ISSN 2347 - 3983
Volume 8. No. 9, September 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter228892020.pdf
https://doi.org/10.30534/ijeter/2020/228892020

Enhanced Data security through Deep Data Classification in the Cloud Computing

Dorababu Sudarsa[1],PeddadaVenkateswara rao[2], Rokesh Kumar Yarava[3], V RavitejaKanakala[4]
[1,2,4]Assistant Professor, Department of Computer Science and Engineering, KoneruLakshmaiah Education Foundation, Vaddeswaram, AP, India.
[3]Assistant Professor, Department of Computer Science and Engineering, Mallareddy Engineering College(A), Hyderabad, India.
[1]dorababu.sudarsa@gmail.com, [2]pvrao@kluniversity.in, [3]rokeshy1@gmail.com,[4]raviteja.kanakala@gmail.com

ABSTRACT

Data is a valued asset especially when data moving to the cloud. Cloud computing model attracts different users because of its various benefits such as scalability, simplicity, expense reduction, and high resource resistance. Cloud computing is a wide-ranging architecture based on different models for providing different software services and hardware services. There may be different types of data and the degree of protection needed for all the data is also may vary. Data privacy and data security is the major area of research and development in the cloud computing. Privacy protection and data leakage became essential thing for so many organizations moving on to cloud. Several classification methods that are expressed various parameters based on different dimensions for providing the data security can be based on the level and are required protection. But, the latest approaches introduced by the cloud, related to distributed resources, multi-tenancy idea, computation outsourcing, high dynamism of the model, data warehousing and the non-transparent style of cloud increase the security and privacy concerns and makes building and maintaining trust among cloud service providers and consumers is a serious security issue. In this paper, we proposes a new approach to improve security of data in cloud computing. It proposes a classification model to classify data before being presented into a suitable encryption system based on the category. Because data in cloud has not the same sensitivity level, encrypting these different types of data with the same algorithms can lead to a lack of security or resources. By this scheme, we can achieve high security for our data, and can optimize the computation cost and resource consumption while ensuring data confidentiality.The efficiency of the proposed classification scheme is analysed with the sample dataset collected.

Key words: Cloud computing, Cloud storage, Data security, Data privacy, Data classification, Data sensitivity.

1. INTRODUCTION

In general data can be represented as set of raw facts. Data is a big asset of any organization. Data could be in any of the forms like numbers, letters, words, images, audio, video etc. Data privacyand security is a key issue for any organization. Data can be structured or unstructured. Majority of organizations maintain their data as structured manner, so that very easy to classify, manage, and access. Usually classification process of structured data found in spreadsheets and databases are less complex and time consuming to manage than those of unstructured data such as source code, documents, and email. Irrespective of whether structured data or unstructured data, it is very important for organizations to manage the data sensitivity. When data classification properly implemented, it ensure that sensitive or confidential data assets are managed with greater mistake than data assets that are considered public or free to distribute.

The cloud computing refers to a model for supportingflexible and convenient, global, on-demand network access to a sharable group of configurable computing resources (e.g., servers, storage, networks, applications, and services) that can be quickly provisioned and released with minimal efforts to manage or service provider interaction [1].It lessens the need of user involvement by complicating technical details such as licenses, software upgrades, and maintenance from their clients [2]. We can understand clearly about the cloud and its environment through the figure1 shown in below.

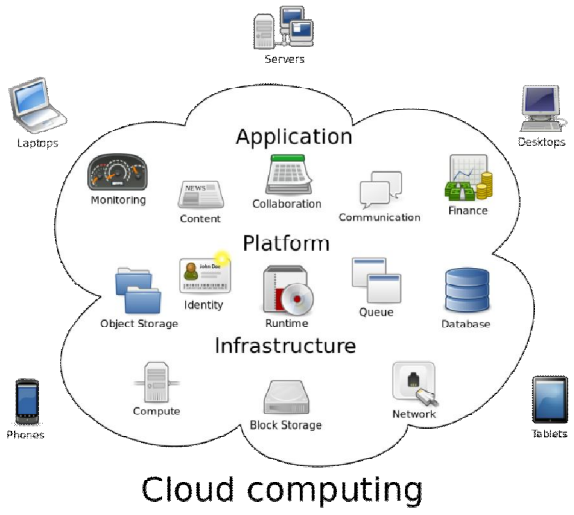

Figure 1: Cloud Computing Environment

Though, the innovative methods presented by the clouds, related to outsourcing computation, multi-tenancy concept, distributed resources, high dynamism, data warehousing and the non-transparent model of cloud maximize the security and privacy issues and makes a critical security challenge [3] forconstructing and hold trust among cloud service providers and cloud consumers.

