IDSアラートに対する誤検知削除方法の提案とその評価

吉村　尚人[1,a)]　池上 雅人[2]　長谷川 智久[2]　原田　隆史[2]　北谷　浩[2]　森井　昌克[1,b)]

概要：ネットワークを不正アクセスから守る手段の一つとして侵入検知システム (Intrusion Detection System: IDS) がある．IDS はシステムやネットワークを監視することで，不正アクセスの兆候を検知し，管理者に通知するシステムである．IDS は多くの誤検知 (false positive) を含む膨大な量のアラートを発することが知られており，その中から実際に行われた不正アクセスを見つけ出し，対策を講じなければならない．そのため，IDS を利用するには高度な専門知識が要求される．しかし，多くのセキュリティ担当者はセキュリティインシデントに関する十分な知識を有していないという現状がある．IDS の発するアラートを削減する手法の一つに Spathoulas らのアラートの出現頻度に関する特徴を用いた手法がある．本稿ではネットワーク型 IDS(NIDS) を対象に，アラートの特徴およびクラスタリングを用いたアラートの削減手法を提案し，IDS の利用を援助することを目的とする．提案手法では既存手法よりも高い精度でのアラートの削減に成功した．

キーワード：IDS，アラート，誤検知，DBSCAN，ネットワークセキュリティ

Proposal and Evaluation of False Positives Reduction Methods for IDS Alerts

Naoto Yoshimura[1,a)]　Masato Ikegami[2]　Tomohisa Hasegawa[2]　Takafumi　Harada[2]
Hiroshi Kitani[2]　Masakatu Morii[1,b)]

Abstract: Intrusion Detection System (IDS) is one of the means to protect the network from unauthorized access. IDS detects signs of unauthorized access and notifies administrators by monitoring systems and networks. IDS is known to produce a large number of alerts including many false positives, and it is necessary to find out the unauthorized access actually happened and respond to it. Therefore, abounding knowledge is required to use IDS. However, many security personnel do not have sufficient knowledge about security incidents. Spathoulas and Katsikas proposed the method to reduce the alerts generated by IDS uses the features of the frequency of the alerts. In this paper, we propose alert reduction methods using the features of alerts and clustering for Network IDS to support the use of IDS. Proposed methods succeeded in reducing alerts with higher accuracy than the existing method.

Keywords: IDS，Alert，False Positive，DBSCAN，Network Security

1. はじめに

インターネットサービスの普及に伴い，企業でも電子商取引や顧客情報をはじめとする機密情報の管理などにインターネットが利用されている．これらはサービスの利便性を向上させる一方で，不正アクセスなどのサイバー攻撃の増加を招いている．

ネットワークを不正アクセスから守る手段の一つとして侵入検知システム (Intrusion Detection System: IDS) がある．IDS はシステムやネットワークを監視することで，不正アクセスの兆候を検知し，管理者に通知するシステムで

[1] 神戸大学
Kobe University
[2] キヤノンマーケティングジャパン株式会社
Canon Marketing Japan Inc.
[a)] yoshimura@stu.kobe-u.ac.jp
[b)] mmorii@kobe-u.ac.jp

ある．IDS は多くの誤検知 (false positive) を含む膨大な量のアラートを発することが知られており，その中から実際に行われた不正アクセスを見つけ出し，対策を講じなければならない．そのため，IDS を利用するにはネットワークやセキュリティインシデントに関する高度な専門知識が要求される．しかし，多くのセキュリティ担当者は十分な知識を有していないという現状がある．

この問題を解決するために，IDS の発するアラートの中から実際の攻撃により生成されたアラートのみを抽出し，正常な通信に対して生成された誤検知アラートを削除する手法が提案されている．Spathoulas，Katsikas（2010）はアラートの出現頻度に関する特徴を用いた誤検知削除手法を提案している [1]．Spathoulas らは DARPA 1999 データセットと IDS の一種である Snort を用いた実験の結果，生成されたアラートから 75%の誤検知アラートを削減した．この手法では単一のアラートが持つ情報だけでなく，複数のアラートから得られる情報を，誤検知の判定に用いている．しかし，著者らがこの手法の実装および誤検知削除性能の評価を行った結果，実際の攻撃により生成されたアラートの見逃しを少なくしようとすると，削除できる誤検知アラート数が小さくなることがわかった．一般に IDS の利用においては，ある程度誤検知を許容してでも攻撃の見逃しをなくすことが求められる．そのため，攻撃アラートの見逃しが少ない条件での誤検知削除の性能が重要となる．そこで本稿では Spathoulas らの手法に加え，クラスタリングを用いたフィルタを追加することで，既存手法よりも高い精度でのアラート削減手法を提案する．ISCXIDS2012[2]，CICIDS2017[3] の 2 つのデータセットを用いた評価の結果，攻撃アラートの見逃しが少ない条件で既存手法よりも多くの誤検知アラートの削減に成功した．

