Detection of Data Leakage in Cloud Computing Environment

Neeraj Kumar
Dept. of CSE
HCST,Mathura-INDIA
neeraj.iiita2009@gmail.com

Vijay Katta
Dept. of CSE
HCST,Mathura-INDIA
vijkatta@gmail.com

Himanshu Mishra
Dept. of CSE
HCST,Mathura-INDIA
himanshu.ims2009026@gmail.com

Hitendra Garg
Dept. of CSE
HCST,Mathura-INDIA
hitendra.garg@gmail.com

***Abstract*—In the recent years internet technologies has become the backbone of any business organization. These organizations use this facility to improve their efficiency by transferring data from one location to another. But, there are number of threats in transferring critical organizational data as any culprit employee may public this data. This problem is known as data leakage problem. In the proposed work, we are suggesting a model for data leakage problem. In this model, our aim is to identify the culprit who has leaked the critical organizational data.**

***Keywords*—Bell-LaPadula model (BLP); Hash Function; AES; Watermark; Message chaining**

I. Introduction

In the current business scenario, data leakage is a big challenge as critical organizational data should be protected from unauthorized access. Data leakage may be defined as the accidental or intentional distribution of private organizational data to the unauthorized entities. It is important to protect the critical data from being misused by any unauthorized use. Critical data include intellectual copy right information, patent information, functional information etc.

In many organizations, this critical organizational data have been shared to many stakeholder outside the organizational premises. Therefore, it is difficult to identify the culprit, who has leaked the data[1][2]. In the proposed work, our goal is to identify the guilty user when the organizational data have been leaked by some agent. In the proposed work, Bell-La Padula security model has been used which provide the analysis and design of secure computer systems. This model is called data confidentiality model.

Bell-LaPadula model mainly focuses on data confidentiality issues and provides controlled access to classified information. In contrast to the Biba-Integrity model which describes rule for the protection of data integrity[3]. In this formal model, the entities in an information system are divided into subjects and objects. The notion of a "secure state" is defined, and it is proven that each state transition preserves security by moving from one secure state to other secure state, thereby inductively proving that the system satisfies the security objectives of the model. The Bell-LaPadula model is built on the concept of a state machine with a set of allowable states in a computer system. A system state is defined to be secure if the only permitted access modes of subjects to objects are in accordance with a security policy. To determine whether a specific access mode is allowed, the clearance level of a subject S is compared to the classification level of the object O to determine if the subject is authorized for the specific access mode. The clearance/classification scheme is expressed in terms of a lattice[4][5] as shown in Figure-1.

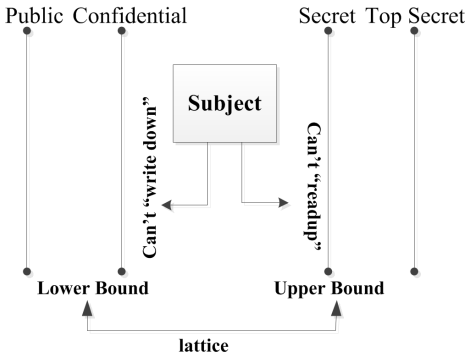

Fig. 1: In the Bell-LaPadula model, each subject S has a lattice of rights

AES algorithm and RSA algorithm shows good performance among different symmetric and asymmetric encryption technique[6] based on different performance factors such as key value, computational speed and tenability. Various experimental factors were also analyzed based on text files used and experimental results proves that DES algorithm consumes least encryption time than AES but in terms of memory usage AES uses least time than DES algorithm. In RSA encryption time is more and also memory usage is very high[7][8]. These techniques are useful for real-time encryption.

In other model, it has been shown a new comparative study between encrypting techniques based on nine factors like key length, cipher type, block size, developed, cryptanalysis resistance, security, possibility key, possible ACSII printable character keys, time required to check all possible keys. Based on these factors AES is better than DES and RSA[9]. It is also been discussed that DES is secret key based algorithm suffers from key distribution and key agreement problems but RSA consumes large amount of time to perform encryption and decryption operation. It have been also observed that decryption of DES algorithm is better than other algorithms in terms of throughput and power consumption[10].

