Vol.2021-DPS-186 No.61
Vol.2021-CSEC-92 No.61
2021/3/16

DCGANとパーティクルフィルタを用いたネットワークトラフィック異常検出

森岡 卓哉[1,a)]　青木 茂樹[1,b)]　宮本 貴朗[1]

概要：本研究では IDS の精度を高めるため，画像処理の分野で高い性能を示している深層学習を IDS に応用する手法を提案する．一般に，深層学習では教師データが必要となるが，深層学習手法を IDS に応用する場合，多種多様な異常データの収集が困難であることが課題となっている．そこで本研究では，深層学習の中でも教師データを必要としない DCGAN を使用して IDS を構築する．DCGAN は学習した画像と類似する画像は生成できるが，学習していない画像に類似する画像は生成できないため，この性質を利用してネットワークトラフィックの異常を検出する．まず，学習期間のトラフィックデータを画像に変換して DCGAN で学習する．次に，別の期間に収集したトラフィックデータを学習時と同様に画像に変換し，変換した画像に類似する画像を DCGAN で生成する．ここで，類似する画像の生成の際に，パーティクルフィルタを利用した最適解の探索アルゴリズムを適用する．そして，DCGAN が類似した画像を生成できるか否かによって，正常な通信と異常な通信を識別する．実験では，MWS データセットと CICIDS2017 データセットを用いて本手法の有効性を確認した．

Anomaly Detection of Network Traffic using DCGAN and Particle Filter

Takuya Morioka[1,a)]　Shigeki Aoki[1,b)]　Takao Miyamoto[1]

Abstract: In this paper, in order to improve detection accuracy of IDS, we propose a method to apply deep learning, which shows high performance in the field of image processing, to IDS. In general, deep learning method requires a large amount of teacher data. However, when applying deep learning methods to IDS, it is difficult to collect a wide variety of anomaly data. In our research, we construct IDS using DCGAN, which does not require teacher data. DCGAN can generate a fake image similar to the learned image, but cannot generate a fake image similar to the unlearned image. We use this property to detect anomaly in network traffic. First, we convert traffic data during learning period into images and learn the images by DCGAN. Second, we convert the traffic data collected in another period into images in the same way as during learning, and use DCGAN to generate an image similar to the converted image. Here, we use a particle filter to search for parameters for image generation by DCGAN. When the generated image and the converted image are similar, we recognize that the traffic data is similar to the learning period and is normal. Otherwise, we recognize that the traffic data is different status from the learning period and is anomaly. In order to verify the effectiveness of the proposed method, we performed an experiment using MWS2018 dataset and CICIDS2017 dataset.

1. はじめに

近年，サイバー攻撃等のネットワーク犯罪の増加に伴い，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:Intrusion Detection System）の研究が盛んに行われている．IDS はシグネチャ型とアノマリ型の 2 種

[1] 大阪府立大学大学院人間社会システム科学研究科
Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University
[a)] sza01286@edu.osakafu-u.ac.jp
[b)] aoki@kis.osakafu-u.ac.jp

類に分けることが出来る．代表的なシグネチャ型 IDS として，Snort[1] や Suricata[2]，TheBro[3] 等が挙げられる．シグネチャ型 IDS は異常を定義したパターンファイルに基づいて異常の検出を行う方式である．しかしパターンファイルに定義されていない攻撃については亜種を含め検出できないという欠点がある．一方，代表的なアノマリ型 IDS としては文献 [4,5,6] の手法が挙げられる．アノマリ型 IDS は正常な通信のみを含むデータを正常状態と定義し，そこから外れた状態を異常として検出する方式である．これにより，シグネチャ型 IDS の問題点であった未知の異常の検出が可能となる．しかし，シグネチャ型 IDS に比べ，誤検知が多く発生することが欠点として挙げられる．

本研究では，アノマリ型 IDS の精度向上のために，画像処理分野において高い識別性能を示している深層学習に注目する．一般に深層学習では，学習時に教師データが必要であるが，ネットワークの異常検出において正常と異常の両方のラベルが付いたデータを取得することは難しいため，一般的な深層学習手法を適用することは難しい．そこで深層学習の中でも教師データを必要としない敵対的生成ネットワーク (GAN:Generative Adversarial Network) に注目する．GAN は文献 [7] で提案された手法であり，Generator と Discriminator という 2 つのニューラルネットワークを用いて，学習した画像に類似する画像を精巧に生成できるモデルである．文献 [8] では，文献 [7] の GAN に対して畳み込みニューラルネットワーク（CNN:Convolutional Neural Network）を組み込むことで，更に精巧な画像を生成することが可能な DCGAN(Deep Convolutional Generative Adversarial Network) を提案している． 文献 [9] では，DCGAN を用いて医療用画像から患者が疾患を持つか否かを判別している．

