

DDoS attacks detection in the cloud using K-medoids algorithm

SeongHo Yoon, Miyoung Kang

School of Cybersecurity, Korea University, Seoul 02841, Korea

shyoon@korea.ac.kr, dasuni@korea.ac.kr

Abstract— A denial of service (DoS) attack means that users cannot use a device or access server or network resources due to malicious cyber threats by an attacker, and Distributed Denial of Service (DDoS) attack is that several zombie PCs attack in the form of DoS. DDoS attacks are increasing daily, and DDoS attacks on the cloud are also growing. In this paper, since IDS/IPS operating outside the cloud does not recognize the inside, we propose a model to detect DDoS attacks by checking network traffic using the K-medoids algorithm, one of the clustering algorithms.

Keywords— DDoS, Clustering, Cloud Computing, Security, TCP/IP

I. INTRODUCTION

A denial of service (DoS) attack means that users cannot use a device or access server or network resources due to malicious cyber threats by an attacker, and Distributed Denial of Service (DDoS) attacks is that several zombie PCs attack in the form of DoS[1].

More than 20% of companies around the world have reported one or more DDoS attacks in their infrastructure[2]. As such, it is easily exposed to DDoS attacks. There are several examples of recent DDoS attacks. Google Cloud announced a trend in the scale of DDoS attacks under the name of "increasing the index of DDoS attacks volume," and revealed in September 2017 that it was attacked at an all-time high of 2.54 Tbps[3]. Also, on February 28, 2018, GitHub.com was attacked by 1.35 Tbps in a DDoS attacks, and the server was temporarily paralyzed[4]. As another example, Amazon AWS Shield Service was attacked by up to 2.3 Tbps in three days in February 2020[5]. Like the Amazon incident, it can be seen that the cloud has become a target for DDoS.

This paper will cover how to detect DDoS attacks on the cloud. As the cloud market develops, attacks on the cloud are frequently occurring. In particular, it also makes a denial-of-service attack against cloud servers. Denial of service attacks not only resource depletion attacks but also financial attacks. Without preparation for these attacks, massive losses can occur if multiple facilities are attacked. Currently, it is not possible to detect denial of service attacks from cloud servers with security equipment. This paper proposes a detection model using the K-medoids algorithm, one of the types of clustering algorithms, and a model that detects denial of service attacks in the cloud by mounting the proposed detection model to a hypervisor in the cloud.

II. BACKGROUND

A. Cloud

Cloud computing is an on-demand model of a network that shares resources. It is a service provider that can minimize the effort for management and can provision and release[6].

Cloud computing is largely divided into IaaS, PaaS, SaaS as service model types, according to the range of computing resources provided to users[7]. First, Infrastructure as a Service (IaaS) is a service that provides users with control over key computing resources, including processing, storage, and networks, including operating systems and applications[7]. It is a computing environment that performs mainly logical and intact functions, consisting of multiple virtual machines (VMs) and containers in one physical device, and is provided with an environment isolated from the rest[7]. Second, Platform as a Service (PaaS) is a service that provides the right to develop the desired system based on the provider's technology and to distribute user-generated outputs to the cloud infrastructure using the provider-supported database, development tool, library, and operating system[7]. Users cannot control the cloud infrastructure, but do have preferences for operating applications and control over them[7]. Third, SaaS (Software as a Service) is a service that provides an environment in which applications can be used through a network from various user devices, and users cannot manage and control cloud infrastructure in which applications run, including settings for specific applications[7].

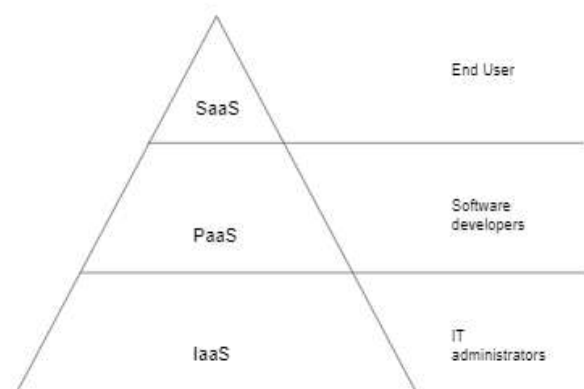


Fig1. Cloud Service Model

Virtualization is a basic technology that supports cloud computing by abstracting hardware resources, and cloud computing doesn't have without virtualization [8]. A cloud that runs on virtualization operates on a hypervisor basis. A Cloud hypervisor is software that enables hardware resources to be shared with multiple virtual machines, as shown in Figure 2 [9].

