ネットワーク型侵入検知システム評価用セッション型データセットに関する一考察

高原尚志[†1]

概要：インターネット上に次々と新たなサイバー攻撃が生産されている今日、サイバー攻撃を検知し、侵入を防ぐことは重要な課題である。侵入検知には、ホスト上のシステムログなどをもとに攻撃を検知するホスト型侵入検知システム(HIDS)とネットワーク上の通信データをもとに攻撃を検知するネットワーク型侵入検知(NIDS)がある。本稿では、NIDS に注目する。NIDS の評価には、評価用データセットを用いることが広く知られている。評価用データセットには、通信データそのものを提供するもの（PCAP 型）と通信データをセッション単位で加工したもの（セッション型）がある。本稿では、この中でもセッション型データセットに着目する。セッション型データセットでは、米国のイリノイ大学から提供されている KDD Cup 1999 Data(KDD99)と京都大学から提供されている Kyoto2016 Dataset などがあるが、本稿ではこの 2 つのデータセットに注目する。上記 2 つのデータセットは、攻撃通信か正常通信かのラベル付けがなされているため（KDD99 の場合は攻撃の種類も示されている）、NIDS 手法を統一した基準に基づいて評価することができ、NIDS の評価では広く用いられている。本稿では、上記 2 つのデータセットについてその特徴を明らかにし、考察を加える。

キーワード：KDD Cup 1999 Data，Kyoto2016，NIDS，PCAP 型, Session 型

A Consideration on Dataset for Evaluation of Network-based Intrusion Detection System

HISASHI TAKAHARA[†1]

Abstract: Today, there are new cyber-attacks on the Internet. It is very important that we detect and protect cyber-attacks. There are HIDS(Host-based Intrusion Detection System) and NIDS(Network-based Intrusion Detection System) in IDS(Intrusion Detection System). In this paper, NIDS is focused on. When NIDS is evaluated, benchmark datasets for NIDS are usually used. In the benchmark datasets, there are PCAP-type and session-type. In PCAP-type raw traffic data is used and in session-type raw traffic data is remade in session-unit. In this paper session-type datasets are focused on. In session-type dataset, KDD Cup 1999 Data(KDD99) and Kyoto2016 Dataset are widely known. These datasets are labeled in whether attack or normal. Thus NIDS methods can be evaluated in collective criteria. In this paper, characteristics of those two datasets are clarified and some considerations about those datasets are given.

Keywords: KDD Cup 1999 Data，Kyoto2016，NIDS，PCAP-type, Session-type

1. はじめに

1.1 研究の背景

インターネット上では日々新たな攻撃が開発され、従来の定義ファイルを利用した攻撃検知手法では、対応することが困難になっている。そこで、近年、人工知能(AI: Artificial Intelligence)の手法を利用した検知システムが開発され、新たな攻撃に備える傾向がある。AI の手法とは、具体的には機械学習(ML: Machine Learning)の手法で、サポートベクターマシン(SVM: Support Vector Machine)、ナイーブベイズ(NB: Naïve Bayes)、決定木（DT: Decision Tree）などが広く知られているが、その応用として、決定木を組み合わせたランダムフォレスト(RF: Random Forest)などのアンサンブル手法も広く知られている[1]。上記の ML 手法は、過去のデータ（学習データ）から現在のデータ（評価データ）を評価する手法で、教師あり学習と呼ばれているが、学習データを用いずに評価データを評価する教師なし学習手法も用いられる機会が多くなりつつある。代表的な教師なし学習手法としては、クラスタリングが広く知られている。

侵入検知システムには、ネットワーク上の通信データをもとに検知を行うネットワーク型侵入検知システム（NIDS: Network-based Intrusion Detection System）とホストのシステムログなどをもとに検知を行うホスト型侵入検知システム（HIDS: Host-based Intrusion Detection System）がある[2]が、本稿では NIDS に焦点を当てて議論を進める。

