社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2003－CSEC－23 (3)
2003／12／19

誤警報を削減し未知のDoS攻撃を検知するNIDS

稲垣 洋介†, 古屋 治之†, 武藤 佳恭††

ネットワークにおける不正侵入を検知するツールには NIDS(Network Intrusion Detection System)がある. 従来の NIDS は, 自身のシグネチャと呼ばれるデータベースとのパターンマッチングにより攻撃を判定する. しかしこの手法では誤警報が多発し実運用に耐えることができない. そこで本研究では, クラスタ解析とニューラルネットワークによる学習を用いた NIDS を考案した. 昨年度の研究では, 本 NIDS の新しいコンセプトを考案し, 幾つかの DoS 攻撃の検知が可能であることを確認した. 今年度は複数の DoS, スキャン攻撃を検知する統一的なパラメータを見出し, これらを用いて11種の DoS, スキャン攻撃に対して検知実験を行った. その結果, 本手法により誤警報をほぼ完全に削減することに成功した.

NIDS for eliminating false positive and detecting unknown DoS attacks

Yousuke Inagaki† ,Haruyuki Furuya† ,Yoshiyasu Takefuji††

NIDS (Network Intrusion Detection System) is a software program for detecting many kinds of attacks including BOF (Buffer Overflow) exploitations, worms, DoS (Denial of Service) attacks, scanning, and so on. Most of NIDS are based on stored signatures, which are stored in a "signature" database with written any attacks in the text style where the system compares the packets against its signature entries. However the signature based approach makes many false positive. In this paper, we have developed the NIDS in the dual way of cluster analysis and training by the neural network. First, we have created a new concept of this NIDS and verified that it could find some DoS attacks. Then, we have discovered the unified parameter set for detecting many attacks. Based on the result of the experiments for detecting DoS and scanning attacks, the proposed approach provides a near-perfect detection and successfully reduces the false positive rate.

1 はじめに

ウィルス, ワーム, DoS, スキャン等サイトに対する不正アクセスは年々増加し, また被害規模も拡大し続けており, その本質的な対策が求められている. ネットワーク内のトラフィックを監視し不正アクセスを検知するツールとして NIDS(Network Intrusion Detection System)がある. 現在市場にある殆どの NIDS はシグネチャ型といわれ, 自身に攻撃のルールが定義してある"シグネチャ"と呼ばれるデータベースを持ち, それらとパケットデータとのパターンマッチングによって攻撃を判定する. しかし, シグネチャ型の NIDS では, 多くの亜種を簡単に作り出すことができる DoS やスキャン攻撃を検知することが困難となる. また, NIDS の重要な問題に誤警報がある. NIDS のシグネチャには攻撃のルールが定義されており, 例えば"データバイト数が 5 バイト以上のパケットを攻撃とみなす"といったルールにある閾値を基に攻撃かどうかを判断する. しかしこれら閾値による判定は,値が閾値を越えたかどうかといった単純なものであるため正常なパケットに対しても攻撃と判定しやすく, そのため誤警報が多発してしまう.

このような問題に対し本研究では, クラスタ解析, ニューラルネットワークによる学習(1)を用いて誤警報を減らし, 未知のDoS, スキャンを検知するNIDSを考案した. 昨年までは大まかなアルゴリズムのベースを開発し, 幾つかのDoS, スキャン攻撃の検知が可能であることを確認した(2). しかしこの段階では, 攻撃毎に検知の設定を変えて検証しており, 汎用性に欠ける面がある. これを受けて今年度は, 実用レベルを目指し広範囲なDoS, スキャン攻撃に対し, 全攻撃の検知に対応する統一的な攻撃の特徴点や, 判定における閾値やサイトに応じたカスタマイズ方法を見出す. また, 誤警報削減と攻撃亜種の検知について既

† キーウェアソリューションズ株式会社 Keyware Solutions, Inc.
†† 慶應義塾大学 環境情報学部 Faculty of Environmental Information, Keio University