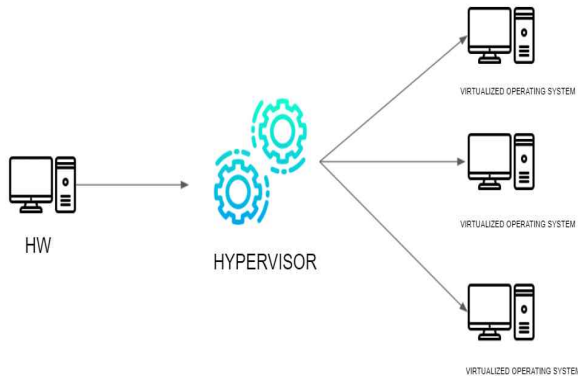


Fig2. Cloud Virtualization

B. K-medoids algorithm

The K-medoids algorithm has emerged to solve the noise or outliers of the K-means algorithm. The results of the comparison of the two algorithms showed that the time is taken for cluster head selection and the spatial complexity of cluster overlap were much better in K-medoids than in K-means. The data set results showed that all data sets such as execution time, insensitivity to outliers, and noise reduction were. In terms of aspects, the K-medoids was better [10]. The K-means algorithm performs calculations by taking all coordinates as centroids, whereas K-medoids take real coordinates as centroids. The K-medoids algorithm classifies groups using the similarity of data. The similarity of data is based on the distance value between data in several dimensions. The K value should be determined in advance by a non-hierarchical clustering method, and the clustering result varies depending on the K value by forming a group as much as the K value.

The procedure for performing the algorithm is shown in Fig 3 and follows the order below.

1. Among n objects, K representative objects are designated. ($n > k$)
2. After designating K-medoids, assign the remaining objects to the medoids with the highest similarity. Here, similarity means a distance measure.
3. Any designated objects other than medoids.
4. Calculate the total cost between the original medoids and any medoids. Here, the total cost is the sum of the values of the distance measurement.
5. If the total cost is negative, replace any medoids with existing medoids.
6. Repeat the above process until there is no change.

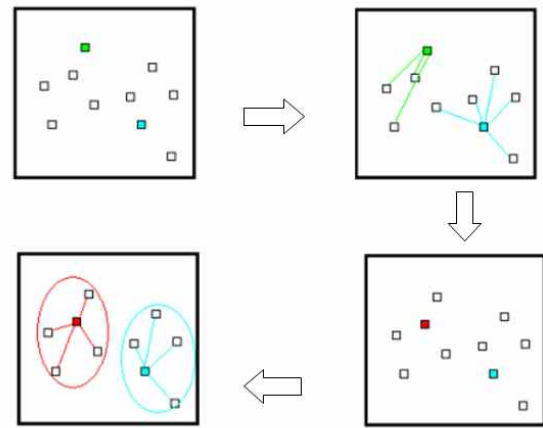


Fig 3. procedure for performing the K-medoids algorithm

III. DDoS DETECTION MODEL

The axis setting is the most important thing in detecting DDoS attacks using the K-medoids algorithm. The classification of DDoS detection items includes the following items [11].

- Average of Packets per flow
- Average of Bytes per flow
- Average of Duration per flow
- Percentage of Pair flows
- Growth of Single flows
- Growth of Different Ports

Based on the above classification, applicable new parameters were derived. As a result, items necessary for DDoS detection were extracted from the K-medoids algorithm as follows.

- Different Source IP and Same Destination IP Pair Count
- Different Source Port and Same Destination Port Pair Count
- Total Data Size
- Total Packet Count

Due to the nature of the DDoS attacks, when an attacker attacks, it continuously changes the Source IP through several zombie PCs and implements the attack. In addition, because it generates enormous traffic, the total packet volume increases, and the total amount of data transmitted also increases. Using these characteristics, different Source IP and Same Destination IP Pair Counts, Total Data Size, and Total Packet Counts are placed on each axis to detect DDoS attacks.

The results output by substituting different Source IP and Same Destination IP Pair Count, Total Data Size, and Total Packet Count into each axis are shown in Fig 4, Fig 5, and Fig 6.

Fig 4 results from outputting the pair of source and destination IPs per second as parameters on the X-axis, the sum

of packets transmitted per second on the Y-axis, and the Count of packets transmitted per second on the Z-axis.

