Vol.2019-IOT-44 No.52
2019/3/8

侵入検知システムのための グラフ構造に基づいた機械学習および可視化

熊谷　将也[1,a)]　松本　亮介[1]

概要：サイバー攻撃の高度化・多様化に伴い，機械学習手法を用いた侵入検知システム (IDS: Intrusion Detection System) の研究が盛んに進められている．これまでにも k 近傍法や One-Class サポートベクターマシン，ニューラルネットワークなど様々な機械学習手法の適用がすでに提案されており，いずれも高い検知率が得られている．ただし，IDS の実運用を考慮する場合には，最終的に判断する管理者が検知結果の検証に多くの時間を費やすことを避けるため，高い検知率だけでなく検知結果に対する説明性を有した手法が求められる．本研究では，複数の要素とそれらの関係性を直感的に理解できるグラフ構造でトラフィックを表現することに着目し，ノードや部分グラフの可視化によって検知結果を説明する 2 種類の機械学習手法を用いた IDS を提案する．1 つ目は，正常と定義したグラフ構造からの崩れを異常として検知する手法であり，各ノードの異常度を示すことができる．2 つ目は，正常/異常のグラフ構造をグラフ畳み込みニューラルネットワークで学習し，その学習モデルによって異常を検知する手法であり，異常の要因となる部分グラフを示すことができる．

Graph Based Machine Learning and Visualization for Intrusion Detection System

Kumagai Masaya[1,a)]　Matsumoto Ryosuke[1]

Abstract: Along with the sophistication and diversification of cyber attacks, intrusion detection systems (IDS) using machine learning have been actively studied. Many machine learning methods (such as the k-nearest neighbor method, One-Class support vector machine, neural network, etc.) have been applied to IDS already, and many cases achieved high accuracy. However, even though highly accurate anomaly detection can be achieved, false detection occurs. Consequently, administrators in actual operations would spend much post-processing time to validate the detection results. Therefore, anomaly detection should explain which factors of input causes its anomaly. We focused on graph structuring of traffic data and proposed two types of IDS using machine learning which can explain the reason of anomaly by visualization of node or subgraph. The first method can detect anomalies from a difference of the graph structure between normal and abnormal, and obtain the degree of an anomaly for each node. The second one can learn the graph structure labeled of normal or abnormal using a graph convolution neural network, and show a subgraph as a cause of anomaly.

1. はじめに

スマホや IoT デバイスの急速な普及により，インターネットはますます身近な存在となっている [1]．ところが，サイバー攻撃の脅威も年々急速に増加していることから，その脅威も私達にとって身近な問題となってきている [2]．そのため，サイバー攻撃が発生した際に直ちに検知することができる侵入検知システム (IDS: Intrusion Detection System) の必要性が高まっている．

特に最近では，機械学習手法を適用した IDS の研究が盛んに行われている．その機械学習手法の学習アルゴリズムは，教師あり学習と教師なし学習の 2 種類に大別される．教師あり学習は，正解ラベル（IDS の場合は攻撃か正常，または攻撃の種類など）が付与されたデータを学習するため，

[1] さくらインターネット株式会社　さくらインターネット研究所
[a)] m-kumagai@sakura.ad.jp

すでに学習した攻撃またはそれに類似した攻撃に対して高い精度で検知ができる．一方，教師なし学習は，正常な状態を仮定して，その正常からどの程度ずれているかを定量的な数値を算出することで異常な状態の検知ができる．実際に，教師あり学習である k 近傍法 [3] やニューラルネットワーク [4]，教師なし学習である One-Class サポートベクターマシン [5] などを IDS に適用した研究例があり，いずれの手法も高い検知率が得られている．

IDS の実運用を考慮する場合には，最終的に判断する管理者が検知結果の検証に多くの時間を費やすことを避けるため，検知結果に対する説明性を有した手法が求められる．また高い検知率であるとはいえ誤検知は発生してしまうため，なぜ誤検知をしたのかを少なくとも推測する手段の提供が必要である．従来の機械学習手法でも，入力したトラフィックがどの既知の攻撃であるか（または類似しているか）を検知結果に対する説明として示すことは可能である．しかし，具体的にトラフィックのどの要素が検知結果に影響していたかを説明することは難しく，なぜ誤検知に至ったのかを推測する手段の提供も困難である．したがって，トラフィックの要素と検知結果との関係性を示すことが機械学習を適用した IDS において課題となる．また，それらの関係性についても，定量的な数値とともに示すことができれば，定性的な説明だけをするよりも説得力が期待できる．このことから，関係性をいかに定量的に示すかも課題として挙げられる．さらに，機械学習を用いた IDS の研究において単一の正常状態を仮定することがしばしばあるが，トラフィックの場合は時間や環境の変化などによって複数の正常状態が存在し得る [6] ため，単一の正常状態の仮定だけでは誤検知をしてしまう可能性がある．そのため，複数の正常状態を考慮した手法を提案することも課題となっている．

