IPv6 DoS Attacks Detection Using Machine Learning Enhanced IDS in SDN/NFV Environment

Chia-Wei Tseng, Li-Fan Wu, Shih-Chun Hsu, Sheng-Wang Yu

Broadband Networks Laboratory, Telecommunication Laboratories, Chunghwa Telecom Co., Ltd., Taoyuan, Taiwan

cwtseng@g.ncu.edu.tw, jacksonwu@cht.com.tw, yjchui@cht.com.tw, sjsheu@cht.com.tw, swyu@cht.com.tw

***Abstract*—The rapid growth of IPv6 traffic makes security issues become more important. This paper proposes an IPv6 network security system that integrates signature-based Intrusion Detection Systems (IDS) and machine learning classification technologies to improve the accuracy of IPv6 denial-of-service (DoS) attacks detection. In addition, this paper has also enhanced IPv6 network security defense capabilities through software-defined networking (SDN) and network function virtualization (NFV) technologies. The experimental results prove that the detection and defense mechanisms proposed in this paper can effectively strengthen IPv6 network security.**

Keywords— IPv6, Traffic Classification, Machine Learning, Decision Tree, IDS

I. INTRODUCTION

The rapid development of mobile networks has accelerated the popularity of IPv6. According to Google's statistics [1], the proportion of global Internet users using IPv6 is approximately 32.16%, which is an increase of about 5.8 % points over the same period last year. The rapid growth of IPv6 traffic makes security issues become more important [2]. At present, most devices for detecting and preventing abnormal behaviors of IPv6 networks are firewalls or IDS. The firewalls majoring to perform access control for known IPv6 service types cannot effectively determine the abnormal packets from the passing traffic. IDS can effectively detect the intrusion of hackers and can provide appropriate compensation control measures and recommendations. It is an indispensable key component of IPv6 security protection. However, various new matching algorithms have been constantly proposed and applied to IDS, and the intrusion detection technology still has room for improvement. More advanced detection technology is needed to be able to handle increasingly complex IPv6 network environments. Therefore, the integration of IT technology based on huge data analysis or artificial intelligence technology based security solutions has gained more attention [3]. The machine learning is an algorithm that may find computers suitable for prediction or mathematical model classification [4]. This method collects a large number of raw data and standard answers to train the data adjustment and to select the corresponding mathematical model. At the same time, the results of the classification are calculated by comparing the verified data to determine whether the model is suitable for prediction or classification. In terms of defense technology, SDN and NFV are hot topics for next generation network. The concept of SDN centralized management and NFV virtualization resources play an important role in IPv6. How to combine SDN and NFV technologies to achieve network evolution and secure the IPv6 environment is the key to future network development [5][6].

This paper utilizes machine learning algorithm to enhance the IDS's ability to identify features to improve the effectiveness of detecting IPv6 DoS attacks. The paper also combined with SDN/NFV technologies to make the implementation of IPv6 security measures more effective. The rest of the paper is organized as follows: in Section II, the background and related works are addressed. Section III describes the architecture of the IPv6 malicious detection system. Section VI illustrates system test scenarios and experimental results. The last section concludes this paper and addresses potential future works.

II. BACKGROUND AND RELATED WORKS

Due to the increasing popularity IPv6, the security issues have received more and more attentions [7]. Network attacks that exist in both IPv4 and IPv6 include reconnaissance, middle-in-the-middle (MITM), and blocking attacks [8]. Reconnaissance is a relatively common attack method. Ping sweep and port scan are two general reconnaissance methods. The MITM attack hijacks communication between two devices and makes them believe that they are connected to each other directly. In the MITM scenario, the attacker can alter packets in transit or drops them or injects new packets. Blocking attacks occur when some malicious event attempts to make a resource unavailable. IPv6-specific attacks are mainly based on IPv6 next header and ICMPv6 attacks [9]. The next header field in the IPv6 indicates the type of the next extended header, but the next header may also be used by hackers to maliciously manipulate the header to attack the IPv6 network. IPv6 next header attacks include RH0, fragmentation, Hop-by-Hop option header and destination option header padding attacks, etc. ICMPv6 is used by IPv6 nodes to perform internet-layer functions. Neighbor Discovery Protocol (NDP) is the key to ICMPv6 and a method commonly used by hackers to attack [10]. NDP depending on the ICMPv6 message is easy to manipulate by the attackers due to its protocol vulnerability. These ICMPv6 packets are achieved using Multicast. Forgery and misuse will let hackers know the information of the LAN nodes, and even further information such as DHCPv6 hosts and routers in the multicast group. This increases the risk of fraud in IPv6 networks. There are already many attack tools that

support IPv6, such as SI6 [11] and THC [12], which can be used to attack the weaknesses of the IPv6 related protocols.

