International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 22 (2017) pp. 12237-12245

Identification of Guilt Agent and Leaked Data by Using MAC-IP

B. Raja Koti, Dr. G.V.S. Raj Kumar and Dr.Y. Srinivas

[1]Research Scholar, [2]Associate Professor, [3]Professor
[1,2,3] Department of Information Technology, Gandhi Institute of Technology and Management (GITAM), Visakhapatnam campus, Gandhi Nagar, Rushikonda, Visakhapatnam-530045, Andhra Pradesh, India.
[1,2]Orcid Id: 0000-0002-2325-2732, 0000-0001-6782-4141

Abstract

In this digital world, there is a phenomenal growth in technology which deals with a huge amount of data. Due to this availability and usage of massive data there are chances of data exposure leading towards leakages or modifications. This indirectly accounts to huge losses with respect to privacy of an organization. In this article, we propose a methodology for detecting the data leakages, identifying the guilt agents and preventing the guilt agents, by using MAC-IP binding technique. This methodology helps to restrict authorized data accessing by unauthorized users and also maintaining the log records for every data transactions.

Keywords: Data Leakage, Guilt Agent, MAC-IP Binding, Data Protection, Data Privacy.

INTRODUCTION

Every organization is highly dependent on data for day to day transactions. So securing the data becomes a major concern for any organizations. The value of the data is incredible, so it should not be leaked or altered. There are number of methods designed in particular aiming towards for the data security by using different encryption algorithms. In-spite of these secured methodologies, there is a big issue of the integrity of the users using these systems. It is very hard for any system administrator to trace out the data leaker among the system users. It creates a lot of ethical issues in the working environments. Therefore Data leakages are considered to be one of the biggest challenging issues for any enterprises.

The aim of Information security is to provide a protective mechanism to protect the data from an unauthorized user. Some of the security concerns arise due to the availability number of web resources emerged due to the growth of computer networks. There exist two types of attacks, viz., passive attack and active attack. The information security can be classified based on three aspects namely Security Attacks, Security Mechanism and Security Service. Security service contains Confidentiality, where the content is to be known only to the sender and receiver, Integrity, concerns with the correctness of data to be considered during transfer, keeping in account of the available resources. Authentication is another aspect that is to be taken care while ensuring the confidentiality.

In-spite of these security constraints, there is every possibility of data leakages. Leakage of data may have an impact on confidential information like finances, shares, stake holders, partners, private details of customers and employees. While dealing with vast information, there is a high probability of the data to be leaked by an unauthorized person. This leakage may occur from within the organization and therefore to safeguard the data, it is necessary to identify the data leakage sources and the guilt agent involved. With the advances in technology, most of the data of the organizations are stored across the globe using sophisticated servers or utilizing the cloud computing technologies for storage purposes. This data that is stored globally can be provision to retrieve only by an authenticated super user or administrator, who may be the head or any superior authority of the organization. This data is known as Master Data. The admin now manages the master data and sends a replica of this data to a group of authorized users. Once this data is sent to a user, it is no more in the control of the admin of what the user does with the data. A user may leak this replicated data to an unauthorized organization.

In such cases, the confidentiality of an organization is being at stake. So, to prevent such loss, this data must be secured such that even if it reaches to an exploiter the data cannot be utilized. In the present era, data leakages have become very frequent, leading to huge losses (in terms of finance and also confidential information) to organizations. Using mathematical probability, we are aware of the probable guilt agents and in some cases we can also find the guilt agent, but we are unable to protect this leaked data. Our method proposes a way to protect this leaked data from the unauthorized personnel and also it suggests a way to find the proper guilt agent with the usage of **MAC-IP Binding** technique.

The proposed methodology helps to identify such data leakages and find out the guilt agent who leaked the data. Also, it provides a platform to showcase how to protect this leaked data from being accessed by unauthorised. The rest of the paper is organized as follows; this paper deals with the recent Literature review in this current scenario, Data Leakages and Guilt Agents identification is presented in the paper, highlights about MAC-IP Binding. Deals with the methodology developed, the results derived are presented and the conclusions derived are summarised.

LITERATURE REVIEW

The survey of Data Leakage techniques [1] highlights that; one can detect the guilty agent without changing the reliability of the original data. An investigation of data leakage is a scary

proposition. Security practitioners have always had to compact with data leakage issues that take place from email and other Internet channels mostly. Data security is the critical part for most businesses, organizations and even in home computer users. Data like client information, payment information, personal files, bank account details is hard to change and potentially dangerous if it falls into the wrong hands. Data lost due to disasters such as a flood or fire results into crushing. In a contrary, if the data is hacked by unauthorized persons, it may lead to more disastrous situations than the data crushing. However, some of this information should be bonded to be within the system. The unauthorized access of this data could lead to many problems for the larger organizations or even in the personal home user also. Today most of the problems associated with the data loss is with respect to stolen of passwords and thereby utilizing these passwords for accessing the individuals bank accounts in an unethical manner.

