ABSTRACT:
Cloud computing is latest advancement in computing, it is all about providing computing as a service. Cloud computing also provides storage as a service to users. Large amount of data can be stored on cloud data centers. Cloud data centers are storage devices or platforms used to store data in cloud computing. Confidentiality of data stored in data centers is of utter importance and data needs to be safeguarded from unauthorized access. Traditional encryption methodologies for data confidentiality were costly and time consuming. Here in this work, we propose data confidentiality by fragmenting data on the basis of sensitivity. This paper focuses on fragmenting data into public data segment (PDS) and sensitive data segment (SDS). Public data segment is not a major concern so focus is only on sensitive data segment. For experimental purpose example of Aadhaar is taken and fragmented on basis of sensitivity. Result show proposed method of data fragmentation gives better performance with complex queries and is cost effective as cost of encrypting whole database is much more than fragmenting sensitive data.

Keywords: Cloud Computing, Data confidentiality, Computing Services, Cryptographic Algorithm, Cloud computing environment and Sensitivity based Fragmentation

[1] INTRODUCTION
Cloud computing is an advancement of various technologies coming together for distributed computing to masses. Cloud computing is combined from technologies such as web 2.0, grid computing, cluster computing and parallel computing. There are various definitions [1] to cloud computing, these various definitions summarize cloud computing as “A pool of virtual resources that are useful and accessible, can be used as resources on demand with or without nominal charges”. Cloud computing is long believed vision associated with providing computing as services, it is compared to electricity a shared pool associated with pay on demand service.

Cloud computing provides resources on demand over a distributed network. Data is stored over clouds in cloud computing. Clouds are not the fog up of environment; they are data centers where data can be stored. There are various kinds of clouds [2, 3, 4] available in cloud computing as per requirements. Below is a description of these various types of clouds.

Kumar Pradeep
Public Clouds: Public clouds tend to be available for everyone to work with, they feature ideal effectiveness. Infrastructure and services are offered by websites over the internet. These kinds of clouds are definitely more liable to security violence.

Private Clouds: Private clouds are used pertaining to individual goal such as for employees of a corporation. A specific organization has to buy the structure in addition to services because of its requirements coming from third party. These kinds of clouds tend to be most secure style of clouds because they are not for open access.

Hybrid Clouds: Hybrid clouds tend to be the combination of both private and public clouds. In these types of clouds, some clouds are for open access and other are for organization customers. Hybrid clouds are useful when the organization is neither tiny nor that large being employing general public clouds for their purpose of data storage.

Community Clouds: A community cloud is a term coined together for neighborhood clouds. Only customer of that neighborhood are authorized to access most of these clouds hence making these clouds secure than general public clouds but also limiting their access to limited number of users. Overall performance of community clouds can be good as a result of restriction on number of users.

Data is stored on one of above type of clouds, customer doesn’t know where and how his data is stored. Data confidentiality is required in cloud computing environment because data is stored at afar distance not known to user through distributed environment. Traditionally cryptography technique and algorithm were used for securing data at data centers. These techniques were time consuming because whole database was needed to be encrypted and decryption keys were needed to be generated; also this method of encrypting whole database was costly. After that data was stored at these distributed places by breaking it into pieces called fragmentation. Fragmentation of data should be done in such a way that integrity and secrecy of data is maintained. Fragmenting whole data is a tedious and costly task. There needs to be a mechanism for fragmenting only important data.

This paper proposes a new method of fragmenting data on the basis of sensitivity by using normalization of relational databases. Tables are categorized based on user requirements relating to performance, availability and serviceability and exported to XML as fragments. After defining the fragments and assigning the appropriate confidentiality levels, cloud service providers store all fragments that must remain unlinked in separate locations. So this way data can be fragmented confidentially. Proposed work focus on data sensitivity. This method divides database tables on basis of sensitivity and store sensitive tables with critically sensitive data. In this way sensitive data will become more secure and data confidentiality is maintained.

