Vol.2018-SE-199 No.16
2018/7/19

ネットワーク型侵入検知の評価用データセットに関する提案

高原尚志[†1]

概要：近年，インターネット上のサイバー攻撃は激しさを増しており，その対策が急務となっている．ネットワーク上の攻撃からコンピュータシステムを守る方法にネットワーク型侵入検知システム（Network-based Intrusion Detection System: NIDS）がある．NIDS に関しては，人工知能（Artificial Intelligence: AI）を用いたものなど多くの研究がなされており，①手法の理論的な開発，②ソフトウエアによる実装，③データセットによる評価という手順で研究を進めて行くのが一般的である．この際，②のソフトウエアによる実装では，既存のソフトウエアを用いるのか自分で作成したプログラムを用いるのかを検討する必要があり，用いるソフトウエアの信頼性が重要である．また，③のデータセットによる評価では，ラベル付きかラベルなしかや公開か非公開か，同一のネットワーク環境で収集されたデータかなどが重要となる．用いるプログラムや評価用データセットの信頼性が担保されなければ，新たな手法を提案しても，その有効性は保証されない．この内，本稿では，評価用データセットについて，研究結果の信頼性を保証するための課題を明らかにし，解決策を示す．

キーワード：サイバーセキュリティ，NIDS，評価用データセット，KDD Cup 1999 Data，Kyoto2016，DARPA Intrusion Detection Data Sets

A Suggestion of Evaluation Data Sets for Network-based Intrusion Detection System

HISASHI TAKAHARA[†1]

Abstract: Today, cyber-attacks on the Internet have become common, so we need to protect PCs from them. One method of protection is called Network-based Intrusion Detection System (NIDS). Steps of research of NIDS are follows: (step1) Development of methods, (step2) Implementation of software, (step3) Evaluation of methods by datasets. In (step2) reliability of software and in (step3) reliability of datasets is very important. If those reliabilities are not guaranteed, the overall reliability of the research is in question. In this paper we focus on the datasets of NIDS and point out problems. Moreover, we suggest some solutions. We hope that this paper contributes the research of NIDS in the future.

Keywords: Cyber security, NIDS, Evaluation datasets, KDD Cup 1999 Data, Kyoto2016, DARPA Intrusion Detection Data Sets

1. はじめに

本論文は，ソフトウエアエンジニアリングシンポジウム2018(SES2018)に投稿（査読）中の論文[1]を加筆修正し，更に発展させたものである．

1.1 背景

インターネット上のサイバー攻撃が激しさを増す今日,その対策が急務となっている.サイバー攻撃を発見する手法にはネットワーク上の通信データを用いて侵入を検知するネットワーク型侵入検知（Network-based Intrusion Detection System: NIDS）とホストのログデータを用いて侵入を検知するホスト型侵入検知（Host-based Intrusion Detection System: HIDS）とがある．この内，本稿では，NIDS について研究を進める上での課題について検討する．

NIDS の研究を進めるうえで,次のステップが考えられる．

(STEP1) 手法の理論的開発
(STEP2) 手法のソフトウエアによる実現
(STEP3) データセットを用いた手法の評価

(STEP1)で開発した手法を(STEP2)で実装し,(STEP3)で評価するのであるが，多くの論文では，開発した手法の有効性を(STEP3)の評価結果に基づいて示している．この場合，(STEP2)で用いるソフトウエアと(STEP3)で用いるデータセットの信頼性が同時に担保されなければ，論文自体の主張は意味がないものとなってしまう.しかし,現在のところ,著者らが知る限り，上記の信頼性の担保について，その課題を詳細に論じた文献はない（評価用データセットについては，古い，冗長的であるなど有効性について論じた文献は多数存在するが，データの信頼性について論じた論文はない）．

1.2 動機

1.1 章で示した通り，用いるデータセットの信頼性は研究全体に大きく影響する柱のひとつである．これが担保されなければ，研究全体の信頼性が担保されず，手法の評価結果など研究結果が意味のないものになってしまう．そこで本稿では， (STEP3)で用いられているデータセットについて,現状を分析して,課題を明らかにする.これにより,今後 NIDS の研究の信頼性を担保することへの貢献を目指す．

†1 新潟県立大学
University of NIIGATA PREFECTURE

1.3 先行研究

Sharafaldin らは，IDS 用データセットに対する課題を指摘し，信頼できるデータセットを作成するため，以下の 11 個の基準を提唱した [2][3].