The problem of data leakage has been an issue of major concern for enterprises since quite a long time [2]. Various techniques for prevention of data leakage as well as detection of leaked data has been drafted and practiced. Traditionally leakages was handled using watermarking [4], [5], wherein a unique code or a signal is securely, imperceptibly, and robustly embedded into each copy distributed among the third parties. Watermarking was initially used in images [6]. Technique of data provenance is also employed for the purpose of detection. Watermarking technique includes modification in original data which is not acceptable in certain cases. Recent works, [7], [8], also presented marks of insertion mechanisms to store the relational data using watermarking techniques. The guilt agent detection model is proposed in this paper is similar to data provenance [9]. In this article, tracing the lineage of S objects implies essentially the detection of the guilty agents. All these suggested solutions are domain specific, those are lineage tracing for data warehouses [10], and assumes some prior knowledge on the way a data view is created out of data sources [11]. The loss of sensitive data and the other forms of initiative data it can be lead to significant loss to financial and reputational damage. While enterprises are now well-aware of these types of danger issues and then data protection has become a hot topic, even though many organizations are not familiar with this content-aware of technologies, and also don't fully understand the business case for protection of their data. With this context in mind from the review of Data Leakage techniques, the current proposal addresses a model to overcome all the Data Leakage issues.

DATA LEAKAGES & GUILT AGENT

Data leakage can be accomplished by simply remembering what was seen, by physical removal of devices like tapes, disks and reports or by restrained ways such as data hiding. This may be electronic, or may be via a physical method. The terms Data Leakage and Information Leakage are synonymous. The term "unauthorized" does not mean only means of intentional or malicious data. Unintentional or inadvertent data leakage is also deemed to be unauthorized. In order to execute the proper protective measures, one must understand the very objective of the data that needs to be protected in prior.

Every organization adopts its own policies and strategies to protect data from external access, but does not protect against theft and accidental disclosure of sensitive data by employees and partners. One should aware that all the data losses are not only due to the result of external, malicious attacks. The chance of disclosure or mishandling of confidential data done by internal members is also a noteworthy factor that results into data leakages. So it is core essential to track every file that enters and leaves via a network for ensuring the confidentiality. Further it is needed to block sensitive data transfers via Universal Serial Bus (USB) drives and other removable media. For that purpose, if security event is triggered, access to a specific endpoint which is unauthorized needs to be blocked instantly. Based on publicly disclosed Data Leakage breaches, the type of data leaked is broken down as follows:

Unauthorized application use:

It is believed that the usage of unauthorized applications leads towards as much as half of an organization's data loss. Percentage of data loss incidents from the usage of unauthorized applications: **70%**.

Mishandling of communal organization systems:

An employee sharing their work device with unauthorized persons without the supervision of a super user or administrator leads to about **44%**of data loss incidents.

Access to unauthorized networks:

The IT professional claims that dealing with restricted ports, during the breakdown of network accounts to malicious attacks. Percentage of data loss incidents due to unauthorized network access is around **39%**.

Remote Access leakages:

Employees working from home, transferring data between work and personal computers results to about **46%** of data leakage incidents.

Misuse of passwords:

Globally, around **21%** of working sector shares their passwords with colleagues. In countries like China, India and Italy, this rate jumps to around **29%.**

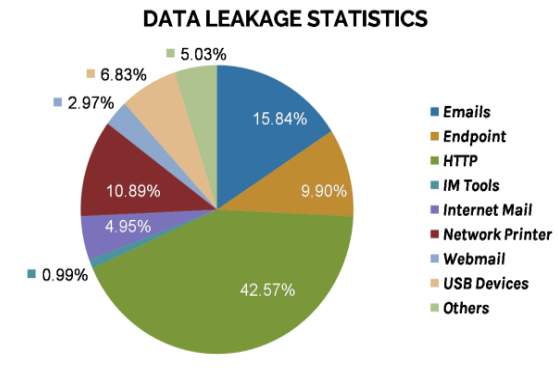

Figure 1: Data Leakage Statistics.

Leak Channels:

For unintentional leaks by employees, the most significant channels were e-mail (23.7%), paper documents (16.2%), removable media (6.7%), and equipment theft/loss (5.3%), with 47.1% of all recorded unintentional leaks going through the network channel.

