Secured Privacy Preservation for Sensitive Data Exposure by using Digital Envelope

Ms. Archana W. Sawarkar
M.E. Student
Department of Computer Science & Engineering
Government Engineering College Aurangabad, India

Abstract— Leakage of data has drastically take place in the recent year. Data leakage problem face by many security companies, analysis institutions and government companies say the survey. From the distinguishable of data leak cases, human mistake are responsible for the data leak from the data leak. Many researcher work on this issues for detecting inadvertent sensitive data leak occurred by the mistakes of human. In the existing system they solve the data leakage problem, existing method implement the method for detecting the data leak. But existing method does not provide the security to the data or the system. The existing method is not secured. Overcome this issue, introduced the method which provides security to the data leak detection system. Also we introduced the concept of digital envelope, in which encrypt the data and send to the DLD system. System used ECC algorithm for security purpose. Here also introduced the bloom filter system for storing the hash values of the data. Finally measure the experimental result plot the graph and compare the result with existing and proposed system.

Key words: Data Leak, Network Security, Privacy, Collection Intersection

I. INTRODUCTION

In today's world, many industries, different institutes and government organization are facing the problem of data leakage. To overcome this problem, many systems designed for the data security by using different encryption algorithms. This is a big problem of the integrity of the users of that systems. It is very complex for any system administrator to discover the data leaker in the system users. It makes a lot many cyber issues in the working office.

Network info drop location (DLD) mostly performs intellectual bundle review (DPI) and seems for any events of tender info styles. DPI is a method to focus on payloads of IP/TCP bundles for reviewing application layer info, e.g., HTTP header/content. Ones the measurements of information found in movement cross the limit then the alarm activated for notification. The popularity framework will be sent on a switch or incorporated into existing system interruption recognition frameworks (NIDS).Direct acknowledge of data leak discovery need the plaintext info. But still, this is undesirable requirement for discover data leakage, because it would not set the classification of the knowledge. Not chances that a discovery system is compromise, at that time it is going to expose the plaintext info. Moreover, the data businessman would have to outsource the data leak discovery to suppliers is also not to uncover the plaintext delicate information to them. At this issue, one wants new info drop discovery arrangements that permit the providers to comb content for holes while not taking within the delicate data. Plaintext information required for data leak in direct acknowledges. Still, this requirement is unsuitable, because it could not find out the classification of the knowledge.

The chance to information security from insider's threat is replacing into more and more essential attributable to the interminable use of the computers and additionally communication systems. Several schemes are assigned for justificatory information from outer attacks but those mechanisms are fail to protect information from unauthorized users could misuse their prerogative in polishing off malicious activities.

In this paper, authors have an inclination to data-leak sensing resolution which would be outsourced associate degreed be deployed in a passing semi honest detection setting. Author, apply and measure fingerprint techniques which raise info privacy all around data-leak catching operations. Author approach depends on a fast and sensible unifacial computation on the spiritualist info (classified documents, sensitive emails, etc.). It allows the info owner to giving protection delegate the content-inspection operation to DLD suppliers whereas not exposing the sensitive information. Implementing author detection technique, the DLD provider is sculptural as associate degree honest but- curious (i.e. semi-honest) antagonist, can entirely gain restricted information regarding the sensitive info from either the discharged digests, or the content being inspected. Victimization, authors technique, an internet service provider (ISP) can perform detection on its customer's Semitic worshiped being traffic firmly and provide data-leak detection as associate add-on service for its customers. In another scenario, individuals would mark their own sensitive information and raise the administrator of their native network to watch info leaks for them.

In this paper further we will see: Section II talks about related work studied till now on topic. Section III current implementation details, introductory definitions and documentations and in addition formally expresses the privacy protection system discussed by this paper. Section IV show conclusions and presents future work.

II. RELATED WORK

In this section we will see work done by the various researchers on privacy preserving of data.

