Int. J. Nonlinear Anal. Appl.
Volume 12, Special Issue, Winter and Spring 2021, 1117-1132
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/IJNAA.2021.5573

An experimental study on cloud honeypot and data visualization using ELK stack

Fakariah Hani Mohd Ali[a], Muhammad Fadhli Mohd Salleh[b,*], Nurul Huda Nik Zulkipli[c]

[a]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Shah Alam, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia.*
[b]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Shah Alam, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia.*
[c]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Melaka Kampus Jasin, 77300 Merlimau Melaka, Malaysia.*

(Communicated by Madjid Eshaghi Gordji)

Abstract

Nowadays, companies have been moving their IT infrastructure from own data centers to specialized public cloud providers. While there are cost benefits, the security issue is one of the major concerns in cloud computing due to the number of companies that use cloud storage to save their personal data keep increasing. Many honeypots have been used in the past, but they were difficult to use due to a lack of data visualization and attack analysis. To learn more about attackers, their motivations and techniques, honeypots are used to investigate how attackers attempt to hack an information system and provide useful insight into potential security flaws. This honeypot allows to monitor attacks by pretending to be actual machines with valuable and sensitive data, such that attackers interact with them. For this research, honeypot was set up on DigitalOcean cloud and the experimental method performs and result of the implementation in this research use real attack since the honeypot deployed on the cloud and exposed to the Internet. The results show that Cowrie honeypot able to collect data that is valuable to security researcher or network administrator for future research to make analysis. It is believed by implementing Cowrie honeypot using ELK stack on cloud platform will assist on detection and prevention for SSH attacks.

Keywords: Honeypot, Cowrie, SSH attacks, ELK Stack, Cloud Computing

[*] Corresponding author

Email addresses: fakariah_hani@uitm.edu.my (Fakariah Hani Mohd Ali), muhdfadhlisalleh@gmail.com (Muhammad Fadhli Mohd Salleh), nurulhuda8450@uitm.edu.my (Nurul Huda Nik Zulkipli)

Received: June 2021 *Accepted:* September 2021

1. Introduction

DigitalOcean is a cloud-based infrastructure provider. It has a simple set-up and is very cost-effective. It enables developers to complete tasks such as spinning up a server which is droplet, in a fraction of the time required by other platforms. The ELK Stack consists of Elasticsearch, Logstash, and Kibana, which is a collection of three open-source applications. The ELK stack provides centralized logging for detecting server or application faults. It gives us the ability to search through all the logs in one spot. It also assists in the discovery of issues across different servers by connecting logs together over a set period of time. This tool falls under the Big Data category since it meets the three primary criteria for this definition: volume, variety, and speed [7].

Honeypot is defined as a mechanism of security that serve as bait or virtual trap to lure attackers. It is a system that will be placed in the real network system to work with other security system to improve the security of the network. Moreover, it is set up to be a decoy that are strategically placed in the network and pretend to be a server or the real system. The attacker might think that the honeypot is the real system and attempt to gain access to the system, but the attacker just entered a faked environment that exposed the information about them. A honeypot is a closely monitored computing resource that wishes to be probed, attacked, or compromised, and a honeypot is an information system resource whose value is derived from the unauthorised or illegal use of that resource [12].

Every new technology brings with its new issues. Virtualization, autonomous computing, grid computing, and a variety of other technologies have all contributed to cloud computing [3]. A major challenge is ensuring that the cloud is secure enough to work as intended. However, there are many threats on cloud computing, such as data breaches, Insider threat, Denial of service attacks, Cloud malware injection and many more. The Spectre and Meltdown attacks, which first surfaced in early 2020, allow attackers to obtain information from the kernel by breaking the barrier between programmes and the operating system [4]. To determine their attack patterns and track hacker's behaviors, honeypot is needed for better analyzing and understanding in order to create more secure systems. A honeypot is a tool used to gather evidence or information to learn as much as possible about attack patterns, hacker motivations, and regularly utilized programmed launched by them [3].

Existing Honeypots come in many shapes and sizes, but lack of data visualization. Although there are many honeypots were used for detection, we are still unable fully secure the computer system and prevent the network from attackers if there are no analysis and action is made. Analysis is difficult to perform without a good data visualization, making it difficult to take any action to prevent new attacks. Data visualization such as dashboard, gives a clear information about the attack on the honeypot, identifies attack patterns, better understanding on new attacks, and better analysis by providing a visual context through maps or graphs. Therefore, the main objective of this research are mainly focus on the experimental study by designing and setting up SSH Cloud honeypot called Cowrie honeypot. After that, a series of analysis on attacks detection and attacks patterns from the honeypot will be carried out and be visualized through dashboard using ELK Stack from different cloud providers located in different country. Besides, data obtained from Cloud honeypot will be recorded all the time. It will be useful for analyzing for a better interpretation of the attacker motive in the system. ELK Stack provides a good analysis of data by showing the data into graph, table, and maps.

2. The concept of honeypot

A honeypot is a computer system that is specifically designed to attract and ”trap” persons who attempt to hack into other computers [5]. It is set up to be a decoy in the network and pretend to be a real system. It lures the attackers and wastes their time if they try to gain unauthorized access to the network. [15] mentions that the honeypot's aim is to detect and learn from attacks, and then utilise that information to improve security. Honeypot also can be used to log malicious activities and learn new threats. [15] stated that it can be used to observe activities of an individual which gained access to the honeypot. Figure 1 shows the basic architecture of honeypot. Honeypot can be categorized into two, which are research honeypot and production honeypot.

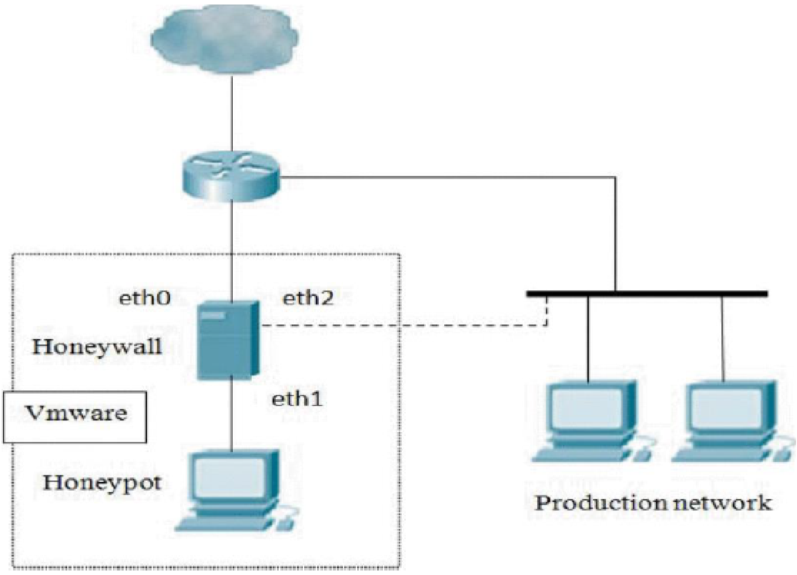


Figure 1: Basic architecture of Honeypot From [16]