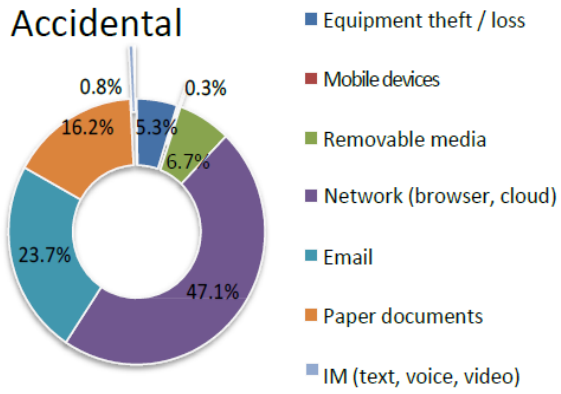

Figure 2a: Data Leakage Accidental.

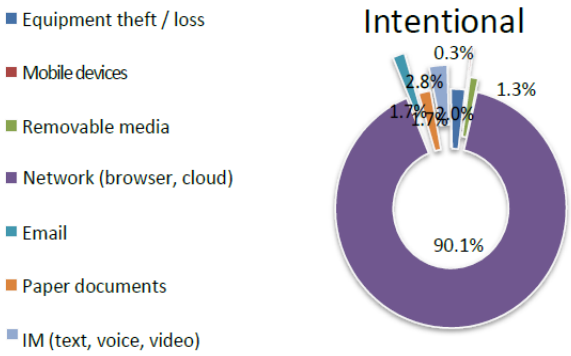

Figure 2b: Data Leakage Intentional.

Malicious leaks prefer the network channel, with more than 90% of the cases involving unauthorized information transfer or disclosure via the Internet (including web services, e-mail, and other online resources). The network channel prevails both in the number of leaks and volume of compromised personal data, accounting for more than **66%** of such incidents (see Fig.3).

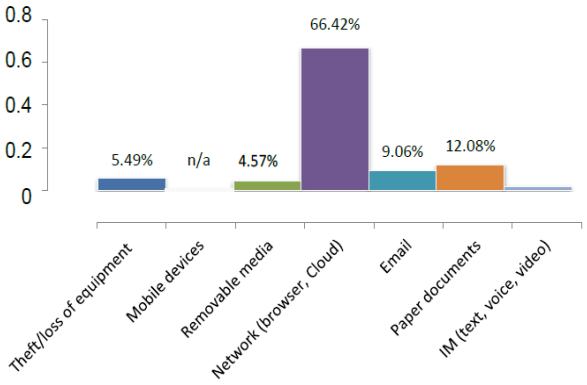

Figure 3: Malicious leaks.

The fact that network is now the main channel for accidental and intentional leaks shows its growing importance among business sectors. Abundance of web-based communication services and annually increasing number of human errors accounts towards the increased share of accidental leakages, which are mainly due to the process involved while distributing the data, posting data over the Internet, etc. On the other hand, offenders less often use knowingly controlled data transfer channels, such as e-mail and instant messengers. In this context, network remains a channel where control and protection tools surpass the capabilities of offenders.

Industry-Specific Map

The industry-specific map, presented in fig.4, portrays a more comprehensive picture of leaks. A bubble size represents a total volume of compromised records by all segment of companies (in millions of personal data records), while vertical axis shows a number of leakages by industry (Industry-specific leakage includes personal data leakages where the exact number of compromised data is known. However, the given data of volume calculated for the industry excludes mega leakages instances with over 10 million records being compromised). The map is divided into three diagrams depending on the size of an affected company (small, medium, and large).

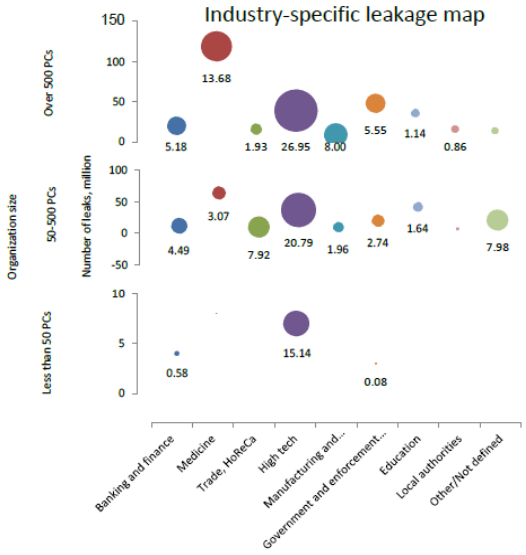

Figure 4: Industry-Specific Map.

