**Analysing Attacker Behaviour Through Honeypot**

**CAPSTONE PROJECT PHASE-1**

**Phase – I Report**

***Submitted by***

**1. 20BCE10083 Arpit Subhash**

**2. 20BCE10290 Manu Yadav**

**3. 20BCE10456 Yash Shukla**

**4. 20BCE10558 Karan Gyani**

**5. 20BCE10831 Sunil Kumar Nishad**

***in partial fulfillment of the requirements for the degree of***

***Bachelor of Engineering and Technology***

****

**VIT Bhopal University**

**Bhopal**

**Madhya Pradesh**

**December, 2023**

****

**Bonafide Certificate**

Certified that this project report titled **“Analysing Attacker Behaviour Through Honeypot”** is the bonafide work of “20BCE10083 Arpit Subhash, 20BCE10290 Manu Yadav, 20BCE10456 Yash Shukla, 20BCE10558 Karan Gyani, 20BCE10831 Sunil Kumar Nishad**”** who carried out the project work under my supervision.

This project report (DSN4095-Capstone Project Phase-I) is submitted for the Project Viva-Voce examination held on 23 December 2023.

**Supervisor**

Dr. G.R. Hemalakshmi

Senior Assistant Professor

School of Computing Science and Engineering

**TABLE OF CONTENTS**

**Sl. No. Topic Page No.**

1. Introduction 4
2. Objective 6
3. Literature review 7
4. Proposed Work 8

4.1. Data

4.2. Data cleaning and processing

4.3 Analysis

5. Result 10

6. Individual contribution 14

7. Conclusion 16

8. References 17

# 1. INTRODUCTION

A honeypot is a computer security mechanism set to detect, deflect, counteract attempts at unauthorized use of information systems. A honeypot consists of data attempts at unauthorized use of information systems. A honeypot consists of data that appears to be a legitimate part of the site, but is actually isolated and monitored, and that seems to contain information of value to attackers, who are then blocked[1].

Lance Spitzner defines honeypots as “an information system resource whose value lies in unauthorized or illicit use of that resource”[2]. Honeypots are a very useful tool for learning about tools, procedures, targets, and methods of attackers.

1.1 Classification of honeypots:

The first classification is based on the **role of honeypot.** According to this classification, honeypots are divided in **server-side honeypots** and **client-side honeypots.** Server-side honeypots are useful in detecting new exploits, collecting malware, and enriching research of the threat analysis (e.g. Conpot[3]). On the other hand, honeypots for client-side attacks are called client-side (e.g. Thug[4]).The prime motive of client-side honeypots is to identify and detect malicious activities across the Internet [5].

The second classification is based on the **level of interaction**. The **low-interaction honeypots** detect attackers using software emulation of characteristics of a particular operating system and network services on the host operating system. Advantage of this approach is in a better control over attacker’s activities, since attacker is limited to software running on a host operating system.

On the other hand, disadvantageous about this approach is the fact that the low- interaction honeypot emulates service, or couple of services, but it does not emulate complete operating system. Examples are HoneyD[7], Dionaea[6]. Honeypots that offer attackers more ability to interact than do the low- interaction honeypots, but less functionality than high-interaction solutions, are called **medium-interaction honeypots**. They can expect certain activity and are designed to give certain responses beyond what a low-interaction honeypot would give. Example is Kippo[8]. In order to get more information about attackers, their methods and attacks, it uses a complete operating system with all services. This type of honeypot is called **high-interaction honeypot**. Main aim of this type of honeypot is to provide the attacker access to a real operating system. Examples of this type of honeypots are HonSSH[10].

1.2 Honeynet:

The concept of honeypots is extended by honeynets. Honeynet can be defined as “a highly controlled network of honeypots” [11]. To successfully deploy a honeynet, honeynet architecture that define honeynet architecture:

Data capture: monitors and logs all activities of attacker within the honeynet.

Data control: purpose of which is to control and contain the activity of attacker.

Data collection: all data are captured and stored in one central location.

1.3 Data analysis:

Data analysis is an ability of honeynet to analyse the date, which is being collected from it [14,15]. Data analysis is used for “understanding, analysing, and tracking the captured probes, attacks or some other malicious activities”[1].

# 2. OBJECTIVE

The "Analyzing Attacker Behavior Through Honeypot" project is designed to deepen our understanding of cyber attacker behaviors, methodologies, and tactics by deploying and monitoring honeypots, which are systems designed to attract and interact with malicious users. By analyzing the attacks on these decoy systems, the project aims to uncover the strategies, techniques, and patterns employed by cybercriminals, thereby contributing to the enhancement of cybersecurity defenses.

