1W-01

機械学習を用いた侵入検知システム改良手法の検討

近松 康次郎 †　平川 豊 ‡　大関 和夫 ‡
†芝浦工業大学大学院理工学研究科　‡芝浦工業大学工学部

1. 研究背景

近年、マルウェアや不正侵入といったサイバー攻撃の増加に伴い、セキュリティ技術の一つである侵入検知システム(IDS)が注目されている。侵入検知システムとは、ホスト上のプロセスやネットワーク上を流れるトラフィックを監視し、不正や異常を検出し管理者に通知するシステムである。

侵入検知システムの主な検知手法としてシグネチャ検知がある。シグネチャ検知は、既知の攻撃手法について特徴的なパターンを事前にデータベースに登録し、監視対象である通信とパターンマッチングを行う。そして登録されているパターンと一致した場合、その通信を不正や異常として検出する。この手法は誤検知率が低く、事前に登録したパターンの既知の攻撃については高い検知率を示す。しかし、未知の攻撃や既知の攻撃を部分的に変更した亜種攻撃については検知が難しいという問題がある。

そこで、シグネチャ検知に機械学習を用いた手法[1][2]が提案されている。これらの手法では、正常な通信と攻撃それぞれに見られる各特徴との類似性から識別するため、未知の攻撃や亜種攻撃に対しても検知が可能となる。しかし、機械学習を用いると単純なパターンマッチングではなくなるため、誤検知が発生するという問題点もある。

機械学習を用いた検知の評価基準として、しばしば false positive や false negative が用いられる[3]。false positive とは、正常な通信を攻撃と誤検知してしまうことで、これに対して false negative とは、攻撃を正常な通信と誤検知してしまうことである。一般的に、false positive より false negative のほうが重大なインシデントの要因となる可能性が高い。

そこで本研究では、攻撃の検知を最優先に考え、機械学習を用いた侵入検知システムの false negative の削減を目指す。

An Approach of Improving Intrusion Detection Systems using Machine Learning
†Kojiro Chikamatsu, ‡Yutaka Hirakawa, ‡Kazuo Ohzeki
†Electrical Engineering and Computer Science, Shibaura Institute of Technology, Tokyo, Japan
‡Information Science and Engineering, Shibaura Institute of Technology, Tokyo, Japan

表１　ラベル別による誤検知数上位５位

	誤検知数	データに含まれる数
back	228	228
warezclient	114	114
pod	30	30
smurf	30	28122
normal	12	9893

2. 事前実験

まず false negative の要因を調査するため、事前実験として、機械学習を用いて検知を行い、誤検知したデータをラベル別に数えた。機械学習アルゴリズムにはニューラルネットワークの多層パーセプトロンを用い、データセットには KDDCup1999[4]の kddcup.data_10_pecent を用いた。

誤検知数の多かったラベル上位５つを表１に示す。ラベル normal は正常な通信を表す。smurf や normal は誤検知数上位５位以内ではあるものの、高い割合で正しく検知されており、false positive や false negative の増加を抑えていた。しかし、上位３つの攻撃については、全く検知できておらず、false negative 増加の要因となっていた。このように、既存の手法では検知可能な攻撃に偏りがあり、false negative に特定の攻撃が大きく影響していることが確認できた。

3. 提案手法

本研究では、false negative の削減を目的とし、特定の攻撃について学習器を分割する手法を提案する。

具体的には以下の通りである。

1. 特定の攻撃について、その攻撃のみを学習させた学習器を作成
2. 特定の攻撃について、その攻撃以外を学習させた学習器を作成
3. 判定の際は、各学習器から結果を出力し、それぞれの結果を統合して最終的な判定とする

