Detection and mitigation of DDoS attack in cloud computing using machine learning algorithm

Aroosh Amjad, Tahir Alyas*, Umer Farooq, Muhammad Arslan Tariq

Department of Computer Science, Lahore Garrison University, Pakistan

Abstract

Cloud computing, with its staggering and on-demand services had revamped the technology so far. Cloud consumers are freely to use the applications and software on the premises of Pay-as-you go concept. This concept decreased the cost and make the services less expensive and more reliable. One of the most important characteristic of cloud structure is on demand self-service. Cloud computing applications can be accessed anywhere at any time with much less cost. As cloud provide its consumers with its tremendous on demand services, besides this it is surviving from the excruciating security issues that are discourteous towards the cloud. There are, as many different attacks that results in making the servers down. One of the most hazardous attack is DDoS. This paper hiloghted the DDoS attack and its prevention technique which results in making the server side less vulnerable. The scenario includes, a transmission of million and trillion of packets in the form of DDoS at cloud-based websites, thus making it differentiated though different hosts. Making use of operating systems such as ParrotSec to make the attack possible. Last step includes detection and prevention through the most e

Keywords: Cloud computing, DDoS, vulnerability, sql injection, mitigation, TCP/IP, UDP, ICMP packets, malicious, exploit

Received on 25 May 2019; accepted on 01 August 2019; published on 12 August 2019

1. Introduction

Cloud computing is a great rewording for centralism of different computer services under one server”. Programs and data are being removed from PCs and desktop computers and moved on in “cloud”. Cloud computing is very strong competitor within IT world and it offers “pay as you go” with low cost services. All large industries and organizations have switched their data into the cloud. Cloud computing has mitigated the different issues regarding time, effort and cost by providing services with least amount, minimum time and less effort. There are different types of services that cloud computing is providing to its users but three are the most effective services which are being used commonly like: SaaS, PaaS, and IaaS. Plus, cloud computing has an additional characteristic as shown in Figure 1. As it is obvious that everything which facilitates people, may have issues also. Cloud computing is facing security issues because there is tremendous amount is data is included within cloud, more the cloud computing is going upward, more issues are coming towards the cloud. And the major and the most concerned issue is Security, what if you have submitted your Jewells in bank and you came to know that there is no security guard present out there, although is necessary to store you jewels in a safe place, as like every user’s data is situated within cloud and different hackers. are attacking out there, having different purposed.[1] There are many different attacks that can cause severe damage to data, present on a cloud, but many authors have highlighted DDoS Attack which affect more badly and further it can be considered as an alarm for cloud consumers. DDoS, simultaneously attack from multiple systems at one single system, millions of packets are being sent at

*Corresponding author. Email: tahiralyas@lgu.edu.pk
one target to slow down its all services. DDos can attack on a network, on an application more over on different websites to make all the services down, service providers and large companies are mainly targeted of DDOS attack and every commercial. And government companies are also targeted.

1.1. Essential characteristics of Cloud Computing:

There are five different characteristics under cloud environment as shown in figure 2.

1.1.1. Broad Network Access: As its name imply, broad network access, is that the cloud is on broad network access, in this case every user can get access of different services provided by cloud, many of the companies get facility to stay updated with their customers and other companies through cloud services but importantly, it depend on the network, that which type of network is being used by an organization. Network is private then the provided information would be within members and in case of public cloud, information would supposed to be public that anyone can access that sort of information plus services provided by cloud. Most likely users offer private cloud network to get rid of security issues because in the case of public cloud user’s information can be leak out so to get rid of from this problem users prefer to be in private cloud.

1.1.2. Rapid Elasticity: It is one of the most important factors within cloud that provide its users with all the facilities in a very reliable way. Cloud is the only that provides all the services very easily and users get access from the cloud more easily on demand. Moreover, cloud have provided their users the great advantage of storage in this way cloud users can get more services according to the free storage (N. Tabassum, M. S. Khan, & Alyas, 2019).

