6X-01

教師なし学習を用いたDDoS攻撃検知に関する一検討

樋場 叶耀 [†1]　小野 大地 [†1]　ギリエ ルイス [†2]　和泉 諭 [†3]　阿部 亨 [†1,†4]　菅沼 拓夫 [†1,†4]

[†1] 東北大学大学院情報科学研究科　[†2] 東北大学電気通信研究所
[†3] 仙台高等専門学校　[†4] 東北大学サイバーサイエンスセンター

1　はじめに

近年，Distributed Denial of Service (DDoS) 攻撃による被害が増加しており，検知手法が数多く提案されている．その中でも，機械学習を用いた教師あり学習による検知手法は高い検知精度が示されているが，訓練データに含まれない未知の攻撃に対しては検知精度が低くなる．教師あり学習では攻撃データを収集するために，実システムに対してDDoS攻撃を実行しなければならない．また，訓練データにはラベル付けされたデータが必要であり、膨大なトラフィックに対して分析を行い，ラベル付けを行う作業は非現実的である。

本研究では，訓練データに攻撃データを必要とせず，正解ラベルがないデータを用いて学習を行う教師なし学習に着目する．先行研究である [1] [2] は，正常データのみでモデル構築を行い，未知の攻撃に対しても検知可能なAutoEncoderによる検知モデルを提案している．AutoEncoderは，シンプルな構造による計算コストの削減を達成しているが，未知の攻撃に対する検知性能は不足している．

本稿では，DDoS攻撃検知に対するEfficient GAN [3] の有効性を検証する．Efficient GANはGenerative Adversarial Network (GAN) を用いた異常検知アルゴリズムであり，検知モデルにEfficient GANを採用したDDoS攻撃検知手法を提案し，未知の攻撃に対する検知性能の向上を図る．

2　関連研究

教師なし学習を用いたDDoS攻撃検知に関する研究としては文献 [1] がある．この研究では，AutoEncoderによる検知モデルを提案している．AutoEncoderは入力データと出力データが一致するように学習を行うNeural Networkであり，正解ラベルを必要としない教師なし学習アルゴリズムである．正常データで学習したAutoEncoderに攻撃データを入力した場合，攻撃データを再構成できない性質を利用し，攻撃の有無を判定する．入力データと出力データのL1ノルムを指標とする再構成誤差が予め設定した閾値を超える場合に攻撃と判定する．この手法は，学習の際に攻撃データを必要とせず，未知の攻撃に対しても検知が可能である．しかし，AutoEncoderはシンプルな構造による計算コストの削減を目的としており，検知性能としては不十分である．

他の検知手法として，文献 [2] ではVariational AutoEncoder (VAE) による検知モデルが提案されている．VAEは潜在変数が確率分布に従うように学習を行うAutoEncoderである．この研究では，評価用データセットにISPの大規模トラフィックを使用しており，VAEが学習に使用される訓練データに多少の攻撃データが含まれていても，検知性能に影響を与えないことが示されており，また，AutoEncoderよりも高い検知精度が示されている．しかし，評価に使用された攻撃はSYNフラッド攻撃のみであり，他のDDoS攻撃による評価が必要である．

3　提案手法

3.1　概要

本研究では，未知の攻撃に対する検知性能の向上を目的とし，検知モデルに文献 [3] のEfficient GANを採用したDDoS攻撃検知手法の検討を行う．GANは複雑な高次元データの分布のモデル化に優れた生成モデルであり，Efficient GANはGANを用いた異常検知アルゴリズムである．GANによるデータ分布のモデル化に着目し，ネットワークトラフィックの定常状態をモデル化することで，異常なトラフィックであるDDoS攻撃を検知する．

3.2　Efficient GAN

Efficient GANはGenerator，Discriminator，Encoderの3つのNeural Networkで構成されている．Efficient GANのアーキテクチャを図1に示す．

Generatorはデータの特徴を圧縮した低次元表現に相当するノイズを入力し，データを生成する．DiscriminatorはGeneratorが生成した偽物のデータと学習で使用する本物のデータの真偽を判定する．この2つのネットワークを交互に競合させ学習を行うことで，Generatorは本物に近いデータを生成できるようになる．Encoderは与えられたデータから特徴を抽出し，ノイズを生成する．Encoderはデータの推論に使用される．

正常データで学習させたEfficient GANは異常データの生成はできないという仮定に基づき，判定データと生成データから再構成誤差を算出し，再構成誤差が予め設定した閾値を超えた場合に異常と

A Study on DDoS Attack Detection Using Unsupervised Learning
Kiyoaki KAYABA[†1], Daichi ONO[†1], Luis GUILLEN[†2], Satoru IZUMI[†3], Toru ABE[†1,†4], and Takuo SUGANUMA[†1,†4]
[†1]Graduate School of Information Sciences, Tohoku University
[†2]Research Institute for Electrical Communication, Tohoku University
[†3]National Institute of Technology, Sendai College
[†4]Cyberscience Center, Tohoku University

