

# 機械学習型NIDS用の詳細な分類を持つデータセットの提案

大久保 洋樹<sup>1,a)</sup> 小林 良太郎<sup>1</sup>

**概要：**近年、インターネットにおける攻撃は増加傾向にあり、これらに対する防御の必要性が高くなっている。その防御手段の1個としてネットワーク型侵入検知システムが挙げられる。ネットワーク通信を監視し、攻撃通信であると判断されると、ホスト側に通知するシステムであり、悪性通信であると判断する機構には機械学習が多分にして用いられている。この機械学習に必要な学習材料として多くのNIDS用データセットが先行研究によって作成されているが、これらは問題点を抱えている。第1に作成日が古く、最新の攻撃データを含んでおらず、同攻撃を検知することが難しい点である。第2に通信の分類が正常・悪性のラベル付け、または悪性通信の大まかな攻撃手段のラベル付けのみであり、実際の攻撃手法を示す詳細な分類が少ないとといった点である。第3に、攻撃シナリオの網羅性が低い点である。これらの問題点を解決するために、新たなデータセットの提案をする。

**キーワード：**NIDS, データセット, 機械学習

## A Proposal of Dataset and Detailed Classification for Network Intrusion Detection Systems

KOJU OKUBO<sup>1,a)</sup> RYOTARO KOBAYASHI<sup>1</sup>

**Abstract:** In recent years, the number of attacks on the Internet has been increasing, and the need for protection against these attacks is becoming more and more important. One of the defensive methods is a Network-based Intrusion Detection System: NIDS. The system monitors network traffic and notifies the host when it is determined to be an attack traffic, and machine learning is often used as a mechanism to determine whether the traffic is malicious or not. Many datasets for NIDS have been created by previous studies as learning materials necessary for this machine learning, but they have some problems. Firstly, they are old and do not contain the latest attack data, making it difficult to detect these attacks. Secondly, the classification of traffic is limited to the labeling of normal and malicious traffic, or the labeling of rough attack methods for malicious traffic, and there are few detailed classifications that show actual attack methods. Thirdly, the coverage of attack scenarios is low. To solve these problems, we propose a new dataset.

**Keywords:** NIDS, Dataset, Machine Learning

### 1. はじめに

近年増加傾向を示すサイバー攻撃への防御手法としてネットワーク型侵入検知システム（Network Intrusion Detection System: NIDS）があげられる。NIDSにおいては機械学習が注目されており、Shanshanらのように様々なアプ

ローチで高精度化を目指す研究が数多く行われている[1]。IEEE Xploreにおいて“NIDS”に一致する論文を検索すると、2018年から2021年の3年間では247件であったのに対しても、2022年から現在では611件（2025年6月20日現在）であり、この数字からも注目度が伺える。ここで、NIDSによる攻撃の検知を含む研究においては、公開されている正常通信と悪性通信を含むデータセットを用いることが一般的である。その理由として、研究者がデータセットの作成にリソースを割く必要がなくなる点や、同一

<sup>1</sup> 工学院大学  
Kogakuin University, Shinjuku, Tokyo 163-8677, Japan  
a) j122059@ns.kogakuin.ac.jp

のデータセットを利用することで他研究との結果の区別が容易になるといった点からである。しかしここで使用されているデータセットには以下のような問題点を抱えている。  
(1) 作成日が古いため、最新の攻撃に対応していない。以下、この問題を prob1 と呼ぶ。