1.1.3. Measured Services: Cloud computing apply metering on the resources and different organizations have to pay for what they have used. In case of optimizing resource utilization it can be leverage by pay per use capabilities. Virtual servers’ instances that stored in the cloud are getting monitored, measured and then reported by the service provider on cloud.(Boroujerdi & Ayat , 2014).

1.1.4. On Demand Self Service: On-demand self-service is a technique through which users can facilitate from different services coming through service providers at any time and at any place. Users get this facility to get services on their demand. Like network storage and server time, as automatically needed without the interference of human. Different companies like Microsoft, IBM and HP are the most prominent companies used as in-demand seller, these organizations only need to provide products on-demand and sell services, resources to their consumers to facilitate them.(Kwiat, Kamhoua, 2018).

1.1.5. Measured Service: Measured service means paying cloud service cost as per their usage. It also known as metered service. In measured services all the problems and faults are being controlled and monitored. Measured service is a term that IT experts apply to distribute computing.

1.1.6. Resource Pooling: Resource pooling is a method within cloud that provide resources to the customers and release on demand, typically PaaS users get the resources from resource pool on the basis on their demand and give the resources back to the resource pool if it does not need. In this way, all the complexities would be mitigated that occur in cloud while using resource pool method. Resource pooling allow the cloud suppliers to provide all the resources to their customers and users. Resource pooling is the way through which every user can get resource that can be accessed on demand.(Jing, Yan, 2019)

1.2. Cloud Service Model:

There are three basic service model

1.2.1. Infrastructure-as-a-Service: IaaS yields the complete computing resources in the form of virtualization over the internet. IaaS users gain resources and complete services through WAN like internet. Users can use the services provided by cloud to install the elements of an application. Generally, IaaS consumers gain services on a pay as you go services, typically by hour, week or month. Some users are to be paid by using virtual machine space they have used. The pay as you go thing remove capital expense to deploy in-house software and hardware(Rudol,Implementasi, 2019).
1.2. Platform-as-a-Service: Platform as a service, where cloud providers give access to their consumers to use the software platform for their systems. All hardware and software require to build such applications which are provided by PaaS provider though different means like dedicated network, VPNs. Users pay by using the platform. PaaS help to develop web-based applications without any cost with less complexity from buying new hardware or operating systems. PaaS make such an environment which provides with the ability of IT resource as a service. Moreover, it provides the complete package of developing and building the complete cycle of web-based applications (Mondal, Hasan, 2018).

1.2.3. Software-as-a-Service: The first and most important one is Software as a service, provides a desktop functionality through web browser and applications are delivered online. The ability provided to the customers running on the cloud, to use the applications is SaaS. The applications are used among various users through different devices like browsers. It not the headache of users from where applications and browsers are accessing on, they do not handle the server-side data and does not control backend infrastructure on cloud like network, OS, storage, servers. SaaS can also be cited as “on-demand software”, where users may have access to application software and programs where as cloud service providers manage the backend infrastructure that help to run applications. (Biswas, Wu, 2019).

1.3. Cloud Development Model

There are four Cloud development model in cloud computing.

1.3.1. Private Clouds. Private cloud is an infrastructure, in which a single organization is supposed to use exclusively, an organization which comprises of multiple consumers. Private clouds are also considered to be as an internal cloud. A private cloud is built and managed by external service providers. It may have an extra control over reliability, performance and on security. Although private clouds are being criticized from being similar with accustomed server farms and even not providing such benefits like capital costs. (Abbas, Alyas, 2019)

1.3.2. Public Cloud. A cloud where services are offered as publicly by service providers thus, they offer many benefits to their service providers which includes negative capital on infrastructure etc. public clouds have complete control over network, data and security which pampered their effectiveness in many businesses. (Iqbal, Adnan, 2016)

1.3.3. Hybrid Cloud. Hybrid clouds is a combination of both public and private cloud. Hybrid clouds have functionalities of both of them (public and private) although hybrid include infrastructure that runs in private cloud and remaining part runs in public cloud. It includes more flexibility because of having both public and private cloud. Moreover, they have enough control over privacy as compare to the public clouds perhaps facilitating on-demand service.