文献 [10] では文献 [9] を応用し，DCGAN を用いてアノマリ型 IDS を構築している．DCGAN は学習した画像と類似する画像は生成できるが，学習していない画像に類似する画像は生成できないため，この性質を利用してネットワークトラフィックの異常を検出している．まず正常な通信のトラフィックデータを画像に変換して DCGAN で学習する．その後，テスト用のトラフィックデータを学習時と同様にテスト用画像へと変換し，変換した画像に類似する画像を DCGAN と最急降下法を組み合わせて生成する．そして，DCGAN で生成した画像とテスト用画像との類似度により異常度を算出して，異常度を基に正常な通信と異常な通信を識別する．

本研究では文献 [10] の異常検出における異常度の算出時に，モンテカルロ法を応用して状態空間の内部状態を推定するパーティクルフィルタを用いることで，検出精度の向上と安定化を目指す．また，様々なサイバー攻撃の検知に対応するために，文献 [10] とは異なる新たなパケットの画像化手法について検討する．実験では，MWS2018 データセットの BOS データセット [11] と CICIDS2017 データセット [12] を用いて本手法の有効性を確認した．

以下，2 節で関連研究，3 節で提案手法，4 節で実験と考察，5 節でまとめについて述べる．

2. 関連研究

2.1 ネットワークの異常検知に関する関連研究

本研究に関連する従来研究として，アノマリ型 IDS に関する文献 [4],[5],[6] の概要を説明する．文献 [4] では，パケットのエントロピーに基づく異常検出手法が提案されている．この手法ではまず，単位パケット数あたりの IP アドレスやポート番号などの出現回数を計測する．次に，出現回数を基に特徴量の出現確率を求め，求めた出現確率からエントロピーを算出する．その後，エントロピーの時系列変化に注目した EMMM 法により，エントロピーが大きく変化する場合を攻撃などが含まれている異常状態として検出している．

文献 [5] では，複数の特徴量の組み合わせによる異常検出手法を提案している．この手法では，異常をトラフィック量の異常，通信手順の異常，通信内容の異常の 3 種類に分け，単位時間あたりのトラフィック量を数値化した特徴量，フロー内のパケットのペイロードのパターンの傾向を数値化した特徴量を学習用データからそれぞれ抽出する．そして新たなデータでこれらの特徴量を抽出し，学習用データの値と閾値以上離れている特徴量が存在する場合に異常であると判断する．

文献 [6] では，ネットワークのトラフィックは複数の正常状態で表されると考え，複数の正常状態を定義し，各状態との違いから異常を検出する手法を提案している．この手法では，異常を含まないデータから単位時間当たりの ICMP や TCP パケット数等を計測してクラスタリングする．メンバが少ないクラスは削除し全てのクラスにおいて閾値以上のメンバ数となるまでクラスタリングを繰り返す．クラスタリング結果を正常状態として定義し，新たに観測されたデータから同様の特徴を抽出し，正常クラスとの距離が閾値以上かどうかで異常の判別を行っている．

2.2 敵対的生成ネットワーク（GAN）に関する関連研究

GAN に関する従来研究として，文献 [7],[8],[9],[10] について述べる．文献 [7] では，学習用データとなる画像を学習し，学習用データに精巧に類似する画像を生成できる敵対的生成ネットワーク (GAN:Generative Adversarial Network) を提案している．GAN は深層学習の中でも，教師データを必要としない手法である．GAN の構成の概要を図 1 に示す．GAN は Generator と Discriminator と呼ばれる 2 つのニューラルネットワークで構成される．Generator は学習用データに酷似したデータを生成するように学習する．一方，Discriminator は Generator が生成したデータと学