When a user stored his data in cloud or to use a cloud service whereever his data is involved, majority times, he cannot pay attention to know the details of the security risks faced there; attacks, vulnerabilities, malware information and security policies established in the cloud. For instance, an attacker can introduce a piece of code into web application to ignore access control mechanisms, to gain free access to all customer's data, authorized data, tokens and also plaintext passwords by cross locate scripting attack. Also, attacks such as denial of service(DoS) attack, flooding attack, Wrapping attacks, XML Signature and others lead to taking the administrative rights of the cloud user or to making data unavailable [4]. In these scenarios the user loses the control of his data and especially the cloud service provider or an external party can read and use it for their benefit. To touch a high privacy and security goals associated to data and cloud services, cloud service provider(CSP) establish a Service Level Agreement (SLA) to the cloud consumers, but unfortunately no standard procedure provided to conceive an SLA. SLA reports associated to the approved services, which is very helpful for both consumers and provider. However, these SLA reports do not fully conform to the consumer needs in terms of security. Several cloud providers does not provide sufficient SLA to assure that user data is fully secure [5]. If cloud computing is feasible alternative to the traditional systems and gain user's trust, cloud's infrastructure must implement and provide a security at a level that ensure the data protection and guarantee to its confidentiality,data integrity and availability. Consequently, a more awareness is required about tools and measures to take in order to avoid malicious activities that target data.

In [15], Frank Simorjay et al. given some clarification on the different parameters or factors from which we should know how to achieve the data security and confidentiality,such as data access control, Authentication, Authorization ,etc. They described about them as follows:

Data access Control

To control the data accessing from the user, we use Authentication and authorization. Sometimes Authentication and authorization are confused with each other and their roles misinterpreted. In reality they are quite different, as shown in figure below:

Figure 2: Difference between Authentication and Authorization.

Authentication usually consists of at least two things: a user ID or username to identify a user and another is a token such as a password, to confirm that the username credential given is valid or not. it verifies that the user is a valid user or not. Authorization is a process of providing the ability to access data set, data file, an application, or some other object to an authenticated user. Assigning the rights for use, modify, or delete items to authenticated users that they can access, needs care to data classification. The difference between Authentication and Authorization is described in the figure2 very well. The validity of data in the sense, data which is recorded and used with respect to related requirements that should be valid over a given period of time. Relevance means, the data captured should be used with respect to relevant requirements. Completeness of data means data should be complete or fulfilled with respect to its usage. Data accessibility deals with the access of data with respect to time, scope and cost. Data consistency deals with the uniformity of the content which is stored with respect to changes made by transactions that used data.

Basic security issues of data comprise confidentiality, availability and integrity. Data confidentiality deals with the privacy of data that comprises authentic and authorized accessibility of sensitive data. Data integrity deals with data consistency and data accuracy which are required to achieve integrity. Data availability issues relating to fool proof storage, storage type, requirements for disaster recovery and also backup plan.Data availability is most important issue to any organization moving to the cloud. Security issues concerning data increases when moving to the cloud. User controls about data, data protection technique and data availability are some of the issues that user requires to know before using the cloud for data storage. The data stored on the cloud should be protected from unwanted malicious disaster that may be man-made or natural. A cloud provider should know and accordingly need to provide measures to achieve data availability.

Data Classification Process

Many organizations understand the need of data classification and want to implement it is a basic challenge. One effective and simple implement of data classification is by using the PLAN, DO, CHECK, ACT model from MOF. The figure3 below shows the tasks that are needed to implement successfully in data classification model.

Select a suitable model that addresses your needs

Many types of processes are present for classifying the data, like manual processes, location-based processes that classify data based on a system's or a user's location, application-based processes such as database-specific classification, and automated processes used by various technologies.They introduced two generalized terminology models based on well-used and industry-respected models. Both of these terminology models, provide three levels of classification sensitivity, which are shown in the table1 below.

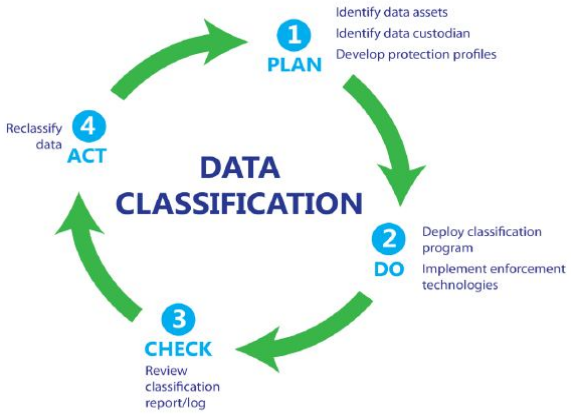

Figure 3: Steps in classification process

Table1: Levels of classification sensitivity

Sensitivity	TerminologyModel 1	TerminologyModel2
High	Confidential	Restricted
Medium	For internal use only	Sensitive
Low	Public	Unrestricted

Data Ownership

To provide the security for the data it's important to launch a clear chain of ownership for all the data assets. The table2 given below identifies different data ownership roles in data classification and their relevant rights.