To solve the problem in which a not a common arrangement of important data digests is used as a part of recognition, authors give a privacy preserving information leak detection explanation. Our method allows the owner of information to assign the detection operation to a semi honest supplier in securely manner and also covering the important data to the supplier. By studying how the administration of internet suppliers can support their clients DLD as a surpluses administration with very effective protection ensures[1].

An author implements the technique for calculating the ability for information leak in network traffic. Instead of trying to get the event of important data this is not possible task in the entire universal case objective is to ascertain and confine its most noteworthy volume. For Hypertext Transfer Protocol (HTTP) they also give an estimation computation, the basic convention for web scanning. For the online journal situation for site the results are good. Paper shows that network traffic is rehashed or can found by external this can be an important advantage of paper[2].

Authors present a system called Panorama, which is used to search and break down malware by getting central characteristic. In this broad of system investigations, panorama sufficiently differ all the malware tests and nearly does not have false positive. Also as a contextual study (by utilizing Google Desktop) they shown structure they given can collect its data received to and preparing conduct, and in particular setting it can send the data back to the remote server. This system will catch its data very accurately and preparing conducts. By spying on client malicious systems can break the security of the client conduct. It can also be done with the trustworthy vendors such as Google Desktop or Sony DRM. Google Desktop already shown that their system can detect data get to and preparing behavior, also they have confirm that it sends back sensitive data to remote Servers in specific settings[3].

Authors give Storages Capsules which is a new way for securing private reports on a personal computer. Storage Capsules are nothing but the encoded record compartments which lets hacked machine to view and modify delegate records without malware having the ability to steal data. It is possible by making a checkpoint of the state of present system and rejecting the output of device before allowing access a Storage Capsule. After this forms to the storage capsule are sent towards a trusted module. The module selected as a trusted then declassifies the storage capsule by re scrambling its contents, and sends it to storage in low respectability environment. Storage Capsules are encrypted record holders that permit a traded off machine to safely view and alter sensitive records without malware having the capacity to steal secret information. The primary constraint Suffer huge losses if private information falls into the wrong hands. One of the essential threats to classification is malicious programming on PCs[4].

Authors introduce Aquifer as an arrangement structure and system for neutralizing unplanned information exposure in advanced edge working structures. In Aquifer, application engineers characterize mystery confinements that guarantee the entire customer interface work process describing the client assignment. Aquifer even after it is shared guaranties past basic permission checks and permits applications to hold control of information. By using the record consent it manages the strategic distance from equivocal read-compose record open masks and additionally legitimately broadcasting the names even when the work process name changes between record open and file write[5].

Author developed convention ECKE-1 as a proper elliptic curve Diffie-Hellman two-party key by taking into account convention using verification of the public key. The framework shows that in contraventions of the author's claims convention ECKE-1 is in the opposite of key compromise impersonation attacks. They also launch the improved convention ECKE-1N which can easily take attacks stated. The execution of enhanced protocols is nearly same as the MQV protocol and also has the same security properties[6].

Author gives effective, privacy preserving development of basic genomic computation for an instance, computing the edit distance and Smith Waterman likeness scores between two successions. Such presented strategies are more secure and are highly practical in the world compared with other solutions. To evaluate their model implementation they use sequences from the Pfam database of protein families and shows the demonstrated certifiable grouping arrangement and related issues in a privacy preserving manner[7].

III. IMPLEMENTATION DETAILS

Here we will see the details of the proposed system. Where we will have system overview in detail, proposed algorithm, mathematical model of the proposed system.

A. System Overview

The following figure 1 shows the architectural view of the proposed system.