●Sharafaldin らが提唱した 11 個の基準

S1.Complete Network Configuration…完備したネットワーク環境の構築

モデムやファイアウォール，スイッチ，ルータなど多様なネットワーク機器や Windows，Linux，Macintosh など多様な OS を含んだネットワーク環境が設定されているか

S2.Complete Traffic…完備した通信データの取得

多様な送信元から多様な送信先への通信データが取得されているか

S3.Labeled Dataset…ラベル付け）

攻撃通信や正常通信など取得したデータに対して，正解のラベル付けがなされているか

S4.Complete Intreraction…完備した相互通信の構築

データセットを取得したネットワークに関して，LAN 内，LAN 間などの相互通信が設定されているか

S5.Complete Capture…完全なキャプチャデータの提供

不必要と思われるデータであっても削除しないなど取得したすべてのデータが提供されているか

S6.Available Protocols…有用なプロトコルの提供

HTTP や FTP，電子メールのプロトコルなど有用なプロトコルが含まれているか

S7.Attack Diversity…多様な攻撃データの提供

最新の攻撃を含む多様な攻撃が含まれているか

S8.Anonymity…匿名性への配慮の制限

プライバシーを考慮して，IP とペイロードのいずれかが削除されていないか

S9.Heterogeneity…異質なデータの提供

通信データだけではなく OS ログやネットワーク機器のログなどを含んでいるか

S10.Feature set…適切な特徴量の抽出

有用な特徴量が抽出されているか

S11.Metadata…メタデータの提供

提供されたデータセットを充分に説明する文書が提供されているか

この内，本稿では S3 ラベル付けに注目して，論理を展開する.

更に，Sharafaldim らは文献[2]の中で，上記の 11 個の基準をクリアしたデータセット[4]を提案し，上記 11 個の基準に照らし合わせて，他のデータセットとの比較を一覧表で示した.（表 1.）（本稿の表 1 では，広く用いられているデータセット（DARPA，KDD99）と日本のデータセット（KYOTO）のみ抜粋）

表 1. IDS 評価用データセットを基準に照らし合わせた一覧（文献[3]の Table1.より引用（一部抜粋））

Table1. Comparing of available datasets based on evaluation framework

		DARPA	KDD'99	KYOTO
Network		Y	Y	Y
Traffic		N	N	N
Label		Y	Y	Y
Interaction		Y	Y	Y
Capture		Y	Y	Y
Protocols	http	Y	Y	Y
	https	N	N	Y
	ssh	Y	Y	Y
	ftp	Y	Y	Y
	email	Y	Y	Y
Attacks	Browser	Y	Y	Y
	Bruteforce	Y	Y	Y
	DoS	Y	Y	Y
	Scan	Y	Y	Y
	Backdoor	N	N	Y
	DNS	N	N	Y
	Others	Y	Y	Y
Anonymity		N	N	N
Heterogeneity		N	N	N
Feature Set		N	Y	Y
Metadata		Y	Y	Y

多田らは，文献[5]の中で，評価用データセットとして重要なこととして，次の 2 点を指摘している.

・最新の攻撃傾向を反映できている

・データの収集期間が長い

また，多田らは，一部のデータセットは，実験用のネットワーク上で悪性通信を疑似的に作り出して作成されており，実際のネットワーク環境を反映できていないとの指摘も行っている.

Mahbod らは，文献[6]の中で，KDD99 について，意図的に攻撃を増やしている，通信データの欠損があるなどの指摘を行い，実際のネットワークから得られる通信データとの乖離を指摘している.

1.4 解決すべき課題

IDS の研究を行うに当たっては，評価用データセットについて，次の 3 つの観点から検証することが重要である.