In the recent years, lots of changes happens in the field of watermarking systems. Digital images are more popular than analog due to easy duplication and transmission on different types of networks. Watermarking is used where authentica-

978-1-4799-6929-6/14 $31.00 © 2014 IEEE
DOI 10.1109/CICN.2014.172

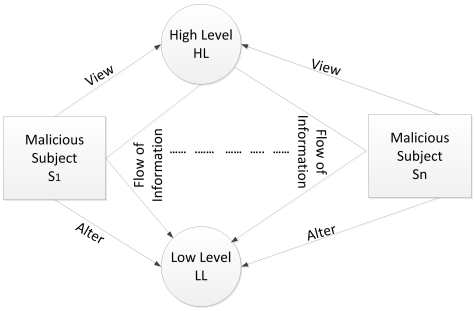

Fig. 2: Information Flow

tion or ownership is needed[11][12]. Watermarking is more efficient tool in ownership claiming and fingerprinting of digital data[11][12]. Watermark can be used to transmit secure message from one place to other place.
The computational cost and time complexity is the measure problems with robust cryptographic algorithms. These techniques use the concept of message authentication. These mentioned techniques ensure that any change in message can be easily traced out (active attack) but it fails in case passive attack. Therefore, one single technique is required for both message confidentiality and authentication. The main concern of the proposed work is to protect the secret information being transmitted.
This paper is structured as follows: In Section II proposed model is discussed. Section III contains Applications and Efficiency Measurement of the proposed model. Results Measurements and conclusion have been discussed in Section IV and Section V respectively.

II. Proposed Model

In this proposed model we are providing the solution for critical Data Leakage problem. The proposed model has been described in the following sections-

A. Secured Environment Infrastructure

We are using the concept of Bell-LaPadula Model for providing secured infrastructure, it is a state-machine model and used to apply access control in different environment such as-
Military security - Army, Air-force, Navy, NATO, NASA etc.
Commercial security- Marketing Sales, Research and development, Human Resource department etc.

In Bell-LaPadula model, information flow will be between the high levels to low level it is shown in Figure-2. We define a state, if the system as a secured environment, and it follows some rule defined, as the allowed access mode of the any subject S, with the any object O is allowed, with respect to defined security policy. To find whether any specific access mode will be allowed, the clearance of a subject S is compared to the classification of the object O. i.e. S =(S1, S2, S3,..Sn), O =(O1, O2, O3,..On) both S and O are combine and creating up the security level used to determine if the subject S is authorized for the specific access mode[13][5].

1) Security Model: Current access set Z=Triplets (S=subject, O=object, A=attribute), and security level is defined as pair of(C, S). C= classification i.e. Public, Confidential, Secret, Top secret. S= category set-i.e. Military, Air force, Defense, R and D. (c1,s1) dominates (c2,s2) iff $c1 \geq c2$ and $S1 \supset S2$. Level-1 dominate Level-2 because they form lattice. The clearance/classification concepts are expressed in terms of a lattice[4][14]. In Bell-LaPadula model there are two basic strategies which defined the foundation for secure access are-

2) Reading down (NRU): A subject S has only read access to objects O whose security level L is below the subject's current clearance level. This prevents a subject from getting access to information available in security levels higher than its current clearance level.

3) Writing up (NWD): A subject S has only write access to objects O whose security level L is higher than its current clearance level. This prevents a subject to pass information from lower level to its current level.

4) Simple Security Property: (NRU- No Read Up) A subject S at a given security level may not read an object O at a higher security level. For any $(S, O, A) \in Z$ if A includes observation, then level(S) must dominate level (O) i.e. Any public user cannot read a top-secret document.

5) "star"-Property: (NWD- No Write-Down) a subject S at a given security level must not write to any object O at a lower security level. If a subject S can observe $O1$ and modify $O2$, then level($O2$) dominates level ($O1$) i.e. cannot copy top secret files into secret files.

