ISSN:0975-9646

Zero private information leak using multi-level security and privileged access for designated authorities on demand

Syama BabuRaj[1], Pretty Babu[2]

Dept.Computer Science & Engg. , Sree Buddha College of Engg.
Alappuzha, Kerala, INDIA
[1]MTech Student, [2]Assistant Professor

*Abstract***— Big Data is the term for a collection of large and complex data sets that cannot process using traditional data processing application. Security and granular access control are the main issues in this. Security means to protect the information from attacker and access control means only legitimate user who has the access rights can access data. Encryption is one of the techniques for providing security. But ordinary encryption and access control mechanism is not feasible in big data because this doesn't provide much security. In order to provide security with access control, multilevel access control is incorporated. In this scheme, access to the database is based on the level of users. Also master key scheme is used to protect the private key and private information. One disadvantage in master key scheme is that the authorized authorities cannot access database even with court search warrant. To overcome this, key splitting method is introduced here. This scheme provides privileged access for designated authorities. Also revocation list is maintained in the database to avoid unnecessary access when the user is revoked.**

Keywords—Big data,Security,Access control,Encryption,Master key,Key splitting,Revocation list

I. INTRODUCTION

Values of quantitative or qualitative variables, belonging to a data item are called data. Database contains collection of data. Collection of large and complex data sets that cannot process using traditional data processing or on-hand database management tools are termed as big data. Big data consist of large amount of data by companies or government about us and our surroundings. Big data size is constantly growing and most of the data captured today is unstructured. Some of the examples of big data are data collected from sensors used to gather climate information, digital pictures and videos, posts to social media sites, purchase transaction records, and cell phone GPS signals. Big data includes data sets of large sizes beyond the capability of commonly used software tools to capture, curate, manage and process the data within an acceptable time limit. It is Big data is difficult for the big data to work with using relational database management system and desktop statistics. Security is one of important challenge. Other challenges are capture, curation, storage, search, sharing, analysis and visualization.

Big data is characterized by three dimensions which is also known as 3V's. These 3V are volume, variety and velocity. Volume of data exceeds traditional processing since large amount of data is created every day. As the data formats vary by video, location of user and modes of interaction the variety of data is huge. The speed of data flowing in and out of the system is referred as velocity of data. Big data is data that is distributed in storage and parallelized in processing. If the data fits in an Excel Spreadsheet, it is small data. If the data fits on a single traditional server like MySQL, it is medium data. If the data is so large and spread across multiple servers, it is big data.

Security and access control are the main important challenges in big data. Security means to protect the information from inside and outside attacker. Access control means only legitimate user who has the access rights can access data. Normal encryption and access control scheme cannot used in the case of big data because they doesn't provide much security. So master key scheme is introduced. But using this scheme authorized authorities cannot access database even with courts search warrant. In order to avoid this problem, master key splitting scheme is incorporated.

II. RELATED WORKS

Security and granular access are important challenges in big data. To provide several securities, mostly we prefer encryption. But normal encryption scheme is not feasible, because if any attacker gets the key for encryption they can easily decrypt the data. To avoid unauthorized access of data from hard disk, disk level encryption [1] scheme is used. Here encrypted data is placed in the disk and whole data portion of hard disk is encrypted. This avoids the outside attackers. Each time authorized access need to decrypt the whole data part of hard disk, after that it again need encryption. As a result number of encryption and decryption is increased.

In mixed cryptographic technique [2], sensitive data is protected from attacker even in multiple levels. It provides secure data storage but doesn't provide access control and key management. In encryption using biometrics [3], both key and biometrics (fingerprint, hand geometry, iris, signature etc) are need. Key alone cannot be used for decryption. If any attacker gets the key he cannot decrypt it. Biometric is needed along with the key for decryption. Here problem occurs when the biometrics are stolen by attacker. Also devices are expensive.