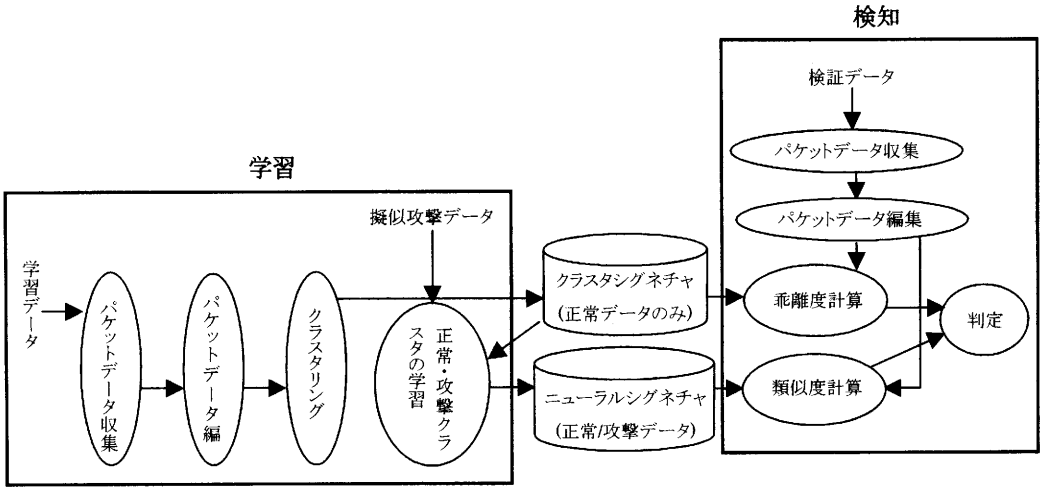

図1 本研究の NIDS 主要機能
Fig.1 Main function of proposed NIDS

存製品との比較を行った.

2 機能概要

本NIDSの主要な機能を図1に示す. 本NIDSは, 統計手法によってパケットヘッダ情報から"内向きのSyn パケット数"等, 設定した条件に合致したパケットの数をカウントする. このような項目を複数個集計し1つのベクトルを生成する. このベクトルを"検知オブジェクト"と呼ぶ. この検知オブジェクトを空間上に表すと普段の正常な場合と攻撃を受けている場合とでは分布状態が変化する. 特にDoSやスキャン攻撃は単位時間において急激にパケット量が増大するため統計的手法を用いると傾向が現れやすくなる. この検知オブジェクトを基に検証時の検知オブジェクトが正常状態や攻撃にどのくらい近いのか判定する. 判定方法はクラスタ解析によって乖離度を求めるものと, ニューラルネットワークのシミュレーションによって類似度を求めるものの2種類を用いる. クラスタ解析では, 予め正常状態のデータをクラスタリングし, クラスタシグネチャを生成する. この正常時のパケットの傾向を基に, 検査すべき検知オブジェクトとの乖離度をベクトル間の距離を計り, 正常状態の集合からどのくらい離れているかを調べて正常, 異常の判定を行う. 一方ニューラルネットワークでは, グループ分けした正常時の検知オブジェクトと擬似攻撃時の検知オブジェクトを学習させて"ニューラルシグネチャ"を生成し, 検証データとそれぞれのグループとの類似度を計算し, 正常, 攻撃クラスタの中でどれに最も近いかを判定する.
但し, 本研究で用いている"クラスタシグネチャ", "ニューラルシグネチャ"は, 従来のNIDSが用いているような攻撃ルールが記載されているテキストベースのシグネチャとは異なり, クラスタ解析とニューラルネットワークの学習によって得られた処理結果を指す.

