# 機械学習を用いたIDSログからの情報抽出に関する考察

2016SE091 矢田智也　2016SE038 川崎陽介
指導教員：沢田篤史

## 1　はじめに

Webサーバ等へ不正にアクセスし，個人情報を盗み出したり，ウイルスに感染させて正常に稼働できない状態に陥れるといったサイバー攻撃が増加しており，迅速な検知や侵入防止，情報漏洩の防止といったサイバー攻撃対策の発展が課題となっている．

サイバー攻撃への対処法である侵入検知システム（IDS：Intrusion Detection System）は常に通過するパケットを監視し，攻撃を検知したらユーザへ通報するシステムである．IDSの検知方法は2種類存在するが，それぞれ誤検知の発生が問題となっている．正常通信を定義しておき，それ以外の通信を攻撃と判断するアノマリ型は攻撃でない通信を攻撃と判断するfalse positiveが多い．攻撃のパターン（シグネチャ）を定義してそれを基に攻撃を検知するシグネチャ型は，攻撃通信を攻撃と判断しないfalse negativeが多い．IDSのセキュリティを向上させるためには定義していないパターンの攻撃への対策が必須である．

本研究の目的は，未知の攻撃を検知し，その攻撃のパターン化を行い，パターンに基づいて異常と考えられる通信を指摘するためのシステムを構築することである．ここで述べる未知の攻撃とは，シグネチャ型IDSにおいて登録されている攻撃パターンと一致しないが，攻撃の恐れが高い通信パターンをいう．未知の攻撃のパターン化には，通信のログを何らかの基準に従って分類し，観測された通信がどれに分類されるかを判別する必要がある．

本研究の研究課題は次の3つである．

1. シグネチャ型IDSにおいて正常通信と判断されたログを分類し，未知の攻撃の候補となりうる通信パターンを抽出する方法の確立
2. 観測された通信が抽出された通信パターンにあてはまるかを判別する方法の確立
3. 上記の2点を実現する検知システムを構築するためのアーキテクチャを定義

課題1では，正常通信と分類された通信ログをクラスタリングすることで，正常通信の枠の中で類似性が高い通信にまとめることが目的である．課題2では，クラスタを基に抽出した重心を使った重心法に基づいて現在監視されている通信がどのクラスタに類似するかを判別する．その後その通信に対してラベル付けを行い，再びクラスタリングを行う．通信の度に行うことで現段階での詳細な枠組みを施したクラスタを得ることが目的である．課題3では上記の2点を達成するために，IDSの共通フレームワークであるCIDF（Common Intrusion Detection Framework）[2]を改良する．これにより，IDSで未知の攻撃を検知するアプリケーションのためのアーキテクチャを設計する．

本研究ではWeka[6]のクラスタリング機能で設定したクラスタに分類し，クラスタの重心のデータを用いて通信のログがどのクラスタに分類するかを判別する機能をPythonで実現する．CIDFを改良したアーキテクチャに基づいて簡単なプロトタイプのシステムを開発し，KDD CUP 99DataSet[1]を加工した正常通信ログから未知の攻撃の可能性の高い通信ログを抽出する実験を行う．

## 2　侵入検知システムとそれに関する課題

### 2.1　IDS（Intrusion Detection System）

IDSとは通信を監視し，サイバー攻撃があるとユーザに警告するシステムであり，検知方法が2つ存在する．シグネチャ型は不正アクセスの攻撃パターンを定義し，攻撃パターンと一致したデータを得ると不正アクセスと判断する．アノマリ型は正常通信のパターンを定義し，それ以外の通信を攻撃と判断する．IDSは監視対象によって2つに分類され，ネットワーク型はネットワーク上の通信データを監視して通常通信と比較して攻撃を検知し，ホスト型はホスト上のデータを監視して異常を検知する．

### 2.2　CIDF

IDSの標準フレームワークであるCIDFはプロトコルとアプリケーションプログラミングインターフェースの開発に使用される取り組み方法であり，侵入検知コンポーネントを他のシステムで再利用することができる．CIDFの構成を以下に示す．