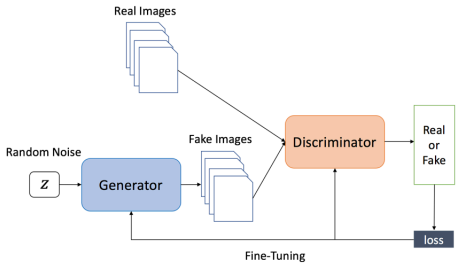

図 1 GAN の構成

習に用いたデータを正しく判別するように学習する．この仕組みにより，Generator と Discriminator は互いを高め合うように学習するため，Generator は学習用データに精巧に類似したデータを生成できる．しかし GAN は学習が不安定であることが問題となっていた．そこで文献 [8] では，GAN の 2 つのネットワークに畳み込みニューラルネットワークを組み込んだ DCGAN を提案している．これにより，GAN の学習を安定させ，より精巧な画像を生成している．文献 [9] では，文献 [8] の DCGAN を用いて医療用画像から患者が疾患を持つか否かを判別している．初めに疾患を持たない健康な患者の医療用画像を学習用データとして DCGAN で学習する．その後，疾患を持つか否か不明な患者の医療用画像をテスト用データとし，学習済みの Generator が生成する画像との誤差を求める．GAN はパラメータを適切に選択すれば学習した画像に類似する画像を生成できるが，学習した画像に類似しない画像はどの様なパラメータを選択しても生成できない．従って，テスト用データが学習用データと類似していれば誤差は小さくなり，学習用データと類似していなければ誤差は大きくなると考えられる．この性質を利用して，求めた誤差を基に異常度を算出することで疾患の有無を判別している．

文献 [10] では文献 [9] を応用し，DCGAN をアノマリ型 IDS として利用している．まず正常な通信のトラフィックデータを画像に変換して，DCGAN で学習する．その後，文献 [9] の手法を参考にして異常度を算出し，異常検出を行っている．文献 [9]，[10] では，パラメータの探索方法として最急降下法を利用していた．しかし，最急降下法でのパラメータ探索では必ずしも大域的な最適解が求められるとは限らず，局所最適解に陥る可能性が高い．このことから，最急降下法では，探索時の初期パラメータによっては検知できない異常が存在する可能性がある．

3. 提案手法

提案手法の概要を図 2 に示す．本手法では，パラメータの探索方法としてパーティクルフィルタを用いることで大域的な最適解を探索し，初期パラメータに依存せずに安定して異常検出する手法を提案する．更にトラフィックデータの画像への変換方法を多様なサイバー攻撃に対応させることによって異常検出精度を向上させる．本手法は学習と異常検出の 2 つのプロセスに分かれている．学習プロセスでは，学習期間のトラフィックデータを画像に変換し，学習用データとして DCGAN で学習する．異常検出プロセスでは正常通信と異常通信を含むトラフィックデータを学習プロセスと同様に画像に変換し，テスト用データとする．その後，学習済みの DCGAN とパーティクルフィルタを用いて正常通信と異常通信を識別する．

3.1 トラフィックデータの画像変換

あるネットワークで送受信されるパケットを pcap 形式でキャプチャしてトラフィックデータとする．収集したトラフィックデータを，検出対象としている攻撃に最適な画像化手法により画像に変換する．以下に 3 つの画像化手法について述べる．

(1) **パケットの到着順に 192 バイト以上のパケットのみを画像化**

パケットの先頭から 192 バイトを 8 ビット単位で 0 から 255 までの値として読み込む．そして変換した数値を 1 画素あたり RGB の 3 つの値に割り当てることで，1 × 64 画素の画像に変換する．これを 64 パケットに対して行い，64 × 64 画素の画像 1 枚を作成する．この手法では，ペイロードが空のパケットが除かれた画像が生成され，ペイロードの中身がより強調される．そのため，この手法ではマルウェアによる通信など，端末同士が送受信するパケットの内容に特徴がある攻撃の検出に適している．

(2) **宛先 IP アドレス毎にパケットを分類した後に，全サイズのパケットを対象に手法 1 と同様に画像化**

この手法では，ネットワーク内の各端末が外部ネットワークから受信するパケットのみで画像を生成するため，多数の攻撃者端末からの通信が多く見られる DDoS 攻撃などの検出を想定した画像化手法である．

(3) **送信元 IP アドレス毎にパケットを分類した後に，全サイズのパケットを対象に手法 1 と同様に画像化**

この手法では，ネットワーク内の各端末が外部ネットワークへ送受信するパケットのみで画像を生成する．そのため，web ページ上での不正な入力フォーム作成による入力内容の詐取など，被害端末が送信するパケットに重要な内容が現れるクロスサイトスクリプティングなどの検出を想定している．

トラフィックデータを画像に変換した例を図 3 に示す．図中の 1 行が 1 パケットを表しており，画像全体で 64 パケットが表されている．白色で表現されている部分はデータが含まれていない部分を表す．画像中の左端部分はパケットのヘッダ部分であり，それ以外の部分はペイロード部分を表している．このようにトラフィックデータを画像に変換することでトラフィックの特徴を可視化することができる．

Vol.2021-DPS-186 No.61
Vol.2021-CSEC-92 No.61
2021/3/16

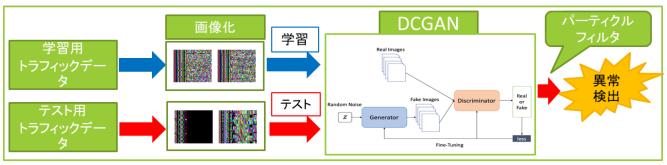

図 2 提案手法の概要

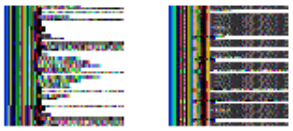

図 3 画像への変換例 (左：正常通信 右：異常を含む通信)

