Journal of Engineering and Applied Sciences 11 (3): 476-481, 2016
ISSN: 1816-949X

A Reference Dataset for ICMPv6 Flooding Attacks

[1]Omar E. Elejla, [1]Bahari Belaton, [2]Mohammed Anbar and [2]Ahmad Alnajjar
[1]School of Computer Science, Universiti Sains Malaysia, Penang, Malaysia
[2]National Advanced IPv6 Centre (NAv6), Universiti Sains Malaysia, Penang, Malaysia

Abstract: IPv 6 network intrusion detection and particularly anomaly-based approaches suffer from lack of reference datasets to be used for the comparison, deployment and performance evaluation of them. Intrusion Detection Systems (IDSs) use these datasets to evaluate their capabilities for detecting different types of attacks by differentiating between normal and abnormal traffics. A few IPv6 datasets have been proposed for this purpose. However, those datasets are either not publically shared due to privacy issues or focused on only one type of attack messages thus it can be exclusively used to test approaches of this message type. The flooding of ICMPv6 packets is a possible attack can be performed to deny the services of an IPv6 victim. ICMPv6 is one of the most important supported protocols of IPv6 where it has several fundamental functionalities such as; neighbor discovery and router discovery processes. Therefore, this study presents a diverse dataset for several ICMPv6 messages' flooding attacks to be used for the evolution of any proposed detecting solution of such attacks. This dataset contains 14 different ICMPv6 Flooding attacks. This research aims to assist various researchers in testing, evaluating and comparing purposes through sharing the generated datasets.

Key words: Intrusion detection, dataset generation, IPv6, ICMPv6, flooding attacks

INTRODUCTION

Eventually, IPv6 will be the main protocol for networks communication because IPv4 became incapable to serve the new demands of IPv4 addresses to users. IPv6 was designed by Internet Engineering Task Force (IETF) witha 128 bit address, more than IPv4's address space to be the sustainable communication protocol. In addition, IPv6 came up with new mechanisms to ease the communication between network nodes. For example, it defined an address auto-configuration feature which allows any new host to initiate its own IPv6 address automatically. Also in terms of security, IETF attempted to avoid some security limitations of IPv4 in the new protocol development. For example, scanning all aliveaddressesin a network by probing them one by one was possible in IPv4 (Elejla *et al*., 2014) but it became impractical in IPv6 because of the huge number of addresses (Durdagi and Buldu, 2010).

Although, much effort has been made by IPv6 developers to make it perfectly secured, there is no network that can befully secured (Elejla *et al* ., 2014). The vulnerabilities in security that have been discovered basically come from issues in IPv6 specifications such as the IPv6 multicast addresses which are the addresses where many nodes are subscribed to. An attacker can use these addresses to perform a reconnaissance attack by sending one packet to one ofthese addresses. Moreover, IETF failed to address some of the known threats in IPv4, such as the availability of performing a flooding attack. Flooding attack aims to overwhelm a targeted network or device with traffic which is more than their capacity to handle in order to deny the service of the victim (Weber, 2006).

IPv6 relies on ICMP protocol more than IPv4 and it is called ICMPv6. Therefore in IPv6, ICMPv6 is amandatory protocol for any IPv6 network due to its functionalities. It replaced not only ICMPv4 functions but also other network related protocols such as the Address Resolution Protocol (ARP) and Internet Group Management Protocol (IGMP). Therefore, any IPv6 node needs to fully implement ICMPv6 messages to be able to communicate properly. Unfortunately, ICMPv6 is considered as an insecure protocol and has low awareness of security issues. Thus, it is vulnerable to many attacks such as the Smurf attack, where the attacker sends a fabricated packet sourced from a different address (victim's address) to a multicast address. The replies from all the subscribed addresses to the multicast address will over whelm the victim and flood it with a big number of packets (Zagar and Grgic, 2006).

As aconsequence to ICMPv6 security issues, mitigation solutions should be proposed. These solutions need to be evaluated and tested using reliable datasets.

Corresponding Author: Omar E. Elejla, School of Computer Science, Universiti Sains Malaysia, Penang, Malaysia

ICMPv6 is an insecure protocol and securing it helps to secure IPv6 generally because of its vulnerabilities as well as the importance of it. Therefore, proposing reliable datasets for ICMPv6 attacks is essential in order to help in conducting an experimental security study for any new solutions. Similarly in IPv4, many datasets (e.g., KDD' (Cmcttugh, 2000) were proposed and being used to evaluate Ipv4 proposed mitigation solutions. A few IPv6 datasets have been proposed till this time. Moreover, these datasets are facing some limitations. Mainly, the absence of a reliable dataset contains different attacks against ICMPv6 protocol (Saad *et al*., 2014a, b). Therefore, this study aims to create a reliable dataset focusingonthe common attacks of ICMPv6 to be used in testing the mitigation solutions.