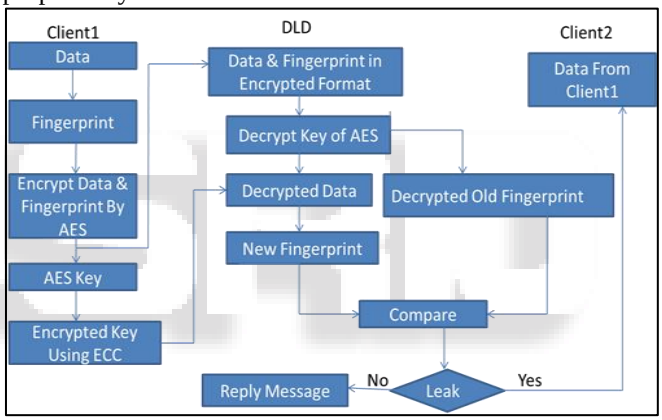

Fig. 1: System Architecture

In proposed system we are using Digital envelope for providing security for data transformation. In digital envelope we are using AES and ECC algorithm for data encryption. Also our system supports multiple file support whereas existing system is limited for only one file. In proposed system we are using SHA 256 algorithm for hash generation. By making use of SHA 256 we are maximizing the efficiency of hash generation of file.

B. Algorithm

In this section we will discuss the algorithm of used in proposed system; AES algorithm and SHA1 algorithm are used for the hashing technique.

1) Algorithm 1: Proposed System Algorithm

- Step 1: Create fingerprint f, which is representation signature of data.
- Step 2: F = {f_1, f_2, f_3,……,f_n} as per data.
- Step 3: f' is fingerprint i.e. generated at DLD side or server side.
- Step 4: f' = {f'$_1$, f'$_2$, f'$_3$,……..,f'$_n$} as per data at DLD
- Step 5: Compare F with F' and if both are equal then send data to destination else discard data.

2) Algorithm 2: AES Algorithm

This is the easy algorithm it can be developed by utilizing the cheap processor and least amount of memory.Very efficient Implementation was a keyfactor in its selection as the AES cipher. The steps involved in AES are

- Key Expansion: Using Rijndael's key schedule Round keys are derived from thecipher key.
- If DistanceToTree(u)>DistanceToTree(DCM) and First-Sending(u) then
- Initial Round: AddRoundKey where Each byte of the state is combined with theround key using bitwise xor.
- RoundsSubBytes: non-linear substitution step ShiftRows: transposition stepMixColumns: mixing operation of each column. AddRoundKey
- Final Round: It contain SubBytes, ShiftRows and Ad-dRoundKey

3) Algorithm 3: SHA1 Algorithm

a) Step 1: Adding Padding Bits

Message is "padded" with a 1 and as many 0's as necessary to bring the message lengthto 64 bits fewer than an even multiple of 512.

b) Step 2: Append Length

64 bits are appended to the end of the padded message. These bits hold the binary formatof 64 bits indicating the length of the original message.

4) Algorithm 4: ECC Algorithm

The Elliptic Curve Cryptographic (ECC) Algorithm followed by following steps:

- Sender and User Calculated B = F = (F1, F2). (Sender sends a message G to User as follows)
- Calculate (F1 * F2) mod N = K.
  Calculate K * G = C, and send C to Sender.
- User receives C and decrypts it as follows:
- Calculate (F1 * F2)mod N = K.
- Calculate (K-1)mod N.(Where N = E)
- K-1*C = K-1*K*G = G.

C. Mathematical Model

1) Input

Browse Dataset

U= {u1, u2, u3,....,un}

Where, U is a set of input data and u1, u2, u3,....,un are the different dataset.

2) Process

Client

C = {C1, C2, C3,....,Cn}

Where, C is represented as a set of Client and C1, C2, C3,….,Cn are the number of Clients.

Bloom Filter

B= { b1, b2, b3,…..,.bn}

Where, B is represent as a set of Bloom Filter Model and b1, b2, b3,…,.bn are set of Bloom Filter Model.

Server S= { s1, s2, s3,….,sn}

Where, S is represent as a set of Server and s1, s2, s3,….,sn are number of server.

3) Output

Final Result FR= {fr1, fr2, fr3,…..,frn}

Where, FR is represent as a set of final result and fr1, fr2, fr3,....,frn are number of final output.