2. 侵入検知システム（Intrusion Detection System: IDS）

侵入検知システム（Intrusion Detection System: IDS）とはシステムやネットワークを監視することで，不正アクセスの兆候を検知し，管理者に通知するシステムである．

IDS は監視対象と検知方式によって分類される．まず監視対象によりネットワーク型 IDS（NIDS）とホスト型 IDS（HIDS）に分類される．NIDS はネットワーク上の通信を監視し，疑わしい通信があれば管理者に通知する．そのため複数台数の端末の監視に適している．HIDS は各ホスト（サーバやクライアント）のイベントを監視し，正常でないイベントを確認すると管理者に通知する．HIDS は NIDS に比べ誤検知率が低いことが知られている．

また検知方式によりシグネチャ型（不正検出型）とアノマリ型（異常検出型）に分類される．シグネチャ型の IDS は，あらかじめ用意されている検出ルールを利用し，ルールと一致した通信を検知し管理者に通知する．シグネチャ型の IDS は既知の攻撃の検知に優れている反面，ルールの定義されていない未知の攻撃を検知できない．アノマリ型の IDS は正常な状態を定義しておき，その定義から外れる状態を検知し管理者に通知する．アノマリ型の IDS はシグネチャ型に比べ誤検知が多いが，未知の攻撃を検知できる可能性がある．

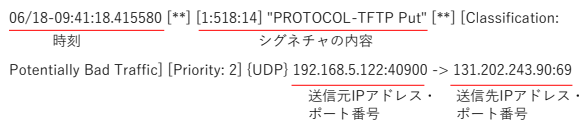

図 1 Snort のアラート
Fig. 1 alert of Snort

本稿ではシグネチャ型の NIDS を対象とした手法を提案する．一般的なシグネチャ型の NIDS が生成するアラートは時刻，シグネチャの内容，送信元 IP アドレス・ポート番号，送信先 IP アドレス・ポート番号などの情報を持つ．図 1 に NIDS の一種である Snort のアラートを例として示す．

3. 関連研究

IDS の誤検知削除手法に関する研究は広く行われている．Hubbslli と Suryanarayanan（2014）はシグネチャ型 NIDS の誤検知削除手法について調査し，まとめた [4]．Hubbslli らによると，誤検知削除手法は大きく以下の 9 つに分類される．

Signature Enhancement: 過去の検知履歴やネットワーク情報などの環境情報をシグネチャの検出ルールに加えることでシグネチャの精度を向上させる．

Stateful Signature: 従来の IDS は単一のネットワークパケットをシグネチャデータベースと照合し検知を行うが，複数のパケットにわたる攻撃の検知に失敗する．そこで過去の状態を保持することで複数のパケットに対して検知を可能にする．

Vulnerability Signature: 一般に IDS のシグネチャはストリングマッチングや正規表現が用いられるが，これらを掻い潜るような攻撃も増えている．これらの攻撃を検出するために，与えられた脆弱性に対して考えうる攻撃を検知するシグネチャをセマンティックを用いて作成する．

Alarm Mining: IDS の生成するアラートに含まれる IP アドレス，ポート番号，プロトコルなどの情報から誤検知であるアラートとそうでないものを定義付けるアラート群を見つけ出す．そのための手法として，クラスタリング，分類，ニューラルネットワーク，頻出パターンマイニングなどが用いられる．

Alarm Correlation: IDS のアラートが持つ情報は，そこからどのような攻撃が行われたかを判断するには不

十分な場合が多い．そのため相関のあるアラートを集約して攻撃シナリオを再構築する．

Alarm Verification: 検証メカニズムを用いて，行われた攻撃が成功したか，また基礎となる攻撃がターゲットネットワークへ影響を及ぼしたかを検証する．

Flow Analysis: IDS は攻撃の無い通常時でもアラートを生成するため，通常時のアラートと異常時のアラートの違いを分析する．

Alarm Prioritization: 各アラートに優先順位を与える．優先順位を与えるのに，ネットワークトポロジや，IDS の履歴，IDS の配置などが用いられる．

Hybrid Methods: 実際のネットワークでは，すべての手法が常に有効とは限らないため，フィルタリング手法とデータマインイング手法を組み合わせた手法を用いる．

