IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, November 2022
Impact Factor: 6.252

Identification of Multiple Attacks in Cloud Environment using Big Data

Dr. Samydurai A[1], Achuthan R[2], Akash K[3], Eber Sheckel E[4]
Associate Professor, Department of Computer Science and Engineering[1]
UG Scholar, Department of Computer Science and Engineering[2,3,4]
SRM Valliammai Engineering College, Chengalpattu, Tamil Nadu, India

Abstract: *Since the Cloud environment is prone to many attacks we are implementing a Big Data based centralized log analysis system to identify the network traffic occurred by attackers through DDOS, SQL Injection, and Brute Force attacks. The log file is automatically transmitted to the centralized cloud server where big data is initiated and uses a tool called Hadoop to process the huge amount of log files that are being sent to big data. If an attacker attacks any files then it will be compared with the attack dataset that is maintained here to detect the attacks that are being used by the attacker. This system also helps in storing the information of all registered users, and their files which get uploaded to the cloud server and downloaded from the server and IP addresses in order to view any attacks that may occur in the future. All this stored information is maintained securely using SQL which is a backend process. Thus we are implementing a system that delivers a very high performance as well as very efficient results in categorizing the attacks. Since the Hadoop tool is being used in the system, this system is able to increase its scalability and achieves a faster detection of attacks like DDOS, SQL Injection, and Brute Force attacks. This system can play a vital role in various organizations to safeguard their privacy. Data integrity, Data confidentiality, and non-repudiation can be achieved by using this system. Finally, we can say that this system avoids any faults that occurred in the previous system that was invented before this.*

Keywords: Brute force attacks, Hadoop, Data Integrity, SQL Injection, DDOS Attacks

I. INTRODUCTION

Big data is generally used to process a huge amount of files or any data within a short period which helps the certain system to improve their efficiency. The challenges include analysis, capture, time duration, search, sharing,storage, transfer, visualization, and privacy violations. The larger data set has more tend more than the smaller dataset because a larger data set may be used to compare the variety of data to prevent diseases in the medical area.to avoid crimes in society, and also to spot business trends. we can implement big data in our project because everyemployee has instructed information so we can make an analysis of thisdata. Big data is a term utilized to refer to the increase in the volume of data that is difficult to store, process, and analyze through traditional database technologies. Big data has the nature of translating the given data to a certain extent by performing some processes for it. The term “big data” is relatively new in IT and business. However, several researchers and practitioners have utilized theterm in previous literature. For instance, referred to big data as a large volume ofscientific data for visualization. Several definitions of big data currently exist. Meanwhile and defined big data is characterized by three Vs: volume, variety, and velocity. The terms volume, variety, and velocity were originally introducedby Gartner to describe the elements of big data challenges. IDC also defined big data technologies as “a new generation of technologies and architectures, designed to economically extract value from verylarge volumes of a wide varietyof data, by enabling the high-velocity capture, discovery, and/or analysis.” specified that big data is not only characterized by the three Vs mentioned above can also extend to four Vs, namely, volume, variety, velocity, and value This 4V definition is widely recognized because it highlights the meaning and necessity of big data.

1.1 System Overview

The aim of this project is to identify the attacks like Brute Force attacks, SQL Injection, and DDOS attacks which could cause major threats to the cloud environment by stealing the sensitive pieces of information of the users like bank account details, user login credentials, and their other privacy details. This project helps the user to avoid any malicious websites

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, November 2022
Impact Factor: 6.252

by sending a warning pop-up message with the help of the SQL log file. This project also holds the information of all the previously registered users in order to avoid any difficulties in the future. This project will help to improve the security features at a very high rate and also provides very high performance and scalability with the help of the Hadoop tool. The reason for the system's high scalability is due to the Hadoop tool which processes a very huge amount of data within a short period.