3.2 DCGAN による学習

学習期間のトラフィックデータを前節の手法により画像に変換し，学習用データとして DCGAN で学習する．DCGAN は GAN に対して畳み込みニューラルネットワークを組み込んだモデルである．GAN は Generator と Discriminator の 2 つの学習器から構成される．Generator は潜在変数 $\boldsymbol{z}$ を入力とし，学習用データに類似したデータを生成するように学習する．Discriminator は Generator によって生成されたデータと学習用データを正しく識別できるように学習する．学習は Discriminator が判別を行う際に発生する誤差を基にこれら 2 つのネットワークを更新していくことで行っていく．次式は学習の過程を数式で表したものである．G は Generator，D は Discriminator，$\boldsymbol{x}$ は学習用データ，$\boldsymbol{z}$ は潜在変数を表す．

$$\min_G \max_D V(D,G) = E_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[logD(\boldsymbol{x})] + E_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[log(1-D(G(\boldsymbol{z})))] \quad (1)$$

$G(\boldsymbol{z})$ は Generator によって潜在変数 $\boldsymbol{z}$ を入力として生成されたデータを，$D(\boldsymbol{x})$ はデータ $\boldsymbol{x}$ が訓練データである確率を表す．Discriminator は $D(\boldsymbol{x})$ を最大化しようと，Generator は $log(1-D(G(\boldsymbol{z})))$ を最小化しようとする．つまり Generator は Discriminator に訓練データと誤って判別されるような精巧なデータを生成することを目的に，Discriminator は Generator が生成したデータを完全に判別することを目的として学習する．このように学習を行うことで，2 つのネットワークは互いに高め合い，Generator は訓練データに精巧に類似したデータを生成することが出来るようになる．DCGAN では 2 つのネットワークに畳み込みニューラルネットワーク (CNN) を適用している．CNN は通常のニューラルネットワークに畳み込み層とプーリング層を追加することで識別精度の向上を図っている．GAN においても CNN を適用することで，従来より高精度に画像を生成することが出来る．

3.3 異常検出

3.2 節で学習した DCGAN を用いて異常を検出する．正常なデータのみを用いて DCGAN で学習すると，学習後の Generator は正常なデータの分布 p に従って画像を生成する．したがって，学習後の DCGAN にテスト用画像 $\boldsymbol{x}$ を入力した際に，$\boldsymbol{x}$ が正常なデータの画像であれば p に従うため，潜在空間内に $G(\boldsymbol{z}) \approx \boldsymbol{x}$ となる潜在変数 $\boldsymbol{z}$ が存在すると考えられる．一方，$G(\boldsymbol{z}) \approx \boldsymbol{z}$ となる $\boldsymbol{z}$ が存在しない場合は $\boldsymbol{x}$ は学習用データに含まれない異常なデータの画像であると判断できる．しかし Generator では，$\boldsymbol{x}$ から $\boldsymbol{z}$ を求めることが出来ない．そこで文献 [10] では，乱数を用いて $\boldsymbol{z}$ の初期値をサンプリングした後に最急降下法で最適解を探索することで，$\boldsymbol{x}$ に対応する $\boldsymbol{z}$ を 1 つだけ算出している．しかしこの手法では，サンプリングした初期値や最適化回数によっては，$\boldsymbol{x}$ が正常なデータの場合でも，$G(\boldsymbol{z}) \approx \boldsymbol{x}$ となる $\boldsymbol{z}$ を求められない場合がある．そこで本手法では，状態推定モデルであるパーティクルフィルタを用いて $\boldsymbol{z}$ の最適解を探索する．

パーティクルフィルタとは，モンテカルロ法を応用して，ある状態空間（ここでは潜在変数 $\boldsymbol{z}$ の分布を表す潜在空間 $\boldsymbol{z}$）の分布を推定するアルゴリズムである．分布が正規分布などの様に単純ではない場合に有効である．このアルゴリズムでは，まず空間内にランダムなパーティクル（粒子）を散布し，各パーティクルの尤度を計算する．そして，尤度が高いパーティクルの近くで新しいパーティクルを散布する手順を繰り返すことにより，空間の内部状態を推定する．本手法では $\boldsymbol{z}$ のパーティクルを多数生成して最適解を探索する．パーティクルフィルタを用いることによって，初期値などのパラメータの影響を受けずに，最適な $\boldsymbol{z}$ を安定して求めることが出来ると考えられる．以下，パーティクルフィルタを用いた異常検出の具体的な手順を示す．