†1 新潟県立大学
University of NIIGATA PREFECTURE

ⓒ2018 Information Processing Society of Japan

新たな検知手法を開発した場合、評価用のデータセットを用いてその性能を評価することが広く知られているが、NIDS で用いる評価用データセットには、加工しない通信データ（PCAP データ）をそのまま収録したデータセット（PCAP 型データセット）と通信データをセッション単位で加工したデータセット（セッション型データセット）がある[3]。PCAP 型データセットは、加工しない個別の通信データを用いているため、それぞれの通信データが後に攻撃通信となるのか、正常通信なのかを判断することが難しい。これに対してセッション型データセットは、セッション単位にまとめるため、セッション全体が攻撃である場合には攻撃通信、正常なセッションであれば正常通信と正解ラベルを付与することが PCAP 型データセットに比べると比較的容易に行うことができる。そのため、PCAP 型データセットには正解ラベルが付与されていないものが多く、セッション型データセットには正解ラベルが付与されているものが多い。正解ラベルが付与されていない場合、攻撃通信か正常通信かの判断は、データセットを用いるユーザに委ねられる。先ほど、PCAP 型データセットは、個別の通信データとなるため、正解ラベルを付与するのが難しいと述べたが、そのデータセットを用いるユーザにとっても正解ラベルを付与する難しさに変わりはない。そのため、正解ラベルが付与されていない場合､ユーザごとに正解(攻撃通信か正常通信かの判断）が異なる可能性があり、この場合、手法を評価する上では大きな課題となる。これに対して、セッション型データセットでは、正解ラベルが付与されている場合が多く、ユーザに依存せず、正解（攻撃通信か攻撃通信かの判断)を統一することができる｡そこで、本稿では、セッション型データセットに着目して議論を進める。

代表的なセッション型データセットには、KDD 1999 Data(KDD99)[4]と Kyoto Data(Kyoto)[5]があるが、本稿ではこの 2 つのデータセットに着目して、特徴量などの特徴やその用途について比較検討する。

1.2 研究の動機

NIDS の研究において､開発した手法を評価する場合に、KDD99 や Kyoto などの既存の評価用データセットが多くの論文で用いられている。しかし、上記のデータセットを用いる根拠や上記のデータセットを用いた場合の開発した手法の性能評価に関するメリットなどを詳細に用いた論文は著者が知る限り存在しない。しかし本来は、なぜそのデータセットを用い、そのデータセットを評価に用いることによって開発した手法の性能をどのように評価できるのかなどについて論じる必要があると考えられる。

そこで本稿では､KDD99 と Kyoto という 2 つの広く用いられているデータセットに焦点を当て、その特徴や評価に用いた場合のメリットなどについて論じる。本来は、PCAP 型のデータセットについても論じる必要があるが、まずはセッション型のデータセットについて議論を深め、PCAP 型データセットを論じる際の試金石とする予定である。

1.3 既存研究

1.3.1 KDD Cup 1999 Data

***概要**

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（DARPA 1998）[6]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[8]がある（図 1）．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．

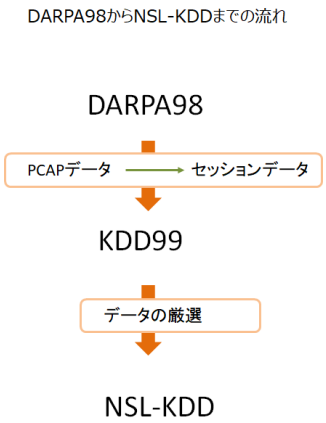


図 1. DARPA98 から NSL-KDD までの流れ
Figure1. Flow from DARPA98 to NSL-KDD

***既存研究**

Lippmann らは、文献[7]の中で、DARPA98 について、取得方法から統計的な分析まで詳細に説明している。

Saffaa らは、文献[9]の中で、KDD99 や機械学習の手法に

ⓒ2018 Information Processing Society of Japan

ついての今までの研究を整理し、機械学習の手法ごとにTPR(True Positive Rate)やFAR(False Alarm Rate)、学習データからモデルを作成するまでの時間などを示している。

Mahbod らは、文献[10]の中で、KDD99 の中で、データが冗長的であるなど KDD99 の課題を指摘し、その解決方法として、KDD99 から冗長データを削除した新たなデータセット NSL-KDD を紹介している。