The project seeks to collect comprehensive attack data from the honeypots to identify common attack vectors, understand the sequence of attacker actions, and discern behavioral patterns. This analysis will not only reveal the most exploited vulnerabilities but also inform the development of more effective security measures and strategies for anticipating and mitigating future attacks. The insights gained will be instrumental in refining existing security protocols, designing robust network architectures, and guiding the development of new defense mechanisms.

Furthermore, "Analyzing Attacker Behavior Through Honeypot" aims to contribute to the cybersecurity community by providing empirical data that can refine threat detection and response mechanisms. The project's findings are expected to offer valuable intelligence for improving security systems, influencing policy making, and guiding future cybersecurity research and education. Ultimately, this initiative is about bolstering the overall security posture by transforming the way we understand, predict, and respond to cyber threats, making it a significant step towards a more secure and resilient digital environment.

# 3. LITERATURE REVIEW

It focuses on two areas, namely the data collection and data analysis in honeypots and honeynets. Research in area of the data analysis has resulted in a number of papers, discussing specific topics, concerning design of data analysis and analysis of specific aspect of attack.

Kaur[12] proposes data analysis module in the third generation of honeypots. This module is based on data access method: a relational model based fast path. Sharms[13] focuses on honeywell and how it is used in honeynet environment to trap attackers. One of the modules of honeywell is data analysis module.

Another interesting paper[14] focuses on design and implementation of virtual honeynet, Its data analysis can be divide into three levels, such as swatch software, Hflow2, and walleye.

Project Modern Honey Network focuses on deploying and managing honeypots and analysis ata freom honeypots. It allows easy analysis of data from low-interaction and medium-interaction honeypots. Unforutnately, this framework doesn’t focus on high interaction honeypots and issues of secure connection between honeypots and itself.

Thonnard[15] proposes the application for analysing one spcific aspect of the honeynet data i.e. the time series of the attacks. The result is identification of the activites of several worms and botnets in the traffic according to time series analysis.

# 4. PROPOSED WORK

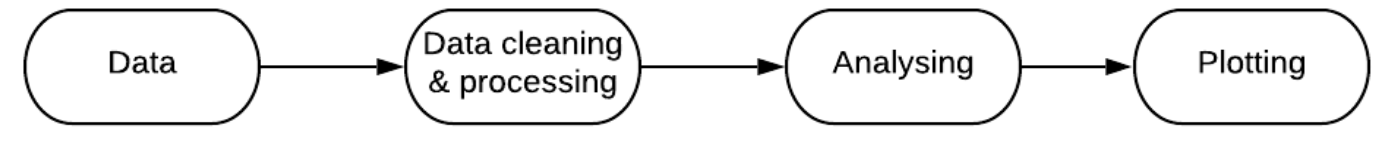
The Work flow that will be carried forth in this project is as shown below in Table.

Figure 1- Flow chart graph

**4.1.Data**

Honeypot data-set used, is taken from AWS Honeypot Database. It is an open-source database including information on cyber-attacks and attempts. Dataset has 451,581 data points collected from 9:53pm on 3 March to 5:55am on 8 September 2013. The data has a total of 15 columns and have been displayed in table 1 below.

The first columns contain date, time. The rest of the columns display specific properties of attacker and host such as country of origin of attack, IP address, latitude, longitude and protocol etc.

Table 1 – Description of data-set

The total number of records and columns have been shown in Table 2.

Table 2. Number of records and columns

**4.2. Data cleaning and preprocessing**

1. There are missing, since values that need to be removed because longitude and latitude are missing, since these are crucial features for the geolocation on world map using geopandas.

2. There are some faulty values of latitudes and longitudes, as latitude cannot be greater than 90 degrees and cannot be less than -90 degrees. So that need to be dropped from data-set, similarly for longitudes.

3. There are columns that are not relevant or redundant to project, so that need to be dropped to get better output.

Example –“cc”, “src”, “unnamed:15”.