- Event Generator
  イベントを取得する．ネットワークベースとホストベースに分類
- Event Analyzer
  侵入時に考えられる構成要素．他のコンポーネントからの情報を分析し，侵害状況を特定
- Event Database
  イベント情報やその分析結果を保存するデータベース．データベースをもとに記録を視覚化してシステムの状況を把握
- Response Unit
  他のCIDFコンポーネントに変わってアクションを実行するように支持する場所（プロセスの強制終了，接続のリセット，ファイルのアクセス許可の変更などが含まれる）

### 2.3 IDS 及び IDS のフレームワークに関する課題

IDS はすべての攻撃を検出できるわけではない．我々が着目する IDS において発生する誤検知は実際には攻撃でない通信を攻撃と認識する false positive と実際の攻撃を攻撃として認識しない false negative の 2 種類に分けられる．アノマリ型は未知の攻撃には強いが正常通信を正確に定義しなければいけないので，false positive が非常に多い．シグネチャ型はアノマリ型に比べて false positive は少ないが，定義していない攻撃やパターンを少し変えた攻撃を防ぐことが困難であり，false negative が多く発生する．

一方で，未知の攻撃を検知するという本研究の目的を解決するためのソフトウェア基盤が整備されていない．ソフトウェアアーキテクチャを定義することで，未知の攻撃を検知するシステムのための共通基盤を整備する．

## 3 クラスタリングを用いた未知の攻撃の検出方法

### 3.1 検出方法の概要

IDS におけるシグネチャからすり抜けた攻撃通信と正常通信が入った通信ログを用意する．その通信ログをクラスタリングし，IDS が正常通信として判断した攻撃通信がクラスタの重心を使用した重心法により，指定したクラスタ数に応じて通信ログを本当の正常通信と攻撃通信に分類する．作成できたクラスタの一つに攻撃の指摘があった場合，そのクラスタに属する通信は攻撃通信である可能性の指摘が可能である．実際に妥当性を検証するために，本研究では KDD CUP 99 Data Set を使用する．

### 3.2 クラスタリング

クラスタリングとはデータ等の集合体を機能やカテゴリに分割することであり，教師なし学習に分類される機械学習の一つである．本研究では任意でクラスタの数を設定でき，計算不可が少なく膨大なデータに対しても素早く計算できる k-means 法を採用した．本研究では，設定するクラスタの数を 2 つと 5 つの 2 種類に設定して検証を行う．

最初の検証において，クラスタを 2 つに設定して検証を行う．これは，正常通信と攻撃通信のクラスタに分類することができるかを検証するためである．次に，クラスタの数を 5 に設定して検証を行う．これは，一般にサイバー攻撃の種類を 4 種類に分類することが多いからである．ゆえに，正常通信と攻撃対象の調査，DoS 攻撃，バッファオーバーフロー攻撃，その他の攻撃の 4 つの攻撃通信を合わせた 5 つのクラスタにおいても正しく分類できるかを検証することが目的である．

### 3.3 特徴値の選定

それぞれの特徴値からすべて相関分析した結果，ピアソンの積率相関に基づいて相関係数を r とし，$0.7 \leq r \leq 1.0$ となる値を相関係数が高いと判断して使用する特徴値とみなす．作成したクラスタを可視化して確認した結果，ログが一方に極端に偏っている特徴値やログの値がすべて等しい特徴値を発見し，使用する特徴値から削除した．これらの結果，使用する特徴値を 17 個用いることに決定した．

## 4 未知の攻撃を検知するためのアーキテクチャ設計

本研究では，未知の攻撃を指摘できるようなアーキテクチャを設計する．IDS を正常通信として通過したパケットをクラスタリングし，処理した結果を貯めて現在着目しているデータが分類したクラスタのどれに該当するかを判別する処理を，IDS のための共通フレームワークである CIDF を改良することで IDS に適したアーキテクチャを設計する．従来の IDS のアーキテクチャを図 1 に示す．