1.3.4. Community Cloud. These are shared infrastructure for specific community of consumers from an organization that have common concerns. It can be operated or managed by one or more than one organization within in community.

1.4. Types of DDoS

there are the following types of DDoS attacks.
1.4.1. SYN Flooded attack. SYN flood attack breakdown an abnormality in TCP connection that includes three-way handshake, normally sync packets are sent towards the other host in order to check whether the host side is active or not, in response it gives back its ACK plus with SYN packet, ultimately through this procedure, a new connection can be build. But in case of SYN flood, the sender sends multiple SYN flood requests and neither respond back to the host ACK and continue to send requests. The host side continue to wait for ACK for each request and bind all the resources for new connection but all in vein and eventually it results in Denial of Service.

1.4.2. Ping of death. This attack includes sending Denial of Service. connection but all in vein and eventually it results in for each request and bind all the resources for new connection but all in vein and eventually it results in Denial of Service.

1.4.3. NTP amplification. Network time protocol pre-press exploits publicly to over-whelm the targeted side with UDP traffic. It involves the procedure query to response system and specific ratio is been defines like 1:20 to 1:200 and in case any of attacker get that list of open NTP server by using any tool like Metasploit can hit the target and can launch a devastated attack of DDoS.(Neupane, Neely, 2019)

1.4.4. HTTP flood. Attacker exploits genuine GET or POST http requests in order to get attack any web server or an application. In case of each single request allocation of max resources, attack is most effective.

1.4.5. UDP flood. According to actual definition, UDP flood is an attack which victimize the server end with the use of user datagram protocol packets. This attack flood on randomly selected ports on a distant host. Host in this way, frequently check the applications listening at port, in case of no application found it might reply with ICMP. So, this process can ultimately lead towards the inaccessibility.

1.4.6. Zero-day attacks. This attack includes all the anonymous or new attacks, it may exploit all those vulnerabilities for them, no patches are been introduced yet. This term is famous among the fell known member of hacker’s community

2. Related work

Cloud Computing provides the easiest way to share information and resources in the field of information technology. It has gain fame because it is scalable and can workout at any platform without any disruption. Cloud computing is reliable, flexible and thus facilitates us in such a way that we can gain any sort of information from anywhere at any time. Although Cloud Computing does not expense at all it works on the basis of Pay-as-you-go concept. Mainly cloud computing proffer three amazing services, such as: SaaS, IaaS, PaaS separately defined as software as a service, infrastructure as a service, platform as a service. SaaS provides services to the end users like google doc provide services, one application and different services and much more.(Mishra, Pilli,& Varadharajan, 2019) PaaS includes an activity such that, service providers like google engine grabs the machine instance and other details which are technically embedded within a machine, from a developer. IaaS is basically providing the backend level of abstraction, having dedicated servers which can be controlled by backend developers. In the context of cloud computing there are three types of clouds: private cloud, public cloud, hybrid cloud where as private cloud is controlled and thus created within an enterprise. The company which is providing cloud services is responsible for deploying and maintaining private clouds, a company which sells cloud services often controls public clouds. Whereas hybrid cloud is the combination of both public and private cloud. It is understood that cloud computing provides enough advantages in spite of that it may face many challenges like Security challenge.(Almseidin, Kovacs, 2016)