Table 2: Data ownership roles and their relevant rights

Role	Create	Modify/delete	Delegate	Read	Archive/restore
Owner	X	X	X	X	X
Custodian			X		
Administrator					X
User*		X		X	

The data owner is the original creator of the data file, who can give ownership and assign a guardian (called custodian) . When a file is created, the owner should assign a classification, they have a responsibility to understand what needs to be classified as confidential data based on their organization's policies. Data of all owners can be auto-classified for internal use only (sensitive) unless creating or owning confidential (restricted) data types. Frequently, the owner's role will change after the data is classified.An **administrator** is a user who is responsible for ensuring the data integrity, but they are not a data owner, custodian, or user. The administrator role includes backup, restoration, maintaining records of the data,, and choosing, obtaining, and operating the devices and storage that stores the data assets. The **asset user** is anyone who is granted access to data or a file, access assignment is often given by the owner to the custodian.

To implement these classification process, we need to consider details about who, where, when, what, and why a data asset would be used, changed, deleted or accessed. Sometimes reclassifications also need to be done. Reclassifying is a changing of classification state of a data when a user or system defines that the data risk profile or importance has changed. This reclassification is very important for ensuring the classification status continues to be current and valid. Reclassification can be done either manually or automatically.

The classification of data process has been used for decades by large organizations such as Microsoft, governments, private organizations, and military entities to manage the integrity of their data. Organizations which are assessing cloud computing for future use or organizations that are presently using cloud services and looking for ways to optimize data management will benefit from this paper. Data Classification is a method of defining different data levels and determining a level of sensitivity to it. It is a crucial activity at different stages as data is being created, stored, modified, or transmitted. The classifications of the data determine up to what extent the data needs to be secured and its value in terms of Business Resources. Data classification can be done based on the different aspects. Some based on the risk related with the disclosure are public, internal, confidential (or highly confidential), restricted, regulatory, or top secret. And some based on its way of creation are user personal data, their usage patterns etc.

To provide the access control and authorization, classifying data based on security levels, many organizations felt it as area of interest by using or providing cloud services. In this paper, we have studied a set of classifications procedures given in the literature survey and recognized a set of parameters based on the security needs for cloud data. We have analysed some sample data sets that provide the security based on their usage and access control with respect to cloud computing environment.In the literature, we mentioned several mechanisms implemented in the cloud infrastructure to protect data. Moreover manymethods were proposed to secure the data at various levels, on different platforms and different domains such as network, hardware, software, hypervisors, storage, access control, classification model to classify the data before being presented into an applicable encryption system. Because data in the cloud has not same sensitivity level, encrypting it using same algorithms can lead to lack of resources or of security. Hence,it is needed to optimize computation cost and the resources consumption while ensuring the data confidentiality.

2.PROPOSED SYSTEM MODEL

The system model of our proposed scheme is given in below figure4. In the cloud virtual environment, the data can be taken from the different local servers to store into the cloud. Before storing that data into the cloud,it is classified based on the application areas belonging that data, then reclassifying it deeply based on its sensitivity and its properties. Finally, the fully classified data can be store into the cloud environment as different type of clusters in such a way that, it is possible to apply security algorithms on a specific data set , so that it can be high protected.