ICMPv6 flooding attacks: ICMPv6 is a backbone protocol in IPv6 where it is responsible for important functionalities such as neighbor discovery. IPv6 protocol does not work properly without a complete implementation of ICMPv6 messages because IPv6 network operations need ICMPv6 messages for many important processes as described in RFC 4443 (Conta *et al*., 2006). ICMPv6 has two types of messages usage. First, Error messages where the type field value ranges from 0-127 and are sent as a response for failed deliveries of messages. The second type is Informational messages which are sent to share needed information between nodes and their range of type field value is from 128-255.

These messages can be misused for several attacks such as ICMPv6 flooding attack. ICMPv6 flooding attack is performed by sending a big number of packets of ICMPv6 to one victim which can be a single node (PC or router) by its address. Also, it can targettheentire IPv6 network by sending the packets to an IPv6 multicast address. An example of the ICMPv6 packets is the Router advertisement message (ICMPv6 type 134) where it can be sent to overwhelm a node, forcing it to generate a new IPv6 address from the given prefix in the message. This results in a denial of service attack for that node because its CPU load increases to 100% in order to handle these messages. Also, these messages can be sent to a multicast address (such as all node address FF02::1) to overwhelm all subscribed nodes in that address.

ICMP flooding attack was known in IPv4 networks in the same way as in IPv6. However, ICMPv4 is less important for communication than ICMPv6 due to the dependencies of IPv6 on ICMPv6 functionalities where it is considered the most important protocol associated with IPv6 (Saad *et al*., 2014a, b). The IPv4 administrators may block most of the ICMPv4 messages on their networks to avoid its vulnerabilities. In contrast, blocking ICMPv6 message cannot be used in IPv6 because IPv6 operations need ICMPv6 messages for many basic functions such as the address resolution process. There by, ICMPv6 Flooding Attack is more critical in IPv6 than in IPv4 and it needs more studies to address it.

Literature review: The existence of datasets is useful especially for evaluating the performance of machine learning, artificial intelligence, genetic algorithm and in the statistics of mitigation solutions. IPv4 mitigation solutions have been tested using several datasets such as DARPA (McHugh *et al*., 2000), KDD and LBNL (LBNL/ICSI., 2004). These datasets are mainly used for testing the ability of solutions to differentiate between normal and abnormal traffic. Usually, researchers choose one or more of these datasets that meet the needs of their solution to be used. Despite the significant contributions of these datasets, they are limited to IPv4 solutions testing where they do not contain IPv6 traffics. This encouraged IPv6 researchers to overcome this limitation bycreating datasets with IPv6 traffic to be used for testing its mitigation solutions.

A few datasets for IPv6 traffic have been proposed. Some of these data sets were created to study IPv6 protocolfor non-security purposes where they do not contain malicious traffic. MAWI Working Group Traffic Archive dataset (WIDE, 2015) is one of these datasets from the WIDE project which contains daily traces of normal IPv6 traffic. It has been used for the security analysis of IPv6 by visualizing the traffic in (Barrera and Van Oorschot, 2009). This kind of datasets is limited to non-mitigation aims of IPv6 researches. Therefore, it cannot be used for testing security solutions.

Recently, a dataset called Ark is released by the Center for Applied Internet Data Analysis (CAIDA) for the purpose of Intrusion Detection Systems (IDSs) testing. Although, it meets a lot of researchers' needs, it has a limitation of having specific traces of traffic and it is not a comprehensive dataset of different kinds ICMPv6 flooding attacks. Moreover, a few features of the traffic are privatized which leads to a limitation indetection efficiency. For example, the source and destination addresses which are the main keys in any detection approach are removed from this dataset. Therefore, it is not suitable for AI-based solutions which are efficient in intrusion detection.

