An Evaluation Framework for Intrusion Detection Dataset

Amirhossein Gharib*, Iman Sharafaldin†, Arash Habibi Lashkari‡ and Ali A. Ghorbani§
Canadian Institute for Cybersecurity (CIC)
University of New Brunswick (UNB), Fredericton, Canada
*agharib@unb.ca, †isharafa@unb.ca, ‡a.habibi.1@unb.ca, §ghorbani@unb.ca

***Abstract*—The growing number of security threats on the Internet and computer networks demands highly reliable security solutions. Meanwhile, Intrusion Detection (IDSs) and Intrusion Prevention Systems (IPSs) have an important role in the design and development of a robust network infrastructure that can defend computer networks by detecting and blocking a variety of attacks. Reliable benchmark datasets are critical to test and evaluate the performance of a detection system. There exist a number of such datasets, for example, DARPA98, KDD99, ISC2012, and ADFA13 that have been used by the researchers to evaluate the performance of their intrusion detection and prevention approaches. However, not enough research has focused on the evaluation and assessment of the datasets themselves. In this paper we present a comprehensive evaluation of the existing datasets using our proposed criteria, and propose an evaluation framework for IDS and IPS datasets.**

***Keywords*—Intrusion Detection, Intrusion Prevention, IDS, IPS, Evaluation Framework**

I. Introduction

Intrusion detection has attracted the attention of many researchers in identifying the ever-increasing issue of intrusive activities. In particular, anomaly detection has been the main focus of many researchers because of its potential in detecting novel attacks. However, its adoption by real-world applications has been hampered due to system complexity as these systems require a substantial amount of testing, evaluation, and tuning prior to deployment. Evaluating these systems using labeled traffic with an extensive set of intrusions and abnormal behavior is ideal but not always possible. Hence researchers normally resort to datasets that are often sub-optimal.

As network behaviors and patterns change and intrusions evolve, it is necessary to move away from static and one-time datasets towards dynamically generated datasets, which not only reflect the traffic compositions and intrusions, but are also modifiable, extensible, and reproducible [1]. Moreover, selecting a suitable dataset is a significant challenge itself. Because on one hand, many such datasets are internal and cannot be shared due to the privacy issues, and on the other hand, those that become avaliable are heavily anonymized and do not reflect the current trends. Because of the lack of certain statistical characteristics and the unavailability of these datasets a perfect dataset is yet to be realized [1], [2]. It is also necessary to mention that due to malware evolution and the continuous changes in attack strategies, benchmark datasets need to be updated periodically [1].

The rest of paper is organized as follows. An overview of the datasets generated between 1998 and 2016 is presented in Section II. Section III discusses the previous evaluation frameworks and provides details of our new framework. Section (IV) presents an assessment and evaluation of the available datasets using the proposed framework.

II. Available Datasets

In this section, some of the existing IDS datasets will be evaluated from the perspective of demonstrating the need for IDS datasets that reflect the characteristics of a worthy dataset.

DARPA (Lincoln Laboratory 1998,1999): This dataset was constructed for network security analysis purposes. Researchers criticized this dataset for issues associated with the artificial injection of attacks and benign traffic. This dataset includes send and receive mail, browse websites, send and receive files using FTP, use telnet to log into remote computers and perform work, send and receive IRC messages, monitor the router remotely using SNMP activities. It contains attacks such as DOS, Guess password, Buffer overflow, Remote FTP, Syn flood, Nmap, and rootkit in this dataset. This dataset does not represent real-world network traffic and contains irregularities such the absence of false positives, and is outdated for the effective evaluation of IDSs on modern networks both in terms of attack types and network infrastructure. Moreover, it lacks actual attack data records [3] [4].

KDD'99 (University of California, Irvine 1998,99): The KDD Cup 1999 dataset was created by processing the tcpdump portion of the 1998 DARPA dataset, which nonetheless suffers from the same issues. This dataset includes more than twenty attacks such as neptune-dos, pod-dos, smurf-dos, buffer-overflow, rootkit, satan, teardrop, to name a few [5].The network traffic records of normal and attack traffics. They are merged together in a simulated environment. This dataset has a large number of redundant records and is studded by data corruptions that led to skewed testing results [6]. NSL-KDD was created using KDD dataset [6] to address some of the KDD's shortcomings [3].

