Intensive Convergence Encryption for Secure Deduplication in Cloud

Priyanka J.
Department of Computer Science and engineering,
IFET college of engineering

Divya A.
Senior Assistant Professor
Department of computer science and engineering
IFET college of engineering

ABSTRACT
Deduplication is used for eliminating the duplicate copy of same data and providing security to them. Replica of same data is obtained by encrypting the same information with different key. Surplus amount of data is being stored in the cloud environment to shrink the data in storage space convergence encryption technique is used. In convergence encryption technique the key will be generated from the file by the hash value so that same file will turn out same cipher text. When the cipher text is indistinguishable it will be identified as a duplicate. The main weak spot of the convergence encryption comes from the fact that key for a given data part can be generated by anyone in a deterministic way. This allows an attacker to easily recuperate the plain text from the key so it critically compromises user's confidentiality. In this paper encrypted text by convergence encryption is made complicated by the bit reversal and RC4 algorithm. so that the data will be more secure in the cloud storage and prevented from the attacker. Deduplication manager will be used for the purpose of detecting the duplicate file and manage the keys. By this even though the Attacker knows the cipher text he cannot obtain the original text from encrypted data it will make the encryption more efficient and make the data secure.

Keywords
dedupication, convergence encryption, confidentiality

1. INTRODUCTION
Cloud storage is becoming very popular and data in the cloud is getting enormous everyday by the usage of people for the personal and organizational purposes. To reduce the amount of data stored in the cloud deduplication concept is used, which will avoid storing two similar copy of same. By this deduplication concept the storage space and the bandwidth are saved and cloud is become more efficient For the security purpose data will be encrypted by the client before uploading the file to the cloud storage because data may be leaked and internal attacks may happen to the file. In the traditional method (figure 1.2) For the security purpose data will be encrypted by the client before uploading the file to the cloud storage because data may be leaked and internal attacks may happen to the file. In the traditional method (figure 1.2)

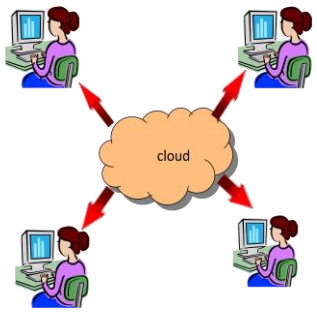

Fig 1.1 cloud storage users

data will be usually encrypted with own key of the user in their system and upload the file to the storage space as the encoded data so the cipher text will be different even though the plain text is same . This will make the deduplication impossible. And occupy large storage space in the cloud

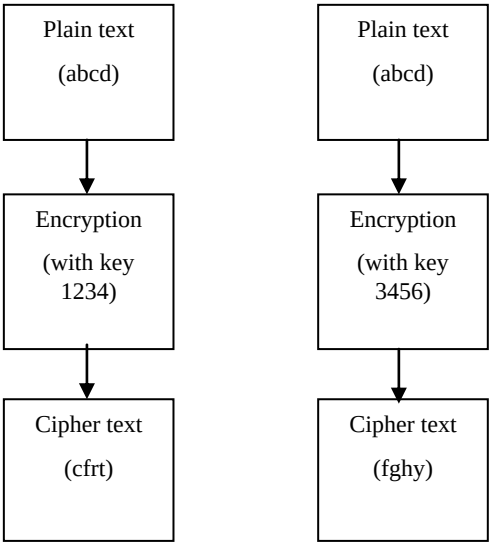

Fig 1.2 traditional method

To make deduplication possible convergence encryption(figure 1.3) is used, in this technique during the uploading of the file key will be generated as the hash value of the original data and then file will be encrypted with the hash value instead of giving the encryption key by user. So the cipher text will be same for identical plain text. By this the data can be compared with the existing file and duplicate data can be avoided from the storage space.

Main weak spot of the convergence encryption comes from the fact that key of a given data segment can be generated by anyone in a deterministic way. By finding the key form the

hash value of the data to avoid this deduplication manager is implemented before uploading the data to the cloud storage.

