# DATA SECURITY IN MOBILE CLOUD ENVIRONMENT BY ADOPTING DATA CLASSIFICATION MODEL

**Deepak. G* and Manjunath. S. S****

* Assistant Professor, Department of Computer Science & Engineering, Dayananda Sagar College of Engineering, Bangalore. Email Id: deepak.dsce@gmail.com, deepak_gopal_rao@yahoo.com
** Professor & Head, Department of Information Science & Engineering, Dayananda Sagar Academy of Technology and Management, Bangalore. Email Id: mnj_ss2002@yahoo.co.in

**ABSTRACT:** The cloud services are openly available for all kinds of the organizations to store and access their data from the cloud using their mobile devices. Mobile cloud security has become a major concern for most of the organizations to store and access their sensitive or critical data from the cloud. In the cloud, data is stored in a centralized manner, where malicious users can easily access the information and modify the data. So far to provide data confidentiality among number of users or tenants, we classify the data into various levels of security and apply our proposed algorithm by means of upgrading the key size based on the levels chosen.

**KEYWORDS**: Cloud, Mobile cloud, DRDP method.

## INTRODUCTION

Cloud computing [1] plays a vital role in every organization's day to day work. Organizations utilize cloud services for storing and accessing the information because of its features like scalability, resource pooling, etc. Nowadays researchers mainly focus on efficiently storing the information and retrieving the information in a mobile cloud environment. Data centers running in a simultaneous cooperated and distributed manner empowers the deployment of Cloud Computing. Each user's data is stored in multiple physical locations redundantly to reduce the data integrity threats. Therefore, to achieve robust and secure cloud data storage system in the real world, distributed protocols for storage correctness assurance is of most importance in achieving. Distributed computing is of developing enthusiasm because of its potential for conveying adaptable capacity and preparing. Cloud security [2] is a noteworthy range of worry that is confining its utilization for specific applications: "Information Confidentiality and Auditability" referred as one of the main deterrents to the selection of distributed computing in the persuasive Berkeley report. While security concerns are keeping a few associations from receiving distributed computing by any means, others are thinking about utilizing a blend of a safe inner "private" cloud, alongside less secure "open" mists. Touchy applications can then be sent on a private cloud, while those without security concerns can be conveyed remotely on an open cloud. In any case, there are issues with this methodology. Presently, the portion of utilizations to mists is generally done on a specially appointed, per-application premise, which is not perfect as it needs thoroughness and auditability. Further, choices are regularly made at the level of granularity of the entire application, which is designated completely to either an open or private cloud in light of a judgment of its general affectability. This wipes out the potential advantages for dividing an application over an arrangement of mists.

Mobile cloud computing [3] is defined on 5th March 2010 at open garden blog states "The availability of cloud computing services in mobile echo system". Mobile cloud computing [4 -5] provides the availability of cloud computing services in a mobile ecosystem, the usage of cloud computing services in combination with mobile devices and mobile internet. In Mobile cloud computing applications runs on a remote server and then sent to the user and also there is no need of powerful configuration for the mobile devices, since the complicated modules are processed on the cloud. The main role of the mobile cloud computing [6] is that the information is available at our finger tips anywhere at any time, so that users can access information through mobile devices.

## LITERATURE SURVEY

Moving data into the cloud offers greater convenience to users since they don't have to care about the complexities of direct hardware management. Amazon Simple Storage Service (S3) and Amazon Elastic Compute Cloud (EC2) [7 - 8], internet-based online services provides huge amounts of storage space and customizable computing resources, this computing shift, eliminates responsibility of local machines for data maintenance. The importance of ensuring the remote data integrity has been highlighted by the following research works [9–13]. Researchers have also proposed distributed protocols [14–16] for ensuring storage correctness across multiple servers or peers.