DEFCON (The Shmoo Group, 2000): Generated in 2000, DEFCON-8 dataset contains port scanning and buffer overflow attacks, whereas DEFCON-10 dataset was created in 2002 using port scan and sweeps, bad packets, administrative privilege, and FTP by telnet protocol attacks. The traffic produced during the"capture the Flag (CTF)" competition is different from the real world network traffic since it mainly consists of intrusive traffic as opposed to normal background traffic. This dataset is used to evaluate alert correlation techniques [7] [8].

978-1-5090-5493-0/16/$31.00 ©2016 IEEE

CAIDA (Center of Applied Internet Data Analysis - 2002/2016): CAIDA consists of three different types of datasets: 1) CAIDA OC48, which includes different types of data observed on an OC48 link in San Jose and provided by CAIDA members, DARPA, NSF, DHS, Cisco; 2) CAIDA DDOS attack dataset, which includes one-hour DDoS attack traffic split of 5-minute pcap files; and 3) CAIDA Internet trace 2016, which is passive traffic traces from CAIDA's equinix-chicago monitor on High-speed Internet backbone. Most of CAIDAs datasets are very specific to particular events or attacks and are anonymized with their payload, protocol information, and destination. This dataset is not an effective benchmarking datasets due to a number of shortcomings, see [9] [10] [11] [12] [2] for details.

LBNL (Lawrence Berkeley National Laboratory and ICSI - 2004/2005): LBNL's internal enterprise traces are full header network traffic recorded at a medium-sized site. This is a dataset without payload and suffers from a heavy anonymization to remove any information which could identify an individual IP [13].

CDX (United States Military Academy 2009): The CDX dataset represents the network warfare competitions can be utilized to generate modern day labeled dataset. In this dataset common attack tools namely Nikto, Nessus, and WebScarab have been used by attackers to carry out reconnaissance and attacks automatically. Benign traffic includes web, email, DNS lookups, and other required services. CDX can be used to test IDS alert rules but it suffers from the lack of traffic diversity and volume [14].

Kyoto (Kyoto University - 2009): This dataset has been created through honypots, so there is no process for manual labeling and anonymization, but it has limited view of the network traffic because only attacks directed at the honeypots can be observed. It has ten extra features such as IDS_Detection, Malware_Detection, and Ashula_Detection than previous available datasets which are useful in NIDS analysis and evaluation. As normal traffic has been simulated repeatedly during the attacks and producing only DNS and mail traffic data, which is not reflecting real world normal traffic so there are no false positives, which are important for minimizing the number of alerts [15] [16] [17].

Twente (University of Twente - 2009): To creating this dataset, three services OpenSSH, Apache web server and Proftp using auth/ident on port 113 have been installed and collected data from a honypot network by netflow. There are some side-effect traffic such as auth/ident, ICMP, and irc traffic which are not completely benign or malicious. Moreover, it contains some unknown and uncorrelated alerts traffic. This dataset is labeled and is more realistic but the lack of volume and diversity of attacks is obvious [18].

UMASS (University of Massachusetts - 2011): The dataset includes trace files which are network packets and some traces on wireless applications [19] [1]. It has been generated using a single TCP-based download request attack scenario. The dataset is not useful for testing IDS and IPS techniques due to the lack of variety of traffic and attacks [20].

ISCX2012 (University of New Brunswick - 2012): This dataset was generated by a dynamic approach and the authors present good guideline for generating realistic and useful IDS evaluation datasets. Their approach consists of two parts. In the first part, Alpha profile carried out various multi-stage attacks scenarios to stream the anomalous segment of the dataset. The second part, Beta profile, which is the benign traffic generator, has generated realistic network traffic with background noise. Real traces are analyzed to create profiles to generate real traffic for HTTP, SMTP, SSH, IMAP, POP3, and FTP protocols. The dataset generated by this approach consists of network traces with full packet payloads and relevant profiles. However, it does not represent new network protocols since near 70% of todays network traffics are HTTPS and there are no HTTPS traces in this dataset. Moreover, the distribution of the simulated attacks is not based on real world statistics [2].

