AWS HONEYPOT ATTACK DATA

[1]

Team:
Satya Sai Jayanth Devineni (G01206829)
Yeshwanth Reddy Bommu (G01197092)
Kuldip Gadapa (G01239022)
Vineel Vishwanth Busi (G01222602)

INTRODUCTION:

According to Online Digital industry experts, a Cyberattack is defined as a trial to obtain unauthorized and unaware access to one's personal and official assets (data, money) through network which is finally lead to misuse and destroy their assets. It is an offensive maneuver that effects computer information systems, computer networks, infrastructure, and personal computer devices. It can be implemented by nations, states, individuals, hackers, groups, or organizations. In our project, we focused on Amazon Web Services (AWS) honeypot attack data and visualizing those cyberattacks [1].

Honeypot is an effective way of tackling the impact of cyberattack on computer infrastructure. It mimics a system to cause a cyberattack. It can be used to identify, deviate and know functionality of cyber criminals but it cannot stop attacks completely. They could be utilized as traps for cybercriminals as they think it is a legitimate target because it has system applications and data. AWS Honeypot resembles as Amazon network to outsiders which is maintained by Amazon IT teams. They monitor traffic for suspicious systems, track the attacks and operations to analyze what they want. Finally, they diagnose their security measures, firewall performances and perform necessary updates [1].

Some of the factors are:

It is both reliable and fast because it integrates with real systems.
- It is low cost virtual system.
- They are platform independent and operating systems.
- They support various network protocols.
- The route network traffic is maintained at any time.

LITERATURE REVIEW

Amazon web service (AWS) honeypot is nothing but a trap point and security mechanism deliberated to tempt the attempted attack and if any source accesses the honeypot, the IP addresses will be recorded. Generally, a honeypot is the distraction for the attacker from their actual attack attempt and it will collect the information of the attacker by observing their request responses and the target hosts. Nowadays the cyber-attacks are immeasurable and more sophisticated to the companies, individuals, industries and government [2].

In 1986, the system admin of UC Berkeley named Clifford Stoll was involved in a process to track the charge for $0.75 of a Unix system at the lab. He used two honeypots to track the attacker. The actual target for the attacker was the nuclear defense secrets and later Clifford Stoll created a fictional department working on "Star Wars" to attract the attacker and he was later arrested.

Since then honeypots became standard and the deception toolkit was launched in 1997. The honeynet project remained as the active security community resource [3].

There was a record of 451,581 attacks in a 6 months duration on AWS honeypots. AWS honeypot deals the attackers in a simple method by attracting the attackers with honeypot then the attacker will encounter the honeypot instead of our servers. The top 10 popular AWS data centers include Sydney, Sao Paulo, California, Mumbai, Frankfurt, London, Paris, Ireland, Singapore and Ohio were placed with the cloud server honeypots by an enterprise security company. Most of the honeypot projects are open source and there is a honeypot project that has extension tools where it will also analyze the data that will be collected by the honeypot [4].

PROBLEM DEFINITION

The key problem is to identify malicious activity that organizations tend to fortify. A honeypot is used for such purpose that will deliberately configure with known vulnerabilities at a location to make more tempting or obvious target for attackers. As honeypot has no production data or don't participate in legitimate traffic on your network and that is how we can record and identify cybercrime. The definition covers a diverse array of systems, from simple virtual machines which offer a few vulnerable systems to build fake networks spanning multiple servers. The goals of honeypot are diverse as they can be used as defense in depth to academic research. The 3 common types of honeypots are pure honeypot, high-interaction honeypot, and low-interaction honeypot [3].

Research honeypots allows close analysis of how hackers do their dirty work. The hacker's techniques on using infiltrate systems, escalate privileges, etc. are scrutinized. They are set up by security companies, academics, and government agencies to examine the threat landscape. However, once the honeypot is detected its value diminishes and it is used by spamming industries to identify spam-catching honeypots [3].

DATASET

Our dataset contains the attack data of the Amazon web services (AWS) containing the following data which include datetime, host, src, proto, type, spt, dpt, srcstr, cc, country, locale, locale abbr, postal code, latitude and longitude. Using this dataset, we can visualize the following which includes the geolocation of the attacked places, presenting the top attackers, detecting attacks by the host, and highly active IP addresses [5].

Nominal	Description
Datetime	Packet Arrival Date (YYYY-MM-DD)
host	Honeypot Server
src	Packet Source
proto	Packet Protocol Type
type	Packet Type
spt	Source Port
dpt	Destination Port
srcstr	Source IP Address
cc	Source Country Code
country	Source Country
locale	Source Location
localeabbr	Locale Abbreviation
postalcode	Postal Code

Ordinal	Description
latitude	Source Latitude
longitude	Source Longitude

This raw data will be then processed into a CSV file containing refined data about AWS honeypot. The dataset will have rows and 15 attributes.

Methodology:

From our dataset, we observed the data measurements (interval, nominal, ratio, ordinal) of our attributes. We would like to perform the quantitative research which is tested and objective. It also has both independent and dependent variables. For our hypothesis we want to use frequencies tables, cross tables and chi squared tests. We would like to use bar charts and line graphs to view the data in a simple way.

TIME PLAN:

The Time plan of the project proposal has tasks and implementation parts. It shows specific tasks to get completed in the desired time. We are implementing code and methods, Analyzing the performance, Project milestone 2 submitted within 04/08/2020, Final Project Report Submission will be submitted on 04/20/2020, later Preparation of Project Presentation will be done. The final document submission will be submitted within 05/04/2020.

References

[1] "Honeypot: Efficient and Cheap Way to Detect LAN Attacks," 2018. [Online]. Available: https://martinhaller.com/it-security/honeypot-efficient-and-cheap-way-to-detect-lan-attacks/.

[2] "AWS Honeypot Data: Visualizing The Threat of Cyberattacks," 2020. [Online]. Available: https://www.sisense.com/whitepapers/gofigure-aws-honeypot-data-visualizing-the-threat-of-cyberattacks/.

[3] "Group01_Report," 2018. [Online]. Available: https://wiki.smu.edu.sg/1718t3isss608/Group01_Report.

[4] "what-is-a-honeypot-a-trap-for-catching-hackers-in-the-act.," 2019. [Online]. Available: https://www.csoonline.com/article/3384702/what-is-a-honeypot-a-trap-for-catching-hackers-in-the-act.html.

[5] "Cybercriminals attack cloud server honeypot in 52 seconds," 2019. [Online]. Available: https://www.cio.com/article/3515424/cybercriminals-attack-cloud-server-honeypot-in-52-seconds.html.

[6] "AWS Honeypot Attack Data," 2018. [Online]. Available: https://www.kaggle.com/casimian2000/aws-honeypot-attack-data/kernels.