Alarm Mining に属する手法として，Liang ら（2015）の K-Means, Fuzzy C-Means（FCM）を用いた手法がある [5]. Liang らはクラスタリングにより誤検知であるアラートが含まれるクラスタと，真に攻撃によるアラートが含まれるクラスタを分離することを試みた．Liang らは DARPA 2000 LLDOS1.0 を用いた実験を行った．評価指標は誤検知の削除率を示す ER，攻撃アラートの見逃し率を示す F-E-R，削除できなかった誤検知の割合を示す M-E-R の 3 つである．K-Means では ER が 64.90%，F-E-R が 6.77%，M-E-R が 35.11%となり，FCM では ER が 80.77%，F-E-R が 16.58%，M-E-R が 19.23%となった．また，岩崎ら（2018）は DBSCAN を用いて，教師なし異常検知による IDS の誤検知削除手法を提案した [6]．岩崎らは攻撃が複数の攻撃手法の組み合わせからなり，実際の攻撃により生じたアラートのシグネチャ別発生量は誤検知であるアラートとは異なる傾向を示すと考え, 一般的な IDS から生成されるアラートは多数の誤検知と少数の攻撃アラートにより構成されることから, アラートに対して異常検知を行うことで攻撃アラートのみを抽出し，誤検知を削除する手法を提案した．岩崎らはこの手法により実際にネットワークを監視して得られたアラートの削除を行い，再現率 100%，適合率 14%，F 値 24%を達成した．

4. 誤検知アラートの削除

本章では，4.1 節で Spathoulas らのアラートの出現頻度を用いた削除手法について述べ, 4.2, 4.3 節では Spathoulas らのフィルタにクラスタリングを用いたフィルタを追加した 2 つの削除手法をそれぞれ提案する．

4.1 アラートの出現頻度を用いた削除法

Spathoulas らにより提案されたアラートの出現頻度を用いた削除手法について説明する．この削除手法はフィルタを用いてアラートの削除を行う．フィルタは 3 つのコンポーネントで構成されており，それぞれのコンポーネントはアラートの調査から得られた，実際の攻撃と正常な通信におけるアラートの出現頻度に関する特徴を利用している．それぞれのコンポーネントで利用される特徴を以下に示す．

NRA（Neighboring Related Alerts）： 実際の攻撃により，送信元/送信先 IP アドレスに類似度を持つひとまとまりのアラート群が生成される．

HAF（High Alert Frequency）： 実際の攻撃により，同一のシグネチャにより生成されたアラートが異常な分布を生じる．

UFP（Usual False Positives）： 誤検知はシグネチャが誤検知を引き起こす頻度で識別できる．

それぞれのコンポーネントでは各アラートが攻撃アラートである可能性を示す評価値を算出する．各アラートが攻撃であるかの判定は，それぞれのコンポーネントで算出された評価値の最大値，最小値，平均値の 3 つのうちいずれかと設定した閾値 th を比べることで行う．閾値を越えなかったアラートについては削除する．

4.1.1.-4.1.3 では各コンポーネントにおける評価値の計算方法について述べる．

4.1.1 NRA の算出法

NRA の評価値は実際の攻撃により，送信元/送信先 IP アドレスに類似度を持つひとまとまりのアラート群が生成されるという特徴を利用して算出される．評価値を算出するにあたり，2 つのパラメータ t_0，l を定義する．t_0 は隣接するアラートを数える際に利用するタイムウィンドウのサイズを表し，l はカウントしたアラート数を 0 から 1 の実数値で表わされる評価値に変換する際に用いる閾値を表す．

各アラート α (i) に対し次の条件を満たすアラート α (j) の数 $\mathrm{nra_i}$ をカウントする．ただし，$\{\mathrm{sip}, \mathrm{dip}\}$ は送信元 IP アドレス sip と送信先 IP アドレス dip の組を表す．

条件 1．$|t_j - t_i| \leq t_0$

条件 2．$\{\mathrm{sip}_i, \mathrm{dip}_i\} \cap \{\mathrm{sip}_j, \mathrm{dip}_j\} \neq \emptyset$

これより定義した閾値 l を用いて, アラート α (i) に対するこのコンポーネントの評価値 b_i^{NRA} は

$$b_i^{\mathrm{NRA}} = \frac{\min(\mathrm{nra}_i, l)}{l} \tag{1}$$

で表される．しかし，Spathoulas らによるとこの評価値は IP スウィープ攻撃やポートスキャンを識別できなかった．そのため Spathoulas らはこれらの攻撃を識別できるような評価値 b_i^{NRAsweep} を用意し，この値が b_i^{NRA} よりも大きい場合 b_i^{NRAsweep} をこのコンポーネントの評価値とした．