4. Preprocess datetime using pandas ‘to\_datetime()’.

**4.3. Analysis**

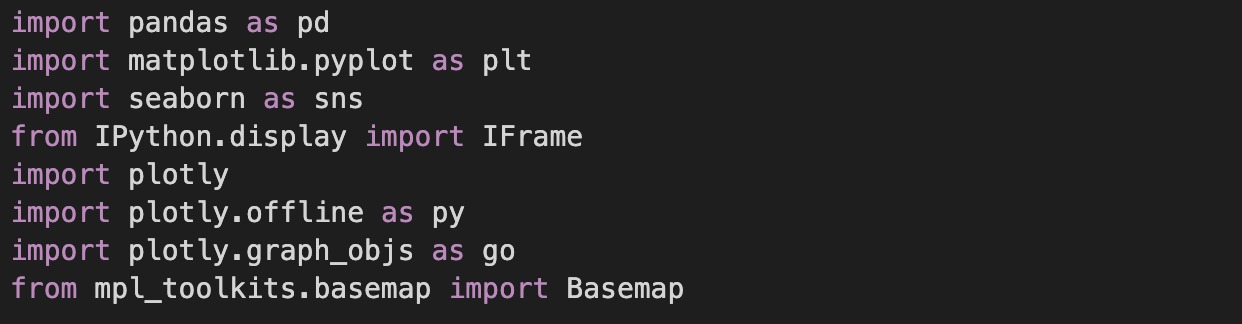
3.1 Geolocation of attacks

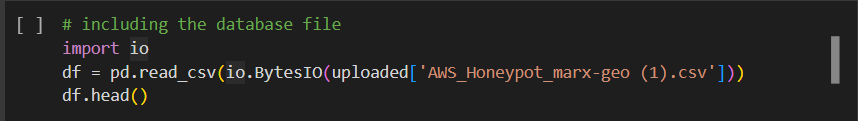
3.2 Top 10 countries from where attack originate.

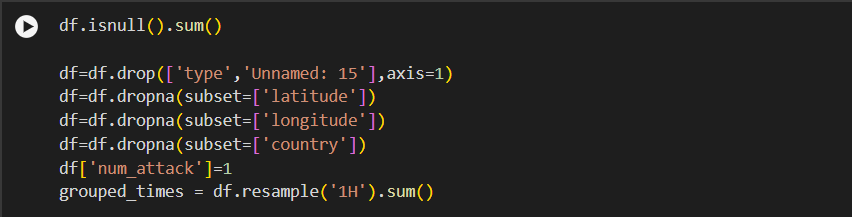
3.3 Top 10 host attacked in world.

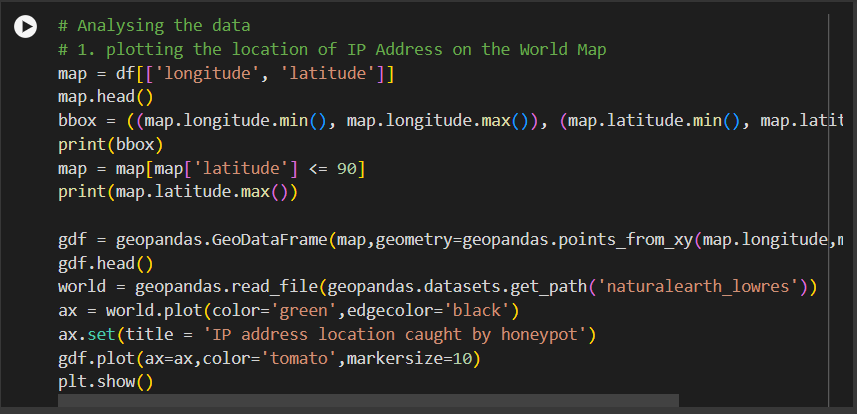
3.4 Frequency of attack in an hour span.

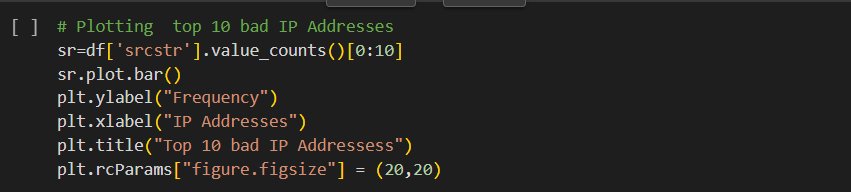
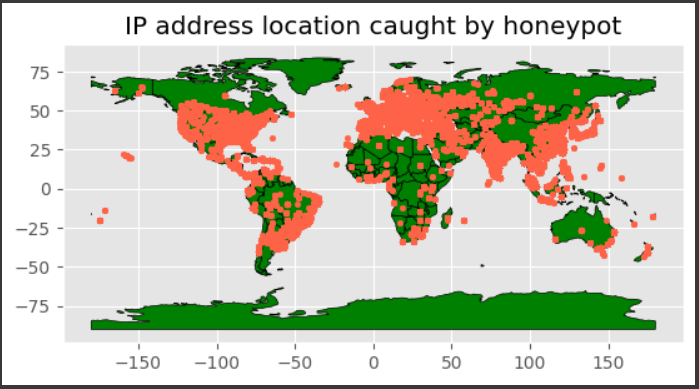
# 5. Results