2.1. Research honeypot vs production honeypot

Research honeypots is a tool to gain information about the hacker's or attacker's community. Their purpose is to identify new threats and learn more about the hacker's motivations and technique. This honeypot makes easier to gain information and be able to see in action what the attacker's intentions and methods they are using. By collecting all the information, security expert can prevent it and block all the attack. A honeypot provides real-time insight into how an attack occurred. Research honeypots are difficult to create and manage, but they are used to collect high amounts of information [15]. [15] stated that they are often utilised by organisations interested in learning more about risks research, such as colleges, governments, the military, or huge enterprises. Meanwhile, Production honeypot is used to prevent potential attacks and implemented inside the production network. They mostly used within an organization and help to mitigate risk like internet threat. They function as extension to Intrusion Detection Systems (IDS) that accomplish an enhanced role of detection [15]. Production honeypot captures a limited amount of information as same as low interaction honeypots. According to [15], A honeypot identifies threats that conventional security systems do not detect, and an IDS requires a database with often updated signatures of known attacks. Production honeypots can improve security for an organization into several steps, which are prevention detection and reaction.

2.2. Levels of interactions in honeypot

Honeypot can be categorized into three levels which are low, medium and high-level interaction. This classification is based on how much an attacker will be able to interact with the honeypots.

2.2.1. Low-Interaction honeypot

The most easily implemented are the low-interaction honeypots. For low-interaction honeypots, basic services such as Telnet and FTP are emulated. Host operating systems are not vulnerable to attack, honeypots with low interaction are fairly safe to run, but cannot be used where a more complex and interactive environment is required, such as an SMTP server [5]. This low-level interaction honeypot only can use one or more simple services that record all communication attempts to a specific service, such as a web or SSH server and provide the simplest passive method of monitoring attack attempts [5]. It does not have any real time monitoring capabilities for this type of honeypot. They are commonly used to collect information and cannot be used in the network to completely track malicious activities. It is easy to install and manage these types of honeypots at a low cost. For statistical purposes, the data obtained from a low-interaction honeypot can be used and various automated attack patterns around the network can be calculated.

2.2.2. Medium-Interaction honeypot

Medium interaction honeypots are more complex and complicated than low interaction honeypots. Since it is middle level, it can handle botnet detection and malware collection. However, a medium level of interaction is just enough for being attractive for an attacker to want to break the system. Even though emulated programs should react in the same way as actual programs, it can be very difficult to emulation a set of software since the same security problems should not be present. If not, the emulation is aborted [5]. A medium engagement level is adequate to keep the attackers interested and allows the administrator the power to select which services to imitate in order to better understand how they can be abused.

2.2.3. High-Interaction honeypot

High-interaction honeypots are more advanced compared to low and medium interaction honeypot since it is a real computer system, having real services and operating systems. The aim of an interactive honeypot is to decide what the intruder is doing after accessing root from the device [5]. These honeypots are hard to configure, manage and maintain, and it is the most risk one when the attacker can gain real access. This level of honeypot requires continuous monitoring, so the attacker can eventually monitor it and will use the attack as a starting point for other attacks [5]. Either the attacker is obsessed with the honeypot, and all of their interactions and keystrokes are captured, resulting in high-quality data about their intentions, or the attacker exploits the fact that they can communicate with the real system to get access, compromising the production network.

2.3. Attacks against SSH protocol

The SSH protocol, also called Secure Shell, is a method for securing remote login from one device to another. It offers numerous different ways for high authentication and maintains communication confidentiality and integrity with strong encryption. The protocol is used in an organization to provide some services such as protected access for users and automated processes, interactive and automatic transfer of files, performance of remote commands, management of network infrastructure and other mission-critical system elements. Furthermore, the SSH protocol is open and well developed, and there are several software libraries that allow the development of SSH clients. Secure Shell refers to port 22 and implements a protocol for login and password authentication, while other more secure methods, such as public key authentication, can be used instead [9]. Nonetheless, SSH protocol attacks have become too common, as an attacker may quickly guess or brute-force the right credential provided by an authorised system user or administrator. There are few attacks and have been found such as brute-force and dictionary attacks, SSH port scanning and SSH penetration attacks.

2.3.1. The brute-force and dictionary attacks

The concept of a brute force attempt is simple when attacker may try every possible value until authentication has been achieved [8]. One of the most effective ways to obtain SSH access to servers is by brute-forcing credential. Brute force means that attacker may be try all possible character of password. By using dictionaries, attacker will use it for large lists of common passwords. By default, most SSH servers would have a limit to the number of authentication attempts that can be attempted per connection, but as with many things involving connectivity, the attacker can bypass them [9]. The attacker performs brute force or so-called dictionary attacks if excessive levels of login attempts are detected.

2.3.2. The SSH port scanner

Port scanner may help attacker to find a weak point to break into the system and can scan to find vulnerabilities. By listening on port 22, the attacker could try to connect and try to break the system by using weak passwords. A port scan is a mechanism in which all ports at an IP address are tested to see if they are open or closed. On the other hand, scan port using Nmap command is one of the methods to determine the state of the port that SSH is running on. Port scan is the first step for going to penetration attacks and do not give an access to a system to the hackers. Same goes to Banner Grabbing that allows an attacker to discover network hosts and running services on the open ports. It is the best technique to gain information about the open ports by simply querying the service port.

2.3.3. The SSH penetration attacks

Malicious activities such as injecting a virus program, installing a bot running a DDoS attack, distributing spam, backdoor, or testing a newly created malware. Such types of operations are referred to as penetration attacks. This makes SSH authorized keys can be stolen and harm the system.

Metasploit is one of the methods for SSH Port Scanner and exploit's tool used for searching system weaknesses. Metasploit also is a malicious software that provides services such as payloads with sets of malicious code, encoders for convert code, post-exploitation module that gather information and gain further access to the target system.