(1) $[-1,1]$ の一様乱数から潜在変数 $\boldsymbol{z}$ を生成する．ここでは，生成する $\boldsymbol{z}$ の数を 1500 個としている．

(2) テスト用画像 $\boldsymbol{x}$ と DCGAN による生成画像 $G(\boldsymbol{z})$ の異常度 $A(\boldsymbol{x})$ を次式を用いて算出する．

$$A(\boldsymbol{x}) = \sum |\boldsymbol{z} - G(\boldsymbol{z}_n)| \quad (2)$$

(3) $A(\boldsymbol{x})$ が小さい順に $\boldsymbol{z}$ を並べ替える．ここで，テスト用画像 $\boldsymbol{x}$ と生成画像 $G(\boldsymbol{z}_n)$ が同一の時に $A(\boldsymbol{x})$ は 0 となり，$\boldsymbol{x}$ と $G(\boldsymbol{z}_n)$ が異なる場合に $A(\boldsymbol{x})$ は大きな値となる．

(4) 上位 500 個の $G(\boldsymbol{z})$ について，$\boldsymbol{z}$ を微小に変化させた変数 1000 個を $\boldsymbol{z}_{new}$ とする．$\boldsymbol{z}_{new}$ を生成する際には $A(\boldsymbol{x})$ が小さい $\boldsymbol{z}$ の近傍で，多数の新しい変数を発生させる．また，新たに $[-1,1]$ の一様乱数から生成した潜在変数 500 個を $\boldsymbol{z}_{new2}$ とする．

(5) $\boldsymbol{z} = \boldsymbol{z}_{new} + \boldsymbol{z}_{new2}$ として手順 2 へ戻る．

以上の手順を 20 回繰り返したところで探索を終了し，そ

表 1 BOS データセットの詳細

データセット	日付	進行度	正常	異常	合計
学習用データ	2017/8/17	2	50,000	0	50,000
テスト用データ	2018/1/23	8	2,409	233	2,642

の際の異常度を $A(\boldsymbol{x}_{20})$ とする．そして，$A(\boldsymbol{x}_{20})$ が閾値以上ならば異常，閾値未満ならば正常とする．

4. 実験

4.1 実験条件

実験には MWS2018 の BOS データセット [11] と CICIDS2017 データセット [12] を使用した．各データセットの詳細については次節以降で説明する．BOS データセットを用いた実験では，3.1 節の手法 1 のみで画像化を行い，文献 [10] の手法と比較することで，パーティクルフィルタを適用することの有効性を確認した．また CICIDS2017 データセットを用いた実験では，サイバー攻撃の種類毎に最適な画像化手法を確認した．

テスト用データに対する正常と異常のラベル付けでは，C2 サーバ（Command & Control サーバ）または攻撃者端末と通信を行っているパケットが 1 枚の画像中に 15 パケット以上存在している場合に異常のラベル，そうでなければ正常のラベルを付与することとした．また学習の際，DCGAN の学習回数は 20 回，バッチサイズは 64 とした．

4.2 実験データセット

4.2.1 BOS データセット

MWS2018 の BOS データセット [11] は，標的型攻撃のマルウェアをローカルネットワーク内の端末に感染させ，その通信をキャプチャした研究用データセットである．BOS データセットはマルウェアとの通信の進行度によって 1 から 8 まで定義されている．進行度 1，2 のデータはマルウェアを実行したが C2 サーバとの通信は発生しておらず，進行度 3，4，5 のデータは C2 サーバとの通信は発生したが C2 サーバとの通信は成立していない状態を示している．進行度 6，7，8 のデータは通信が発生し，C2 サーバとの通信も成立している状態である．実験では，2017/8/17 に収集された進行度 2 のトラフィックデータを異常が含まれていない正常なトラフィックデータとして学習に用いた．また，2018/1/23 に収集された進行度 8 のトラフィックデータを異常を含むトラフィックデータとしてテストに用いた．本実験で用いた学習用データの画像は 50000 枚，テスト用データの画像は 2642 枚である．表 1 に，ラベル付けした学習用データとテスト用データの画像枚数の詳細を示す．

4.2.2 CICIDS2017 データセット

CICIDS2017 データセット [12] は，攻撃者端末から組織内ネットワークへのサイバー攻撃を想定した研究用データセットである．含まれている攻撃は，ブルートフォースアタック，DoS，DDoS，クロスサイトスクリプティング，ボットネット，ポートスキャンである．2017/7/3 から 2017/7/7 までのトラフィックをキャプチャしており，7/3 のトラフィックは全て正常なものである．それ以外の日の通信は攻撃通信と正常通信の両方を含むトラフィックである．