$$\mathrm{NRA}_i = \max(b_i^{\mathrm{NRA}}, b_i^{\mathrm{NRAsweep}}) \tag{2}$$

b_i^{NRAsweep} は b_i^{NRA} におけるカウントするアラートの条件 2 を以下の条件 2' のように変更した方法で算出する．

条件 2'．条件 2'.1-条件 2'.4 のうちいずれかを満たす．

- 条件 2'.1. $\mathrm{sip}_i \equiv \mathrm{sip}_j$ and $\mathrm{sub}(\mathrm{dip}_i) \equiv \mathrm{sub}(\mathrm{dip}_j)$
- 条件 2'.2. $\mathrm{sip}_i \equiv \mathrm{dip}_j$ and $\mathrm{sub}(\mathrm{dip}_i) \equiv \mathrm{sub}(\mathrm{sip}_j)$
- 条件 2'.3. $\mathrm{dip}_i \equiv \mathrm{sip}_j$ and $\mathrm{sub}(\mathrm{sip}_i) \equiv \mathrm{sub}(\mathrm{dip}_j)$
- 条件 2'.4. $\mathrm{dip}_i \equiv \mathrm{dip}_j$ and $\mathrm{sub}(\mathrm{sip}_i) \equiv \mathrm{sub}(\mathrm{sip}_j)$

ただし，sub(ip) は IP アドレスが所属するサブネットのネットワークアドレスを表す．

4.1.2 HAF の算出法

HAF の評価値は，実際の攻撃によるアラートは，同一のシグネチャによって生成されたアラートにおいて異常な分布を生じるという特徴を利用して算出される．シグネチャに関するアラートの出現頻度が攻撃によるアラートでは通常の通信による誤検知アラートに比べ大きくなることに基づいて評価値を計算する．評価値を算出するにあたり，パラメータ m を定義する．m は求めた頻度を 0 から 1 の実数値で表わされる評価値に変換する際に用いる閾値を表す．

各アラートにおいて最小時間差 mtd_i を計算する．mtd_i は各アラートと，そのアラートと共通のシグネチャにより生成されたアラートで最も時間的な距離が小さいものとの時間差と定義する．共通のシグネチャにより生成されたアラートが存在しない場合は $\mathrm{mtd}_i = \infty$ とする．mtd_i を用いてシグネチャに関するアラートの出現頻度は次のように算出される．

$$f_i = \frac{1}{1 + \mathrm{mtd}_i} \tag{3}$$

シグネチャごとに出現頻度の平均値を計算し，$\mathrm{sfa}_{\mathrm{sid}}$ とし，評価に用いる頻度 nf_i を以下のように求める．

$$\mathrm{nf}_i = \frac{f_i}{\mathrm{sfa}_{\mathrm{sid}_i}} \tag{4}$$

最後に，m を用いてこの頻度を 0-1 の値に変換して，このコンポーネントの評価値は以下のようになる．

$$\mathrm{HAF}_i = \frac{\min(\mathrm{nf}_i, m)}{m} \tag{5}$$

4.1.3 UFP の算出法

UFP の評価値は，誤検知はシグネチャが誤検知を引き起こす頻度で識別できるという特徴を利用して算出される．評価値を算出するにあたり，パラメータ n を定義する．n は求めた頻度を 0 から 1 の実数値で表わされる評価値に変換する際に用いる閾値を表す．評価値の算出には，攻撃のない期間を利用する．攻撃のない期間において，アラート全体に対し各シグネチャにより生成されたアラートが含まれる割合を計算し，$f^{\mathrm{af}}_{\mathrm{sid}}$ とする．同様に，実際に判定を行うデータを一定時間ごとに区切り，区切られたデータごとに各シグネチャにより生成されたアラートが含まれる割合を計算し，$f^{\mathrm{c}}_{\mathrm{sid}}$ とする．最後に，n を用いてこのコンポーネントの評価値は以下のようになる．

$$\mathrm{UFP}_i = \frac{\min(\frac{f^{c}_{\mathrm{sid}_i}}{f^{\mathrm{af}}_{\mathrm{sid}_i}}, n)}{n} \tag{6}$$