1. 信頼性
2. 最新性
3. 有効性

信頼性の観点からは，データセットに攻撃通信と正常通信と

いう正解ラベルを付与する必要があると考えられる[2][3]が，ラベルが付加されていないデータセットを評価用として用い場合，”どの通信を攻撃通信と見なすかが画一的でない”という問題が生じる．また，ラベルのないため，提供されているデータをすべて攻撃通信として扱い，自組織のネットワークから得たデータを正常通信として組み合わせたデータセットを用いる場合が多く見受けられる．この場合，”自組織から得られた通信データが非公開であるため読者による結果の検証が困難である”，”ネットワーク環境が異なるデータでの検証になってしまう”などの課題が生じる可能性がある．

最新性の観点からは，現在でも世界的に KDD99 が IDS 評価用ラベル付きデータセットとして広く用いられているが，KDD99 は 1998 年に収集された DARPA98 をセッションデータに加工したものであるため，最新性という面が課題となっている．

有効性という観点からは，評価用データセットの攻撃通信の割合が実際の攻撃をモニタした場合と極端に異なる場合（KDD99 の場合，約 8 割が攻撃通信）があるため，評価用データセットでの評価が，実践に有効か否かという有効性の面で課題がある．

1.5 本研究の貢献

1.4 章の課題を踏まえて本論文の学術的貢献は以下の通りである．

・評価用データセットの信頼性に関する現状（課題）を指摘し，その解決策を示す
・評価用データセットの最新性に関する現状（課題）を指摘し，その解決策を示す
・評価用データセットの有効性に関する現状（課題）を指摘し，その解決策を示す

特に，最新性と有効性に関しては，著者らがシンポジウムなどで専門家から意見を聞いた結果を踏まえて，基準として具体的な数値を示す．これは，著者らが知る限り，世界で初めての提案である．

更に，次の提案も合わせ行う．

・継続性と最新性を両立するための解決策
・再現性を保証するための解決策

以上により，開発された手法が有益であるということが，客観的基準をもって，検証可能にすることを目指す．

1.6 関連研究

Atilla らは，2010 年から 2015 年に発行された 65 の科学雑誌の 149 の論文をサーベイして，侵入検知（Intrusion Detection System: IDS）や機械学習研究（Machine Learning Research: MLR）に用いられたアルゴリズム（手法）やソフトウエア，データセットなどを示している[7].

文献[7]によれば，(STEP3)で用いられるデータセットの内 64.9%は KDD Cup 1999 Data[8][6][9](以降，KDD99 と称す)である．また，University of New Brunswich(UNB)の Canadian Institute for Cybersecurity において KDD99 の欠点を改良した NSL-KDD[6][10][11]とそのもととなったデータセットである DARPA 1998 Dataset[12]を合わせると 80.5%の論文が使用していることになる．特に，用途を IDS に限った場合（表 2.の*印がついたデータセット）には、96.5%の論文が KDD99 系(KDD99+NSL-KDD+DARPA98)のデータセットを使用しているという結果となった。つまり，文献[7]で見る限り、ほとんどすべての IDS 系の論文が KDD99 系のデータセットを使用していることとなる．また，京都大学が提供しているデータセットである Kyoto(Kyoto2006+)[13][14][15][16]も，全体の 1.5%（IDS 用途で 1.8%）で世界的に用いられているという結果となった（表 2.）．

表 2. 多く用いられているデータセット
*印は IDS 用データセットを表す．用いている論文数が 3 より少ないものは省略．（文献[7] Talble.9 より完全引用）
Table2. Most used Datasets

Dataset Name	Article Count
KDD99*	133
NSL-KDD*	23
DARPA*	9
Iris	8
Glass	5
Breast Cancer	5
Synthetic Data	5
Porker Hand	5
Image Segmentation	3
ISCX*	3
Wine	3
Kyoto*	3

2. データセットを用いた手法の評価の現状

2.1 評価用データセット

2.1.1 評価用データセットの用途

本稿では，正解ラベルを比較的容易に付すことができるセッション型データに注目し，現在も世界的に IDS 関連の多くの論文で用いられている KDD99 と京都大学のハニーポットを用いた Kyoto2016 という 2 つの評価用データセットについて述べる．加えて，参考のため，正解ラベルは付されておらずセッション型データセットではないが，日本で配布されている代表的な評価用データセットの 1 つである MWS データセットについても触れる．

2.1.2 KDD Cup 1999 Data

KDD Cup 1999 Data は，米国国防高等研究計画局（Defense Advanced Research Project Agency: DARPA）と米国空軍研究所