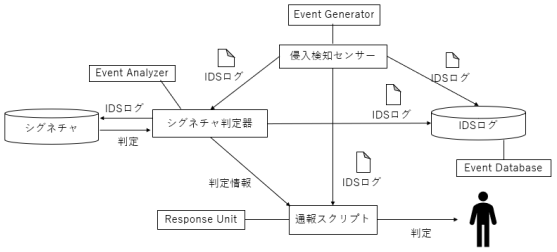


図 1 従来の IDS の動的構造

図 1 は従来の IDS のアーキテクチャの動的構造を示す．侵入検知センサーに検知された通信はデータベースにあるシグネチャと合致するか判定され，すべてのログを保存するデータベースに格納される．通信とシグネチャが合致した場合を攻撃通信とみなし，ユーザへ通報される．

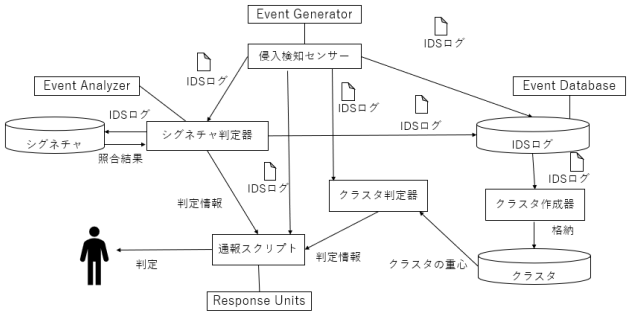


図 2 CIDF を拡張したアーキテクチャの動的構造

それに対して，本研究で提案するアーキテクチャの動的構造を図 2 に示す．すべての IDS のログをクラスタリングし，正常通信をクラスタに分けてクラスタの重心を求める．その重心を基に，現在流れている通信がどのクラスタに所属するかを重心法によって判定し，再び流れてくる通信を含めたすべてのログをクラスタリングして再学習し，重心を再度求めて判定する．これらの操作を繰り返し行うことで，より緻密な正常通信の分類が可能となり，もし分類したクラスタが未知の攻撃通信の疑いができた場合，まとめて攻撃通信の指摘が可能である．

我々が提案した IDS のアーキテクチャにおいて追加した機能を説明する．IDS ログのデータベースから出力されたログデータをクラスタリングに使用するためにクラスタ作成器を用意する．このクラスタ作成器でクラスタリングを実行し，クラスタの重心を求める．IDS ログのデータベースとは別に，クラスタリングを実行したデータを格納したデータベースを用意することで，クラスタとクラスタの重心をデータベースに格納することができる．そのクラスタの重心を用意したクラスタ判定器に出力することで，クラスタの重心を基に現在流れている IDS ログを監視し，未知の攻撃と仮定した攻撃を検知できるかどうかを判定することができる．

## 5 評価

### 5.1 KDD CUP 99 Data Set

不正アクセスに対する研究において扱われるデータセットとして使用される KDD CUP 99 Data Set は，約 500 万件のフルセットとその中から 10% を抽出した約 50 万件の 10% データセットからなり，学習用と評価用のデータセットが存在する．学習用データには 22 種類の攻撃と正常通信が含まれ，評価用データには 39 種類の攻撃と正常通信が含まれる．この内 17 種類は学習データにない攻撃であり，学習データにあるが評価用データにない攻撃も 2 種類ある．本研究では，学習用データセットと評価用データセットを使用する．

### 5.2 Weka

Weka は世界中で多くの機械学習の研究で使用されている，Java 言語によるオープンソースのデータマイニングのフリーソフトである．k-means 法には，Weka における実装である SimpleKMeans にて検証し，クラスタの数は 2 つと 5 つの合計 2 回検証を行う．