(2) 正常通信、悪性通信の詳細なラベル付けがされていない。あるいは、攻撃シナリオの分類がされていない。以下、この問題を prob2 と呼ぶ。

(3) 攻撃シナリオの網羅性が低い。あるいは、通信の収集環境がオンライン、オフラインどちらか一方のみであり限定的である。以下、この問題を prob3 と呼ぶ。

それぞれについて詳しく述べる。prob1において、脆弱性データベースの1種である CVE (Common Vulnerabilities and Exposures) の2024年発行件数は40,009件であり、作成日が1年古いデータセットはこれらの攻撃データを含んでいないため、これを元に作成したNIDSは約4万種類の攻撃の検知が難しくなる[2]。prob2について、ラベルがDDoS、BruteForce等の攻撃シナリオのみであった場合、SYNフラッドやGET/POSTフラッド等の種類の特定に時間がかかり、対策の完了までの間に被害が拡大する可能性がある。prob3においては、攻撃シナリオの網羅性が低い場合、登録シナリオ以外の攻撃が来た場合に無防備になる。そこで我々は、上記の問題を解決するデータセットを新たに提案する。このデータセットは、外部に公開しているハニーポットから取得したデータであるオンライン攻撃通信と、ローカル環境に対してペネトレーションツールを用いて取得したデータであるオフライン攻撃通信を含む。オンライン攻撃通信は、ハニーポットのログファイルを解析し、その特徴から導ける攻撃をラベルとして付与する。また、正常通信として研究室で実際に行われている通信を含む。これにより、最新の攻撃を含み、詳細なラベル付けがされており、網羅性が高いという特徴を持つこのデータセットは、NIDSにおける高精度化を目指す研究に貢献できると考える。

本論文の構成は以下のとおりである。第2章では、NIDSに関する先行研究と既存データセットの特徴等について述べる。第3章では提案したデータセットの概要と構築環境について述べる。第4章では統計情報を述べ、第5章では取得したデータの妥当性の評価を行い、第6章で考察を述べる。最後に、第7章でまとめを述べる。

## 2. 関連研究

既存データセットを利用したNIDS関連の研究は近年その数を増やしている。平野らはKyoto 2016 Datasetを使用しており、近松らはKDD CUP 1999 Dataを使用している[3][4]。この章では、既存データセットの研究とその問題点について述べる。

### 2.1 KDD CUP 1999 Data

KDD CUP 1999 Dataはカリフォルニア大学アーバイン校より公開されているデータセットである[5]。これは“KDD-99 The Fifth International Conference on Knowledge Discovery and Data Mining”におけるコンペティションの課題として使用されたデータセットで、MIT リンカーン研究所による典型的な米空軍 LAN を模倣したローカルネットワークの Tcpdump データ9週間分を元にして構築されている。訓練データは7週間分のデータから、これらを処理した約500万件の接続データを使い、テストデータには2週間分のデータを処理して入手した約200万件の接続データを使っている。

攻撃シナリオはDoS(サービス拒否攻撃)、R2L(リモートマシンからの不正アクセス)、U2R(権限昇格を目指した攻撃)、プロープ攻撃(ポートスキャン等)といった4個の主要なカテゴリに分類された22種類の攻撃シナリオを持っている。特徴として、新しい攻撃のほとんどが既知の攻撃の亜種であり、既知の攻撃のシグネチャでこれらが検知できるという考え方から、一部攻撃シナリオはテストデータにのみ含まれている。

このデータセットの問題点として、prob1とprob3があげられる。最終更新日が1999年10月28日であり、かなり古いことがわかる。また、1個のカテゴリに対して攻撃の数は多いものの、攻撃の種類がほかのデータセットと比べると少ないため、網羅性が低いと言える。

### 2.2 NSL-KDD dataset

NSL-KDD datasetはKDD CUP 1999 Dataの改良版として公開されたデータセットである[6][7]。このデータセットにはKDD CUP 1999 dataと比較したとき、大きく3個の利点が挙げられる。第1に、データに冗長なレコード、重複レコードが含まれないため、分類器の偏りを防ぐことが可能になっている点である。第2に、訓練データとテストデータのレコード数が適切であるため、編集することなく実験可能となっている点である。第3に、データセットを用いた研究結果に一貫性があるため、比較が容易といったものである。

しかし、攻撃シナリオ等については変更がないためKDD CUP 1999 Dataと同様の問題を抱えている。

### 2.3 Kyoto 2016 Dataset

Kyoto 2016 Datasetは“Traffic Data from Kyoto University’s HoneyPots”によって公開されているデータセットである。KDD CUP 1999 Data、Kyoto 2006+ Datasetが古くなったことを受け作成されたデータセットでもある[8][9]。京都大学に設置されているハニーポットのデータを使用しているため、この名前で公開されている。特徴量は24種類から構成されている。一方で、外部に公開されて