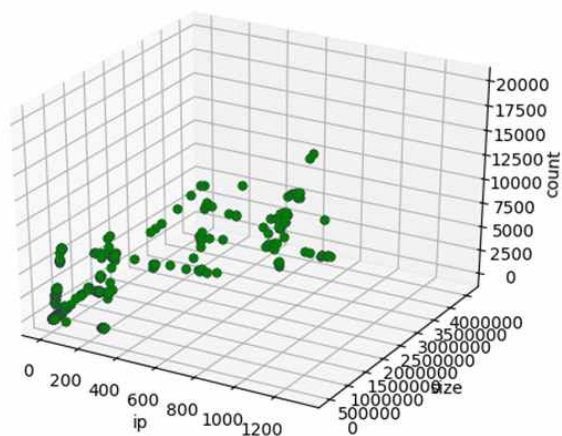


Fig 4. Set IP pair, data size, packet count on the X, Y, and Z axes, respectively

Fig 5 is the result of outputting the Count of packets transmitted per second as parameters on the X-axis, the sum of packets transmitted per second on the Y-axis, and the pair of source and destination IPs per second on the Z-axis.

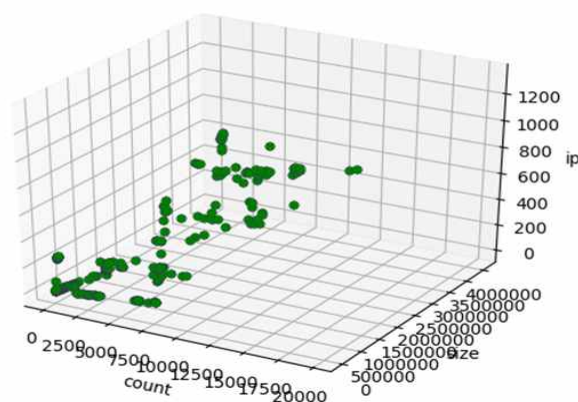


Fig 5. Set packet count, data size, IP pair on the X, Y, and Z axes, respectively

Fig 6 is the result of outputting the sum of packets transmitted per second as parameters on the X-axis, the Count of packets transmitted per second on the Y-axis, and the pair of source and destination IPs per second on the Z-axis.

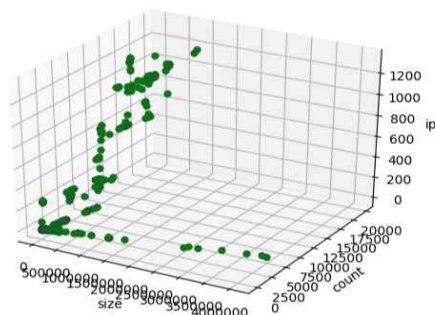


Fig 6. Set data size, packet count, IP pair on the X, Y, and Z axes, respectively

When the sum of bytes of packets transmitted per second is set as a parameter on the X-axis, the Count of packets transmitted per second is set on the Y-axis as a parameter, and the Z-axis is set as the Count of sources IP of packets transmitted per second. It is easy to distinguish between normal traffic and anomalous network traffic.

Apply it to the K-medoids algorithm over time intervals. The dataset is represented by a total of 1,001 data, of which 852 attacks and 149 attacks are included. Accuracy evaluation for this experiment measures True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), respectively, and True Positive Rate (TPR) and False Positive Rate (FPR)[12]. The measurement results are shown in the table.

TABLE I
MEASURING ALGORITHM PERFORMANCE

CONTENT	RESULT
TRUE POSITIVE	143
FALSE POSITIVE	30
TRUE NEGATIVE	822
FALSE NEGATIVE	6
TRUE POSITIVE RATE	0.959
FALSE POSITIVE RATE	0.035

IV. DDoS DETECTION MODEL IN CLOUD COMPUTING

Place the detector inside the hypervisor that monitors network packets passing through the hypervisor. The detector monitors each virtual machine's IP header and TCP/UDP packets. The overall structure of the proposed model is shown in Fig 7.

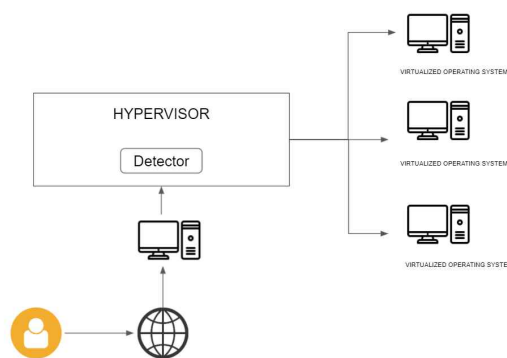


Fig 7. overall structure of DDoS detection model in cloud

Detectors in hypervisors collect and detect key data, such as the Count of packets transmitted per second, the total size of data transmitted per second, and IP pairs, then cluster using the K-medoids algorithm to detect DDoS attacks. IDS/IPS, which operates outside the cloud, cannot detect attacks because it does

not recognize the inside of the cloud, and using that detector for hypervisors will enable the detection of DDoS attacks inside the cloud.