IPv6 address types are relatively complex, which also makes traditional network security protection technologies face severe security challenges. For this reason, the emergence of SDN and NFV technologies has brought new development opportunities for IPv6 network security defense. The integration with IPv6 makes it easier to expand and build the network, and also simplifies the infrastructure of end-to-end communication technology. [13] discusses the main architectures of SDN and illustrate how IPv6 can be deployed and integrated in SDN technologies using OpenFlow mechanisms. [14] reviews different detection techniques that are available to prevent DDoS attacks, characteristics of these techniques and issues that may arise using SDN and NFV techniques. To strengthen the detection ability of abnormal traffic, the machine learning method is used to label the samples of the data set to construct a model of attack behavior as the basis for identification [15]. Decision Tree (DT) is a commonly used machine learning method for establishing classification systems based on multiple features. The model results are presented in a tree shape, which is easy to understand and highly interpretable. The model also has a mechanism for variable selection and missing value filling and can handle classification and regression related problems. [16] addresses the design pattern of 5G micro operator and proposes a Decision Tree Based Flow Redirection (DTBFR) mechanism to redirect the traffic flows to neighbor service nodes. The pattern allows users of different µOs to be concatenated via SDN technology and then realizes the rapid connection of network to effectively enhance the interconnectivity of networks. [17] uses decision trees for abnormal traffic detection and classification and provide appropriate countermeasures to protect wireless nodes in the network and target nodes from DDoS attacks.

In order to improve the real-time detection efficiency and identification accuracy of IPv6 abnormal traffic. This paper considers the use of a decision tree algorithm and the addition of IPv6 features to construct an IPv6 network security detection system. At the same time, the paper also uses the technical advantages provided by SDN/NFV to effectively control the connection ability of IPv6 abnormal traffic.

III. Proposed IPv6 Malicious Detection System

The design and implementation of the IPv6 network security system is described in this section.

A. System Architecture

Figure 1 is an example to illustrate the network architecture of the IPv6 security system. The architecture adopts the open source platform, and uses its abstraction and integration of virtual computing, network and storage resources to construct a prototype in the SDN/NFV enabled environment. The system combines signature-based IDS detection and machine learning classification to identify possible IPv6 attacks through traffic analysis and automatic defense measures against different types of malicious threats. As shown in Figure 1, the system collects and analyzes the IPv6 network traffic through the detection module, and defense through the protection module to reduce

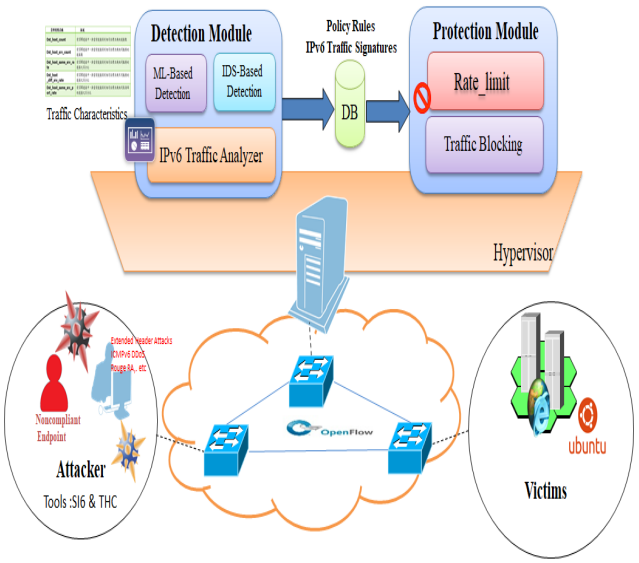

Fig. 1 IPv6 Network Security System Architecture

the impact of the possible malicious attacks. The detection and the protection module are the main functions provided by the IPv6 network security system. The detection module determines whether the IPv6 traffic flow is normal, and the characteristics of the IPv6 packet are captured by the traffic analyzer and are provided to the IDS and machine learning detection components for processing. The protection module is mainly for the treatment of IPv6 abnormal traffic and can be controlled by rate limit or blocking.

