Security analytics for data centers

Law, Security

Abstract

Data center and cloud business is growing very fast. Hence, it gets attraction of attackers to breach the security. Attacker tries different types of attacks such as Brute Force attack to crack password, Distributed Denial of Service (DDOS) attack to make resources unavailable from actual user. Also, attackers try to breach the security to get confidential data. Due to popularity of cloud, security and privacy become biggest concern in front of every organization. Hence, it is necessary for organization to prevent their resources and data from attackers. Attacks can be prevented using multiple ways such as configure firewalls to allow specific Internet Protocol (IP) address, use machine learning model to detect attack patterns, monitor network traffic etc. Visualizations plays an important role in monitoring network traffic and to represent large amount of data. Generate an alert based on condition and then act on it immediately.

I. Introduction

Cloud providers are offering different services based on ‘ pay as you go’ model. Hence, many businesses are shifting their business to cloud. As effect, lots of data is generated on daily basis. Security and privacy of data is utmost important for any organization. Security analytics can help detect different security attacks and ultimately helps organization to prevent from disasters.

Logging can be a security administrator’s best friend. It’s like an administrative partner that is always at work, never complains, never gets tired, and is always on top of things. If properly instructed, this partner can provide the time and place of every event that has occurred in the system. Each system collects its own logs and keep track of events such as login details, rpm package installation details etc., All of them can combine with each other to give a complete and a clear picture about all the events in the system. This data can also be used for the detection of anomalous activity both in real-time, as well as re-actively during an incident-response event. We can use the system logs and run deep learning model to detect different security attacks.

Different security services can be configured properly to prevent attacks. Security service like fail2ban [1] can be used to prevent brute force attack [8]. It blocks source IP temporarily after certain number of failed login attempts. Security service like wazuh [2] can help detect integrity of file and to generate alerts in real time.

II. System Design

A. General Overview

The overall architecture of our system is shown in Figure 1. Log aggregator service RSYSLOG [6], collects system logs and forwards it to Admin Log Virtual Machine (VM). Admin Log VM uses Fluentd [7] to upload those logs to Amazon S3 [3]. Kafka [5] is queuing service, which reads those logs from S3 and store into queue. Spark streaming [4] reads queue data and start processing it. Once required data is extracted, different security analytics can be performed like anomaly detection. At the end, all the processed data is pushed to Elasticsearch, Logstash, Kibana (ELK) stack. Security administrator use Kibana dashboards to monitor all the data in all the data center.

B. Functionalities Offered

1. Fault Tolerance: The ability of the system that allows it to continue operating despite the failure of some of its components. In our case even if one or more component fails then also system can continue its execution. In Figure 1, you can see backup Admin Log VM to handle fail over scenario.

2. Load Balancer: Load balancer allows us to efficiently distribute network traffic across multiple hosts in data center. We have cluster of Admin Log VM and all hosts sends the data to the load balancer and then load balancer distributed that data based on network traffic to each VM.

3. Heterogeneity of Hosts: Log forwarder service is deployed on different hosts such as Linux, Windows etc. which forwards those logs to Admin Log VM.

4. Security Analytics: Machine learning model is deployed on Spark Streaming server which allows to process log data and run security analytics such as anomaly detection.

5. Fail2ban Service: It is intrusion prevention service to protect hosts from brute force attacks. It blocks source IP temporarily, when it exceeds certain failed login attempts within specified time-period.

6. Visualization: Kibana dashboards are used to visualize all the gathered log data.

III. Process

A. Gather Logs Data and Upload it to Amazon S3

Hypervisor Logging is based on client server architecture. All Hosts (admin hosts, management hosts and game seat hosts) are considered as a client and Admin Log VM is considered as a server. Figure 2 shows the overall workflow of hypervisor logging. All the hypervisor uses RSYSLOG [6] to forward the logs to Admin Log VM. All hosts are configured to send the data to Admin Log VM’s Virtual IP (VIP) on port 514 using UDP protocol. Admin log VM listens on port 514 for UDP packets. Once packets are received, it stores the data in local files. Fluend [7] monitors the tail of these local files and upload the data to S3 as soon as its available.

B. Intrusion Detection System

Fail2ban is an Intrusion Detection System service which protects hosts from brute force attack [8]. We have created jenkin pipeline to deploy fail2ban [1] on hypervisors. We have configured fail2ban to block source IP for 5 minutes after 5 failed login attempts. We also configured additional rules in IP tables to exclude IP range within NVIDIA infrastructure.

C. Integrity Checking, Rootkit Detection, Time-Based Alerting System

We have service called WAZUH, which allows us to perform log analysis, file integrity check, rootkit detection, time-based alerting [2]. We have installed and configured wazuh agent service on hypervisors and network devices to send the data to wazuh-manager installed on Admin Log VM. Wazuh manager sends global configuration to all its agent to do certain tasks such as integrity check, rootkit detection based on OS/device types. Once wazuh-manager receives the data from its agent, it starts running different rules on that data. If any rule succeeds, it generates an alert in real-time and sends an email notification to security administrator.

Deep Learning is a branch of Machine Learning that focuses on learning complex relationships in data through high-level abstractions. It comprises of a set of algorithms and models that revolve around a graph-like structure between the input and the target output. The graph also contains a collection of nodes that capture latent features and high-level information contained within the data for modeling the relationships between the inputs and the outputs. Training such models requires dedicated hardware (such as GPUs) and specialized optimization techniques. A big chunk of deep learning research is dedicated to these two problems. Deep learning research is also focused on coming up with new graph structures, operations and model types. There are broadly two types of deep learning models.

The first type focuses on learning a series of transformations that converts the input to the target output. For input x and y, it tries to learn a cascaded function f(x) that approximates y. These type of deep learning models are called deep neural networks. Deep Neural Networks (DNNs) are artificial neural networks with many hidden layers (some people even consider more than 1 hidden layer as deep). Some examples of DNNs include multilayer perceptron, autoencoder, recurrent neural network, etc.

The second type focuses on learning a probability distribution between the variables (which may or may not be divided into input and output) and hidden/latent variables. Some examples are Deep Belief networks, Probabilistic Autoencoders and Deep Boltzmann Machines. These can be either directed or undirected probabilistic graphical models.

E. Kibana Visualization

Kibana dashboards are used to visualize the data. ELK stack allows us to load data into elastic search. We can define index pattern which allows discovery of data. Once we discover required data we can visualize it easily. We have dashboards to monitor network traffic, check login failures, daily log volume, Vulnerability and DDOS attack. Figure 3 shows the number of login events per data centers and graph in left corner shows that number of hosts are banned after failed login attempts in red color.

IV. Conclusion

Security and privacy of data is big concern. Logging can be a security administrator’s best friend. If properly instructed, this partner can provide the time and place of every event that has occurred in the network or system. Security analytics can be used to detect certain security attacks. Security services like fail2ban, wazuh can help prevent attacks. Even after preventing majors, if attack is successful then alerting can be used to take immediate actions. Visualization helps security administrator to monitor the data in real time.