2.1 検知オブジェクト

NIDS が学習や検知に用いる検知オブジェクトは, 条件(フィルタリング条件)に合致したパケットの数を単位時間毎にカウントし, それらを複数種まとめて1つのベクトルとする. フィルタリング条件には, パケットの向き(外向き, 内向き), 対象ポート番号, サーバ, 通信プロトコル等を設定する.
昨年度の研究では, 攻撃毎に検知オブジェクトの項目(パケット項目)を変更して検証を行っていた. そこで今年度は, 広範囲なDoS, スキャンをまとめて検知することができる統一的なオブジェクトを見出す. 今回, 表1にある11種のDoS, スキャン攻撃を検知する統一的な検知オブジェクトを選定する. 初めに実際に攻撃を仕掛けそのときのネットワークトラフィックをモニタリングし, どのようなパケットが発生するのか調べた. 各攻撃において現れた特徴的なパケットの項目を1つにまとめて, 複数種の攻撃に対応した検知

表1 対象攻撃

Table 1 Focused attacks

攻撃		使用ツール
スキャン	Connect scan	Nmap
	Syn scan	Nmap
	Xmas scan	Nmap
	Fin scan	Nmap
	UDP scan	Nmap
DoS	Synflood	Synパケット送信プログラム
	Smurf	Tfn
	Pingflood	Tfn
	Unreach	Tfn
	UDPflood	Tfn
	Fragment	Synパケット送信プログラム

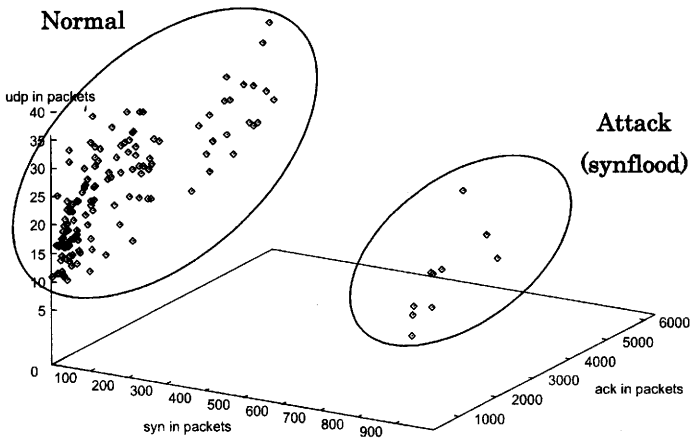

図2 正常と攻撃の分布状態 (Synfloodの場合)

Fig.2 Plots of normal and attack pattern (in case of Synflood)

表3 検知オブジェクト

Table 3 Detecting objects

	名前	説明
1	UDP in	外部から内部へのUDPパケット数
2	Fragment in	外部から内部へのFragmentパケット数
3	ICMP echo in	外部から内部への ICMP echo パケット数
4	ICMP unreach in	外部から内部への ICMP echo パケット数
5	ICMP unreach out	内部から外部への ICMP Unreach パケット数
6	Syn in	外部から内部への Syn パケット数
7	Ack in	外部から内部への Ack パケット数
8	Fpu in	外部から内部への Fin/psh/urg パケット数
9	Fin in	外部から内部への Fin パケット数
10	AR out	内部から外部への AR パケット数
11	TCP K in	外部から内部への TCP K(ポート2000から60000)パケット数
12	TCP K out	内部から外部への TCP K(ポート2000から60000)パケット数

オブジェクト(統一検知オブジェクトと呼ぶ)を設定した. この統一検知オブジェクトが複数種の攻撃を検知しかつ, 正常状態を正常と認識すれば良い. 選定した 12 個の要素からなるベクトルは, 正常時には殆ど 0 となる項目が多いので変動域の大きい代表的な項目を幾つか用いて図 2 のようにグラフ化し, 見た印象で正常と攻撃が分かれているか確かめた. 最終的に決定した検知オブジェクトを表 3 に示す.