Nelcileno らは、文献[11]の中で、KDD99 の 41 個の特徴量を IG(Information Gain)[1]の値にもとづいて、ランク付けしている。また Wang らは、文献[12]の中で、KDD99 の 41 個の特徴量について、KDD99 で示されている 4 つの攻撃カテゴリ（DoS、Probe、R2L、U2R）ごとに、IG の上位 10 個の特徴量を示している。

1.3.2 Kyoto Data

***概要**

Kyoto2016[3][13]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．

***既存研究**

Song らは、文献[14]の中で、Kyoto2006+の特徴量について説明している。また、Song らは、文献[15]の中で、Kyoto2006+を作成するためのハニーポットにおけるデータの収集方法について解説をしている。更に、Song らは、文献[16]の中で、Kyoto2006+について、正常通信や既知攻撃、未知攻撃などの割合やアクセスして来た国や地域の割内など、統計的な分析を行うとともに、特徴量を 14 個に絞った経緯などについて説明している。

多田らは、文献[3]の中で、Kyoto2006+を更に発展させたKyoto2016 Dataset(Kyoto)について、収集期間や統計的な分析などを行っている。

1.4 解決すべき課題

解決すべき課題としては、KDD99、Kyoto を個別に論じた文献はあるが、両者を比較検討した論文は、著者が知る限り存在しない。そのため、両者を比較して、その特徴を明らかにすることによって、次の事柄を明らかにする。

・開発した手法の評価にどのデータセットを用いるべきか
・用いた場合どのような性能をどのように評価できるのか

これにより、開発した手法について、用いるデータセットを論理的に選択できるようにする。

1.5 本研究の貢献

本論文の学術的貢献は以下の通りである。

・広く用いられている NIDS 評価用データセットである KDD99 と Kyoto について比較検討し、手法の評価に用いる際の参考資料を提供する
・開発した NIDS の手法の性能を評価するために、どのようなデータセットをどのような理由で用いるべきかについて明らかにする

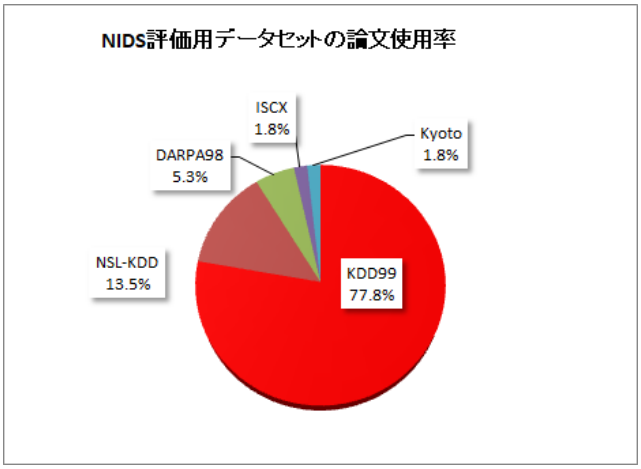
1.6 関連研究

1.6.1 NIDS 評価用データセットの使用率

Atilla らは，文献[17]の中で、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．

文献[17]によれば，NIDS の評価用に用いられるデータセット（PCAP 型データセットも含む）の内 77.8%は KDD99 である(図 2).

図 2. NIDS 評価用に用いられているデータセットの割合
Figure 2. Ratio of Dataset for Evaluation of NIDS



また，KDD99 と 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 とを合わせると 91.2%の論文が使用していることになる．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， Kyoto も， 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.）.

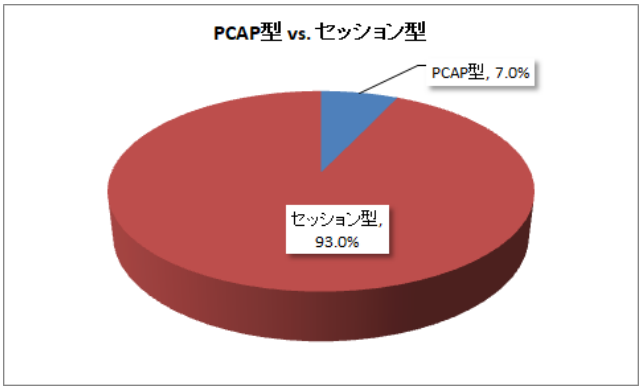
ⓒ2018 Information Processing Society of Japan

