Secure and QoS Aware Architecture for Cloud Using Software Defined Networks and Hadoop

Abhijeet Desai[1], Nagegowda K S[3]
Department of Computer Science and Engineering, PESIT
Bangalore, 560085, India
abhijeetdesai14@gmail.com, nagegowda@pes.edu

Ninikrishna T[2]
Dept. of Electronics and Communication Engg, CMRIT
Bangalore, 560037, India
t.ninikrishna@gmail.com

Abstract—Cloud services have become a daily norm in today’s world. Many services today are been migrated to the cloud. Although it has its own benefits it is difficult to manage due to the sheer volume of data and the various different types of services provided. Adhering to the Service Level Agreement (SLA) becomes a challenging task. Also the security of the cloud is very important since if broken all the services provided by the cloud distributor are at risk. Thus there is need of an architecture which is better equipped with security as well as adhering to the quality of service (QoS) written in the SLA given to the tenants of the cloud. In this paper we propose an architecture which will be use software defined networking (SDN) and Hadoop to provide QoS aware and secure architecture. We will also use Kerberos for authentication and single sign on (SSO). In this paper we have shown the sequence of flows of data in a cloud center and how the proposed architecture takes care of it and is equipped to manage the cloud compared to the existing system.

Keywords—Software defined networks (SDN); Cloud computing; Hadoop; Big data; Kerberos; Quality of service (QoS);

I. INTRODUCTION

Cloud computing is a model which gives on-demand access to shared computing resources. It provides for customers to use third party resources to process their tasks and store their data. There are three basic models in cloud computing Infrastructure-as-a-service (IaaS), Platform-as-a-service (PaaS) and Software-as-a-service (SaaS). In IaaS the clients can get physical and virtual machines from the provider for their organization. In PaaS the cloud service providers provides computing platforms like databases and development tools. In SaaS the cloud provider provides softwares such as games and email services to the customer.

Software defined networking (SDN) is a method of networking which provides abstraction by separating the control plane from the data plane in the network. The data plane will be the one which is responsible to forward the packets in the network. They will be dumb, meaning they won’t be having any intelligence and will do whatever the control plane asks it to do. The data plane will consists of rules according to which it will forward the packets. The data plane consists of programmable switches. The control plane is the master of the data plane. The control plane is intelligent and programmable and it sends rules to the data plane which the data plane has to follow. There can be one or many controllers controlling thousands of switches.

Openflow is a protocol which is been used by SDN widely to send messages from control plane to data plane and vice versa. Northbound apis are used to between the controller and applications to communicate with each other. Southbound apis are used between the controller and the underlying switches to provide granular control of the switches. Any switch which supports openflow protocol or similar protocols which are helpful to provide abstraction can be used with a compatible controller. Most of the switches support openflow protocols and it has become a de facto standard in many ways. There are several open source controllers like POX, NOX, floodlight, beacon and openaylight. We can emulate SDN switches with the help of mininet emulator. Using mininet we can create several openflow enabled switches and we can use some controller say openaylight to add intelligence to them.

Hadoop is used for distributed processing. Using commodity hardware we can combine the processing power of several machines and perform tasks parallel to get faster output. By the virtue of Hadoop distributed file system (hdfs) and mapreduce it has become easy to perform computations on several machines by writing programs. These programs are automatically scalable which means there is no need to write different program for working on 100 machines and different for 10000 machines as Hadoop takes care of it. The central node called as the name node will be controlling the other nodes which are called as data nodes. Hadoop is very useful for handling of big data due to its parallel processing.

The objective of this paper is to propose an architecture which provides better security to the cloud and also takes care of QoS requirements. Thus we combined SDN and Hadoop to fulfil these needs since we have seen in our previous experiments in papers [1] and [2] that we can use SDN and Hadoop to get network characteristics and to set rules to the incoming traffic. We are also using Kerberos to enhance the security and provide authentication and single sign on (SSO).
II. RELATED WORK

Several attempts have been made to combine the power of SDN along with cloud. Yen [3] proposes a mathematical model of SDN based on cloud computing. The paper [3] uses queuing theory as the bases of computations and shows how to achieve better load balancing and monitoring services. Chase [6] shows how bandwidth allocation happens among various virtual machines in cloud and SDN.