D. Experimental Setup

For building the system we used Java (version jdk 8) and Netbeans (version 8.1) as a development tool. Our system can work on any basic hardware so there is no need of a specific hardware to run the system.

IV. Result and Discussion

A. DataSet

Dataset used in this system is the text file.

B. Results

In table 1 shows the time required for implementing the different methods of the proposed system. Time calculated for different methods, the different kinds of methods are fingerprint encryption, encryption of data, sending the data, digital envelope, compare data fingerprint.

	Encryption of data	Encryption of hash	Encryption of key using ECC	Encrypted Fingerprint	Sending data
D 1	300 Ms.	200 Ms.	450 Ms.	285Ms	310 Ms.
D 2	350 Ms.	305 MS.	55 Ms.	45 Ms.	50 Ms.

Table 1: Time Measurement

The following figure shows the graph of the proposed system. Plot the graph from the above table.

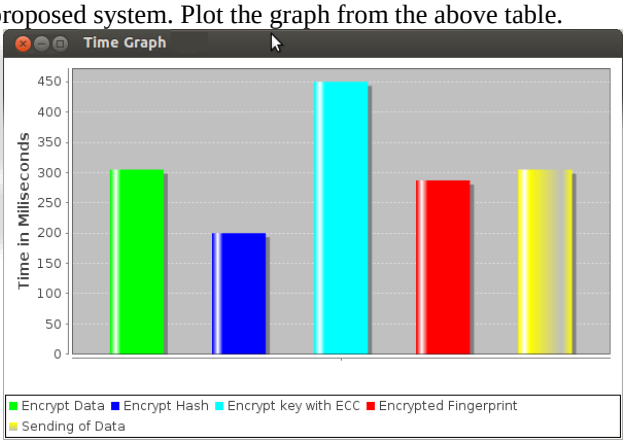

Fig. 2: Time Graph for First Dataset

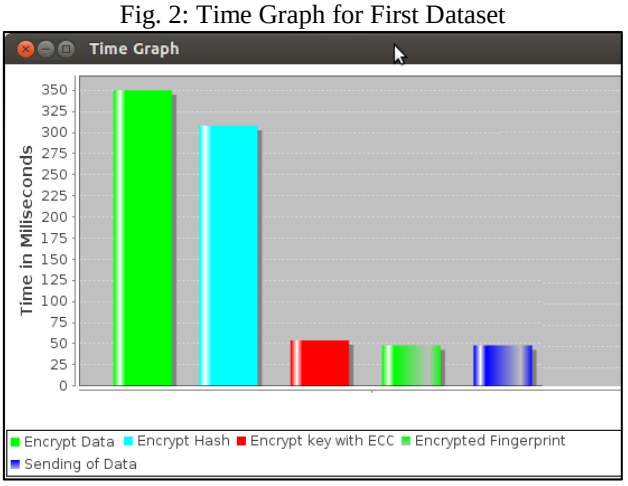

Fig. 3: Time Graph For Second Dataset

Figure 2 and figure 3 show the time graph for two different datasets. In graphs the time taken for each phase/process is shown for two different data sets.

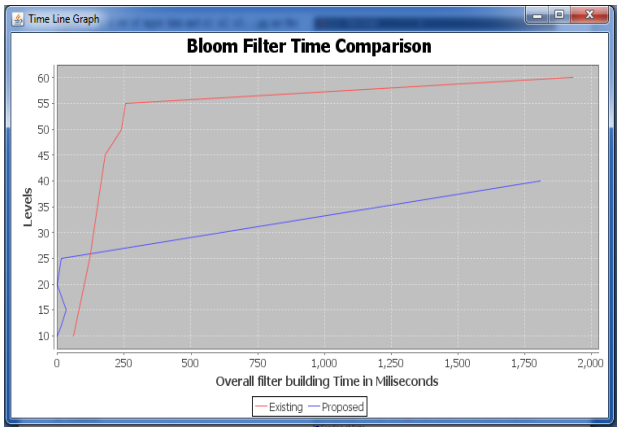

Fig. 4: Time comparison of Bloom Filter Generation

In the figure 4 We have considered the overall filter building time of existing vs. proposed system where x-axis have filter building time in seconds and y-axis contains no. of levels of the system used for building that filter.