B. Machine Learning Enhanced IPv6 Intrusion Detection

The paper utilizes the decision tree method as the mechanism of traffic classification. DT is a learning method used for classification and regression in machine learning. It can provide high execution efficiency and clear classification features, which is conducive to data-based learning models. This paper focuses on the types of DoS attacks. IPv6 DoS-related features are shown in Table I. Table II shows the statistics of the number and proportion of different data categories in the IPv6 dataset recorded by the softflowd [18].

TABLE I. IPv6 Features

Feature Name	Description
IPv6_src_addr	Source IPv6 address in IPv6 header
IPv6_dst_addr	Destination IPv6 address in IPv6 header
IPv6_next_header_type	Next header type in IPv6 header
ICMPv6_type	Type of ICMPv6
durations	Duration time of flow
Count_pkt	Total number of IPv6 packets sent by sender
Src_bytes	Packet Size
diff_time_pkt	Interval between the same type of IPv6 packet sent by sender

TABLE II. IPV6 DATASETS

Type	Total	Training Set	Test Set
Normal	13,356 (51.3%)	9,349	4,007
RA Flooding	2,706 (10.4%)	1,874	832
Fragment Flooding	3,583 (13.8%)	2,504	1,079
Extension Header DoS	2,792 (10.7%)	1,971	821
TYP SYN Flooding	3,603 (13.8%)	2,530	1073

The DT algorithm used in the system is the classification and regression Tree (CART) algorithm [19]. The algorithm input is the training set *D*, the threshold of the Gini coefficient, and the threshold of the number of samples. The output is a decision tree *T*. The CART classifies the data from the root node at the top level down to two nodes, and the data is determined at each node and proceeds to the next branch, which continues to the final classification of the data. The operation procedure of the CART algorithm is described as follows:

(1) For the current node's data set as *D*, if the number of samples is less than the threshold or has no features, the decision subtree is returned, and the current node stops recursion.
(2) Calculate the Gini coefficient of the sample set *D*. If the Gini coefficient is less than the threshold, the decision tree subtree is returned, and the current node stops recursion.
(3) Calculate the Gini coefficient of each feature value of the current node for each feature on the data set *D*.
(4) Among the calculated Gini coefficients of each feature against the data set *D*, the feature A with the smallest Gini coefficient and the corresponding feature value a are selected. According to this optimal feature and feature value, the data set is divided into two parts D_1 and D_2, and the left and right nodes of the current node are established at the same time. The data set *D* of the node is D_1, and the data set *D* of the right node is D_2.
(5) Repeat steps 1-4 for the left and right child nodes to generate the entire decision tree.

For a given sample *D*, assuming there are *C* categories, P_i is the proportion of a single classification to the overall classification set, then the Gini coefficient of sample *D* is expressed as equation (1).

$$Gini\ (D) = 1 - \sum_{c=1}^{C}(P_i)^2 \quad (1)$$

For sample *D*, if *D* is divided into two parts D_1 and D_2 according to a certain value a of feature *A*, then under the condition of feature A, the Gini coefficient of *D* is expressed as equation (2).

$$Gini\ (D,A) = \frac{|D_1|}{|D|} Gini\ (D_1) + \frac{|D_2|}{|D|} Gini\ (D_2) \quad (2)$$

The model trained using the CART algorithm is very intuitive for IPv6 attack traffic. As long as the attack characteristics are completely consistent and defined in the classification, they can be accurately classified in the attack traffic. If the features are inconsistent, other smaller features can be searched through the calculated Gini coefficient to achieve the best data classification.

IV. EXPERIMENTS

This section mainly describes the results of the IPv6 abnormal traffic classification test, and also verifies the prototype function of the IPv6 attack traffic detection and defense system. Table III summarizes the hardware and software for conducting the experiments.