III. LITERATURE SURVEY

Alexander Tolstoy & et.al.,(2016).[1]: Today we witness the appearance of some additional to Big Data concepts:data lakes and fast data. Are theysimply the new marketing labels for the old BigData IT or really new ones? Thus the key goal of the paper is to identify the relationship between these three concepts, giving special attention to their application to information security (IS) issues. The reason lies in the fact that volumes of IS-related information is one thing, but the real problem for securingenterprises' IT infrastructure assets is the speed with which things related to IS happen. Thus to conclude that this paper deals with Big Data which improves this system's scalability

Hussein T.Mouftah & et.al.,(2016). [2] : The digitalization of our day-to-day activities has resulted in a huge volume of data. This data, called Big Data, is used by many organizations to extract valuable information either to take marketing decisions, track specific behaviors or detect threat attacks. The processing of such data is made possible by using multiple techniques, called Big Data Analytics, which allow getting enormous benefits by dealing with any massive volume of unstructured, structured and semi-structured content that is fast changing and impossible to process using conventional database techniques. However, while Big Data represents an immense opportunity for many industries and decisions makers, it also represents a big risk for many users. This risk arises from the fact that these analytics tools consist of storing, managing and efficiently analyzing varied data gathering of all possible and available sources. To conclude, This paper provides a detailed review on challenges of data privacy by providing some protection techniques.

Karim Abouelmehdi & et.al.,(2016). [3]**:** With the growing development of data, it has become increasingly vulnerable and exposed to malicious attacks. These attacks can damage essential properties such as confidentiality, integrity, and availability of information systems. To deal with these malicious intents, it is necessary to develop effective protection mechanisms. In this paper we will indicate the main risks arising in Big Data and existing security mechanisms, we focus on Hadoop security and its components because it remains the required Framework for the management and processing of big data. To conclude that this paper helps to minimize the number of risks that could possibly occur in the system,

Elisa Bertino & et.al.,(2016). [4]**:** Recent technologies, such as IoT, social networks, cloud computing, and data analytics, make today possible to collect huge amounts of data. However, for data to be used to their full power, data security and privacy are critical. Data security and privacy have been widely investigated over the past thirty years. However, today we face new issues in securing and protecting data, that result in new challenging research directions. Some of those challenges arise from increasing privacy concerns with respect to the use of such huge amount of data, and from the need of reconciling privacy with the use of data. Other challenges arise because the deployments of new data collection and processing devices, such as those used in IoT systems, increase the attack potential. Thus to conclude that this paper has discussed the relevant concepts and approaches for Big Data security and privacy, and identified research challenges to be addressed to achieve comprehensive solutions to data security and privacy in the Big Data scenarios.

IV. METHODOLOGY

The proposed system makes use of four modules that work together to build an efficient attack identification system which contains modules like Cloud Establishment, User Interface, Logfile Generation, Logfile Analysis Using Bigdata, Attack Analysis, And Categorization. The behavior and working of the above-mentioned modules are represented in the following figures.

4.1 Cloud Establishment

This module deals with the establishment of the cloud which is done by creating an app on the Dropbox cloud. Netbeans IDE is used here to run java code as a mediator to link the cloud server. After completion of the connection, the administrator can upload the files to the cloud server by proving the login details. Then Users can request these uploaded files from the cloud server. Thus Both the File upload and the file request are handled by the main Cloud Server. This

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, November 2022
Impact Factor: 6.252

system uses Dropbox cloud to improve this system's security feature by generating tokens once in a while in order to prevent any entry of an unauthorized person to access the cloud.