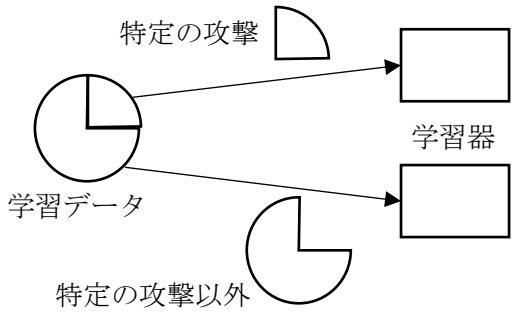

図１　提案手法の学習

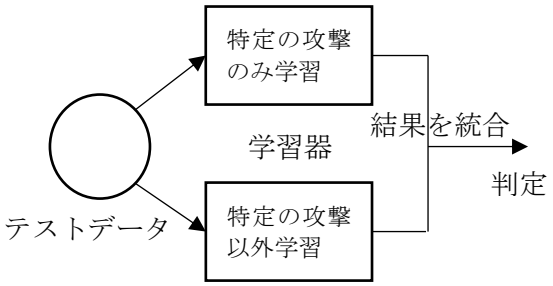

図２　提案手法の判定

結果の統合は論理和で行う。すなわち、どちらか一方の学習器が攻撃と判定した場合、攻撃と判定する。また、分割する攻撃としては、単一の学習器で検知した際に最も誤検知数の多かった攻撃を用いる。

4. 評価

単一の学習器を用いる既存手法と二つの学習器を用いる提案手法を比較・評価する。分割する攻撃については、事前実験の結果に基づき back とする。

機械学習アルゴリズム及びデータセットには事前実験と同様のものを用いた。

ここで、精度とはデータを正しく判定できた割合、適合率とは攻撃と判定したデータのうちデータが実際に攻撃だった割合、再現率とはテストデータに含まれる攻撃データのうち攻撃と判定したデータの割合である。

表２　評価結果

	既存手法	提案手法
精度	0.99123	0.98961
適合率	0.99977	0.98949
再現率	0.98931	0.99767
false positive	9	421
false negative	424	92

表２に評価結果を示す。false negative を削減することができ、再現率が向上している。しかし、適合率と再現率はトレードオフの関係があるため、適合率は低下している。また、削減できた false negative に比べて、増加した false positive の割合が大きいため、精度についても若干低下している。

表３　normal，back の誤検知数

	既存手法	提案手法
normal	5	414
back	245	3

あるデータ群に対する既存手法と提案手法の normal，back の誤検知数を表３に示す。back の誤検知数が減少していると同時に、normal の誤検知数が増加していることから、back のみを学習した学習器は back に対する検知能力は高いが、normal に対する検知能力は分割前の学習器よりも劣っていると考えられる。このため、精度を低下させることなく、false negative を削減するには、特定の攻撃のみを学習させる学習器に正常な通信に対する検知能力も求められる。

5. まとめ

特定の攻撃について学習器を分割することにより、false negative を削減し再現率を向上させることができた。しかし、分割した学習器は特定の攻撃だけでなく、正常なデータに対しても正しく判定できなければ false positive が増加してしまい、全体としての精度が下がってしまう。

今後の課題としては、他の攻撃についての分割や分割数をさらに増やした場合の評価、分割した学習器自体の精度向上等が挙げられる。

参考文献

[1] 小池泰輔，梅澤猛，大澤範高，“ランダムフォレストアルゴリズムを用いたネットワーク侵入検出システムの性能解”，情報処理学会第 76 回全国大会講演論文集，4Z-2，Vol.2014，No.1，pp.619-620，2014-03-11

[2] 高原尚志，櫻井幸一，“KDD CUP 99 Data Set を用いた異なる学習データによる機械学習アルゴリズムの評価”，コンピュータセキュリティシンポジウム論文集，Vol.2015，No.3，pp.457-464，2015-10-14

[3] 小宅宏明，宮地玲奈，川口信隆，重野寛，岡田謙一，“機械学習によるネットワーク IDS の false positive 削減手法”，情報処理学会論文誌，Vol.45，No.8，pp.2104-2112，2004-08-15

[4] UCI KDD Archive: KDD Cup 1999 Data，https://kdd.ics.uci.edu/database/kddcup99/kddcup99.html，2017-12