ニューラルネットワークを用いた不正アクセス被害予測方式における予測精度の向上

1G−4

鴨田 浩明† 馬場 達也† 小久保 勝敏† 松田 栄之† 矢口 博之‡
㈱NTTデータ 開発本部† 東京電機大学 情報社会学科‡
e-mail:{kamochan, baba, kokubo, matu}@rd.nttdata.co.jp† yag@ia.dendai.ac.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の中でも特に、不正アクセスの特徴（シグネチャ）を使用しないanomaly detection方式のIDSが注目され、現在研究が進められている[1,2]。anomaly detection方式のIDSでは異常として検出されたアクセスが及ぼす被害まではわからない為、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。

そこで、anomaly detection方式のIDSにより異常と判断されたパケットを、ニューラルネットワークを用いて解析することにより、ホストが受ける被害の種類を予測する方式について研究を進めてきた[3]。本稿ではニューラルネットワークを用いた被害予測方式において、特徴選択手法を用いた入力項目の最適化に関する検証結果について報告する。

2. 不正アクセス被害予測方式

これまで、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測方式について検討を進めてきた。以下に不正アクセス被害予測方式の概要を説明する。

被害予測方式とは、過去の不正アクセスの特徴と被害との対応を学習させることにより、被害の種類を予測する方式である。被害の種類を予測するに当たり、過去の不正アクセス事例等の調査から、異常なアクセスにより引き起こされる被害を表1のように分類した。以下にニューラルネットワークを用いた場合の学習の手順を示す。

1. IP/ICMP/TCP/UDP等のプロトコルの各フィールドの特徴を情報としてもつ入力項目を作成する。
2. 表1に示す被害の種類を出力項目とする。
3. 上記の入力項目と出力項目から成るニューラルネットワークを構成する。
4. 被害の種類を判別することが出来る既知の不正アクセスデータを用いて学習を行う。

Improvement of Prediction Accuracy on a Method to Predict Damage from Unauthorized Access Using Neural Networks.
KAMODA Hiroaki†, BABA Tatsuya†, KOKUBO Katsutoshi†, MATSUDA Shigeyuki† NTT DATA CORPORATION†
YAGUCHI Hiroyuki‡ TOKYO DENKI UNIVERSITY‡

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

3. 特徴選択の必要性

これまでの検討では、各プロトコルのほぼ全てのフィールドの情報を用いてニューラルネットワークによる学習を行ってきた。全てのフィールドの情報を入力項目として持たせることの利点として、既知の不正アクセスと特徴の大きく異なる未知の不正アクセスが現れた場合でも、再学習を行うことにより容易に対処が可能である点が挙げられる。一方で既知の不正アクセス、あるいは既知の不正アクセスに特徴が類似した未知の不正アクセスの被害を予測することを考える場合には、不必要な情報が入力項目に含まれていることにより、予測の精度が低下するという問題点もある。

そこで既知の不正アクセスに対象を絞り、予測精度の向上を目的とした入力データの整理を行った。まず各プロトコルの各フィールドの値をドメインエキスパートの知識に基づきメタデータ化するというAI的アプローチを行った。さらにメタデータ化された入力項目に対して特徴選択を行った。特徴選択により不正アクセスの特徴から被害の種類を予測する為に必要な最小個の入力項目が得られる。特徴選択の結果得られた入力項目をニューラルネットワークに与えることで，学習効率や予測精度の向上が期待できる。

4. 特徴選択とニューラルネットワークによる実験

今回、特徴選択の手法としては、CSM(Cartesian Space Model)に基づく特徴選択法[5]を採用することとした。CSMに基づく特徴選択を用いることにより、ニューラルネットワークとAI的アプローチを結びつけ、両手法の利点を活かすことが出来る。