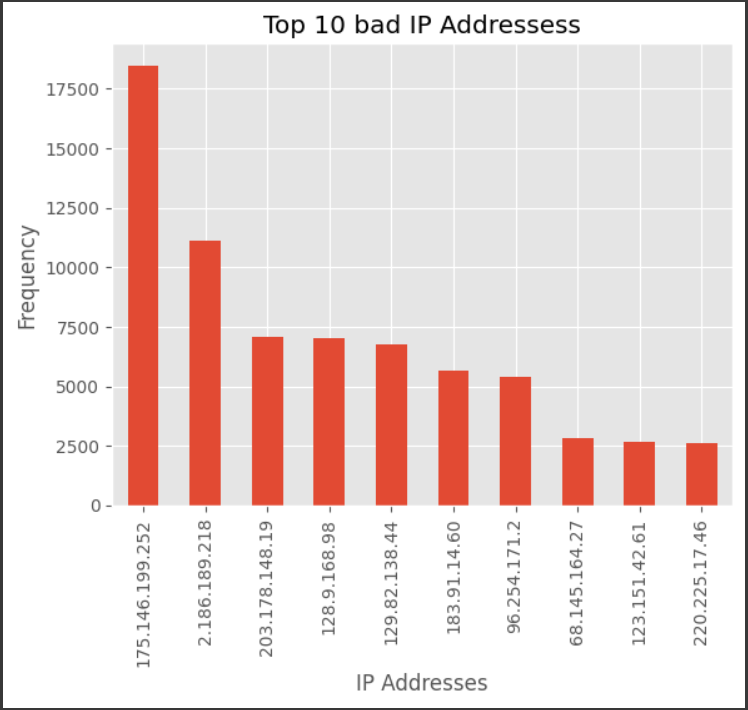


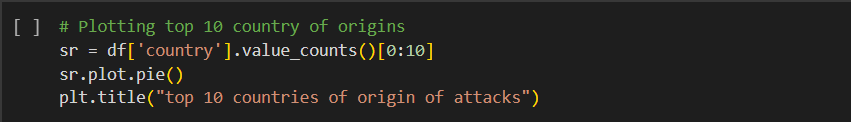


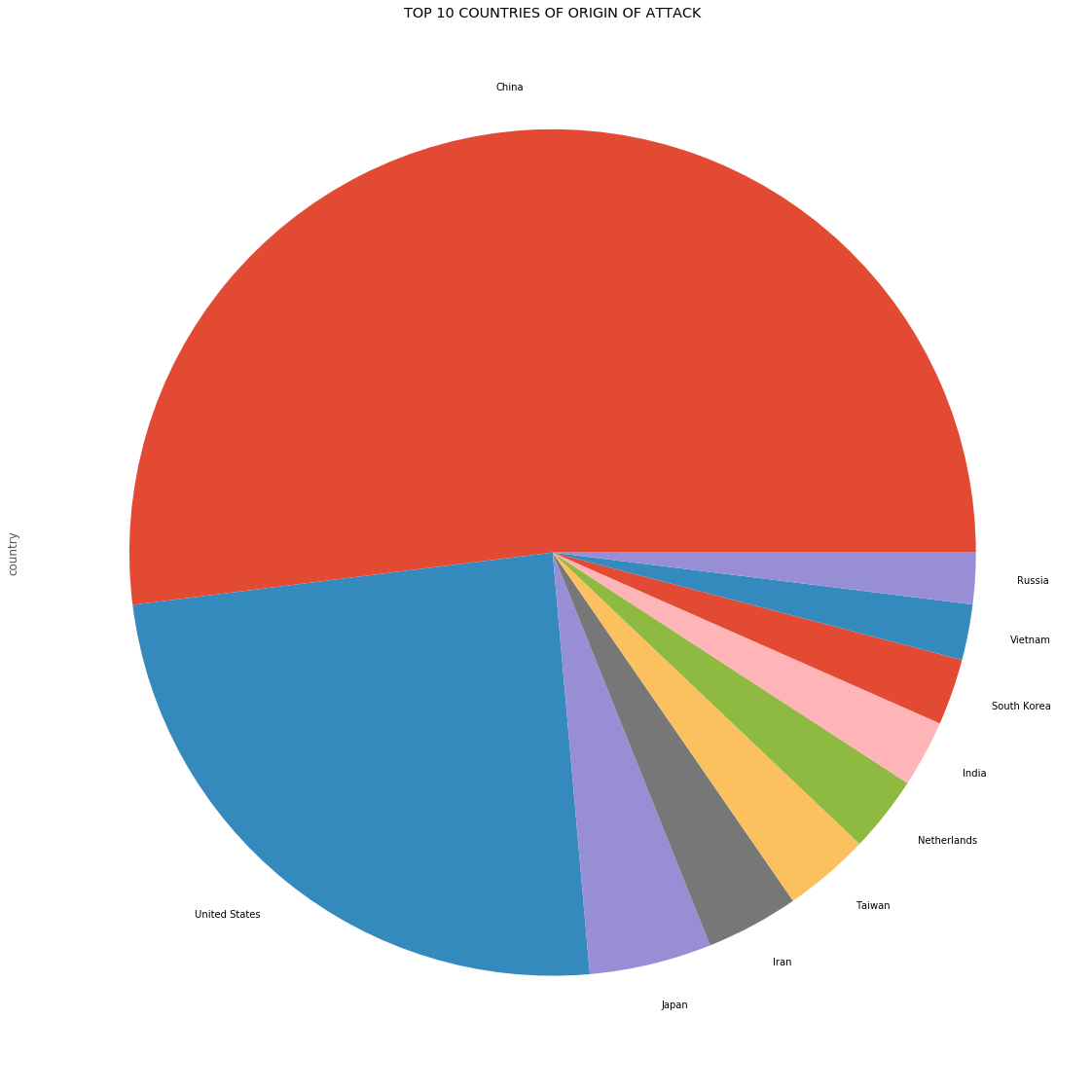


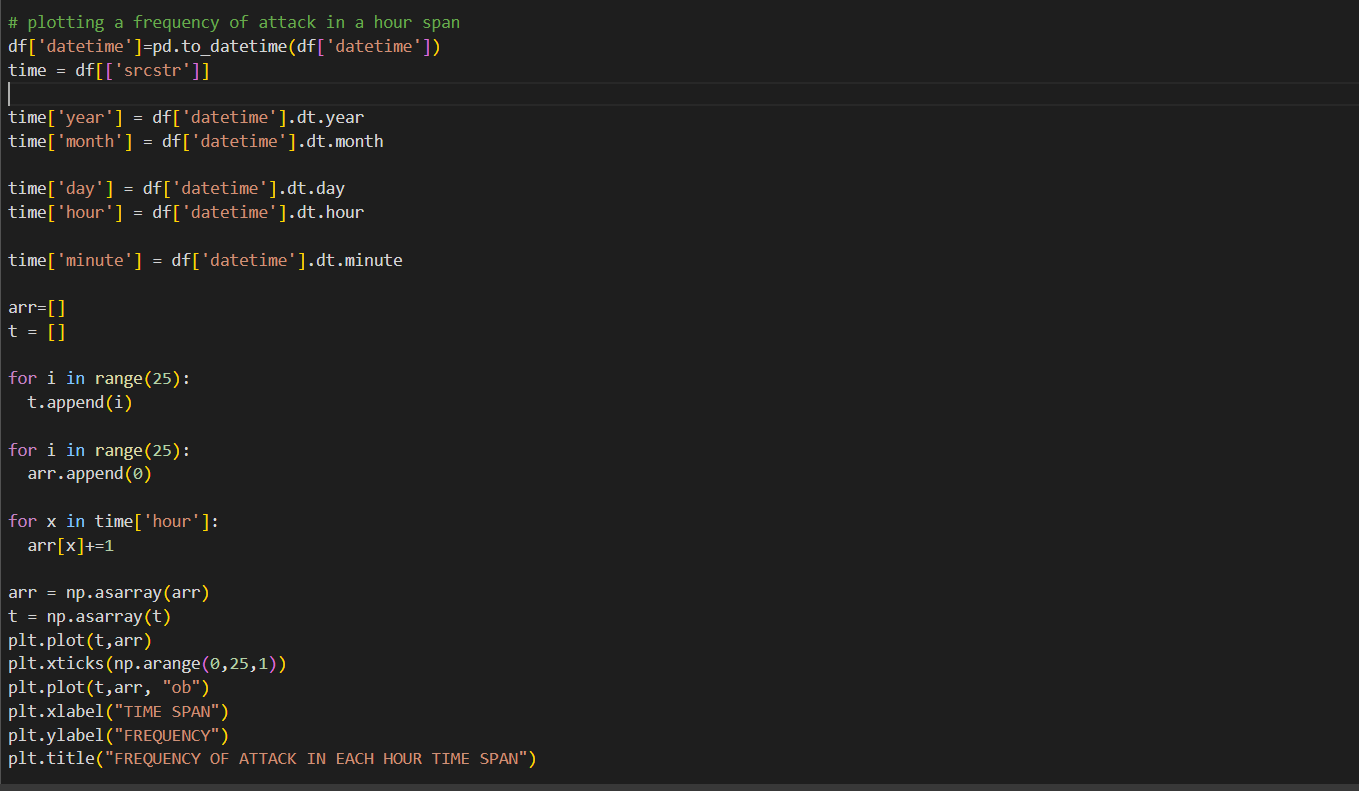


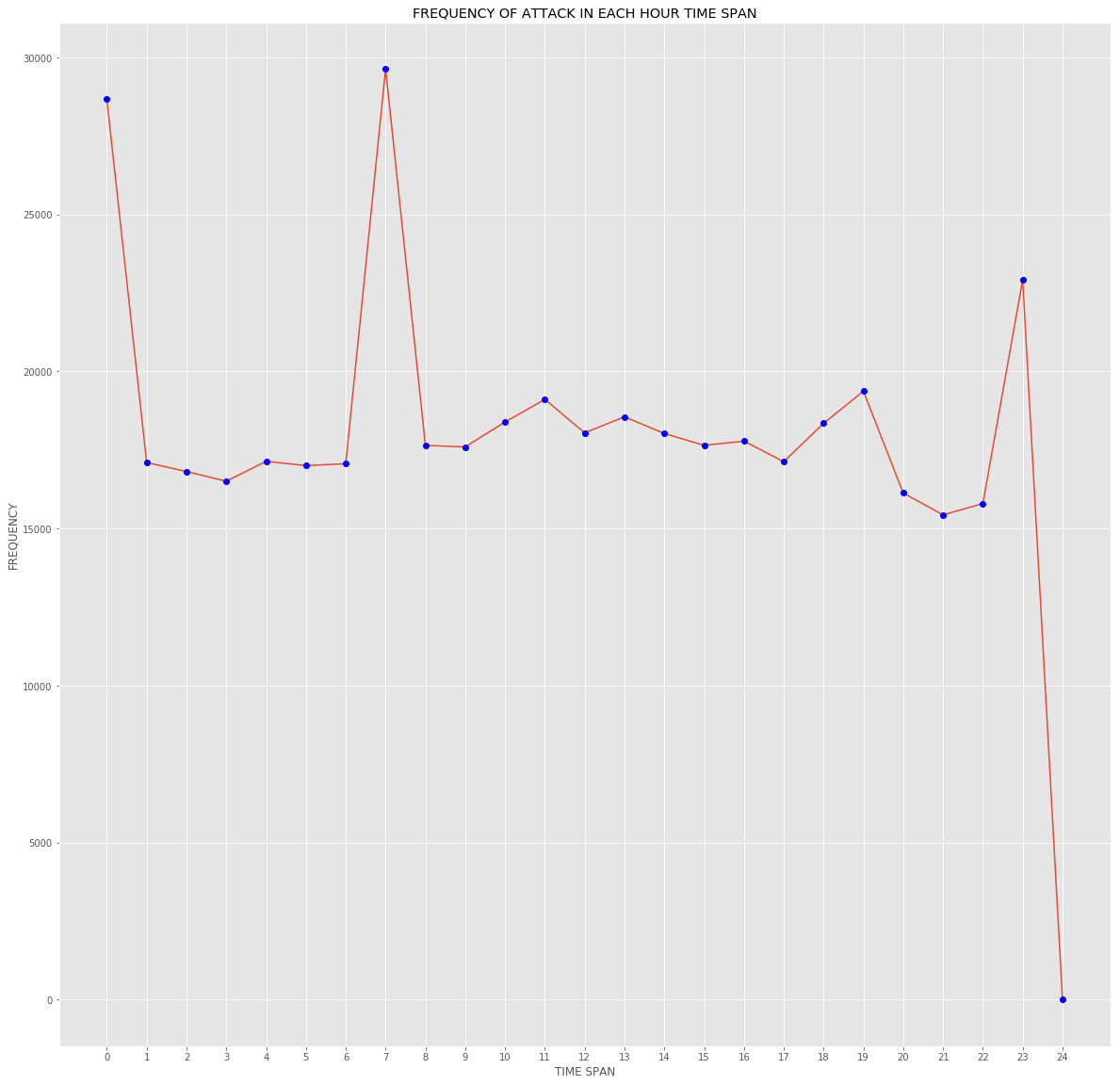


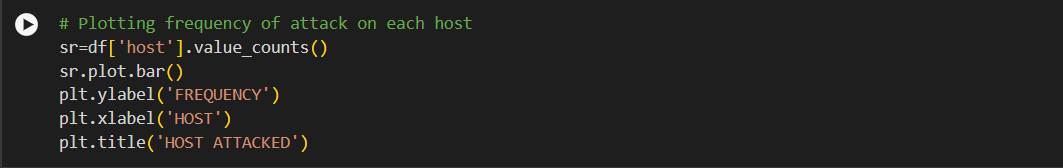


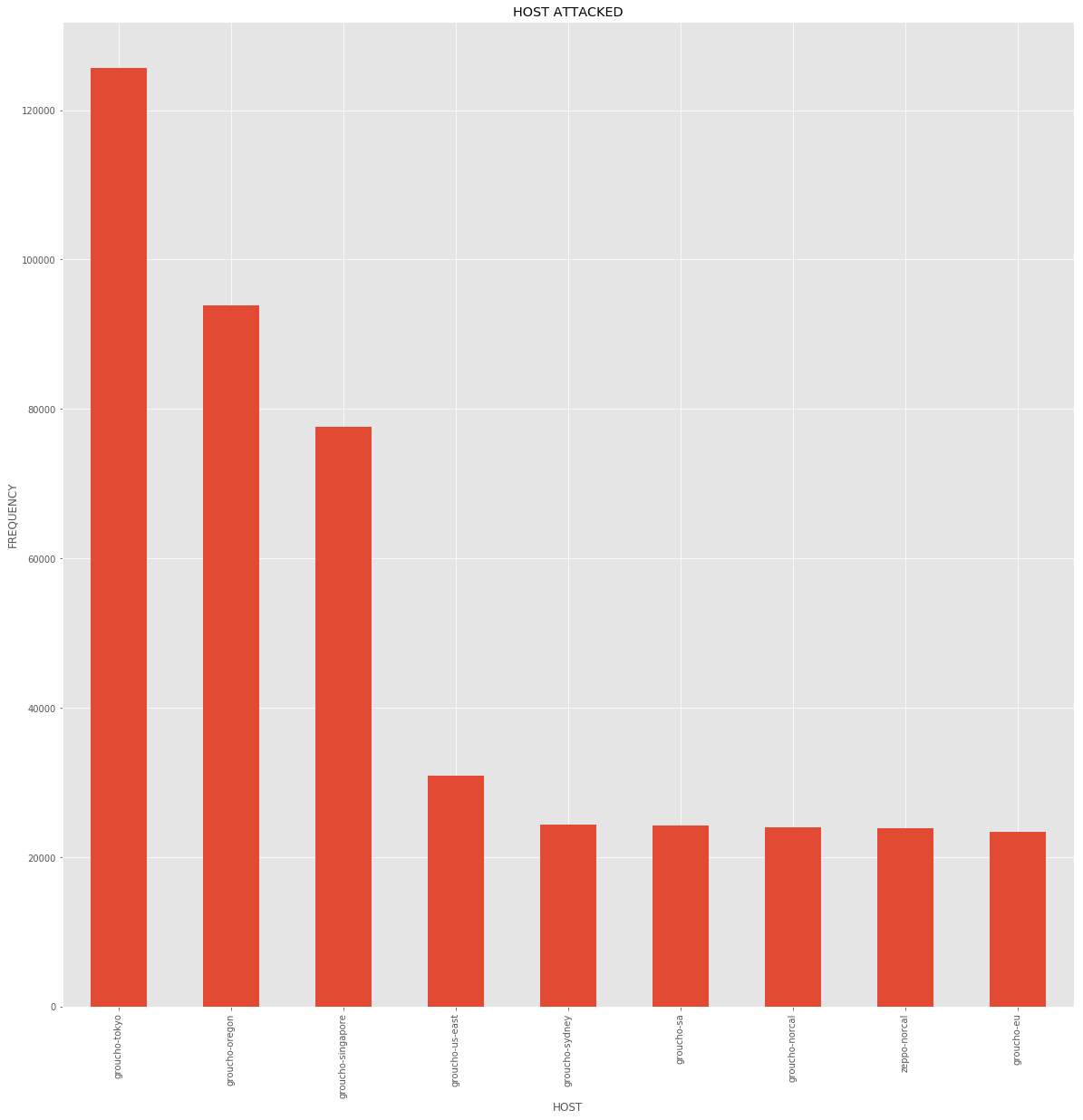












# 6. Individual Contribution

**Arpit Subhash (20BCE10083) contribution:**

Arpit's contribution has been invaluable, focusing on the application of statistical concepts to dissect the dataset comprehensively. His proficiency in statistical analysis has illuminated patterns