2.4. The usage of ELK stack and cowrie honeypot

The term of stack can be described as a grouping of software products to function together like the famous LAMP stack consisting of Linux, Apache, MySQL and PHP.. The ELK stack is an acronym used to describe a stack that comprises of three popular researchs: Elasticsearch, Logstash, and Kibana. The stack consists of multiple open source software in order to build a system monitoring. Cowrie honeypot is one of latest honeypot that will focus on brute-force attack and ELK Stack is needed to display all the data from cowrie honeypot.

2.4.1. The ELK stack

The ELK (Elasticsearch, Logstash, and Kibana) stack is a full-stack system that enables actionable insights in real time from practically any form of organised and unstructured data source. Elasticsearch is a distributed, open source search and analytics engine that can handle textual, numerical, geographical, structured, and unstructured data. Logstash is a server-side data pipeline that collects data from various sources at the same time, processes it, and transmits it to Elasticsearch. Kibana is an Elasticsearch data visualization and management application that also serves as the user interface for monitoring, managing, and protecting an Elastic Stack cluster. ElasticSearch expands to massive volumes as well, in order to obtain sensor data from industrial devices, machines, and other IoT (Internet of Things) device output and scale the system by adding clusters and nodes [14]. ELK is a popular stack for managing log files and gathering them as JSON files. It is very flexible, scalable, and customizable. Besides, it also be able to be used in a simple or complex manner because it benefits both basic and complicated operations [14]. Geoip is another plugin for Logstash that converts IP addresses to longitude and latitude, and searches its database for city and county names.

2.4.2. Cowrie honeypot

Cowrie is an SSH and Telnet honeypot with a medium engagement level that is used to record brute force attacks and SSH requests. Cowrie makes use of a Python code base that is maintained and publicly accessible on GitHub. Cowrie is a honeypot that simulates SSH and Telnet services with medium to high interactivity in order to record brute force attacks and shell operations performed by hackers. Despite its usefulness in protecting servers, honeypot cowrie has a key disadvantage in that it still presents data in the form of a log system, which is regarded inefficient for network administrators to undertake log monitoring [1]. These honeypots are frequently linked to the Internet in order to monitor the tools, scripts, and hosts used by password guessing attackers. Cowrie, like other honeypots, will log or analyse attacks performed against it while deceiving the attacker into thinking they are in a server. A honeypot records evidence of attacks, and the system can provide data on the number of attacks that occur each month [11]. This enables the honeypot administrator to obtain an understanding of the types of attacks being tried, their overall success or failure rate, and the geographical location of the IP address from which a given attack originates. Cowrie also distributed honeypots, which were used to log SSH interactions in a MySQL database [6].

3. Proposed design

As mention before, this research aims to designing and setting up SSH Cloud honeypot and visualize the results using the ELK Stack. This proposed design is about Cloud honeypot and Data Visualization to prevent SSH attacks. There are three important parts which are detection, defense and report of attacks as shown in Figure 2 as improvement of the proposed design. This design will focus on the main three parts as mentioned and will be discussed in section.

3.1. Detection parts

The first part is detection. To mitigate SSH attacks, SSH detection is the most important part because the attacks are extremely easy to execute and difficult to detect. Detection is needed for detecting any malicious activity in the system. The inner detection design as shown in Figure 3 Detection phase will focus on these attacks: Brute force & dictionary attack, SSH port scan and SSH penetration.

In process of detection, it is used to detect SSH attack such as brute-force attacks and dictionary attacks, SSH port scan and SSH penetration attacks. The process of detection as follows:

I. The proposed design will check the username and password after user or attacker insert username and password. The system will determine who is the legitimate user before access the system. If attackers attempt to login the system more than five times, the system will send an alert to email and provide some information about the attacker. SSH encrypts communication between two hosts and also uses public-key cryptography for user authentication. [4].

If not the legitimate user, the system will specify the type of SSH attacks, whether it is Brute-force attacks and dictionary attacks, SSH port scan or SSH penetration attacks based on the behavior of attacks as shown in Table 1.

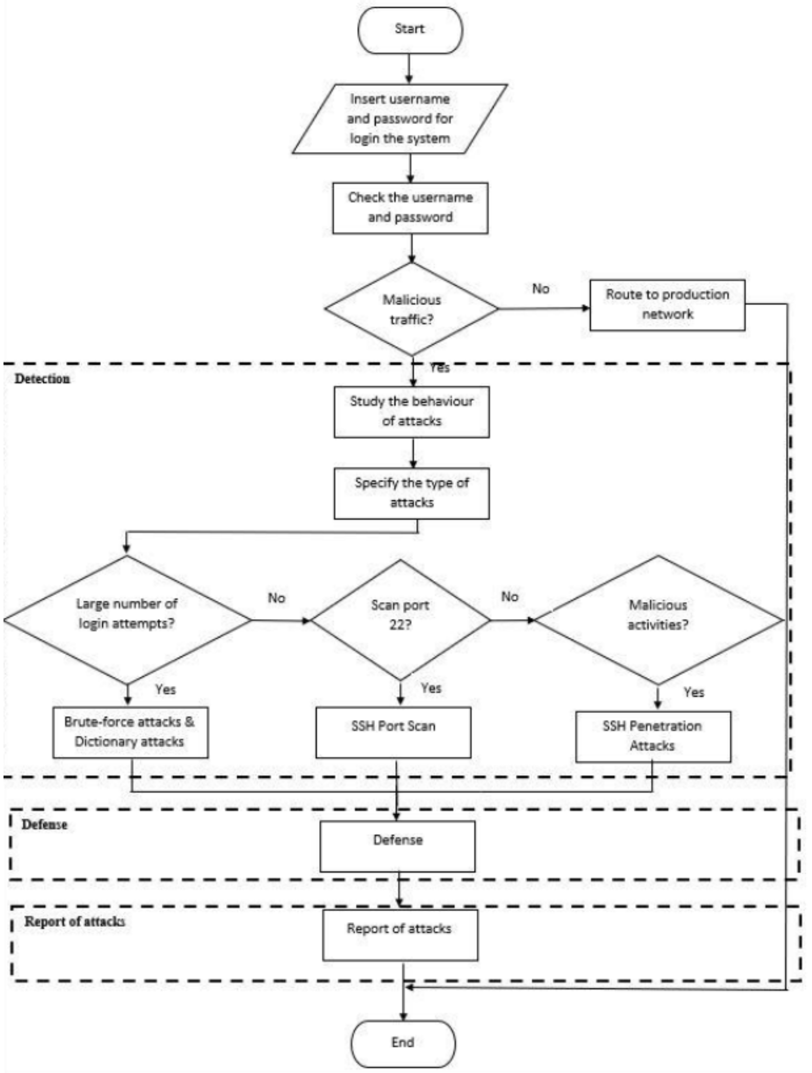

Start
Insert username and password for login the system
Check the username and password
Malicious traffic?
No
Route to production network
Yes
Detection
Study the behaviour of attacks
Specify the type of attacks
Large number of login attempts?
No
Scan port 22?
No
Malicious activities?
Yes
Yes
Yes
Brute-force attacks & Dictionary attacks
SSH Port Scan
SSH Penetration Attacks
Defense
Defense
Report of attacks
Report of attacks
End