表 2. 多く用いられているデータセット
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 より一部抜粋）
Table2. Most used Datasets

Dataset Name	Article Count
KDD99	133
NSL-KDD	23
DARPA	9
ISCX	3
Kyoto	3

また、文献[17]によれば、セッション型 NIDS 評価用データセットの NIDS の評価に用いられる割合は、使用される割合は、PCAP 型が 7.0%、セッション型が 93.0%で、ほとんどの論文が NIDS の評価にセッション型データセットを用いているという結果になった（図 3)。

図 3. PCAP 型とセッション型の論文における使用率
Figure3. Ratio of PCAP-type and Session-type in papers



1.6.2 NIDS 評価用データセットの必要要件

Sharafaldin らは，文献[18][19]の中で、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の必要要件として, 以下の 11 個の基準を提唱した.

●Sharafaldin らが提唱した 11 個の基準

1.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

2.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

3.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

4.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

5.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

6.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

7.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

8.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

9.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

10.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

11.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、1.1 でも指摘したように、基準 3 のラベル付けがなされているか否か、つまり攻撃通信か正常通信かの正解ラベルが提供されているか否かが重要と考え、正解ラベルが提供されている KDD99 と Kyoto に注目して議論を展開する.

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.）. なお、この内、本稿の表 1 では，広く用いられているセッション型データセットである KDD99 と京都大学が提供する日本のセッション型データセットである KYOTO のみ抜粋した。

ⓒ2018 Information Processing Society of Japan

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[19]の Table1.より引用（一部抜粋））

Table1. Comparing of available datasets based on evaluation framework

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

多田らは，文献[3]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[10]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．

2. KDD99 と Kyoto の比較検討

KDD99 は、NIDS の評価用データセットとして広く知られているが、以下のような課題が指摘されている。

***KDD99 の課題**

・1999 年に作成されたため、最新の攻撃に対する評価ができない[3]

・ACM の SIGKDD(The community for data mining, data science and analytics)[20]が主催した DATA Mining の競技会である KDD Cup 1999[21]用に作成されたため、攻撃通信を疑似的に加えているため、実際のネットワーク環境を反映できていない[3]

・意図的に通信データを加えており、通信データが冗長的である[10]

上記の課題を解決するため、京都大学のハニーポットを用いて作成されたのが Kyoto2016 Dataset(Kyoto)である。Kyoto は、以下の特徴を持つ[3]。

***Kyoto の特徴**

・最新の攻撃傾向を含む（2015 年 12 月までのデータ）

・実際のトラフィックから作られている（ハニーポットで収集されたデータのみ使用）

・長い収集期間（2006 年 11 月から 2015 年 12 月までのおよそ 10 年間）

Kyoto では、KDD99 の 41 個の特徴量から、NIDS において正常通信と攻撃通信の識別に大きく影響する 14 個の特徴量を抽出して使用している[15]。これにより、KDD99 との継続性を維持することができると考えられる。

***攻撃通信と正常通信の割合**

また、KDD99 における攻撃通信と正常通信の割合では、攻撃通信と正常通信の割合がおおよそ 8 対 2 であり、KDD99 の冗長性などを修正した NSL-KDD では、おおよそ 3 対 7 と割合が逆転しており、Kyoto ではおおよそ 8 対 2 となっている。シンポジウムなどでの専門家の議論では、実際のネットワークでの攻撃の割合は通常は 1%未満のことが多く、実際に運用されているネットワーク環境を反映させるには、Kyoto の作成者である国立情報学研究所の高倉弘喜教授との検討では攻撃通信の割合を 5%未満に抑える必要があるとの結果となった。このため、NIDS の評価に KDD99 や NSL-KDD、Kyoto を用いる場合には、攻撃を種類別に抽出するなどの工夫が必要となると考えられる。