### 5.3 評価方法

我々が提案するアーキテクチャに基づいて実装することが可能であることを示すために構築したプロトタイプシステムの評価を図 3 に示す手順で行う．

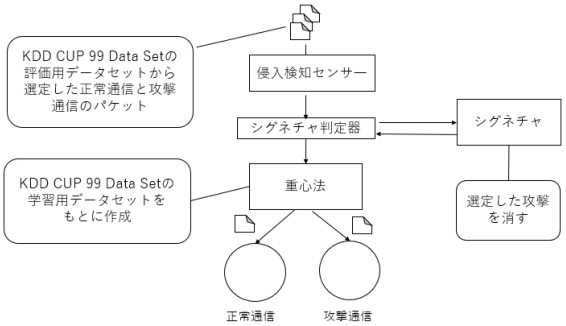


図 3 評価の手順

- KDD CUP Data Set のデータの中から一部の攻撃を取り出し，取り出した攻撃を正常通信と仮定する．
- 実際にクラスタリングできることを示す．
- パケットとクラスタの類似度を測ることで未知の攻撃に近い攻撃を検出できることを示し，その評価を行う．

実装後の評価を行うために，KDD CUP 99 Data Set の評価用データセットから選定した攻撃通信と正常通信が入ったパケットを用意する．シグネチャから選定した攻撃を意図的に外し，IDS が既知の攻撃として認識できなくする．その後 Weka を用いて学習用データセットを基に設計したシステムによりクラスタの重心を求める．作成したパケットを流し，正常通信と学習用データセットにない 17 種類の攻撃通信がクラスタの重心を使用した重心法により本当の正常通信と攻撃通信に分類できるかを判定することで評価を行う．

### 5.4 評価指標

機械学習の評価基準として用いられる指標に，正例と負例のデータを正しく識別した割合である正解率（精度，Accuracy），正例と判断したもののうち，真に正例であったものの割合である適合率（Precision），正例全体のうち，正例と判定された割合である再現率（Recall）がある．またそれぞれの指標には一長一短があるので，これらを総合的に評価する指標として F 値（F-measure）がある．

### 5.5 検証結果

設定するクラスタ数を 2 つと 5 つの 2 種類で検証し，5 つのクラスタでの結果を示す．クラスタリングで使用するデータセットは IDS によって正常通信と判断されたログとし，クラスタリングを行う際に使用されたログを A，A からログを選定して攻撃のラベルを外したものを B と呼称し，以下に具体的な検証方法を示す．

1. A の特徴値を本研究において設定した値まで削減
2. 削減した A をクラスタリング機器にかける
3. 定義したクラスタ数に応じてクラスタを作成
4. B がどのクラスタに類似するかを重心法により判定
5. 判定結果を基に正常通信と攻撃通信の正解率，適合率，再現率，F 値を検出

本研究における判定器としてどのクラスタに分類するかを判定するシステムを実装した．判定器はクラスタの重心，IDS ログを入力して判定結果を出力させた．実装したシステムを基に KDD CUP 99 Data Set を使用して判定を行った結果，上記の仕様通りに判定できた．本研究の検証ではクラスタリング機器に Weka を，判定器は Anaconda の Spyder を使用して作成した．データセットは学習用データセットを使用してクラスタリングし，重心を求めた．

評価用データセットから意図的に正常通信の 20 個と攻撃通信の 80 個の合計 100 個を抽出し，攻撃のラベルを削除して判定器にかけた結果，TP（True Positive）が 100%，TN（True Negative）が 52%，FP（False Posotive）が 0%，

FN（False Negative）が 47.5% となった．評価指標を求めるために TP，TN，FP，FN の 4 つの値を使用し，以下の各指標の定義式によって結果を得た．

表 1　通信ログの判別結果

| | TP | TN | FP | FN | 合計 |
|---|---|---|---|---|---|
| (個) | 20 | 42 | 0 | 38 | 100 |

- 正解率 $= \dfrac{TP+TN}{TP+FN+FP+TN}$
- 適合率 $= \dfrac{TP}{TP+FP}$
- 再現率 $= \dfrac{TP}{TP+FN}$
- F 値 $= \dfrac{2 * \text{再現率} * \text{適合率}}{\text{再現率} + \text{適合率}}$

表 2　評価指標の結果

| | 正解率 | 適合率 | 再現率 | F 値 |
|---|---|---|---|---|
| (%) | 62 | 100 | 34.5 | 50.7 |