6) Read-Only: The subject can only read the object.

7) Append-Only: The subject can only write to the object but it cannot read.

8) Execute-Only: The subject can execute the object but can neither read nor write.

9) Read-Write: The subject has both read and writes permissions to the object.
The Read and Write access of bell-LaPadula model is explained in Figure-3 where every subject and object has their own clearance and classification level respectively by which they are able to access the document.

10)Tranquility Principle: The tranquility principle of the Bell-LaPadula model states that the classification of a subject or object does not change while it is being referenced[15].

B. Creating Watermark

In this model server will add an image logo to all the stored documents and this image logo represents the organization. As we know that each intensity value in the image ranges from 0 to $(2^{24} - 1)$, and for each of the three components of color image as RED, GREEN and BLUE ranges from 0 to $(2^8 - 1)$. Each character, has their ASCII values ranges from 0 to $(2^8 - 1)$. So, any text can be Inserted into the document by replacing the intensity value of pixel location, with the ASCII value of character, which is needed to be hide and transmit with the documents[16][17].

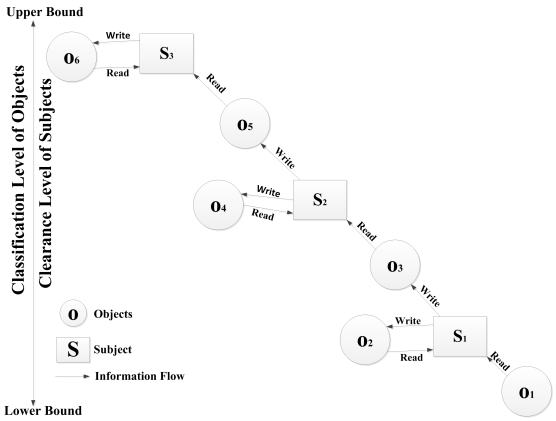

Fig. 3: Read and Write access provided by the Bell-LaPadula model

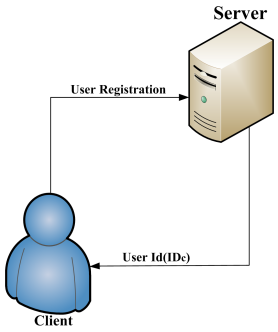

Fig. 4: Client Server Registration

So, first applying any cryptographic algorithm to text and then embedding the resulting ASCII to document. This process ensures the necessary security. The key idea behind the implementation of this technique is to embed secret message into the document with a computationally secured and time effective manner. Instead of using a high weighted cryptographic algorithm like RSA, an effective light weighted algorithm like AES can be used with authentication scheme like SHA-512[18][19]. The main focus of the proposed model is that only the registered user will be able to access the critical document otherwise non-registered user has to first register itself with the server it is shown in Figure-4. This Watermarking technique discuss how and where to place the authentication code in the critical document D. The server securely will maintain a server directory table for each registered client's id which is shown in TABLE-1. The input to the algorithm is original document D, secret message (ID_C) and 128 bit key (K) used in AES-128 encryption scheme which produce the cipher text C as an output.

In second phase of watermark embedding includes another input IV, (initial value), which is 512 bits long, and used for generating 512 bits long message authentication code M, it embedded into the document D. The process of watermark embedding in document is describe in following phases-

TABLE I: Server Directory Table

SNo.	Client-Id	SHA-512(Hash)	(m,n)
1.	ID_{C1}	M1	– – – – –
2.	ID_{C2}	M2	– – – – –
3.	ID_{C3}	M3	– – – – –
4.	—	– – – – –	– – – – –
5.	—	– – – – –	– – – – –
n.	ID_{Cn}	Mn	– – – – –

1) Phase I: Calculation of all parameters:

- Calculate the cipher text, C, by using secret message (ID_C), Encryption Key K and AES-128 encryption Algorithm. It will be implemented using block cipher techniques[20].
- Calculate message authentication code, M, using ID_C, initial vector (IV) and SHA-512 scheme. Notice that, M is generated by using secret message (ID_C), not by using the cipher text C. This will confuse the intruder, and will provide the extra level of security [19].
- Calculate positioning pixels in the document D as:
  - Row positioning pixel, m = $I(1,1)+2$
  - Column positioning pixel, n = $I(1,2)+2$