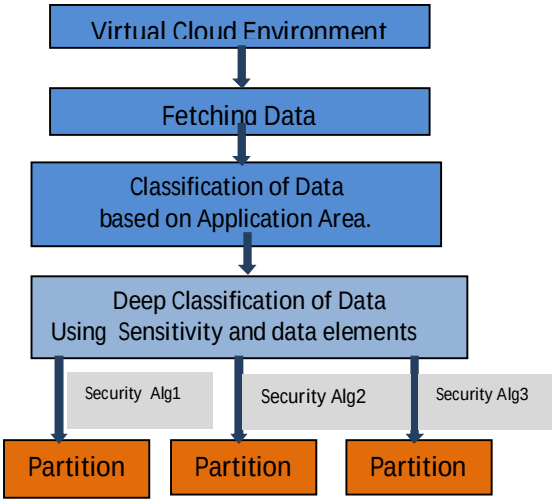

Figure 4: Cloud simulation Environment

Data security is the key issue in cloud computing. Though, several experimentations had been done to solve this issue,still different data leakage problems are facing in the cloud. In this paper we concentrate on the data stored in the cloud and to secure it by adopting a classification method. Weobserved that different types of data are loading through the cloud infrastructure. Some customers need more security and some needs more confidentiality and protection than other for their data. But reality exist in the in cloud environment is different; a cloud provider accomplishes the same models or mechanisms for security for different types of data, or sometimes simply they uses same security procedures for allclient's needs. Itleads to waste of time, a loss of cost; and a misuse of resources. We identified some of the research work existing in the literature related to this paper.The main idea is the classification of data before applying to any security technique that optimize its security.

Barker et al. proposed a new privacy data model based on data classification. They recommendto build a system that should have the capability of protecting user privacy by using the three fundamental aspects such as purpose, visibility and granularity. They spited granularity into the four sub-aspects (None, Existential, Partial, and Specific). The visibility is also divided into the five sub-aspects (None, Owner, House, Third party, All/Work). And for the Purpose, they proposed five alternatives aspects (Single, Reuse same, Reuse selected, Reuse any, and Any) [6].

3.LITERATURE SURVEY

In [1], the sensitive data is outsourced to a cloud service provider with the authorisation of block level modifications. Unplanned mutual belief is established between data owner and user by using third party. Pearson discussed policies and assessment procedures in [2] for privacy enhancement methods and tools. Privacy is provided in terms of legal obedience and user belief, data leakage for sensitive data. Ji Hu and Klein A, in [3] gave a benchmark to secure the data-in-transit in the cloud. And also protecting the data during migration is discussed via benchmark for the encryption overhead and security. Additional encryption is needed for high security but it needs more computation. So a benchmark provides balance for the security and also encryption overhead.Tetsuya M, Kazuhiro S and Hirotsugu, K. provided a large scale search system for the information exchange between the internet communities that leads to formation of hidden Channels [4]. An agent based security model used to control data from hidden channel is presented. It may solve the problem of data leakage in the cloud.In [5] discussed aboutthe privacy issue by holding data control to user to increase confidence, also discussed cloud computing attacks and some provisions to overcome from the same.

In [6], aninnovative patient-centric framework is presented for data access control of Patient Health Records,where data is stored in semi-trusted servers. Attribute based encryption techniques used to encrypt each patient's Health Record file to achieve fine-grained and scalable data access control.In [7], data characteristics are studied with respect to the online social networks.They have identified the information and its leakage at the time of data sharing in the network. For this purpose they also used third party check.

Data protection organisation is proposed in [8] by considering several aspects. A case study is inspected by providing the solution for data protection in terms of questionaries' like who needs protection, what has to be protected etc. A restriction to data accessibility by the third party applications is proposed in the form of framework in [9]. Policies are applied for restricted access and therefore the privacy of user confidential data is achieved from the third party applications.Data privacy is achieved to the user that makes use of the hotspots for the internet access in [10].

In [11],a three dimensional vision for data taxonomy is proposed. This classifies the data as per visibility, and Granularity. Various levels of data are defined along these dimensions to achieve the data privacy. Taxonomy for social data is presented in [12], which classifies the data based on the way data is generated in the social network, and hence the privacy and access rights have to be applied. Data classification at various phases in a social network is presentedin [13], which classifies the data based on security parameter confidentiality in a network. This classification applies in several phases of data like collection, processing, dissemination and invasion.

In [14], data security andvarious security issues areanalysed, and a trust based solution is proposed for the same. Here data security concerns have identified and attempted to provide security by classifying the data. Many dimensions are identified along which security and protection level can beapplied to different data. Sometimes security need to check in the different levels like PaaS,SaaS and IaaS,Dr.N.Srinivasu et al. in[20] found threats at all these levels and resolved them by using the security mechanisms..

In [16], RasmeetKour et al. proposed a classification mechanism to classify thedata depending upon the security elements of the data.Thiswill divide data into two parts-sensitive and non-sensitiveby improved bagging and boosting technique.In [17], Kumar Pal Singh et al. proposed a mechanism for data classification based the data

sensitivity. Depends on the sensitive rating of the data in terms of confidentiality, integrity and availability, data can be classified into three categories:low impact, moderate and high impact. Hence the security algorithms can be applied on to the data based on suitability. It is also giving good results.In[15], RizwanaShaikh et al. given the data classification mechanism through that they classified the date based on the three properties: Access control, Content and Storage. Each property again divided into sub-properties as given in figure5.This method given detailed classification up tosome level. The value of data is identified based on their usage and access control restrictions.