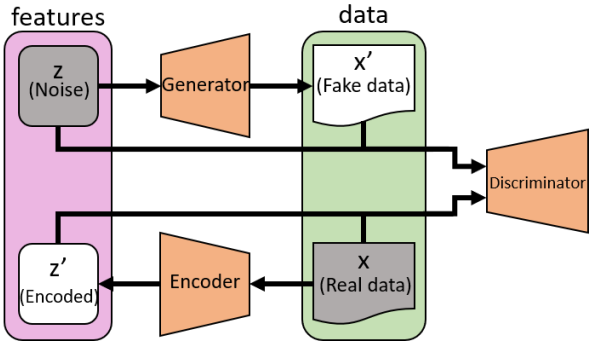

図 1: Efficient GAN のアーキテクチャ

判定する．再構成誤差は判定データと生成データの L1 ノルムである．

3.3 学習と攻撃検知の手順

Efficient GAN による学習と攻撃検知の手順を以下に示す．

(1) 正常データのみを使用し，Generator, Discriminator，Encoder を学習
(2) 判定するデータを Encoder に入力し，ノイズを生成
(3) (2) で生成されたノイズを Generator に入力し，データを生成
(4) (3) で生成されたデータと判定データから再構成誤差を算出
(5) 再構成誤差が予め設定した閾値を超える場合に攻撃と判定

4 予備実験

提案手法について，CICIDS2017 dataset [4] によるモデルの評価実験を行った．このデータセットは，実験用にネットワーク環境を構築し，バックグラウンドトラフィックから収集した正常データと複数の比較的新しい攻撃ツールによって収集した DDoS 攻撃データが含まれる．フローレベルで処理されたトラフィックは各フローの統計情報と正常，または攻撃のラベルが付与されている．

今回の実験では学習に正常データを使用し，Efficient GAN による検知モデルを構築する．推論では正常データと攻撃データを使用し，3.3 章で述べた再構成誤差の閾値を変化させることで，Precision-Recall (PR) 曲線と Receiver Operatorating Characteristic (ROC) 曲線を求める．加えて，PR 曲線では各閾値での Precision の平均である Average Precision (AP) を計算し，ROC 曲線では曲線の下の領域面積である Area Under the Curve (AUC) を計算する．また，ROC 曲線における破線は検知モデルがランダムな予測をしている場合を表す．実験結果を図 2 と図 3 に示す．

実験結果から AP は 0.79，AUC は 0.91 となった．AP と AUC は 1 に近いほど，攻撃データに対する検知性能が高いことを示し，Efficient GAN による攻撃検知が可能であることが確認できた．

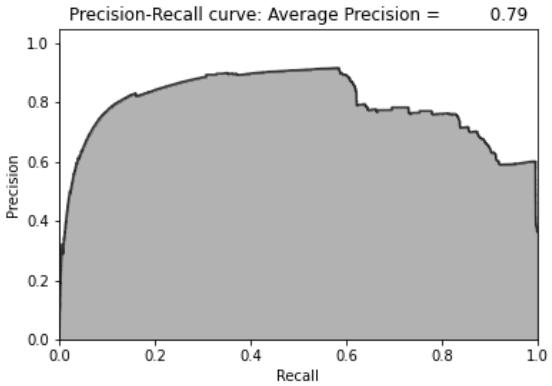

図 2: PR 曲線

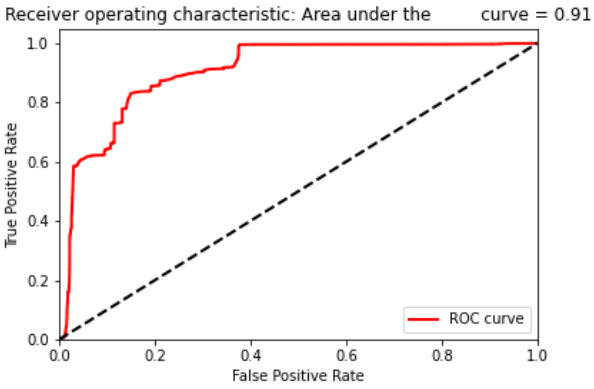

図 3: ROC 曲線

5 おわりに

本稿では，教師なし学習である Efficient GAN による DDoS 攻撃検知手法について検討を行った．予備実験により，Efficient GAN を用いた DDoS 攻撃検知が可能であることが確認できた．今後は，他の教師なし学習アルゴリズムとの比較，検知モデルに Efficient GAN を採用した DDoS 攻撃検知フレームワークの提案および実装を行う．

参考文献

[1] K. Yang et al.: “DDoS Attacks Detection with AutoEncoder,” NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium, pp. 1-9, 2020.
[2] Q. P. Nguyen et al.: “GEE: A Gradient-based Explainable Variational Autoencoder for Network Anomaly Detection,” 2019 IEEE Conference on Communications and Network Security (CNS), pp. 91-99, 2019.
[3] H. Zenati et al.: “Efficient GAN-Based Anomaly Detection,” arXiv:1802.06222v2 [cs.LG], 1 May 2019.
[4] I. Sharafaldin et al.: “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization,” Proc. 4th Int. Conf. Inf. Syst. Secur. Privacy (ICISSP), pp. 1-8, 2018