and crucial information, significantly enhancing the project's depth. Arpit's role extended beyond analysis; he adeptly transformed complex data into understandable visual representations, making the results accessible and impactful for all team members. Moreover, his collaborative spirit shines as he continues to work closely with teammates, exploring and implementing various machine learning techniques to uncover deeper insights and predictive patterns. His forward-thinking approach aims not just to analyze the current data but to leverage it for predictive analytics, contributing to a more innovative and informed project outcome. Arpit's dedication to continuous learning and application of advanced analytical techniques is pivotal in driving the project towards cutting-edge research and application.

**Manu Yadav (20BCE10290) contribution:**

Manu's major contribution lies in the critical analysis of the dataset, where his keen analytical skills came to the forefront. He delved deep into the data, uncovering key insights that are vital for the project's direction and success. Manu's ability to identify patterns, anomalies, and trends has provided a substantial understanding of the underlying dynamics of the data. His collaborative approach was evident as he actively engaged in discussions with teammates, sharing findings and brainstorming on their implications. This not only fostered a dynamic team environment but also ensured that these insights were incorporated effectively into the project's strategy. Furthermore, his dedication to rigorously validating and cross-referencing results has helped maintain the project's integrity and credibility, making his role indispensable to achieving comprehensive and reliable outcomes.