GUILT AGENT

A guilt agent is a person who transmits authorized data (or information) to an unauthorized organization or person. This guilt agent is a definite employee or a person with access to this sensitive data of the company. Recent years have seen an increasing number of agents being developed to extend the sphere of human. Those agents, with their autonomous reasoning and decision-making capability, can engage in complex interactions on behalf of their owners. There is no single-agent system in specific. Instead, agents usually live in a society of agents, which is known as multi-agent system. Usually, agents in MAS (multi-agent system) represent various stakeholders, each with distinct interests and objectives. They try to pursue their own objectives, even at the cost of others. Various modalities of identifying guilt agents can be possible in E-mails, Chatting, using USBs, and smart phones. However in this article we have confined ourselves for analyzing the cases of identifying guilt agents in specific to file transfer.

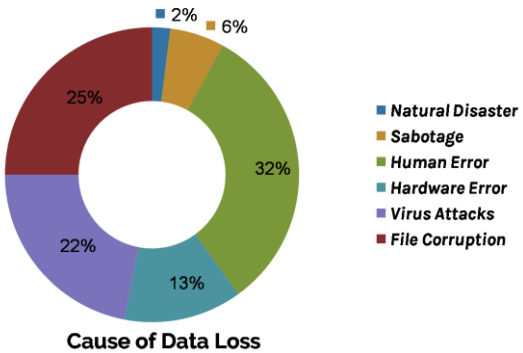

Figure 5: Cause of Data Loss.

MAC-IP Binding

Before starting off with the concept of MAC-IP Binding, an introduction to MAC address and IP address is presented in the subsection of this section.

MAC address:

A media access control address (MAC address) of a computer is a unique id given to network interfaces for communications at the data link layer of a network segment.

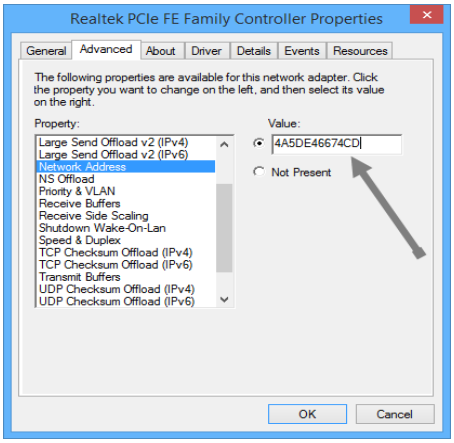

Figure 6: A device's MAC identity.

As shown above Fig.6, a unique MAC address is assigned to a network device. Every hardware device on a local network address has a unique MAC address. A MAC address is often assigned by a network interface controller (NIC) manufacturer and is stored in the hardware.

IP address:

An Internet Protocol address (IP address) is a numerical label assigned to each device (e.g., computer, printer) that participate in a computer network that uses the Internet Protocols for communication with other servers.

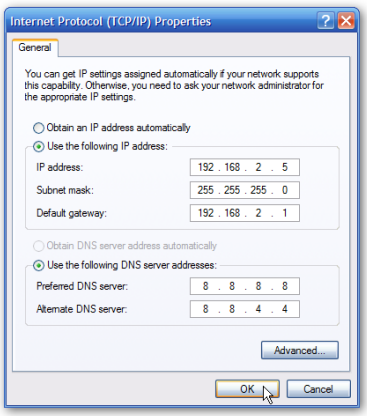

Figure 7: IP address.

An IP address is categorized into two versions: Internet Protocol Version 4 (IPv4) which is a 32-bit number and most commonly found in use today and Internet Protocol Version 6 (IPv6) which is a 128-bit number. IP addresses are written and displayed in human-readable notations like 192.168.256.1 (IPv4) and 2001:bd7:0:123:0:4567:7:1 (IPv6).

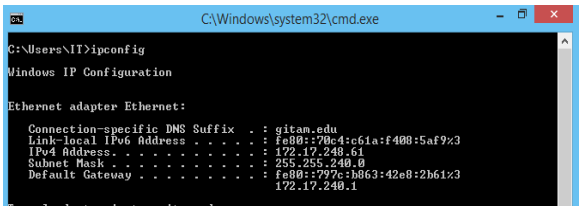

Figure 8*:* Physical address and IP address.

The above Fig.8 depicts an IPv4 and an IPv6 Address alongside a physical address, i.e. a MAC address in a command prompt.

MAC-IP Binding:In the method of MAC-IP Binding, A Dynamic Host Configuration Protocol (DHCP) server assigns an IP address to a device, every time it connects to the network. A MAC address will already be assigned to a network interface. Binding an IP address avoids IP address changing with reconnection and altering of that IP address. Once an IP address and MAC address are bound, the IP address will be reserved for that device and that device will get that IP address every time it connects to the router.