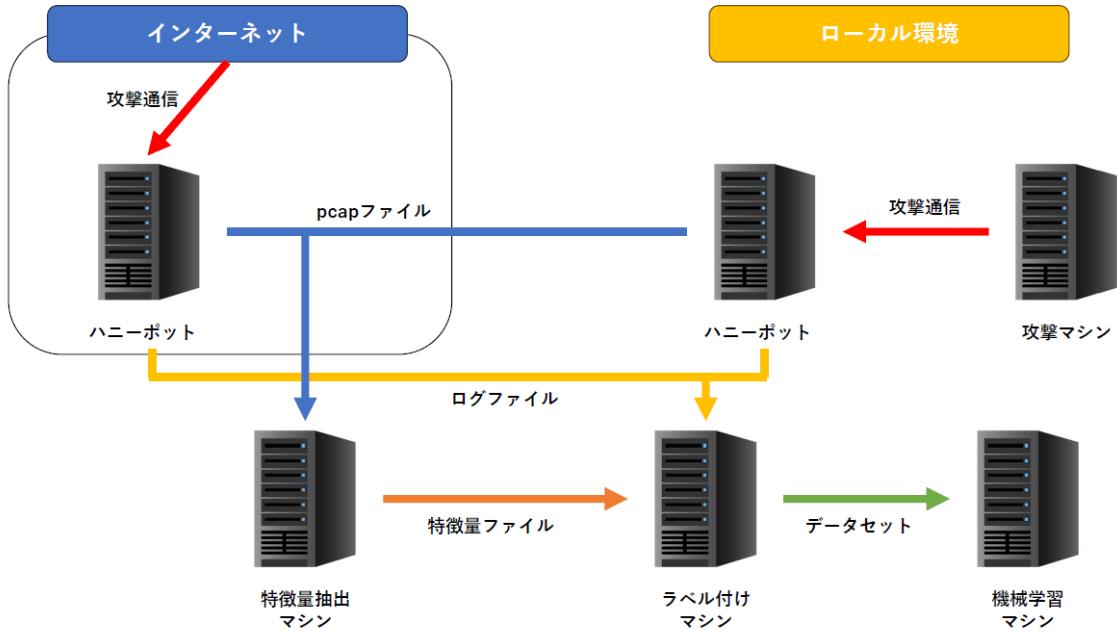


図 1 構築環境  
Fig. 1 Environment Setup

いるハニーポットを使用しているため、攻撃シナリオについて言及がなく、検知の有無等の「正常」「悪性」のラベルのみとなっているため、prob2 を抱えている。

#### 2.4 UNSW-NB15 Dataset

UNSW-NB15 Dataset はニューサウスウェールズ大学サイバーレンジラボの IXIA PerfectStorm ツールによって、Moustafa らが作成したデータセットである [10]. tcpdump を利用し、約 100GB のトラフィックを用いて作成された。このデータセットには Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, Worms からなる 9 種類の攻撃シナリオが指定かつラベル付けされているものの、詳しい分類まではされていないため prob2 を抱えている。

#### 2.5 CSE-CIC-IDS2018

CSE-CIC-IDS2018 は Communications Security Establishment (CSE) と the Canadian Institute for Cybersecurity (CIC) の共同プロジェクトによって作成されたデータセットである [11]. また、このデータセットは、機密保持の観点から社内データセットが外部に公開できないといった入手性の低さや、既存データセットの統計特性の問題の解決にも焦点を当てている。また、CICFlowMeter を用いたパケットの処理により 80 種類の特徴量をデータとして保持している。また、BruteForce, DoS, Web 攻撃, Exploit, Botnet 攻撃などの攻撃シナリオが含まれている。

このデータセットはほかの既存データセットに比べて、作

成日が新しいものの、prob2 と prob3 を抱えている。攻撃シナリオの種類が UNSW-NB15 と比較しても少なく、詳細なラベル付けがされていないためである。

### 3. データセットの作成

本データセットで使用する通信データは、tcpdump を使い pcap ファイルとして保存している。また、pcap ファイルから特徴量を抽出する際には CICFlowMeter[12] を利用している。CICFlowMeter を用いることで 85 種類の特徴量を抽出することができる。