Baucke [4] shows how SDN can be used in Cloud to improve the security, establish VPN and also to have support for quality of service (QoS). Yan [5] shows how distributed denial of service attacks can be prevented in cloud using SDN. Paper [5] proposes an architecture which can be used to detect various distributed denial of service attacks and compares it with the existing system and which all features does the new architecture brings.

In [1] which is written by us we showed with results how SDN can be used along with Hadoop to get and set the quality of service (QoS). We also showed how Hadoop gives feedback to the SDN controller and how it is easy to get the network characteristics. In [2] we along with Mantur using the basic architecture as in [1] were able to enhance it and provide firewall rules and have a statistical anomaly based attack detection system.

In general we have seen instances of SDN used with cloud. We have also seen instances of Hadoop and SDN used together. In this paper we propose an architecture which can be used to enhance the security and support QoS.

III. PROPOSED CLOUD ARCHITECTURE

The proposed cloud architecture is as shown in figure 1. Following are the components of the proposed architecture.

A. Demiliterized Zone (DMZ)

The DMZ consists of its own topology and open flow enabled switches and an SDN controller. There should be more than one SDN controller in order to have a failsafe mechanism in case one controller is compromised or can’t handle the traffic. The packets from the load balancer from the outside world come to the switches where proper rules are programmed and according to which the packets are taken care of. If appropriate rules are not present the packet information is purged to the SDN controller which will decide on the action to be taken on the packets. The SDN controller can act as a firewall or we can install a separate firewall server if needed in the DMZ. The SDN controller will also be continuously creating new rules and making itself smart by learning from the past experiences and from the feedback given by Hadoop. In our previous paper [2] we can get a gist of how the firewall and statistical based anomaly detection will work.

B. Authentication server/Ticket granting server (AS/TGS)

When a new user tries to access the cloud the SDN controller will send the required information to the authentication server which will then authenticate the
user. The user can then ask permission to access the required servers and based on the Ticket Granting server the ticket granting ticket can be issued with the help which the user can be authorized to access specific servers. Also due to the nature of the Kerberos we will obtain a single sign on service in which the user can be authenticated once to the authentication server and then it will be able to access all the services in the servers to which it has authorization. The user uses the shared secret between it and the application servers in the data centers to provide a secure transmission of data.

C. Smart Load balancer

The filtered packets from the DMZ are sent to the smart load balancer. This as any load balancer does distribute the load of the packets on the openflow switches. Apart from that the DMZ SDN controller sends control information to the smart load balancer as to send the packet information to the Hadoop. The smart load balancer sends the required information used to calculate the required characteristics necessary for the administrator of the cloud management.

D. SDN which is inside the cloud

This SDN controller will be having more management related functionalities compared to the DMZ SDN which has more security functionalities. However there will be some security related tasks also such as related to authorization. The SDN Controller will be getting a feedback from Hadoop regarding security and management. An example is an authenticated user trying to access services which it is not authorized to. Also this SDN Controller will be responsible for the quality of service of the flows. As we know that the cloud tenants will be having Service Level Agreement (SLA) with the cloud vendors using these SDN controllers we can easily adhere to the agreements and allocate or reserve required resources to the tenants. Also due to a central point of control the management of the systems will be far easier and efficient. We will also have a failsafe mechanism or have backups of the SDN controller. These controller will be continuously synchronized and will be sending heartbeat to show it is alive that is up and running. The mediar will take care of synchronization and several other simple things such as distribution of the control from the SDN controller to a particular switch.