2) Phase II: Placement of cipher C and authentication code M into image:

- Replace the pixel value starting from (m, n) in the original document, with the value of cipher text C. Each block in cipher text will change exactly 16 pixel bits in the original document D.
- Replace the last 64 pixel bits, with the authentication code M calculated, in reverse order. This will confuse to intruders and provide the extra level of security. Finally the watermarked document WMD will be generated as the output for this process as shown in Figure-5.

C. Sending WMD to Client

In this phase the created Watermarked document WMD will be send to the requested client along with server's public key certificates(PKC_{server}), which verify the integrity of the genuine source of document.

This Server will use the nonce (C_{nonce}) in order to protect the man in middle attack. The send document will be encrypted with the public key of the client (PU_C) and contain the $hash_S$ which is created by the server. The process of sending WMD to client is shown in Figure-6.

Client will receive the send document and open with the help of his private key PR_C and verify the document, by creating the hash of received WMD document. If the received $hash_S$ is equal to the created $hash_C$ (ie. $hash_S$ = $hash_C$) then the document is not altered in between and document integrity is maintained[17].

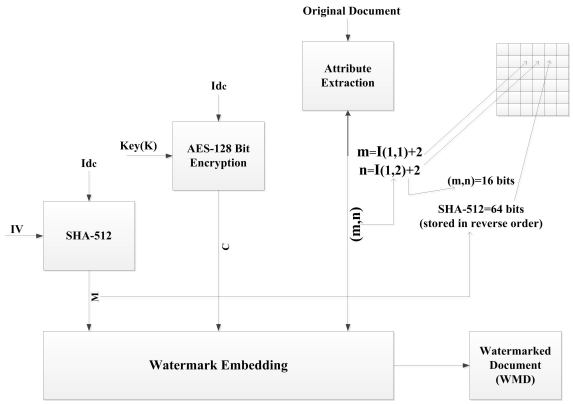

Fig. 5: Process of Watermark Embedding

Fig. 6: Process of sending WMD to client

D. Detecting the Client id

In this scenario suppose clients has leaked the document by some means and this document has value for that organization in terms of money and reputation. In this step we are focusing on detecting the client's who has leaked the documents. The watermark is extracted from the received watermarked document(WMD) by applying reverse procedure of above proposed scheme. This is shown in Figure-7 and processed in two phases-

1) Phase I: Point out the placement of cipher C and authentication code M.

- Server will use the table where point (m, n) is stored. Here we are using the same (m, n) value for all the clients and it will be secret.
- Calculate the m and n value, and then point out the starting position of C in watermark Document WMD.
- Find the authentication code M is in the last 64 pixels in reverse order.

2) Phase II: Calculation and Verification of Secret Message ID_C.

- Secret message ID_C is calculated by using K and AES-128 decryption algorithm.
- Server will verify this ID_C by matching in the stored directory table and also verify the ID_C as a correct message, by calculating authentication code M' from ID_C. If M and M' both are equal, the extracted message is correct. And the client will be identified who has leaked this data. The overall working of the proposed model is shown in the Figure-8.

