1S-01 ニューラルネットワークを利用した不正アクセス被害予想方式の検討

鴨田 浩明　馬場 達也　小久保 勝敏　松田 栄之
㈱NTT データ　開発本部　技術開発部
e-mail:{kamochan, baba, kokubo, matu}@rd.nttdata.co.jp

1. はじめに

近年 Web サイトでの改竄事件等が相次ぎ、不正アクセスの対策技術が重要視され、不正アクセスの検出を目的とする IDS(Intrusion Detection System) が開発されている。IDS には、misuse detection 方式と、anomaly detection 方式の二種類の検知方式がある。前者は、ネットワーク上のパケットやホスト上のログを signature と呼ばれる既知の不正アクセスの特徴を記述したデータベースと照合することによって不正アクセスを検知する方式である。後者は、正常アクセスの範囲を定義し、その範囲から外れるアクセスを異常として検知する方式である。

misuse detection 方式では、signature として登録されていない未知の手法の攻撃は、検知することが出来ないという問題点がある。一方 anomaly detection 方式では未知の手法の攻撃を検知することが可能になるという利点があることから、現在研究が進められている [1,2]。しかし、anomaly detection 方式では異常として検知されたアクセスが及ぼす被害まではわからない。そこで、実際にどのような被害を及ぼすアクセスであったのかを判別する為には、管理者が手作業で調査を行わなければならないという問題点がある。

本稿では、anomaly detection 方式の IDS により異常と判断されたパケットから、ニューラルネットワークを用いて、被害の種類を予測する方式を提案する。

2. 不正アクセス被害予測方式の検討

2.1. 不正アクセスによる被害の分類

本研究では過去の不正アクセス事例等の調査から、異常なアクセスの被害を表 1 の様に分類し、異常なアクセスにより引き起こされる被害が、表 1 のいずれかに分類されるものとして、被害の予測を行う方式の検討を行った。anomaly detection 方式の IDS において異常と判断されたアクセスが引き起こす被害を予測することが可能となれば、管理者は異常検知後に迅速な対応をすることが出来るようになる。例えば、被害の種類に応じて、メールによる通知のみを行う、ファイアウォールと連携して不正パケットのフィルタリングを行う、発信源追跡システム [3] 等と連携し不正アクセス発信源の自動追跡を行う等の処置を自動的に選択することが可能となる。

表 1: 被害の分類

被害の種類	具体的な被害
実害無し	被害はないが、異常なアクセス
環境情報漏洩	システム稼動の有無や、アクティブなポート、その他ユーザー情報等が取得される
性能低下	システムリソースや、ネットワークの帯域幅等が消費させられる
システムダウン	システムや、サービスが停止させられる
不正操作	権限のないコマンド実行や、ファイルの改竄等が行なわれる

2.2. 被害予測方式の基本概念

インターネットで利用されるプロトコルの仕様は RFC で定められており、プロトコルヘッダのフィールド毎に取り得る値の範囲や、フィールド間の関係等が規定されている。プロトコルヘッダのフィールドに入っている値やフィールド間の関係が、定められた仕様に違反している場合にはホストに被害を及ぼすおそれがある。さらにパケットに含まれるフィールドの内容によっては被害状況との関連が知られており、特定のホストに対して送信されたパケットに含まれるフィールドの内容を分析することで、そのホストが受ける被害を特定することが可能である。

被害の特定は、過去の不正アクセスデータを調査し、不正アクセスパケットのフィールドの内容と被害の関係を厳密にルールとして記述することで、可能となる。しかし、ルールに記述されていない特徴を持つ不正アクセスが行われた場合には被害を特定することが出来ない。そこで、ニューラルネットワークを用い、過去の不正アクセスの特徴箇所と被害の対応を学習させる方式を提案する。不正アクセスパケットのフィールド情報には連続的な値を取るものと、フラグの様に真偽値を取るものがあるが、本方式では、ニューラルネットワークを用いることで、連続値と真偽値の両方を扱うことを可能とした。

2.3. ニューラルネットワークの学習

ニューラルネットワークには、プロトコル毎のヘッ

A Study of a Method to Predict Damage from Unauthorized Access Using Neural Networks
KAMODA Hiroaki, BABA Tatsuya, KOKUBO Katsutoshi, MATSUDA Shigeyuki
Department of Information Technology,
Research and Development Headquarters,
NTT DATA CORPORATION

ダフィールドの情報を、連続値と真偽値に分けて、入力項目として与える。また既知の不正アクセスに関する被害は特定されている為、具体的な被害の種類を表1の分類に従いニューラルネットワークの出力項目として与え、学習を行う。これにより、既知の不正アクセスの特徴と、不正アクセスが及ぼす被害とを結びつけるニューラルネットワークを構築する。