Cloud computing has faced many disruptions of DDoS attack, DDoS attack effect the services of cloud computing in such a way that attacker find some vulnerabilities in the software to affect the services provided by cloud computing. Secondly some attackers attack for the sake of consuming bandwidth and all resources. There are different prevention techniques of DDoS attack, the author in illustrated that the daily base scan of machine is foremost in case if there is any abnormal behavior within the system, it can be detected. By maintaining all the software and protocols the weaknesses and abnormalities of computer can be mitigated.(Soliman, Salama, 2019)

Different techniques of forensic in different phases to apply forensic technique in order to prevent DDoS attack. The author has completely defined the process of cloud computing and discussed its different models moreover traditional forensic and forensic investigations both are brought into the scene and compared with a throughout analysis. Forensic investigation in cloud computing is supposed to be very tough and time taking as there are a lot of devices which are connected within cloud computing that needs to be forensically tested. Cloud computing has a large variety of tool and may applications that are used for the analysis in forensic investigation.(Neupane, Neely
at el. 2019)

Security is a main concern within cloud computing to make it secure, cloud is very vast technology, stores a vast amount of data and transactions are done remotely within private or public network. Denial of service attack and distributed attacks kind of attack which gets down the performance of cloud services. The author in [21], aims in making survey analysis on this major issue on MANET and cloud computing previous research has concluded that the classification methods are obligatory. Author has done deep study on the DDOS attack and its drawbacks and concluded that it is very necessary to mitigate DDOS from cloud environments totally. (Abbas, Adnan, 2019)

DDoS attack prevents the availability of the services provided by cloud infrastructure. DDOS attacks by using black hole algorithm and neural network. For the purpose of detection author experiment a large data set of 12500 samples and test samples approx. 12500 and this experiment is performed ten times.DDOS attack is initiated through different workbenches which would go through fog defender into the cloud. Moreover, different rules are being pertained within fog defender to prevent and protect the cloud from DDOS. For less response time and less resource utilization author have worked within one policy that, mitigation and prevention is only applied on the edge of network not in cloud in this way resource and time would be lessen. The approach which author has used is only for TCP and HTTP traffic. (Alosaimi, M. Alshamrani, 2016)

3. Proposed Methodology

The first and major task is Information Gathering, information gathering is a process in which we can find out the different vulnerabilities of the victim machines in order to attack at victim site. Information Gathering involves all the information about the ongoing services, open or closed ports and other weaknesses. In this case attacker might get information about the weak points of victim and can attack conveniently. Every service that cloud computing provides, have specific port number like: http uses port no 80, 20 to 23 is being used by TCP and UDP performing different functionalities, ftp runs on 990 but sometimes on 21 as well and moreover. Summing up, information gathering is such a process which provides all the relevant information about the vulnerabilities of any system so that attacker might attack accordingly. Nmap scanner is a tool, which is used for the purpose of information gathering, it only needs an IP address of the victim machine, once it has that IP, it starts scanning the whole system that shows different activities, different services that another system is providing, open and closed ports thus all those activities which are going on can be shown up even though operating system which is used by another

system would also be shown. Once the opened port is found. An attack would be generated, in our case we are using Distributed denial of service attack, which includes further different attacks like ping of death, DDOS is an extremely worst attack that ruin all the services of the system. DDOS apply million number of packets that would be merely intolerable and make the all services down. ParrotSec have a mechanism, it includes shell or terminal through which all the operations can be performed by applying commands as like other operating systems such as Kali and Ubuntu, ParrotSec is also command line interface and by giving commands ParrotSec perform all the operations thus at terminal, the command PING IP would be applied and thus more than 65000 packets would be sent to the victim site and all the services became down. This is how attack would be generated. Next phase is detection phase, In this scenario, cloud-based website would be targeted, Nmap would scan the all vulnerabilities of our target website in this way it’s all abnormalities would bring into the scene. Once the weakened ports are shown, a python-based script would be generated which includes DDoS. Once DDoS do attack on the specific website it may brought their services down onto their knees. Next phase is detection phase, Sniffing is the procedure, which monitors and capture every packet during transmission. It monitors all the inbound and outbound traffic on the internet and make analysis of each packet. Sniffing overall sniff every packet in order to analyses it properly. In this case, sniffing would be done at server side. There are many tools for sniffing purposes but most of the important one is Wireshark. Wireshark do sniff each and every packet deeply. After the overall analysis of packets, a large data set has been created which implies at classifier. Random forest and naïve Bayes are well- known classifiers and both are being compared which displays excellent results as shown in experimental setup. There are as many other classifiers through which detection can be possible like Naïve Bayes, SVM, KNN and k-means etc., but in spite