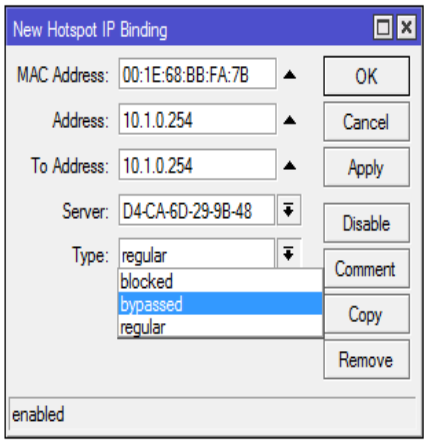

Figure 9: The process of MAC-IP binding.

As the above Figure9 depicts, a Network is managed at the host server and forwarded to a router/modem which consists of both a MAC and an IP address. Both these addresses are bound at this stage and when users try to access, they receive a sub-protocol of this IP address. At the user's end, the MAC ID of the system and the IP address assigned to this system are bound again. As mentioned, when MAC-IP binding is done, if the system is networked with the same IP address as the bounded IP address, only then, the user gets access to the Internet.

METHODOLOGY

In our proposed methodology, we use the technique of MAC-IP binding together with maintaining a log file at the server side to identify the guilt agent and protect the data that is being transferred without authentication. Even though, it is secure, there is a chance of risk of data. To eliminate, this risk our model proposes a methodology.We need to protect copyrighted information against security threats that are done by authorized employee freedom of movement and new communication channels.

Algorithm for the proposed model

Step 1: Start

Step 2: Get IP& MAC

Step 3: Bind and store in server

Step 4: Any request from client

Step 5: Server check bind and response

Step 6: If bind is not matched, server is alerted

Step 7: If bind match also server will respond to client

Step 8: Server gets IP and MAC of that client

Step 9: If client sent any file to another client

Step 10: Repeat step 4 to 7

Step 11: Server record all the move of its users

Step 12: Sever identify that miss match request

Step 13: And checks the all log records of it

Step 14: Finalize the data leaked or not

Step 15: Exit

Linking MAC-IP addresses

Initially, the IP address and the MAC address of the device are linked or bounded and stored to the host server as a file. The concept of MAC-IP binding reduces the alteration of IP or MAC addresses meaning a system's MAC bounded with an IP address connects to the network only with the bounded IP.

Data transmission using protocols

The data in an organization is transmitted through out using protocols. The file records each and every move of the data in the protocol and the users that retrieve this data gives a maximum assurance that this data stays within the authenticated organization.

Combining the MAC-IP address log file

The bounded MAC-IP addresses that are stored as log file in the host server are now placed with each record of data movement in to file. For e.g. if User1 accesses the data, his IP and MAC address are linked with this log file recorded file and action was record in to recorded file in server .

Identifying the guilt agent

Now, when the MAC-IP address is linked with the log file in the server, when there is an unauthorized record of move(s) in the protocol, that record is detected. Using the MAC-IP address linked to the particular log file and the time stamps allotted by these records, the guilt agent can be detected when he makes a move without the permission of the super-user.

Protecting the leaked data

Even after finding the guilt agent, the transfer of data is successful to the external person/ company. To protect this data from being accessed by the unauthorized users, our method proposes the following: The MAC-IP addresses that are linked to the data in the log file always checks for the correct MAC and IP address. If it is leaked and transferred to another organization or person that are trying to access this data, the file of data detects a mismatch in its own MAC and IP addresses and realizes it is out of its original user or authorized user. Now, this authorized user that receives files containing in that data is either corrupted or encrypted using encryption algorithms.

In the below Fig.10, let's assume that two different types of organizations are there which were connected by internet network. Two types of paths drawn, are the communications in between them. Green line was correct communication because the bind value was matched so that data will share among them.

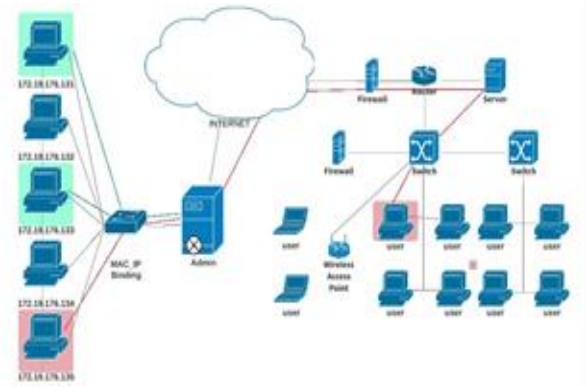

Figure 10: The proposed methodology diagram.

