AWS HONEYPOT ATTACK DATA

Team Members:
Satya sai Jayanth Devineni G01206829
Yeshwanth Reddy Bommu G01197092
Kuldip Gadapa G01239022

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].

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].

RESEARCH QUESTIONS

1. What are the factors that cause the cybercriminals and how they attack data?
2. How can a honeypot be capable of diverting the hackers from true data?
3. How the unauthorized activity is captured, and cybercrime associated spoofs are blocked?
4. Which locations are mostly prone to cyber attacks and what do they aim for?
5. The most common timings of attack and IP addresses that are significant to attack?

6. How the AWS team controls and monitors the cyber attacks and how would they improve their system security and stabilities accordingly.

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, localeabbr, 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].

AWS_Honeypot_marx-geo

datetime	host	src	proto	type	spt	dpt	srcstr	cc	country	locale	localeabbr	postalcode	latitude	longitude
3/3/13 21:53	groucho-oregon	1032051418	TCP		6000	1433	61.131.218.218	CN	China	Jiangxi Sheng	36		28.55	115.9333
3/3/13 21:57	groucho-oregon	1347834426	UDP		5270	5060	80.86.82.58	DE	Germany				51	9
3/3/13 21:58	groucho-oregon	2947856490	TCP		2489	1080	175.180.184.106	TW	Taiwan	Taipei			25.0392	121.525
3/3/13 21:58	groucho-us-east	841842716	UDP		43235	1900	50.45.128.28	US	United States	Oregon	OR	97124	45.5848	-122.9117
3/3/13 21:58	groucho-singapore	3587648279	TCP		56577	80	213.215.43.23	FR	France				48.86	2.35
3/3/13 21:58	groucho-tokyo	3323217250	TCP		32628	2323	198.20.69.98	US	United States	Illinois	IL	60661	41.8825	-87.6441
3/3/13 21:59	groucho-oregon	3730416887	TCP		6000	1433	222.89.164.247	CN	China	Henan Sheng	41		34.6836	113.5325
3/3/13 22:07	groucho-singapore	3738622573	TCP		6000	3306	222.214.218.109	CN	China	Sichuan Sheng	51		30.6667	104.0667
3/3/13 22:12	groucho-oregon	3683919430	TCP		6000	1433	219.148.38.70	CN	China	Hebei	13		39.8897	115.275
3/3/13 22:14	groucho-singapore	1007884304	TCP		6000	1433	60.19.24.16	CN	China	Liaoning	21		41.7922	123.4328
3/3/13 22:14	groucho-tokyo	3639889826	TCP		6000	1433	216.244.79.162	US	United States	Washington	WA	98168	47.4891	-122.2908
3/3/13 22:20	groucho-oregon	1965603898	TCP		9907	1433	117.40.188.58	CN	China	Jiangxi Sheng	36		28.55	115.9333
3/3/13 22:20	groucho-sa	3672981807	TCP		33367	22	218.237.65.47	KR	South Korea	Seoul	11		37.4906	127.02
3/3/13 22:20	zeppo-norcal	3672981807	TCP		33367	22	218.237.65.47	KR	South Korea	Seoul	11		37.4906	127.02

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.

TIMEPLAN

The Time plan of the project proposal has tasks and implementation parts. It shows particular tasks to get completed in the desired time and it has required amount of data is processed in CSV file first, implementing code and methods, Analyzing the performance, Preparation of Project Report, Preparation of Project Presentation, Final Project Report Submission. The project may take up to 5 weeks.

References

[1] "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/.

[2] "Singapore Management University," 2018. [Online]. Available: https://wiki.smu.edu.sg/1718t3isss608/Group01_Report.

[3] J. Fruhlinger, "CSO," 01 April 2019. [Online]. Available: https://www.csoonline.com/article/3384702/what-is-a-honeypot-a-trap-for-catching-hackers-in-the-act.html.

[4] "CIO," 10 April 2019. [Online]. Available: https://www.cio.com/article/3515424/cybercriminals-attack-cloud-server-honeypot-in-52-seconds.html.

[5] "Kaggle," [Online]. Available: https://www.kaggle.com/casimian2000/aws-honeypot-attack-data/kernels.