画像処理の分野では，入力画像のどの部分が重要であるかを定量的数値およびそれに基づいた可視化によって判定結果を説明する手法がすでに提案されている [7]．IDS においても，検知結果に対してトラフィックのどの要素が影響していたかを定量的数値に基づいた可視化によって説明できれば，課題を解決する 1 つのアプローチとなる．そこで本研究では，複数の要素とそれらの関係性を直感的に理解できるグラフ構造でトラフィックを表現することに着目し，定量的な数値に基づいたノードや部分グラフの可視化によって検知結果の説明ができる機械学習手法を用いた IDS の実現を目的とする．そもそもトラフィックは，IP やポート，フラグなど様々なフィールド情報という要素で構成されている．また，通信プロトコルが存在するため，正常な通信であれば各フィールド情報間は特定の相関関係をもった時系列的挙動をする．例えば，TCP 通信のコネクション確立に利用される 3 ウェイハンドシェイクにおいて，SYN と ACK は時系列的に連動した動きをする．これらの前提条件から，各フィールド情報の時系列データをノードとして捉えることで，トラフィックをそれらの相関関係としてグラフ構造で表現することができる．このようにトラフィックをグラフ構造で表現することができれば，単一の正常なグラフ構造を定義してそのグラフの崩れから異常を検知する教師なし学習や，正常/異常のグラフ構造を学習したモデルによって異常を検知する教師あり学習に応用が可能となる．そして，定量的な数値に基づいたノードや部分グラフの可視化によって検知結果を説明することができれば，本研究で挙げた主な課題を解決することが期待できる．ただし本稿において，複数の正常状態を考慮した手法の提案についての課題は，教師あり学習によるアプローチによってのみ解決を図る．

本稿の構成は，次のとおりである．2 章では，グラフ構造に基づいた説明性を有する異常検知の先行研究を整理する．3 章では，本研究で利用するデータセットの概要やそのデータに施す前処理手法についての説明をはじめ，2 種類の提案手法の詳細を説明する．4 章では，提案手法の実験による評価を行い，5 章で本稿のまとめと今後の課題を述べる．

2. 関連研究

機械学習によって高精度な異常検知ができたとしても，誤検知は発生してしまう．もしその異常検知を実運用に適用した場合，最終的に判断する管理者は検知結果の検証に多くの時間を費やすことになると予想される．したがって，入力のどの要素に起因した異常なのか，または正常な場合とどこが異なるのかなどを説明できる異常検知が必要となる．ここでは，グラフ構造に基づいた説明性を有する異常検知の先行研究を整理する．

そもそも私達の身の回りに存在するデータは，ソーシャルネットワークやセンサーネットワークなど要素間の相互関係に意味を持つものが多い．そのため，複数の要素とそれらの関係性をノードとエッジとして直感的に理解できるグラフ構造で表現し，それを異常検知に利用することは自然な発想である．実際に，グラフ構造に基づいた異常検知は，レビューサイトのフェイクレビュー検知 [8] や株式取引の異常パターン検知 [9]，ソーシャルネットにおけるスパム検知 [10] など，すでに様々な分野のデータを利用した研究が報告されている [11].

本研究と同様にグラフ構造に基づいた異常検知を IDS に応用した先行研究も存在する．Ding ら [12] は，ネットワークコミュニティ間の通信の振る舞いを監視し，コミュニティ構造に反する通信を異常として特定する手法を提案した．彼らの手法は，どのネットワークコミュニティ間で異常が発生しているかを結果的に説明することはできるが，どのような異常であるかを具体的にグラフ構造の要素によって説明することが目的ではなかった．また，可視化

による説明性の提供という観点に着目した研究例も存在する．Tong と Lin[13] は，残差行列が明示的に非負であることを要求した行列分解によって，異常の要因となるエッジの存在を可視化した．Akoglu ら [14] は，グラフ構造の中から特定の関係性を持つグループ（例えば，特定の異常を持つグループ）を見つけ出し，それらを部分グラフとして可視化する手法を提案した．彼らの手法は，異常の要因をエッジや部分グラフの可視化によって説明できるが，各要因の定量的な評価を目的とするものではなかった．