[2] RELATED WORK

As security of data in cloud computing is a key issue a lot of research has been done in the field of security at data center and in transit phase. These various proposals [4, 6, 7, 8, 9, 10] giving a useful insight into the previous work done towards data security and confidentiality in cloud computing. Grossman [4] described the Case for cloud computing which explains public
and private clouds and distinguishes between clouds that provide on demand computing instances and those providing on demand computing capacity. Computing instance is software as a service and computing capacity is infrastructure as a service. It gives information that clouds are scalable, economical and simple.

Jhawar et al. [6] proposed security requirements for resource management in cloud computing. This paper addresses the problem of guaranteeing security with additional consideration on reliability and availability issues for management of resources in cloud environment. This model considers security issues in virtual machine allocation. Security constraints from users as well as service providers are considered. Bin-packing allows VMs to be placed on top or beside each other on physical hosts. Silva et al. [7] proposed a methodology for management of cloud computing using security criteria. This model proposes GQM methodology to develop a cloud computing security metrics hierarchy; a security index is produced that describe security level. In second step security index is used to compute an allocation index that helps in setting management priorities with a security bias. So they avoid under using of resources and increase response time. GQM is goal-question-metric, goal is conceptual level, question is operational level and metric is quantitative level. Problem with this model is that it is highly dependent on human intervention; further comparison of strategies for management of cloud computing in this model can be extended.

Zissis et al. [8] proposed securing e-Government and e-Voting with an open cloud computing architecture. This report describes that e-government and e-voting can become more secure by using cloud computing with cryptography. This report was based on the principle that system security cannot be stronger than its weakest link. By using cloud computing user’s terminal can be put under security perimeters. Public key encryption could be added in this model to overcome threats in e-voting and scope of e-government could be further broadened.

Hudic et al. [9] featured data confidentiality using fragmentation in cloud computing. This paper proposes fragmentation of data in cloud computing environment and still having the feature of confidentiality. They focused on fragmentation of data with constraints other than encryption. Dimitros Zissis et al. [10] proposed introduction of a Trusted Third Party, tasked with assuring specific security characteristics within a cloud environment. They used cryptography, specifically public key infrastructure operating in concert with SSO and LDAP for ensuring integrity, authentication and confidentiality of data and communication involved. This paper presents a horizontal level of services available to all involved entities.

Traditional methodologies of cryptography and fragmenting whole database were time consuming affairs and cost of these methodologies was on high side. Our approach of fragmenting data on basis of sensitivity is different from traditional approaches by taking into consideration only critical data. Proposed approach takes less time than adaptive approaches and is cost effective because only sensitive data is fragmented and data that is not important is not taken into consideration.

[3] PROPOSED MODEL FOR DATA CONFIDENTIALITY

Proposed model is about securing the data while it is residing in the storage location. This inherently triggers the need for designing ways to store and retrieve data. This model fragment data on basis of sensitivity in two categories first is public data segment (PDS) which
has no value for intruder. Second category of sensitive data segment (SDS) is of more importance. It needs to be saved from access by intruder and mapped to PDS for providing useful information. Sensitivity of data is divided into three categories low, mid and extreme. An illustration is Aadhaar account information, divided into tables below.

Table 1: Basic_Info
| Aadhaar Number (PK) | Name | DOB | Gender | Address | Mail-id |

Table 2: Bank_Info
| Aadhaar number | Bank Name | Account Number (PK) |

Table 3: Account_Info
| Account Number (PK,FK) | Password | Password Question and Answer |

Table 4: Transaction_Info
| Aadhaar Number (PK) | Password | Transaction Password |

Table 5: Aadhaar_Account_Info
| Aadhaar Number (PK) | Account Number (PK) |

So if an intruder gets unauthorized access to Basic_Info it is not of much importance for that person. Bank_Info alone is not useful but Transaction_Info is extremely sensitive because an online transaction can be done only by this table. So transaction_Info is sensitive and important data. This model store related data at different locations but mapping is done only at any update or query. The sensitivity for Aadhaar is as follows Basic_Info and Bank_Info-Low sensitivity, Account_info-Mid sensitivity, Transaction_Info-Extreme sensitivity, Aadhaar_Account_Info-Mid sensitivity. Fragmentation is done by taking security constraint into consideration and mapping is done by sensitivity concept as given in tables below.