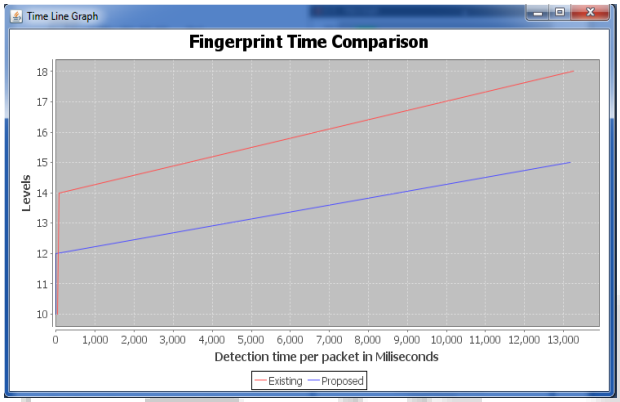

Fig. 5: Time comparision of Fingerprint Generation

In the above figure 5, We have considered fingerprint time comparison of existing vs. proposed system where x-axis have detection time per packet in miliseconds and y-axis contains no. of levels of the system used for building that fingerprint..

V. CONCLUSION AND FUTURE SCOPE

In this paper proposed the secured data leak detection model. Here introduced the system which secured the DLD by using concept of digital envelope, which makes use of the ECC algorithm and AES algorithm for data encryption. In this system implements hash, secured the hash method by using AES algorithm. Also implements the digital envelope technique for security purpose. Finally the experimental result of the proposed system is more secure than the existing system.

ACKNOWLEDGMENT

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. We are thankful to the authorities of Dr. Babasaheb Ambedkar Marathwada University and concern members of Journal, organized by, for their constant guidelines and support. We are also thankful to the reviewer for their valuable suggestions. We also thank the college authorities for providing the required infrastructure and support. Finally, we would like to extend a heartfelt gratitudeu to friends and family members.

REFERENCES

[1] Xiaokui Shu, Danfeng Yao, Member, and Elisa Bertino, ``Privacy-Preserving Detection of Sensitive Data Exposure", In IEEE transactions on information forensics and security, vol. 10, no. 5, May 2015.
[2] K. Borders and A. Parkas, "Quantifying information leaks in outbound web traffic," in Proc. 30th IEEE Sump. Secure. Privacy, May 2009, pp. 129-140.
[3] H. Yin, D. Song, M. Agile, C. Kruegel, and E. Kirda, "Panorama: Capturing system wide information flow for malware detection and analysis," in Proc. 14th ACM Conf. Compute. Commun. Secur., 2007, pp. 116-127.
[4] K. Borders, E. V. Weele, B. Lau, and A. Prakash, "Protecting confidential data on personal computers with storage capsules," in Proc. 18th USENIX Secur. Symp., 2009, pp. 367-382.
[5] A. Nadkarni and W. Neck, "Preventing accidental data disclosure in modern operating systems," in Proc. 20th ACM Conf. Compute. Commun. Secur., 2013, pp. 1029-1042.
[6] Shengbao Wang, Zhenfu Cao, Maurizio Adriano Strangio and Lihua Wang, "Cryptanalysis and Improvement of an Elliptic Curve Diffie-Hellman Key Agreement Protocol", In IEEE Transaction on December 14, 2007.
[7] S. Jha, L. Kruger, and V. Shmatikov, "Towards practical privacy for genomic computation," in Proc. 29th IEEE Symp. Secure. Privacy, May 2008, pp. 216–230