The red line indicates miss match of bind value so Admin will record those two user IP & MAC address to verify what they were sharing among them and also identify if user was guilt agent or not. So whenever Admin will get this miss match alerts then that Admin will check and identify data that leaked and also who did it.

EXPERIMENTAL RESULTS

Get IP & MAC Address from the system and Bind those two addresses by using SHA1 algorithm and store in server.

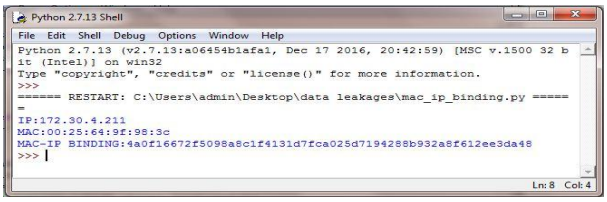

Figure 11: MAC-IP Binding.

Any request from client.

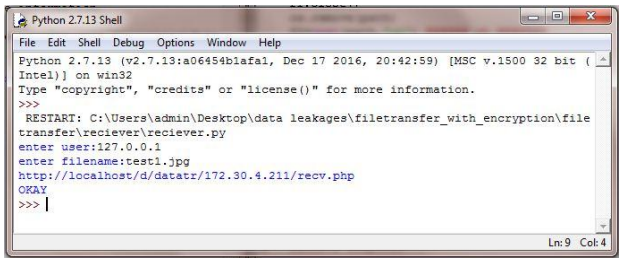

Figure 12: Request and response from the server.

Server verify bind and response.

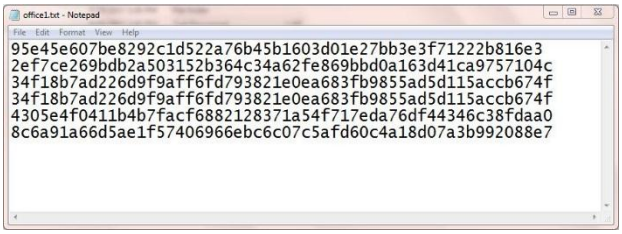

Figure 13: All user bind values saved in file.

If bind is not match server gets alert and it stored in the server

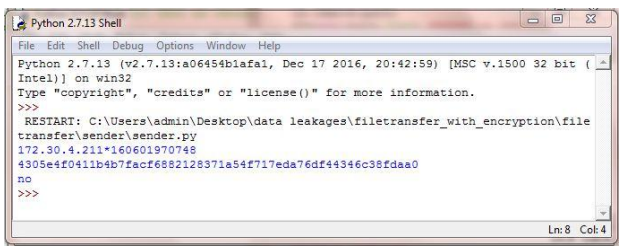

Figure 14: Bind value is not matched.

If bind match response to client.

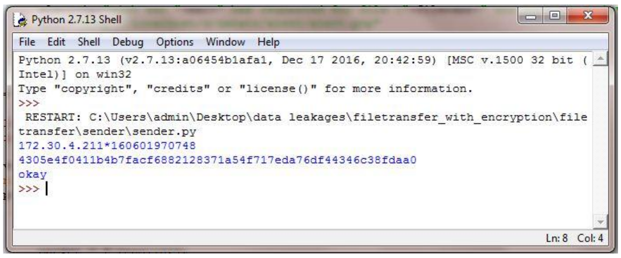

Figure 15: Bind value is not match.

Server gets IP and MAC of that client in case of not matched of the bind value.

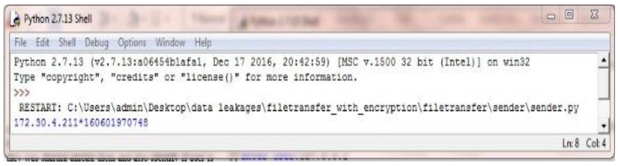

Figure 16: Server will get IP and MAC address.

If client sent any file to another client.

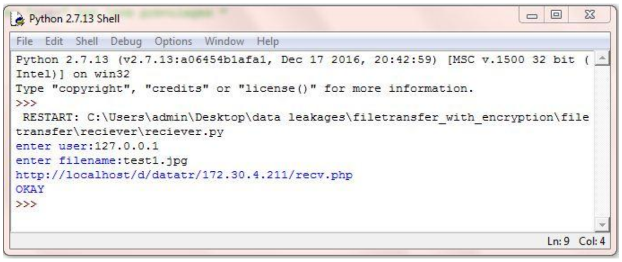

Figure 17: Client Request and response from the server.

Server record all the move of its users.