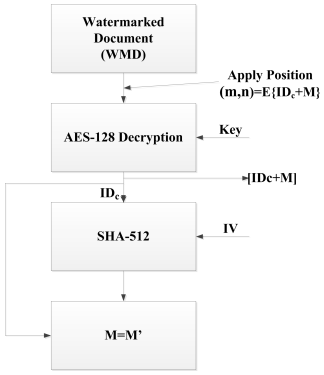

Fig. 7: The Process client id detection form WMD

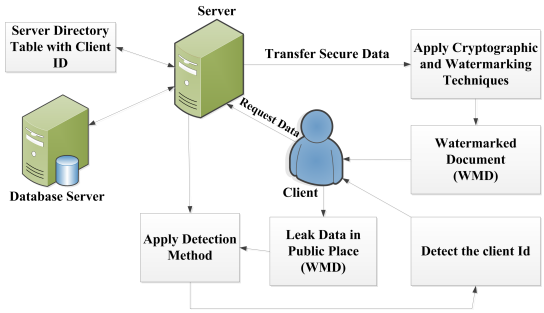

Fig. 8: Overall Working of the Proposed Model

III. Applications and Efficiency Measurement

The proposed method is efficient to use with any size of documents. Here, we have used only client ID_C with the original document. Therefore, in this case if the hidden message is short, the changes made to the original document is also short, resulting in less change in original document, and hence intruder will not be able to analyze the secret.The proposed technique can be used with the different types of image such as Binary images, Gray scale images and Color images etc. so it can be said that this technique is best suited to transmit symmetric keys in secure manner. The technique is very economical, because it uses light cryptographic algorithm AES-128 with SHA-512 to provide double security with half computational time[20][21]. If RSA is used then it leads to two major problems[22]. First its key length is very high, i.e., of 1024 bits, Second it will be more difficult to calculate exponential computations than simple computations needed in AES which uses the key of length only 128 bits long.The technique proposed in this paper uses the mixing concept to generate message authentication code, and put it in reverse order to confuse the intruders[17][22].

IV. Results Measurement

In this model we have chosen the AES model because it is faster in the process of encryption and decryption. To crack the 128-bit AES key using a well-known brute force attack it would take 1 billion years. AES is the successor of DES as standard symmetric encryption algorithm for US federal organizations. AES accepts keys of 128, 192 or 256 bits (128

bits is already very unbreakable), uses 128-bit blocks (so no issue there), and is efficient in both software and hardware. It was selected through an open competition involving hundreds of cryptographers during several years. If we compare these algorithms in terms of encryption and decryption we find that time taken in AES encryption of message over different size of the message is less than DES and RSA. It can be verified in Table-II and Table-III by comparing the encryption and decryption time of AES with DES and RSA Algorithm[20][22].In the Table-III it shows the decryption time for different size of the messages AES takes less time over DES and RSA Algorithm. In Table-IV where we analyze the different factors

TABLE II: Comparison of various packet sizes for DES, AES & RSA algorithm (Encryption Time)

Sno	DES	AES	RSA	Data Size
1	3.0	1.6	7.3	153KB
2	3.2	1.7	10.0	118KB
3	2.0	1.7	8.5	196KB
4	4.0	2.0	8.2	868KB
5	3.0	1.8	7.8	312KB

TABLE III: Comparison of various packet sizes for DES, AES & RSA algorithm (Decryption Time)

Sno	DES	AES	RSA	Data Size
1	1.0	1.1	4.9	153KB
2	1.2	1.2	5.0	118KB
3	1.4	1.24	5.9	196KB
4	1.8	1.2	5.1	868KB
5	1.6	1.3	5.1	312KB

TABLE IV: Analysis of various factors

Factor Analyzed	DES	AES	RSA
Development Years	1977	2000	1978
Key-Length (Bits)	56	128,192,256	$\leq$1024
Nature of Algorithms	Symmetric	Symmetric	Asymmetric
Encryption/Decryption(Speed)	Low	High	Medium
Nature of Security Attacks	Inadequate	Highly Secured	Highly Secured

which will shows the characteristics of the DES, AES and RSA Algorithms[23].

V. Conclusions and Future Scopes

The proposed technique will provide better security against data leakage problem. We can detect the data leaker in real time by using this method. It also protect different types of active and passive attacks. The proposed technique is computationally cost effective in terms of time and space uses. Therefore, this can be useful in distributed computing environment to protect data from data leakage. The proposed technique is based on symmetric algorithm, therefore it is infeasible to extend this model for web environment where multiple number of users frequently accessing the data object. We can also implement this technique for asymmetric cryptography.

References

[1] Rohit Pol, Vishwajeet Thakur, Ruturaj Bhise, and A Kat. Data leakage detection. *International Journal of Engineering Research & Application*, 2(3):404–410, 2012.