表 2 手法 1，[10] の手法を用いた時の CICIDS2017 データセット画像の枚数

データセット	検知対象	正常	異常	合計
学習用データ	-	80316	-	80316
テスト用データ	DoS	31	30	62
	ブルートフォース	35	41	76
	クロスサイトスクリプティング	38	27	65
	DDoS	31	30	61
	Portscan	31	4	35

表 3 手法 2 を用いた時の CICIDS2017 データセット画像の枚数

データセット	検知対象	正常	異常	合計
学習用データ	-	97096	-	97096
テスト用データ	DoS	36	41	77
	ブルートフォース	33	30	63
	クロスサイトスクリプティング	33	31	64
	DDoS	20	25	45
	Portscan	20	20	40

表 4 手法 3 を用いた時の CICIDS2017 データセット画像の枚数

データセット	検知対象	正常	異常	合計
学習用データ	-	84176	-	84176
テスト用データ	DoS	31	30	61
	ブルートフォース	33	30	63
	クロスサイトスクリプティング	38	27	65
	DDoS	30	25	55
	ポートスキャン	30	30	60

学習用データとして 7/3 のトラフィックデータを使用した．テスト用データとしては，それ以外の日のトラフィックデータを使用した．テスト用データの中には正常通信に加えて，攻撃通信として 5 種類の攻撃（DoS，DDoS，ポートスキャン，ブルートフォースアタック，クロスサイトスクリプティング）が発生した時の通信が含まれている．3.1 節で述べた各手法で画像化して異常検出した場合と文献 [10] の手法を適用した場合の 4 パターンで，これらの攻撃を検知する実験を行った．表 2，表 3，表 4 に各手法毎に正常および異常のラベル付けをした学習用データとテスト用データの画像枚数の詳細をそれぞれ示す．

4.3 実験結果と考察

4.3.1 BOS データセットを用いた実験結果

図 4 に本手法と既存手法である文献 [10] の手法の ROC 曲線および AUC 値を示す．本手法の AUC 値は 0.71，文献 [10] の手法の AUC 値は 0.63 となり，本手法の方が高精度な結果が得られた．これは異常検出における異常度の算出時にパーティクルフィルタを用いることにより，入力したテスト用画像 $\boldsymbol{x}$ に対応する潜在変数 $\boldsymbol{z}$ をより正確に求めることが可能となり，識別精度が向上したためであると考えられる．一方，識別できなかった要因として，テスト用

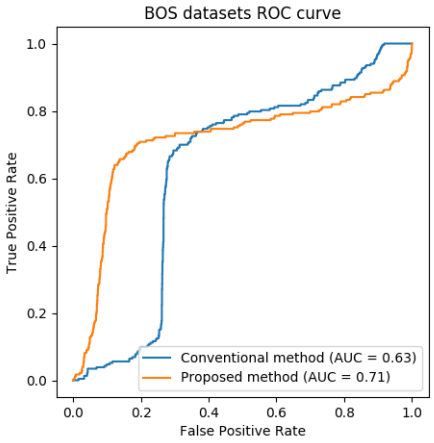

図 4 BOS データセットを用いた時の実験結果

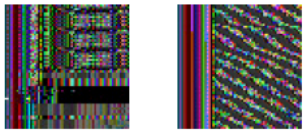

図 5 生成画像の比較 (左:学習用データ　右:誤識別したテスト用データの正常画像)

データの正常画像の中に学習用データの画像と類似していないものが存在することが挙げられる．図 5 に学習用データと誤識別したテスト用データの正常画像の例を示す．これは学習用データを増加させ，DCGAN が生成できる画像の種類を増やすことで改善できると考えられる．

4.3.2 CICIDS2017 データセットを用いた実験結果

CICIDS2017 データセットを用いた実験を ROC 曲線および AUC 値によって評価した結果を図 6，図 7，図 8，図 9，図 10 に示す．図 6 は DoS 攻撃の検知についての実験結果である．DoS 攻撃の検知については，手法 3 での AUC 値が 0.90 と一番良い結果になった．手法 3 での生成画像を図 11 に示す．図 11 の左側の正常画像と右側の異常画像を比較すると，異常画像ではどのパケットにおいてもペイロードの中身が同じものになっているために画像に縦線が現れていることを確認できる．また，異常なペイロードを持つパケットが等間隔で見られるなど，ペイロードの空白部分が明確に異なっている．このことにより，他の手法よりも検知結果がよくなったと考えられる．