ここで、外部に公開したハニーポットのような不特定多数からの攻撃を“オンライン攻撃”，ローカル環境で特定の攻撃を自ら仕掛けるような攻撃を“オフライン攻撃”と呼称した場合、関連研究のまとめを表 1 に示す。本研究では網羅性を高めるために、オンライン攻撃とオフライン攻撃の両方を用いる。

構築した環境を図 1 に示す。

#### 3.1 オンライン攻撃

本研究では、オンライン攻撃として外部にハニーポットを公開しデータを収集する。今回使用するハニーポットとして、GitHub 上で公開されている“T-Pot”を採用した [13]. T-Pot は Docker によって構成されている、20 種類以上のハニーポットをまとめたマルチハニーポットプラットフォームである。

表 1 既存データセットの網羅性  
**Table 1** Coverage of Existing Data Sets

データセット名	オンライン攻撃	オフライン攻撃	攻撃シナリオ数	攻撃種類
KDD CUP 1999 Data		V	4	22
NSL-KDD dataset		V	4	22
Kyoto 2016 Dataset	V			
UNSW-NB15 Dataset		V	9	0
CSE-CIC-IDS2018		V	7	0

表 2 オフライン攻撃の収集状況  
**Table 2** Progress of Offline Attack Collection

攻撃シナリオ	攻撃種類	種類数
DoS	Slowloris, SYN フラッド, GET/POST フラッド等	18
Probe	Nmap, Netcat, Masscan	3
Fuzzing	BruteForce, Directoryscan, SQLinjection	3

### 3.1.1 収集環境

- ハニーポット構築環境
  - ホスト OS: Ubuntu 22.04 LTS
    - \* CPU: Intel Core i7-10700
    - \* メモリ: 16 GB
  - 使用ソフトウェア: VirtualBox
  - ゲスト OS: Ubuntu 24.04.1
    - \* T-Pot version Hive
    - \* CPU プロセッサ数: 4
    - \* メモリ: 10 GB

### 3.2 オフライン攻撃

本研究では、オフライン攻撃として被攻撃側と攻撃側をローカル環境で構築しデータを収集する。収集データの形式をできる限りオンライン攻撃と均一にするため、同様に T-Pot を被攻撃側として構築し、攻撃側から攻撃を仕掛ける。

#### 3.2.1 収集環境

- 被攻撃側
  - ホスト OS: Windows 11
    - \* CPU: Intel Core i5-1250P
    - \* メモリ: DDR4 16 GB
  - 使用ソフトウェア: VirtualBox version 7.0.14
  - ゲスト OS: Ubuntu server 24.04 LTS
    - \* T-Pot version 24.04 Hive
    - \* CPU プロセッサ数: 4
    - \* メモリ: 16 GB

#### ● 攻撃側

- ホスト OS: Windows 11
  - \* CPU: Intel Core i7-1195G7 @2.90GHz
  - \* メモリ: DDR4 8 GB

- 使用ソフトウェア: VirtualBox version 7.0.14

- ゲスト OS: Kali Linux-2024.1

- \* CPU プロセッサ数: 2

- \* メモリ: 2 GB

#### 3.2.2 オフライン攻撃の収集状況

2025年6月27日現在のオフライン攻撃の収集状況は表2の通りである。DoSには、一般的に使われているwebサーバーへの攻撃から、特定のゲームサーバーへの攻撃のような特定の環境への攻撃種類も含まれている。Fuzzingでは、FFUFとWfuzzの2種類のツールを使い、攻撃を実行した[14][15]。この際に被攻撃側にDVWA 2.5を構築しFuzzingの攻撃対象を明確にした[16]。また、攻撃時のワードリストにはrockyou.txtを始めとした許可無しで使用できるリストを用いた。

#### 3.3 攻撃種類の特定

T-Potでは、攻撃を検知しその総数や攻撃の詳細等様々な情報を見ることができるが、攻撃の種類は特定することができない。一方で、T-Potを構成する各ハニーポットは対象となるサービスが絞られているため、それらハニーポットのログとtcpdumpを用いてキャプチャしたpcapファイルを元に攻撃の種類を特定することができた。

