A LITERATURE REVIEW ON
DATA LEAKAGE DETECTION IN CLOUD COMPUTING

Dr. B. Indira
Dept. of Information Technology
Sreenidhi Institute of Science and Technology
Hyderabad, India
bindira@sreenidhi.edu.in

Divya Mary CH
Department of Information Technology
Sreenidhi Institute of Science and Technology
Hyderabad, India.
divyamarychj7@gmail.com

ABSTRACT

Cloud computing is "an Internet-based computing," where different services such as servers, storage and applications are delivered to an organization's computers and devices through the Internet. In today's era, it is most interesting technology which is offering services to its users on demand over the internet. Since cloud computing stores the data, security has become the main obstacle which is disturbing the deployment of cloud and stores their personal data. So that data storage security is required. In this paper I discussed about different techniques used before to detect the data leakage by referring number of papers. This paper analyses and explores some of the methods and techniques, how data is leaked and detected in the cloud environment. And, it concentrates mainly on information security.

KEYWORDS: Cloud computing, leakage detection, guilty agent, steganography and cryptography.

1. INTRODUCTION

As cloud computing is a model of business computing, sometimes the sensitive data must be confidential and must be shared with a trusted third party the main thing is to detect when the distributor's sensitive data has been leaked by the agents. So, to secure the data Perturbation is a very useful technique where the data is modified and made less sensitive before given to the agents [1].

The traditional method is to handle leakage detection is water marking [4]. The fake objects act as a type of watermark for the whole thing. The fake objects do not correspond to real files but they appear as realistic to the agent. If those fake objects are revealed then the distributor may become confident that the agent is guilty. Data leakage happens every day where the confidential business information and sensitive data are leaked out [12]. Now-a-days technology is always available for everyone that anybody can hack the data easily. To prevent the hacking, the security methods should be provided in the cloud which are better than traditional techniques to stop unauthorized access of data which effects on productivity and growth of the organization [13]. to ensure such high level of data security, the cryptography process is used by the service providers and the another method which helps in hiding the information where the cryptography cannot handle, the steganography technique is used [14]. We have some cryptographic techniques like 3DES, SHA1 & MD5 to secure the sensitive data. If a distributor has given confidential data to the third parties. Some of the data is leaked and found in unauthorized place [5].

Cyber-attacks are the network attack which refers to disturb and alter someone's privacy and it is the way of accessing the secret data of a country. To prevent these types of attacks we need to install a good firewall and antimalware software and devices. The most followed approach to prevent this is developing the IDS system [17]. Another technique
to provide data security is the data leakage detection using system call provenance and data leakage detection system for map reduce computational security. To address the issue of data leakage in the cloud environment stated above, my objective in this paper to review the literature on data protection & privacy in cloud, and the results of the review are presented.

2. PROBLEMS AND SOLUTIONS

Why data leakage is a problem in the cloud?

The cloud computing is the most interesting technology which offers the services like storing the data in cloud via internet. As cloud stores the most confidential business information it needs the security which disturbs the deployment of cloud. If the sensitive information gets leaked then the organisation could face the difficulties and huge loss.

Usually the distributor’s sensitive data must be shared to the third party, in the earlier days they were using perturbation technique to secure the data before given to the agents. But there is a confidentiality issue, to overcome this they approached the watermarking technique[1] which uses the fake objects to detect the guilty agent and data provenance problem.

If the information is too large it is very difficult to detect the leakage. Hence JOSE presented an algorithm which can be implemented using GPU. It converts the words into fixed size phrase to perform CRC computation in the host processor[2].

SANDIP & KULKARNI implemented a Data Leakage Detection Model to overcome the problems in the Watermarking technique which cannot gives the secure data transmission and distribution[4]. This model is implemented by improving the probability of leakage identification by developing unassuming techniques like steganography[5].