（Air Force Research Laboratory: AFRL）のもとでマサチューセッツ工科大学（Massachusetts Institute of Technology: MIT）Lincoln laboratory の The DARPA Intrusion Detection Evaluation Group が作成配布した世界初の NIDS 評価用標準データセットである 1998 DARPA Intrusion Detection Evaluation Data Sets（以降 DARPA 1998 と称す）をもとに作成された NIDS 評価用データセットで，University of California Irvine の Machine Learning Repository で公開されている．DARPA 1998 がパケットキャプチャファイル形式であるのに対し，KDD99 はこれを加工し，セッションデータ形式として供給されており，現在でも多く論文で NIDS の評価用データセットとして使用されているが，日本国内の論文では，①データが古い，②冗長的である，③攻撃通信の割合が多く現実的でないなどの理由からあまり使用されていない．KDD99 の②冗長的である，③攻撃通信の割合が多く現実的でないなどの欠点を修正したデータセットとして，UNB(University of New Branswich)の CIC(Canadian Institute for Cybersecurity)から提供されている NSL-KDD がある．現在では，NIDS 評価用データセットとしてこちらを用いた論文も多くみられる．

考察

KDD99 は，1998 年に取得された DARPA98 のデータを基本にしているため，データが古く，評価手法が現在の攻撃に対応しているかが判定できないという意見がある．しかし，過去の手法との比較など研究の一貫性を保つ上で，現在も多くの論文で評価用データセットとして採用されている．

2.1.3 Kyoto2016

Kyoto2016 は，その前身である Kyoto2006+に 2015 年 12 月までのデータを追加したものである．Kyoto2006+は，KDD99 が古くなったのを受け，京都大学に設置されているハニーポットで 2006 年 11 月から 2009 年 8 月までの期間に収集された通信データをもとに，KDD99 の特徴量の内特に影響が大きい 14 個を抽出して，セッションデータに加工し，既存の攻撃，未知の攻撃，正常通信の 3 つのラベル付けをして配布されているものである．Kyoto2006+におけるセッションデータへの加工時の課題について，改良も加えている．

考察

KDD99 の特徴量の中で特に影響が大きな 14 個を引き継いだデータセットであり，KDD99 のデータが古いという課題に対応して，2015 年 12 月までのデータを収集している．そのため，開発した手法の最新の攻撃に対する検証も行うことができる．

2.1.4 MWS **データセット**

マルウエア対策研究人材育成ワークショップ(MWS)[17]は，日本でデータセットを配布している代表的な団体である．MWS では，ポット観測データや研究者コミュニティから提供されたデータを「研究用データセット」として提供している[18]．2017 年現在，配布されているデータセットは，次の通りである（サイト[17]より引用）．

BOS 2014～2017

総務省実証事業「サイバー攻撃解析・防御モデル実践演習の実証実験の請負」にて実施し、研究者コミュニティから提供された組織内ネットワークへの侵害活動を観測したデータ

FFRI Dataset 2013～2017

株式会社 FFRI で収集したマルウェアの動的解析ログ

NICTER Dataset 2013～2017

サイバー攻撃観測・分析・対策システム NICTER で収集したダークネットトラフィックデータ、メールサーバに届いたダブルバウンスメールのデータ

CCC DATAset 2008～2013

マルウェア検体を収録したボット観測データ群であり、CCC 運営連絡会が運用するサイバークリーンセンターハニーポットで収集したマルウェア検体とウイルス対策ソフト 6 製品での検知名をリスト化したデータ

D3M (Drive-by-Download Data by Marionette) 2010～2015

研究者コミュニティから提供された Web 感染型マルウェアデータ

NCD in MWS Cup 2014

MWS Cup 2014 会期中に収集したホワイトデータセット

PRACTICE Dataset 2013

総務省「国際連携によるサイバー攻撃予知・即応に関する実証実験」（略称：PRACTICE）の挙動観察システムで、マルウェアを長期観測した際の通信トラヒック（マルウェア感染後の通信挙動）を示すデータ

PRACTICE (AmpPot) Dataset 2015

インターネット上のオープンなサーバ (DNS, NTP 等) を踏み台にして通信を増幅させることでサービス妨害を行う分散反射型サービス妨害攻撃 (DRDoS 攻撃) を観測したデータセット