SSH、Telnetの攻撃を対象としたハニーポットであるCowrieを例に挙げる。Cowrieのログには外部からの通信が記録されているため、同一IPアドレスから短時間に連続してログイン失敗ログがあるならばBruteForceとラベル付けを行った。その際のログファイルの一例を図2に示す。また、同一IPアドレスかつ同一時間である通信をpcapファイルから取り出し、紐づけることで通信データにもラベル付けを行った。

```
{
  "eventid": "cowrie.session.connect", "src_ip": "170.64.232.193", "src_port": 48934, "dst_ip": "172.23.0.2", "dst_port": 22, "session": "08bb7bcd685e",
  "eventid": "cowrie.client.version", "version": "SSH-2.0-Go", "message": "Remote SSH version: SSH-2.0-Go", "sensor": "184fe4b373ee", "timestamp": "2021-06-10T12:50:27Z",
  "eventid": "cowrie.client.kex", "hash": "084386fa7ae5039bcf6f07298a05a227", "hashAlgorithms": "curve25519-sha256@libssh.org,ecdh-sha2-nistp256",
  "eventid": "cowrie.session.closed", "duration": "8.0", "message": "Connection lost after 8.0 seconds", "sensor": "184fe4b373ee", "timestamp": "2021-06-10T12:50:27Z",
  "eventid": "cowrie.session.connect", "src_ip": "117.242.151.57", "src_port": 34651, "dst_ip": "172.23.0.2", "dst_port": 23, "session": "38db4ee5ec7",
  "eventid": "cowrie.session.closed", "duration": "13.403619527816772", "message": "Connection lost after 13 seconds", "sensor": "184fe4b373ee", "timestamp": "2021-06-10T12:50:27Z",
  "eventid": "cowrie.session.connect", "src_ip": "170.64.232.193", "src_port": 43098, "dst_ip": "172.23.0.2", "dst_port": 22, "session": "2e949dd3b5b8",
  "eventid": "cowrie.client.version", "version": "SSH-2.0-Go", "message": "Remote SSH version: SSH-2.0-Go", "sensor": "184fe4b373ee", "timestamp": "2021-06-10T12:50:27Z",
  "eventid": "cowrie.client.kex", "hash": "0a07365cc01fa9fc8260ba4019af499", "hashAlgorithms": "curve25519-sha256,curve25519-sha256@libssh.org",
  "eventid": "cowrie.login.failed", "username": "root", "password": "1q2w3e4r", "message": "login attempt [root/1q2w3e4r] failed", "sensor": "184fe4b373ee", "timestamp": "2021-06-10T12:50:27Z",
  "eventid": "cowrie.session.closed", "duration": "2.2", "message": "Connection lost after 2.2 seconds", "sensor": "184fe4b373ee", "timestamp": "2021-06-10T12:50:27Z"
}
```

図 2 Cowrie ハニーポットのログ例

Fig. 2 Example of Cowrie Log File

### 3.4 正常通信の採取

正常通信は、機械学習型 NIDSにおいて、正常・悪性の検知のための判断材料として必要である。そこで、正常通信として研究室で実際に行われている通信を採取する。これには HTML を始めとした、普段使用されている通信が含まれている。このネットワークでは、すべての PC がスイッチを経由して通信を行っていることから、スイッチのミラーポートを使用し、通信の取得を行っている。

## 4. 統計情報

この章では入手した通信の統計情報を示す。オンライン攻撃通信では、ある期間に T-Pot で取得した攻撃の詳細な内訳を示す。オフライン攻撃通信では、取得したデータの注意点について示す。

### 4.1 オンライン攻撃通信

オンライン攻撃通信では、T-Pot を用いて悪性データを取得している。ここで、2025年6月10日から6月14日までの5日間に取得したデータ数と各ハニーポットの内訳を図 3 に示す。1日に2万件以上の攻撃通信を入手することができており、その内訳は5日間すべてにおいて cowrie ハニーポットが占める割合が最も多く、次いで Dionaea である。Cowrie は、SSH または Telnet に対する攻撃を収集するハニーポットであるため、同5日間におけるポートごとの攻撃数はポート 22 (SSH)、ポート 23 (Telnet)、ポート 2323 (Telnet 代替ポート) が多くなっている。他に攻撃数が多いポートとして、445 (SMB)、80 (HTTP) 等が挙げられる。