Data Classification Properties		
Access Control	**Content**	**Storage**
Frequency of Access Frequency of Update Visibility and Accessibility Retention	Precision Accuracy Reliability/Validity Degree of Completeness Consistency and Auditability	Storage-encryption Communicationencryption Integrity Access Control Backup and recovery plan Data Quality Standards

Figure.5: Data Classification Properties

Content of data have several properties with respect to its modification such as Precision/ Accuracy, Reliability/ Validity, Degree of Completeness, Consistency, Auditability,
Data storage policies can be applied based on the norms and constrictions applied to the different types such as Storage-encryption, Communication-encryption, Integrity, Access Control, Backup and recovery plan, Data Quality Standards. Access Control category defines the access restrictions we can applied on data. That includes Frequency of access, Frequency of update, Visibility and Accessibility, Retention etc.

They Simulated the classification of data based on personal data set elements like name, addresses etc. is taken as sample data and analysed. They have used the subjective criteria to classify them and based on that, security requirements for the storage and communications can be incorporated. These criteria can be transformed to values and threshold can be set for objective evaluation.

Garigipati,N.et al. presented a study on the data security and the query privacy in cloud in [27], thru that we can get some security issues. In [28] Keerthi,G et al. presented the source based privacy for the confidential data in the cloud based e-healthcare systems. In [29], Mane,P.M et al. presented a method to achieve confidentiality and efficient access control of data in cloud using CP_ABE.

After classifying the data, we need to apply the security mechanisms to obtain the high security for our data. For instance, in [21], N.Srinivasu et al. proposed a security algorithm named Honey Encryption which is a different technique that solve the brute force problems also.Some techniques are used to enforce the security mechanisms like used in [24].Babukarthik et al.proposed a mechanism for data owners to execute their security policies to confirm data confidentiality and integrity, which also enable trusted data sharing through untrusted cloud providers.In[25],different decisional learning parity with noise (DLPN) named as key-Ordered DLPN based security algorithm is proposed, here DLPN is extended to an even-odd-order mechanism for the encryption, in this, odd and even bits are input integer values for the key generation algorithm in that security method.

In centralized controlling cloud, the Host Controller(HC) manages the resources such as virtual machine (VM) and physical hosts across hosts in the Data Center (DC).If HC is down,the services hosting by several hosts in DC also will down. Hence, decentralized VM management used to avoid by taking one of the hosts from DC as HC to keep DC in running.The host's in DC does VM migration across various hosts incorporate the security mechanism on hosts to protectdata in migration[22]. It is mandatory to allocate the resources and using good and suitable scheduling algorithm to complete the given task in the cloud in minimal amount of time. In [23], S.Phani Praveenet. al, proposed a method with two phases as resource allocation and tasks scheduling. Efficient resource allocation is done using social group optimization scheme and tasks scheduling using shortest-job-first scheduling algorithm. In [30],Naresh et al. presented the method to optimize the resourceschedule in the cloud.Sometimes we need to prevent the data attacks like, In [26], Chaitanya,G.K., et al. proposed a method to prevent data theft attacks inIaaS cloud through the trusted computing

4.PROPOSED CLASSIFICATION METHOD

From the above literature, we can understand that data classification is very essential method to enhance the data security in the cloud both at storage level as well as access level. So, in this paper we proposed a mechanism for achieving the security to the next level, by classifying the data into some more deep.
Data can be classified into different categories based on its sensitive level or rating as low sensitivity, moderate sensitivity and high sensitivity. The data should be stored and maintained based on its context. The sensitivity level can be defined to the different type of data whether it is public, private, personal, official, health related data, criminal issues data etc.For example, some sample data is shown in the table3 and table4 with their sensitive levels.

Table 3: Different Type of data and its Sensitivity (For sample)

Type of data & its Sensitive level	
Type of data content	**Sensitivity level**
Public data	Low
Private data	Moderate
Personal data	High(Restricted)
Official data	Moderate
Health related data	High
Criminal issues data	High
Hospital Data	High
Research Data	High

In addition to the sensitive level, we described some properties of data in terms of security, integrity, availability of data as said in [15]. If the data classification is applied according these properties and based on the above specified

sensitive level at the storage level of cloud, then very easy to maintain and access it. After the data classification is done with these properties and sensitivity, we can apply the suitable security algorithms to achieve optimal security of the out sourced data. Instead of applying the security mechanisms on the entire data blindly without concentrating on the type and sensitivity of the data, if we apply based on the specific type and sensitivity of the data, we can achieve the optimized security. Once the data is outsourced to the cloud, in general auditing can be done to verify the data integrity as done in [18].