Ide ら [15] は，L_1 制約付きの疎構造学習を用いてグラフ構造を推定し，2 つのグラフ構造の比較によって異常度を計算する手法を提案した．彼らの手法は，片方のグラフ構造を単一の正常状態で仮定した場合に外れ値検知として利用することができる．また，各ノードの異常度が計算できるため，検知結果に対する定量的な説明が可能である．しかしながら，実際のトラフィックは時間や環境の変化などによって複数の正常状態が存在する可能性がある [6] ため，単一の正常状態を仮定する外れ値検知では対応が難しい．Gibberd[16] は，疎構造学習によるグラフ構造化を利用した変化点検知手法を提案した．彼らの手法は，現在のグラフ構造と少し前のグラフ構造を比較しているため，単一の正常状態を仮定する必要はなくなる．ところが変化点検知は，緩やかに変化する異常であった場合に異常の検知が難しい．また，前の状態からの変化で異常を検知するため変化した要素はわかるが，その異常自体の内容を説明するのに適していない．

3. 提案

本研究では，トラフィックの要素と検知結果との関係性を定量的に示すという課題を解決するため，まずトラフィックをグラフ構造で表現することに着目し，定量的数値に基づいたノードや部分グラフの可視化によって検知結果を説明する 2 種類の IDS を提案する．

提案手法 1 は，あらかじめ正常状態と仮定した単一のグラフ構造からの崩れを異常として検知する手法である．本手法では，ノードごとに定量的な異常度を計算することができ，その異常度が特定のしきい値を超えたノードを可視化する．また，本質的に Ide ら [15] が提案した異常検知手法と同等であるが，データ選定や前処理を検討し，トラフィックの異常検知に応用したものである．ただし本手法は，2 章で言及した複数の正常状態を考慮した手法の提案について課題を解決するものではなく，その課題の解決は提案手法 2 に委ねる．

提案手法 2 は，正常/異常のラベルを付与したグラフ構造をグラフ畳み込みニューラルネットワーク（GCNN: Graph Convolutional Neural Network）で学習し，学習モデルで異常を検知する手法である．検知結果に対する説明は，入力したグラフ構造の部分グラフの可視化によって行う．ここで得られる部分グラフは，検知結果に対する定量的影響度を示すことが可能である．また GCNN では，グラフ構造中の各部分グラフと学習対象（正常/異常）との関係を学習する側面を持つため，複数の正常状態の学習と捉えることができ，提案手法 1 で未解決であった課題の解決が期待できる．

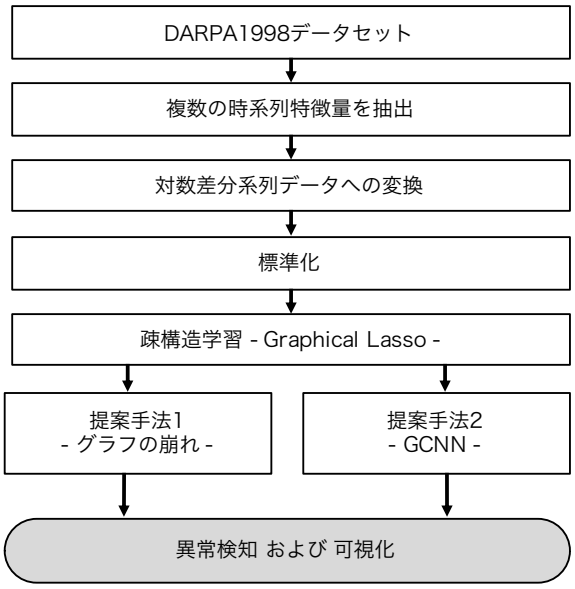

図 1　提案手法の概略図

3.1 データセットと前処理

提案手法（図 1）では，ラベルが付与されたパケットキャプチャ型のトラフィックデータが必要であるため，DARPA1998 データセットを使用する [17]．DARPA1998 は，仮想的に構築した空軍基地ネットワークに対するサイバー攻撃をシミュレートし，Tcpdump でトラフィックを収集したデータセットである．学習用データとして，月曜日から金曜日までの 5 日間を 1 週間分とした 7 週間分が提供されている．トラフィックの中から説明変数となり得るフィールド情報（送受信 IP や TCP/UDP ポート，TCP フラグやパケットサイズなど）を複数選択し，各変数でフィルタした 1 分単位のパケット数を時系列特徴量として用意する．得られた各時系列特徴量の変化率から異常を検知するため，それぞれを対数差分系列に変換し，標準化を施す．トラフィックの変化は，輻輳や異常の影響がなければ，短い時間スケールにおいて平均や分散が大きく変化しないとされている [18] ことから，確率分布として捉えることができる．今回作成した各時系列特徴量も特定の確率分布に従うデータとして，次節以降の方法論を考える．