Balajee Maram & et al [17] have proposed symmetric cryptographic algorithm that uses Image and Double reflection perturbation method. Double-reflecting Data-Perturbation method plays a vital role in encryption and decryption. The proposed algorithm shuffles data and is converted into image and from the receiver-side, the image is converted into ASCII and then data.

Keke chen & Et al [18] have proposed the Geometric Data Perturbation (GDP) method & several aspects of GDP method. They have also highlighted well-known data-mining models that delivers comparable level of quality over the geometrically perturbed data set over the original data set. They have mainly proposed a multi-column privacy evaluation framework for evaluating the effectiveness of geometric data perturbation with respect to different level of attacks.

Subhasri and et al [19] has proposed the new level of data security solution using the Reverse Caesar cipher algorithm using ASCII of full 256 characters. The purpose of this method is to solve the security issues in multi level encryption for both cloud providers and cloud consumers using cryptographic methods. Cloud computing reduces operating cost and increases the efficiency of computing. Even though efficiency increased, still there is security threat for the data that is stored in third party area especially in Internet. Due to data security issue with cloud computing many business organization have fear in storing their data in Cloud.

Iehab AL Rassan, Hanan Al Shaher [20], Mobile devices with Internet capabilities have increased the use of mobile cloud computing. Due to hardware limitations in mobile devices, these devices can't install and run applications which requires more CPU processing or memory. The author proposes and implements a new user authentication mechanism of mobile cloud computing using fingerprint recognition system. Fingerprint images of mobile users can be captured and processed using mobile phone camera to access mobile cloud computing . The implementation of the proposed solution in different mobile operating systems and devices show security enhancement in mobile cloud computing with accepted performance level.

A. Gholami and E. Laure [21] has discussed the design and implementation of a security framework for Biobank Cloud, a platform that supports the secure storage and processing of genomic data in cloud computing environments. The proposed framework is built on the cloud privacy threat modeling approach [22 - 23] which is used to define the privacy threat model for processing next-generation sequencing data according to the DPD. This solution includes a flexible two factor authentication and an RBAC access control mechanism, in addition to auditing mechanisms to ensure that the requirements of the DPD are fulfilled.

E. Ayday, J. Raisaro, U. Hengartner, A. Molyneaux, and J.-P. Hubaux [24], have discussed several privacy issues associated with genomic sequencing. They have also described several open research problems of outsourcing to cloud providers, genomic data encryption, replication, integrity, and removal of genomic data along with giving suggestions to improve privacy through collaboration between different entities and organizations.

The literature reveals data encryption and decryption is done by taking entire message or document without data classification. The entire message or document set may contain sensitive and non-sensitive information, encrypting the non-sensitive information with the same method will increase the time complexity, as the non-sensitive and sensitive information, both are encrypted using the same encryption and decryption method. The above problem can be solved by classifying the data based on their sensitivity as 'levels' by adopting different key lengths. This helps in faster encryption and decryption rate by preserving the privacy policy of the information.

## PROPOSED SYSTEM

In the proposed system we apply Double-Reflecting Data Perturbation Method (DRDP) to generate Intermediate key $I_k$, where, the original data by X and Y axis is to achieve the perturbed data for some confidential attribute. In this method, the randomization function plays a crucial rule, and if the function is not properly chosen it may degrade the clustering quality. The distortion operation performed to the confidential attribute is given by:

$$opj = \rho Aj + (\rho Aj - aj) = 2\,\rho Aj - aj \quad \ldots\ldots (1)$$

Where Aj ($1 \leq j \leq n$) is a confidential attribute and a j ($1 \leq j \leq n$) is an instance of Aj. $\rho Aj$ is defined by the following formula:

$$\rho Aj = |(\max Aj + \min Aj)/2| \quad \ldots\ldots (2)$$

Where max Aj and min Aj are respectively the maximum value and minimum value of attribute Aj. Before we adopt this method we classify the data set into the attribute class such as sensitive, Non Sensitive and we further classification the data into 3 levels such as Level 1, Level 2 & level 3.
The privacy of data is measured by variance between the actual and perturbed values using the formula:

$$A = \text{VAR}(A - A') \text{ VAR}(A) \ldots\ldots (3)$$

**LEVEL 1:** The basic security level concerned with the general type of data like photos, videos etc. which do not need high degree of sensitiveness. So this level the security is basic security which is used by most of the product online.
**LEVEL 2:** Sensitive level is designed for the data with medium sensitive degree like personal files, videos, pictures, documents etc.
**LEVEL 3:** The other level of security is Restricted Level. All the data which the user wants to be highly secure are comes under this category. Data like Financial Transaction, Secret documents of organization, Military data etc. all comes under this.

**Table 1:** Notations Used in the Algorithm

| **Symbols used & Terminologies** | | |
|---|---|---|
| $S_k$ - Shared-Secret-Key(16 byte long) | S - Session-Key | $I_K$ - Intermediate-Key |

***Algorithm for Encryption:***
**Step 1:** The Permanent shared secret key $S_k$ is shared between the sender and receiver.

**Step 2:** Session key S is generated for a particular session using the below formula:
**random_string = char(floor(94*rand(1, length)) + 32);** where Length is 16 character long.

**Step 3:** Intermediate-Key $I_K$ will be generated based on both $S_K$ and S by following steps:
**a.** Shuffle all the characters in each sentence by means of using Double-reflecting Data-Perturbation method. Here we use . (Dot) as a delimiter.
**b.** Arrange all the characters into 4X4 matrices except the last character
**Total. No of Characters % 16**
**c.** On applying transpose of the matrix we get $\mathbf{M = M^T}$
Four characters in each row in every matrix, will be shuffled according to Double-reflecting-data-perturbation method. So the characters in 1st row, 2nd row, 3rd row and 4th row in 1st matrix will be arranged as a paragraph. Then from 2nd matrix and so on.

**Step 4:** The first 16 characters are xored ⊕ with one-time Session-Key.
**Step 5:** Next 16 characters are xored ⊕ by Intermediate-Key, and so on.
**Step 6:** Now the Intermediate-Key is appended to cipher-text and transmitted to the destination

In the algorithm discussed above, the Intermediate key length depends on the level of security which user wants to apply to the document or a part of the document. The length of Intermediate-Key will be 16 bytes for LEVEL 1 security, for LEVEL 2 security Intermediate-Key will be 32 bytes, and for the LEVEL 3 [Restricted Level] is 48 bytes. The LEVEL 3 data is more secured than compared to other level .The crucial information are encrypted at LEVEL 3 so that, it is difficult for a hacker to breach the security.