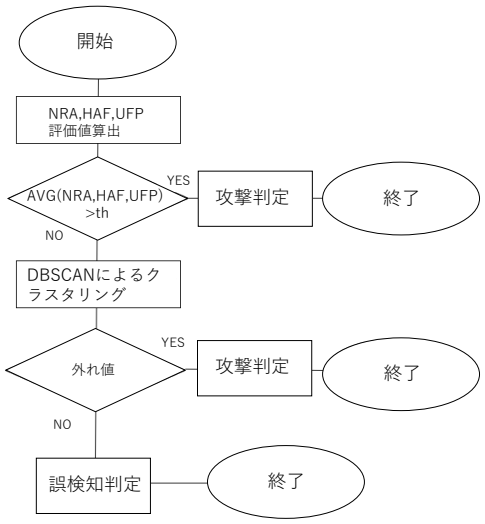

図 2 DBSCAN を用いた削除法
Fig. 2 Reducing method using DBSCAN

4.2 DBSCAN を用いた削除法

本節では 4.1 節で述べた Spathoulas らの手法にフィルタを追加することで見逃し率を小さくした場合での精度向上を可能にする手法を提案する．新たに追加するフィルタには DBSCAN を用い，Spathoulas らのフィルタで攻撃と判定されなかったアラートに対して適応することで，攻撃アラートの見逃しを削減する．このフィルタは，一般的な IDS のアラートにおいて，実際の攻撃によるアラートの割合は低いことを利用し，外れ値を検知することで，攻撃アラートを抽出することを目的としている．そのために外れ値を検知できるクラスタリングアルゴリズム DBSCAN を用いる．DBSCAN は密度をもとにしてクラスタリングを行い，クラスタを構成する際の閾値となる最大半径と最小の点数を決定することで，クラスタに分類されない外れ値を検知することができる．

DBSCAN を適用する際の特徴量として [シグネチャ ID，送信元 IP アドレス，送信元ポート番号，送信先 IP アドレス，送信先ポート番号] を用いる．ただし，IP アドレスに関してはドット付き 10 進数から，10 進数への変換をう．この手法では，Spathoulas らのフィルタで攻撃と判定されたものおよび Spathoulas らのフィルタで攻撃と判定されなかった中から DBSCAN で外れ値として検知されたものを攻撃と判定し，その他のアラートを削除する．この手法で行う処理のフローを図 2 に示す．

4.3 K-Means と DBSCAN を用いた削除法

4.2 節で述べた DBSCAN を用いた誤検知削除法は，4.1 節で述べた既存の手法に比べ，攻撃アラートの見逃し率が低い場合に，より精度の高い誤検知削除性能を示した．しかし，DBSCAN によるクラスタリングでは，外れ値に分類されない攻撃や，外れ値に分類される誤検知がみられ，より厳密な判定法が望ましいと考えた．そこで，4.1 節で

説明したフィルタによる判定結果を，後のフィルタでの判定時に考慮することで精度を向上させる方法を提案した．この削除法では実際の攻撃により生じたアラートは，シグネチャや，IP アドレス，ポート番号に類似性があると考えた．削除手法で行う処理の流れを示す．

ステップ 1．すべてのアラートを対象に K-Means によるクラスタリング

ステップ 2．アラートの出現頻度を用いたフィルタ（4.1 節）

ステップ 3．K-Means で得られた各クラスタにおいてステップ 2 で攻撃と判定されたアラートの割合 AR を算出

ステップ 4．DBSCAN のクラスタリング結果と K-Means で得られた各クラスタの AR を用いた判定

それぞれのステップで行う処理について説明する．

4.3.1 ステップ 1

このステップでは，実際の攻撃により生じたアラートはシグネチャ，IP アドレス，ポート番号に類似性があると考え，攻撃アラートの多いクラスタと少ないクラスタを作成する．クラスタリングアルゴリズムには K-Means，特徴量にはアラートの持つ [シグネチャ ID，送信元 IP アドレス，送信元ポート番号，送信先 IP アドレス，送信先ポート番号] を用いた．K-Means のアルゴリズムを以下に示す．

(1) 初期値として k 個のクラスタ中心を用意する．
(2) 各点を中心との距離が最も近いクラスタへ割り振る．
(3) 各クラスタの中心を求める．
(4) クラスタ中心の変化がなくなるまで（2）と（3）を繰り返す．

4.3.2 ステップ 2

このステップでは，4.1 節で述べた Spathoulas らのアラートの出現頻度を用いたフィルタによるアラートの攻撃判定を行う．このステップで評価値が閾値を超えたアラートについては攻撃であると判定する．閾値を超えなかったアラートについては次のステップに進む．

4.3.3 ステップ 3

このステップではステップ 1 で得られたクラスタごとに，ステップ 2 で攻撃と判定されたアラートが含まれる割合 AR を算出する．AR は各クラスタに含まれるアラートが攻撃によるアラートである可能性を示す指標となる．このステップでは，AR の値が極端に大きくなるか小さくなるクラスタが得られることが理想となる．

4.3.4 ステップ 4

このステップでは，ステップ 2 で攻撃と判定されなかったアラートに対して DBSCAN によるクラスタリングを行い，その結果とステップ 3 で求めた AR を用いて，攻撃であるかの判定を行う．

$$b = \begin{cases} \mathrm{AR} & (\text{クラスタ＝外れ値}) \\ \mathrm{AR} - 1 & (\text{その他}) \end{cases} \tag{7}$$