TABLE III. HARDWARE/SOFTWARE SPECIFICATIONS

Hardware / Software	Specification / Version	
Manufacturer	DELL INC. PowerEdge R230	
Hardware Spec	CPU	Intel Xeon Processor E3-1200
	Memory	DDR-4 32G
	Interface	1000 Base-T
Software Spec	Operating System	Ubuntu Linux 16.04
	Software List	OVS 2.11.1
		RYU 4.3
		MariaDB 10.0.38
		Apache 2.4
		Suricata 4.1.3
		Keras 2.2.5
		IPv6 THC 3.7

A. IPv6 Anomalous Traffic Classification

The system combines signature-based IDS detection and machine learning. This paper utilizes DT mechanism to classify IPv6 DoS traffic types, thereby strengthening the deficiency of signature-based IDS recognition capabilities. The DT algorithm will set multiple decision points. Each decision point will perform a binary classification of the training data, and after each binary classification, the data impureness of the classification is calculated. Each decision rule in the classification method provides a threshold from the specified feature to make a split on IPv6 traffic into the greater subset and less subset. Table IV shows the classification results of Gini DT, which has an accuracy rate of 99.3%. Only a few of these normal traffic are predicted as TCP SYN flooding.

TABLE IV. CONFUSION MATRIX FOR DT CLASSIFICATION RESULT

Prediction Truth	Normal	RA Flooding	Fragment Flooding	Ext. Header DoS	TYP SYN Flooding
Normal	3964	6	0	0	8
RA Flooding	6	806	0	0	0
Fragment Flooding	0	0	5	0	0
Ext. Header DoS	0	0	0	837	0
TYP SYN Flooding	35	0	0	0	1073

B. IPv6 Network Security System Function Test

This paper focuses on the detection and protection mechanisms of IPv6 malicious traffic. As shown in Figure 2, this scenario takes IPv6 Telnet attack as an example to test IDS IPv6 detection function and OpenFlow blocking function. The second scenario is that the system uses a decision tree mechanism to classify unknown packets or suspicious IPv6 traffic flows. As shown in Figure 3, this experiment takes the IPv6 DoS attack as an example and uses the speedtest tool on the website to detect the attacker's connection status. In this case, the attacker uses the THC tool to launch a DoS attack from the user side, and adopts the SDN strategy of reducing the transmission rate after the system decision tree classification module.

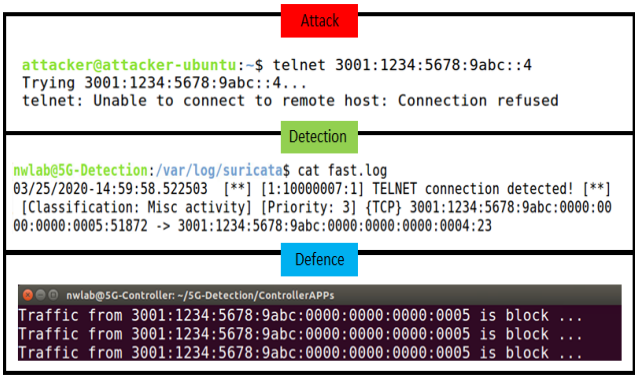

Fig. 2 Example of IDS Detection and OpenFlow Blocking Scenario

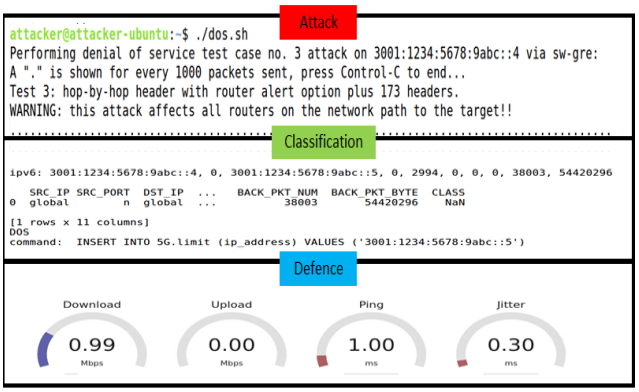

Fig. 3 Example of ML Classification and Rate Limiting Scenario

V. CONCLUSIONS

This paper is applied to detect and analyze IPv6 DoS attacks under the SDN/NFV environment and propose corresponding defense solutions for IPv6 abnormal traffic. Research results of this paper can be applied to the security monitoring of 5G edge networks in the future, and strengthen the security control of IPv6 environment. Our future work will involve more IPv6 attack types and analyses of the IPv6 network security system.

