PRIVACY PRESERVATION OF CLOUD DATA USING K-MEAN CLUSTERING

Amanpreet Kaur[1], Dr. Jyotsna Sengupta[2]
[1]Student of Master of Philosophy, [2] Professor
[1] Department of Computer Science
[1]Punjabi University Patiala, Punjab, India

Abstract: Data privacy is major concern in cloud computing. Most of the methods preserve the sensitive data like address, phone number, etc. and fewer methods to preserve quasi identifiers. In anonymization and randomization there is information loss during reverse process of preservation. I have worked to check the information loss problem by using K-mean clustering. Database of employee is considered. Three quasi identifiers are preserved during experiment. In gender attribute preservation, age attribute is taken because gender is Boolean. The whole age record is divided into two clusters and the first and the second cluster centroid are assigned to male and female respectively. In case of age preservation, 3 clusters are formed. Initially the difference between maximum and minimum age is calculated and if that difference is small then consider it as a noise, otherwise calculate the remainder of maximum and the minimum centroid and if its value is small, then take it as a noise and merge it in the age data. If both conditions are not applicable then take the remainder of maximum cluster with 10 (constant), and that becomes the additive noise. To preserve the area code, find the difference between maximum and minimum centroid and then multiply it to the original zip code data. So, the privacy of attributes is preserved. During reverse process the age noise is subtracted from preserved age data and the altered area code is divided with noise and original data is received. So the information loss is null in the proposed method.

Index Terms **– Cloud Computing, K-mean, Quasi identifiers and Privacy Preservation.**

1. INTRODUCTION

A cloud implicates a distinct Information technology environment that is usually constructed for the ambition of remotely provisioning scalable and measured IT devices. This term is originated as an emblem for the Internet that is a network of networks presenting group of decentralized IT resources. The logo of a cloud is commonly used to represent the Internet in a collection of blueprints and common documentation of Web-based architectures. The main idea of cloud computing is mainly based on a fundamental principal of reusability of IT capabilities. Cloud computing is on-demand model which have shared pool of computing resources, servers, storage, applications and services. It is a computing paradigm, in which we have large pool of systems or devices which are inter-connected in private or public networks, to provide dynamically expandable or adaptable infrastructure for application, data and storage of file. The infrastructure of cloud is maintained by the cloud provider rather than individual cloud customer.

Top benefits of Cloud Computing

Why is the cloud computing so vast or popular? Here are some common reasons that show why organizations are moving to cloud computing services:

- IT Cost reduction

Perhaps, the most significant benefit of cloud computing is IT cost savings. Cloud computing may decrease the cost of maintaining and managing our IT systems. Instead of purchasing costly systems and equipment for business, you can decrease the costs by adopting the resources of cloud computing service provider.

- Flexibility

Cloud computing allows users or employees or customers to be more flexible or malleable in their work practices. If you need connect to your data when you are off-site, you can access to your virtual office, easily and quickly. Organizations can scale up or raise as computing needs expands and scale down again when demands decrease.

- Reliability

Cloud computing takes data backup, emergency recovery and business continuity or stability simpler and less expensive, because data and information can be mirrored (copied) at several redundant (duplicate) sites on the cloud provider's network.

- Security

Cloud computing servers provide much more security than local servers. You never worry about loss of critical or important data and business applications because of any kind of natural disaster or full-on computer crash as your data or information is kept in the cloud so you can access it without worrying about what happens to your machine.

- Backup and Recovery

As all your whole data is stored or kept in the cloud so backing it up and restoring the data is relatively much easier than storing on a physical device and retrieving from same.

- Easy Information Access

After registering yourself in the cloud, you can easily access the data or information from any location, where there is a connection of Internet. This beneficial feature lets you go beyond the time zone and the geographic location issues.

- Document control

Document control is must to keep a secure business. You never realize what can happen if a document gets into the wrong or fraud hands, even if it is in the hands of an untrained employee.

