番(Telnet)の通信が10月7日を以降に約1.5倍に増加している。

Telnetは暗号化されていないため、セキュリティ上のリスクが高いプロトコルであり、この増加は何らかの攻撃や不正アクセスに関連している可能性があると考えられる。さらに、図3および図4のクラスタリングの結果からも、ポート23番の通信が増加していることが明らかになった。この増加は単なる偶発的なものでなく、特定の時間帯や条件で集中して発生していることを示しており、ハマスによる侵攻後サイバー攻撃との関連性が疑われる。

5 おわりに

本研究では、市場に影響を与える経済的や社会的な社会情勢の変化がネットワーク通信内容に与える影響を、k-meansクラスタリングや主成分分析を用いて調査した。結果から、一例では

あるが、社会情勢がネットワーク通信に影響を与えた遠因を示唆できた。しかし、調査に用いたログデータや分析手法は改善の余地が多分にあり、今後の研究について課題を残している。今後は分析対象のデータ量の増加や多数要因の関係性を調査および分析していく。

参考文献

- [1] 情報セキュリティ白書 2024 (IPA), https://www.ipa.go.jp/publish/wp-security/eid2eo0000007gv4-att/2024_ALL.pdf (閲覧日:2024年10月18日)
- [2] 石井友基, 後藤厚宏. "プレッシャーによるサイバー攻撃兆候検知に向けた検討", 情報処理学会第79回全国大会講演論文集, pp.611-612, 2017年
- [3] 前田龍司, 川端純弥, 平原蒼生, 守屋海斗, 阿部祐輔, 塩崎雅基, 永田正樹. "ネットワークログと社会情勢を活用したサイバー攻撃の動向調査", 第86回全国大会講演論文集, pp.345-346, 2024年