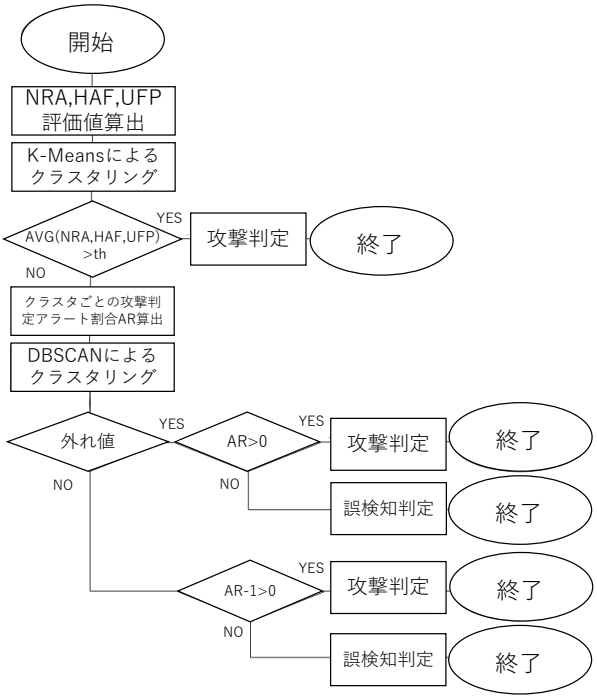

図 3 K-Means と DBSCAN を用いた削除法
Fig. 3 Reduction method using K-Means and DBSCAN

$$\text{判定} = \begin{cases} \text{攻撃} & (b > 0) \\ \text{誤検知} & (\text{その他}) \end{cases}$$

ステップ 2 で攻撃と判定されたアラートとステップ 4 で攻撃と判定されたアラートを除いたアラートを誤検知と判定して削除する．

5. 評価

既存手法である Spathoulas らによるアラートの出現頻度を用いた削除手法と，本稿で提案する DBSCAN を用いた手法および K-Means と DBSCAN を用いた手法について実装および評価を行った．データセットには ISCXIDS2012[2]，CICIDS2017[3] の 2 つを利用した．両データセットはともに pcap ファイル，ラベル付きフローデータを含んでいる．IDS によるアラートは NIDS の一種である Snort[7] を用いて，データセットの pcap ファイルから生成されたものを用いた．また出現頻度を用いた削除法において必要な，攻撃のない期間のアラートは両データセットにおいて攻撃の行われていない観測初日のデータを用いた．Snort によって生成されたアラートの内訳を表 1 に示す．

spathoulas らは NRA，HAF，UFP の用いるパラメータおよび判定に用いる閾値 th に，攻撃アラートの見逃しを小さくしつつより多くの誤検知を削除するものを選択し固定していたが，今回は提案した手法との比較のため th に関しては 0 から 1 の値で変動させグラフを描画した．

5.1 評価指標

評価に用いる指標について説明する．表 2 において TP

表 1　データセットの内訳
Table 1　Datasets

	全体	攻撃	誤検知
ISCXIDS2012	229113	9865	219248
CICIDS2017	435296	162341	272955

表 2　評価指標
Table 2　Evaluation index

判定	ラベル	
	攻撃	誤検知
攻撃	TP	FN
誤検知	FP	TN

表 3　見逃し率ごとの FPR(%)
Table 3　FPR(%) for each missing rate

	見逃し率	既存手法 [1]	DBSCAN	K-Means + DBSCAN
ISCXIDS2012	1%	77.2	70.2	61.5
	2%	70.5	38.9(見逃し率 1.3%)	25.5(見逃し率 1.35%)
	5%	58.6	-	-
	10%	45.2	-	-
CICIDS2017	1%	99.2	98.0	99.7
	2%	98.0	92.6	83.5
	5%	94.6	73.9	73.6
	10%	75.4	62.1(見逃し率 7.4%)	62.5(見逃し率 7.3%)

は実際の攻撃を正しく攻撃と判定，FN は実際の攻撃を誤検知と判定（見逃し），FP は誤検知を攻撃と判定，TN は誤検知を正しく誤検知と判定したことを表す．今回はこれらを用いて，見逃し率と，FPR を定義する．見逃し率は実際には攻撃によって生じたアラートを誤検知と判定した割合，FPR は攻撃と判定したアラートの中に含まれる誤検知アラートの割合を表す．そのためこれらの値は小さいほうが良い値となる．

$$\text{見逃し率} = \frac{FN}{TP + FN}$$

$$\text{FPR} = \frac{FP}{FP + TN}$$

5.2　結果

3 つの誤検知アラート削除法を実装し，得られた見逃し率 (%) と FPR(%) の結果を示す．

図 4, 図 5 には提案手法の結果を示し，既存手法も含めた結果を図 6，図 7 に示す．また，表 3 において見逃し率を 1，2，5，10%としたときの FPR(%) を示す．