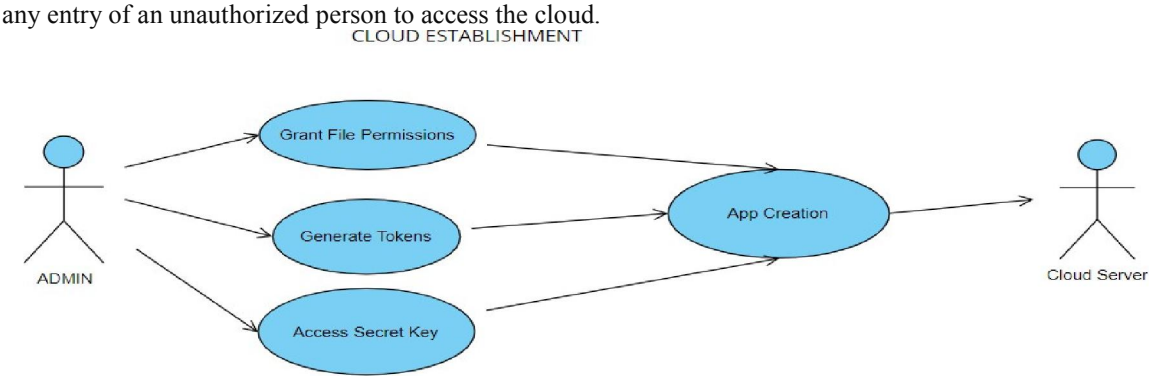

Figure 1: Cloud Establishment

4.2 User Interface

In this module a user interface gets implemented for interaction between the user and system. a dialog box will be generated for various processes like server creation, user registration, user login, file uploading, file downloading Attack identification, etc. when a user's IP gets blocked a pop-up message gets appears to notify the user for warning purpose. This module helps the user to easily interact with the system which simplifies certain processes that take place here.

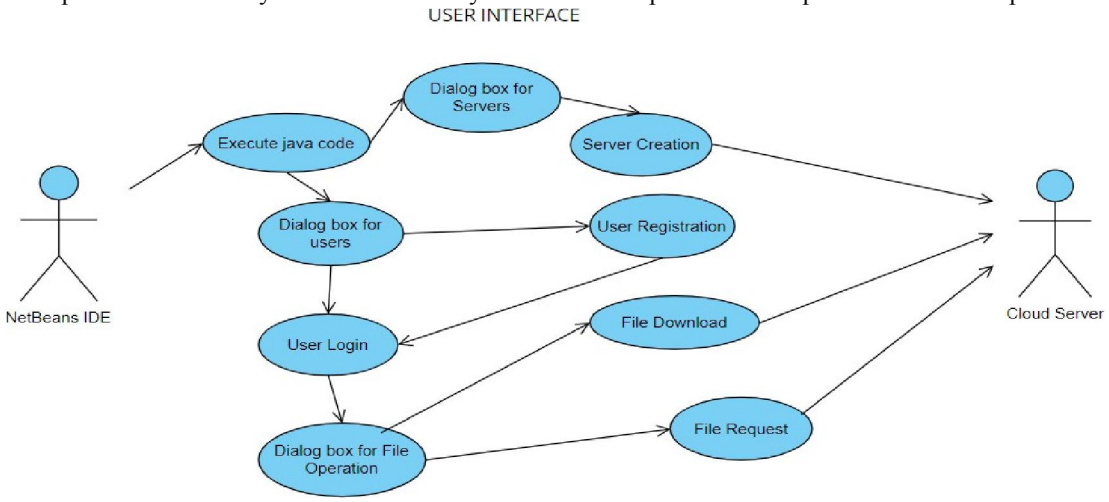

Figure 2: User Interface

4.3 Log File Generation

In this module A SQL server log file gets generated for every process that gets implemented in this system like server creation, user registration, file uploading, file downloading, attack identification, and Blocking of the illegitimate user's IP, etc. This generated Log file helps to maintain the records of a database containing users' information and their system's information as well. In this log file, other activities also get maintained like time entity, request entity, and List of all blocked clients, and files as well.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, November 2022
Impact Factor: 6.252