但し，MWS データセットの取得及び使用に際しては，（社）情報処理学会コンピュータセキュリティ研究会 MWS 組織委員会との間で使用条件などに関する契約（覚書など）を交わす必要があり，また，データに関して，ラベルづけはされていない．

考察

日本で配布されている代表的な評価用データセットである．最新の攻撃傾向を反映し，トラフィックデータやマルウエアの解析ログなど多様なデータを提供しているが，攻撃通信と正常通信の正解ラベルが付されていないため，研究者ごとに正解が異なる可能性がある．

2.2 提案

評価用データセットに関して，信頼性，最新性，有効性という 3 つの観点から検討を行ってきた．その結果として，評価用データセットによる検証を有益なものにするためにそれぞれの観点からの提案を行う．加えて，継続性と最新性の両立，再現性についても提案を行う．

（提案 1.）信頼性に関する提案

正解ラベルが付与されている必要がある

（提案 2.）最新性に関する提案

シンポジウムなどで専門家からの意見を得た結果, できれば 3 年前以降のデータ, 少なくとも 5 年前以降のデータが含まれている

（提案 3.）有効性に関する提案

シンポジウムなどで専門家からの意見を得た結果, 攻撃通信の割合は，多くとも 5%程度に抑える

なお，（提案 1）は，文献[2][3]の sharafaldin らの意見を採用したものである．また，（提案 2）に関しては，多田らの文献[5]（2017 年出版で 2015 年の攻撃傾向について最新の攻撃傾向と標記）とシンポジウムなどで専門家から，「2017 年時点で 2015 年の攻撃傾向は最新のものと言える」という意見を得たものを反映したものである．（提案 3）に関しては，「実際のモニタでは攻撃通信の割合は 1%以下であることが多く，特に攻撃を受けているときにはそれよりも高くなるため, 評価用データセットの攻撃通信の割合としては 5%程度であれば，実践的と言える」という複数の専門家の意見を反映させたものである．

（提案 4）継続性と最新性の両立に関する提案

また，既存の研究からの継続性という観点から，現在でも，評価用データセットとして KDD99 が使用されている．そこで，背景となっている既存の論文の評価結果との比較という観点と最新の攻撃に対する評価を合わせて行うため，KDD99 と Kyoto2016 など複数のデータセットを用いて，それぞれ評価を行うことを加えて提案する．

（提案 5）再現性に関する提案

更に, 科学技術論文の基本である読者による検証可能性の観点からは，評価用データセットに関して，すべて公開されているデータセットを用いることも合わせて提案する．

このようにすることによって, 過去の評価結果との比較や他者による検証が可能である，信頼性と最新性，有効性を兼ね備えた評価が行えると考える．

3. まとめ

本稿では，NIDS の研究にあたり，現状を分析し，その課題を，評価用データセットについて，信頼性，最新性，有効性の 3 つの観点から指摘し，その解決策を提案した．信頼性の観点からは，結果の一貫性を見据えて，正解ラベル付きの評価用データセットを用いるべきであるという提案を行い, 最新性の観点からは, シンポジウムなどの結果から 3 年前以降のデータを含むデータセットを用いることを提案，有効性の観点からは，こちらもシンポジウムなどの結果から, 攻撃通信の割合が多くとも 5%程度であるデータセットを用いるなど，具体的な数値基準を示して提案した．

また, 既存の論文の評価結果との比較と最新の攻撃傾向に対する評価を合わせて行うため，KDD99 と Kyoto2016 など複数のデータセットでの評価も提案した．

更に, 科学技術論文の原則である読者による検証が可能性を保証するため, 評価にはすべて公開されているデータセットを用いることを提案した．

今後，自らの提案手法も含め，既存の様々な手法について，本稿で提案した評価方法を実践することによって，信頼性，最新性，有効性を兼ね備えた検証を行っていく予定である．

謝辞 本研究は JSPS 科研費 JP17K00187 の助成を受けたものです．この場を借りて，感謝の意を表します．

参考文献

[1] 高原尚志, “ネットワーク型侵入検知における研究結果の検証に関する一検討,” Software Engineering Symposium (SES2018) (査読中).

[2] Iman Sharafaldin, Arash Habibi Lashkari and Ali A. Ghorbani, “Toward Generating a New Instrusion Detection Dataset and Intrusion Traffic Characterization,” In Proceedings of the eth International Conference on Informnation Systems Security and Privacy (ICISSP 2018), pp.108-116, (2018).