ADFA (University of New South Wales - 2013): To create ADFA dataset, authors installed Apache, MySQL and Tikiwiki for offering web server, Datasabse server, remote access and FTP server. This dataset includes FTP and SSH password bruteforce, Java based Meterpreter, Add new Superuser, Linux Meterpreter payload and C100 Webshel attack vectors. It contains normal training and validating data and 10 attacks per vector [21]. In addition to lack of attack diversity and variety of attacks, the behaviors of some attacks in this dataset are not well separated from the normal behavior [22] [23].

III. Framework

Over the years network attacks have evolved from a simple low-scale to a more sophisticated large-scale problem. During this period there has been considerable research and development into attack detection strategies, but only limited research on testing these techniques against realistic data. One of the key reasons was the legal and privacy issues associated with sharing captured data. As a result, the majority of the published works are evaluated by:

Replaying publicly available datasets (F1): One obvious way to test and evaluate a detection system is to replay traffic of real attacks [24] [25].

Generating traffic (F2): Artificial generation would seem to be the only practical way to generate both attack and benign traffic. Unfortunately, both hardware and software traffic generators are far from being ideal for simulating such attacks. Curl-loader is one of the open source tools to generate artificial traffic.

Testbed design strategies: An essential requirement for deploying synthetic traffic traces is to first have an experimental setup and a traffic generation software. There are three commonly used testbed design strategies:

- Simulation (F3): In this method, attackers, targets and network devices are all simulated [26]. NS-2 and Opnet are two examples of network simulators.

- Emulation (F4): It is a step forward in realism over simulation. Real machines are used as attackers and targets, and only the network topology is recreated in software [27].

- Direct physical representation (F5): In this technique, the desired network topology is built by physically arranging a network of routers, switches, and computers [28].

TABLE I: Comparison of traffic generation strategies

	F1	F2	F3	F4	F5
Reconfigurability	Yes	Yes	Yes	No	No
Low cost	Yes	Yes	Yes	Yes	No
Monitoring	No	No	No	No	Yes
Attack diversity	No	No	Yes	Yes	Yes
Scalability	No	Yes	Yes	No	No
Interactive traffic	No	No	Yes	Yes	Yes

The literature reviewed above highlights the shortcomings of various approaches used for obtaining realistic datasets. Table I provides a comparative analysis of various traffic generation strategies. Although significant studies have been done on IDS dataset generation, little research conducted on the evaluation and assessment of IDS and IPS datasets. In proposing a new framework, we have taken into account this evaluation and assessment that has been done.

Scott *et al.* presented three major criteria in datasets: redundancy, inherent unpredictability and complexity or multivariate dependencies [29]. Heidemann and Papadopoulos used trace data in research to find common problems that cut across types of data and defined four aspects namely Privacy and Anonymization, Unavailable type of data such as local-observation or local-inference, developing new techniques, and moving target and coverage. They suggested that one of the most important aspects even when some data already exist is continued observation [30].

Ghorbani *et al.* discussed the IDS and evaluation criteria of these systems and believe that datasets are valuable as sets in this domain. But as there are some issues in creation of these datasets such as synthetic traffic generation or difficulties in collecting real attack scripts and victim software, all of them suffer from the fact that they are not good representatives of the real world traffic [31].

Nehinbe outlined an evaluation based on previous works regarding some aspects such as privacy issues, approval from owners, documentation problems, labeling and anonymity [1]. Shiravi *et al.* defined evaluation criteria with six aspects: realistic network, realistic traffic, labeled dataset, total interaction capture, complete capture, and diversity of attacks. Also, they assessed five previous datasets based on these aspects [2].

Given the shortcomings of the existing datasets, we believe that a comprehensive and wholesome framework for generating IDS/IPS bench marking dataset is needed. The next of this section defines the features of such framework. In all, we define eleven features as follows.

1. Complete Network configuration: Having a complete computer network, in fact is the foundation of an offline dataset to represent the real world. Several attacks have revealed their true faces only in a perfect network which has all equipment such as number of PCs, server(s), router, and firewall. So it is necessary to have a realistic configuration in the testbed to capture the real effects of attacks.