## 6　考察

### 6.1　クラスタリングの妥当性の考察

クラスタリングを行うことにより，攻撃判定されなかった攻撃通信に対し複数のクラスタに分類，指摘することが可能である．また，シグネチャ型における未知の攻撃に対する脆弱性は本研究での方法で指摘可能であり，アノマリ型 IDS のにおける誤検知の多発という弱点も，シグネチャ型 IDS を使用前提しているので比較的少ないと思われる．

### 6.2　提案したアーキテクチャの妥当性の考察

IDS ログのデータベースから出力されたログデータをクラスタリングに使用するためにクラスタ作成器を用意し，この作成器によってクラスタリングを実行し，クラスタの重心を求めた．IDS ログのデータベースとは別に，クラスタリングを実行したデータを格納したデータベースを用意することで，クラスタとクラスタの重心をデータベースに格納することができた．クラスタの重心を用意したクラスタ判定器に出力することで，クラスタの重心を基に流れてくる IDS ログを監視し，未知の攻撃と仮定した攻撃を検知できるかどうかを判別することができた．

### 6.3　関連研究との比較

高原の研究では KDD CUP 99 Data Set の特徴値を 41 個のうち 38 個使用しているのに対し，我々の研究では 17 個使用している．使用する特徴値を可能な限り削減することで高原の研究では誤検知率が 15%，正解率が 72% に比べて，本研究での提案方法では誤検知率は 0% であったが，正解率は 62% となり正解率が低い結果となった．しかし，クラスタ数を増やせば正解率などが上昇したので，クラスタ数とクラスタリング精度を向上させることでより高い精度を得られると思われる．

## 7　おわりに

IDS は未知の攻撃に弱く，その対策を立てなければならないので，本研究では未知の攻撃を検知し，攻撃のパターン化を行い，パターンに基づいて異常と考えられる通信を指摘するためのシステムの構築を示した．

我々が提案する方法は，IDS によって正常通信と判断された IDS ログをクラスタリングしてクラスタの重心を求め，現時点で流れる攻撃を重心法を用いてどのクラスタに類似するか判定し，正常通信と判断された攻撃通信を検知する方法である．我々はクラスタの判別処理を CIDF を改良して未知の攻撃を検知するアプリケーションのためのアーキテクチャを設計した．その後プロトタイプシステムを構築し，擬似的なログデータを用いて実験したところ，通信ログの判別結果による正解率は 62% となった．

今後の課題は，未知の攻撃に対するよりクラスタ数やクラスタリング精度の最適化と攻撃通信への指摘の発展である．攻撃通信の指摘はユーザが行う前提であるので，よりユーザの選択を広げられるように通信の可視化や自動判別を作成することで，検知率はより向上すると考えられる．また，本研究で作成したプログラムは非常に単純であるので，より現実的な事例に対応可能なプログラムを記述し，実用性について検証する必要がある．

## 参考文献


[1] KDD Cup 1999 Data, `http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html`, 2015.

[2] Porras.P, Schnakenberg.D, S.Staniford-Chen, Stillman.M and Wu.F, “ The Common Intrusion Framework Architecture, ” `http://www.isi.edu/gost/cidf/drafts/architecture.txt`, 1998.

[3] 小池宏明，宮地玲奈，川口信隆，重野寛，岡田謙一，“ 機械学習によるネットワーク IDS の false positive 削減方法, ” 情報処理学会論文誌，Vol. 45，No. 8，pp. 2105-2112，2004.

[4] 沢田篤史，高倉弘喜，岡部寿男, “ 開放型大規模ネットワークのための IDS ログ監視支援システム, ” 情報処理学会論文誌，Vol. 44，No. 8，pp. 1861-1871，2003.

[5] 高原尚志，櫻井幸一, “ KDD CUP 99 Data Set を用いた異なる学習データによる機械学習アルゴリズムの評価, ” Computer Security Symposium 2015，pp. 457-464，2015.

[6] 高原尚志, “ Random Forests と K-Means 法によるハイブリッド式アノマリ検知方式, ” Computer Security Symposium 2016，pp. 1019-1026，2016.

[7] 武田圭史, “ 侵入検知システムに関する研究の現状, ”情報処理，Vol. 42，No. 12，pp. 1-6，2001.