To reach the authentication and access control the ALP[authentication and leak prediction model] is introduced to overcome the privacy issues and data leakage in the condition of text mining.[6]

Bell-LaPadula model[7] is used for the secured infrastructure which is used to apply the access control and which also protect from active and passive attacks, it has used the symmetric cryptography and can also implemented for asymmetric.

To stop the unauthorized access the kernel module is implemented to allow the user to check the sending process. This novel DLP model helps the user that which data can leave their systems[8]. For the improvement of the system performance a mediated certificateless encryption scheme which solves the key escrow and certificate revocation problem, also steganography is used to reduce unauthorized access[9].

3. LITERATURE REVIEW

“Papadimitriou Hector Garcia-Molina”. In this paper, it is described that how the sensitive data get leaked and how to detect the guilty agents. It presents algorithms for distributing objects to agents to identify the leaker. The author evaluated the strategies in different data leakage scenarios [1], this paper mainly focuses on data allocation problems that how the distributor can share the data with agents. The four instances of the problem that are depending on the type of data requests and whether fake objects are allowed, and the data requests are of explicit and simple. The use of fake objects has come from the trace records in mailing. These can help to identify improper use of data. This paper evaluates the algorithms and simulates the data leakage problems using python, describes the allocation strategies.

“Jos Kolenchery Digital Surgicals inc Austin,USA”. This paper states the technique which is used to detect virus patterns and data leakage in messages. Depending on the length of the message, the complexity of detecting the leakage is increases.
The web security is essential to secure the data. In past, the traffic inspection appliances and application security software installation was done. Web security involves pattern matching on different web and email traffic. In this paper the cloud based web security service architecture for fast phrase matching is presented [2]. It is a different technique to secure the cloud services.

"Iliana Lankoulov, Maya Daneva". This paper describes the implementation of data leakage detection model. As the water marking technology does not provide the complete security, it includes the comparison of the previous watermarking and the model of the data leakage detection system. And the different techniques of water marking system are discussed [4]. And also different 5 modules of data leakage detection system are explained. The results of these two models are compared, and the guilt model is analysed by explaining the impact of probability. Mainly this model provides security as well as tracking system is developed.

" Sandip A.kale, Prof S.V Kulkarni". This paper attempts to add the privacy preservation with leak detection by the authentication and confidentiality to solve the risk of the privacy. In this paper, the documents of information are clustered in the format of text mining; they are represented in trees or graphs. To provide authentication, the Redactable signature scheme is used. To the leakage free redactable signature generation module [5], the clustered documents of trees will be given as input, atlas they are displayed to the cloud user. In this paper, they have used four algorithms to generate keys and signature verification. This model helps to provide security to cloud service.

"Ms.N. Bangar Anjali, Ms.P. Rokade Geetanjali, ms.R. Shekhar Swati, Prof N B Kadu. " This paper contains the implementation of data leakage detection model. This paper deals the secured data transmission and leakage detection [4]. It describes about providing security to our data, by adding fake objects to the actual file the leaker would be detected. In this paper, the data allocation strategies are proposed to improve the identification of leakage. His algorithm used to implement the distribution strategies, which can leads to chance of identifying the guilty agent.

"Mohamad Farhatullah ". The aim of this paper is to find out the culprit who has revealed the confidential data of an organization. In this paper, the Bell-La Padula security model has been used which is known as data confidentiality model his Bell-la padula model is built in contrast to the Biba Integrity model which are mainly focuses on confidentiality and integrity respectively it is alson used to apply the access control to provide secure infrastructure. In this paper [7], the author analysed the watermark[4] and watermark document is created and send to the requested client along the server's public key, then client will be received the document and opens with his private key. If any client has leaked the data then we can detect them by doing the reverse process of watermark. The author selected the AES model algorithm because it is very efficient in software and hardware. It is completely based on symmetric algorithm which gives better security. It takes more time and space can be useful in distributed computing and also implemented for asymmetric cryptography.