Figure 2: Proposed design

Table 1: Behavior of three SSH attacks

No.	Behavior of Attack	Types of SSH Attacks		
		Brute Force & Dictionary Attacks	SSH Port Scanner	SSH Penetration Attack
1.	Number of login attempts	Large number of login attempts	Large number of login attempts	• One time of login attempts • Large number of login attempts
2.	Number of TCP flow packets	High number of TCP flow packets	Small number of TCP flow packets	Random number of TCP flow packets
3.	Source of attack	• Real IP address • Spoofed IP address	• Real IP address • Spoofed IP address	• Real IP address • Spoofed IP address

Data Mining Hierarchical Clustering Method [16] is another detection method. Given a set of ‘n’ objects to cluster and a distance matrix of ‘nn’. The following are the steps in the Hierarchical

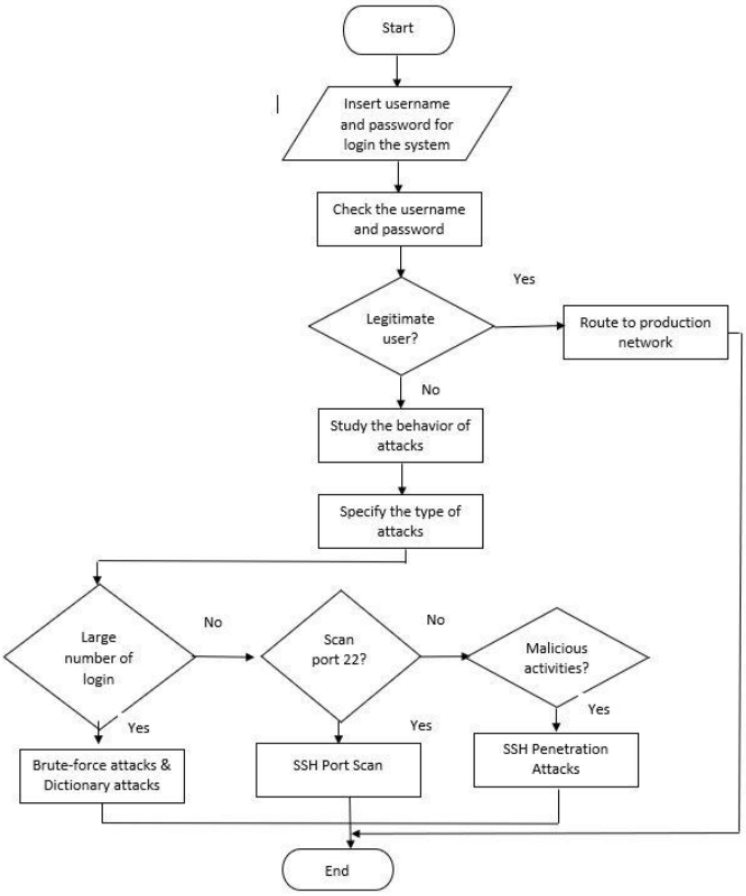

Start
Insert username and password for login the system
Check the username and password
Legitimate user?
Yes
Route to production network
No
Study the behavior of attacks
Specify the type of attacks
Large number of login
No
Scan port 22?
No
Malicious activities?
Yes
Brute-force attacks & Dictionary attacks
Yes
SSH Port Scan
Yes
SSH Penetration Attacks
End

Figure 3: The inner design on detection

Clustering Method [10]:

Assign each item to its own cluster, such that if we have 'n' items currently, then we will have 'n' clusters, each with only single item. Besides, allow the similarities the clusters to match the similarities between the items contained within them.

1. Find the clusters that are the most similar to one another and combine them into a single cluster.

2. Measure the similarities between the new clusters with old clusters.

3. After that, we can repeat step 2 and step 3 until all items are clustered into the single cluster with size 'n'.

Step 3 can be done in a variety of ways, including single-link clustering, complete-link clustering, and average-link clustering. Single link clustering's purpose is to find the shortest distance between any data point in one cluster and any data point in another. The diameter or maximum technique (also known as complete-link clustering) aims to find the shortest distance between any two data points.

To evaluate the effectiveness of this method are as follows [17]:

A true positive (TP) detection is one that has been correctly classified as malware. The better the outcome, the higher the real positive. When a detection is wrongly reported as benign, it is called a false negative (FN). A benign application that has been adequately determined as benign

