

# 不正侵入検知におけるセキュリティアラートの シグネチャ別発生量に着目した誤検知削減手法

岩崎 信也<sup>†1</sup> 角田 朋<sup>†1</sup> 関口 悦博<sup>†1</sup> 小西幸洋<sup>†2</sup> 大鳥 朋哉<sup>†2</sup> 薦田 憲久<sup>†3</sup>

**概要:** 近年、企業や公的機関などのサイバーセキュリティは重要度を増しており、インシデント対応の迅速化が求められている。また、攻撃の巧妙化、脆弱性の発見、攻撃トレンドが絶えず変化することから、パターン定義による不正検知が難しくなり、教師データによる学習を必要としない教師なし異常検知型手法の重要性が増している。本研究では、不正侵入検知における誤検知を、機械的に削減する手法を提案する。提案手法では、アラートごとに、過去一定時間の当該アラートと同一の攻撃元 IP アドレスのアラートを抽出する。抽出したアラートからシグネチャの発生量をパターンとして集計し、クラスタ分析手法により同様のパターンもつアラートを分類し、「攻撃」と「問題なし」を判定する。アラートに提案手法を適用した結果、再現率 100%、適合率は 14%、F 値は 24%となった。

**キーワード:** サイバーセキュリティ、インシデント対応、クラスタ化、DBSCAN、異常検知

## 1. はじめに \*

近年、企業や公的機関などのサイバーセキュリティは重要度を増しており、インシデント対応の迅速化が求められている。インシデント対応では、セキュリティ機器が発生させるアラートを調査する。このアラートは多量かつその大部分が、実際には正常な通信であり、誤検知であることが分かっている[1]。インシデント対応の迅速化として、誤検知のアラートを機械的に削減する試みが行われている。本研究では、セキュリティ機器が発生させるアラートのシグネチャに着目し、教師なし異常検出手法を利用することで、誤検知を削減する。

## 2. 誤検知の削減

セキュリティアラートは、侵入検知装置やファイアウォールなどのセキュリティ機器が、ネットワークトラフィックを監視し、攻撃の可能性が高いイベントを検知した際に発生させるものである。図 1 はセキュリティ機器が発生させたアラートの例である。シグネチャとは、検知したアラートが、どのような攻撃の可能性があるかを示している。

```
0123456,THREAT,wildfire-virus,1,2017/06/29 21:13:11,203.0.113.3,192.168.12.201,  
発生時刻 攻撃者IPアドレス  
0.0.0.0,0.0.0.0,From_Internet_To_MTA_SMTP,,smtp,vsys1,Untrust,  
DMZ,ethernet1/1,ethernet1/2,MSS,2017/06/29 21:13:11,347809,1,28943,  
25,0,0,0x2000,tcp,drop,"a.xls" Virus/Win32_VGeneric.any,medium,  
シグネチャ  
client-to-server,2289,0x0,PA,192.168.0.0-  
192.168.255.255,0,0,0,0,0,0,InternetFW,I000PAL01
```

図 1 アラートログの例

アラートは、少しでも攻撃の可能性があると発生するが、実際の攻撃につながるのはいずれであり、攻撃ではない正常な通信を誤って攻撃と検知する誤検知が多い。このため、

セキュリティ機器がアラートを発生させた際に、誤検知のアラートを機械的に削減する試みが行われている。削減手法の一つに、教師なし異常検知手法がある[2]。教師なし異常検知手法は、アラートが多数の誤検知と少数の攻撃から成ると仮定し、似たような傾向を持つアラートと、ユニークな傾向を持つアラートを分類する手法である。この方法は、入力アラートのみから分類することができ、他の異常検知手法と比べ精度が劣るが、事前に教師データなどで学習する必要がない。セキュリティ分野では、攻撃の巧妙化、脆弱性の発見、トレンドの変化などで、攻撃の傾向が比較的短期間で変化することから、教師なし異常検知手法を利用した研究が盛んである[3]。そのため、本研究では、教師なし異常検知手法を利用する。

## 3. シグネチャ発生量に着目した異常検知手法

近年の攻撃は、複数の攻撃手法を組み合わせることが多く、一度の攻撃で、多種のシグネチャのアラートを発生させる。このため、「攻撃」につながるアラートのシグネチャ別発生量は、「問題なし」のアラートとは違う傾向を持つと考えられる。そこで、本研究では、セキュリティ機器が発生させたアラート群に対し、発生したアラート 1 件ごとに、過去一定時間の同一攻撃者 IP アドレスのアラートを抽出し、抽出したアラートのシグネチャ別発生量から、検討すべきアラートを削減する手法を提案する。教師なし異常検出手法により、アラートのシグネチャ別発生量を分類することで、攻撃・誤検知のアラートを判定し、誤検知を削減できると期待できる。

具体的には以下の手順となる。

- (1) アラートごとに過去一定時間同攻撃者のアラート抽出対象となるアラートごとに、以下の複合条件でアラート郡からアラートを抽出する。

- ・アラート発生時刻から過去一定時間[tw]内のアラート
- ・アラートの攻撃者 IP アドレスが同一

- (2) シグネチャ別の発生量の集計

抽出したアラートから、シグネチャ別の発生量を集計す

\*†1 株式会社日立システムズ 研究開発本部

Hitachi Systems, Ltd. Research & Development Division

†2 株式会社日立システムズ ネットワークセキュリティサービス事業部  
Hitachi Systems, Ltd. Network Security Services Division

†3 コーデソリューション株式会社  
Code Solutions Co., Ltd.