3.2 疎構造学習

各時系列特徴量間に存在する相関関係のグラフ構造を求める（構造学習）ため，多変量正規分布を想定したガウス型グラフィカルモデル（以下 GMM と略記）を利用した．

GMM において，精度行列 Λ は変数間の直接相関を表し，構造学習はこの Λ を推定する問題と考えることができる．ただし，データに含まれているノイズの影響を排除するため，疎な Λ を求める工夫が必要となる．今回は，疎な Λ を推定する方法として式 (1) の L_1 制約項付きの最尤方程式を解くことを考える．

$$f(\Lambda; S, \mathrm{tr}) = \arg\max_{\Lambda}(\ln\det\Lambda - \mathrm{tr}(S\Lambda) - \rho||\Lambda||_1) \quad (1)$$

ここで，S は標本共分散行列，tr は行列の対角和，det は行列式を表す．ρ は正則化パラメータであり，どの程度を Λ を疎な構造にするかを決定する．$||\Lambda||_1$ は $\sum_{i,j=1}^{M}|\Lambda_{i,j}|$ により定義される．特に本研究では，ブロック勾配法 [19] を用いることで式 (1) を効率よく解くことができる Graphical Lasso に着目する [20], [21]．$\partial f/\partial\Lambda = 0$ をブロック勾配法で解くため，ある特定の変数 x_i に関係する行と列が最後にくるように並び替えた Λ とその逆行列 Σ を次のように分割する．

$$\Lambda = \begin{pmatrix} L & l \\ l^T & \lambda \end{pmatrix}, \Sigma \equiv \Lambda^{-1} = \begin{pmatrix} W & w \\ w^T & \sigma \end{pmatrix} \quad (2)$$

式 (2) における各変数の次元は，$W, L \in R^{(M-1)\times(M-1)}$，$w, l \in R^{M-1}$，$\lambda, \sigma \in R$ である．また，標本共分散行列 S も同様に分割し，行列の微分公式や Λ が正定値であることを利用することで，式 (1) を L_1 制約付き回帰問題に帰着させ，効率よく解くことができる．この Graphical Lasso を用いて，各時系列特徴量を 30 分のウィンドウサイズで疎なグラフ構造に変換し，1 分毎に更新するグラフ構造の時系列データを作成する．

3.3 異常度の定義

提案手法 1 では，単一に定義した正常状態のグラフ構造と時系列的に並んだグラフ構造とを順に比較することで，その違いを異常として判断することを考える．前節までの計算において，正常状態のグラフ構造とその比較対象となるグラフ構造は，それぞれ確率分布 $p_A(x)$ および $p_B(x)$ と考えることができる．このように 2 つの確率モデルが与えられている場合，一般に差異の尺度として用いられる Kullback-Leibler 距離（以下 KL 距離と略記）が挙げられる．特定の変数 x_i に着目したとき，$p_A(x_i|z_i)$ および $p_B(x_i|z_i)$ の KL 距離の期待値を分布 $p_A(z_i)$ によって計算すると式 (3) となる．

$$d_i^{AB} \equiv \int dz_i p_A(z_i) \int dx_i p_A(x_i|z_i) \ln\frac{p_A(x_i|z_i)}{p_B(x_i|z_i)} \quad (3)$$

ただし，$z_i \equiv (x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_M)^T$ と定義する．また，そもそも GMM では正規分布を想定しているため，式 (3) は解析的に求めることができる．

$$\begin{aligned} d_i^{AB} &= w_A^T(l_B - l_A) \\ &+ \frac{1}{2}\left\{\frac{l_B^T W_A l_B}{\lambda_B} - \frac{l_A^T W_B l_A}{\lambda_A}\right\} \\ &+ \frac{1}{2}\left\{\ln\frac{\lambda_A}{\lambda_B} + \sigma_A(\lambda_B - \lambda_A)\right\} \end{aligned} \quad (4)$$