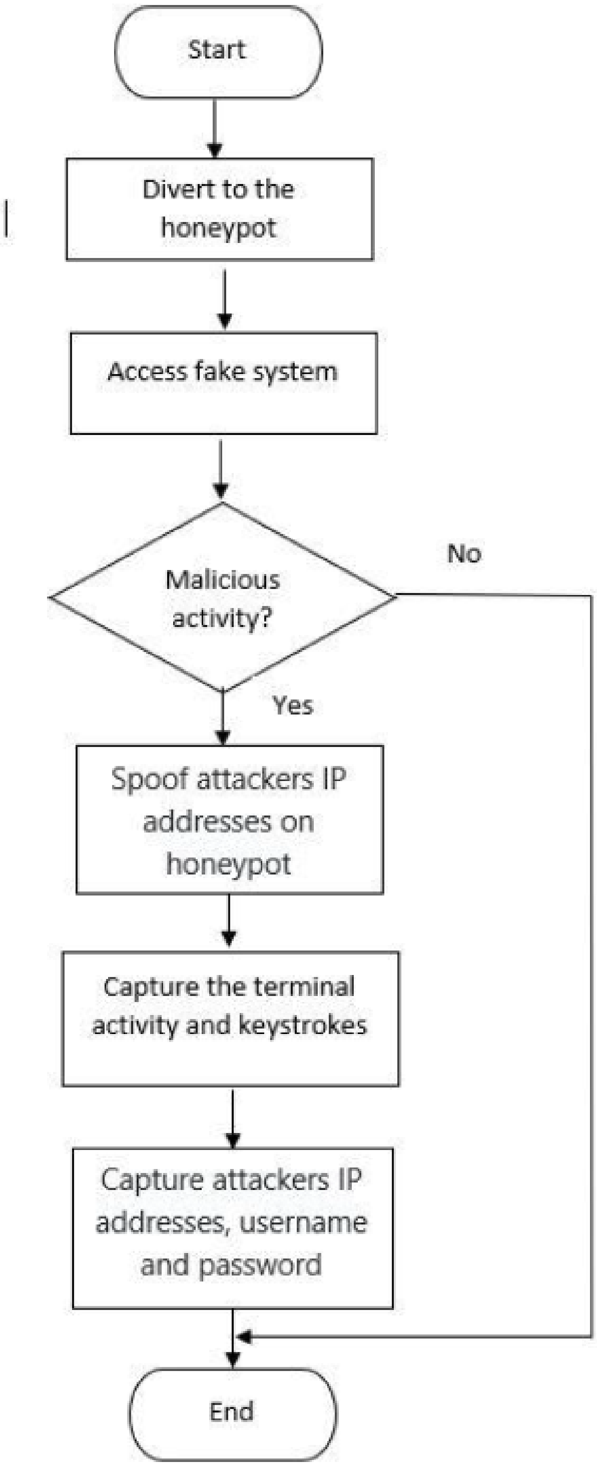

Start
Divert to the honeypot
Access fake system
Malicious activity?
No
Yes
Spoof attackers IP addresses on honeypot
Capture the terminal activity and keystrokes
Capture attackers IP addresses, username and password
End

Figure 4: The Inner design on Defense

is referred to as a true negative (TN). A false positive (FP) is a harmless programme that has been incorrectly classified as malware. The ratio of benign files incorrectly classified as dangerous to all benign files in the set is known as the false positive rate, whereas the false negative rate is the ratio of malware files correctly classified to all malware files in the testing set.

1) True Positive Rate (TPR)

$$TPR = \frac{TP}{TP + FN} \tag{1}$$

2) False Positive Rate (FPR)

$$FPR = \frac{FP}{FP + TN} \tag{2}$$

3) Overall accuracy will be calculated as the proportion of the total number of correct predictions divided by the total number of forecasts.

$$Precisin = \frac{TP}{TP + FP} \tag{3}$$

4) Recall:

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

The detection results will be shown and explained in the next paper.

3.2. Defense parts

Secondly, are the Defense parts. Defense is one of the most important parts, where it is used to protect and block Brute-force attacks & Dictionary attacks, SSH Port Scan and SSH Penetration attacks before it harms the network. The proposed defense design as shown in Figure 4. The attacker will be diverted to the honeypot system and moved to fake systems to interact with honeypot. If Cowrie honeypot detects any malicious activity, the system will spoof attackers IP addresses and moved them to fake system. In the honeypot, all terminal activity such as commands and keystrokes will be recorded as a log files until the attacker log out from the system. By using Cowrie honeypot features, all interaction is captured into a log files and their sessions can be viewed in real time. Other than that, the honeypot will capture their interaction details such as IP address, username, and password.

3.3. Reports of attacks

Finally, is the reporting medium of the attacks. The report of attacks is used to record the type of SSH attacks has been detected. The process of the report of attacks as shown in Figure 5. The report attacks will have the following details of attacks:

a. Type of SSH attacks - There are three type of SSH attacks: Brute-force attacks & Dictionary attacks, SSH port scan and SSH penetration attacks.

b. Number of login attempts – The number of failed login attempts, either small or large number of login attempts.

c. Severity level – Risk of attack, either severe attacks or not-so-severe attacks that determine the level of risk to the system.

d. Attackers session – All keystrokes and commands used by attackers that interact with honeypot.

e. Attacker source – The address that the packet was sent from the attacker, either it is using a real IP address or spoofed IP addresses.

f. User credential – The username and password used by attacker to login to the system.

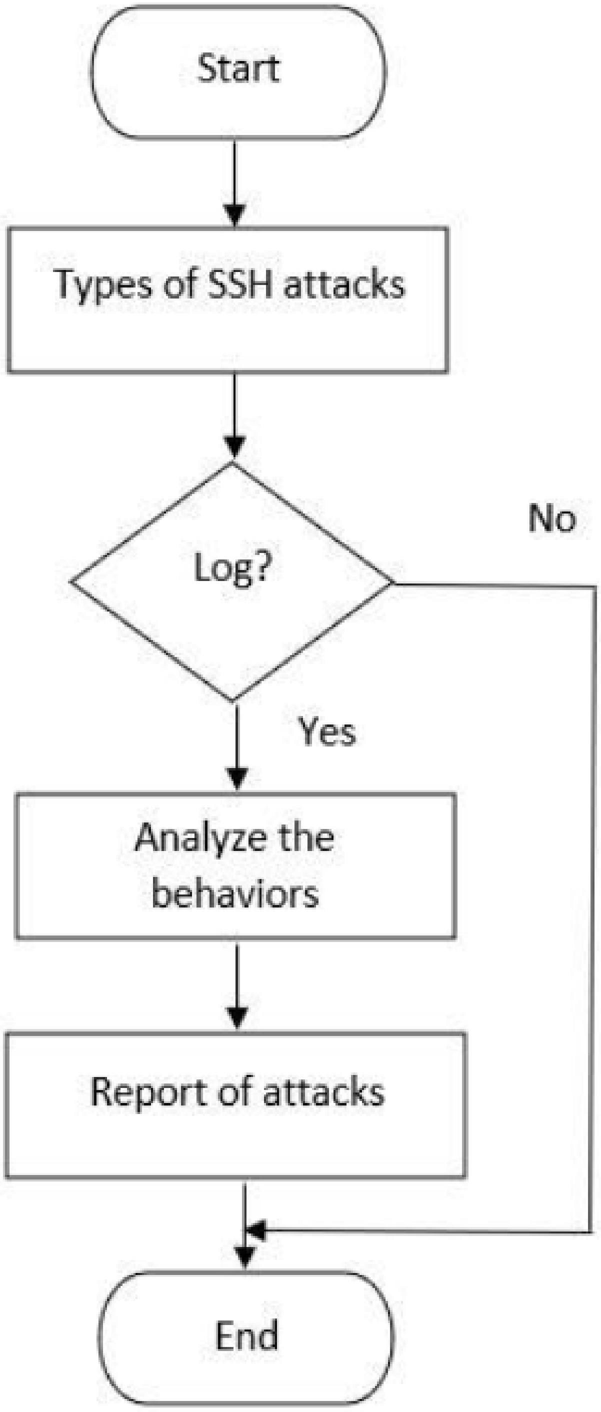


Figure 5: The Inner Process of Reporting the Attacks

4. The experimental setup

This section explained the setup of the experiment. Figure 6 shows the system configuration used to test, detects, prevent attacks and report of attacks. This includes alert network administrator using email and analyse the information of attacker from virtual honeypot. It consists of two virtual machines that running on DigitalOcean Cloud named as vm1 and vm2, but for sensor-my and sensor-cn, will be running on IPServerOne. The approach of this research is to build a medium-interaction honeypot which is implemented on virtual machine with different cloud server location in order to expose to the Internet and get real attack. Vm1 and vm2 are running on cloud server that placed in Singapore, while sensor-my is in Malaysia and sensor-cn is in China.