[2] Rupesh Mishra and DK Chitre. Data leakage and detection of guilty agent. *International Journal of Scientific & Engineering Research*, 3(6), 2012.

[3] Kenneth J Biba. Integrity considerations for secure computer systems. Technical report, DTIC Document, 1977.

[4] David Elliott Bell. Bell–la padula model. *Encyclopedia of Cryptography and Security*, pages 74–79, 2011.

[5] Mukesh Singhal and Niranjan G Shivaratri. *Advanced concepts in operating systems*. McGraw-Hill, Inc., 1994.

[6] AL Jeeva, Dr V Palanisamy, and K Kanagaram. Comparative analysis of performance efficiency and security measures of some encryption algorithms. *International Journal of Engineering Research and Applications (IJERA) ISSN*, pages 2248–9622, 2012.

[7] E Thambiraja, G Ramesh, and Dr R Umarani. A survey on various most common encryption techniques. *International journal of advanced research in computer science and software engineering*, 2(7):226–233, 2012.

[8] Bruce Schneier, John Kelsey, Doug Whiting, David Wagner, Chris Hall, and Niels Ferguson. Performance comparison of the aes submissions, 1999.

[9] Hamdan Alanazi, BB Zaidan, AA Zaidan, Hamid A Jalab, M Shabbir, Yahya Al-Nabhani, et al. New comparative study between des, 3des and aes within nine factors. *arXiv preprint arXiv:1003.4085*, 2010.

[10] Aman Kumar, Sudesh Jakhar, and Sunil Makkar. Distinction between secret key and public key cryptography with existing glitches. *Indian Journal of Education and Information Management*, 1(9):392–395, 2012.

[11] Hitendra GARG and Suneeta AGARWAL. A secure image based watermarking for 3d polygon mesh. *SCIENCE AND TECHNOLOGY*, 16(4):287–303, 2013.

[12] Hitendra Garg and Suneeta Agrawal. Uniform repeated insertion of redundant watermark in 3d object. In *Signal Processing and Integrated Networks (SPIN), 2014 International Conference on*, pages 184–189. IEEE, 2014.

[13] CISSP Susan Hansche, CISSP John Berti, and Chris Hare. *Official (ISC) 2 guide to the CISSP exam*. CRC Press, 2003.

[14] D Elliott Bell and Leonard J La Padula. Secure computer system: Unified exposition and multics interpretation. Technical report, DTIC Document, 1976.

[15] David Elliott Bell. Looking back at the bell-la padula model. In *ACSAC*, volume 5, pages 337–351, 2005.

[16] Frédéric Deguillaume, Sviatoslav V Voloshynovskiy, and Thierry Pun. Method for the estimation and recovering from general affine transforms in digital watermarking applications. In *Electronic Imaging 2002*, pages 313–322. International Society for Optics and Photonics, 2002.

[17] Stallings William and William Stallings. *Cryptography and Network Security, 4/E*. Pearson Education India, 2006.

[18] Achal Kumar and Vibhav Prakash Singh. Digital watermarking using color image processing using images for transmitting secret information.

[19] JJK RUANAIDH and T PUN. Rotation, scale and translation invariant spread spectrum digital image watermarking. *Signal processing*, 66(3):303–317, 1998.

[20] NIST FIPS Pub. 197. *Announcing the Advanced Encryption Standard (AES)*, 2001.

[21] Máire McLoone and John V McCanny. Efficient single-chip implementation of sha-384 and sha-512. In *Field-Programmable Technology, 2002.(FPT). Proceedings. 2002 IEEE International Conference on*, pages 311–314. IEEE, 2002.

[22] Jakob Jonsson and Burt Kaliski. Public-key cryptography standards (pkcs)# 1: Rsa cryptography specifications version 2.1. 2003.

[23] B Padmavathi and S Ranjitha Kumari. A survey on performance analysis of des, aes and rsa algorithm along with lsb substitution technique. *International Journal of Science and Research*, 2(4), 2013.