Another recent dataset has been proposed by Saad *et al*. (2014b) which focuses on ICMPv6 Flooding Attack. This dataset has two main limitations. Firstly, it focuses only on ICMPv6 ECHO request (ICMPv6 type 128) flooding attack. Therefore, it can only be used to evaluate solutions targeting this kind of attacks. The

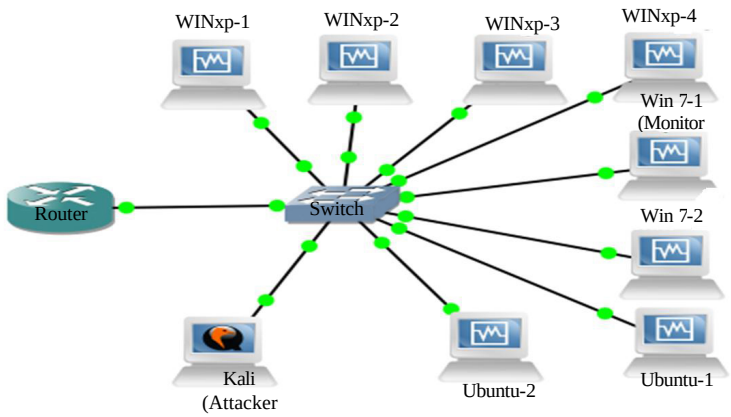

Fig.1: Testbed network architecture

Table 1: Testbed devices name, physical address, and IPV6 addresses

Node name	Physical address	Global address	Link-local address
Router	C4-01-12-C8-00-00	2000::C601:12FF:FEC8:0	Fe80::C601:12FF:FEC8:0
WINxp-1	02-00-00-00-00-01	2000::FF:FE00:1	Fe80::FF:FE00:1
WINxp-2	02-00-00-00-00-02	2000::FF:FE00:2	Fe80::FF:FE00:2
WINxp-3	02-00-00-00-00-03	2000::FF:FE00:3	Fe80::FF:FE00:3
WINxp-4	02-00-00-00-00-04	2000::FF:FE00:4	Fe80::FF:FE00:4
Win7-1 (Monitor)	02-00-00-00-00-05	2000::FF:FE00:5	Fe80::FF:FE00:5
Win7-2	02-00-00-00-00-06	2000::FF:FE00:6	Fe80::FF:FE00:6
Ubuntu-1	02-00-00-00-00-07	2000::FF:FE00:7	Fe80::FF:FE00:7
Ubuntu-2	02-00-00-00-00-08	2000::FF:FE00:8	Fe80::FF:FE00:8
Kali (Attacker)	B8-D0-00-00-00-00	2000::BAD0:FF:FE00:0	Fe80::BAD0:FF:FE00:0

second limitation is that this dataset is not available for other researchers because it is not shared online for the privacy of the reachers network. Najjar and Kadhum (2015) have released the most recent dataset which have the advantage that labelsthe traffic as normal and abnormal traffic. This labeling gives the opportunity for supervised learning AI techniques to be applied on it. However, it is not comprehensive where it has only two types of ICMPv6 flooding attacks. In addition, it is alsonot publically published to be used by others.

MATERIALS AND METHODS

The proposed synthetic dataset

Topology: The proposeddatasetwas created using a virtual network of Graphical Network Simulator 3 (GNS3) (Grossman *et al*., 2013) containing 9 interconnected PCs, Cisco router and switch. The GNS3 is an open source emulator of computer networks by connecting real and virtual devices in order to allow the design of different network topologies. In addition, it uses Dynamips emulator to run real Cisco IOS images (Neumann, 2014). Figure 1 shows the network topology that was used to generate this dataset. Different OS images (such as Ubuntu, Kali and Windows XP, Windows 7) have been installed on Oracle virtual boxes toeffectively construct a real and diverse traffic. The majority of these OSswas Windows because of its wide use in today's life where the desktop market share of Windows is >90% of the global market.

Router interface has been configured with the EUI-64 addressing method to generate IPv6 addresses of nodes. Physical addresses of the stations have been modified to make the process of analyzing, tracking and labeling of the dataset traffic easier for us. For example, Kali station's physical address has been changed to contain the string "BAD" in its IPv6 address which makes sense because it is used as an attacker node. Table 1 shows the details of nodes' physical addresses and the generated IPv6 addresses for each node in the topology.

Traffic generation and capturing: Usually, the datasets that are used to evaluate IDSs have two types of traffic classes: normal (without any attacks performed) and abnormal (with attacks performed). Therefore, capturing the traffic has two stages; First, pure normal traffic iscaptured from the network using Wireshark installed in the monitor station. Wireshark Software (Combs, 2008) is a packet sniffer and analyzer used to capture the traffic from the network. The 48 h (from 14, Nov. 2015 at 1 pm to 16, Nov. 2015 at 2 pm) of normal traffic was captured when using it without any malicious activity in the network.