また、CSMに基づく特徴選択を用いて、ニューラルネットワークによる被害予測の精度を向上させる為には、試行錯誤的な作業が必要不可欠となる。これは、

ニューラルネットワークに用いる入力項目の組み合わせを、特徴選択を用いて様々に変化させながら、最も予測精度の高くなる組み合わせを検証していく必要がある為である。そこで実験では以下に示す手順を繰り返し行い予測精度の高くなる組み合わせを検証した。

[実験手順]

1. ニューラルネットワークの入力項目に対して特徴選択を行う。
2. ニューラルネットワークはトランスポート層のプロトコル (ICMP、TCP、UDP) 毎に別々のものを構成する (IP ヘッダの情報は全てのニューラルネットワークに共通に含まれる)。
3. 既知の不正アクセスデータをランダムに 2 組に分け、1 組のデータを用いて学習を行い、残りの未学習のデータを用いて、正しく被害を予測出来るかどうかを評価する。ここで、既知の不正アクセスデータとは、CVE[4] をもとにして人工的に作成したデータである。
4. 上記作業をニューラルネットワーク毎に 3 回行い、その平均を実験結果とする。

今回の実験はトランスポート層以下のプロトコルの特徴を悪用した不正アクセスに対象を絞って行った。ただし、これらの不正アクセスには「不正操作」という被害をもたらすものが存在しない為ニューラルネットワークの出力項目は「不正操作」を除く 4 種類で構成した。

5. 実験結果

特徴選択を行った前後で最も予測精度が向上した場合の入力項目数の変化を表 2 に示す。

表 2: 入力項目数の変化

	ICMP	TCP	UDP
特徴選択前	33	54	39
特徴選択後	22	14	16

全てのニューラルネットワークに共通に含まれる IP ヘッダの情報において、特徴選択を行うことにより絞り込まれた入力項目は、フラグメントオフセット値、パケット長、TTL、送信元アドレス、宛先アドレス、チェックサム、オプションであった。特徴選択の前後での被害予測の実験結果を、表 3, 表 4 に示す。

6. 考察

実験結果に示されているように、特徴選択を行った後では、全ての項目において被害予測の精度が向上すると共に、ICMP、TCP においては、収束するまでの学習回数も減少した。これは、特徴選択を行うことにより、被害を予測する為に必要な入力項目を効率よく絞り込むことが出来た為であると考えられる。

表 3: 実験結果（特徴選択前）

	ICMP	TCP	UDP
実害無し (%)	99.7	100.0	100.0
環境情報漏洩 (%)	96.6	97.5	84.2
リソースの消費 (%)	100.0	86.8	100.0
システムダウン (%)	97.4	98.2	95.4
平均学習回数	494.0	1717.0	1054.7

表 4: 実験結果（特徴選択後）

	ICMP	TCP	UDP
実害無し (%)	100.0	100.0	100.0
環境情報漏洩 (%)	100.0	100.0	91.3
リソースの消費 (%)	100.0	100.0	100.0
システムダウン (%)	97.5	98.7	100.0
平均学習回数	453.7	1004.0	3330.7

一方で UDP においては、特徴選択を行うことにより学習回数が大幅に増加するという結果を得た。これは、安定的な特徴を見出すことが出来なかったことが原因である可能性があり、今後の研究において検証していく必要がある。

7. まとめ

本稿では、これまで検討を進めてきたニューラルネットワークを用いた不正アクセス被害予測方式において、特徴選択を行うことにより予測精度を向上させることが可能であることを示した。

今後は、被害予測方式を組み込んだプロトタイプを作成し、実環境での評価を行っていく予定である。また、アプリケーション層の各プロトコルにおいても、異常アクセスがホストにもたらす被害を予測する方式の検討を行う予定である。

8. 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] 鴨田浩明 他. **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**. 情処 62 全大講演論文集 (3), pp.283-284, March 2001.

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

[5] 矢口博之. **カルテシアン空間モデルに基づく知識獲得支援システム**. 人工知能学会誌 Vol.11 No.1, pp75-85, January 1996.