本方式では、過去の不正アクセスと被害の関係を学習することで、異常と判断されたパケットがホストに及ぼす被害を予測する。未学習のデータが入力として与えられた場合でも、過去の例から自動的に類似する特徴を探し出し、最も可能性の高い被害を予測することが可能になる。また、ニューラルネットワークでは再学習を行うことが容易であり、サイト毎の環境に応じた最適なニューラルネットワークを構築することが可能となる。

3. 不正アクセス被害予測方式の有効性検証

本方式の有効性を検証する為、トランスポート層以下のプロトコル (IP/ICMP/TCP/UDP) の仕様に違反した不正アクセスに対象を絞り検証を行った。実験は、CVE[4] に掲載されている不正アクセスを実際に発生させ、不正アクセスパケットの情報を収集した。収集したパケットの情報をニューラルネットワークへ入力できる形式に変換し、表1に示す出力項目とあわせて、バックプロパゲーション法による学習を行った。ニューラルネットワークはプロトコル毎に別々のもの (IP/ICMP, IP/TCP, IP/UDP 用の3つ) を構成した。図1に、IP/TCP 用のニューラルネットワークを示す。

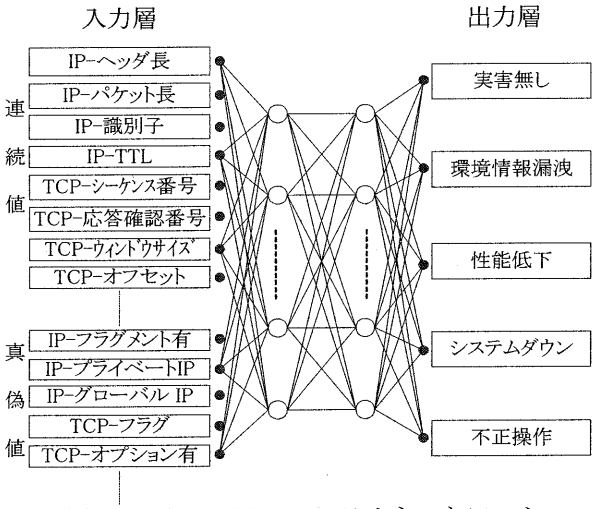

図1. IP/TCP 用ニューラルネットワーク

実験は、leave-one-out 方式を用いて行い、出力として80%以上の確率で算出された被害を、ニューラルネットワークの予測結果とした。実験結果を図2に示す。

実験結果からも分かるように、比較的高い精度

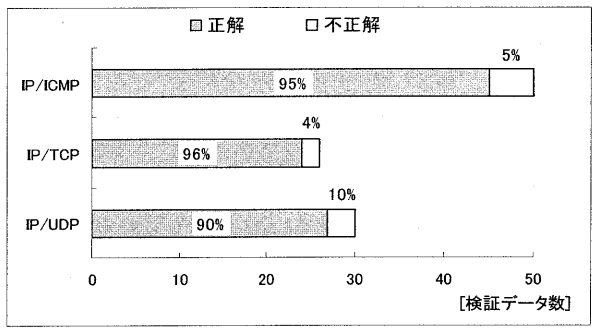

図2. 実験結果

で被害を正しく予測することが出来た。これは、IP/ICMP/TCP/UDP の各プロトコルは、ヘッダフォーマットが厳密に規定されている、フィールド数等は固定であるなど、ニューラルネットワークが被害予測に適した学習モデルであったことが理由であると考えられる。

4. まとめ

本稿では、異常として検知されたパケットをニューラルネットワークを用いることにより、異常パケットが及ぼす可能性のある被害を予測する方式を提案した。また、方式の有効性を検証する為に、トランスポート層以下のプロトコルに対象を絞り実験を行った。その結果、高い正答率が得られ、本方式による被害予測の有効性を示した。

一方で今回の検証では、小規模な環境で実験を行った為、十分なサンプルデータを得られていない。今後、実環境でのデータの収集を行い、検証を進めていく必要がある。また、アプリケーション層プロトコルの仕様に違反した不正アクセスに対しても、その被害を予測することが可能となる方式の検討を行う予定である。

5. 謝辞

本研究は、通信・放送機構（TAO）の委託研究テーマ「不正アクセス発信源追跡技術に関する研究開発」の一環として行われているものである。

参考文献

[1] 馬場達也 他. **不正アクセス検知のためのプロトコルチェック方式の検討**. 情処 61 全大講演論文集 (3), pp.257-258, October 2000.

[2]P.A.Porras and P.G. Neumann. *EMERALD: Event monitoring enabling responses to anomalous live disturbances.* In Proceedings of the 20th National Information System Security Conference, pp.353-354, October 1997.

[3] 竹爪慎治 他. **不正アクセス発信源追跡アーキテクチャの一検討**. 情処 60 全大講演論文集 (3), pp.287-288, March 2000.

[4]MITRE Corporation. *Common Vulnerabilities and Exposures.* http://cve.mitre.org/