ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2020.0292

DATA MINING APPROACH AND SECURITY OVER DDOS ATTACKS

M. Arvindhan[1] and Bhanu Prakash Ande[2]
[1]*School of Computing Science and Engineering, Galgotias University, India*
[2]*Department of Computer Science, Gambella University, Ethiopia*

Abstract

The benefit of on-demand services is one of the most important benefits of using cloud computing; therefore, the payment method in the cloud environment is pay per use. This feature results in a new type of DDOS attack called Economic Denial of Sustainability (EDoS), where as a result of the attack the customer pays the cloud provider extra. Similar to other DDoS attacks, EDoS attacks are divided into different groups, such as bandwidth-consuming attacks, specific target attacks, and connections-layer-exhaustion attacks. In this study, we propose a novel system for detecting different types of EDoS attacks by developing a pro le that learns from normal and abnormal behaviors and classifies them. In this sense, the extra demanding resources are allocated only to VMs that are found to be in a normal situation and thus prevent attack and resource dissemination from the cloud environment.

Keywords:

DDoS Attacks, EDoS Attacks, Cloud Computing, Machine Learning Detection

1. INTRODUCTION

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.

Cloud systems are widely exposed to various types of security threats, due to their multi-tenant nature that allows multiple virtual machines (VMs) owned by different clients to share a single physical infrastructure . While some other attacks aim to compromise confidentiality and integrity of data, DDoS attacks are a major threat to the availability of network systems and services [2]. Traditionally, DoS attack leads to unavailability of computer resources to genuine users, by flooding the victim with unwanted traffic. Authors in [1] said that the DDoS attack is similar to a DoS attack, but the impact of this attack is more destructive than the latter, because it involves many compromised and distributed systems usually known as botnets. Detection and mitigation are simply insufficient against the different methods of DDoS attacks now employed by hackers [2]. Thus in this paper, we resume recent security mechanisms against DDoS attacks in the cloud computing environment, and we compare them according to some comparison criteria.

The rest of the paper is structured as follows. Section 2 presents the problem of DDoS attack in a cloud computing environment. Section 3 is devoted to the solutions of DDoS attack problems. In section 4, we provide a comparative study of the recent security mechanisms against DDoS attacks.

1.1 DDOS THREATS IN ENVIRONMENT CLOUD COMPUTATION

The major threat to cloud-based accessibility is the distributed Denial-of-Service attack. DDoS attack is more dangerous than DoS attacks, since DoS attacks usually overwhelm a targeted network with one machine and one connection to the internet. DDoS attack, however, is an attack in which several compromised computer systems (bot) attack the target possibly causing service failure.

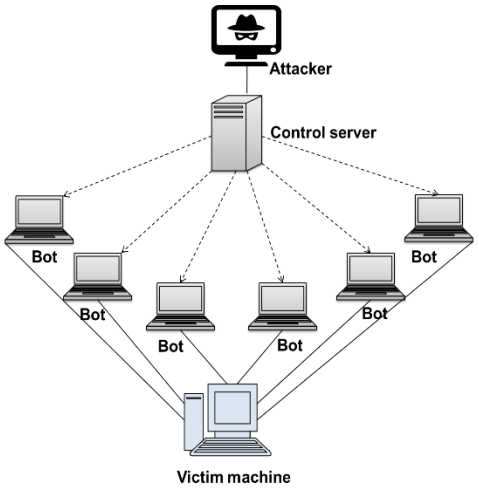

Fig.1. DDoS attack through botnet

The Fig.1 shows botnet-based DDoS attack. In DDoS attacks, some of these botnets use and try to overwhelm the opportunities provided by the cloud service. Target resources can include the CPU, memory, or even the network it operates in. As a consequence, cloud may become sluggish and legally inaccessible. When applied to the cloud background, DDoS attacks and their interpretation are entirely different [4]. These attacks were very productive where the attackers took advantage of cloud functionality (auto scaling, pay-as-you-go billing, and multi-tenant). As stated in [5], these features provide the advantage of operating on a single physical server more than VM from separate VM owners and enables a cloud user to use services without physically purchasing them. If one VM is infected with malicious software and a DDos attack on a physical host is initiated by that VM, it can cause problems for other VMs on the same host. In turn, with the same resources available on the cloud computing.