The virtual honeypot is configured to open SSH port to attract attacker to run port scan on vm1, sensor-my and sensor-cn. The SSH service, which runs on port 22, is one of the functions of the services [13].

Cowrie honeypot was used to trap attacker in a system which records every keystroke, logging every malicious activity and attacker's IP address. Cowrie honeypot will analyse attacker's behaviour and sort the data according to the most attempted password, username, top command input and top IP address by country.

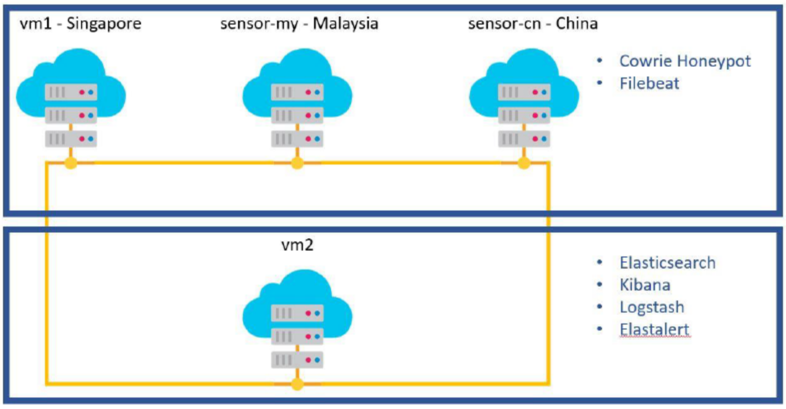


Figure 6: The network diagram of the experiment

The Elastic stack is a collection of three software components from the same developer: ElasticSearch, Logstash, and Kibana. They are made to function together to form a complete log management and visualisation tool. This ELK Stack will be installed on vm2 and Filebeat as an agent to forward the data from honeypot to ELK stack on vm1, sensor-my and, sensor-cn. With different cloud server location, we can create large data visualization and able to determine the best solution for various SSH attacks.

This proposed design will be implemented to detect three types of SSH attacks as mentioned and specify the types of SSH attacks, whether it is Brute force & Dictionary attacks, SSH port scan or SSH Penetration attacks. Cowrie honeypot will analyse the attack and visualize the data on Kibana Dashboard. Alert system such as Elasalert will be used for detecting malicious activity and send an alert through email.

5. Results and discussions

The results were collected in three different locations of honeypot – Malaysia, Singapore, and China. Dashboard will be presented as the result based on the honeypot's server location. The experiment was conducted, and 714,105 individual log entries were successfully collected. The statistics gathered from various honeypot services are divided into three categories: traffic, method and SSH.

A. Traffic

Cowrie honeypot uses the MaxMind GeoLite database to map IP addresses to geolocations that configured as filter on Logstash. There are some differences in general traffic data aimed towards honeypots. By looking at individual countries, China is the most country with higher committed cyber-attacks. From the result, "sensor-my" which is honeypot in Malaysia location stand out as the honeypot with most targeted, higher SSH and Telnet connections.

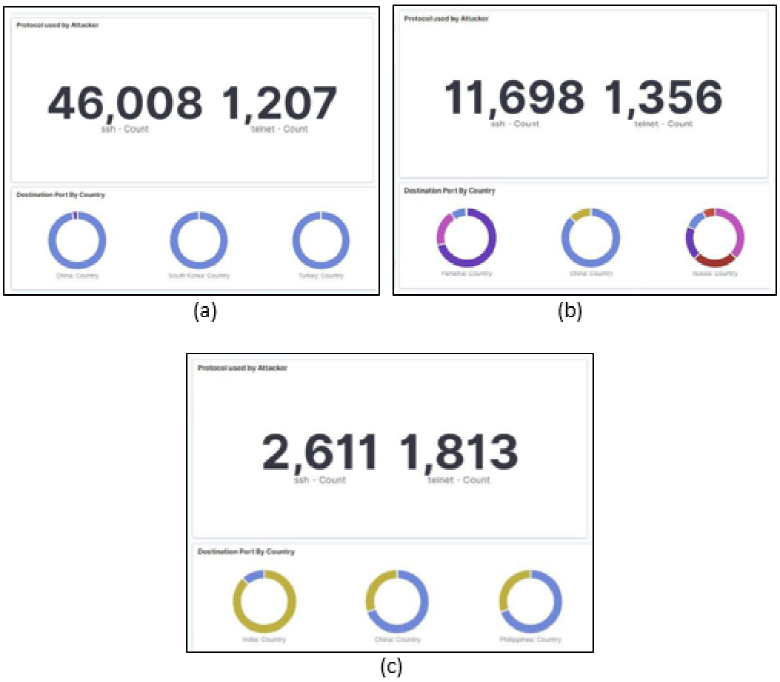


Figure 7: Number of Connection on (a) Malaysia honeypot, (b) Singapore honeypot and (c) China honeypot

B. Target

On every machine, the honeypots expose the same services. Furthermore, the machines have same settings, with the different of the IP address. As a result, it is reasonable to assume that all machines have comparable behaviors and targets. In general, an attacker does not use every port equally, concentrating on the standard ports for SSH, HTTP, and SMTP. Nonetheless, the data shows in figure 8 that non-standard ports that are unassigned or application-specific, which are also targeted.