### 4.2 オフライン攻撃通信

オフライン攻撃通信では、現在 DoS, Fuzzing, Port-scan の3種類を主に収集している。各攻撃の性質上、pcap ファイル上で多量のトラフィックを取得しても特徴量を抽出する際にその数が大きく減ることがあるため、その数が機械学習の時に外れ値とならない様に調整する必要がある。特徴量を抽出後の要素数が機械学習に影響が出ると考えた100未満とならないように調整している。

## 5. 評価

収集した攻撃通信は機械学習を用いてその妥当性を評価する。今回は python のライブラリである sklearn を使用し、学習方法にはランダムフォレストを採用した。訓練データは各ラベルにおいて同数とし、テストデータは訓練データに含まれていない物を利用する。また、判定後に正解ラベルと比較しその正答率をもとに評価を行う。

### 5.1 同じ攻撃におけるオンライン、オフラインの判定

オンライン攻撃とオフライン攻撃に差異があるのかを機械学習を用いて確かめた。Cowrie ハニーポットから収集した実際の SSH BruteForce を含む通信をオンライン攻撃通信として用意した。また、ローカルで構築したハニーポットに対して Hydra を用いた SSH BruteForce を実行し、入手した通信をオフライン攻撃通信として用いた。この機械学習では、オンライン攻撃通信、オフライン攻撃通信から抽出した各 40 個のデータを使用した。それぞれ 7 割を訓練データ、3 割をテストデータとし、渡されたテストデータでオンライン攻撃通信、またはオフライン攻撃通信であるかの判定を行い、その正答率を出力する。結果は表 3 の「状況 1」である。

### 5.2 同じ攻撃シナリオにおける判定

同じ攻撃シナリオの通信は、特徴が似ている場合が多いため、その種類を機械学習で判定することができるのかを確認した。攻撃シナリオは DoS を採用し、DoS ツールを用いてオフライン環境下で通信を収集した [17]。このツールは GET/POST や Slowloris 等の汎用的なものから、特定のサービスを対象とした攻撃手法まで提供している。実行可能な種類をすべて実行しデータを収集したが、処理後にデータ数が極端に少なく、学習に影響を及ぼすと考えた種類については除外した。最終的に 18 種類の DoS 通信を得し、種類ごとにラベル付けを行なった。各ラベルにおけるデータ数は 137 であり、訓練データは 7 割、テストデータは 3 割とした。機械学習では、テストデータを正しい DoS の種類に判定できたその正答率を求めた。結果は表 3 の「状況 2」である。また、比較対象として異なる攻撃シナリオで