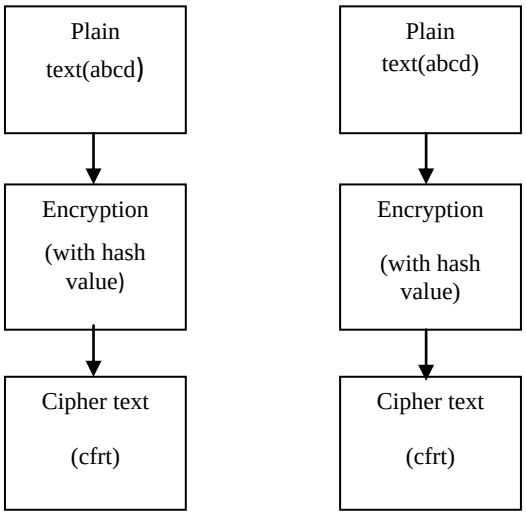

Fig 1.3 convergence encryption

In the deduplication manager encrypted data with the convergence encryption is again undergone bit reversal and then encrypted with the RC4 algorithm. Then the key will be stored and managed in the database. After the encryption, the cipher text will be stored in the cloud. It will make the data more efficient and secure. Then block level deduplication technique is used for the deduplication because it will be more efficient than the file level deduplication technique and comparing the file blocks in the storage environment.

2. RELATED WORKS

Jin Li, Xiaofeng Chen, Mingqiang Li[1] proposed Dekey,a new edifice where the user no need to remember all the key.

Dekey will be use to manage the key securely across multiple server . Pasquale Puzio[3] proposed a block level deduplication, which is made in sort to store all the key in the metadata and digital signature is used for protection. Wee Keong Ng[7] showed how to Construction data deduplication protocol by the hash function and collision resistant mechanism. private data provably secure in the simulation based framework. Mihir Bellare [2] carry deduplication concept, in which a group of associated client (company employee) encrypt their data with the help of key server(KS). It will evade brute force attack. Jia Xu[8] presented a concept to allows enclosed amount multi-time leakage of the intended file before and after their scheme starts execute. It will evade divide and conquer, poison attack and secure data from the inside attacker. Nesrine Kaaniche[9] shows a Asymmetric encryption of the data and then the deduplication is avoided by Merkle tree properties . Opposing to unauthorized access to data and any data confession during sharing process, done by two levels of access control proof.

Jiawei Yuan, Shucheng Yu[5] proposed a concept of Public and constant cost storage integrity activity idea with secure deduplication based on technique including polynomial-based authentication tag and homomorphic linear authentication. Mihir Bellare[2] introduced a concept of key server which will help the group of associated client(company employee) encrypt their data.

Dr. Ajit, Preeti Kalra, Sonia Dhull[11] proposed a novel important technology to prevent illegal copying of data. Digital watermarking can be applied to audio, video, text or images. Shai Halevi[6] proposed system have authentication and anonymous models, a map is created for each file that portray how to renovate a file from chunk. The file is encrypted using asymmetric key pair. Jorge Blasco [10] gives a novel explanation based on Bloom filters that gives a flexible, scalable, and provably protected solution to the weaknesses of deduplication. It will make the data secure. Bloom filters have advantage of memory and time efficiency, since they involve less space than other data structures to store essentials in the set, and less time to execute membership queries.

3. SYSTEM IMPLEMENTATION

In this paper, the data will be uploaded and processed with the convergence encryption in the client side and then it will be moved to the deduplication manager for secure deduplication . the file is spitted to multiple blocks as the block size of 4kB then the file will be checked to find whether it is already existing in the cloud storage space. This can be done by the function of finding the matched and mismatched block in the new file from the existing file in the storage space by the string matching function.