る.例えば,図 2 において,アラート 4 を対象としたとき,同一攻撃者かつ,過去一定時間内のアラート 1,アラート 2 が集計対象となる.これらのアラートのシグネチャ別の発生量をカウントすると,シグネチャ A が 1 回,シグネチャ B が 2 回となる.対象となるアラートごとに,同様に全てのアラートで,シグネチャの発生量を集計する.

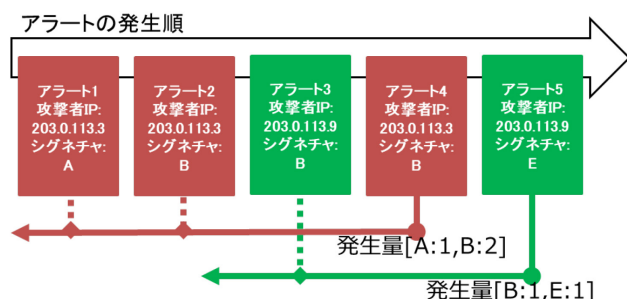


図 2 アラートのシグネチャ別の集計

### (3) アラートの分類

(2)の集計結果を元に,クラスタリング手法の一つである DBSCAN[4]を利用して,アラートを分類する. DBSCAN は,入力データの密度により,クラスタを構成するアルゴリズムである.距離閾値と,クラスタの最小対象数をパラメータとし,機械的に最適なクラスタ数を決定する.また,入力データの内,どのクラスタにも分類できないものを,外れ値として分類することができる.

### (4) 判定

本手法では,(3)の結果,クラスタに分類されたアラートを「問題なし」と判定し,外れ値に分類されたアラートを「攻撃」と判定する.これにより,同様のシグネチャ発生量のパターンが多数存在するアラートは,「問題なし」として判定する.

## 4. 評価

実際にネットワークを監視しているセキュリティ機器 2 台のアラート 7 日分をデータセットとして,提案手法を適用し誤検知の削減を試みた.対象のデータセットは表 1 である.ここでの「攻撃」・「問題なし」は,セキュリティ分析担当者が,対象のアラートを判定したものである.

表 1 データセットの分析担当者による判定結果

| データセット | 攻撃 | 問題なし  |
|--------|----|-------|
| A      | 7  | 7023  |
| B      | 9  | 13223 |

評価では,分析担当者と提案手法による判定結果と比較し,再現率・適合率・F 値を算出した.なお,評価時の過去一定時間[*tw*]は,セキュリティ分析担当者の意見の下,30 分とした.

本提案手法を適用した結果が,表 2,表 3 となった.

表 2 提案手法による判定結果

|             |          | データセット A |      | データセット B |       |
|-------------|----------|----------|------|----------|-------|
|             |          | 攻撃       | 問題なし | 攻撃       | 問題なし  |
| 提案手法<br>の判定 | 攻撃       | 7        | 96   | 9        | 32    |
|             | 問題<br>なし | 0        | 6927 | 0        | 13191 |

表 3 提案手法の評価

|     | データセット A | データセット B | 平均   |
|-----|----------|----------|------|
| 再現率 | 1.00     | 1.00     | 1.00 |
| 適合率 | 0.07     | 0.22     | 0.14 |
| F 値 | 0.13     | 0.36     | 0.24 |

表 2,表 3 のとおり,データセット A, B 合わせて 16 個の攻撃アラートを,全て適切に攻撃と判定し,攻撃の再現率は 100%となった.ユニークなシグネチャパターンのアラートが外れ値として分類され,攻撃を正しく判定できたと考えられる.再現率が 100%であることは,サイバーセキュリティでは,攻撃を見逃さないため重要となる.しかし,誤って「問題なし」を「攻撃」と判定したのは,128 件である.これは攻撃アラート数の 8 倍になり,攻撃の適合率は平均して 14%,F 値は 24%となった.この要因は,「問題なし」のアラートに,通常は発生せずユニークであるが正常な通信を検知したアラートが含まれているためと考えられる.

## 5. まとめ

本研究では,不正侵入検知における誤検知を,機械的に削減する手法を提案した.提案手法では,アラートごとに,過去一定時間の当該アラートと同一の攻撃元 IP アドレスのアラートを抽出する.抽出したアラートからシグネチャ別の発生量を集計し,クラスタ分析手法により,同様のパターン进行分类し,「攻撃」と「問題なし」に判定した.アラートに提案手法を適用した結果,再現率 100%,適合率は 14%,F 値は 24%となった.

## 参考文献

- [1] 藤田直行,“侵入検知に関する誤検知低減の研究動向,” 電子情報通信学会論文誌 B, vol. 89, no. 4, pp. 402–411, 2006.
- [2] D. E. Denning, “An Intrusion-Detection Model,” *IEEE Trans. Softw. Eng.*, vol. SE-13, no. 2, pp. 222–232, Feb. 1987.
- [3] 巽康平, 秋山豊和, “トラフィックパラメータのクラスタリングによる異常検知手法の特性調査,” 研究報告インターネットと運用技術 (IOT), vol. 2013, no. 5, pp. 1–6, 2013.
- [4] M. Ester, M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” *Proc. Second Int. Conf. Knowl. Discov. Data Min.*, pp. 226–231, 1996.