5.3　考察

表 3，図 6，図 7 から同様の見逃し率を実現した場合，既存手法と比較して提案手法ではより小さい FPR となることが分かる．つまり，より多くの誤検知を削除できているといえる．これより，見逃し率が小さい領域で既存手法よりも良い精度で誤検知の削除が達成できたといえる．一般に IDS ではある程度誤検知を許容してでも見逃しを減らす

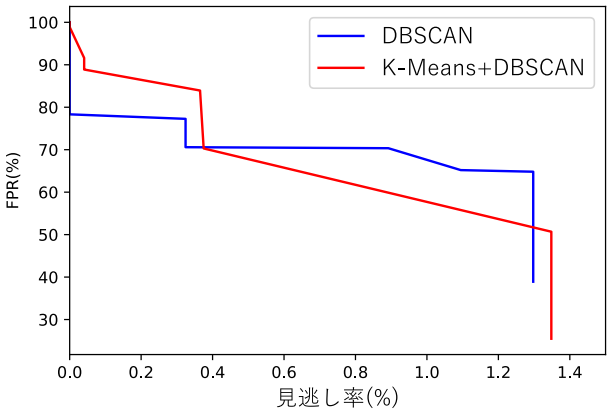

図 4　ISCXIDS2012 データセットにおける提案手法の比較
Fig. 4　Comparision of the proposed methods on ISCXIDS2012

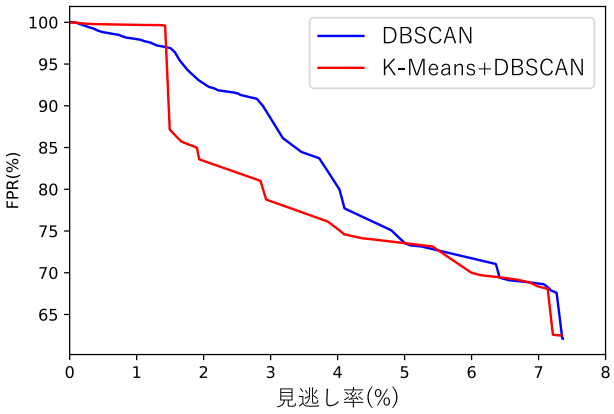

図 5　CICIDS2017 データセットにおける提案手法の比較
Fig. 5　Comparision of the proposed methods on CICIDS2017

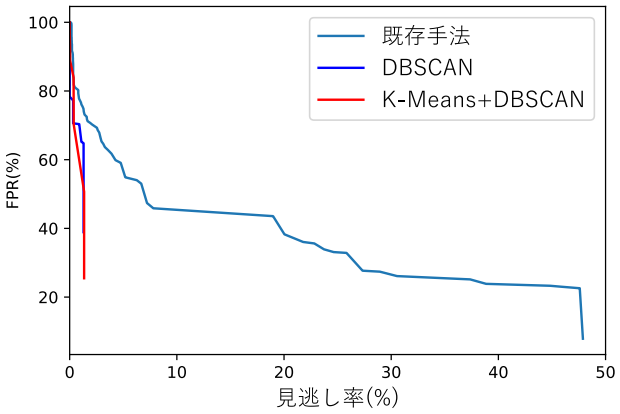

図 6　ISCXIDS2012 データセットにおける手法の比較
Fig. 6　Comparision between previous method and proposed methods on ISCXIDS2012

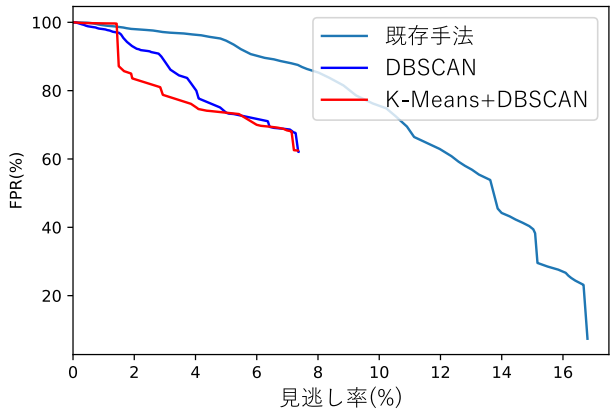

図 7　CICIDS2017 データセットにおける手法の比較

Fig. 7　Comparision between previous method and proposed methods on CICIDS2017

ことが重要視されているため提案手法は有用であると考えられる．また提案案手法では出現頻度を用いたフィルタにおける th を変動させても見逃し率が小さい領域に集中し，安定性が高いことも分かった．