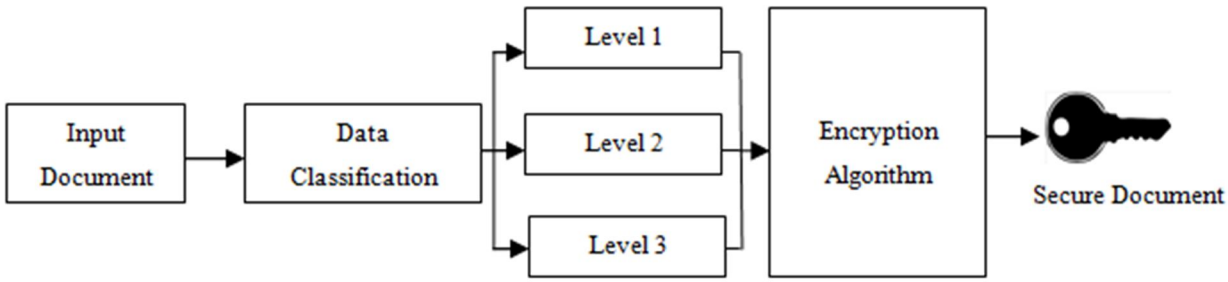


**Figure 1:** System Architecture

All the data is encrypted and aggregated as shown in figure 1, and with different key lengths, identifying the pattern of encryption is difficult even if hacker gets the data. The security defined in this method is difficult to break as all the encrypted data is aggregated, sent and the original message will have distributed data of different sensitivity. The attribute such as Occupation, Race, Sex or public information are not sensitive hence it can be encrypted using LEVEL 1 , but Age , Work class, Capital-gain, Salary are sensitive hence based on sensitivity of information the level of security with which the field has to be encrypted is decided . The '$' symbol is appended between the LEVEL 1 encrypted Data, '#' symbol is appended between the LEVEL 2 encrypted Data, '^' symbol is appended between the LEVEL 3 encrypted Data. As Data is encrypted differently based on sensitivity of data in a single file and is merge together. So to clearly distinguish the encrypted data for different levels of security the special characters are used.

**Algorithm for Decryption**

The Encrypted data between '$' will be decrypted using the 16 byte Intermediate key followed by the decryption method .similarly the data between the '#' will be decrypted by 32 byte intermediate key and data between '^' will be decrypted using 48 byte intermediate key. After this first stage decryption it is fed as an input for the decryption algorithm to decrypt further, to get the original message. The sequence of steps used to decrypt the original message is discussed below:

**Step 1:** The receiver calculates the Session-Key using Intermediate-Key and Shared-Secret-Key in the following way:
$\mathbf{S_{11}=S_{k11} \oplus I_{K11}, S_{12}=S_{k12} \oplus I_{k12}}$**, So on.**
**Step 2:** The first 16 characters are $\oplus$ with one-time Session-Key.

**Step 3:** Next 16 characters are $\oplus$ by Intermediate-Key, and so on.

**Step 4:** All characters will arranged in 4X4 matrices (M) except last characters.
**Total Number of Characters % 16**
**Step 5:** All 4 characters in each row in every matrix will be reshuffled according to Double-reflecting-data-perturbation method.

**Step 6:** All matrices will be transposed $\mathbf{M = M^T}$

**Step 7:** Now all the characters in each row in each matrix including left-over characters are arranged as a paragraph.

**Step 8:** All the characters in each row (dot is delimiter for each row) will be reshuffled according to Double-reflecting-Data-Perturbation method, to get the original text message.

**COMPARISON AND RESULTS**

The proposed algorithm includes xor $\oplus$ operation for encryption and decryption, the time taken for encryption and decryption is very less in processing. The proposed algorithm's time of encryption and decryption will directly depend on the size of the document provided. As the data inside a single document is encrypted using different key length (Intermediate Key) the pattern of encryption is highly difficult to trace by the hacker.

**CONCLUSION**

In 1st phase of proposed method, all the characters in the sentence of the document are converted into ASCII. The characters are shuffled according to their ASCII value by Double Reflecting Data Perturbation Method. The privacy of data is measured by the variance between the actual and the perturbed values using the equation 3 mentioned above. It has been analyzed that the privacy or the security level of the confidential data is improved a lot by the proposed method for Encryption and Decryption.

## REFERENCES

[1] Cloud Computing: saas, paas, iaas, virtualization, business models, mobile, security and more by Dr.Kris Jamsa, "vgpdf.oceanbooks.eu/cloud-computing-saas-paas-dr-kris-37837041.pdf".
[2] A review on cloud computing security issues & challenges by F. A. Alvi1, B.S Choudary N. Jaferry , E.Pathan, IJAST, 2014.