" Ryan K L Ko, Alan Y S Tan, Ting Gao ". This paper shows the new ways of detecting data leakages occurred it and user side while having the cloud services. the authors have used different approaches of DLP they proposed am mantrap approach to DLP to prevent data leakage from the kernel systems of all systems. His paper describes the implementation of the physical mantrap access control techniques in the Linux kernel by designing the kernel space mantrap dlp architecture. To achieve this they have modified the udp-blocking tcp-blocking which are used to capture the data or message and allow the message to allow or not. so it notifies users to allow or block the sending of data everything is in users control. If the user may not know the outgoing format than the content detection technique would be approved. Than the user may able to send the more message if more information is provided.

" Louai A. Maghrabi ". In this paper [9] the author listed out the all the threats of the data security over the cloud. It describes how many types of threats are there in the security and it express the security issues which are mainly related to privacy confidentiality, integrity of the sensitive information of the person or the organisation. Finally paper stated that as the
rapidly emerging computing the new issues may appear as increase of cloud services to the user.

"A. Madhuri, T.V. Nagaraju". This paper [10] mainly concentrates on the issues related to security of data of cloud computing. Basically the cloud service provides shares the data to rusted agents [1] to protect the author proposed a system architecture which provides the data integrity and service availability and it states the security risks in cloud computing environment like access network load, data security, data location, segregation etc. The author used the secret sharing algorithm to overcome risks above mentioned and it explains about DEPSKY MULTI- CLOUD MODELS which addresses the availability and confidentiality of data in the data storage system.

"Mohis M, Devipriya V S ". Many of cryptographic technique and traditional methods were used the ore security provide confidentiality and access contralto cloud but they to failed [1] to address the confidentiality and privacy issues. In this paper[11] the author proposed a system with modified certificates public key. Encryption more security and a Steganographic method which provides inside the cloud. In this system, the user enciphers the data before uploading in to cloud. As the steganography technique is used for hiding the data under the image. Some of the data embedding methods are introduced to give more security inside the cloud. This method solves the key escrow problem and embedding module with steganography will overcome the unauthorized access of attacks on confidential information.

“Prof. Sushilkumar N. Holambe, Dr. Ukhas B. Shinde, Archana U. Bhosale”. This paper presents data leakage detection system using diffrenet6 allocation strategies [12] and it allows only trusted third parties. It detects sensitive data to be not leaked by adding fake records [1,4] The authors developed a model for finding guilty agents [5] by the data allocation strategies and calculating the guilty probability to sensitive information.

“Vinay kumar pant, Jyoti Prakash, Amit Asthana”. This paper describes how to protect data by using the Cryptography band steganography techniques .In this paper, the author used the three steps data security model as it uses the RSA algorithm and secondly steganography techniques to hide data within the image. It may be also applicable to video, audio or text. To reach the confidentiality, data integrity and access control [9, 7] and multi tenancy to protect from cyber-attacks.

“Ankit Dhamija, Research Scholar, Prof Vijay Dhaka”. This paper proposed a design for cloud architecture which secures the sensitive data which shared from an organisation and to the servers of the CSP. It presents the combination of cryptographic and stenographic techniques. It deals with the data migration in the cloud. In the previous techniques the message will be integrated twice instead of using single level encryption algorithm which can hide large amount of data [13]. To improve, this paper proposed secure cloud migration architecture using cryptography and steganography [SCMACS]. It consists of five steps which are equally partitioned as cryptographic and steganography process. To encrypt the user files they have used an effective approach of encryption using ASCII codes as it becomes cipher text. To approach this paper proposed Least Significant Bit method in which bits are located in image pixels. Each pixels have presented the red, green, blue elements, the text will be hidden the inside the image for each integer the RGB values get separated and LSB equals to zero. When 8 bits of characters are processed it jumps to next character and process is repeated is until the whole message is hidden. The extraction of message from image carrier as constructs back. It’s very safe to use this approach for sensitive data.