If the block is already existing in the storage space it will be given a pointer from the reference of the existing data else the block is not matched with the existing content it will be allocated a new space in the cloud. So that the space can be saved in the storage and we can use the cloud in more efficient way. After all the data is being compared with the existing data file and deduplication. the duplicated file block will be given a encryption for security because the internal theft may occur in the cloud. the encryption is done in two stage as bit reversal function and the RC4 algorithm. The bit reversal algorithm is all the encrypted bit will be changed from 0 to 1 and 1to 0 then this process will be followed by the RC4 algorithm. RC4 algorithm will normally uses the 64 bit and 128 bit key sizes and involve the key Scheduling Algorithm and the Pseudo-Random Generation Algorithm.

In this algorithm the secret key will be generated and run in the key Scheduling Algorithm (KSA). KSA will use the secret key to scramble the array. finally the key stream is XORed with the plain text and gives out the encrypted cipher text. Then the data will be stored in the cloud.

3.1 Encryption for equivalent cipher text:

The user will login into cloud in their respective account and upload the data to the cloud. After the uploading of the file it will undergo convergence encryption. So same plain text will always produce the same cipher text.

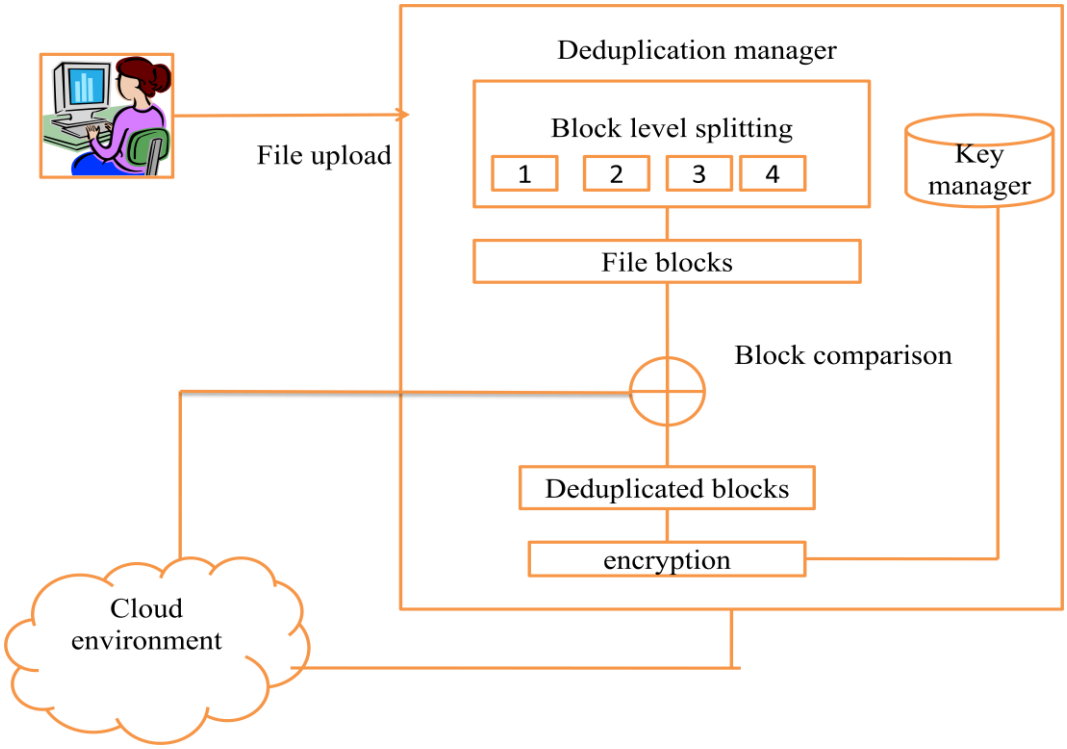

Fig 3.1 block level deduplication

After the convergence encryption file will be moved to the deduplication manager.