2. Complete Traffic: Traffic is a sequence of packets from a source that can be a host, router, or switch to a destination, which may be another host, a multicast group, or a broadcast domain. Based on the traffic generation techniques it is possible to have realistic, pseudo-realistic, or synthetic traffic in a dataset. The pseudo-realistic has partially the real world traffic, such as having simulated human behavior traffics with real attack scenarios.

3. Labeled dataset: While a dataset for evaluating different discovery mechanisms in this domain is important, tagging and labeling data are also important. If there are no correct labels, without a doubt, it is not possible to use a dataset and the results of the analysis also are not valid and reliable. For example, in network datasets, after converting pcaps to netflows then it is possible to have reliable labels for flows which are more useful and understandable for users. But, These are labeled dataset which does not clearly state the name and type of the attacks and only labeled them as benign or malicious. In other words, it is possible to have unlabeled, partially-labeled, and fully-labeled datasets.

4. Complete Interaction: For the correct interpretation of the results evaluation, one of the vital features is amount of available information for anomalous behaviour. So, having all network interactions such as within or between internal LANs is one of the major requirements for a valuable dataset.

5. Complete Capture: Even in a complete traffic dataset, it is essential to capture all traffic for the researchers who want to evaluate their proposed detection systems. It seems some of the datasets are capturing traffic partially and removing part of the traffic which is Non-functional or not labeled while it is very influential to have all traffic together to calculate the false-positive percentage of an IDS system.

6. Available Protocols: There are many different types of traffic some of which are vital for testing an IDS system such as Bursty traffic which is an uneven pattern of data transmission and can cover some protocols such as HTTP and FTP. Interactive traffic includes sessions that consist of short request and response pairs such as applications involving real-time interaction with users (e.g., web browsing, online purchasing). In latency sensitive traffic the user has an expectation that data will be delivered on time such as VOIP and Video conferencing. In Non-Real-time traffic such as news and mail traffics, timely delivery is not important. A complete dataset should have both normal and anomalous traffic.

7. Attack Diversity: In recent years, threats have expanded their scopes into intricate scenarios such as application and app attacks. The type of attacks is changing and updating daily. So, having the ability to test and analyze IDS and IPS systems by these new attacks and threat scenarios is one of the most important requirements that an off-line dataset should support. We categorized attacks into seven major groups based on the 2016 McAfee report, Browser-based, Brute force, DoS, Scan or enumeration, Backdoors, DNS, and other attacks (e.g., Heartbleed, Shellshock, and Apple SSL library bug).

8. Anonymity: The privacy compromising issues occurs when both the IP and payload are available. So, most of the datasets removed their payload entirely which decreases the usefulness of the dataset especially for some detection mechanisms such as deep packet inspection (DPI).

9. Heterogeneity: In IDS domain, it is possible to have different sources for creating a dataset such as network traffic, operating systems logs, or network equipment logs. A

homogeneous dataset with one type of source can be useful for analyzing a specific type of detection systems while a heterogeneous dataset can be used for a complete test covering all aspects of the detection process.

10. Feature set: The main goal of providing a dataset is its usability for other researchers to test and analyze their proposed system. One of the main challenges is how to calculate and analyze the related features. It is possible to extract features from different type of data sources such as traffic or logs using feature extraction applications.

11. Metadata: Lack of a proper documentation is one of the main issues in available datasets in this area. Most of the datasets do not have documentation or even if they have it is not complete. Insufficient information about the network configuration, operating systems for attacker and victim machines, attack scenarios, and other vital information can detract from the usability of a dataset for researchers.

Equation (1) is useful to measure the proposed framework. In this equation, *W* as a flexibility coefficient is the weight of each feature which can be defined based on the organization request or type of the IDS system that has been selected for test. As we have eleven features in our framework, we should define eleven *W* for any scenario. *V* is the coefficient of each sub-factor that can be defined based on experiences or distribution of sub-factors in different scenarios. We have two features with sub-factors: attacks and protocols. In these features *V* should be defined for each different sub-factor as well. Also, F is the appearance of the specific factor and sub-factor in the dataset that can be binary (0 or 1) or multi-valued.

$$\sum_{i=1}^{n} W_i(\sum_{j=1}^{m} V_j * F_j) \quad (1)$$

Where n is the number of features that in our framework was 11 and m is the number of coefficients for each factor. In proposed framework, for two factors "attacks" and "protocols" value of m is 7 and 5 respectively but for the other factors $m = 1$. To better understand the equation, two datasets analysis in the next section with reliable value of *W* and*V*.