Table 2: Performed attacks and their details

Name	THC command	ICMPV6 Flooding attack Packets	Target
Attack 1	Denai l6 eth0 FE80::FF:FE00:1 1	Echo request	WINxp-1
Attack 2	Denai l6 eth0 FF02::1 1	Echo request	All Nodes (FF02::1)
Attack 3	Flood_dvertise6 eth0 FE80::FF:FE00:1	Neighbor advertisement	WINxp-1
Attack 4	Flood_advertise6 eth0	Neighbor advertisement	All nodes (FF02::1)
Attack 5	Flood_solicitate6 eth0 FE80::FF:FE00:1	Neighbor solicitation	All nodes (FF02::1)
Attack 6	Flood_solicitate6 eth0 FF02::1	Neighbor solicitation	All nodes (FF02::1)
Attack 7	Flood_router26 eth0 FE80::FF:FE00:1	Router advertisement	WINxp-1
Attack 8	Flood_router26 eth0	Router advertisement	All nodes (FF02::1)
Attack 9	Flood_mld6 eth0 FE80::FF:FE00:1	MLD report	WINxp-1
Attack 10	Flood_mld6 eth0	MLD report	All routers (FF02::2)
Attack 11	Flood_mld26 eth0 FE80::FF:FE00:1	MLDv2 report	WINxp-1
Attack 12	Flood_mld26 eth0	MLDv2 report	All MLDv6 capable Routers (FF02::16)
Attack 13	Flood_mldrouter6 eth0 FE80::FF:FE00:1	MLD router advertisement	WINxp-1
Attack 14	Flood_mldrouter6 eth0	MLD router advertisement	All snoopers (FF02::6a)

This capturing started within the starting of the network to ensure that all the ICMPv6 initial messages between nodes are captured.

THC-toolkit (Heuse, 2013) is a complete tool for IPV6 attacks and has beenused to generate the malicious traffic of ICMPv6 flooding attacks in the network. THC-toolkit is a command line tool already installed and built-in with Kali OSs. There are many ICMPv6 flooding attacks in the THC-toolkit that are performed in this dataset from the Kali station. Since this dataset is intended for network security and intrusion detection purposes, it would not be complete without performing a set of attack scenarios.

Therefore, 14 different ICMPv6 flooding attacks were performed separately in the network using THC-toolkit. All the performed attacks were captured and logged using Wireshark from the monitor device (Win 7-1). One THC-toolkit command is executed in each attack scenario for a small period of time, targeting its victim. The attacks scenario was divided into two scenarios according to their victim; the first scenario was targetingan IPv6 address (WINxp-1) and the second one targeting more than one IPv6 (such as All Nodes FF02::1). While the attack is in action, traffic was being captured in a separated file to be labeled after that. Table 2 shows the performed attacks, the THC's commands used and the targeted victims of each attack.

RESULTS AND DISCUSSION

Network interaction information and anomalous traffic are important for post-evaluation and the correct interpretation of results. Therefore, it is an essential and a major requirement for the dataset to have this information. As mentioned before, the attacks are performed from aknown IPv6 address that helps to validate the datasets by knowing the source of the malicious traffic. Therefore, each row of packet is labeled according to the source IPv6 address in addition to its times. As a result from these experiments, 14 stamped rows of packet capture (PCAP) files containing malicious traffic were created. Each file contains different attack traffic and itis named according to the attack's nameas shown in Table 1. Also, the 48 h of normal traffic was captured in a separated file which contains normal packets between nodes. In total, there are 15 Packet Capture (PCAP) files that were created for this dataset (available at:http://omarusm.aalnajjar.com).

Data preprocessing and labeling: Having the full payload of traffic is beneficial for any detection system thatrelies on analyzing payloads (deep packet inspection). Moreover, it helps researchers to investigate attacks behaviors for a deeper understanding of them. Therefore, the full PCAP files of the normal traffic and the abnormal traffic are available in our dataset. Moreover, a CSV file is used to record all packets headers of the traffics (normal and abnormal) and label them becauselabelinga dataset is ofimmense importance in the evaluation of the detection mechanisms and perfect training of IDSs. Figure 2 shows a snapshot from the CSV file of the proposed dataset.

To create a CSV file from the PCAP files, then umbers of features have to be chosen and extracted from the available information of the packets. The most relevant features for the performed ICMPv6 flooding attacks were chosen in ourdataset using the Wireshark filtering scheme. In total, the CSV files containthe extracted features such as IPv6 address, physical addresses and port numbers. In addition, it has a class feature to label each traffic record according to the attacks' names shown in Table 2. This labeling helps to simplify the evaluation of the Intrusion Detection Systems and provides realistic and comprehensiveness to the dataset.