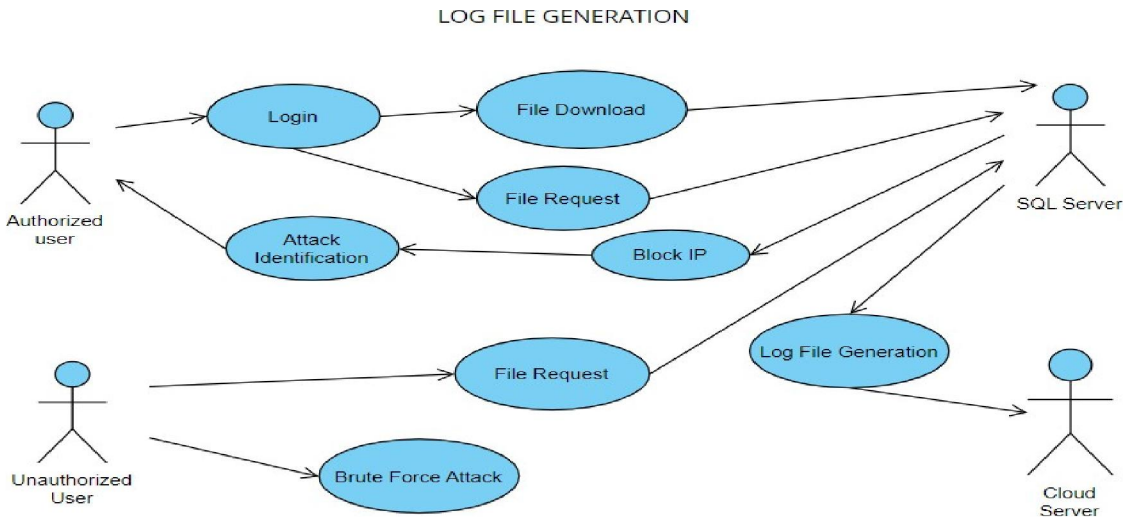

Figure 3: Log File Generation

4.4 Log File Analysis Using Bigdata

This module plays a major role in processing all the files that have been uploaded to the cloud server and downloaded from the cloud server using the big data tool called Hadoop. All of the personal information that has been collected from the user for the login process gets stored in the SQL log file and this information gets maintained securely over there. Whenever administrator wants these stored data they can easily access these data from the SQL server without any difficulties. Thus this module helps the system to improve its security feature.

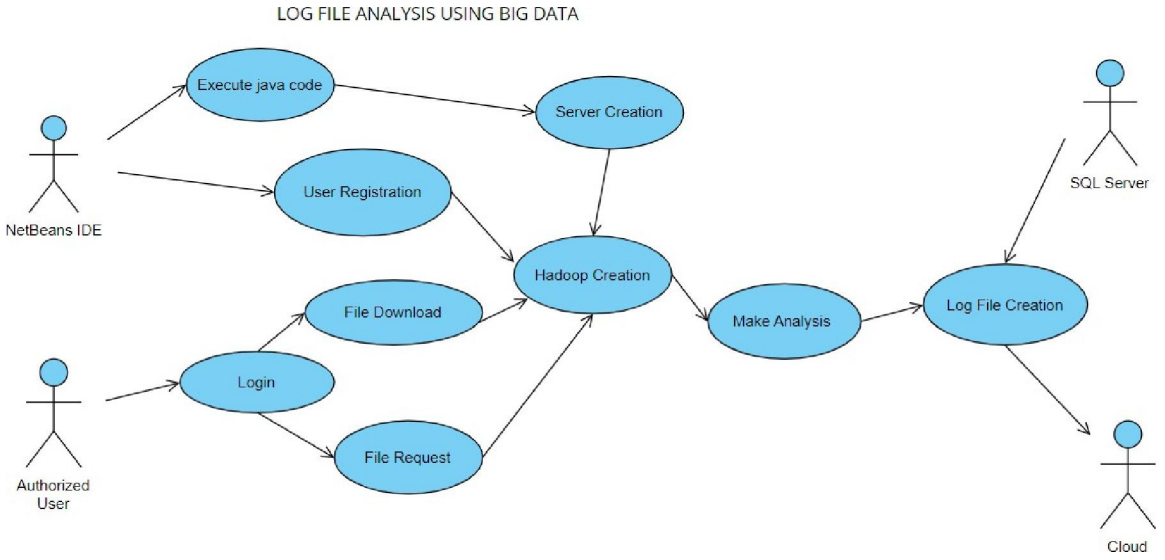

Figure 4: Log File Analysis Using Bigdata

4.5 Attack Analysis And Categorization

This module will play a major role in identifying the attacks like Brute Force Attacks, SQL Injection, and DDOS attacks. The Hadoop tool has been connected to this module which makes the system easily analyze and identify the attacks within a short period. When an illegitimate user has been found their IP gets blocked by this module so that there will be no more occurrence of such users in the future. The SQL log file gets linked to this module so when an illegitimate user's IP gets blocked this information gets maintained over the SQL log file for future reference.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, November 2022
Impact Factor: 6.252

