RESEARCH ARTICLE OPEN ACCESS

Proposed Approach for Detection of Suspicious Activity using Provenance Data in Cloud Environment

Minal Bharat Pokale, Dr. Sandeep Chaware
(Computer Engineering, Savitribai Phule Pune University/MMCOE, Karve Nagar, Pune)

Abstract:

Cloud computing is one of the powerful computing that connect the whole world. Huge amount of data get uploaded over cloud so that it can be accessible by the user at any time and any point. Security must be concerned related to confidentiality of data, integrity of data, availability etc. Data stored over cloud is physically not accessible to the user. Data modification can be done by unauthorized user or by some malicious activity so user needs to be ensure that their data is secure or not. Therefore mechanism is required where user can check if integrity of its data is maintained or compromised. There are various methods are available such as mirroring but it require more storage space. Sometimes we need TPA to verify the data. In this, user check the integrity of its data store over cloud environment and detect the suspicious activity. For this new scheme "Data Provenance" is used, that collect the history data of user. On the basis of which, suspicious activity is detected. Provenance data is used to collect the history data and identify the user behaviour. Using this we reduce the need of TPA and replication of data item on client side for integrity checking as checking part is done over cloud infrastructure.

***Keywords* — TPA (Third Party Auditor), Provenance Data, SHA (Secure Hash Algorithm), Cloud Server.**

I. INTRODUCTION

Cloud Computing changed the world around us. Since data is getting bigger and needs to be accessible from any point on any devices, people moving their data to the cloud environment. Therefore storing data to the cloud is gaining popularity. Cloud Computing is nothing but the virtual pool of various resources. Centrally all the information of customer is stored over cloud so that it accessible from anywhere. As cloud computing is pool of resources, these resources are offered to the user via internet. Cloud computing offers storage services. In recent year, storage in cloud computing gained popularity among both companies and private users. However the issues such as availability, confidentiality, reliability and interoperability is need to be considered. But the most important point that need to be considered is security how organizations gives assurance about the security of the document or information that stored by user over cloud. Cloud Computing is an important concept in computer development. This concept refers to the use of capacity and storage of computer and servers over the internet. Hardware Software resources used by user at remote locations is managed by third party. Examples of cloud services are online file uploading, storing, social networking site, google drive, web mail etc.

Clouds can provide many types of services like applications (e.g., Google Apps, Microsoft online) ,infrastructures (e.g., Amazon'sEC2,Nimbus), and platforms to help developers write applications (e.g., Amazon's S3, Windows Azure).We store sometimes sensitive data over cloud, for example, medical records and social networks. The user validity is who stores the data is also verified. The cloud is not totally secure because anyone can do modification and violates the integrity of the data store over cloud. So the data must be stored in the form that is not readable to the outsider or unknown person.

II. LITERATURE SURVEY

Maintaining security in cloud computing is basic objective of most of the researches. TPA is used to verify the integrity of data. So many techniques are used such as checksum, mirroring for error correction and replication of data to maintain the integrity. But all these techniques have some advantages and disadvantages that consider in the following survey.

Muhammad Imran, Helmut Hlavacs, Inam Ul Haq, Bilal Jan, Fakhri Alam Khan,Awais Ahmad, proposed a scheme which check the integrity using provenance data and verification over cloud environment[15]. There must have a mechanism that check integrity of user data is violated or not. In this article author reduces the need of TPA. Auditor is not required for verification of user's data. Author proposes a scheme in which user is able to check the integrity of its data over cloud storage. For this it uses new concept of data provenance. It reduces the need of TPA and replication of data item on client side.

Bhale Pradeepkumar Gajendra, Vinay Kumar Singh, More Sujeet, proposed system in which for securing user document Identity Based Encryption technique is used and for generating respective hash value of the data store over cloud using MD5. Three entities are involve in this client, cloud admin, Third Party Auditor. There is no practical method for authentication of the user, It is higher exposure to the risk and it is less secure[16].

Sultan Aldossary, William Allen, proposed a scheme in which it listed out the issue such as data loss,malicious insiders,insecure interface and APIs,account and service hijacking, data location and denial of service. For this issue what are the possible solution are also listed out. For example using strong API for access control, analyzing data during runtime, encryption technique to make content unreadable, making strong access control and authentication. When data is transmitted preventing user from sharing their sensitive data, using two factor authentication[1].One issue need to be considered some organization need their system to be available all the time because available is important as their provide some critical services. Cloud service providers offers resources that are shared among many client. If an attacker uses all the resources, then it is unable to available to others so it leads to DoS and slow the system performance.