Another major DDoS attack was faced by Amazon EC2 cloud servers. These incidents of attack have resulted in heavy downtime, business losses, and many long-term and short-term effects on victim business processes.

A report by Verisign iDefense Security Intelligence Services shows that cloud and SaaS (Software as a Service) are the most

attacked target of DDoS attacks in the last quarter. More than one-third of all reported mitigations of the DDoS attack were on cloud services. One of the DDoS attack's most important piece of advice in the cloud is "economic losses". Research in predicts the total financial loss of around 444 K USD due to a DDoS attack. There are other Neustar studies that present the Q1 2015 financial loss results. The total financial loss in this study reaches 66 K USD/h. When applied to the cloud context, DDoS attacks and their interpretation are entirely different.

The disparity occurs primarily because of the results of a target server attack. Clouds as a service infrastructure (IaaS) run customer services inside. Server virtualization is the key to the cloud elastic and on-demand capabilities, where VMs are acquiring more and more capacity when needed, and when idle, and returning unused resources. The heavy organ donation trend of cloud com-putting is due to the on-demand [6].

Computing and availability capabilities of resources helps the cloud network to provide deep services by scaling as and when a VM needs. As there is ample amount of on-demand funds available in the cloud, a VM will not suffer a resource loss. This "elasticity" or "auto-scaling" feature results in a DDoS attack based on economic losses known as Economic Denial. This feature of "elasticity" or "auto-scaling" results into economic losses based DDoS attack which is known as Economic Denial of Sustainability (EDoS) attack or Fraudulent Resource Consumption (FRC) attack.

1.2 PAY AS YOU GO TO THE ACCOUNT

Because of its slimmer resource accounting and billing model, the on-demand service model has become quite appealing for consumers. "Pay-as-you-go" model allows the use of services by a cloud user without purchasing them actually. A VM owner may want more on-the-fly resources to be added or removed as required. Certain benefits of using the cloud platform include efficient use of hardware and no need for arrangements such as electricity, room, cooling and key tenancy. Pricing or accounting plays a major role when DDoS attacks in the cloud are understanding. Cloud instances are mostly paid on an hourly basis and the minimum accounting period is therefore one hour. It is possible to allocate resources on a fixed basis, pay-as-you-go basis and through auctions. Likewise, total size and total data (in and out) transfer are used to measure capacity and network band-width. It is very clear that these models are, and continue to evolve, "pay-as-you-go" models. [8]

1.3 MULTI-TENANCY

Multi-tenancy offers the advantage of operating on a single computer server more than one VMs from separate VM owners. Multi-tenancy is a way to increase the use of infrastructure and higher ROI (Return on Investment). On a single physical machine, an individual user may want more than one VM running similar or different applications.

1.4 DDOS ATTACK SCENARIO IN CLOUD

There will be many servers in an infrastructure cloud capable of running VMs in virtualized multi-tenant environments. In addition to targeting "Denial of Service," attackers may be aimed at attacking aspects of cloud consumer economic sustainability. Discussions on this attack began immediately after cloud computing was started There are several other articles that have called this attack Fraudulent Resource Conservation (FRC) attacks. Attackers implant bots and trojans extensively over the Web on infected machines and threaten cloud services with attacks on Distributed Denial of Service. When the target service is hosted in the cloud, DDoS takes the form of an EDoS attack. Exist (also known as "Booters") organizations that provide their users with a network of bots to launch DDoS threats on their competing websites. Such attacks motivations vary from economic competition, political rivalry, blackmail between countries to cyber wars.

2. DDOS ATTACKS SOLUTIONS

The DDoS attacks are mostly botnet driven attacks where a botnet controller directs a large number of automated malware driven bots to launch the attack. We show directly visible attack effects as well as attack effects which are not directly visible or become visible post-attack. Direct attack effects include service downtime, economic losses due to the downtime, auto-scaling driven resource or economic losses, business and revenue losses, and the downtime and related effects on services which are dependent on the victim service.