[3] Iman Sharafaldin, Amirhossein Gharib, Arash Habibi Lashkari and Ali A. Ghorbani, “Towards a Reliable Intrusion Detection Benchmark Dataset,” Journal of Software Networking, Vol.2017, Issue 1, pp.177-200, (2018).

[4] “IDS 2017 Datasets Research Canadian Institute for Cybersecurity UNB (online),” available from <http://www.unb.ca/cic/datasets/ids-2017.html> (accessed 2018-04-26).

[5] 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，Volume 58, No.9, pp1450-1463, (2017).

[6] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set,” Proc. the 2009 IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISA 2009), pp.53-58 (2009).

[7] Atilla Ozgur, Hamit Erdem, “A review of KDD99 dataset usage in intrusion detection and machine learning between 2010 and 2015 (online),” available from <https://peerj.com/preprints/1954/> (accessed 2018-04-26).

[8] “KDD Cup 1999 Data,” available from < http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html> (accessed 2018-04-26).

[9] Saffa O. Al-mamory, Firas S. Jassim, “Evaluation of Different Data Mining Algorithms with KDD CUP 99 Data Set,” Journal of Babylon University, Pure and Applied Sciences, vol.21, no.8, p.2663-2681 (2013).

[10] “NSL-KDD Datasets Research Canadian Institute for Cybersecurity UNB (online),” available from < http://www.unb.ca/cic/datasets/nsl.html> (accessed 2018-04-26).

[11] S. Revathi, A. Malathi, “A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection,” International Journal of Engineering Research & Technology (IJERT), vol.2, Issue 12, p.1848-1853

[12] “MIT Lincoln Laboratory DARPA Intrusion Detection Evaluation (online),” available from <https://www.ll.mit.edu/ideval/data/1998data.html> (accessed 2018-04-26).

[13] “Traffic Data from Kyoto University's Honeypots (online),” available from <http://www.takakura.com/Kyoto_data/> (accessed 2018-04-30).

[14] Jungsuk SONG, Hiroki Takakura, and Yasuo Okabe, “Description of Kyoto University Benchmark Data (online),” available from < http://www.takakura.com/Kyoto_data/BenchmarkData-Description-v5.pdf> (accessed 2018-04-26).

[15] Jungsuk Song, Hiroki Takakura, Yasuo Okabe, Masashi Eto, Daisuke Inoue, Koji Nkao, “Statistical Analysis of Honeypot Data and Building of Kyoto 2006+ Dataset for NIDS Evaluation,” In Proceedings of The 1st International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS’11), pp.29-36, (2011).
[16] Jungsuk SONG, Hiroki TAKAKURA, Yasuo OKABE, “Cooperation of Intelligent Honypots to Detect Unknown Malicios Code,” in Proceedings of Workshop on Information Security Threats Data Collection and Sharing (WOMABT), pp.31-39, (2008).
[17] “マルウェア対策研究人材育成ワークショップ 2017 (MWS2017),” available from <https://www.iwsec.org/mws/2017/> (accessed 2018-04-30).
[18] 高田雄太, 寺田真敏, 村上純一, 笠間貴弘, 吉岡克哉, 畑田充弘, “マルウエア対策のための研究用データセット ～MWS Datasets 2016～ (online),” available from <http--www.iwsec.org-mws-2016-20160714-takata-dataset.pdf>
[19] 株式会社 FRRI, “FFRI Dataset 2017 FFRI Dataset 2017 FFRI Dataset 2017 のご紹介(online),” available from <https://www.iwsec.org/mws/2017/20170606/FFRI_Dataset_2017.pdf> (accessed 2018-05-10).
[20] 笠間貴弘, “NICTER DATASET 2017(online),” available from <https://www.iwsec.org/mws/2017/20170606/NICTER_Dataset_2017.pdf> (accessed 2018-05-10).
[21] 高田雄太, 秋山満昭, “MWS 2017 意見交換会 D3M (Drive-by Download Data by Marionette) データセット説明(online),” available form <https://www.iwsec.org/mws/2017/20170606/d3m.pdf> (accessed 2018-05-10).