In Stored data separation [4], sensitive data is protected by providing strong access control on user data and isolate code running on behalf of different user. Its architecture include Web stack (virtual web server). Client contacts the application using SSL with the help of dispatcher it assign a web stack. User authenticator authenticates the user based on the ID. Query restrictor restricts the view of the database based on the ID. Access to the database is based on the ID of the user. Thus access control can be given but it does not provide much security. In secure storage through authenticated encryption [5], Initialization Vector (IV) is used and computes MAC of the data using this IV. After computing the MAC, data is encrypted and stored. When the same data is fetched, it is deciphered and computes MAC using IV. Then compare the stored MAC with computed MAC. If the values are same, then they offer security.

In multi level access control [6], divide the user according to their access level. Top secret user is the highest level user who has access to the most restricted data. Unclassified user is the low level user who has access to public view of data. Secret level is the middle level. Security is provided by allowing multi level access control. Each user level has its own public and private key. Using private key, encryption is done and using public key (present at each user) decryption is done. Normally this key is present in the database. But here master key is used. All private keys for different levels and private information are encrypted using master key. Master key is stored in master key server not in the database. Only encrypted data is stored in database.

PROBLEM STATEMENT

For providing access control and security, multi level access control is used. Access to the database is based on the level of user. In this master key scheme is incorporated. In this scheme all keys and private information is encrypted using master key for providing security and master key is stored in master key server, not in database. Master key server is a trusted server. So security administrator and law enforcement authorities cannot access data even with court's search warrant. Once the key is given to the user it cannot taken back.

III. Proposed Scheme

For providing access control, we use the architecture in multi level access control. Fig.1 shows the architecture for multi level storage. This architecture contains three levels (L1, L2, and L3). Each level has its own private key (R_{L1}, R_{L2}, R_{L3}) and public key (U_{L1}, U_{L2}, U_{L3}). For providing security all private key along with public key of level L1 and user private information is encrypted using master key. L1 data and U_{L2} are encrypted using R_{L1}. L2 data and U_{L3} are encrypted using R_{L2}. L3 data is encrypted using R_{L3}. These encrypted data are transmitted to the database server. L1 user has its public key (U_{L1}). Using U_{L1}, L1 data is decrypted. After decrypting L1 user gets the public key for Level 2 i.e. U_{L2}. Using that U_{L2}, L2 data is decrypted. After decrypting, L1 user gets the public key for Level 3 i.e. U_{L3}. Using that U_{L3}, L3 data is decrypted. Thus the L1 user gets L1, L2 and L3 data. In the case of L2 user, they have its public key U_{L2}. Using U_{L2}, L2 user decrypts its data. After decrypting L2 user gets U_{L3}. Using that U_{L3}, L2 user can decrypts the L3 data. Thus L2 user gets L2 and L3 data, but L2 user cannot access L1 data. In the case of L3 user, they have the public key U_{L3}. Using that U_{L3}, L3 user can decrypts the L3 data. But L3 user cannot access L1 and L2 data. UI, L1, L2 and L3 are stored in the database.