図 7，図 8 はそれぞれブルートフォースアタック，クロスサイトスクリプティングの検知についての実験結果である．ブルートフォースアタック，クロスサイトスクリプティングの検知では，手法 2 での AUC 値が 0.94，0.96 と一番良い結果になった．図 12 に手法 2 での正常トラフィック画像とブルートフォースアタック発生時のトラフィック画像を示す．図 12 より，DoS 攻撃検知の際と同じく，ペイロードの中身や空白部分が正常画像と異常画像で大きく異なるために検知精度が上昇したと考えられる．ブルートフォースアタックについては攻撃者の端末からローカルネットワーク内の端末へ大量に類似するパケットを送ることから，手法 2 で作成された異常画像が正常通信時と大きく異なったと考えられる．

図 9 は，ポートスキャンの検知についての実験結果である．ポートスキャン検知においては，図 9 より，手法 1 と文献 [10] の手法では AUC 値がそれぞれ 0.80，0.62 となっている．しかし，手法 1 ではポートスキャンの異常画像が 4 枚しか生成できず，検知精度を確認するには不十分であった．また，それ以外の手法では検知に失敗している．この理由としては，手法 2 と手法 3 で画像化した際，ポートスキャン時に観測されるパケットと正常通信時のパケットの中に類似しているものが存在することがあげられる．図 13 に手法 2 によるポートスキャン時のトラフィック画像と学習用データのトラフィック画像を示す．このように，ペイロード部分の空白が多い画像は正常通信においても存在するため，ポートスキャン画像の異常値が小さくなっている．ポートスキャン時はローカルネットワーク内に短時間で大量のパケットが送られるという特徴があるため，検知精度を改善するためにはトラフィックを 64 パケット毎に画像化するのではなく，単位時間毎に画像化する手法を検討することが必要であると考えられる．

図 10 は，DDoS 攻撃検知についての実験結果である．手法 3 での AUC 値が 0.71 となり，一番精度が高くなった．手法 3 に比べて，手法 2 での実験ではポートスキャンの時と同様に，異常画像に空白が多くなったため，画像の異常値が低くなったと考えられる．また，他の攻撃と比べて全手法で検知精度が下がった要因の 1 つとして，テスト用データの正常画像に学習用データと異なるものが多かったことが挙げられる．これについては学習用データの追加や DCGAN のパラメータ調整によって，DCGAN が生成できる画像の種類を増やすことで改善できると考えられる．

4.3.1 節での BOS データセットを用いた実験結果と，本節での実験の結果から，DCGAN とパーティクルフィルタを用いた異常検知により，従来手法よりも異常検知精度が向上することを確認できた．また，複数の画像化手法の比較により，それぞれの攻撃の種類に対応した最適な画像化手法が存在することも確認できた．一方で，ポートスキャンのようにうまく検知できない攻撃も存在した．今後の課題としては，トラフィックデータの画像化方法についての更なる検討，各攻撃に対応した画像化手法を組み合わせて総合的に異常を検知する方法の検討が挙げられる．

5. おわりに

本論文では，深層学習手法の 1 つである DCGAN と状態推定モデルであるパーティクルフィルタを用いて，ネットワークの異常を検知する手法を提案した．実験では 2 種類

