Real-Time DDoS flood Attack Monitoring and Detection (RT-AMD) Model for Cloud Computing

Alaa Alsaeedi  
Computer Science department, computer science and engineering  
AALSAEEDI0034.stu@uj.edu.sa

Omaimah Bamasag  
Cybersecurity department, College of computer science and engineering  
obamasag@uj.edu.sa

Asmaa Munshi  
Cybersecurity department, College of computer science and engineering  
ammunshi@uj.edu.sa

ABSTRACT
In recent years, the advent of cloud computing has transformed the field of computing and information technology. It enabled customers to rent virtual instances and take advantage of various services on-demand with the lowest costs. Despite the advantages offered by cloud computing, it faces several threats; an example is DDoS attack which is considered one of the most serious ones. This paper proposes a real-time monitoring and detection of DDoS attacks on the cloud using machine learning approach. Naïve Bayes, K-Nearest Neighbor, and Random Forest machine learning classifiers have been selected to build predictive models. This model will be evaluated on the cloud for its accuracy and efficiency.

CCS CONCEPTS
- Security and privacy; • Intrusion/anomaly detection and malware mitigation; • Intrusion detection systems.

KEYWORDS
DDoS attack, IDS, Machine learning, Cloud computing

ACM Reference Format:
Alaa Alsaeedi, Omaimah Bamasag, and Asmaa Munshi. 2020. Real-Time DDoS flood Attack Monitoring and Detection (RT-AMD) Model for Cloud Computing. In The 4th International Conference on Future Networks and Distributed Systems (ICFNDS ‘20), November 26, 27, 2020, St.Petersburg, Russian Federation. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3440749.3442606

1 INTRODUCTION
Recently, cloud computing has gained much attention as it has widespread impacts across different fields, such as information technology, software engineering, data storage, and the way organizations are dealing with their IT processes. Cloud environment provides resources to customers in a virtual way with high efficiency and low cost. For example, it enables users to experiment with software before purchasing them and use storage capacity at a low cost compared to buying it in traditional ways. Cloud computing environment is a shared environment (multi-tenancy) between more than one customer where they use the same physical resources. Shared environment concept may threaten the security and availability. However, the Cloud Services Provider (CSP) must have the ability to ensure the security and availability of resources to maintain the commitment to customers, called the Service Level Agreement (SLA). There are many risks to cloud security caused by system and application vulnerabilities, such as account hijack, malicious insiders, data loss and denial and distributed denial of service. Denial of Service (DoS) is a single attack that targets the system to prevent the customers from accessing the services and resources, while Distributed Denial of Service (DDoS) is a simultaneous DoS attacks on the same target. These attacks have a substantial negative impact on the confidentiality, integrity, and availability of data. The availability of cloud services is one of the most critical CSP goals. DoS and DDoS attacks are the main threats leading to cloud services unavailability. This paper will explore machine learning for real-time detection of the DDoS attacks on the cloud environment. The research methodology will be as follows:

1. Collect DDoS attacks dataset
2. Design a mechanism that detects these attacks by monitoring the network traffic, detect if the traffic has normal behavior or not.
3. Implement and evaluate a new anomaly-based DDoS attacks detection method (RT-AMD) by using machine learning algorithms.
4. Test the effect of (RT-AMD) on cloud environment.

The contribution of this paper is to evaluates machine learning algorithms for the dataset that we have collected, investigates results with related works. Furthermore, we improve results and reach real-time attack detection by using incremental learning.

The rest of the paper is organized as follows. Section 2 describes background information on the cloud and DDoS attacks needed for the research; Section 3 presents an overview and analysis of related work; Section 4 presents proposed solution, and Section 5 presents the conclusion and future work.

2 BACKGROUND
The National Institute of Standards and Technology (NIST) defines cloud computing as “a model for enabling convenient, resource pooling, ubiquitous, on-demand access which can be easily delivered with different types of service provider interaction”[2]. Cloud environment is characterized by many features, such as manageability, scalability, availability, security, on-demand, expedient, ubiquitous, multitenant, elasticity, and stability. The services delivery in a cloud environment is categorized into three main models; Infrastructure as a Service (IaaS), Platform as a Services (PaaS), Software as a Service (SaaS), and Everything as a service (XaaS) either in public,
private, community or hybrid cloud, as defined by NIST [2]. This section presents a brief overview of the cloud computing model, DDoS flood attacks type, and security challenge and effects faced by cloud environment.