![Figure 4. Proposed Model](EAI Endorsed Transactions on Scalable Information Systems 07 2019 - 10 2019 | Volume 6 | Issue 23 | e7)
all of that. The most powerful classifier is "Naïve Bayes". In this work, naïve Bayes is used for prediction of packets while the DDoS attack at application layer. Naïve Bayes being a simple but efficient method, can predict the situation according to the given data. We have trained our analyzed data set within naïve Bayes and further with a cross validation method of 60 folds, we generated a new prediction based data set which gives output in different forms like: true positive rate, true negative rate, false positive rate, false negative rate and much more in order to understand the packets which travelled form source to destination. Naïve Bayes being a predicted approach gives the output as true positive and false negative. False negative is considered as an alarm for network users. Naïve Bayes and random forest detected false negative rate as a DDoS packet and true positive rate as a normal packet.

4. Simulation and Results

4.1. Data pre-processing

Data pre-processing is most efficient technique within data mining which makes the raw data into an easy and understandable form. As real-time data is inconsistent and incomplete but preprocessing is most useful technique through which less useful data can be converted into useful data. Weka has many preprocessing filters. Among all of them a single filter is chosen such as: normalize.

Basically normalization is a process of making data un-redundant able or removing all the redundancies and duplication from the data.

4.2. Training data set

Training of data set includes the development of machine learning model. Training data is usually used to train an algorithm. In a generic means, training data is certain amount of percentage from the overall dataset including testing dataset. To make a separation between testing and training of dataset is much important for building a machine learning based model. A machine learning based model is necessary although to make further prediction against the trained dataset.

4.3. Prediction algorithm

So far, after training and testing of data set, different algorithm are generated hereby in order to predict Many of the problems. In this case, detection of DDoS (malicious or non-malicious) packets are to be considered.

4.4. Prediction of naïve Bayes

This figure shows true positive and false positive rates. False positive rate is considered to be an alarm of DDoS or false packets where as true positives are the normal one. Here, the average mean of true packets are 0.982 while malicious one is about 0.01.

4.5. Proposed formula for naïve Bayes

\[ P(b|c) = \frac{P(c|b) \cdot P(b)}{P(c)} \]

Where, \( P(c|b) \) is the likelihood

\[ P(b) = \text{class prior probability} \]

\[ P(c) = \text{predictor prior chance} \]

This graph shows the detected clusters resulted from naïve Bayes.
5. Conclusion

This research comprises of DDoS attack its detection and prevention. The first and major thing to do is finding vulnerability of ports, the weak ports can be easily get hacked though. We had chosen ParrotSec operating system as it is effective and most reliable operating system. As DDoS transmit million number of packets towards the destination end, so we had chosen one cloud-based website to attack. Once the attack had been established, website gets down Machine learning plays a vital role in detection as well. The analysed data is being trained in the most common yet powerful tool ‘weka’. By applying pre-processing with the use of “discretize” filter. Thus the next step is quite interesting and useful in prediction in detection. We have used these two algorithms and compare within a single platform thus concluding the positives with naïve Bayes. Through complete analysis it came to know that naïve Bayes is stronger than random forest. False positive rate is an alarm for the transmission of packets within a network. And naïve Bayes prediction is far greater than random forest. Naïve Bayes detected the false rate of packets and true rate of packets efficiently than that of random forest.

References