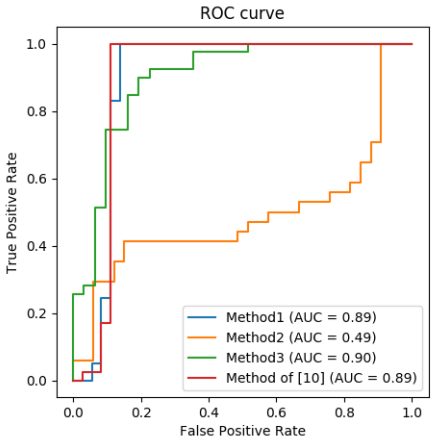

図 6 DoS 攻撃検知についての実験結果

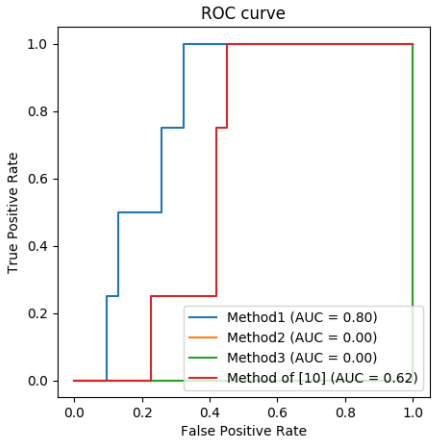

図 9 ポートスキャン攻撃検知についての実験結果

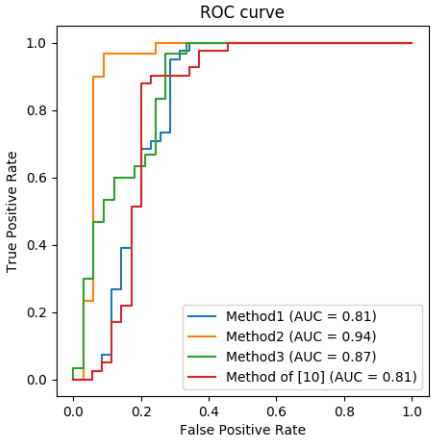

図 7 ブルートフォース攻撃検知についての実験結果

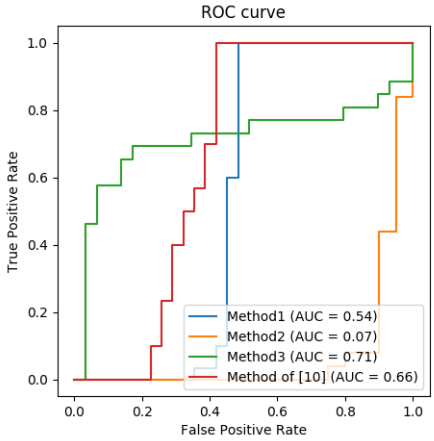

図 10 DDoS 攻撃検知についての実験結果

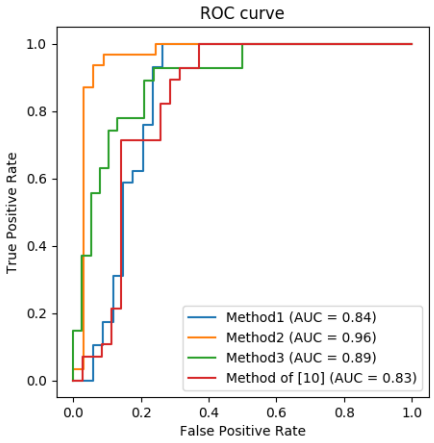

図 8 クロスサイトスクリプティング検知についての実験結果

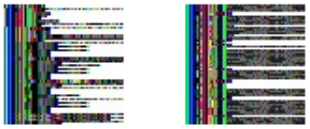

図 11 手法 3 による生成画像の比較 (左:正常画像　右:DoS 攻撃発生時の画像)

図 12 手法 2 による生成画像の比較 (左:正常画像　右:ブルートフォースアタック発生時の画像)

のデータセットを用いて本手法の有効性を確認した．実験結果より，本手法が従来手法よりも高い検知精度となり，有効性を確認することができた．また，異常の種類に応じた最適な画像化手法を用いることで，検知精度を更に向上できることも確認できた．今後の課題としては，DCGAN のパラメータ調整やトラフィックデータの画像変換方法の更なる検討などが挙げられる．

参考文献

[1] Snort, <https://www.snort.org/>(参照 2021-02-04) .

図 13 手法 2 による生成画像の比較 (左:学習用データの正常画像 右:ポートスキャン発生時の画像)

[2] Suricata, <http://suricata-ids.org/>(参照 2021-02-04) .
[3] The Bro, <http://www.bro.org/>(参照 2021-02-04) .
[4] 小島俊輔，中嶋卓雄，末吉敏則：エントロピーベースのマハｒラノビス距離による高速な異常検知手法，情報処理学会論文誌 ， Vol.52, No.2, pp.656-668(2011).
[5] 佐藤陽平，和泉勇治，根元義章：複数の検出モジュールの組み合わせによるネットワーク異常検出の高精度化，信学技報， NS2004-144, pp.45-48(2004).
[6] 平松尚利, 和泉勇治, 角田裕：複数の通常状態を用いたネットワーク異常検出, 信学技報, CS2006-32, pp.61-66(2006).
[7] Lan Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio : Generative adversarial nets. In Advances in Neural Information Processing Systems, pp 2672-2680(2014).
[8] Alec Radford, Luke Metz, and Soumith Chintala : Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv preprint arXiv:1511.06434(2015).
[9] Thomas Schlegl, Philipp Seebock, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs :Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In International Conference on Information Processing in Medical Imaging, pp146-157(2017).
[10] 日置裕士，青木茂樹，宮本貴朗：DCGAN を用いたネットワークトラフィックの異常検出，Computer Security Symposium 2018 講演論文集，2C2-1，pp.341-347(2018).
[11] 高田雄太，寺田真敏，松木隆宏，笠間貴弘，荒木粧子，畑田充弘：マルウェア対策のための研究用データセット〜MWS Datasets 2018〜，情報処理学会論文誌，Vol.2018-CSEC-82，No.38，pp.1-8(2018).
[12] I. Sharafaldin, A. Habibi Lashkari and A. A. Ghorbani: Toward generating a new intrusion detection dataset and intrusion traffic characterization, *Proc. 4th ICISSP*, pp. 108-116(2018).