Akashdeep Bhardwaj , GVB Subrahmanyam , Vinay Avasthi , Hanumat Sastry, proposed a scheme in which author proposes various security algorithm and point out which one is having a good performance. In this security algorithm for cloud such as asymmetric algorithms, symmetric algorithm consider. In asymmetric algorithm two different keys are used one is public and another is private. Using public key all the documents are encrypted and only correct private key is used to decrypt the file. It includes algorithms such as RSA, Diffie-Hellman and ECC[2]. In symmetric algorithm single key is share among multiple users to encrypt and decrypt the data. It has somany drawback once the attacker know the key it can hack the whole the file. It includes AES,3DES,RC6, Blowfish. In this author focuses on the symmetric algorithm. Author determined performance of various algorithm on the bases of time required for reading the file, encrypting,creating,sending and receiving the file. It is found that AES and MD5 has good performance.

Charmee et.al. Proposes a scheme in which it can give the various integrity checking method for remotely storing data on cloud. Provable Data Possesion (PDP),Proof of Retrievability (PoR) ,MD5 based,Encryption Algorithm,RSA based, generating meta-data are all integrity checking methods[3]. Author also mention drawback of this technique such as data recovery is not supported, cannot be used in original form, no public auditability, no support for dynamic operation, it is observed that various methods are available and can be used on the basis of file size and requirement.

Swapnali et.al. Proposes a scheme in which author signifies that users must be concerned about the integrity of its data as user data can be modified or attack by the outside attacker[4]. Therefore new scheme is used for auditing the file and checking integrity of the data stored over cloud using Third Party Auditor. Auditing scheme make used of AES, SHA and RSA algorithm. It cannot audit the file dynamically and whenever user send challenge at that time only the file is audited.

Muhammad et.al.[5]proposes a scheme in which it becomes challenging to manage provenance information because of the relationships that exist within Cloud layers and the creator object.Cloud computing follows a layered architecture where each layer targets a particular domain of end users.In such a layered architecture, provenance (the metadata that describes the derivation history of the object) of the individual layers is of significant importance to establish trust and authenticity[5]. In a typical Cloud environment, each layer provides important provenance information which usually targets a particular domain of clients e.g. Cloud provider uses infrastructure provenance to track resource utilization.

Muhammad et. al. [6]focus of this paper is provenance data for Cloud IaaS .Cloud providers can optimize resource utilization and energy consumption by finding patterns in their usage. One way of finding such patterns is to study the history of Cloud resources activity. This approach is known as Cloud provenance. Provenance can also be used to track errors and faults in Cloud services. In this author develop provenance framework for research cloud in order to find the history of the resources usage.

Adam Bates et al. [9] provide a mechanism to use provenance as an access control for cloud environments. However, their work assumes that provenance meta-data is provided by end hosts. In case of incorrect provenance, the system might suffer critical problems. In our work, we automatically collect and manage provenance meta-data inside the cloud for the different layers. We also focus on the violation of data integrity and its verification.

Juels et.al.focuses on the existing schemes of data integrity in cloud such as Provable data Possession (PdP) [13], Proof or Retreivabilty (PoR) [13], High Availability Integrity Layer (HAIL) [14], and using Third Party Auditors rely on methods like key generation algorithms, cryptographic techniques and replication of data. Schemes in the PdP and PoR category work on file orblock level and require computation overhead because of key generation algorithms. Schemesin the HAIL category rely on replication of data items which adds to the huge storage overhead.TPA adds privacy issues regarding data because of the involvement of third party[12].

Limitations:

1. Existing scheme require high computation overhead because of key generation algorithm.
2. Data loss, malicious insider, service and account hijacking and denial of service all these issues are come and need different solution to solve the above problems.
3. Use of third party to check the integrity is quite expensive for user as user need to pay for that,it does not audit dynamic data also data recovery is not supported.
4.It cannot detect the modification, modified data is consider to be the part of original data, checksum does not provide which series of the data is corrupted ,it does not have capability to verify the violations, it does not consider deduplication ,static data is required, data is encrypted before uploaded to the server.