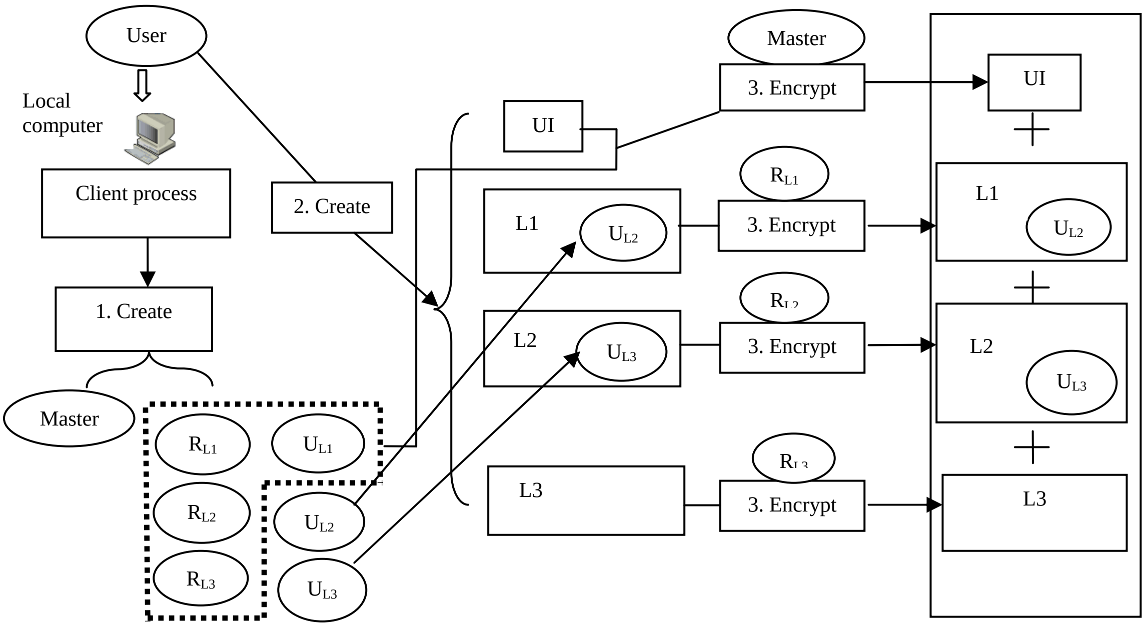

Fig. 1. Multi level storage architecture

Master key is used for providing security. Storing the private key and private information in database is not good, because attacker can easily attack. In order to avoid this, master key is introduced. All private information, private key for all the level and public key of level $1(U_{L1})$ are encrypted using master key. Master key is stored in master key server not in the database server for security. Fig.2 shows master key server. This server contain a master key table which includes hash of user's particular information and corresponding master key. Each user create master key and hash value. This is stored in the master key table.

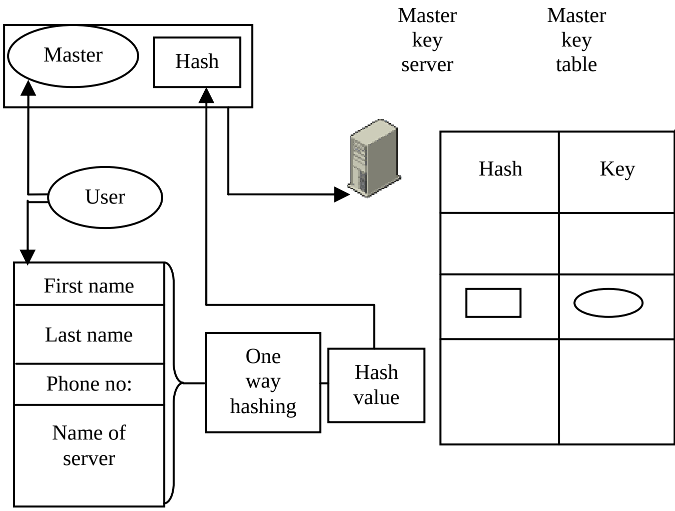

Fig.2. Master key Server

Fig.3 shows how the user recover the master key for sending data. If a user want to send data, it first create the master key using hash value of user's particular information. User request the master key to the master key server with hash value. Then search for the hash value in master key table and response with corresponding master key. Thus the master key recovered.