IV. EVALUATION

Table III shows our assessment of eleven available datasets which have been listed and explained in section II based on related documents and research. Some of the features values are not shown because of the lack of metadata and complete documentation. For applying the proposed framework to each dataset and calculating the score, it is necessary to define the *W* and *V* in a realistic scenario. Here we have selected two famous datasets KDD99 and KYOTO, to evaluate the framework.

As *W* is related to the organization or type of the IDS systems, we consider different values such as [0.05, 0.05, 0.1, 0.05, 0.05, 0.25, 0.25, 0.05, 0.05, 0.05, 0.05] in our scenario. As Protocol and attack factors have more values for us, we defined higher weighs (0.25) to them. Now, we should define two different distributions for the attack and protocol to define the *V* values for each factor. Based on McAfee report [32] distribution of seven attack categories are Browser (36%), Bruteforce (19%), DoS(16%), Scan(3%), DNS (3%), Backdoor(3%), Others (20%). It is necessary to mention that as SSL attacks are seasonal attacks and will not have fixed value in all times, we mix this attack with "Others" in our distribution. So, the *V* values for attack factors will be [0.36, 0.19, 0.16, 0.03, 0.03, 0.03, 0.20].

One of the other shortcomings of the available datasets is the distribution of the protocols. The rapid growth of Internet, has changed the protocols distributions very quickly and finding a valid document will be more difficult. So, we observed the traffic of our research center for one month to find this distribution. For six group of protocols usage percentage and distribution were http(10%), https(74%), ssh(2%),ftp(6%),email(1%), and other(7%). So, the *V* vaules for protocol factor will be [0.1, 0.74, 0.04, 0.08, 0.04].

Table II shows the value of the KDD and KYOTO datasets in this scenario which are 0.56 and 0.85, respectively. Figure 1 shows the comparison between KDD and KYOTO datasets based on the binary values from Table III and the scores from Table II.

TABLE II: KDD99 and KYOTO datasets scores

Dataset	Calculation	Score
KDD	0.05*1+ 0.05*0 + 0.1*1 + 0.05*1 + 0.05*1 + 0.25 * (0.1 + 0.0 + 0.04 + 0.08 + 0.04) + 0.25 * (0.0 + 0.19 + 0.16 + 0.03 + 0.0 + 0.0 + 0.2) + 0.05*0 + 0.05*0 + 0.05*1 + 0.05*1	0.56
KYOTO	0.05*1 + 0.05*0 + 0.1*1 + 0.05*1 + 0.05*1 + 0.25 * (0.1 + 0.74 + 0.04 + 0.08 + 0.04) + 0.25 * (0.36 + 0.19 + 0.16 + 0.03 + 0.03 + 0.03 + 0.2) + 0.05*0 + 0.05*0 + 0.05*1 + 0.05*1	0.85

V. CONCLUSIONS

In this paper, we have studied the exist datasets for the test and evaluation of IDSs, and presented a new framework to evaluate datasets with the following characteristics: Attack Diversity, Anonymity, Available Protocols, Complete Capture, Complete Interaction, Complete Network Configuration, Complete Traffic, Feature Set, Heterogeneity, Labeled Dataset, and Metadata. The proposed framework considers organization policy and conditions using a coefficient, W, which can be defined separately for each criterion. In the future, we plan to generate and make available a new dataset that we will support all the above criteria.