“Chandu Vaidya, Simran Said, Suyog Chadawar”. The method, which this Paper presents the data leakage detection and dependable storage service in cloud computing. Which is simple and which serve the organizations and clients. Which analyses the problems related to the cloud and data leakage and integrity problems. To prevent these things, this paper proposed and flexible scheme by using hashing mechanisms. It uses the MD5, SHA with 3DES algorithm. The main focus of this paper is correct an identify the data correction.

"Er. Nisha Yadav, Dr. Amit Sharma ". This paper [16] focuses on security of data in cloud to protect
the information by unauthorized access misuse, denial of service attack. This uses CHAP protocol and RSA algorithm to provide authentication. This paper proposes model of data security which includes authorization through EAP-CHAP protocol.(Challenge handshake authentication protocol) encryption or decryption by RSA algorithm.

"Saeed M. Alqahtani, Robert John". To detect and prevent of a network and deal with millions of users per second several Intrusion Detection Systems have been developed. This paper [17] used the SNORT & SURICATA IDS systems. The main aim of this paper is to compare the functionality of this both IDS systems to detect the cyber-attack and this paper also proposes a fuzzy logic engine to reduce the false alarms and study is done by using the testing ISCX data set. IDS systems are used to monitor the network traffic to destroy the cyber-attack which produces the alarms against this attack. I this ISCX Data Set and My Cloud IDS systems are described. If the high false rate is there then IDS fuzzy classifier will be built. The SNORT IDS or SUDICATA IDS generates the alerts and takes the sensitive decisions to overcome the malicious traffic. It also compares the SNORT IDS and SURICATA IDS. And FL-SNORT IDS is better than SNORT IDS. FL-SURICATA is better than SURICATA IDS and finally the SNORT IDS is better than SURICATA IDS. The performance of the system gives the higher level of accuracy, sensitivity. By fuzzy technique unwanted alerts are removed. Its states, in future the FL based IDS with genetics algorithm may implemented as the SNORT IDS and SURICATA IDS systems may not be able to understand the changing the networks and attacks.

"Abir Awad,Sara kadry,Gururaj Maddodi, Saul Gill, Brian Lee". In this paper, the design and implementation of PEEPER (Provenance and policy based data exfiltration detection) is uses the data leakage which uses the operating system called provenance [18]. A Linux kernel module is used [8] to maintain authenticity and integrity. They are complex to setup and operate as all may not work in real time so this paper contribute to design and implement and approach to enable the real time data exfiltration and detection using open source software. The architecture proposed is based on data centric approach [8]. It focuses to overcome the drawback of the already proposed models that is for each operating system the user profile may vary so data tracking using systemically provenance the peeper-data exfiltration detection is to detect the unauthorized movement of confidential data real time. The proposed system consists of collection, correlation and cogitation which is implemented using progrger, the open sourced code which records the time and date and the correlation component which generates the meta events which is implemented using open source engine SEC and cogitation states which determines how the files are manipulated. Data leakage is considered within the same machine the system called data records by progrger to rename the call via named pipe and they have implemented the SCP(Secure Copy Encryption) and the paper states to explore how paper can be combined with software defined networking in future.

“Sakshi Chhabra, Ashutosh Kumar Singh". This paper proposed a method which distributes the load of data i the chunks and the data undergo map reduce by filtering and reducing the data with hadoop and framework. The data can be reduced up to 70% which will build the secure by using the data reduce detection. By using the S_MAX algorithm which gives the probability of > 0.4 to find the guilty agent. The aim of this paper is to secure the reduced the data and the data is leaked. It shows the weather forecasting data which accumulated from the website of Government of India. The author evaluated with the help of map reduce, hadoop which reduce the data from GB to MB and using the S-MAX algorithm the guilty agent is detected.