2.1 Cloud computing: Services Models
1. IaaS: is a service delivery model that allows consumers of the cloud to rent a computing infrastructure (Compute, network, and storage) resources such as "hypervisor". These services are provided on-demand, on a "pay-as-you-go", and without the need of spending a vast amount of cost, time and effort [3].
2. PaaS: is a services delivery model that allows the consumers to rent the platforms they need to build and control the applications without installing any tools. A popular PaaS provider is Google App Engine [4].
3. SaaS: is a service delivery model that allows cloud consumers to use any remote services online without the need to install the application [5].
4. XaaS: also known as Anything as a service (AaaS), is about providing anything or everything as a service. Such as Security as a Service (SecaaS), and Routing as a Service (RaaS) [6].

2.2 Cloud computing: Services Development Model
1. Public cloud: A deployment model that offers services in a shared environment over the internet at a low cost. However, many privacy and security concerns are related to the public cloud, making this deployment option not preferable for the organizations, especially business ones [7].
2. Private cloud: A deployment model that offers personal customer services either over the internet or on-premises. It enhances privacy and security but with high cost [8].
3. Community cloud: A deployment shared between multiple organizations. Community cloud shares effort and cost management between the corresponding organizations [8].
4. Hybrid cloud: A blend of two or more cloud types. It’s more secure than public cloud and lower in cost than private cloud [9].

2.3 DDoS flood Attacks Type
DoS attacks launched in a distributed manner to speed up the resources’ consumption for one or many targets called Distributed Denial of Service attacks (DDoS) [10]. DDoS flood attack is categorized into two levels based on targeted protocols either network/transport level or application level [11]. Network/transport level DDoS flood attacks aim to consume bandwidth of hosts or networking equipment by high volumes of data through using network and transport layer protocols such as TCP, UDP, ICMP, and DNS [12]. Application level DDoS flood attacks aim to consume services resources or starvation of resources to disrupt customers through establishing requests by using application layer protocols such as HTTP GET and POST requests [12]. In this study will focus on network/transport level that it’s a core layer of network architecture.

2.4 Cloud security challenges and effects
There are many security challenges that CSP faces, the main of which is the trust that must be in place between CSP and cloud customers. The trust is how the provider can protect the customer data from any breach [1]. One of the popular features of the cloud environment is multi-tenancy and virtualization. Many customers share physical resources, which is considered a big challenge in making this environment secure [9]. DoS is a cyber-attack where an attacker aims to make the systems and servers unavailable, preventing customers from accessing the servers and resources. Some of these effects, in 2009, DDoS attacks launched causing discontinued network services of most popular websites such as Live Journal, Facebook, Amazon, and Twitter [13]. In 2010 and 2011, the effects of DDoS attacks spread on large scale, among more than 75,000 computer systems in 2500 organizations and 4 million computers in 100 countries [14]. In the first quarter of 2013, the volume rate of attacks reached 48.25 gigabyte per second (Gbps). February 2018 witnessed the largest DDoS attack that targeted GitHub with volume rate 1.3 terabytes per second (Tbps.) [15]. In 2020, attacks rose by 20% according to Check Point [16].

3 RELATED WORK
Several mechanisms have been proposed to detect DDoS flooding attacks. These mechanisms have been classified based on several factors, such as the anomalies type, deployment environment, performance, level of security required, and the cost [13]. One of these classifications, which is based on identifies unusual events by monitoring the network traffic to distinguish the normal from abnormal behaviors, is called an intrusion detection system (IDS) [13]. IDS can be further classified into Signature-based, Anomaly-based, and Hybrid detection [13].

3.1 Signature-based detection
Signature-based detection is called Knowledge-based or Rule-based detections, also known as Misuse Detection. This approach is suitable for the detection of known attacks by comparing captured behavior.

Tanriverdi et al. [17] proposed a signature-based detection method to detect web-based attacks using blockchain-based web attack detection model. The signature lists in this study are automatically updated by blockchain technology. The advantage of this proposed method is that it proved that signature-based detection could be used against zero-day attacks. Bakshi et al. [18] proposed a signature-based IDS method to distinguish between the normal and abnormal traffic in VMs. It used Snort to analyze the collected traffic to determine the attack, the virtual server then drops packets coming from the specified IP address. Modi et al. [19] presented a signature-based framework to detect the known attacks and derivatives of known attacks. It also used Snort tool to detect the known attacks from network traffic. The detected attack was input to signature DB to predict derivatives attack by using signature apriority.
3.2 Anomaly-based detection

Behavior-based detections is useful to detect unknown attacks. These techniques compare the observed behavior with normal behavior to detect the abnormal events.