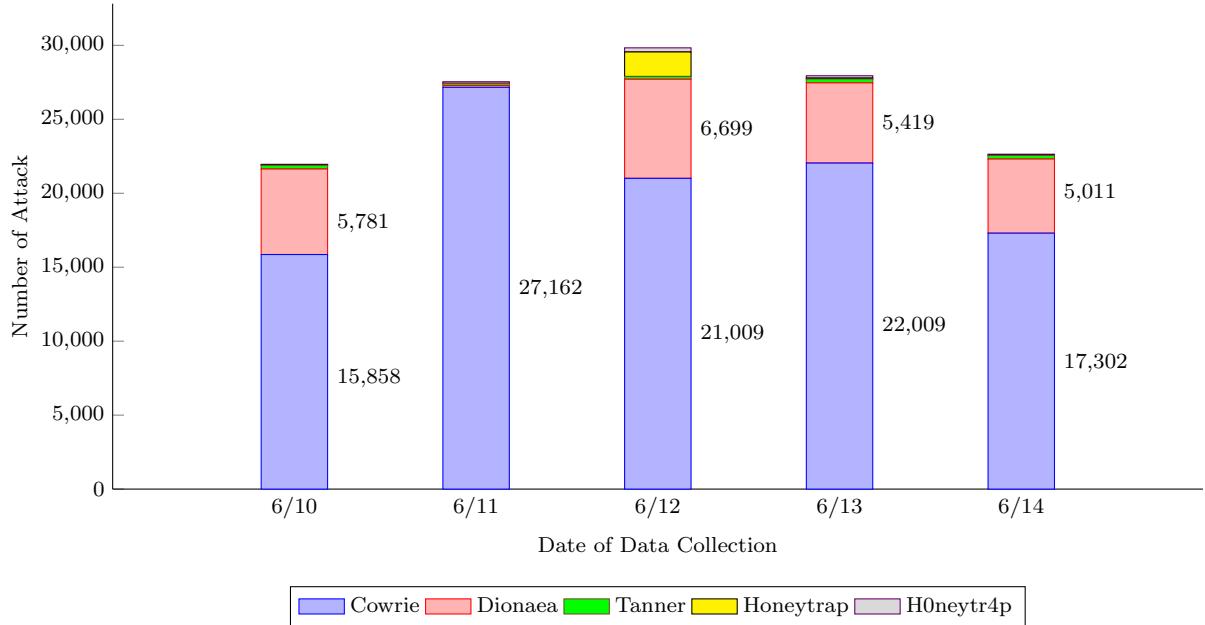


図 3 5 日間のオンライン攻撃通信入手数

Fig. 3 Number of Online Attack Communications Obtained in 5 Days

ある Port-Scan を採用しデータを収集した。この際、種類数は 3 種類、各データ数は 530 であった。結果は「状況 3」である。

## 6. 考察

表 3 を基に考察を行う。「状況 1」では、同じ攻撃シナリオにおけるオンライン攻撃とオフライン攻撃の違いについて求めた。3 回実行した結果、全て正答率 1.0 であった。このことから、同一の攻撃シナリオであっても、オンライン攻撃とオフライン攻撃では明確に違いがあることがわかる。この結果は、オフライン攻撃通信のみ含まれたデータセットでは、オンライン攻撃の検知が難しくなるといった問題を引き起こす。また、今回オフライン攻撃通信を取得する際に使用した Hydra ツールと、オンライン攻撃の際に使用されたツールが違う可能性を始めとした他の要因も考えられる。その上で、全てのオンライン攻撃通信、オフライン攻撃通信が同じではないとも考えられることから、両方の環境で通信を取得する必要性を示すことができた。

「状況 2」、「状況 3」では、同じ攻撃シナリオ下の攻撃を分類することができるのかについて求めた。「状況 2」では、MHDDoS を用いて 18 種類の DoS 攻撃を含む通信を収集し、ラベル付けしたのちに、分類器がテストデータを正しい種類に分類できたか、その正解率を求めた [17]。その結果、正答率は平均して 0.87 であった。原因として種類数が 18 と多い一方で、各データ数が 137 と少ないことが考えられる。「状況 3」では、種類数 3 に対して各データ数が 530 と多いことから正答率もほぼ 1.0 となっている。

これらの結果から、種類数が多い攻撃シナリオはそれぞれ

より多いデータを集めることで正答率を上げることができると考える。また、本データセットが機械学習に有用であることを示している。

## 7. まとめ

本研究では、機械学習型 NIDS 用の既存データセットの問題点を解決するため、実際の攻撃通信（オンライン攻撃）とローカル環境での攻撃通信（オフライン攻撃）の両方を含むデータセットの提案をした。オンライン攻撃では、ハニーポットで誘引した攻撃をログファイルと pcap ファイルをもとに種類を特定しラベル付けを行うことで実用性を高めている。また、収集したデータをもとに機械学習を行いその妥当性、有用性を示した。

機械学習型 NIDS における課題として、攻撃時に人間が気付かない微小な外れ値を入力することで機械学習に悪影響を与える“Adversarial Attack”が 2004 年に Jhon Graham Cummings によって報告されている。近年では、Wang らによる NIDS への Adversarial Attack の研究や、Mingqiang らによる重要なインフラの IDS を攻撃性が保たれたまま検知回避する研究が行われている [18][19]。本研究ではオンライン攻撃としてハニーポットに誘引した攻撃を使用しているため、この攻撃に対する対策ができていない。今後の課題として対策を考える必要がある。また、サイバーセキュリティにおける攻撃は日々進化しており、この提案データセットが完成した後もその例外ではない。いずれは、problem を抱えることになる。したがって、学習データを自動的に選別し自動的にデータセットを更新する機構を作成する必要がある。