式 (4) は，A と B は入れ替えることで d_i^{BA} も同様に求められる．2 つのグラフ間の異常度は，最終的に式 (5) で定義する．それにより，各ノード i ごとの異常度として定量的に示すことができる．

$$a_i \equiv \max\left\{d_i^{AB}, d_i^{BA}\right\} \quad (5)$$

3.4 グラフ畳み込みニューラルネットワーク

本節では，提案手法 2 で利用する GCNN について述べる．GCNN は，画像を入力とした問題に有効な CNN と同様の処理を，グラフ構造を入力とした問題にも対応させるために研究されてきた手法である．ここでは特に，化学分野において報告された Neural Fingerprint（NFP）[22] に着目する．NFP は，分子構造をグラフと捉えて学習することで，水溶度や毒性の高精度な予測を可能にしている．また，NFP では中間層として Fingerprint（部分グラフの集合）を生成することができ，どの部分グラフが予測結果に影響したかを評価することができる．そのため，毒性の学習であれば毒性に最も影響した部分グラフを特定，可視化することができる．本研究では，Chainer Chemistry[23] で提供されている NFP を利用した．

ノード v とエッジ u で構成されたグラフにおいて，あるノード v_i とそれを中心に近傍数 r で隣接するノードとの結合情報を式 (6) で表現する．

$$n_i^{(r)} = \sum_j v_j^{(r)} u_{(i,j)k} \quad (6)$$

$u_{(i,j)}$ は，ノード i,j 間のエッジを表すベクトルである．$n_i^{(r)}$ は，グラフ構造における部分グラフに相当する．また，$n_i^{(r)}$ を入力とした重み W_c およびバイアス b の単層ニューラルネットワークに対して，活性化関数を施す．

$$v_i^{(r+1)} = \sigma\left(n_i^{(r)} W_c^{(r)} + b^{(r)}\right) \quad (7)$$

活性化関数には，シグモイド関数 σ を利用する．各部分グラフに対して重み（画像処理分野におけるフィルタ）を適用している点が CNN における畳み込み操作と類似している．また，本手法におけるネットワークの深さは，近傍数 r に相当し，部分グラフの大きさを決定する．Fingerprint は，softmax 演算を施した実数値のベクトルとして生成する．ここでの softmax 演算は，インデックス付け操作と捉えることができ，また CNN におけるプーリング操作と類似している．Fingerprint では，各ビットがそれぞれ部分グラフを表し，各ビットの値は結果に対する影響度を表すことになる．異常を学習対象とした場合，異常に対する部分グラフの影響度を 0〜1 の実数をとして Fingerprint を形成

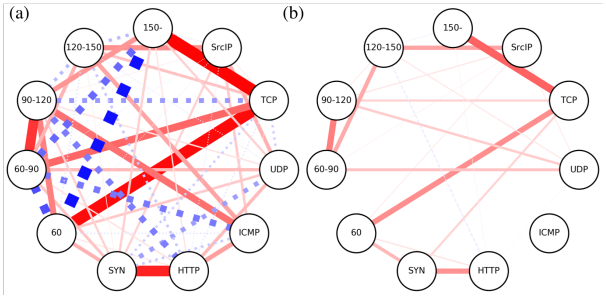

図 2　トラフィックデータから作成したグラフ構造：Graphical Lasso の (a) 適用前および (b) 適用後

し，正常を学習対象とした場合はその逆となる．

$$f = \mathrm{softmax}(v_i^{(r+1)} W_a^{(r)}) \tag{8}$$

そこで提案手法 2 では，元の DARPA1998 データセットの正常および異常のラベルを利用して，グラフ構造の時系列データにもラベルを付与し，NFP で学習することによって異常状態の予測および異常の要因となる部分グラフの可視化を行なう．

4. 評価

4.1 トラフィックデータのグラフ構造化

トラフィックデータから作成したグラフ構造を図 2 に示す．時系列特徴量には，送信元 IP 数，各プロトコル（TCP，UDP，ICMP，HTTP）別パケット数，SYN フラグを含むパケット数，および 5 種類のサイズ別パケット数の合計 11 種類を選択した．グラフ構造の描画には，精度行列から算出した偏相関係数を利用した．

$$r^{i,j} \equiv -\frac{\Lambda_{i,j}}{\sqrt{\Lambda_{i,i}\Lambda_{j,j}}} \tag{9}$$

実線は $r^{i,j} > 0$，点線は $r^{i,j} < 0$ を表し，線の太さは $r^{i,j}$ の絶対値の大きさを表す．正則化パラメータ ρ は，グリッドサーチによって決定した．Graphical Lasso を適用する前のグラフ構造（図 2(a)）は，全ての特徴量間に相関が存在することを表す完全グラフとなった．一方，Graphical Lasso を適用した後のグラフ構造（図 2(b)）は，疎なグラフ構造となった．疎なグラフ構造は，完全グラフと比べてノイズの影響を排除できるため，異常検知への利用の観点からノイズに対して頑強になると考えられる．