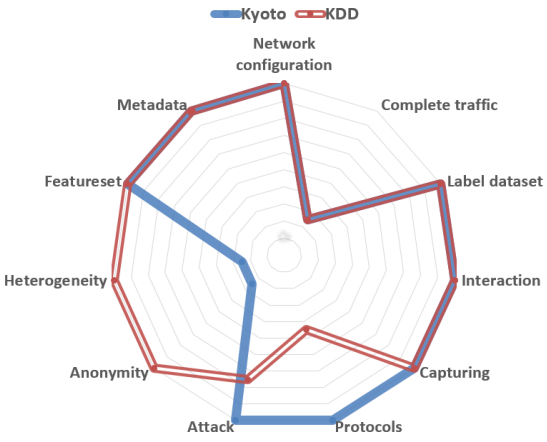

Fig. 1: KDD99 and KYOTO datasets evaluation

TABLE III: Comparison between datasets based on proposed evaluation framework

	Network	Traffic	Labeling	Interaction	Capturing	Protocols					Attack Diversity							~~Anonymity~~	Heterogeneity	Feature Set	Metadata
						http	https	SSH	FTP	Email	Browser	Bruteforce	DoS	Scan	Backdoor	DNS	Others				
DARPA	YES	NO	YES	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	NO	YES	NO	NO	No	No
KDD'99	YES	NO	YES	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	NO	YES	NO	NO	Yess	Yes
DEFCON	NO	NO	NO	YES	YES	YES	NO	YES	NO	NO	NO	NO	NO	YES	YES	NO	YES	-	NO	NO	NO
CAIDAs	YES	YES	NO	NO	NO	-	-	-	-	-	NO	NO	YES	YES	NO	YES	YES	YES	NO	NO	Yes
LBNL	YES	YES	NO	NO	NO	YES	NO	YES	NO	NO	-	-	-	YES	-	-	-	YES	NO	NO	NO
CDX	NO	NO	NO	YES	YES	YES	NO	YES	YES	YES	NO	NO	YES	YES	NO	YES	-	-	NO	NO	NO
KYOTO	YES	NO	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	NO	NO	YES	YES
TWENTE	YES	YES	YES	YES	YES	YES	NO	YES	YES	NO	NO	YES	NO	YES	NO	NO	YES	-	-	NO	YES
UMASS	YES	NO	YES	NO	YES	YES	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	YES	-	-	NO	NO
ISCX2012	YES	NO	YES	YES	YES	YES	NO	YES	YES	YES	YES	YES	YES	YES	YES	NO	YES	NO	YES	NO	YES
ADFA2013	YES	YES	YES	YES	YES	YES	NO	YES	YES	YES	YES	YES	NO	NO	YES	NO	YES	NO	-	NO	YES

REFERENCES

[1] J. O. Nehinbe, "A critical evaluation of datasets for investigating idss and ipss researches," in *IEEE 10th International Conference on Cybernetic Intelligent Systems (CIS)*, Sept 2011, pp. 92–97.

[2] M. T. Ali Shiravi, Hadi Shiravi and A. A. Ghorbani, "Toward developing a systematic approach to generate benchmark datasets for intrusion detection," *Computers and Security*, vol. 31, no. 3, pp. 357 – 374, 2012.

[3] J. McHugh, "Testing intrusion detection systems: A critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory," *ACM Trans. Inf. Syst. Secur.*, vol. 3, no. 4, pp. 262–294, Nov. 2000.

[4] C. Brown, A. Cowperthwaite, A. Hijazi, and A. Somayaji, "Analysis of the 1999 darpa/lincoln laboratory ids evaluation data with netadhict," in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, July 2009, pp. 1–7.

[5] I. U. University of California, "Kdd cup 1999," 2007. [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/

[6] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the kdd cup 99 data set," in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, July 2009, pp. 1–6.

[7] J. O. Nehinbe, *A Simple Method for Improving Intrusion Detections in Corporate Networks*. Springer Berlin Heidelberg, 2010, pp. 111–122.

[8] T. S. Group, "Defcon 8, 10 and 11," 2000. [Online]. Available: http://cctf.shmoo.com/

[9] "Caida data set oc48 link a (san jose, ca)," 2002. [Online]. Available: https://www.caida.org/data/passive/passive_oc48_dataset.xml

[10] "Caida ddos attack dataset," 2007. [Online]. Available: https://www.caida.org/data/passive/ddos-20070804_dataset.xml

[11] "Caida anonymized internet traces 2016 dataset," 2016. [Online]. Available: https://www.caida.org/data/passive/passive_2016_dataset.xml

[12] E. P. Proebstel, "Characterizing and improving distributed network-based intrusion detection systems(nids):timestamp synchronization and sampled traffic," Master's thesis, University of California DAVIS, CA, USA, 2008.