Comparison with other datasets: A few IPv6 datasets have been already created for the research community. These datasets achieved the satisfaction of some researchers, especially those who created it. In contrast, many of them, if not all, failed to satisfy others' objectives. For example, the CAIDA dataset is

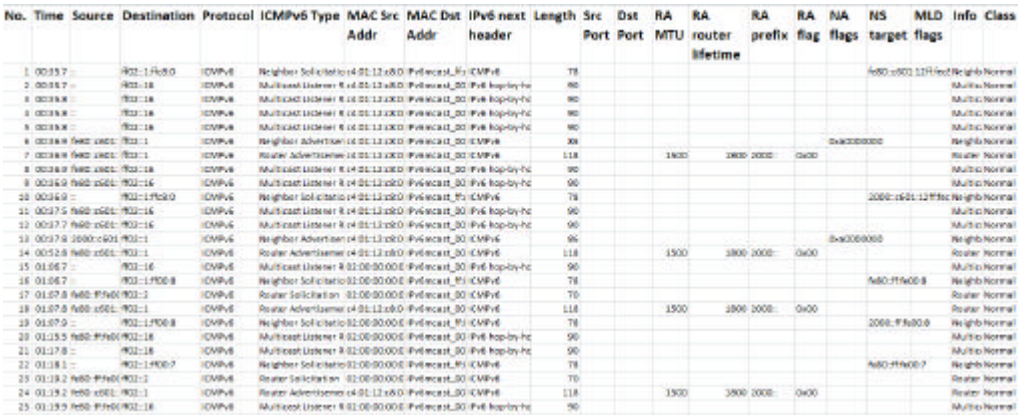

Fig. 2: Snapshot from the labeled traffic

Table 3:Comparison between the proposed dataset andvarious datasets

Dataset name	Network configuration Information availability	Labeled dataset	Complete capture	Diverse attack scenarios	Onlineavailability
MAWI	NO	NO	YES	NO[1]	YES
CAIDA	NO	NO	NO[2]	NO	YES
Saad *et al.*	YES	NO	YES	NO[3]	NO
Najjar and Kadhum	YES	YES	YES	NO[4]	NO
Our Dataset	YES	YES	YES	YES	YES

[1]Does not have malicious traffic; [2]Important information has been removed such as IP addresses; [3]Has only one kind of flooding attack (ECHO request); [4]Has two flooding attacks

availableonline for other researchers however, its traces are modified for its author's privacy issues and thus, their protocol information and sources and destinations' IPv6 addresses are completely removed. This limits others who are interested in such information from using the dataset. Table 3 summarizes a qualitative comparison between the aforementioned IPv6 datasets and also our dataset.

This research has tried to overcome issues that have been explored on the other datasets as shown in Table 3. Most of the available datasets are unlabeled yet. Labeling is clearly important to be used in the evolution of different intrusion detection approaches such as supervised IDSs. These datasets cannot be used for these IDSs because they are not labeled while labeling is the main need for this kind of IDSs. Another issue of the existing datasets is that they do not contain various attacks' scenario such as Najjar and Kadhum's dataset contains only two flooding attacks scenarios. Thus, it can exclusively be used to evaluate approaches which are targeting that kind of attacks only where most of the intrusion detection approaches are trying to detect many kinds of attacks which these datasets are not fit for.

CONCLUSIONS

IPv4 had different datasets that were used for different purposes such as the evaluation of security approaches. Similarly, IPv6 needs to have such datasets to be used for the same purposes. By reviewing the literature, we found that there are a few datasets that have been created for IPv6 traffic. These datasets have some limitations such as being specific for only one kind of testing purpose and also not being publicly available for others to use in more researches. Thus, we saw that there is a need for new datasets that are focusing on IPv6 traffic to help in its improvements.

ICMPv6 is a basic protocol in any IPv6 network which is vulnerable to several attacks. One of these attacks is the flooding attack which sends a huge number of its packets to a victim (device or network). By exploring the available IPv6 dataset it was clear that there is no a dataset containing these kinds of attacks to be used for the evaluation of their detection approaches. That led this work forward into creating a diverse dataset containing normal and abnormal traffic to be a reference for evaluating detection approaches of ICMPv6 flooding attacks. This dataset contains 14 different abnormal traffics of ICMPv6 flooding attacks and 48 h of normal traffic including network initialization messages.