References

[1] google IPv6 statistics, [Online]. Available : https://www.google.com/intl/en/ipv6/statistics.html

[2] L. Almeida, P. Carvalho, C. Jacome, M. Monteiro and M. Cabral, "An In-Depth Analysis of the Last Twenty Years About IPv6 Security," 2018 IEEE 10th Latin-American Conference on Communications (LATINCOM), Guadalajara, 2018, pp. 1-6.

[3] Le, L., Shin, B., Lin, B. P., Tung, L, "Applying Big Data, Machine Learning, and SDN/NFV to 5G Traffic Clustering, Forecasting, and Management," 2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft), Montreal, QC, Canada,2018, pp. 168-176.

[4] Baldi, P., Brunak, S.: Bioinformatics. A Machine Learning Approach. Second Edition, Cambridge, MA: MIT Press, 2002.

[5] C. Li, Q. Wu, H. Li and J. Zhou, "SDN-Ti: A General Solution Based on SDN to Attacker Traceback and Identification in IPv6 Networks," ICC 2019 - 2019 IEEE International Conference on Communications (ICC), Shanghai, China.

[6] C. Tseng, S. Chen, Y. Yang, L. Chou, C. Shieh and S. Huang, "IPv6 operations and deployment scenarios over SDN," The 16th Asia-Pacific Network Operations and Management Symposium, Hsinchu, 2014, pp. 1-6.

[7] A. M. Taib, N. A. Othman, R. S. Hamid and I. H. A. Halim, "A Learning Kit on IPv6 Deployment and its Security Challenges for Neophytes," 21st International Conference on Advanced Communication Technology (ICACT), PyeongChang Kwangwoon_Do, Korea (South), 2019, pp. 419-424.

[8] Carlos Caicedo, J. Bejar, "IPv6 Security Analysis," Technical Report 2014, [Online]. Available : https://www.researchgate.net/publication/263444409

[9] E. Vyncke, K. Chittimaneni, M. Kaeo and E. Rey, "Operational Security Considerations for IPv6 Networks," draft-ietf-opsec-v6, [Online]. Available : https://tools.ietf.org/id/draft-ietf-opsec-v6-16.html

[10] Tao Zhang and Zhilong Wang, "Research on IPv6 Neighbor Discovery Protocol (NDP) security," 2016 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu, 2016, pp. 2032-2035

[11] SI6, [Online]. Available : https://www.si6networks.com/tools/ipv6toolkit/

[12] THC, [Online]. Available : http://manpages.ubuntu.com/manpages/trusty/man8/thc-ipv6.8.html

[13] Chia-Wei Tseng, Y. Yang and L. Chou, "An IPv6-enabled Software-Defined Networking architecture," 2013 15th Asia-Pacific Network Operations and Management Symposium (APNOMS), Hiroshima, 2013, pp. 1-3.

[14] H. D. Zubaydi, M. Anbar and C. Y. Wey, "Review on Detection Techniques against DDoS Attacks on a Software-Defined Networking Controller," 2017 Palestinian International Conference on Information and Communication Technology (PICICT), Gaza City, 2017, pp. 10-16.

[15] S. Das and M. J. Nene, "A survey on types of machine learning techniques in intrusion prevention systems," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, 2017, pp. 2296-2299.

[16] C. Tseng, Y. Huang, F. Tseng, Y. Yang, C. Liu, and L. Chou, "Micro Operator Design Pattern in 5G SDN/NFV Network," Wireless Communications and Mobile Computing (WCMC), vol. 2018, July 2018.

[17] S. Lakshminarasimman, S. Ruswin and K. Sundarakantham, "Detecting DDoS attacks using decision tree algorithm," 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN), Chennai, 2017, pp. 1-6.

[18] softflowd, [Online]. Available : http://manpages.ubuntu.com/manpages/bionic/man8/softflowd.8.html

[19] Classification and Regression Tree, [Online]. Available : https://wiki.q-researchsoftware.com/wiki/Machine_Learning_-_Classification_And_Regression_Trees_(CART)