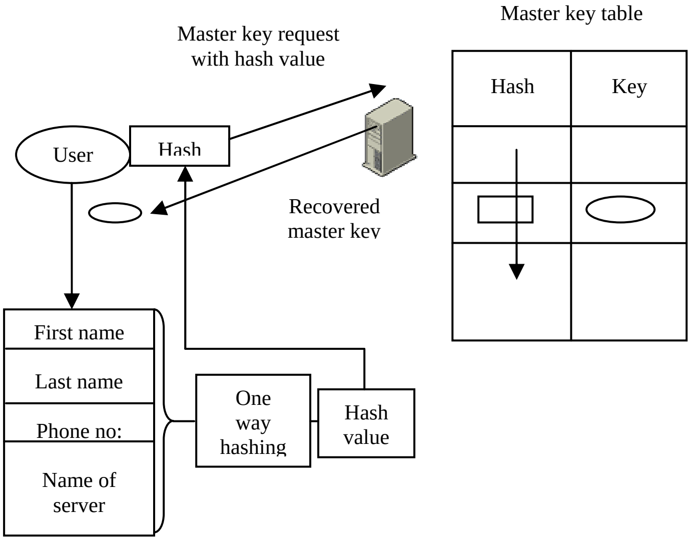

Fig.3. Recovered Master key

Master key provides security by encrypting the private information and private key. No one can access database when we use the master key. This cause problem when trusted authorities like government need to access the database even with court search warrant. For providing access to the designated authorities we use master key splitting scheme. Master key is split into two parts and one part is given to the database server and other part is given to the designated authorities. Individual part of the master key has no use. For accessing the database the two key parts is combined and thus master key can reconstructed. Fig.4. shows the master key splitting scheme.

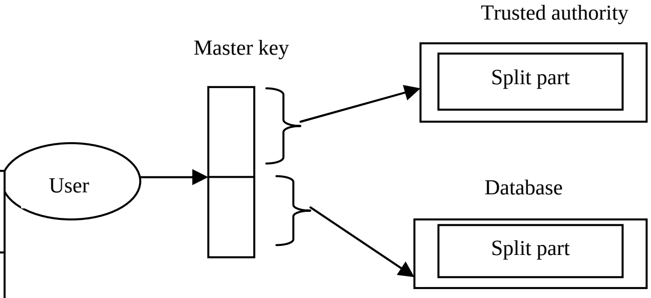

Fig.4. Master Key Splitting

When the designated authorities want to access the database, it first sends a master key share request to the database. Then database server response with corresponding master key share part. Using that key part designated authorities can reconstruct the master key. Using the reconstructed master key designated authorities can access the database.

It is important to check whether the reconstruction of split key results the actual master key without revealing the key part. For that purpose zero knowledge protocol is used. In this paper, a revocation list is introduced to avoid unnecessary access to the database. Once the key is given to the user it cannot taken back when the user is revoked. For this a revocation list is maintained in the database server and access to the database is based on the revocation list. If any revoked user request for accessing, server check whether the user is revoked or not. If the user is revoked, then database server doesn't provide access.

IV. IMPLEMENTATION

This scheme helps to provide security to the big data. When compared to the normal access control mechanism, multi level access control is more powerful. This helps to provide access to the different user based on the level of the user. Private information and private key protection is important. For that purpose, a master key is introduced along with multi level access control. This master key helps to encrypt all user private information and also private key. User is creating this master key and this mater key is stored in the master key server for providing more security to this scheme.

Master key splitting scheme is introduced to provide access to the designated authorities. User split the master key and one split is given to application server and other one is given to designated authorities. Both application server and authorized authorities send their share parts to master key server. Master key server contains hash of user's particular information and corresponding master key. Using the share parts master key server reconstruct the master key using Shamir's algorithm. After that master key server check whether the reconstructed master key is present in the master key table. If it is present then the reconstructed master key is correct. After the reconstruction master key splits in the master key server is deleted. This also provide security because if any one attacks this master key server, then they cannot get the splits.

Shamir's secret sharing algorithm

- Goal is to divide some data *D* into n pieces D1,D2….Dn in such a way that:
  1. Knowledge of any k or more D pieces makes D easily computable.
  2. Knowledge of any k -1 or fewer pieces leaves D completely undetermined
- This scheme is called (k,n) threshold scheme. If k=n then all participants are required together to reconstruct the secret. Suppose we want to use (k,n) threshold scheme to share our secret S where k < n.
- Choose at random (k-1) coefficients $a_1,a_2,a_3...a_{k-1}$, and let S be the a0

  $$f(x) = a_0 + a_1 x + a_2 x^2 + ... + a_{k-1}{}^{k-1}$$
- Construct n points (i,f(i)) where i=1,2…..n
- Given any subset of k of these pairs, we can find the coefficients of the polynomial by interpolation, and then evaluate a_0=S , which is the secret.