2.1 ATTACK STATISTICS

Most security solutions vendors in the industry are quantifying and analyzing denial of service attacks. There are a variety of other studies on the effect and growth of cloud-based DDoS/EDoS attacks. It was also expected that there will be a major target change from conventional servers to cloud-based services [8] for DDoS attackers, and the 2015 Q1 results have even proven this [11].

2.2 TAXONOMY OF DDOS ATTACKS SOLUTIONS

Because there are many forms of DDoS attack, this attack does not have a general solution. In the literature, DDoS attack has been widely studied. Some works classify DDoS attack defense mechanisms into three parts: detection of intrusion, prevention of intrusion and response to attacks (mitigation) [12] [13].

- Attack prevention (challenge response, hidden server/port, restrictive access and resource limit)
- Attack detection (anomaly detection, source/spoof trace, count based filtering, bot cloud detection, resource usage) [14].
- Attack mitigation (resource scale, victim migration, OS resource management, software defined network, DDoS mitigation as service) [15], [16].

In other studies [17] [18] the authors discussed defense mechanism of DDoS attacks based on the location where defense mechanism is deployed. Thus, the classification presented in [19] is based on two criteria.

- Based on where defense mechanism is applied (Near to source of attack, near to the destination of attack, at intermediate routers, hybrids)
- Based on when defense mechanism is applied (before attack, after attack, during attack) [20].

3. TAXONOMY OF DDOS SOLUTIONS

The cloud works related to DDoS protection were extensively surveyed and prepared as a taxonomy. We have included many of the DDoS protection works in conventional networks to assist with the particular direction of research. We prepare this taxonomy by maintaining an understanding that this work would serve the purpose of providing a clear, detailed and complete picture of the literature space solutions, different ideas, and approaches. Fields of taxonomy are given a nomenclature for the classification of different [21][22].

Many studies have been involved in preventing DDoS attacks Tools in recent times. Such approaches are designed to help a victim server continue to serve requests in the presence of attack, such as resources scaling approaches, resource management methods, resource relocation methods, network-based mitigation methods specified by software, etc. [23][24]. The most commonly used approaches for mitigating DDoS attacks in cloud computing environment are: Approaches of resource scale: this class includes all methods that aim to resolve DDoS attacks, allowing database availability with scaled resources, such as resource scaling techniques and resource management techniques.

Software defined networking based mitigation methods: Where the capabilities of SDN (e.g. software based traffic analysis, logical centralized control, dynamic updating of forwarding rules, and global view of the network) make it easy to detect and react to DDoS attacks rapidly [7][25]. There are a considerable number of works tries to benefit the maximum of SDN advantage to mitigate DDoS attack in cloud environments.

Table.1. Comparative table of detection mechanisms against DDoS attacks

Traditional methods	**Data mining and learning methods**	**Important metrics to benchmark the solutions**
Anomaly (Request count threshold) [14]	Adaptive learning of detecting model	The learning algorithm allows to adapt a detection model to, network changes. The proposed algorithm reduces the possibility to consider legitimate traffic as malicious, and minimize false positive and negative
Anomaly (Shannon entropy of source IP + Packet arrival rate) [15]	Fuzzy logic	Fuzzy system used to determine the attack statue in cloud environment, and to detect the attacks at its earlier stage. Fuzzy logic system was easy to implement in cloud computing environment.
Anomaly (covariance method) + Signature [4]	Metaheuristic algorithm (Ant Lion Optimization) + artificial neural network	The proposed method improves the accuracy of DDoS detection.
Anomaly (Request count threshold) [10]	Data mining (LS-SVM, Naive Bayes, K-nearest, Multilayer perception)	The proposed method improves the accuracy of DDoS detection.
IP trace back + Port hopping + Reputation management [6]	Threat intelligence (IT)	The proposed method able to detect unknown threats. With this method there is no false positive.
Anomaly (Request count threshold) [5]	Neural network algorithm (BP)	The proposed method improves the DDoS detection rate.