Kemp et al. [20] presented a solution to detect Slow HTTP attacks based on machine learning techniques. It selected eight classification algorithms as predictive models: Random Forest, decision trees, K-Nearest Neighbor, Multilayer Perceptron, RIPPER, Support Vector Machines, and Naïve Bayes. The authors used Weka machine learning toolkit to build the models. ANOVA was used to compare between the values of Slow attack detection among the eight models. They evaluated the models by Area Under the receiver operating characteristic Curve (AUC), Receiver Operating Characteristic (ROC) curve graphs, True Positive Rate (TPR) and False Positive Rate (FPR). Singh et al. Authors in [21] proposed a Multilayer Perceptron with a Genetic Algorithm (MLP-GA) based method that detected HTTP DDoS attacks on incoming traffic. The paper identified four features to detect application layer attacks. First is the number of HTTP count, referring to the count number of requests per IP address. It assumed that any single IP address sending more than 15-20 HTTP GET/POST request is considered an attack. Second is the number of the IP addresses, referring to the number of IP addresses in small windows time. Attacks are assumed with more than 20 IPs in windows time. Third was constant mapping function, the ports used by attacker is different from legitimate users as one used by attacker is varied and remain open. Fourth was fixed length, as codes with fixed frame length is considered as attack. Filho et al. [22] proposed the Smart Detection system its online approach to DoS/DDoS attack detection. The detection approach used the Random Forest Tree algorithm to classify various types of DoS/DDoS attacks such as TCP flood, UDP flood, HTTP flood, and HTTP slow. The detection rate of attacks because it around 93%. Choi et al. [23] proposed a method to integrate between detection of DDoS flood attacks and MapReduce processing in a cloud computing environment. The proposed framework consists of three parts. In the first part, the packet and log collection module (PLCM) analyses packet transmission and web server logs. Second, the pattern analysis module (PAM) produces the pattern for DDoS attacks detection. Finally, the detection module (DM) detects DDoS attacks by comparing it with a normal behavior model. The detection rate of this method is 88%. Aboorujilah and Musa [24] presented a cloud-based flood attack detection using covariance matrix approach. The proposed detection was divided into training and testing phases. A training phase was aimed to construct normal network traffic profile, and the testing phase was to detect any abnormal traffic by deviation between the normal and any other network traffic. The normal traffic is captured from end users browsing the Internet in their cloud, whereas the flooding attack traffic is generated by using PageRebooter tool. Detection performance was evaluated by using the confusion matrix and present results for internal and external cloud environment. The proposed method achieved 86.77% as detection rate with internal cloud and 84.52% in external cloud.

3.3 Hybrid-based detection

Hybrid-based detection works by combining the detection techniques mentioned above. The performance of this detection depends on the types of techniques chosen.

Hatef et al. [25] proposed a hybrid intrusion detection approach in cloud computing (HIDCC). Applied detection was a combination of signature-based and anomaly-based detection techniques. Snort tool was used for known attacks (signature-based detection) and used the Apriori algorithm to generate a derived attack pattern. Both clustering and classification algorithms are applied for the undetectable attack through snort. The clustering module receives and determined the input packet based on the sample vector. According to the found cluster, the classifier module determines the final class of the packet through algorithm C4.5 as a decision tree classifier. The proposed method was implemented in a real-time environment and achieved an accuracy of 99.38%. Saleh et al. [26] introduced a Hybrid Intrusion Detection Strategy (HIDS), applied in a real-time manner, and deployed as three stages. First, the Naïve Base feature selection (NBFS) technique was employed to reduce sample data dimensionality. Second: optimized Support Vector Machines (OSVM) was used to reject the noisy input sample as it might cause misclassification. Finally: the attacks are detected by Prioritized K-Nearest Neighbors (PKNN) classifier. The proposed method achieved an accuracy of 95.77%. This proposed scheme takes time in the first and second stages at feature selection and outlier rejection before attack detection.

As we mentioned in the previous section, the emergence of the cloud computing concept has led to an increase in security violation risk. There are few studies developed to detect anomaly of network-level DDoS flooding attacks on the cloud environment. Moreover, there are few very few studies that proposed real-time detection with a high detection rate. We will employ different classifiers that suit our needs to build detection models. We will then evaluate these models by comparing them against two main factors: detection rate and performance. Naïve Bayes, K-Nearest Neighbor, and Random Forest classifiers have been selected to build predictive models. Selected classifiers are found to have better detection results than others based on reviews of the related works [20, 22, and 26].