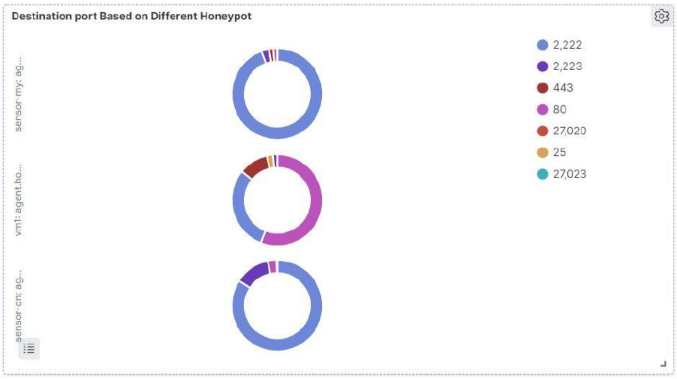


Figure 8: Destination port based on different honeypot's locations

C. Daytime Evaluation

The fundamental idea is to combine incident timestamps and time zones collected through fingerprinting methods. The findings indicate the distribution of attacks throughout the day for any country in local time based on the attack's geographic location.

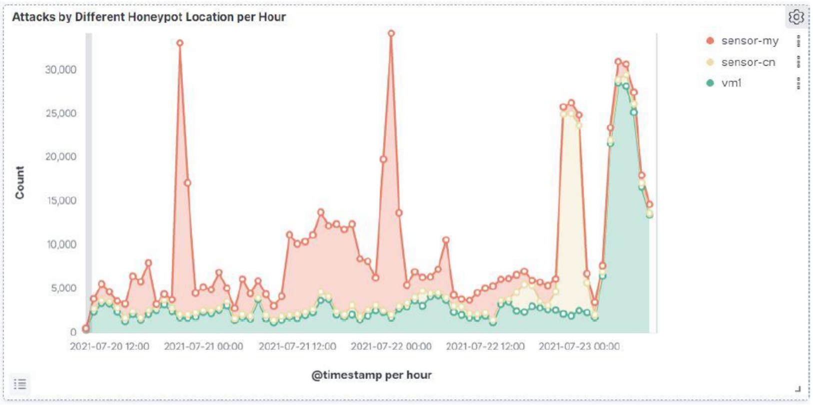
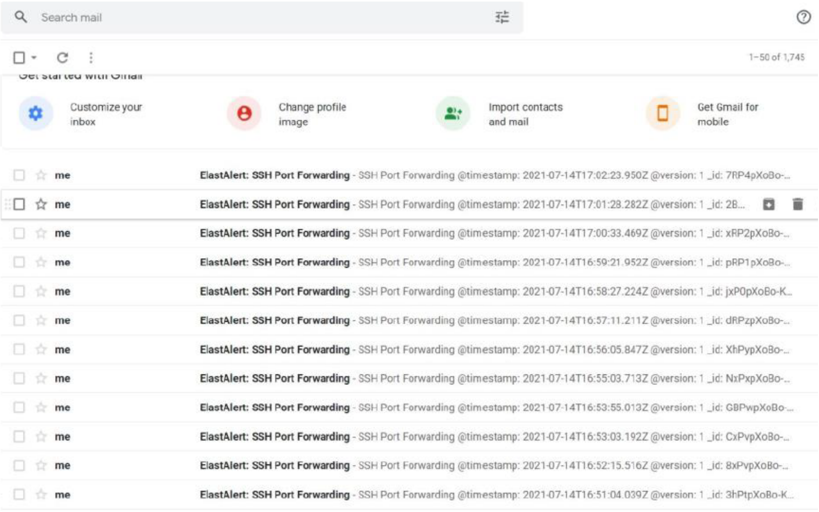


Figure 9: Destination port based on different honeypot's locations

Figure 9 shows that the most attacks happen is between 11:00 PM to 12:00 PM for all honeypots, but Malaysia is different when the rate of attack is increasing between 9:00 PM to 10:00 PM.

D. Alert notification

Figure 10 and Figure 11 shows that alert is successfully functioned and send to the email. There are many rules were set based on various events and attacks. Based on Figure, attacker is trying to delete something from Singapore honeypot (vm1) and Elastalert will send an alert through email to provide some information about attacker.



me ElastAlert: SSH Port Forwarding - SSH Port Forwarding @timestamp: 2021-07-14T17:02:23.950Z @version: 1 _id: 7RP4pXoBo-...

me ElastAlert: SSH Port Forwarding - SSH Port Forwarding @timestamp: 2021-07-14T17:01:28.282Z @version: 1 _id: 2B...

me ElastAlert: SSH Port Forwarding - SSH Port Forwarding @timestamp: 2021-07-14T17:00:33.469Z @version: 1 _id: xRP2pXoBo-...

me ElastAlert: SSH Port Forwarding - SSH Port Forwarding @timestamp: 2021-07-14T16:59:21.952Z @version: 1 _id: pRP1pXoBo-...

me ElastAlert: SSH Port Forwarding - SSH Port Forwarding @timestamp: 2021-07-14T16:58:27.224Z @version: 1 _id: jxP0pXoBo-K...

me ElastAlert: SSH Port Forwarding - SSH Port Forwarding @timestamp: 2021-07-14T16:57:11.211Z @version: 1 _id: dRPzpXoBo-...

me ElastAlert: SSH Port Forwarding - SSH Port Forwarding @timestamp: 2021-07-14T16:56:05.847Z @version: 1 _id: XhPypXoBo-...

me ElastAlert: SSH Port Forwarding - SSH Port Forwarding @timestamp: 2021-07-14T16:55:03.713Z @version: 1 _id: NxPxpXoBo-...

me ElastAlert: SSH Port Forwarding - SSH Port Forwarding @timestamp: 2021-07-14T16:53:55.013Z @version: 1 _id: GBPwpXoBo-...

me ElastAlert: SSH Port Forwarding - SSH Port Forwarding @timestamp: 2021-07-14T16:53:03.192Z @version: 1 _id: CxPvpXoBo-...