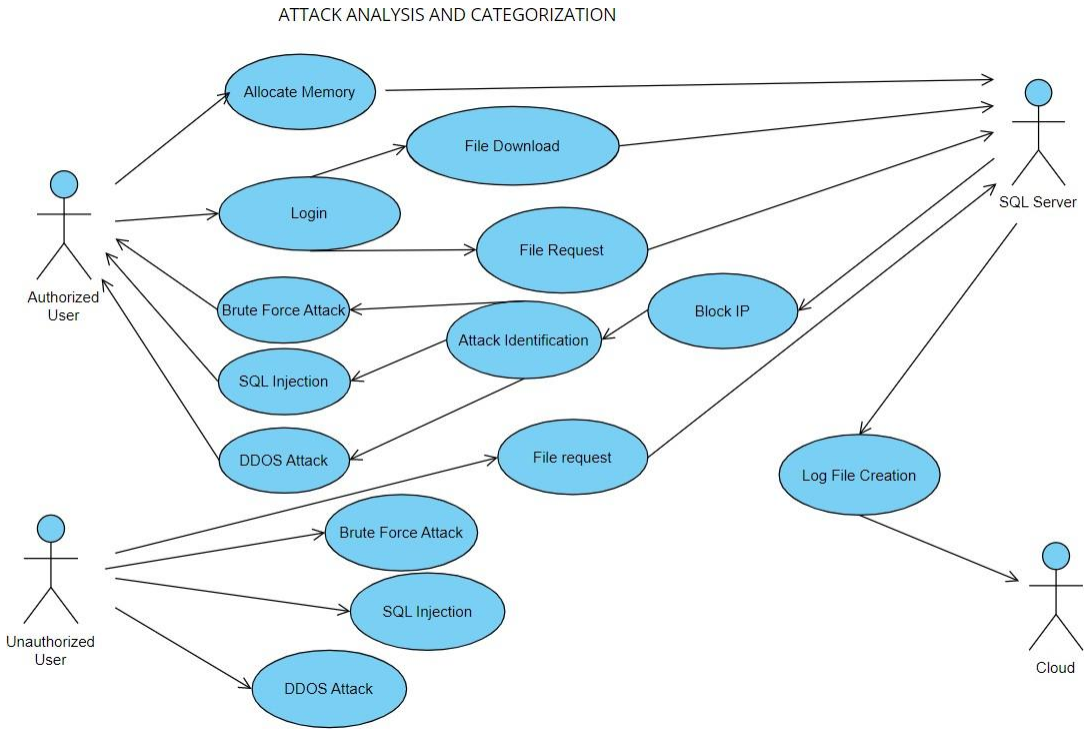

Figure 5: Attack Analysis And Categorization

V. EXISTING SYSTEM

The Existing does not get implemented with the Bigdata tool. Since this system does not use the Bigdata tool, only a limited amount of files or any data can be processed. Time taken to perform certain activities like file downloading and uploading will be increased as well. This system is not able to block a certain user's IP if that user tends to be illegitimate. Even if that attack has been identified there is a high percentage of chance of occurrence of these attacks since there is no blocking process involved in this system.

5.1 Disadvantages

- Generally, once the Attack is identified, that particular request is declined. There is no further blocking system of that IP implemented.
- The performance that has been produced by this existing system is not that high and efficient
- Only A limited amount of data can be used by this existing system since it does get implemented with the concept of Big Data.