Table 4: Sample Data content and its Sensitivity.

Data content	Sensitive level
Voter name	L
Voter Address	H
Voter age	M
Voter gender	M
Voter ID	M
Voting booth name	H
Movies name	L
Story of the movie	H
Schedule of movie	H
Director of movie	M
Profit of movie	H
Remunerationof director	M
CoursesOfferingUniv,	L
Course fee	M
Course duration	M
Course Content	H
Course material	H
Questionson the course	H
Shopping malls addr,	L
Malls staff details,	H
Malls profit	H
Mall offer coupons	H
Vaccine name	L
Vaccine life time	M
Vaccine formula	H
Vaccine purpose	M
Patient Name	M
Patient age	M
Patient disease	M
Patient report	H
Doctor of Patient	M
Treatment taking	M
Govt officials name	L
Govtofficials Designation	M
Govt officials address	H
Govt officials schedule	H
Govt officials salary	H
Govt officials files	H

Table 5: Data classification parameters: Man= Mandatory; R= Restricted; Mod= Moderate; FU= Frequency of Update; FA= Frequency of Access; CE= Communication Encryption; SE= Storage Encryption; Int.= Integrity; Con/Aud=Cosistency/ Audibility; DoC= Degree of Completeness; R/V= Reliability/Validity; P/A= Precision /Accuracy. H-High

Data set / properties	Properties related with Security				Properties related with datacontent				Properties require for cloudStorage			
Data Elements	**FA**	**FU**	**Visibility**	**Accessi Bility**	**P/A**	**R/ V**	**DoC**	**Con/ Aud**	**SE**	**CE**	**Int.**	**AC**
Name	More	Never	All	All	Hig h	High	Man	Yes	Less	Moderate	Less	No Control
Mother name	Moderate	Never	R	R	H	High	Man	Yes	Strong	Strong	Strong	RBAC
Father Name	Moderate	Never	R	R	H	High	Man	Yes	Strong	Strong	Strong	RBAC
Permanent Address	Moderate	Never	R	R	H	High	Man	Yes	Strong	Strong	Strong	RBAC
Commn Address	Moderate	Less	R	R	H	High	Man	Yes	Strong	Strong	Strong	RBAC
Qualificatn	Moderate	Less	All	R	H	High	Man	Yes	Strong	Strong	Strong	RBAC
Profession	Less	Less	All	R	H	High	Man	Yes	Strong	Strong	Strong	RBAC
Salary	Moderate	Less	R	R	H	High	Man	Yes	Strong	Strong	Strong	RBAC
Designation	Moderate	Less	R	R	H	High	Man	Yes	Strong	Strong	Strong	RBAC
Phone	Moderate	Less	R	R	H	High	Man	Yes	Strong	Strong	Strong	RBAC
Mobile	More	Less	R	R	H	High	Optional	Yes	Strong	Strong	Strong	RBAC
DOB	Less	Never	R	R	H	High	Man	NA	Strong	Strong	Strong	RBAC
Place of Birth	Less	Never	R	R	H	High	Optional	Yes	No	Less	Less	RBAC
Caste	Less	Never	R	R	H	High	Optional	Yes	Strong	Strong	Strong	RBAC
Sex	Moderate	Never	All	All	H	High	Man	NA	Less	Moderate	Less	No Control
Nationality	Less	Never	All	All	H	High	Man	NA	No	Less	No	No Control
Blood	Less	Never	R	R	H	High	Optional	Yes	Less	Moderate	Less	RBAC

Group												
Height	Less	Never	R	R	H	High	Optional	Yes	Less	Moderate	Less	RBAC
Weight	Less	Never	R	R	H	High	Optional	Yes	Less	Moderate	Less	RBAC
Marital status	Moderate	Less	All	All	H	H	Man	Yes	Less	Moderate	Less	No Control
Adhaar No	Less	Never	R	R	H	H	Man	Yes	Strong	Strong	Strong	RBAC
PAN	Less	Never	R	R	H	H	Man	Yes	Strong	Strong	Strong	RBAC
Voter Number	Less	Never	R	R	H	H	Man	Yes	Strong	Strong	Strong	RBAC
Passport No	Less	Less	R	R	H	H	Man	Yes	Less	Moderate	Less	RBAC
SSN	Moderate	Never	R	R	H	H	Man	NA	Strong	Strong	Strong	RBAC
Email	More	Less	R	R	H	H	Man	Yes	Strong	Strong	Strong	RBAC
Physical Charactrstics	Less	Never	No-one	No-one	H	H	Optional	Yes	Moderate	Strong	Moderate	No-one
Dept.	Moderate	Less	All	R	H	H	Man	yes	Moderate	Strong	Moderate	No-one
Eye Color	Less	Less	No-one	No-one	H	H	Optional	NA	Strong	Strong	Strong	No-one
Biometric	Less	Less	No-one	No-one	H	H	Man	NA	Strong	Strong	Strong	No-one
Income	Less	Less	R	R	H	H	Man	yes	Strong	Strong	Strong	No-one
No.of Children	Less	Less	R	R	H	H	Man	yes	Strong	Strong	Strong	No-one
Health Status	Less	Less	R	R	H	H	Man	yes	Strong	Strong	Strong	No-one
Police Cases	Less	Less	R	R	H	H	Man	yes	Strong	Strong	Strong	No-one
Office data	Less	Less	R	R	H	H	Man	yes	Strong	Strong	Strong	No-one