4. DATASET DESCRIPTION

In fuzzy clustering, a single particle represents a cluster center vector. In other words, each particle $part_l$ is constructed as follows:

$$part_l = (v_1, v_2, \ldots, v_i, \ldots, v_c)$$

Network congestion. Another detectable network flow is network congestion. When a hot topic arises, the number of new users and the number of old users will increase significantly. There are three characteristics: first, because of a large number of new users, N should be very positive; Secondly, because of the wide range of hot topics, the old users are likely to access it, so the N value should be higher than 1. Finally, even if there are many new users, the constant c will be a smaller value because it follows the TCP/IP protocol with the normal users.

$$NAFV = -N \times A \times F \times V << -1$$

where, F indicates the degree of interest of external users to a particular hot topic: if $R_k \in (0, 1]$ and it is a particular topic, where it can imply that older users are more concerned with the topic; otherwise, it implies that new users are more concerned about the topic [8].

ANN: ANN is efficient for large datasets and the number of hidden nodes in the network is considered as free parameter. Once a network has been structured for a particular application, that network is ready to be trained. There are two approaches to training, supervised and unsupervised. The most often used ANN is a fully connected, supervised network with backpropagation learning rule. This type of ANN is excellent at prediction and classification tasks [9].

HMM: HMM is a powerful mathematical approach, which is designed to model complex data sequences. This classification system is a special type that aims to find each state's posterior probability given a series of measurements, and predicts the state with the highest likelihood.

SVM: An SVM is an exclusionary genetic algorithm formally defined by a hyperplane separator. The nonlinear datasets can be effective. The number of support vectors differs on the problem of optimization occurring, and each supporting vector is sometimes a subset of data. It utilizes kernel processes in the nonlinear datasets to formulate the concept of linearity.

5. RESULTS AND DISCUSSIONS

In Table.2 and Fig.2, the result shows the training accuracy with respect to the number of data. Training accuracy is estimated for the classifiers which includes SVM, HMM and HMM-SVM. The result depicts that the proposed method has the highest training accuracy. In Table.3 and Fig.3, the result represents the testing accuracy of the classifiers with respect to the number of data. Testing accuracy also high for the proposed method when comparing with the other existing classifiers [3] [10].

Table.2. Comparison values of Training Dataset

Performance Metrics	HMM - SVM (%)	ANN (%)	HMM (%)	SVM (%)
Accuracy	95.3	85.5	92.0	85.3
Error	6.6	16.2	9.8	16.4
Sensitivity	97.1	90.1	94.2	89.8
Specificity	91.9	77.3	88.0	77.3
Precision	96.2	88.1	93.9	88.1
F1-score	96.7	89.0	94.1	88.9
Kappa	87.4	66.0	80.4	65.6

Table.3. Comparison values of Testing Dataset

Performance Metrics	HMM-SVM (%)	ANN (%)	HMM (%)	SVM (%)
Accuracy	93.4	83.8	90.2	83.6
Error	6.7	16.5	10	16.7
Sensitivity	95.2	88.3	92.4	88
Specificity	90.1	75.8	86.3	75.8
Precision	94.3	86.4	92.1	86.4
F1-score	94.8	87.3	92.3	87.2
Kappa	85.7	64.7	78.8	64.3