3.2 Block level splitting of data

After convergence encryption in the client side the file will be uploaded to the deduplication manager. In dedeuplication manager, first the data will be segmented into blocks for the performing the operation of block level deduplication. The file will be segmented by the string function. Each block will be segmented as 4KB in size so that the duplication can be identified in high level and more space can be saved. Each block will have the cryptographic hash function exclusive of considering the data type of the block. Each block hash value will be maintained in the index.

```
static IEnumerable<string> Split(string str, int chunkSize)
{
    return Enumerable.Range(0, str.Length /4)
        .Select(i => str.Substring(i * chunkSize, chunkSize));
}
```

File segmentation

By this code the data will be segmented into chunks and broken into 4KB data blocks. It will store the data block in the data base then the index value will be stored in the hash table

3.3 file deduplication

After the block level splitting of the data in the deduplication manager the comparison between the data block will be made among the existing blocks in the cloud. If the data is already present in the cloud the reference of the paper ill be given as a reference and if the data block is not present in the storage space new space will be allocated for the new blocks of data in the cloud storage environment.

After the block level comparison the data will be maintained in the database for the encryption The secret key will be generated and then the key stream is produced after that it will be XORed with the plain text finally the encrypted data will be stored in the cloud environment. By this encryption the data will be safe.

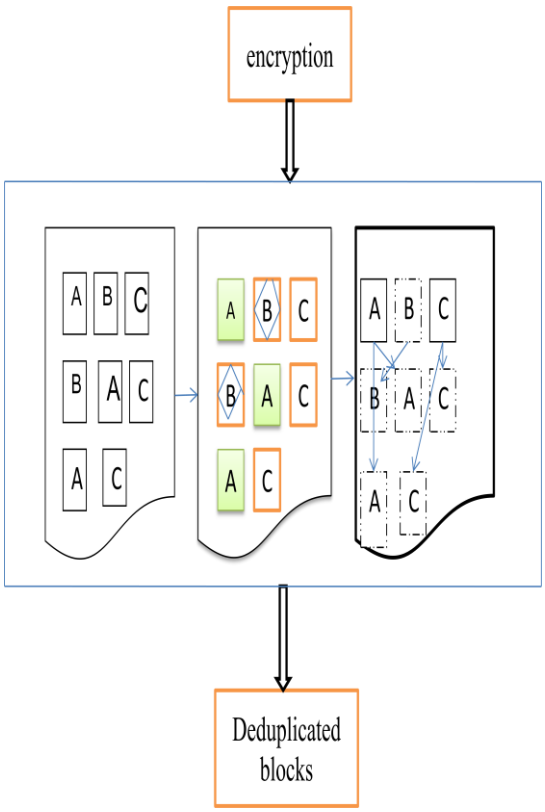

Fig 3.2 file deduplication in cloud storage

3.4 Encryption For security

Encryption is done in the deduplicated file block for security because internal attack may occur in the file . This will be

done after the process of deduplication in the uploaded file . bit reversal function is applied in the encrypted data and after that RC4 algorithm is used.