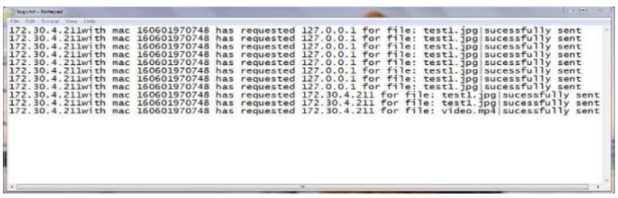

Figure 18: All request and sent files.

Sever stores the all log records of requests and responses.

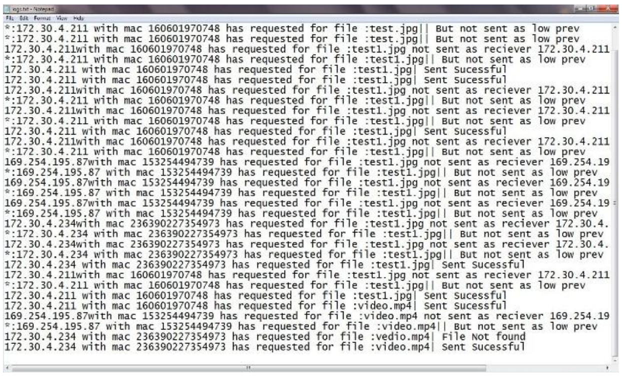

Figure 19: All requests and response.

Sever identify that miss match requests and verify the all the log records of it Finalize the data status.

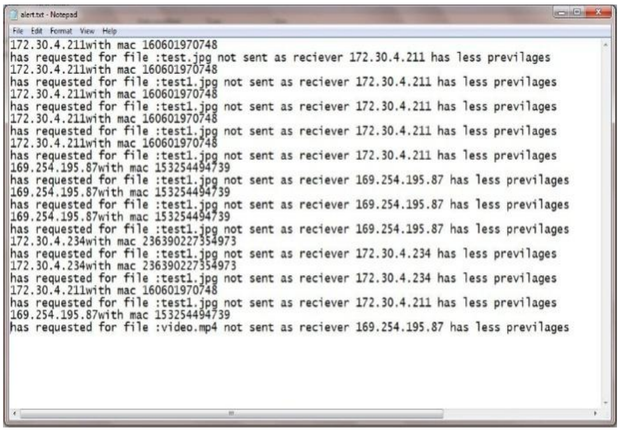

Figure 20: Miss Match bind values requests.

CONCLUSION

This paper introduces the MAC-IP binding technique with an application of this technique to identify the guilt agent(s) in a particular organization and also proposed a method to protect the leaked data.

In this presented article, all the records are stored manually and alert of a message was also known by observations only. As an extension to this article, it is needed to develop a novel method that can solve the problem of manual intervention. Further, if MAC-IP binding is linked to Block-chain technology, an attempt can be made to reduce the manual maintenance of the record. This concept of Block-chain can be combined with MAC-IP for generating more efficient results, thereby helping to identify the guilt agents and also protect the leaked data.

REFERENCES

[1] B. Raja Koti, Dr. G.V.S. Raj Kumar, and Dr. Y. Srinivas, "A Comprehensive Study And Comparison Of Various Methods On Data Leakages", International Journal of Advanced Research in Computer Science, Volume 8, No.7, July – August 2017, pp-627-631, ISSN No: 0976-5697.

[2] Panagiotis Papadimitriou, "Data Leakage Detection", IEEE Transactions On knowledge And Data Engineering, Vol. 23.

[3] P. Papadimitriou and H. Garcia-Molina, "Data leakage detection", Technical report, Stanford University, 2008.

[4] Rakesh Agrawal, Jerry Kiernan, "Watermarking Relational Databases", IBM Almaden Research Center.

[5] S. Czerwinski, R. Fromm, and T. Hodes,"Digital music distribution and audio watermarking".

[6] J.J.K.O. Ruanaidh, W.J. Dowling, and F.M. Boland, "Watermarking Digital Images for Copyright Protection," IEE Proc. Vision, Signal and Image Processing, vol. 143, no. 4, pp. 250-256, 1996.

[7] R. Sion, M. Atallah, and S. Prabhakar, "Rights Protection for Relational Data," Proc. ACM SIGMOD, pp. 98-109, 2003.

[8] Y. Li, V. Swarup, and S. Jajodia, "Fingerprinting Relational Databases: Schemes and Specialties," IEEE Trans. Dependable and Secure Computing, vol. 2, no. 1, pp. 34-45, Jan.-Mar. 2005.

[9] P. Buneman, S. Khanna, and W.C. Tan, "Why and Where: A Characterization of Data Provenance," ICDT 2001, 8th International Conference, London, UK, January4-6, 2001, Proceedings, volume 1973 of Lecture Notes in Computer Science, Springer, 2001.