4.2 異常検知および可視化

4.2.1 提案手法 1

図 3 は，DARPA1998 データセットにおける 7 週目の金曜日のデータについて異常検知を行った結果である．図 3(a) は，トータルパケット数のデータであり，色付きの部分が異常を表している．7 週目の金曜日のデータに含まれる異常は，80 番ポートに対する Syn-flood 攻撃，および複数の送信元 IP から 1 つの送信先 IP に対して行われる

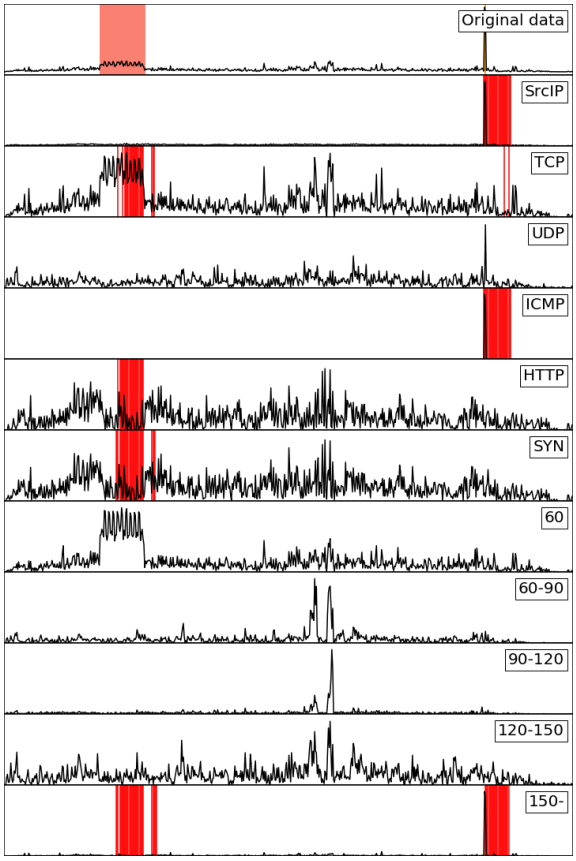

図 3　DARPA1998 の正解データおよび各時系列特徴量と提案手法 1 によって検知した異常の可視化結果

Smurf 攻撃である．図 3(b)〜(l) は，今回選択した 11 種類の時系列データであり，こちらも色付きの部分が提案手法 1 によって検出した異常で表している．実際はノードごとに定量的な数値で異常度が得られ，しきい値を決めることによって色付けしている．

最初の異常は，TCP，HTTP，SYN，パケットサイズ（150bytes 以上）で検知しており，80 番ポートに対する Syn-flood 攻撃であると推測できる．次の異常は，送信元 IP，ICMP，パケットサイズ（150bytes 以上）で検知しており，複数の送信元 IP から 1 つの送信先 IP に対して行われる Smurf 攻撃であると推測できる．このように各ノードの異常度を個別に可視化することによって，異常と判断した理由の説明を提供することができる．それにより検知結果を検証する管理者は，そのノードの種類や値によってある程度の推測が可能となる．また，今回は便宜上上記の 11 種類のノードを設定したが，実際の管理者の経験則に従ってノードの種類を決定することでさらに説明性のある結果になると予想される．ただし，図 3 の結果から一部検知の遅れや過剰検知も確認できた．遅れや過剰検知に関しては，グラフを作成する際に設定したウィンドウサイズが原因の 1 つとして考えられる．今回はウィンドウサイズを 30 分で設定しているため，緩やかに変化する異常の場合に前後最大 30 分の異常が検出できない可能性がある．ま

表 1　GCNN の学習パラメータ

パラメータ	値
Fingerprint 長	100
最大近傍数	2
隠れ層サイズ	100
バッチサイズ	32
エポック数	1000

た，突発的に変化をする明らかな異常の場合は，前後最大 30 分を過大評価してしまう可能性がある．しかしながら，ウィンドウサイズを小さくすれば相関関係のグラフ構造を作成するための十分なデータが得られなくため，異常検知自体がうまくいかなくなることが考えられる．そのため，ウィンドウサイズを小さくする場合は，時系列データを集計するデータの粒度も同時に小さくするなどの工夫が必要になる．