5. ANALYSIS

By observing the literature survey, we can say that, some mechanisms are providing security only based on the security algorithms on the entire data, some are providing only based on classification, some are providing based on both classification and security mechanisms. Though they are providing security, which may not give as optimal, since they are doing only basic level classification The classification process given in this proposed method will definitely will give optimal level security for our data which is outsourced to the cloud, since here classification is doing deeply and then applying the security algorithms on the sensitized and classified data. To perform the data classification automation, we should use any data classification algorithms. If we apply the data classification algorithm and data security algorithm, definitely we can reduce the execution cost of both secure data storage as well as retrieval.

6. CONCLUSION

Data security and privacy is one of the major issues facing while data out sourcing to the cloud storage. Even though somany classification methods exist in the literature survey, they cannot reach to the optimal level security for our data. We proposed amethod, in which we can identify the different type of data with their sensitivity, and we can divide the data into smaller to identify their sensitivity clearly, so that we can store the data into the cloud as clusters and sub-clusters, hence it is possible to access very easily with the proper access rights. Moreover,we identified a set of parameters based on these and data sensitivity, classification can be done deeply and then suitable security algorithm can be applied on a specific data set, not on the entire data. And then data can be stored intothe cloud.So, here we providing security levels based on type of content and accessibility, and in cloud storage as per the required confidentiality and accessrestrictionsforthedataspecified.Wehaveanalyzedfewdataelements andclassifiedthembasedonthe proposed parameters. The classification provisions can be given for storage and communication encryption, integrity and access control mechanisms. Also a regularized backup plan can be decided for disaster and recovery.

REFERENCES

1. AyadBarsoum and Anwar Hasan, "Enabling Dynamic Data and Indirect Mutual Trust for Cloud Computing Storage Systems", IEEE Transactions on Parellel and Distributed Systems, Dec. 2013 (vol. 24 no. 12), pp. 2375-2385.
2. Pearson S, "Taking account of privacy when designing cloud computing services", Software Engineering Challenges of Cloud Computing, pages, 44 – 52, Vancouver, BC,2009.
3. Ji Hu and Klein A, "A Benchmark of transparent data encryption for migration of web application in cloud", Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, pages 735-740, Chengdu,2009.
4. Tetsuya M, Kazuhiro S and Hirotsugu, K., "A system for search, access restrictions and agents in the Clouds", Ninth Annual International Symposium on Applications and the Internet Cloud, Pages 201-204, Japan,2009.
5. Descher M, Masser P, Feilhauer T, A Min Tjoa and Huemer D, "Retaining data control to the Client in Infrastructure Cloud", International Conference on Availability, Reliability and Security, pages 9-16, Dornbirn,2009.