2.2 クラスタ解析

クラスタ解析では, 正常状態の検知オブジェクトをクラスタリングし, 検査対象となる検知オブジェクトとの距離を求め, 距離が離れているものを異常と判定する. 今回は, ベクトル各要素の変動域が異なるためユークリッド距離を用いるとスケールの大きな要素に強く影響を受けてしまい正しい評価ができなくなる. そこで距離計算には, マハラノビス距離を用いた. マハラノビス距離は, 分布密度を考慮し密なところほど距離を長く, 疎なところほど距離を短く計るようベクトル各要素のスケールを揃えて計算する. 正常, 異常の判定にはマハラノビス距離から存在確率を求める. 存在確率は, 検査対象が正規分布上の何処に位置するのかを示したものでマハラノビス距離が大きくなれば存在確率は小さくなる.

2.3 ニューラルネットワーク

ニューラルネットワークは, 入力層, 出力層の行列と伝達関数の中間層からなる. 学習させる入力値は, ベクトルで与えられ出力層からの出力値と教師信号の誤差が小さくなるよう入力層と出力層の行列を更新してゆく. 2 つの正常クラスタと 11 個の攻撃クラスタを学習し,学習完了後シミュレーションによってそれぞれのクラスタに対する類似度を求める. シミュレーションでは1つの検知オブジェクトの入力に対して各クラスタ 13 個に対する類似度が出力される. 類似度は 0 から 1 の範囲で表され, 1 に近いほど類似性が高くなる. 学習では正常, 攻撃の検知オブジェクトを与え, それらの出力値と教師信号を比較し出力値が教師信号に近づくように計算する. 教師信号は, クラスタ数と同じ 13 個の要素からなるベクトルで(1,0,0,0,・・・), (0,1,0,0,・・・)のように該当するクラスタを1, それ以外の要素を全て 0 で表す.

3 検証

3.1 正常, 攻撃検知試験

検証では, 本研究の NIDS が, 正しく正常と攻撃を判定するのか確かめる. 実験は, 実環境で収集した正常状態のネットワークのパケットと, 閉じた環境で収集した攻撃のパケットを組み合わせて実環境上で攻撃が行われている状態を擬似的に作り判定した. 判定の基準には, 独自に定義した"誤警報率", "攻撃見逃し率"を用いて認識精度を評価した. 検証結果を表 4 に示す.

3.1.1 認識精度の定義

本検証では正常, 攻撃データをNIDSに投入した際の判定結果は以下の4パターンに分かれる.

(1) 検証用正常データに対し正常と判定したもの:a
(2) 検証用正常データに対し攻撃と判定したもの:b
(3) 検証用攻撃データに対し攻撃と判定したもの:c
(4) 検証用攻撃データに対し正常と判定したもの:d

これらを用いて認識精度を以下のように定義した.

- 攻撃見逃し率=c/(c+d)
- 誤警報率=b/(b+c)

3.2 サイト規模別攻撃試験

検知オブジェクトはパケット数をカウントして生成するため, ネットワークのトラフィック量によってサイトの違いが顕著に現れる. そこで本研究では正常と攻撃のトラフィック量を変えてそれらを組み合わせた場合, 正常と攻撃を区別するか検証した. 結果を表 5~6 に示す.

表4 各攻撃に対する検証結果
Table 4 Experimental results for each attacks

(%)

検査データ		クラスタ解析		ニューラルネットワーク		
		誤警報率	攻撃見逃し率	誤警報率	攻撃見逃し率	攻撃種類の誤認識率
正常データ		0	−	0	−	−
スキャン	Connect scan	−	0	−	0	40 (Syn scan)*
	Syn scan	−	0	−	0	30 (Connect scan)
	Xmas scan	−	0	−	0	−
	Fin scan	−	0	−	0	−
	UDP scan	−	0	−	0	−
DoS	Synflood	−	0	−	0	40 (Fragment)
	Smurf	−	100	−	0	60 (Pingflood)
	Pingflood	−	100	−	0	40 (Smurf)
	Unreach	−	100	−	0	−
	UDPflood	−	0	−	0	−
	Fragment	−	0	−	0	−

＊カッコ内は誤認識した攻撃名

表5 サイト規模.と同規模の検証データを用いた際の検証結果
Table 5 Experimental results in case of same scale of site data and testing data

(%)