III. PROPOSED SYSTEM

In proposed system, user first register itself over cloud server. User can upload the file over cloud server. On that file user can perform various activities such as insert, delete, update, modify. These data of user is stored over one centralized database in cloud environment. Proposed system uses the new scheme"Provenance data" in which history data of user is collected. On the basis of which suspicious activity is detected. Supposed regular used over cloud server upload his sensitive data using one IP address. If same user is uploading the data using same IP address but there is no suspicious thing is recognized then we cannot say it is thread or hacker.

For example if we collect user's history data and observe that it can delete its file mostly in evening, making update to the file mostly in afternoon and inserting the file in morning. From this data we make analysis that user does not do any changes to the file in midnight. Supposed in midnight that file is get hack or corrupted then that activity is may be suspicious. Origin of that thread is find out. From which IP address that changes get made is tracked and blocked that IP permanently.

In proposed system we reduce the need of TPA and replication of data item over client side. Integrity is maintained by calculating hash of the data store over cloud. If some file is get alter then hash of that is automatically get changed from which we can say that the file is corrupted. For maintaining integrity we can make used of secure has algorithm. Proposed system is reduced the need of TPA. Everything is done over cloud infrastructure so we don't need to pay extra charges for TPA.

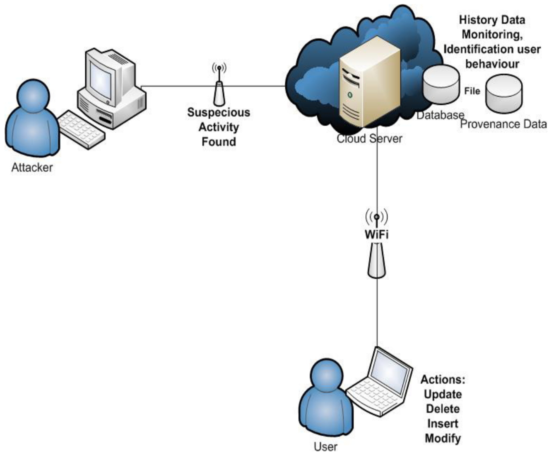

Fig : Detecting Suspicious Activity using Provenance Data over Cloud Environment

The System is designed to detect suspicious activity if any attacker modify, delete, and insert data of user store over the cloud server. This suspicious activity is detected. This activity is detected on the basis of IP. That IP is get blocked by the server. Suspicious activity is detected on the basis of user behaviour, for this history data of user is collected for this provenance data scheme is used. On the basis of which the malicious activity is detected. For that various malicious traffic, detection and prevention techniques are available such as anomaly detection technique, signature scan technique, intrusion detection and prevention system, quality of service matric. Differentiates between normal and malicious activity is first studying the normal behaviour of users, resources, second create pattern for these activity and third any behaviour that deviates from this pattern is considered malicious. Integrity of data stored over the cloud server is maintained by generating hash of the file.

Features of proposed architecture is as follows

1. Securely send files. User authentication is done using generating OTP

2. Observing user's behaviour, suspicious activity is detected and blocked by the server.

3. Ensure the integrity of sensitive data.

4. Everything is done over cloud infrastructure so there is no need to pay for third party to verify the integrity of the data store over cloud.

Proposed system architecture has the following components:

User or data owner: The task of user is to insert, delete, update and modify file that is uploaded by it over cloud server

Cloud Environment: It collect user history data. Observe the user behaviour and find out the suspicious activity over it. If suspicious thing is detected by the server then server blocked that IP address permanently. Server keep the track of IP address from which regular user make changes in its file. Data provenance is used for collecting history data, origin of that data. Integrity is also maintained over cloud by using SHA algorithm.

Attacker: This module is kept, if it make some changes in file over cloud then that activity is track or observe by the server and blocked that IP.

Provenance is used in different domain by scientists and researchers to trust, track back, verify individual input and output parameters to services and sub process information,

IV. ALGORITHMIC STEP USED

Algorithmic Flow of proposed system:

```
Step 1: User Authentication
If(User==verified)
{
User will get access
}
Else
{
Do not get access
}
```

Step 2: Provenance data over cloud infrastructure (History data its origin),user behaviour is identified
Step 3: if suspicious activity found, blocked that IP by server
Step 4: Integrity of store data is maintained using SHA, which calculate the hash value to keep the data secure. If data gets changed the hash value is also changed.
Step 5: Data is verified and suspicious activity is blocked by server.