6. MingLi,ShuchengYu,YaoZheng,KuiRenandWenjingLou,"ScalableandSecureSharingofPersonalHealthRecordsinCloudComputing using Attribute-based Encryption", IEEE transaction on parallel and distributed systems, pages 131-43 vol. 24, issue 1, 2012.
7. Balachander Krishnamurthy and Craig E. Wills, "Characterizing Privacy in Online Social Networks", Proceedings of the first workshoponOnline social networks, WOSN '08, Pages 37-42, ACM New York, 2008.
8. Mike Dutch, A Data Protection Taxonomy, Storage Networking Industry Association, June2010.
9. Yuan Cheng, Jaehong Park and Ravi Sandhu, Preserving User Privacy from Third-party Applications in Online Social Networks, Proceedings of the 22nd international conference on World Wide Web Companion, Pages 723-728. Geneva, Switzerland,2013.
10. Ningning Cheng, Xinlei (Oscar) Wang, Wei Cheng, Prasant Mohapatra, ArunaSeneviratne, Characterizing Privacy Leakage of Public WiFi Networks for Users on Travel, IEEE International Conference on Computer Communications, Italy,2013.
11. Ken Barker, Mina Askari, Mishtu Banerjee, KambizGhazinour, Brenan Mackas, Maryam Majedi, Sampson Pun, and Adepele Williams, A Data Privacy Taxonomy, Advanced Database Systems and Applications Laboratory, University of Calgary, Canada,2009.
12. Bruce Schneier, A Taxonomy of Social Networking Data, The IEEE Computer And Reliability Societies, August2010.
13. Sergio Donizetti Zorzo, Rodrigo Pereira Botelho, Paulo Muniz de Ávila, Taxonomy for Privacy Policies of Social Networks Sites, Published Online, Social Networking, 2013, 2, 157-164 October 2013(http://www.scirp.org/journal/sn).
14. RizwanaShaikh and Dr. M. Sasikumar, "Security Issues in Cloud Computing: A survey. International Journal of Computer Applications 44(19):4-10, April2012.
15. Frank Simorjay "Data classification for cloud readiness" Microsoft Trustworthy Computing, 2014 Microsoft Corporation
16. RizwanaShaikh and Dr. M. Sasikumar "Data Classification for achieving Security in cloud computing" International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).1877-0509 © 2015.
17. RasmeetKour, SupartiKoul and Manpreetkour, "A Classification Based Approach For Data Confidentiality in Cloud Environment" 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS),2017
18. Kumar Pal Singh, Dr. VinayRishiwal and Dr) PramodKumar,"Classification of Data to Enhance Data Security in Cloud Computing" IEEE 2018.
19. Rokesh Kumar Yarava and Rajendra Prasad Singh, "Efficient and Secure Cloud Storage Auditing Based on the Diffie-Hellman Key Exchange",*International Journal of Intelligent Engineering and Systems, Vol.12, No.3, 2019.*
20. N. Srinivasu, O. SreePriyanka, M. Prudhvi and G. Meghana," Multilevel classification of security threats in cloud computing", International Journal of Engineering & Technology, 7 (1.5) (2018) 253-257
21. N Srinivasu, MasoodSahil, Jeevan Francis, Sure Pravallika, "Security enhanced using honey encryption for private data sharing in cloud", International Journal of Engineering & Technology,2017, Vol:7, Issue:1.1, 675-678.
22. Suresh B Rathod and Vuyyuru Krishna Reddy ,"NDynamic Framework for Secure VM Migration over Cloud Computing", Journal of Information Processing Systems, January 2017, 13(3):476-490
23. S. Phani Praveen, K. ThirupathiRao, B. Janakiramaiah ,"Effective Allocation of Resources and Task Scheduling in Cloud Environment using Social Group Optimization", Arabian Journal for Science and Engineering ,Issue 8/2018
24. R.G. Babukarthik1, J. Satheesh Kumar, J. Amudhavel, " SECURE DATA STORAGE AND SHARING IN CLOUD: VM SCHEDULING", 2017 , IIOABJ , Vol. 8 Issue 2, 186-190 .
25. Tarasvi Lakum and B.ThirumalaRao," A Key-Ordered Decisional Learning Parity with Noise (DLPN) Scheme for Public Key Encryption Scheme in Cloud Computing", International Journal of Advanced Computer Science and Applications(IJACSA), Volume 10 Issue 11, 2019.
26. Chaitanya,G.K.,Amarendra,K.,Aslam,S.,Soundharya,U .L.&Saikushwanth,V. 2019, " Prevention of data theft attacks in infrastructure as a service cloud through trusted computing", International Journal of Innovative Technology and Exploring Engineering, Vol.8, no.6 Special Issue 4, pp.1278-1283.
27. Garigipati,N. &Krishna,R.V. 2019, "A Study on data security and query privacy in cloud", Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019,pp.337.
28. Keerthi,G. &Kiran,P.S. 2019," Source based privacy for confidential data in cloud based e-healthcare systems", International Journal of Innovative Technology and Exploring Engineering, vol.8, no.7, pp.1310-1313
29. Mane,P.M, & Rani, C.M.S. 2019,"Achieving confidentiality and effective access control of cloud data using copher text policy based attribute based encryption", Journal of Computational and TheoriticalNanoscience, vol.16, no.12, pp.5063-5066
30. Naresh, Jaya Lakshmi,A. &Reddy,V.K. 2019," Resource optimization using cloud scheduling", International Journal of Innovative Technology and Exploring Engineering, vol.8, no.6, S2, pp.184-189.