Zero knowledge protocol helps to check whether the reconstruction of master key split results the actual master key. This helps to checks the integrity of the master key. Here a random number is used. This random number is known to mater key server, authorized server and application server. Mater key server encrypt this random number using the reconstructed master key and send this to the user. User decrypts this using the master key and send this decrypted values to application server and authorized server. If the decrypted random number matches the real random number, then the reconstructed master key is same as the original.

Authorized authorities gets the user data by asking master key share of that user to the application server. Application server send the share and authorized authorities can reconstruct the master key using Shamir's algorithm. Only reconstruction should be done by authorized authorities. Application server cannot do this and individual share of master key has no use. This also improve security.

Revocation list is incorporated with this scheme. This list is maintained in the database for avoiding the accessing of revoked user. This revocation list helps to improve the performance by controlling revoked users access. When compared to ordinary security scheme, this scheme provides more security by protecting the private information and provides strong access control. This scheme helps for the private key protection.

V. Security Analysis

In this section, we compare our scheme with security feature of the one proposed. This scheme provides more security and also provide access to the designated authorities.

- Access control: Compared with previous scheme, multi level access control is more powerful. Here access to the database is based on the level of each user.
- Security: In this scheme, master key is used for protecting the information. All the keys are encrypted using this master key and this master key is stored in separate server which give more security.
- Key protection: In our scheme all the keys are encrypted using master key. In order to protect the master key, this master key is placed in a separate server, which is a trusted server.
- Privileged access: When compared with previous scheme, this scheme provide privileged access to the designated authorities. Master key is kept in a separate sever. So it is difficult to access the database by authorized user. But our scheme introduces master key splitting scheme and thus privileged access is possible for the authorized user.
- Integrity: When compared to previous scheme, our scheme also provides integrity by checking whether the reconstruction of master key splits results the actual master key.
- User revocation: Also our scheme introduces a revocation list. This revocation list helps to avoid unnecessary access to the database by the revoked user. By using this revoked user cannot access the database.

VI. Conclusion

Big data is a collection of large and complex data set. Security is main important challenge in big data. Now a day’s protecting our private information is crucial. To provide security, multi level access control with the help of master key scheme is used. Also this master key helps to protect the key which is used for encrypting the information. This master key is stored in a separate server. Master key server is a trusted server. So no one can access the master key server. This became a problem when the government authorities want to access the database with court’s search warrant. For providing access to the designated authorities, we proposed a master key splitting scheme. This scheme helps the trusted authorities like government to access the database. Also revocation list is maintained in database server to avoid unnecessary access to the database, when the user is revoked. This scheme provides security and access control to the large data.

REFERENCES

[1] Iqra Basharat,Farooque Azam,Abdul Wahab Muzaffar;" Database Security and Encryption: A Survey Study", *International Journal of Computer Applications (0975 – 888)Volume 47– No.12, June 2012*

[2] Laszlo Hars, "Discrption:Internal Hard-disk encryption for secure storage," *Computer*, vol.40, n0.6, pp.103-105, 2007

[3] Umut Uludag, Sharath Pankanti, Salil Prabhakar, and Aniket K. Jain, "Biometric Cryptosystems: Issues and Challenges," proceedings of the IEEE, vol.92, no.6, pp.948-960,2004

[4] Bryan Parno,Jonathan M. McCune, Dan Wendland, David G. Andersen, and Adrian Perrig, "CLAMP:Practical Prevention of Large-Scale Data Leaks," *Proceedings of IEEE Symposium on Security and Privacy*, pp.154-169, may 2009.

[5] Fangyong Hou, Dawu Gu, Nong Xiao, and Yuhua Tang, "Secure Remote Storage through Authenticated Encryption," *Proceedings of International Conference on Networking, Architecture, and Storage, pp. 3-9, June 2008.*

[6] J acob W. Keister, Hiroshi Fujinoki, Clinton W. Bandy, and Steven R. Lickenbrock," SoKey: New Security Architecture for Zero-Possibility Private Information Leak in Social Networking Applications,"Proceeding of IEEE,2007