一方で，提案手法で大幅な誤検知削除が難しい点も明らかになった．実際に結果をみると，FPR を ISCXIDS2012 では 25.5%，CICIDS2017 では 62.5%より下げることができていない．さらなる誤検知の削除には，既存手法で見逃し率を大きくすることでしか対応できないことがわかった．

また，2 つの提案手法間の比較において，ほとんどの領域では K-Means と DBSCAN を用いた手法の FPR が低くなっていた．これより，K-Means と DBSCAN を用いた手法では，DBSCAN のみを用いた手法でクラスタリング時に外れ値に分類されなかった攻撃や，外れ値に分類された誤検知を，正しく分類できたと考えられる．しかし，見逃し率が特に小さい領域において，DBSCAN のみを用いた手法のほうが FPR が低くなっていた．これは，K-Means と DBSCAN を用いた手法では，既存手法である出現頻度を用いたフィルタでの判定結果をその後のステップでも大いに利用しており，th の値が小さいときに既存手法の精度が低くなることに起因していると考えられる．

6.　まとめと今後の課題

本稿では，IDS の利用においてアラートの分析を困難とする要因である誤検知の多さを解決するために，NIDS の生成するアラートの中から誤検知を削除する手法を 2 つ提案した．提案手法では Spathoulas らによるアラートの出現頻度を用いたフィルタにクラスタリングを用いたフィルタを追加した．既存手法と提案手法 2 つを実装し，2 つのデータセット ISCXIDS2012,CICIDS2017 を用いた評価を行ったところ，提案手法は見逃し率が小さい領域において，既存手法よりも多くの誤検知を削除できることが分かった．一般的な IDS の利用においてはある程度誤検知を許容してでも見逃しを減らすことが重要視されるため，見逃し率の小さい領域でより多くの誤検知を削除できた提案手法は有用性が高い．しかし提案手法では，見逃し率を大きくしてでもより多くの誤検知を削除するという使用法はできない点も明らかとなった．既存手法においても FPR が小さい領域において見逃し率が大きくなっていたため，FPR が小さい業域での精度向上についても検討していく．

提案手法同士の比較では，DBSCAN のみを用いた手法は見逃し率が特に小さい領域でより良い性能を示し，その他の領域では K-Means と DBSCAN を用いた手法がより良い性能を示した．すべての見逃し率においてより良い精度で誤検知を削除できる手法の考案が今後の課題となる．

本稿で提案した手法は，ともにラベルの付いたデータを必要としない教師なし学習を用いていたが，[8，9] などで教師あり学習である機械学習手法を用いた誤検知削除法も研究されている．本稿で提案した手法を教師あり学習を用いて改善する方法についても検討していきたい．

参考文献

[1]　Georgios P.Spathoulas，Sokratis K.Katsikas: Reducing false positives in intrusion detection systems，Computers & Security，Volume 29，Issue 1，pp.35–44，Feb，2010.

[2]　Ali Shiravi，Hadi Shiravi，Mahbod Tavallaee，Ali A. Ghorbani： Toward developing a systematic approach to generate benchmark datasets for intrusion detection，Computers & Security，Volume 31，Issue 3，pp.357–374，May，2012

[3]　Iman Sharafaldin，Arash Habibi Lashkari，and Ali A. Ghorbani： Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization，4th International Conference on Information Systems Security and Privacy (ICISSP)，Portugal，Jan，2018

[4]　Neminath Hubballi，Vinoth Suryanarayanan，： False alarm minimization techniques in signature-based intrusion detection systems: A survey，Computers & Security，Volume 49，pp.1–17，Aug，2014

[5]　Liang Hu，Taihui Li，Nannan Xie，Jiejun Hu: False positive elimination in intrusion detection based on clustering，2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)，Aug，2015

[6]　岩崎信也，角田朋，関口悦博，小西幸洋，大鳥朋哉，薦田憲久： 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除手法，Computer Security Symposium 2018，Oct，2018

[7]　Snort: Snort-Network Intrusion Detection & Prevention System，available from 〈https://www.snort.org/https://www.snort.org/〉(accessed 2019-08-09)

[8]　Jegede T.J，Asanbe M.O：Post analysis of Snort intrusion files using data mining techniques: Decision tree and Bayesian network，Villanova Journal of Science，Technology and Management Vol. 1，　No. 1，2019

[9]　P Wei，Z Zhang and B Chen： A method of eliminating false alarm based on deep learning，Journal of Physics: Conference Series，Volume 1087，Machine Learning and Artificial Intelligence，2018