- Unlimited Storage for use

Storing data or information in the cloud allows you almost unlimited storage capacity. Hence, you should not worry about running out of storage space or expanding externally your current storage space availability.

- The cloud remains always on

Many times you might make the inappropriate mistake of forgetting some important file at office or work place. The cloud remains always on, so that if you have an Internet connection at that time then you can get the applications or data you need from literally anywhere.

2. CLOUD COMPUTING SERVICE MODELS

Cloud Providers offer services that can be grouped or divided into three main categories.

1. **Software as a Service (SaaS):** It is a traditional model of distribution of software, in which software is bought for and installed on personal desktop. Sometimes it is also known as Software-as-a-Product (SaaP). In SaaS model, a whole application is offered to the user, as a service on demand. SaaS is a model in which applications are hosted by a service provider and made accessible to customers or users over a network (Internet). A single instance of the service provided by service provider runs on the cloud & multiple or various end users are serviced. On the user side, there is no need for software licenses, while for the service provider, the costs are lowered, as only the single application needs to be maintained as well as hosted. Today SaaS is provided by companies like Google, Microsoft, Salesforce, Zoho, etc. Examples of SaaS are instant messaging from AOL, Google's Gmail and Apps, Yahoo and Google.
2. **Platform as a Service (PaaS):** Cloud computing has expanded to introduce platforms for creating or making and running or executing custom web-based applications is called PaaS i.e. Platform-as-a-Service. PaaS is an enlargement of the SaaS model. In this type of service, development environment or a layer of software is enclosed and provided as a service on which another higher level of service can be created. The user has the independence to make his own applications, which executes on the infrastructure of provider. To meet requirements of scalability and manageability of the applications, PaaS providers provide a predetermined combination of application servers and OS. The PaaS model creates all of the facilities needed to support the whole life cycle of making and distributing services and web applications, with no downloads or installation of software for IT managers, developers or end users, entirely available from Internet. Examples of PaaS include Google's App Engine, Microsoft's Azure, Force.com and Salesforce's Force.com
3. **Infrastructure as a Service (Iaas):** The core resources used for computing are software and hardware components. IaaS offers basic storage space and computing capabilities services over the network to end users to run arbitrary software. Servers, log access, monitoring, storage, networking equipment, security, processing, load balancing, data centre space and other computing resources are merged and made available to customers to handle heavy workloads. In IaaS, end users access services as well as resources through WAN. Customers can buy any of needed services and they have to pay only for his resources usage. The main idea behind IaaS is resource virtualization. It permits the customer to have his own guest OS on upper part of infrastructure given by the cloud service provider or in words, we can say that the end user would typically form his own software on the top of infrastructure. The administration, formation, and maintenance are the responsibilities of the cloud service provider. Some of the common examples of IaaS are Amazon, 3 Tera, GoGrid etc.

3. DEPLOYMENT MODELS OF CLOUD COMPUTING:

A cloud deployment models are used to represent a particular type of environment of the cloud which is distinguished primarily by size, ownership, and access. There are mainly four cloud deployment models:

- *Public Clouds*

A public cloud is a publicly connectable cloud environment controlled by the third-party cloud provider. The cloud provider is accountable for the formation or production and maintenance of the IT resources and public cloud. All users can publically access the services of the public cloud like using Google search engine to get information about unknown fact, using Amazon application to buy anything according to your need or using other social applications like Facebook, Gmail, Yahoo etc to connect to the outer world.

- *Community Clouds*

It is also same as public cloud except that there is limited access to a specific or particular community of cloud users or customers of that organization which has shared policy, security requirements and mission. It may be collectively owned or controlled or managed one or more companies in the community i.e. by the third-party cloud provider and the members of the community that provides a public cloud but limited access.

- *Private Clouds*

A private cloud is controlled by a single organization. Private clouds exclusively designed for single organization to use technologies of cloud computing by providing the centralized access to different IT resources from different locations, parts, or departments of that organization. The usage of a private cloud can altered when boundaries of organization and trust are applied and defined.