4 PROPOSED SOLUTION

Cloud computing depends on virtualization technology through the hypervisor by creating multiple virtual machines (VMs) corresponding to the same physical system (host system) [2]. In other words, the host system runs multiple VMs that communicate together via the virtual switch. Any VM can be affected by any attacks from outside. If this happens, it will harm the rest of VMs because all VMs are connected. The attackers will be able to consume all resources by sending malicious data to the rest of VMs, leading to servers being down. Therefore, securing VMs means securing all host resources. An overview of the proposed environment is shown in Figure 1. The proposed framework consists of two main components: Monitoring and Detection.

- **The monitoring component** is responsible for monitoring the network traffic for requests coming to the webserver on the cloud and extract the traffic from the network log. If the traffic from the log is found in Blacklist, the alert with
appropriate information is sent to the cloud admin. If it is not in Blacklist, it will move into the next component, detection.

- **The Detection component**: uses trained classifiers to detect if the incoming traffic behavior is normal or abnormal. When abnormal behavior occurs, it will alert the system, send information to the admin, then update the Blacklist with the new traffic information. Otherwise, it can access the cloud and benefit from its services.

Our proposed model will be evaluated using a dataset of abnormal normal traffics called CAIDA “DDoS Attack 2007” [27] Center for Applied Internet Data Analysis Dataset. This dataset contains around one-hour collection of anonymized (abnormal) traffic from a DDoS attack on August 4, 2007.

Wireshark is one of the most popular open-source network analyzer tools under the GNU General Public License (GPL) [28]. We will use this tool to capture normal traffics. The dataset contains information for each packet which is the number of packets in traffic, source address, destination address, protocol type (ICMP, TCP, and DNS), packet length, packet timestamp and label to determine whether the traffic is normal for “0” or attack for “1”. Our dataset had 166448 records for DDoS flood attack traffics and 327318 records for normal traffics. Training dataset will be on python, which is one of the most important and most popular programming languages. Python is the best option for analyzing data and developing algorithms as it is characterized by an interactive environment and a wide range of open-source libraries [29]. Python has rich libraries that are in the field of machine language such as the scikit-multiflow, which will be used in our research. Scikit-multiflow [30] is an incremental learning models for real-time response. Testing dataset will be on cloud environment, where we will choose an appropriate environment that suits the purpose of the study.

5 CONCLUSION

Cloud computing faces many risks, the main of which is a security violation. Information loses or servers disruption which directly affects customer’s trust. One of the most important causes of the server’s failure is DDoS flood attacks. We propose a framework for real-time monitoring of DDoS consisting of two main components: monitoring and Detection. Naive Bayes, K-Nearest Neighbor, and Random Forest algorithms have been selected to evaluate the proposed model. The contribution of this paper is to evaluates machine learning algorithms for the dataset that we have collected, investigates results with related works. Furthermore, we improve results and reach real-time attack detection by using incremental learning. Testing will be conducted on the cloud to ensure the effectiveness of the proposed framework. For future work, we aim to expand RT-AMD tool to detect application level DDoS flooding attacks.