"Alok Rajan, mansi Bhonsle". The main focus of the proposed system in this paper [20] is to avoid information access from cloud data storage centres by unauthorized agents. This paper presents the technique which is more focused and the idea is, before the data is travelled over the network the user data will conceal inside another media file and the encryption is applied. As the system is a multi-layer system user may save data as per the priority. If the information is more important than it will be hidden in images or audio file. Otherwise uses cryptography [11][14]. To implement the system they have used
java, html file, My Sql and TOMCAT for web container. For testing it is used METALAB. The proposed technique uses the H_LASB to hide the message [14]. It is a unique system which has tested with illegal access and hacking which gives more security than existing systems. It does not offer steganography technique with AES in cloud to save and share data. It states, in future the mobiles going to facilitate the service.

“Barjesh Kochar, Sanjiv Saxena, Archana B Saxena ”. This paper provides a new aspect and proposes an attempt to develop a mathematical model for assess the impact of loss due to data leakage in cloud [21].

“Savitha D Torvi, Dr. K B Shivakumar, Rupam Das ”. As the stenographic technique helps to hide the data using the image/audio or text. The use of image carrier requires more storage. Hence, in this paper [22] they proposed a unique data security using text steganography (UDSTS). Which is able to transmit and receive the messages hidden in text format as *.DOC*,*.RTF* E-MAIL/MESSAGE BODY.etc. The XOR encryption is done to encrypt the text with password the plain text is in ASCII format. Using font based and colour based steganography the text file will be hidden with same number of characters with fake text. Font based technique is for lower message size and colour based is quite faster than it.

" Kang II Choi, Jung Hee Lee, Bhum Cheol Lee ". This paper [23] presents a novel method to secure the privacy information for the cloud based video storage systems (CVSS). This paper presents how the video stream data is transferred from the network connected carrier (IP CCTV) will be stored. The paper proposed an approach where the actual video is encoded and decoded which is stored the subscribe key even the system is under hacking CVSS; still it maintains the privacy protection of the video data. The CVSS contain the outgoing video data protection (IVDDP) and (OVDP). The IVDPP block creates a secured image for applying PP algorithm vice versa to the original image. Flowcharts of received and sent video image the CVSS creates a privacy image and transfers the original image to network camera control centres. DCVSS is a distributed system which handles the group of IP cameras.

The separate privacy protection key algorithm is applied and geographically separated DCVSS. So that if DCVSS hacked id doesn’t affect the other DCVSS. In receiving the decryption process is done. By this we can prevent the huge data leakage. In this the DCVSS with cloud based network function virtualization system also presented and implemented.

" Md. Habibur Rahman, Nazrul Islam, Mehedy Hasan Rafa Jan, Shariful and Mohammad Motiur Rahman”. This paper implements the double stage encryption algorithm for multimedia content security with random key generation approach[24].It is an advanced approach, to analyse the issue that the regular side channel attack for hacking the sensitive & confidential image and audio or video . In this method, the encryption process is done twice. At first the multimedia message is encrypted by the DES, AES&RSA using symmetric key. And then the cipher text 1 is then again encrypted using randomly generated symmetric key. The decryption process is also done in two ways. By this the chances of exposition of key is very low. This paper mainly focuses to improve the security of multimedia user data. The encryption process is described in pseudo code and it is decrypted in decryption. According to the size of multimedia data takes different time to generate the different cipher text. This provides the more security for the multimedia data in cloud. It is too hard to hack the multimedia data a of cloud for any unauthorized person even for the system administrator also.

CONCLUSION:

The objective of writing this paper is to summarize an overview of the data leakage detection in cloud computing, and it gives the brief summary of the techniques used for the data security on cloud computing referred from the papers or journals. The main purpose is to know how to detect and prevent to overcome data security threats. This paper discusses the different models and methods how to get rid of security issues and to make the cloud services efficient and lossless.
REFERENCES

THE IMPACT OF LOSS OF TRUST DUE TO DATA LEAKAGE IN CLOUD. 2016 International Conference on Computational techniques in Information and Communication Technologies (ICCTICT).

[22] Savitha D Torvi, Dr. K B Shivakumar, Rupam Das. AN UNIQUE DATA SECURITY USING TEXT STEGANOGRAPHY. 2016 International Conference for Sustainable for Global Development (INDIACom)