4.2.2　提案手法 2

正常/異常の学習には，3.4 節で示した Fingerprint を生成するためのネットワークに加えて，全結合層および活性化関数 relu を利用した．また，正常/異常の 2 値分類を目的としているため，最終層の活性化関数にはシグモイド関数，損失関数には交差エントロピーを利用した．GCNN の学習に利用した主なパラメータを表 1 に示す．訓練データおよび検証データは，70:30 の比率で分割した．その結果，検証データにおいて 98%の精度で異常を予測できることを示した．この値は，先行研究と同程度に高い値である [5]．図 4 は，SYN-flood 攻撃および Smurf 攻撃を異常として検知した際の Fingerprint および影響度の高い部分グラフの一例である．部分グラフの結果から，HTTP と SYN に関わる Syn-flood 攻撃，および送信元 IP と ICMP に関わる Smurf 攻撃であることがそれぞれ推測できる．ノードだけに注目すれば，提案 1 の異常検知とほとんど同様の結果である．ただし提案手法 2 では，異常の要因となるノードだけでなくそれらの相関関係までを示すことができるため，ノード単体の異常の可視化に比べてより具体的な説明性を提供できる．また GCNN における学習では，グラフ構造の中の各部分グラフと学習対象（正常/異常）との関係を学習する側面を持つため，複数の正常状態も部分グラフに分けた状態で学習していると考えられる．そのため，単一の正常状態で仮定している提案手法 1 よりも時間や環境の変化などに頑強な検知ができていると考えられる．ただし，その効果を実験により評価を行うには，それに適したデータセットや条件を用意する必要があるため，今後の研究課題とする．

5.　おわりに

本研究では，トラフィックデータから抽出した 11 種類の時系列特徴量を疎構造学習を用いてグラフ構造化できることを示した．得られたグラフ構造を利用し，トラフィックデータの異常を検知する 2 種類の手法を提案した．提案した 2 種類の手法は，異常の検知だけでなく，異常となった要因を定量的な数値に基づいたノードまたは部分グラフの可視化によって説明できた．ただし，より複雑な異常を検知および可視化する場合は，ノードとなる時系列特徴量を増やす，またはより効果的なものに変更するなどの工夫が必要となると予想される．また，複数の正常状態を考慮した手法として GCNN を用いた手法を提案したが，実際に単一の正常状態を仮定したものと比較した検証は行っていないため，今後の研究課題とする．また提案手法 2 によって生成された Fingerprint は，クラスタリング手法を利用することで，既知のどの攻撃に近いのかなど，より具体的な説明への応用が期待できる．今回のデータセットには，パケットキャプチャ型のトラフィックデータが必要であるため DARPA を利用したが，データの古さや実際のネットワークトラフィックとの違いなどが指摘されているため，今後は最新のトラフィックデータに適用した場合の有効性を検討する．

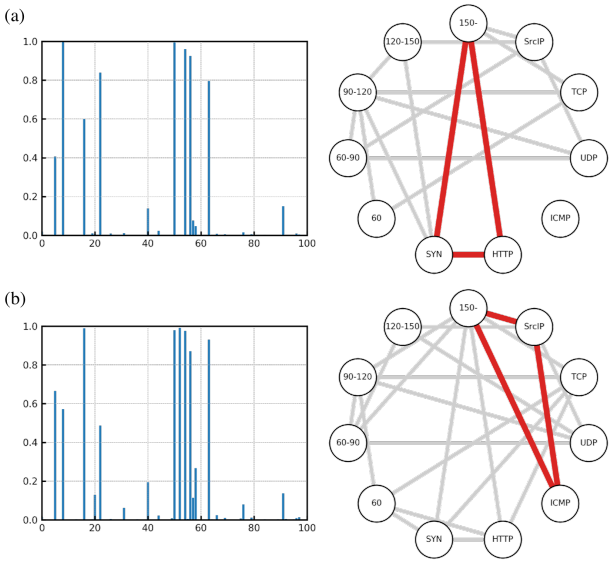

図 4　Fingerprint および異常の要因となった部分グラフの一例：(a)SYN-flood 攻撃, (b)Smurf 攻撃

参考文献

[1] 総務省：平成 29 年版　情報通信白書 (online)，入手先 ⟨http://www.soumu.go.jp/johotsusintokei /whitepaper/ja/h30/index.html⟩ (2018.12.26).
[2] 情報通信研究機構：NICTER 観測レポート 2017 (online)，入手先 ⟨https://www.nict.go.jp/cyber/report /NICTER_report _2017.pdf⟩ (2018.12.26).
[3] Y. Liao and V. Rao Vemuri: Use of k-nearest Neighbor Classifier Forintrusion Detection, *Computers Security*, Vol. 21, pp.439-448 (2002)
[4] J. Kim, J. Kim, H. L. T. Thu and H. Kim: Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection, *In Proceedings of 2016 International Conference on Platform Technology and Service (PlatCon)*, pp.1 - 5(2016).