[13] V. P. A. G. Boris Nechaev, Mark Allman, "Lawrence berkeley national laboratory (lbnl)/icsi enterprise tracing project," 2004.

[14] T. C. R. F. E. D. W. J. A. C. M. G. C. Benjamin Sangster, T. J. OConnor, "Toward instrumenting network warfare competitions to generate labeled datasets." Usenix: The Advanced Computing System Association, 2009.

[15] J. Song, H. Takakura, Y. Okabe, M. Eto, D. Inoue, and K. Nakao, "Statistical analysis of honeypot data and building of kyoto 2006+ dataset for nids evaluation," in *Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*. ACM, 2011, pp. 29–36.

[16] H. T. M. Sato, H. Yamaki, "Unknown attacks detection using feature extraction from anomaly-based ids alerts," in *Applications and the Internet (SAINT), 2012 IEEE/IPSJ 12th International Symposium on*, July 2012, pp. 273–277.

[17] C. H. R. Chitrakar, "Anomaly based intrusion detection using hybrid learning approach of combining k-medoids clustering and naive bayes classification," in *8th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, Sept 2012, pp. 1–5.

[18] A. Sperotto, R. Sadre, F. Vliet, and A. Pras, "A labeled data set for flow-based intrusion detection," in *Proceedings of the 9th IEEE International Workshop on IP Operations and Management IPOM09*, 2009, pp. 39–50.

[19] U. of Massachusetts Amherst, "Optimistic tcp acking," 2011. [Online]. Available: http://traces.cs.umass.edu/

[20] B. N. L. Swagatika Prusty and M. Liberatore, "Forensic Investigation of the OneSwarm Anonymous Filesharing System," in *ACM Conference on Computer and Communications Security (CCS)*, October 2011.

[21] G. Creech and J. Hu, "Generation of a new ids test dataset: Time to retire the kdd collection," in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, 2013, pp. 4487–4492.

[22] M. Xie and J. Hu, "Evaluating host-based anomaly detection systems: A preliminary analysis of adfa-ld," in *Image and Signal Processing (CISP), 2013 6th International Congress on*, vol. 03, 2013, pp. 1711–1716.

[23] M. Xie, J. Hu, and J. Slay, "Evaluating host-based anomaly detection systems: Application of the one-class svm algorithm to adfa-ld," in *2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2014, pp. 978–982.

[24] R. M. Bill Buchanan, Flavien Flandrin and J. Graves, "A methodology to evaluate rate-based intrusion prevention system against distributed denial-of-service ddos," 2011.

[25] T. A. Ahmed Ejaz, Mohay George and B. Sajal, "Use of ip addresses for high rate flooding attack detection," pp. 124–135, 2010.

[26] J. Mirkovic, S. Fahmy, P. Reiher, and R. K. Thomas, "How to test dos defenses," in *Conference For Homeland Security, 2009. CATCH '09. Cybersecurity Applications Technology*, March 2009, pp. 103–117.

[27] M. A. Rajab, J. Zarfoss, F. Monrose, and A. Terzis, "My botnet is bigger than yours (maybe, better than yours) why size estimates remain challenging," in *Proceedings of the First Conference on First Workshop on Hot Topics in Understanding Botnets*. USENIX Association, 2007, pp. 5–5.

[28] J. Yu, H. Kang, D. Park, H.-C. Bang, and D. W. Kang, "An in-depth analysis on traffic flooding attacks detection and system using data mining techniques," *J. Syst. Archit.*, vol. 59, no. 10, pp. 1005–1012, 2013.

[29] P. Scott and E. Wilkins, "Evaluating data mining procedures: techniques for generating artificial data sets," *Information and Software Technology*, vol. 41, no. 9, pp. 579 – 587, 1999.

[30] J. Heidemann and C. Papdopoulos, "Uses and challenges for network datasets," in *Cybersecurity Applications Technology Conference For Homeland Security,CATCH'09*, March 2009, pp. 73–82.

[31] L. W. Ghorbani Ali and T. Mahbod, "Network intrusion detection and prevention: Concepts and techniques," New York, LLCC, 2010.

[32] "Mcafee threat report," 2016. [Online]. Available: http://www.mcafee.com/ca/resources/reports/rp-quarterly-threats-mar-2016.pdf