Table 6: Frag_Transaction_Info
| Datafragmentation_ID (PK) | Password |

Table 7: Frag_Transaction_Sensitivity_Info
| Datafragmentation_ID (PK, FK) | Transaction Password |

Table 8: Frag_Transaction_Mapping_Info
| Aadhaar Number (PK) | Datafragmentation_ID (PK) |

Sensitivity for each table stored in Cloud database is calculated. In case of high sensitivity input table is split into number of tables. If a table is of mid sensitivity then it will be checked for security level and if security level is high it will be split into number of tables. In case of low sensitivity there is no split of tables. Mapped table that connects the new split tables to cloud database is created. Then mapped table is stored at unused server. Mid level sensitive and extreme sensitive tables are assigned to unused servers. Tables which are still sensitive
after fragmentation are stored with tables critical for Cloud database. Result will be in form of tables fragmented on basis of sensitivity.

[3.1] PROPOSED ALGORITHM FOR DATA FRAGMENTATION

1. Initialize number of Tables (N) and number of resources (M)
2. Set data servers used to False
3. Initialize j=1
4. Repeat for j=1 to m
5. Read the sensitivity level of each table from metadata
6. If Table [j] sensitivity is low then assign the table [j] to unused server
7. Else if Table [j] sensitivity is mid and required security level is high split the Table [j]
8. Else if Table [j] is of extreme sensitivity split the Table [j]
9. If required security is high split table at one more level
10. Create mapped table for split table
11. Store split table and mapped table on unused server
12. Repeat step 4 to 10 until 1<j>n
13. Initialize number of fragmented tables to f
14. Assign mid-level sensitive tables to unused servers
15. Assign extreme level sensitive tables to unused servers
16. If after fragmentation Table [f] sensitivity is not of mid-level then store the table on server where midlevel sensitive table is stored
17. Else store table on the server where critical table is stored
18. Repeat step 16 to17 for k=1 to f

[4] RESULTS

Traditional methodology of securing data at data center was encryption. To compare proposed methodology with traditional methodology a database with 24 master tables and 58 transactions is considered. First line in the figure below is drawn on the basis of traditional methodologies and then line based on proposed data fragmentation methodology is drawn. There were 46 critical entries, 92 sensitive entries and 12 mapper tables. Figure 1 shows that as complexity of queries increases performance of proposed methodology in comparison to traditional methodologies increases and proposed methodology start giving better performance.

![Figure 1: Performance Analysis of Data Fragmentation](image-url)
[5] CONCLUSION

This paper investigates issue of data confidentiality in cloud computing environment. For data confidentiality this paper proposes an algorithm for sensitivity based data fragmentation. Though it is difficult to implement this model in real time environment, it gives multi tier security to data in data center. In this model for an intruder to get access to data he needs to get access to sensitive data stored with critical tables which is a tough task. Figure 1 shows that as complexity of queries increases this model produces better performance than traditional encryption technology. Also encrypting the whole database is a tedious and costly task instead database administrator can focus on securing sensitive data. Thus data fragmented on basis of sensitivity makes data more confidential and is cost effective.

[6] ACKNOWLEDGMENTS

Work of this magnitude is only possible with the hand of help from many people. First and foremost, I wish to express my deepest gratitude to my supervisor Dr Dilbag Singh (Associate professor, Dept. of Computer Science and Applications, Chaudhary Devi Lal University, Sirsa) for his untiring guidance. I want to express my deepest gratitude to my parents and friends who have always supported me in whatever decision I have made. Finally I am thankful to GOD for blessing me much more than I deserve.
REFERENCES