表 3 各状況における学習結果  
**Table 3** Learning Results in Each Situation

状況	判定対象	種類数	各データ数	正答率 (3 回分)
1	SSH BruteForce (Cowrie, Hydra)	2	40	1.0, 1.0, 1.0
2	DoS (MHDDoS)	18	137	0.857, 0.880, 0.865
3	Port-Scan (nmap, masscan, netcat)	3	530	0.998, 0.996, 1.0

謝辞 本研究の一部は、JSPS 科研費 23H03396 の支援により行った。

## 参考文献

- [1] S. Tu, M. Waqas, A. Badshah, M. Yin and G. Abbas, “Network Intrusion Detection System Based on Pseudo-Siamese Stacked Autoencoders in Fog Computing,” Proceedings of the IEEE Transactions on Services Computing, Vol. 16, pp. 4317-4327, 2023.
- [2] “JerryGamblin.com,” <https://jerrygamblin.com/2025/01/05/2024-cve-data-review/> (Accessed 2025-06-20).
- [3] 平野誠, 八榎博史, “機械学習を用いた攻撃検知に関する学習手法の精度評価,” 第 81 回全国大会講演論文集, pp. 461-462, 2019.
- [4] 近松康次郎, 平川豊, “ニューラルネットワークを用いた侵入検知システム改良手法の検討,” 第 81 回全国大会講演論文集, pp. 455-456, 2019.
- [5] “KDD CUP 1999 Data,” <https://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html> (Accessed 2025-06-20).
- [6] M. Tavallaei, E. Bagheri, W. Lu and A.A. Ghorbani, “A Detailed Analysis of The KDD CUP 99 Data Set,” Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications, pp. 1-6, 2009.
- [7] “UNB, UNIVERCITY OF NEW BRUNSWICK,” <https://www.unb.ca/cic/datasets/nsl.html> (Accessed 2025-06-23).
- [8] “Traffic Data from Kyoto University’s Honeypots,” <https://www.takakura.com/Kyoto.data/> (Accessed 2025-06-23).
- [9] 多田竜之介, 小林良太郎, 嶋田創, 高倉弘喜, “NIDS 評価用データセット: Kyoto 2016 Dataset の作成,” 情報処理学会論文誌, Vol. 58, No. 9, pp. 1450-1463, 2017.
- [10] N. Moustafa and J. Slay, “UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set),” Proceedings of the 2015 Military Communications and Information Systems Conference (MIICIS), pp. 1-6, 2015.
- [11] “UNB, UNIVERCITY OF NEW BRUNSWICK,” <https://www.unb.ca/cic/datasets/ids-2018.html> (Accessed 2025-06-23).
- [12] UNBCIC, “CICFlowMeter,” <https://github.com/UNBCIC/CICFlowMeter> (Accessed 2025-06-24).
- [13] Telekom-security, “tpotce,” <https://github.com/telekom-security/tpotce> (Accessed 2025-06-24).
- [14] ffuf, “ffuf,” <https://github.com/ffuf/ffuf> (Accessed 2025-06-27).
- [15] xmendez, “wfuzz,” <https://github.com/xmendez/wfuzz> (Accessed 2025-06-27).
- [16] digininja, “DVWA,” <https://github.com/digininja/DVWA> (Accessed 2025-06-27).
- [17] MatrixTM, “MHDDoS,” <https://github.com/MatrixTM/MHDDoS> (Accessed 2025-07-07).
- [18] M. Wang, N. Yang, N.J. Forcade-Perkins and N. Weng, “ProGen: Projection-Based Adversarial Attack Generation Against Network Intrusion Detection,” Proceedings of the IEEE Transactions on Information Forensics and Security, Vol. 19, pp. 5476-5491, 2024.
- [19] M. Bai, et al., “Adversarial Attack against Intrusion Detectors in Cyber-Physical Systems With Minimal Perturbations,” Proceedings of the 2024 IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA), pp. 816-825, 2024.