- *Hybrid Clouds*

A hybrid cloud is a type of cloud environment which can be composed of two or more cloud models which can be different deployment models. For example, a cloud user choose sensitive data services of private cloud model and other less sensitive data services of a public cloud. The resultant combination of different models is a hybrid deployment model.

4. PRIVACY PRESERVATION METHODS

The privacy of cloud data has to be preserved or maintained anywhere and anytime. There are various methods that can put forward to check the of privacy preserving. So, the work of preserving the privacy takes us in both tracks: privacy preservation of the data or information and preservation of the privacy when you choose some kind of third party auditing to make sure the correctness of data. Now some methods of privacy preservation are given as under:

- **Anonymity-based system:** It is also called de-identification of data. To preserve and achieve the privacy in the cloud Jiang Wang et al. developed the system which is Anonymity system which use anonymity algorithm. It eliminates the identifying variables like address, phone number, Aadhar card number etc. Different developers develop different methods of anonymity by varying in there ease of use, cost, complexity, robustness. This algorithm controls and processes the data as well as anonymises all or some or few part before releasing that information in the cloud. Cloud service provider uses its background knowledge to make the association between unspecified and specific knowledge and extract needed knowledge or data from anonymous data. In case of privacy preservation this technique or approach is different from traditional or classic cryptography technique as it is simple as well as flexible enough gets rid of key management. Because of its simplicity, anonymous change in accordance with its attributes and also depends upon service provider. This technique is applicable for limited number of cloud services only. This method preferred only if depends upon automating the anonymisation or automisation. The records or data in the database can be divided into following categories:

i. Explicit identifiers- These are the attributes or variables which helps in recognizing a particular person uniquely. These are harmful to expose publically as the personal information can be leaked if these identifiers are not protected. For example: Name, address, social security number, Aadhar No etc.

ii. Quasi identifiers- These are the type of variables which are not harmful; but whenever these attributed are combined with other attributes while identifying the particular individual person from the group of people. For example: age, gender, date of birth, place of birth, area zip code, city etc.

iii. Sensitive identifiers- The attributes which contains sensitive value of owner of data is called sensitive identifiers. This type of data is usually released publically and can be required by researchers.