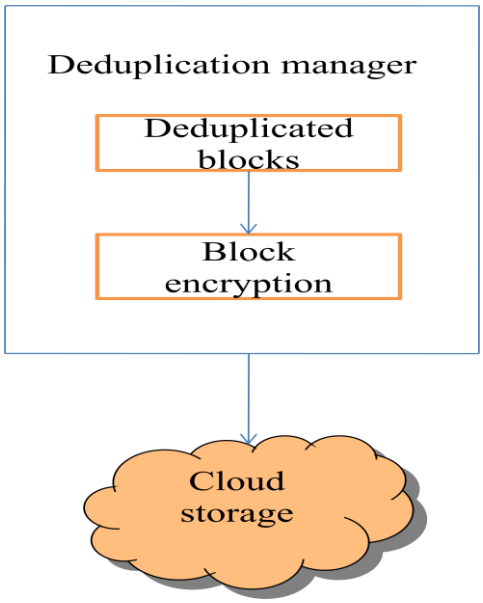

Fig 3.3 Encryption in deduplicated file

RC4 Algorithm will involve Key Scheduling Algorithm and the Pseudo-Random Generation Algorithm

```
public static string RC4(string input, string key)
{
   StringBuilder result = new StringBuilder();
   int x, y, j = 0;
   int[] box = new int[256];

   for (int i = 0; i < 256; i++)
    {
      box[i] = i;
   }


    for (int i = 0; i < 256; i++)
   {
      j = (key[i % key.Length] + box[i] + j) % 256;
      x = box[i];
      box[i] = box[j];
      box[j] = x;
   }

   for (int i = 0; i < input.Length; i++)
   {
      y = i % 256;
      j = (box[y] + j) % 256;
      x = box[y];
      box[y] = box[j];
      box[j] = x;

      result.Append((char)(input[i] ^ box[(box[y] +
box[j]) % 256]));
   }
   return result.ToString();
}
```

RC4 algorithm

4. CONCLUSION

This paper ensures the confidentiality and integrity of the client and also secures the file from the internal attacks in the cloud. Block level deduplication is used which will be more efficient than the file level deduplication in saving the bandwidth and the storage space. The additional layer encryption using RC4 algorithm is carried out in this concept, to make the data more secure and the key will not be shared with anyone else other than the deduplication manager. Deduplication manager will carry out all the deduplication operation after the deduplication function the data will be stored in the cloud. and the key will be managed in the separate database.

This intensive convergence encryption will not allow the attacker to hack the file from the cloud even though they know the cipher text they cannot retrieve the plain text fromthe cipher text so that we can attain the secure deduplication in the cloud environment.

5. REFERENCES

[1] Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou"Secure deduplication with efficient and reliable convergent key management"

[2] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Serveraidedencryption for deduplicated storage", in Proc. 22nd USENIXConf. Sec. Symp., 2013.

[3] Pasquale Puzio , Refik Molva , Melek O¨ nen , Sergio Loureiro "cloudedup: server deduplication with encryption data for cloud storage"

[4] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication", in Proc. 32nd Annu. Int.mConf. Theory Appl. Cryptographic Techn., 2013, pp. 296–312.

[5] Jiawei Yuan , Shucheng Yu"secure and constant public cloud storage auditing with deduplication"[6] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems", in Proc. ACM Conf. Comput. Commun. Security, 2011, pp. 491–500.

[6] W. K. Ng, Y. Wen, and H. Zhu, "Private data deduplication protocols in cloud storage", in Proc. 27th Annu. ACM Symp. Appl. Comput., 2012, pp. 441–446.

[7] J. Xu, E.-C. Chang, and J. Zhou, "Weak leakage-resilient clientside deduplication of encrypted data in cloud storage", in Proc. 8th ACM SIGSAC Symp. Inform., Comput. Commun. Security, 2013, pp. 195–206.

[8] Monique Laurent, Nesrine Kaaniche "A Secure Client Side Deduplication Scheme in Cloud Storage Environments"

[9] Jorge Blasco,Roberto Di Pietro,Agustin Orfila,Alessandro Sorniotti "A Tunable Proof of Ownership Scheme for Deduplication Using Bloom Filters"

[10] Dr. Ajit, Preeti Kalra, Sonia Dhull"Digital watermarking"

[11] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, "A secure data deduplication scheme for cloud storage", Tech. Rep. IBM Research, Zurich, ZUR 1308-022, 2013.

[12] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, "Secure data deduplication", in Proc. 4th ACM

Int. Workshop Storage Security Survivability, 2008, pp. 1–10.

[13] K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan, "Sedic: Privacy- aware data intensive computing on hybrid clouds", in Proc. 18th ACM Conf. Comput. Commun. Security, 2011, pp. 515–526.

[14] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication", in Proc. 32nd Annu. Int. Conf. Theory Appl. Cryptographic Techn., 2013, pp. 296–312.

[15] M. Bellare, C. Namprempre, and G. Neven, "Security proofs for identity-based identification and signature schemes", J. Cryptol., vol. 22, no. 1, pp. 1–61, 2009.

[16] K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan, "Sedic: Privacy- aware data intensive computing on hybrid clouds", in Proc. 18th ACM Conf. Comput. Commun. Security, 2011, pp. 515–526.