[3] Shahryar Shafique Qureshi, Toufeeq Ahmad, Khalid Rafique, Shuja-ul-islam, “mobile cloud computing as future for mobile applications – Implementation methods and challenging issues”, Proceedings of IEEE CCIS, 2011.
[4] Le Guan, Xu Ke, Meina Song and Junde Song, “A Survey of Research on Mobile Cloud Computing”, 10th IEEE/ACIS International Conference on Computer and Information Science, 2011.
[5] Shahryar Shafique Qureshi, Toufeeq Ahmad, Khalid Rafique, Shuja-ul-islam, “mobile cloud computing as future for mobile applications – Implementation methods and challenging issues”, Proceedings of IEEE CCIS, 2011.
[6] Markus Schüring, Georgios Karagiannis, “Mobile Cloud Computing: Resource Discovery, Session Connectivity and Other Open Issues”, IEEE, 2011.
[7] Amazon.com, “Amazon Web Services (AWS),” Online at http://aws. amazon.com, 2008.
[8] N. Gohring, “Amazon’s S3 down for several hours”, online at http://www.pcworld.com/business center/ article/142549/amazons s3 down for several hours.html, 2008.
[9] A. Juels and J. Burton S. Kaliski, “PORs: Proofs of Retrievability for Large Files,” Proc. of CCS ’07, pp. 584–597, 2007.
[10] H. Shacham and Etal, “Compact Proofs of Retrievability”, Proc. of Asiacrypt ’08, Dec. 2008.
[11] K. D. Bowers, A. Juels, and A. Oprea, “Proofs of Retrievability: Theory and Implementation,” Cryptology ePrint Archive, Report 2008/175, 2008, http://eprint.iacr.org/.
[12] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, “Provable Data Possession at Untrusted Stores”, Proc. of CCS ’07, pp. 598–609, 2007.
[13] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, “Scalable and Efficient Provable Data Possession,” Proc. of SecureComm ’08, pp. 1– 10, 2008.
[14] T. S. J. Schwarz and E. L. Miller, “Store, Forget, and Check: Using Algebraic Signatures to Check Remotely Administered Storage”, Proc. of ICDCS ’06, pp. 12–12, 2006.
[15] M. Lillibridge and Etal, “A Cooperative Internet Backup Scheme”, Proc. of the 2003 USENIX Annual Technical Conference (General Track), pp. 29–41, 2003.
[16] K. D. Bowers and Etal, “HAIL: A High-Availability and Integrity Layer for Cloud Storage”, ACM.
[17] Balajee Maram, Dr. Challa Narasimham, “Enormous Symmetric Cryptography Algorithm Using Image and Double Reflecting Data Perturbation Method”,euroessays.org/wp-content/uploads/2014/03 /EJAE-129.pdf, 2014.
[18] Keke Chen & ling Liu, “Geometric data perturbation for privacy preserving outsourced data mining” , knowledge and Information system journal, springer-Verlag -2011.
[19] Subhasri et al., “Multilevel encryption for ensuring public cloud”, International Journal of Advanced Research in Computer Science and Software Engineering 3(7), July - 2013, pp. 527-532.
[20] Iehab AL Rassan, Hanan Al Shaher, “Securing Mobile Cloud Using Finger Print Authentication”, International Journal of Network Security & Its Applications, Vol.5, No.6, 2013.
[21] A. Gholami and E. Laure, “Advanced cloud privacy threat modeling”, The Fourth International Conference on Software Engineering and Applications, SEAS-2015.
[22] Gholami, J. Dowling, and E. Laure, “A security framework for population-scale genomics analysis”, in High Performance Computing Simulation (HPCS), 2015 International Conference on, pp. 106–114, July 2015.
[23] Gholami, A.-S. Lind, J. Reichel, J.-E. Litton, A. Edlund, and E. Laure, “Privacy threat modeling for emerging biobankclouds”, Procedia Computer Science, vol. 37, no. 0, pp. 489 – 496, 2014. The 5th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN- 2014).
[24] E. Ayday, J. Raisaro, U. Hengartner, A. Molyneaux, and J.-P. Hubaux, “Privacy-preserving processing of raw genomic data”, in Data Privacy Management and Autonomous Spontaneous Security, vol. 8247 pp. 133147, Springer Berlin Heidelberg, 2014.