**Yash Shukla (20BCE10456) contribution:**

Yash’s contribution to the project has been multifaceted, involving an in-depth analysis of honeypots and honeynets, their classification, functionality, and models. His expertise has been instrumental in dissecting the architecture and deployment strategies, significantly enriching the project’s content. Additionally, Yash has played a crucial role in enhancing the project’s presentation, contributing to the PowerPoint creation with engaging content that effectively communicates complex information. His insights into ethical and legal considerations, along with a commitment to current cybersecurity developments, have greatly enriched the project’s depth and relevance.

**Karan Gyani (20BCE10558) contribution:**

Karan's role in the project has been central to data management and preparation. He undertook the critical task of understanding the dataset intricacies and meticulously processed it by removing redundant columns and handling null values, ensuring data integrity and usability. Karan's keen attention to detail was further demonstrated as he converted data types, like changing INT64 to date-time objects, thereby streamlining the data for more efficient analysis. His efforts have significantly reduced complexities for subsequent analysis by team members, setting a solid foundation for insightful data exploration. Moreover, Karan's proactive approach in preparing and cleaning the data has accelerated the project's progress, allowing for a more focused and effective analysis phase. His contributions have not only enhanced the quality of the data but also facilitated a smoother and more collaborative working environment for the team.

**Sunil kumar Nishad (20BCE10831) contribution:**

Sunil has been pivotal in laying the foundational research for the project by diligently collecting previously published papers and gathering critical references, ensuring a robust knowledge base and facilitating a smoother workflow for the team. His meticulous approach to sourcing relevant literature has provided valuable historical and contemporary insights, enhancing the project's analytical depth. Additionally, Sunil has been actively involved in the analysis process, applying his analytical acumen to interpret data and trends, which has been crucial in identifying patterns and formulating strategies. His contribution has significantly boosted the project's integrity and has been instrumental in driving the project towards innovative and informed outcomes.

# 7. CONCLUSION

It discusses different types of honeypots such as production honeypots. It also looked at factors that should be considered while implementing a honeypot. For example, the level of interaction of your honeypot depends on what you want to use it for. The legal issues surrounding honeypots and their implementation were examined, and throughout it mentioned the advantages of honeypots.

Honeypots are a relatively new technology that is becoming increasingly popular, and will become even more so as commercial solutions become increasingly popular, and will become available that are easy to use and administer. The honeypots and honeynets are unconventional tools. Their main purpose is to collect data and subsequently analyse them to learn information about attackers and their methods, tools and objectives.

Therefore, it proposes as framework for analysis of data collected from honeypots and honeynets. Because they can be used to collect information on attackers and investigations.

# 8. REFERENCE

[ 1 ] Spitzner, L.: The honeynet project: Trapping the hackers, in *IEEE Security &*

*Privacy, 1(2)*, pp. 15-23, 2003.

[ 2 ] Spitzner, L.: Honeypots: Tracking Hackers, in *Addison Wesley*, pp. 1-430,

2002.

[ 3 ] The Conpot Project. 2015: http://conpot.org/

[ 4 ] Thug. 2015. http://buffer.github.io/thug/

[ 5 ] Biswas, J.: Analysis of Client Honeypots, in *International Journal of*

Computer Science & Information Technologies, 5(4), 2014.

[ 6 ] Dionaea. 2015: http://dionaea.carnivore.it/

[ 7 ] HoneyD, 2015: http://www.honeyd.org/

[ 8 ] Kippo, 2015: https://github.com/desaster/kippo

[ 9 ] Joshi, R. C., Sardana, A. (Eds.): Honeypots: A New Paradigm to

Information Security, in *CRC Press*, 2011.

[ 10 ] HonSSH, 2015: https://github.com/tnich/honssh/

[ 11 ] Abbasi, F., Harris, R.: Experiences with a Generation III virtual Honeynet,

in Telecommunication Networks and Applications Conference (ATNAC),

2009 Australasian, 2009.

[ 12 ] Kaur, M. M., Singh, D. Attack Methodology Analysis By Using Honeypot

Technology.

[ 13 ] Sharma, N., Sran, S. S.: Detection of threats in Honeynet using Honeywall,

in International Journal on Computer Science and Engineering, 3(10),

3332-3336, 2011.

[ 14 ] Shi-wei, Y., Xiu-shuang, M., & Wei-dong, W. A. N. G.: Core Functions

Analysis and Example Deployment of Virtual Honeynet, in *Computer*

*Science*, 3, 021, 2012.

[ 19 ] Thonnard, O., Dacier, M.: A framework for attack patterns' discovery in

honeynet data, in Digital investigation, 5, pp. 128-S139, 2008.