REFERENCES

[1] Ghaffari, F., Gharaei, H., & Arabsorkhi, A. (2019, April). Cloud Security Issues Based on People, Process and Technology Model: A Survey. In 2019 5th International Conference on Web Research (ICWR) (pp. 196-202). IEEE.
[2] Zissis, D., & Lekkas, D. (2012). Addressing cloud computing security issues. Future Generation computer systems, 28(3), 583-592.
[3] Lombardi, F., & Di Pietro, R. (2011). Secure virtualization for cloud computing. Journal of network and computer applications, 34(4), 1113-1122.
[4] Sun, D., Chang, G., Sun, L., & Wang, X. (2011). Surveying and analyzing security, privacy and trust issues in cloud computing environments. Procedia Engineering, 15, 2852-2856.
[5] Fan, H., Hussain, F. K., Younas, M., & Hussain, O. K. (2015). An integrated personalization framework for SaaS-based cloud services. Future Generation Computer Systems, 53, 157-173.
[6] Ali, M., Khan, S. U., & Vasilakos, A. V. (2015). Security in cloud computing: Opportunities and challenges. Information sciences, 305, 357-383.
[7] Modi, C., Patel, D., Borisaniya, B., Patel, A., & Rajarajan, M. (2013). A survey on security issues and solutions at different layers of Cloud computing. The journal of supercomputing, 63(2), 561-592.
[8] Kapadiya, U. (2017). How About Ddos Attacks in Cloud Computing: A Survey. International Journal of Advanced Research in Computer Science, 8(3).
[9] Singh, S., Jeong, Y. S., & Park, J. H. (2016). A survey on cloud computing security: Issues, threats, and solutions. Journal of Network and Computer Applications, 75, 200-222.
[10] Douglis, J., & Ditrokota, A. (2003). DDoS attacks and defense mechanisms: classification and state-of-the-art. Computer Networks, 44(5), 643-666.
[11] Zarzar, S. T., Joshi, J., & Tipper, D. (2013). A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks. IEEE communications surveys & tutorials, 15(4), 2046-2069.
[12] Network layer: https://www.globaldots.com/blog/ddos-quick-start-guide
[13] Kaur, P., Kumar, M., & Bhandari, A. (2017). A review of detection approaches for distributed denial of service attacks. Systems Science & Control Engineering, 5(1), 301-320.
[14] Li, S. H., Kao, Y. C., Zhang, Z. C., Chuang, Y. P., & Yen, D. C. (2015). A network behavior-based botnet detection mechanism using PSO and K-means. ACM Transactions on Management Information Systems (TMIS), 6(4), 1-30.
[15] Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering, 160(1), 3-24.
[16] Cybersecurity report 2020: https://pages.checkpoint.com/cyber-security-report-2020.html
[17] Tanrıverdi, M., & Tekerek, A. (2019, November). Implementation of Blockchain Based Distributed Web Attack Detection Application. In 2019 1st International Informatics and Software Engineering Conference (UBMYK) (pp. 1-6). IEEE.

[18] Bakshi, A., & Dujodwala, Y. B. (2010, February). Securing cloud from ddos attacks using intrusion detection system in virtual machine. In 2010 Second International Conference on Communication Software and Networks (pp. 260-264). IEEE.

[19] Modi, C. N., Patel, D. R., Patel, A., & Rajarajan, M. (2012). Integrating signature apriori based network intrusion detection system (NIDS) in cloud computing. Procedia Technology, 6, 905-912.

[20] Kemp, C., Calvert, C., & Khoshgoftaar, T. (2018, July). Utilizing netflow data to detect slow read attacks. In 2018 IEEE International Conference on Information Reuse and Integration (IRI) (pp. 108-116). IEEE.

[21] Singh, K. J., & De, T. (2017). MLP-GA based algorithm to detect application layer DDoS attack. Journal of information security and applications, 36, 145-153.

[22] Lima Filho, F. S. D., Silveira, F. A., de Medeiros Brito Junior, A., Vargas-Solar, G., & Silveira, L. F. (2019). Smart detection: an online approach for DoS/DDoS attack detection using machine learning. Security and Communication Networks, 2019.

[23] Choi, J., Choi, C., Ko, B., & Kim, P. (2014). A method of DDoS attack detection using HTTP packet pattern and rule engine in cloud computing environment. Soft Computing, 18(9), 1697-1703.

[24] Aborujilah, A., & Musa, S. (2017). Cloud-based DDoS HTTP attack detection using covariance matrix approach. Journal of Computer Networks and Communications, 2017.

[25] Hafez, M. A., Shaker, V., Jabbarpour, M. R., Jung, J., & Zarrabi, H. (2018). HIDCC: A hybrid intrusion detection approach in cloud computing. Concurrency and Computation: Practice and Experience, 30(3), e4171.

[26] Saleh, A. I., Talaat, F. M., & Labib, L. M. (2019). A hybrid intrusion detection system (HIDS) based on prioritized k-nearest neighbors and optimized SVM classifiers. Artificial Intelligence Review, 51(3), 403-443.

[27] The CAIDA “DDoS Attack 2007” Dataset https://www.caida.org/data/passive/ddos-20070804_dataset.xm

[28] Munz, G., & Carle, G. (2008, April). Distributed network analysis using TOPAS and wireshark. In NOMS Workshops 2008-IEEE Network Operations and Management Symposium Workshops (pp. 161-164). IEEE.

[29] Oliphant, T. E. (2007). Python for scientific computing. Computing in Science & Engineering, 9(3), 10-20.

[30] Montiel, J., Read, J., Bifet, A., & Abdessalem, T. (2018). Scikit-multiflow: A multi-output streaming framework. The Journal of Machine Learning Research, 19(72):1–5.