サイト規模	誤警報率		攻撃見逃し率	
	クラスタ解析	ニューラルネットワーク	クラスタ解析	ニューラルネットワーク
小規模	0	0	0	0
中規模	0	1.8	0	0
大規模	0	0	0	0

表6 サイトの規模別誤警報率
Table 6 Experimental results of false positive rate for each scale site

(%)

サイト規模	学習用攻撃データ	クラスタ解析	ニューラルネットワーク
小規模	小規模攻撃	0	0
中規模	中規模攻撃	0	1.8
	小規模攻撃	0	1.8
大規模	大規模攻撃	0	0
	小規模攻撃	0	33

3.3 既存 NIDS との比較

誤警報の削減と攻撃亜種の検知において，本研究の NIDS と既存 NIDS の比較を行った．比較には，フリーウェアの Snort を用いた．

3.3.1 誤警報試験

FTP のファイル転送時において，クライアントサーバ間で高いポート番号同士の通信が発生する．これを NIDS は攻撃と誤認識することがある．検証ではこのようなパケットを擬似的に発生させて Snort と比較した．検証した結果，Snort は誤警報を出したのに対し，本研究の NIDS では検証データを全て正常と判定した．

表7 攻撃亜種検知結果
Table 7 Detecting for variation of attacks

攻撃亜種	Snort	本研究の NIDS
Connect scan	×	○
Syn scan	×	○
Synflood	×	○
Smurf	×	○

3.3.2 攻撃亜種の検知

4 種類の DoS, スキャン攻撃において, 攻撃ツールのソースコードを変更して亜種を作り, それらを用いて Snort と本研究の NIDS を比較した. 結果を表 7 に示す.

4 結論

3章で示した三つの検証によって以下のことが確認された.

- 今回用いた検知オブジェクトによって11種の攻撃と正常の判別をすることができた.
- ニューラルネットワーク, クラスタ解析を用いることによって攻撃亜種を検知することができた.
- クラスタ解析で用いたマハラノビス距離は, 変動域が0の要素が含まれると計算できない. 今回の検証では, ICMPに関する項目の変動域が全て0であったためこれらを除外して計算した. そのため一部の攻撃を見逃すこととなった.
- サイトの規模を変えた場合, 学習させる攻撃データも規模に応じて変更することで攻撃を検知することを確認した
- 今回用いたスキャン攻撃は, 10秒間で約1000個のポートを調べるものである. これを各サイトに対して規模を変更せずに学習させたが攻撃見逃しが最大8%しか起こらず, 100倍程度の規模のサイトであればスキャンの検知が可能であることが確認された.
- 既存 NIDS Snort と比較して誤警報を削減し攻撃亜種を検知できることが確認された.

5 まとめ

本研究の目的はクラスタ解析, ニューラルネットワークを用いた NIDS において, 広範囲な DoS, スキャン攻撃を検知する統一的な検知オブジェクトやクラスタ,ニューラル判定での設定値やサイト規模に対するカスタマイズ方法を見出すことであった. 検証の結果, 11 種の攻撃に対してそれぞれの亜種と正常データを正しく判定する検知オブジェクトと学習や検知に用いる設定値を決定することができた. また, サイト規模を変更した場合, 同等の規模の攻撃を学習させれば正常と攻撃の判別ができた. 既存の NIDS と比較して誤警報を削減し, 攻撃亜種を検知することを確認した. 今後の課題として実環境での検証を行い, 規模だけでなく使用しているサービスや Firewall の設定等の違いによる様々な正常パターンに対する検知を行う必要がある. また広帯域に対応した処理性能の検証も必要となる.

6 参考文献

(1) 武藤 佳恭, 斉藤 孝之, 応用事例ハンドブック ニューラルコンピューティング p348 共立出版(株) 2001
(2) Ruo Ando, Yoshiyasu Takefuji, "Two-Stage Quantitative Network Incident Detection for Asynchronous Data Inspection", proceedings of SCI 2003 The 7th World Multiconfernce on Systemics, Cybernetics and Informatics, 2003.