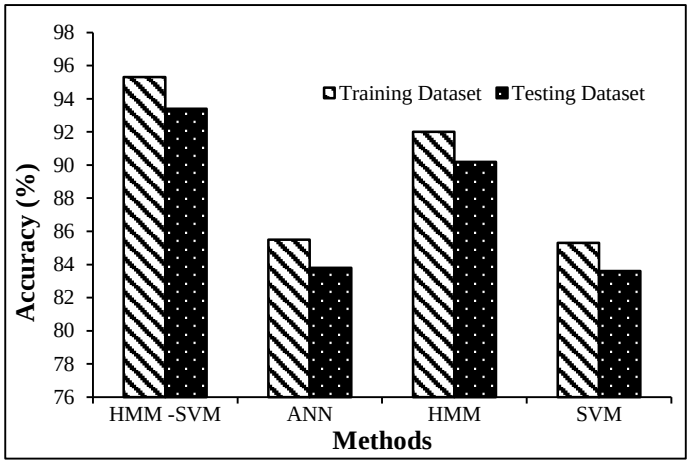

Fig.2. Comparison of Accuracy between training and test dataset

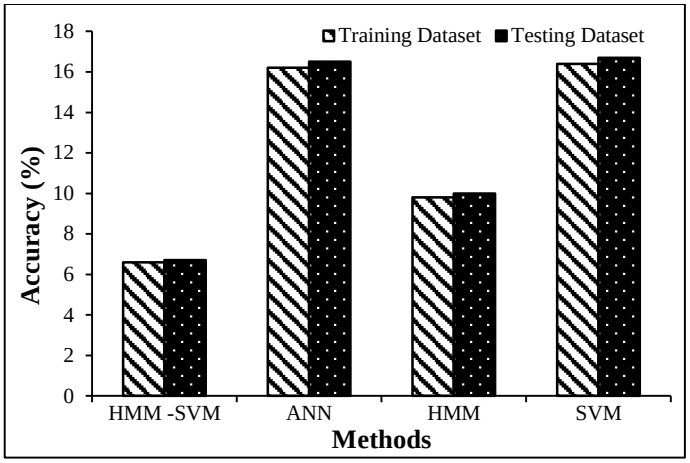

Fig.3. Comparison of error between training and test dataset

6. CONCLUSION

Cloud computing environment security is a serious problem that the scientific community should take into consideration. DDoS attack is one of the harmful attacks aimed at cloud technology systems being available. Throughout this paper, presented a detailed cloud environment analysis of DDoS attacks. In addition, have discussed some of the security mechanisms suggested in the literature against DDoS attacks. We plan to research these attacks in detail in future work in order to develop methods for the prevention and/or detection of DDoS attacks. The HMM-SVM is useful for calculating employability smoothness of data provided in a simple manner. Through the hybridized HMM-SVM, employer can easily filter the best applicants based on their education skill and personnel development skills. The proposed CSFS and the HMM-SVM used the questionnaire approach on the features acquired through the gathered data. Chi-Square, Gini Index, knowledge gain and correlation coefficient methods are used to pick the features we are using, and the CSFS algorithm is used to select the best one. The HMM and SVM classifier hybridization is used for the classification process and has achieved 93.4% accuracy. The results of the experiment are analyzed and correlated with existing classifiers such as SVM, HMM, and ANN.

REFERENCES

[1] B.R. Kandukuri, V.R. Paturi and A. Rakshit, "Cloud Security Issues, in: Services Computing", *Proceedings of IEEE International Conference on Services Computing*, 2009, pp. 517-520, 2009.

[2] L.M. Kaufman, "Can Public-Cloud Security meet its Unique Challenges?", *IEEE Security and Privacy*, Vol. 4, No. 8, pp. 55-57, 2010.

[3] W. Xing, E.P.R. Guo and S. Goggins, "Participation-Based Student Final Performance Prediction Model through interpretable Genetic Programming: Integrating Learning Analytics, Educational Data Mining and Theory", *Computers in Human Behavior*, Vol. 47, pp. 168-181, 2015.

[4] Q. Yan, R. Yu, Q. Gong and J. Li, "Software-Defined Networking (SDN) and Distributed Denial of Service (DDoS) Attacks in Cloud Computing Environments: A

Survey, Some Research Issues, and Challenges", *IEEE Communications Surveys and Tutorials*, Vol. 18, No. 1, pp. 602-622, 2015.

[5] G. Somani, M.S. Gaur, D. Sanghi, M. Conti, "DDoS Attacks in Cloud Computing: Collateral Damage to Non-Targets", *Computer Networks*, Vol. 109, No. 2, pp. 157-171, 2016.

[6] A. Badr and A. William, "*Proactive Approach for the Prevention of DDoS Attacks in Cloud Computing Environments*", Springer, 2017.

[7] L. Feinstein, D. Schnackenberg, R. Balupari and D. Kindred, "Statistical Approaches to DDoS Attack Detection and Response", *Proceedings of IEEE International Conference on Information Survivability Conference and Exposition*, pp. 303-314, 2003.

[8] K. Bunkar, U.K. Singh, B. Pandya and R. Bunkar, "Data Mining: Prediction for Performance Improvement of Graduate Students using Classification", *Proceedings of IEEE International Conference on Wireless and Optical Communications Networks*, pp. 20-27, 2012.

[9] A. Keramati, R.J. Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari and U. Abbasi, "Improved Churn Prediction in Telecommunication Industry using Data Mining Techniques", *Applied Soft Computing*, Vol. 24, pp. 994-1012, 2014.