V.CONCLUSIONS

This paper proposes a DDoS attacks detection model inside the cloud because IDS/IPS operating outside the cloud do not recognize the inside of the cloud. In addition, the accuracy of the proposed detection method was derived by simulating the actual attack. Furthermore, this model is equipped with a hypervisor in the cloud to monitor network packets and detect DDoS attacks in the cloud.

A technique for detecting abnormal traffic by clustering network traffic using the K-medoids algorithm was used as a detection technique. Lastly, in the paper, three axes were used for detection. The X-axis represents the Count of bytes transmitted per second as a parameter. The Y-axis represents the Count of packets transmitted to the parameter Z-axis per second. By setting the axis to the Count of source IPs of packets transmitted per second, each axis can be applied to detect DDoS attacks successfully.

ACKNOWLEDGMENT

Basic Science Research Program supported this research through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2019R1I1A1A0106293913)

REFERENCES

- [1] Understanding Denial-of-Service Attacks, <https://us-cert.cisa.gov/ncas/tips/ST04-015>
- [2] Kaspersky Labs, Global IT security risks survey 2014 - distributed denial of service (DDoS) attacks, 2014, (<http://media.kaspersky.com/en/B2B-International-2014-Survey-DDoS-Summary-Report.pdf>)
- [3] Exponential growth in DDoS attack volumes , <https://cloud.google.com/blog/products/identity-security/identifying-and-protecting-against-the-largest-ddos-attacks>
- [4] February 28th DDoS Incident Report, <https://github.blog/2018-03-01-ddos-incident-report/>
- [5] AWS said it mitigated a 2.3 Tbps DDoS attack, the largest ever, <https://www.zdnet.com/article/aws-said-it-mitigated-a-2-3-tbps-ddos-attack-the-largest-ever/>
- [6] NIST Cloud Computer, <https://csrc.nist.gov/Projects/Cloud-Computing>
- [7] SaaS vs PaaS vs IaaS: What's The Difference & How To Choose, <https://www.bmc.com/blogs/saas-vs-paas-vs-iaas-whats-the-difference-and-how-to-choose/>
- [8] Virtualization vs. Cloud Computing: What's the Difference?, <https://www.businessnewsdaily.com/5791-virtualization-vs-cloud-computing.html>
- [9] What is a Cloud Hypervisor, <https://www.vmware.com/topics/glossary/content/cloud-hypervisor>
- [10] Preeti Arora, Deepali, Shipra Varshney, Analysis of K-Means and K-Medoids Algorithm For Big Data, Procedia Computer Science, Volume 78, 2016, Pages 507-512, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2016.02.095>.
- [11] R. Braga, E. Mota, and A. Passito, "Lightweight DDoS flooding attack detection using NOX/OpenFlow," 2010 IEEE LCN, pp. 408-415, Denver, CO, Oct. 2015.
- [12] A simple guide to building a confusion matrix, [https://blogs.oracle.com/ai-and-datascience/post/a-simple-guide-to-build-a-confusion-matrix#:~:text=True%20Positive%20\(TP\)%20is%20an,incorrectly%20predicts%20the%20positive%20class.](https://blogs.oracle.com/ai-and-datascience/post/a-simple-guide-to-build-a-confusion-matrix#:~:text=True%20Positive%20(TP)%20is%20an,incorrectly%20predicts%20the%20positive%20class.)



Seong Ho Yoon received B.S. degree from the School of Computer Engineering, Dongguk University, Korea in 2020. He is currently an M.S. candidate in the School of Cybersecurity, Korea University. His current research interests are in Network Security and secure software engineering.



Miyoung Kang received the M.S. degree in the Department of Computer Science and Engineering at Dongguk University and the Ph.D. degree in the Department of Computer Science and Engineering at Korea University, Seoul, Korea. She is a research professor at the graduate school of Sybersecurity at Korea University, Seoul, Korea. Her research interests are Formal Methods, process algebras, software-defined networking (SDN), and security in networks.