[10] Y.Cui and J.Widom, "Lineage Tracing for General DataWarehouse Transformations," The VLDB J., vol. 12, pp. 41-58, 2003.

[11] JaymalaChavan and Priyanka Desai "Data Leakage Detection Using Data Allocation Strategies" International Journal of Advance in Engineering and Technology (IJAET), Volume 6 issue 6, Nov 2013.

[12] B. Hauer, "Data and Information Leakage Prevention within the Scope of Information Security," in *IEEE Access*, vol. 3, no. , pp. 2554-2565, 2015.

[13] Sandip A. Kale C, Prof.S.V. Kulkarni C, "Data Leakage Detection: A Survey", IOSR Journal of Computer Engineering (IOSRJCE) ISSN: 2278-0661 Volume 1, Issue 6 (July-Aug 2012), PP 32-35.

[14] Q. B. Hani and J. P. Dichter, "Data leakage preventation using homomorphic encryption in cloud computing," *2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT)*, Farmingdale, NY, 2016, pp. 1-5.

[15] DivyaChaube, Sonali Gandhi, Priyanka Gupta, Rajesh Kolte, "Implementation of Guilt Model and Allocation Strategy for Data Leakage Detection", International Journal of Scientific and Research Publications, Volume 5, Issue 4, April 2015 , ISSN 2250-3153

[16] S. Peneti and B. P. Rani, "Data leakage prevention system with time stamp," *2016 International Conference on Information Communication and Embedded Systems (ICICES)*, Chennai, 2016, pp 1-4.

[17] V. Vijayalakshmi, T. Rohini, S. Sujatha and A. Vishali, "Survey on detecting leakage of sensitive data," *2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave)*, Coimbatore, 2016, pp. 1-3.

[18] Nikhil Chaware, PrachiBapat, RitujaKad, ArchanaJadhav, S.M.Sangve, "Data Leakage Detection", International Journal of Scientific Engineering and Technology (ISSN : 2277-1581) Volume No.1, Issue No.6, pg : 272-273

[19] Nikhil chaware, prachibapat, ritujakad, archanajadhav, Prof.s.m.sangve, "Data leakage detection",Journal of information, knowledge and research in Computer engineering, Issn: 0975 – 6760, page 534-535

[20] Ghagare Mahesh, YadavSujit, KambleSnehal, NangareJairaj, ShewaleRamchandra, "Data Leakage Detection", International Research Journal of Engineering and Technology (IRJET),ISSN: 2395 -0056, Page 1056-1060

[21] Kumar, A. Goyal, A. Kumar, N. K. Chaudhary and S. Sowmya Kamath, "Comparative evaluation of algorithms for effective data leakage detection," *2013 IEEE Conference on Information & Communication Technologies*, JeJu Island, 2013, pp. 177-182.

[22] Rakesh Agrawal, Jerry Kiernan, "Watermarking Relational Databases", IBM Almaden Research Center.

[23] S. Czerwinski, R. Fromm, and T. Hodes,"Digital music distribution and audio watermarking".

[24] J.J.K.O. Ruanaidh, W.J. Dowling, and F.M. Boland, "Watermarking Digital Images for Copyright Protection," IEE Proc. Vision, Signal and Image Processing, vol. 143, no. 4, pp. 250-256, 1996.

[25] R. Sion, M. Atallah, and S. Prabhakar, "Rights Protection for Relational Data," Proc. ACM SIGMOD, pp. 98-109, 2003.

[26] Y. Li, V. Swarup, and S. Jajodia, "Fingerprinting Relational Databases: Schemes and Specialties," IEEE Trans. Dependable and Secure Computing, vol. 2, no. 1,

pp. 34-45, Jan.-Mar. 2005.

[27] P. Buneman, S. Khanna, and W.C. Tan, "Why and Where: A Characterization of Data Provenance," ICDT 2001, 8th International Conference, London, UK, January4-6, 2001, Proceedings, volume 1973 of Lecture Notes in Computer Science, Springer, 2001.

[28] Y.Cui and J.Widom, "Lineage Tracing for General DataWarehouse Transformations," The VLDB J., vol. 12, pp. 41-58, 2003.

[29] A novel data leakage detection,Priyanka Barge, PratibhaDhawale, Namrata Kolashetti3 Ass. Prof., Department of Computer Engineering, NIRMALA CHOUHAN International Journal of Modern Engineering Research (IJMER) Vol.3, Issue.1, Jan-Feb. 2013 pp-538-540 ISSN: 2249-6645.