VI. PROPOSED SYSTEM

Netbeans IDE has been used in this system to run the java code for various processes like user registration, Attack identification, etc. This system also stores the pieces of information of all registered users and their files which get uploaded to the cloud server and downloaded from the server using an SQL log file. The registered user can securely perform certain activities like file uploading and downloading using Dropbox cloud. Even if an illegitimate user tries to get in, they will be easily get found with the help of a dataset containing all the user's information which gets maintained on the SQL log file. This system has been implemented with a Bigdata tool called Hadoop. This Hadoop tool helps to process a huge amount of files so that File downloading and uploading can be finished within a short period. Thus this system provides high scalability as well as improved performance and efficiency.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, November 2022
Impact Factor: 6.252

VII. IMPLEMENTATION

This system uses 5 modules.

1. Cloud Establishment
2. User Interface
3. Log File Generation
4. Log File Analysis Using Bigdata
5. Attack Analysis And Categorization

These modules collaborate to make this system more efficient. Each module has different functions for this system and the details of the working of these modules have been explained. This system gets implemented by using NetBeans IDE which helps to process the java code for various processes like attack identification, storing user pieces of information, and connecting to the cloud server. At first in the Cloud Establishment module, DropBox cloud has been established by creating an app where all the files get stored here so that in the future it will be useful for the user to download it securely. This Dropbox can perform various activities like file insertion, deletion, and modification of the files which has been stored there. In the User Interface module, A dialog box gets created for the admin to upload the file to the server by giving the appropriate name and password as well as for the users to register their details to sign in to download files from the server by giving the already registered name and password. In the Log File Generation module, A SQL log file gets maintained to store all the information of the user like their IP, login credentials, etc to make sure to prevent any illegitimate user from getting signed in to any web browser. This SQL log file also maintains the dataset of already found attacks in the system to prevent any further occurrence of these attacks in the future. In Log File Analysis Using the Bigdata module, A Hadoop tool has been used in this module to process huge amounts of files for downloading, uploading as well as storing so the time to perform all these activities gets reduced which increases the system performance as well as its scalability. In the Attack Analysis And Categorization module when a user tends to sign in by giving the name and password the system checks by matching these login credentials with the details that have been already stored in the SQL log file to find any presence of the illegitimate user. If the user gives any incorrect login credentials they tend to be recognized as illegitimate thus their IP gets blocked and the attacks that they used will be categorized by analyzing the attack dataset that has been maintained in the SQL log file once the attack has occurred matches with the dataset in SQL log file then warning will be shown to the user by creating a pop message of that identified attack which can be Brute force attack, SQL Injection attack, and DDOS attack.

VIII. CONCLUSION

For every different Cloud Data, most servers need to deploy several servers for Attack Detection not yet deployed for most effective systems. In the current Manualdetection of Attacks process, There is no scope for scalability of the detection database. In thecase of a heavily loaded transaction, the attacker either steals the data ortries to deny genuine requests. Nowadays there are multiple Attacks happening in cloud environments but there is no that is no Auto detection process that is meantto prevent these attacks. Thus the project concludes that through this system we identify the attacks and log files separately. we can also view the attacker's name, IP address, file details, and type of attack details will be shown to the admin.

8.1 Future Scope

This system has a very high potential for future modifications and improvements and is easyto implement so that it can be easily deployed in other areas such as various websites like Shopping websites, Educational websites, Amount transactions websites, email id as wells as Online apps used for video conferencing, phone calls, etc. The introduction of newer technologies could also be a further future improvement let's say the use of advanced versions of the tools that had been implemented in this system could lead to something more productive and a really effective method if deployed. As we are well aware the world is going digital now and it makes sense to improve security features in all the tools that we have been using in our everyday life. Hence This system can help to keep the various organizational information hidden thus outsiders will not have the opportunity to have these details bringing disclosure to unwanted people. It also improves the trust between the users and the company by maintaining their products' security at a higher level which helps society to develop at a very high rate.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, November 2022
Impact Factor: 6.252

ACKNOWLEDGEMENT

We sincerely express our gratitude in depth to our esteemed Founder Chairman & Chancellor **Dr. T. R. Paarivendhar, Thiru. Ravi Pachamoothoo Chairman**, **Mrs. Padmapriya Ravi** Vice Chairman, **Ms. R. Harini** Correspondent, **SRM VALLIAMMAI ENGINEERING COLLEGE** for the patronage on our welfare rooted in the academic year.