User Authentication Algorithm

Step 1: User first register itself, enter name, password and email.
Step 2: login using username and password.
Step 3: verification on server side.
Step 4: Authentication and require OTP.
Step 5: OTP generation using user information
Step 6: Login with OTP
Step 7: Generate OTP on server side using user information that is store over cloud server
Step 8: Verify OTP if OTP is matched then it is valid otherwise not a valid user
Step 9: OTP login successful.
Step 10: Login in website by username and OTP.

V. CONCLUSION

We introduce the concept of provenance data in cloud environment. Using this approach we can keep the user data securely and by observing user behaviour, system can find out the suspicious activity. Using provenance data can find out origin of that activity and blocked that IP permanently. Using this system there is no need to used TPA, no need to pay extra charges for auditing purposes. As everything is done over cloud environment space requirement is also reduced.

In future we can enhance this to use the load balancing concept to reduce the burden over cloud as data gets increases day by day.

REFERENCES

1. Sultan Aldossary,William Allen,"Data Security, Privacy, Availability and Integrity in Cloud Computing: Issues and Current Solutions"IEEE,Vol. 7, No. 4, 2016

2. Akashdeep Bhardwaj , GVB Subrahmanyam , Vinay Avasthi , Hanumat Sastry,"Security Algorithms for Cloud Computing",sciencedirect 2016.

3. Charmee V. Desai , Prof. Gordhan B. Jethava,"Survey on Data Integrity Checking Techniques in Cloud Data Storage",IJARCSSE,Volume 4, Issue 12, December 2014.

4. Swapnali More, Sangita Chaudhari,"Third Party Public Auditing scheme for Cloud Storage"ScienceDirect 2016.

5. Muhammad Imran , Helmut Hlavacs , Fakhri Alam Khan , Saima Jabeen ,Fiaz Gul Khan , Sajid Shah , Mafawez Alharbi,"Aggregated provenance and its implications in clouds",Elsevier Nov 2017.

6. Muhammad Imran,Helmut Hlavacs,"Provenance Framework for the Cloud Environment (IaaS)".

7. Sivathanu G, Wright CP, Zadok E.,"Ensuring Data Integrity in Storage: Techniques and Applications. In: Proceedings of the 2005 ACM Workshop on Storage Security and Survivability. StorageSS"05. New York, NY, USA: ACM; 2005. p. 26–36

8. Patterson DA, Gibson G, Katz RH. "A Case for Redundant Arrays of Inexpensive Disks (RAID)", SIGMOD Rec. 1988;17(3):109–116.

9. Bates A, Mood B, Valafar M, Butler KRB. "Towards secure provenance-based access control in cloud environments," In: Bertino E, Sandhu RS, Bauer L, Park J, editors. CODASPY. ACM; 2013. p. 277–284.

10. Cheng-Kang Chu, Sherman S. M. Chow, Wen-Guey Tzeng, Jianying Zhou, and Robert H. Deng, Senior Member "Key-Aggregate Cryptosystem for Scalable Data Sharing in Cloud Storage" IEEE Transactions on Parallel and Distributed Systems. Volume: 25, Issue: 2. Year: 2014.

11. Perrig,"Practical techniques for searches on encrypted data," In IEEE Symposium on Security and Privacy. IEEE Computer Society, Washington, DC, 44–55,2000.

12. Reza Curtmola, Juan Garay, Seny Kamara, and Rafail Ostrovsky,"Searchable symmetric encryption: Improved definitions and efficient constructions", In CCS. ACM, New York, NY, 79–88,2006.

13. Juels A, Kaliski Jr BS," PORs: Proofs of retrievability for large files. In: Proceedings of the 14th ACM conference on Computer and communications security", Acm; 2007. p. 584±597.

14. Bowers KD, Juels A, Oprea A. "HAIL: A high-availability and integrity layer for cloud storage. In: Proceedings of the 16th ACM conference on Computer and communications security," ACM; 2009. p.187±198.

***15.**Muhammad Imran, Helmut Hlavacs, Inam Ul Haq, Bilal Jan, Fakhri Alam Khan,Awais Ahmad," Provenance based data integrity checking and verification in cloud environments",ResearchArticle,may 2017*

***16.** Bhale Pradeepkumar Gajendra, Vinay Kumar Singh, More Sujeet ,"Achieving Cloud Security using Third Party Auditor, MD5 and Identity-Based Encryption ",IEEE,2016*