E. Hadoop

We know that Hadoop is used for processing big data. Since we are dealing with cloud the data will obviously be big data. Hadoop will be useful in both security and management of the cloud. The DMZ SDN controller will give instructions to the smart load balancer on which packet information to send to Hadoop. Using parallel processing using mapreduce, hdfs and several other constructs used in Hadoop, it will calculate the statistical anomaly based attacks as in our previous paper [2]. Also the network capacity planning can be done and the network characteristics can be computed using custom groups such as characteristics per tenant group. In our previous paper [1] we can see the network characteristics based on several factors such as IP address, protocol type, number of sessions and bandwidth consumption. The administrator can specify and customize as to what information he will need for better management and control of the cloud. Also from the core data center where we have the servers, they can give data to Hadoop. This data will be considered as high priority data since it is directly coming from the servers where the actual processing or data storage happens. The Hadoop will give a feedback to the DMZ SDN controller and the SDN controller inside the data center for better management and security. In our papers [1] and [2] we can see an example of the feedback which can be given. Due to continuous feedback the SDN controllers become smarter and secure.

F. Core data center

This will be the main application servers where the services will be running. It can also be storage centers where huge amount of data is stored. Also these can have virtual machines (VM) which provide infrastructure to some organization.

IV. SEQUENCE OF EVENTS

The flow of the data in the architecture is shown using the sequence diagram as shown in Fig. 2.

A. Session Setup

The initial session is setup between the external user and the cloud. Following are the steps that happen during this initial flow.
1. The user/client sends a request to the DMZ SDN controller for a ticket granting ticket (TGT) to access the servers in the core data center.
2. The DMZ SDN controller forwards this request to the authentication server (AS).
3. The authentication server responds with a ticket to the Ticket Granting Server (TGS) along with data like client ID and timestamp. This ticket is sent to the client by the DMZ SDN controller.
4. The ticket received by the client can be decrypted only if the client is legitimate. The ticket is then sent to the TGS which in some cases can be the same as the authentication server.
5. The TGS will respond with a session key which can be used by the client and the servers to exchange data.
6. The same session key is also used in single sign on since the TGS will be sharing the information about the session key to all the servers which the user wishes to access.

B. Data Flow

After the session setup the data flows from the DMZ SDN controller to the internal SDN controller to the servers and vice versa.

C. Background processing

The background processing happens behind the scenes parallel to the data flow and session setup. This is very important since this is used to improve the quality of service of the cloud and also used to enhance the security.

1. The DMZ SDN controller instructs the smart load balancer to capture specific parts of packets for flows and send it to Hadoop. The information captured from packets can vary depending on the source IP or specific flow which the DMZ SDN controller suspects.
2. The core data center that is the servers also can give a feedback to Hadoop informing how to refine the capture of information from the packets and what is useful to the servers. This feedback is set as high priority since it directly comes from the servers where the processing of data happens.
3. The Hadoop will be sending feedback to the DMZ SDN controller and the internal SDN controller to get the characteristics of the data and do a capacity planning as well as the security analysis.
V. CONCLUSION

As we know that Software defined networking (SDN) is an emerging technology, we are trying to find out more applications of it in various fields. In this paper we showed how we can apply SDN to cloud computing and brought out a concept of integrating the features of SDN and Hadoop into cloud computing.

The architecture we proposed can be used to set the quality of service (QoS) requirements for tenants of the cloud using the concept of SDN. Using SDN the segregation of the flows thus controlling the QoS can be accomplished. We can set priority for particular tenants and meet SLA demands when required.

Further this architecture increases the security of the cloud by employing Kerberos and the SDN in tandem with Hadoop. The security aspect will be seamlessly integrated and it will be easier to detect and react to threats due to the programmability of the network. Long term attacks or snooping can be detected by using Hadoop analysis. Also due to Kerberos we can employ single sign on (SSO) due to which it will enhance the user experience.

The data processed in Hadoop can also be used to get the characteristics of data and can be used by the administrator to manage the data effectively. As an example the administrator can use the data for capacity planning since he/she will be knowing which servers are used more often and are loaded and how to handle the traffic efficiently.

REFERENCES