The presence of abnormal traffic in the dataset would improve the detection of unknown attacks such as "Zero Day" attacks. Therefore, the generated dataset is created to be a reference fore valuating the solutions that aim to detect ICMPv6 flooding attacks. In addition, the most relevant features of these attacks have been extracted

from the trafficsand labeled as normal or attack traffic. This preprocessing stage aims toprepare the traffic as input for the approaches that depend on the features, such as artificial IDSs. Thus, this dataset can help evaluate the detection accuracy of the securing mechanisms of IPv6 protocol against ICMPv6 flooding attacks.

The proposed dataset is limited to flooding attacks based on ICMPv6 messages. Therefore in the future work, we plan to improve its diversity and make it more comprehensive by including more IPv6 attacks within its traffic. The IPv6 protocol is vulnerable to different attacks other than flooding attacks. Hence, these attacks will be added in the dataset the future. Moreover, we intend to update and increase the number of machines that are performing the abnormal activity as well as raise the relative portion of malicious traffic. Additionally, the number of machines that are performing normal activities will be increased. Adding these activities aims to add more reality to the dataset traffic and it is necessary for labeling and validating the dataset.

ACKNOWLEDGMENT

The reachers would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the study and to thank the Computer Science School in Universiti Sains Malaysia (USM) for providing the facilities and support to setup the dataset. This research is partially supported by USM Global Fellowship.

REFERENCES

Barrera, D. and P.C. van Oorschot, 2009. Security visualization tools and IPv6 addresses. Proceedings of the 6th International Workshop on Visualization for Cyber Security, October 11, 2009, Atlantic City, NJ., USA., pp: 21-26.

Combs, G., 2008. Wireshark 0.99.5 release notes. https://www.wireshark.org/docs/relnotes/wireshark-0.99.5.html.

Conta, A., S. Deering and M. Gupta, 2006. Internet Control Message Protocol (ICMPV6) for the Internet Protocol Version 6 (IPV6) specification. http://tools.ietf.org/pdf/rfc4443.pdf.

Durdagi, E. and A. Buldu, 2010. IPV4/IPV6 security and threat comparisons. Procedia-Social Behav. Sci., 2: 5285-5291.

Elejla, O.E., A.B. Jantan and A.A. Ahmed, 2014. Three layers approach for network scanning detection. J. Theoret. Applied Inform. Technol., 70: 251-264.

Grossman, J., B. Marsili, C. Goudjil and A. Eromenko, 2013. GNS3 graphical network simulator. https://apps.ubuntu.com/cat/applications/precise/gns3/.

Heuse, M., 2013. THC IPv6 attack tool-kit. https://www.aldeid.com/wiki/THC-IPv6-Attack-Toolkit.

LBNL/ICSI., 2004. LBNL/ICSI enterprise tracing project. http://www.icir.org/enterprise-tracing/.

McHugh, J., 2000. Testing intrusion detection systems: A critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln laboratory. ACM Trans. Inform. Syst. Secur., 3: 262-294.

Najjar, F. and M.M. Kadhum, 2015. Reliable behavioral dataset for IPv6 neighbor discovery protocol investigation. Proceedings of the 5th International Conference on IT Convergence and Security, August 24-27, 2015, Kuala Lumpur, Malaysia, pp: 1-5.

Neumann, J.C., 2014. The Book of GNS3: Build Virtual Network Labs Using Cisco, Juniper and More. No Starch Press, San Francisco, CA., USA., ISBN-13: 9781593275549, Pages: 272.

Saad, R.M.A., A. Almomani, A. Altaher, B.B. Gupta and. Manickam, 2014. ICMPv6 flood attack detection using DENFIS algorithms. Indian J. Sci. Technol., 7: 168-173.

Saad, R.M.A., S. Manickam, E. Alomari, M. Anbar and P. Singh, 2014. Design and deployment of testbed based on ICMPv6 flooding attack. J. Theoret. Applied Inform. Technol., 64: 795-801.

University of California, 2011. KDD cup 1999 data. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

WIDE., 2015. MAWI working group traffic archive. http://mawi.wide.ad.jp/mawi/.

Weber, J., 2013. IPv6 security test laboratory. Master Thesis, Ruhr-University Bochum, Germany.

Zagar, D. and K. Grgic, 2006. IPv6 security threats and possible solutions. Proceedings of the World Automation Congress, July 24-26, 2006, Budapest, Hungary, pp: 1-7.