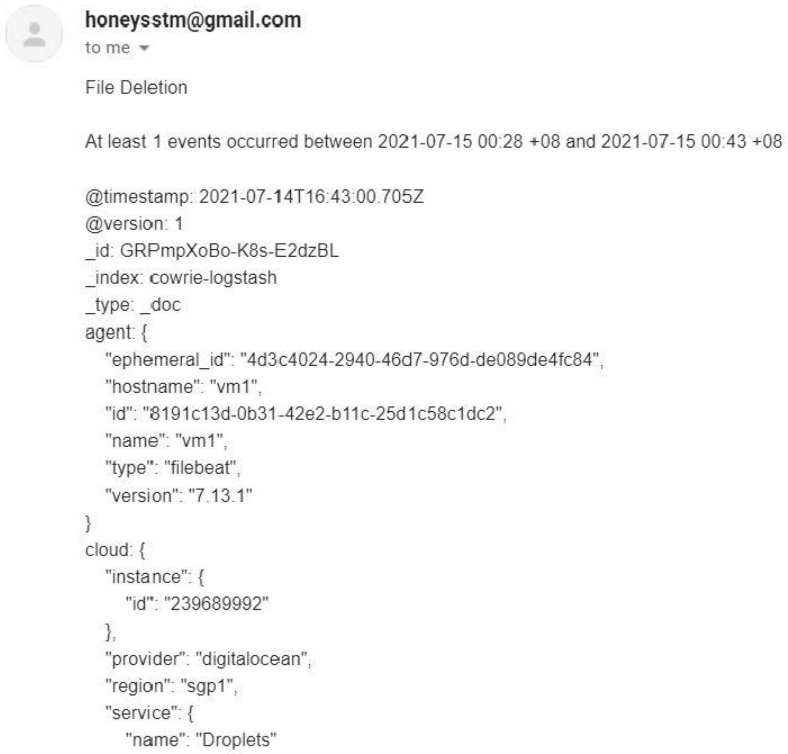
me ElastAlert: SSH Port Forwarding - SSH Port Forwarding @timestamp: 2021-07-14T16:52:15.516Z @version: 1 _id: 8xPvpXoBo-...

me ElastAlert: SSH Port Forwarding - SSH Port Forwarding @timestamp: 2021-07-14T16:51:04.039Z @version: 1 _id: 3hPtpXoBo-K...

Figure 10: Alert notification through email

6. Conclusions

This research is purposely to log malicious activities and learn new threats and as a tool that are meant to be attacked or interacted to provide an environment that can tracts all the activity



honeysstm@gmail.com
to me

File Deletion

At least 1 events occurred between 2021-07-15 00:28 +08 and 2021-07-15 00:43 +08

```
@timestamp: 2021-07-14T16:43:00.705Z
@version: 1
_id: GRPmpXoBo-K8s-E2dzBL
_index: cowrie-logstash
_type: _doc
agent: {
   "ephemeral_id": "4d3c4024-2940-46d7-976d-de089de4fc84",
   "hostname": "vm1",
   "id": "8191c13d-0b31-42e2-b11c-25d1c58c1dc2",
   "name": "vm1",
   "type": "filebeat",
   "version": "7.13.1"
}
cloud: {
   "instance": {
      "id": "239689992"
   },
   "provider": "digitalocean",
   "region": "sgp1",
   "service": {
      "name": "Droplets"
```

Figure 11: Example of file deletion notification

in the network. The research is based on medium interaction honeypot that to detect, defense and provide a report of attacks. The aims of this research are to protect network from SSH attacks and to contribute to an understanding of information security in cloud computing. Finally, the research has succeeded one of its objectives which is to alert the network administrator when detecting malicious activity by sending an email.

The analysis produced a number of significant results. First, a large number of attacks were recorded in a short amount of time, which is surprising given that the honeypot IP addresses were not made public previously. This means that even if an Internet-connected device's network address is not specified, mass scanners will find it in a short time. It is likely that only a few systems per country are targeting cloud providers for weak or exposed systems. All the results of detection and attack patterns in each honeypot can be visualized through dashboard using ELK Stack.

Acknowledgement

This research was supported by Cybersecurity and Digital Forensics Research Interest Group and Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA.

References

[1] D. Afriyantari, P. Putri and A. Rachmawati, *Honeypot cowrie implementation to protect SSH protocol in ubuntu server with visualisation using kippo-graph*, International Journal of Advanced Trends in Computer Science and Engineering, 8 (6) (2019), https://doi.org/10.30534/ijatcse/2019/86862019.

[2] A. Bryk, *Cloud Computing Attacks: A New Vector for Cyber Attacks*, (2020), Retrieved from Apriorit: https://www.apriorit.com/dev-blog/523-cloud-computing-cyber-attacks.

[3] R. B. Buyya, *Cloud computing: Principles and paradigms*, John Wiley & Sons, 87 (2010).

[4] I. D. Cahyani, *Sistem keamanan enkripsi secure shell (ssh) untuk keamanan data*, J. Tek. Elektron. Fak Tek. Uni. Pandanaran, (2011) 1–8.

[5] M. P. Dhruvi Vadaviya and D. M. Abdul Jhummarwala, *Malware detection using honeypot and malware prevention*, International Journal of Computer Engineering and Technology (IJCET) (2019) 1-9.

[6] S. Dowling, M. Schukat and E. Barrett, *Improving adaptive honeypot functionality with efficient reinforcement learning parameters for automated malware*, J. Cyber Secur. Tech., 2 (2) (2018) 75–91.

[7] E. Fontana, *ELK stack — Elasticsearch*, (2020), Retrieved from Betacom: https://medium.com/betacom/elk-stack-elasticsearch-5bfbfebccb7f .

[8] L. M. Harry Doubleday and H. Janicke, *SSH honeypot: Building, deploying and analysis*, International Journal of Advanced Computer Science and Applications(ijacsa), (2016).

[9] G. P. Ioannis Koniaris and P. Nicopolitidis, *Analysis and visualization of SSH attacks using honeypots, Zagreb*, Croatia: IEEE. , (2013).

[10] D. Kavyashri, *Different types of data mining clustering algorithms and examples*, (2018). Retrieved from DWgeek.com: https://dwgeek.com/various-data-mining-clustering-algorithms-examples.html/

[11] S. Paliwal, *Honeypot: A trap for attackers*, International Journal of Advanced Research in Computer and Communication Engineering, (2017).

[12] A. Ramya, *Securing the system using honeypot in cloud*, International Journal of Multidisciplinary Research and Development, (2015) 172-176.

[13] S. Rani and R. Nagpal, *Penetration testing using metasploit framework : An ethical approach*, Int. Res. J. Eng. Technol., 6 (8) (2019) 538–542.

[14] W. Rowe, *What is the ELK Stack?*, (2019). Retrieved from bmc blogs: https://www.bmc.com/blogs/elk-stack/.

[15] S. Sharma, *Detection and analysis of network & application layer attacks using maya moneypot*, 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence). Noida, India: IEEE, (2016).

[16] P. A. M. Solomon Zemene, *Implementing high interaction honeypot to study SSH attacks*, 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI). Kochi, India: IEEE , (2015).

[17] N. Syuhada Selamat, *Polymorphic malware detection based on dynamic analysis and supervised machine learning*, MSc dissertation, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 2021. Accessed on: 1 August 2021.