[10] L.P. Macfadyen and S. Dawson. "Mining LMS Data to Develop an "Early Warning System" for Educators: A Proof of Concept", *Computers and Education*, Vol. 54, No. 2, pp. 588-599, 2010.

[11] Xutao Zhao, "Study on DDoS Attacks based on DPDK in Cloud Computing", *Proceedings of 3rd IEEE International Conference on Computational Intelligence and Communication Technology*, pp. 1-5, 2017.

[12] S. Aqeel, L. David, L. Yan and D. Mohammed, "An Efficient DDoS TCP Flood Attack Detection and Prevention System in a Cloud Environment", *IEEE Access*, Vol. 5, pp. 6036-6048, 2017.

[13] G. Somani, M.S. Gaur, D. Sanghi, M. Conti and R. Buyya, "DDoS Attacks in Cloud Computing: Issues, Taxonomy, and Future Directions", *Computer Communications*, Vol. 107, pp. 30-48, 2017.

[14] B.B. Gupta and O.P. Badve, "Taxonomy of DoS and DDoS Attacks and Desirable Defense Mechanism in a Cloud Computing Environment", *Neural Computing and Applications*, Vol. 28, pp. 3655-3682, 2017.

[15] N. Agrawal and S. Tapaswi, "Defense Schemes for Variants of Distributed Denial-of-Service (DDoS) Attacks in Cloud Computing: A Survey", *Information Security Journal: A Global Perspective*, Vol. 26, No. 2, pp. 1-13, 2017.

[16] A. Rukavitsyn, K. Borisenko and A.Shorov, "Self-Learning Method for DDoS Detection Model in Cloud Computing", *Proceedings of IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering*, pp. 1-6, 2017.

[17] S.M. Himadri, H.M.D. Tariq, H.M.D. Bellal, R.M. Ekhlasur and R. Hasan, "Enhancing Secure Cloud Computing Environment by Detecting DDoS Attack using Fuzzy Logic", *Proceedings of International Conference on Electrical Information and Communication Technology*, pp. 7-9, 2017.

[18] R. Kesavamoorthy and K.R. Soundar, "*Swarm Intelligence based Autonomous DDoS Attack Detection and Defense using Multi Agent System*", Springer, 2018.

[19] W. Alosaimi and K.A. Begain, "A New Method to Mitigate the Impacts of the Economical Denial of Sustainability Attacks against the Cloud", *Proceedings of 14th Annual Post Graduates Symposium on the Convergence of Telecommunication, Networking and Broadcasting*, pp. 116-121, 2013.

[20] H. Wang, Q. Jia, D. Fleck, W. Powell, F. Li and A. Stavrou, "A Moving Target DDoS Defense Mechanism", *Computer Communications*, Vol. 46, pp. 10-21, 2014.

[21] T. Karnwal, T. Sivakumar and G. Aghila, "A Comber Approach to Protect Cloud Computing against XML DDoS and HTTP DDoS Attack", *Proceedings of International Conference on Electrical, Electronics and Computer Science*, pp. 1-5, 2012.

[22] T. Anderson, T. Roscoe and D. Wetherall, "Preventing Internet Denial-of-Service with Capabilities", *Proceedings of ACM International Conference on Computer Communications*, pp. 39-44, 2004.

[23] M. Masood, Z. Anwar, S.A. Raza and M.A. Hur, "EDoS Armor: A Cost Effective Economic Denial of Sustainability Attack Mitigation Framework for E-Commerce Applications in Cloud Environments", *Proceedings of 16th IEEE International Conference on Multi Topic*, pp. 37-42, 2013.

[24] Q. Jia, H. Wang, D. Fleck, F. Li, A. Stavrou and W. Powell, "Catch Me If You can: A Cloud-Enabled DDoS Defense", *Proceedings of IEEE International Conference on Dependable Systems and Networks*, pp. 264-275, 2014.

[25] Q. Chen, W. Lin, W. Dou and S. Yu, "CBF: A Packet Filtering Method for DDoS Attack Defense in Cloud Environment", *Proceedings of IEEE International Conference on Dependable, Autonomic and Secure Computing*, pp. 427-434, 2011.