また、KDD99 がすべてのデータとその一部（10%）を抽出したデータ及び評価用データのみであったのに対して、Kyoto は日にちごとに個別のデータが用意されているため、工夫次第では様々な種類の評価が可能となると考えられる。なお、KDD99 でも攻撃を種類別に抽出して手法を評価するなどして、評価にバリエーションを与えることもできる。

***KDD99 と Kyoto 併用の提案**

最後に、KDD99 は長年広く用いられているデータセットであるが、近年、上で示した課題が指摘されている。そこで、KDD99 で評価を行った過去の手法との比較を最新の攻撃が入ったデータセットで行えるように Kyoto が提供されている。従って、KDD99 での評価と Kyoto での評価を併用することにより、過去の手法の評価との比較を行いながら、最新の攻撃に対する評価も加えた評価が行えるものと考える。

ⓒ2018 Information Processing Society of Japan

3. まとめ

本稿では、NIDS 評価用データセットの中で、セッション型データセットとして広く用いられている KDD99 と Kyoto について、それぞれの解説と 2 つのデータセットの比較検討を行った。その結果、KDD99 と Kyoto を併用することにより、過去の研究結果を加味しながら、最新の攻撃への対応評価が行えることを示した。

今後は、具体的な手法を用いて、上記を実践して行く予定である。その際に、課題などがあれば、適宜解決して、論文として合わせて報告する予定である。

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．

文　献

[1] 荒木雅弘, “フリーソフトではじめる機械学習入門,” 森北出版株式会社(2014).

[2] Glass-Vanderlan, Tarrah R., Iannacone, Michael D., “A Survey of Intrusion Detection Systems Leveraging Host Data (online),” available from <https://arxiv.org/pdf/1805.06070.pdf> (accessed 2018-07-02).

[3] 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).

[4] “ KDD Cup 1999 Data,” available from < http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html> (accessed 2018-08-08).

[5] “Traffic Data from Kyoto University's Honeypots (online),” available from <http://www.takakura.com/Kyoto_data/> (accessed 2018-08-09).

[6] “MIT Lincoln Laboratory DARPA Intrusion Detection Evaluation (online),” available from <https://www.ll.mit.edu/ideval/data/1998data.html> (accessed 2018-08-09).

[7] Richard P. Lippmann, David J. Fried, Issac Graf, Joshua W. Haines, Kristoper R. Kendall, David McClung, Dan Weber, Seth E. Webster, Dan Wyschogrod, Robert K. Cunningham, and Marc A. Zissman, ”Evaluating Intrusion Detection Systems: The 1998 DARPA Off-line Intrusion Detection Evaluation,” Proceedings DARPA Information Survivability Conference and Exposition(DISCEX'00), Volume2, pp.12-26(2000).

[8] “ NSL-KDD Datasets Research Canadian Institute for Cybersecurity UNB (online),” available from < http://www.unb.ca/cic/datasets/nsl.html> (accessed 2018-06-28).

[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] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set,” Proc. 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISA 2009), pp.1-6 (2009).

[11] Nelcileno Araujo, Ruy de Oliverira, Alilton Akira Shinoda, Bharat Bhargava, “Identifying Important Characteristics in the KDD99 Intrusion Detection Dataset by Feature Selection using a Hybrid Approach,” Proceeding of 2010 17th International Conference on Telecommunications, pp.552-558(2010).

[12] Wei Wang, Sylvain Gombult, Thomas Guyet, “Toards Fast Detecting Intrusions: Using Key Attributes of Network Traffic,” Proceeding of 2008 The Third International Conference on Internet Monitoring and Protection(ICIMP), pp.5-11(2008).

[13] “Traffic Data from Kyoto University's Honeypots (online),” available from < http://www.takakura.com/Kyoto_data/ > (accessed 2018-08-09).

[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 2018-08-09).

[15] 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).

[16] 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).

[17] 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)

[18] 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).

[19] 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).

[20] “SIGKDD (online),” available from < http://www.kdd.org/> (accessed 2018-08-09).

[21] “SIGKDD KDD Cup 1999 Computer network intrusion detection(online),” < http://www.kdd.org/kdd-cup/view/kdd-cup-1999/Data> (accessed 2018-08-09).

ⓒ2018 Information Processing Society of Japan