[5] G. Li, Z.Yan, Y. Fu and H. Chen: Data Fusion for Network Intrusion Detection: A Review, *Security and Communication Networks* Vol.2018, pp.1-16(2018).
[6] V. Paxson and S. Floyd: Why don't know how to simulate the Internet, *In Proceedings of the 1997 Winter Simulation Conference (WSC)*, pp.1037-1044 (1997).
[7] M. Ribeiro, S. Singh and C. Guestrin, " why should i trust you?" explaining the predictions of any classifier, *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD16)*, (2016).
[8] G. Wang, S. Xie, B. Liu, and P. S. Yu: Review Graph Based Online Store Review Spammer Detection, *In Proceedings of the 11th IEEE International Conference on Data Mining (ICDM)*, pp.1242 - 1247 (2011).
[9] Z. Li, H. Xiong, Y. Liu, and A. Zhou: Detecting Blackhole and Volcano Patterns in Directed Networks, *In Proceedings of the 10th IEEE International Conference on Data Mining (ICDM)*, pp.294 - 303 (2010).
[10] H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. Choudhary: Towards Online Spam Filtering in Social Networks, *In Proceedings of the 19th Annual Network Distributed System Security Symposium*, (2012).
[11] L. Akoglu, H. Tong and D. Doutra: Graph-based Anomaly Detection and Description: A Survey, *In Proceedings of the 11th SIAM International Conference on Data Mining (SDM)*, Vol. 29, Issue. 43, pp.626-688 (2015).
[12] Q. Ding, N. Katenka, P. Barford, E. D. Kolaczyk, and M. Crovella: Intrusion as (Anti)Social Communication: Characterization and Detection, *In Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, pp.886 - 894 (2012).
[13] H. Tong and C.Y. Lin: Non-negative Residual Matrix Factorization with Application to Graph Anomaly Detection, *In Proceedings of the 11th SIAM International Conference on Data Mining (SDM)*, pp.119 - 130(2011).
[14] L. Akoglu, J. Vreeken, H. Tong, D. H. C. N. Tatti, and C. Faloutsos: Mining Connection Pathways for Marked Nodes in Large Graphs, *In Proceedings of the 13th SIAM International Conference on Data Mining (SDM)*, (2013).
[15] T. Ide, A. C. Lozano, N. Abe and Y. Liu: Proximity-based Anomaly Detection using Sparse Structure Learning, *In Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (2014).
[16] A. J. Gibberd and J. D. B. Nelson: High dimensional changepoint detection with a dynamic graphical lasso, *In Proceedings of 2009 SIAM International Conference on Data Mining*, (2009).
[17] MIT: DARPA1998Dataset (online), 入手先 ⟨https://www.ll.mit.edu/r-d/datasets/1998-darpa-intrusion-detection-evaluation-dataset⟩ (2018.10.23).
[18] 原 聡：機械学習における解釈性, 人工知能 (2018).
[19] O. Banerjee, L. E. Ghaoui, and G. Natsoulis: Convex optimization techniques for fitting sparse Gaussian graphical models, {it In Proceedings of International Conference Machine Learning, pp.89 - 96 (2006).
[20] J. Friedman, T. Hastie and R.Tibshirani: Sparse Inverse Covariance Wstimation with the Graphical Lasso, *Biostatistics* Vol.9, No,3, pp. 432 - 441 (2008).
[21] 井出 剛，杉山 将：異常検知と変化検知 (2017).
[22] D. Duvenaud, D. Maclaurin, J. Aguilera-Iparraguirre, R. Gomez-Bombarelli, T. Hirzel, A. Aspuru-Guzik, R. P. Adams: Convolutional Networks on Graphs for Learning Molecular Fingerprints, *NIPS'15 Proceedings of the 28th International Conference on Neural Information Processing Systems* Vol.2, pp.2224-2232 (2015).
[23] Preferred Networks: Cheiner Chemistry, 入手先 ⟨https://github.com/pfnet-research/chainer-chemistry⟩ (2018.12.11).