We express our sincere gratitude to our respected Director **Dr. B. Chidhambararajan, M.E., Ph.D.,** for his constant encouragement, whichhas been our motivation to strive towards excellence.

We thank our sincere Principal **Dr. M. Murugan, M.E., Ph.D.**, for his constant encouragement, which has been our motivation to strive towards excellence. We extend our hand of thanks to our Head of the Department, **Dr. B. Vanathi, M.E., Ph.D.**, Professor for her unstinted support. We are thankful to our project coordinator, **Mrs. G. Sangeetha, B.E., M.E.,** for providing us with the essential facilities. We are grateful to our internal guide **Dr. A. Samydurai, M.E., Ph.D.,** Associate Professor without whose invaluable guidance and encouragement, this project would not have been a success.

We also like to thank all **Teaching and non-teaching staff members** of our department, for their support during the course of the project. We finally thank our friends and family for their support during the course of the project.

REFERENCES

[1]. D. Fisher, "'venom' flawin virtualization software could lead to VM escapes,data theft," 2015. [Online]. Available: https://threatpost. com/venom-flaw-in- virtualization-software-could-lead-tovm- escapes-data-theft/112772/, Accessed

[2]. on:May 20, 2015.

[3]. Z. Durumeric, et al., "The matter of heartbleed," in Proc. Conf. Internet Meas.Conf., 2014, pp. 475–488.

[4]. K. Cabaj, K. Grochowski, and P. Gawkowski, "Practical problems of internetthreats analyses," in Theory and Engineering of Complex Systems and Dependability. Berlin, Germany: Springer, 2015, pp. 87–96.

[5]. J. Oberheide, E. Cooke, and F. Jahanian, "CloudAV: N-version antivirus in the network cloud," in Proc. USENIX Secur. Symp., 2008, pp. 91–106.

[6]. X. Wang, Y. Yang, and Y. Zeng, "Accurate mobile malware detection and classification in the cloud," Springer Plus, vol. 4, no. 1, pp. 1–23, 2015.

[7]. P. K. Chouhan, M. Hagan, G. McWilliams, and S. Sezer, "Network-based malware detection within virtualized environments," in Proc. Eur. Conf. ParallelProcess., 2014, pp. 335–346.

[8]. M. Watson, A. Marnerides, A. Mauthe, D. Hutchison, and N.-ul-H. Shirazi, "Malware detection in cloud computing infrastructures," IEEE Trans. Depend. Secure Comput., vol. 13, no. 2, pp. 192 205, Mar./Apr. 2016.

[9]. Fattori, A. Lanzi, D. Balzarotti, and E. Kirda, "Hypervisor based malware protection with Access Miner," Comput. Secur., vol. 52, pp. 33–50, 2015.

[10]. T. Mahmood and U. Afzal, "Security analytics: Big data analytics for cybersecurity: A review of trends, techniques and tools," in Proc. 2nd Nat. Conf.Inf. Assurance, 2013, pp. 129–134.

[11]. C.-T. Lu, A. P. Boedihardjo, and P. Manalwar, "Exploiting efficient data mining techniques to enhance intrusion detection systems," in Proc. IEEE Int. Conf. Inf. Reuse Integr., 2005, pp. 512–517.

[12]. Alexander Tolstoy & et.al.,(2010).Application of Big Data, Fast Data andData Lake Concepts to Information Security Issues

[13]. Hussein T.Mouftah & et.al.,(2010).Big Data Analytics: Security andPrivacy Challenge 13.Karim Abouelmehdi & et.al.,(2020).Big Data Emerging Issues: Hadoop Security and Privacy

[14]. Elisa Bertino & et.al.,(2018). Big Data Security and Privacy

[15]. Azzam Mourad & et.al.,(2018). How to Distribute the Detection Load among Virtual Machines to Maximize the Detection of Distributed Attacks in the Cloud?

[16]. Michael R. Watson & et.al.,(2018).Malware Detection in Cloud Computing Infrastructures