- **Architecture for Privacy preserving:** Privacy preservation architecture helps to preserve the privacy of data or information of user. Single user or customer cannot access the whole data which is stored or kept in cloud. This method avoids or restricts the risk of various kinds of internal and external attacks to user data. The architecture of cloud storage grants privacy without any need of trust the external cloud administrators and the corporate administrators. This architecture contains user interface, data management unit, user engine, encryption proxy, cloud database and rule engine. With the help of user interface, end user can send request to access the database in cloud and administrators obtain the request which was sent to user engine and then to rule engine and at last to cloud database in the form of RPC/XML request. By using encryption procedure and allotting secure identities to every request as well as response at every stage, the privacy is preserved and also maintains the system readable usage. Encryption scheme is simple to carry out but there is difficulty in giving machine readable or discernable access rights.
- **Access control for privacy preserving:** Miao Zhou et al has discovered an approach that is flexible enough and consider the privacy through access control or rights granted to each end user in the environment of cloud. Certain types of attributes or parameters are associated with each user of cloud that determines the access control or rights of user. He has used two tier architecture encryption model for data privacy in which there are two phases i.e. surface phase and base phase. In base phase, the owner of the data does local encryption, which is attribute-based, on outsourcing data. On the other hand, in surface phase server of the cloud performs operations on outsourced data by SRM (Server re-encryption Mechanism). The data encrypted by owner is re-encrypted by server dynamically by using SRM when owner of data requests. Either the existing user of cloud is nullify or new one is created request for SRM is provided. The access policies or rights remains hidden to the server. In this way, the privacy of user data is preserved by providing full access control to the owner.
- **Authorization system for privacy preserving task:** David W. Chadwick et al. recommended a policy for preserving the privacy of data and based upon authorization infrastructure for service cloud. For data protection, it allows a user to set his own access or privacy policies and cling it so that unauthorized person should not access your data. By this it will be sure that the data of user is secure and protect from misuse. And also show controlled access of user over data. Policy Enforcement Points (PEPs) and Policy Decision Points (PDPs) are used to make decisions about the authorization and later on applying these decisions on data. Master PDP is discovered (launched) to figures out as well as resolves the issues or problems among various decisions of PDPs. Obligation service is granted as a component of authorization infrastructure, with the help of which the owner of data can decide which data access is authorized and which is unauthorized. In this approach encryption of outsourced data is not necessary as the cloud service provided is trusted. The overall focus is needed on security threats which can be from cloud providers and dividing the whole infrastructure into different distinct services and all those services are running on specific virtual machine. This will enhance the performance of software.
- **A privacy preserving Data Outsourcing:** It is another approach to preserving privacy as well as confidentially of data of owner by using graph privacy constraint method. The nodes of graph represent attributes whereas the links between attributes represent confidentially. Among the complete group of attributes, sensitive attributes are subpart or subset. The information about attributes must not share with external party. The relationship among these attributes is drawn and then fragmented vertically. For performing fragmentation, graph coloring method can be used. One fragment is handed over the owner whereas the remaining fragments are placed at appropriate location at external server. The relation among fragments can be reconstructed with the help of common id. While performing fragmentation, it is compulsory to investigate the total workload is kept smaller or minimized at source and also confidentially constraints are not violated by the fragment of server. The process of fragmentation can be achieved on the basis of certain types of metrics like Min-Query, Min-Attr, and Min-Cond. These matrices make sure that the outsourced data is secured from attacks of third party. Thus the fragmentation can handle the privacy efficiently and effectively.
- **PccP Model in cloud:** it is another technique for privacy preservation in cloud. Three-layer architecture is used for this model i.e. consumer layer, network interface layer and privacy preserved layer. At the lower level, consumer layer is present and all requests of cloud access from different users can be submitted through this layer. The next layer in this model is network layer which is sometimes called address mapping layer. This layer translates the requests of users usually by changing the original IP address for the privacy preservation. And it also provides mapping among manipulated and original or real IP address. At the top of model, privacy preserved layer is presented and it mainly used to generate unique or distinct user cloud identity. The sensitive data can be preserved with the help of this layer by using Privacy Check Mechanism. By this mechanism, we can determine what amount data can be made transparent in the cloud and also determine access control.
- **Dynamic Metadata Reconstruction:** Metadata exploitation possibility is determined by Adeela Waqar et al. If the attacker gains information or knowledge about metadata from somewhere, then there could be risk of data privacy. To overcome this problem of privacy preservation, a framework has been proposed. Initial step of this method is to separate the metadata in the cloud and then merge the attributed of segregated data into a group and place it in private form which can be either exclusively (fully) private or partially privates or sometimes non-private depending upon data sensitivity. The next phase after data classification is table splitting. In this phase, the table of database is splits horizontally as well as vertically. This dividing part ensures the normalization of database. Ephemeral referential consonance is the next phase which comes after reconstruction of metadata. It makes sure that the data is secured in the cloud before and after dividing the database table. This whole process shows the efficiency of using this method.
- **Public Auditing for Secured Data Storage:** C. Wang et al. developed a method of privacy preservation to carry out or perform public auditing on cloud data. The traditional cryptographic measures in cloud computing are not sufficient to secure data or to achieve security. The ubiquitous nature of the data and data outsourcing is the main reasons. So they use the concept of Third Party Auditing (TPA). Random masking and Homomorphic authenticator make sure that TPA never gains any information during auditing process. So TPA is trustable as well as capable to connect to the cloud storage for auditing process. If any risk is there then the report of auditing brings it out. This system uses two phases and four algorithms. The first phase is making up by SigGen and KeyGen algorithms called Setup. In this secret parameters initialization and generation of verification of metadata takes place. After this, in second phase, the auditing process takes place in Audit phase

and also assures the data correctness in the cloud. VerifyProof and GenProof algorithms are used in second phase. This approach ensures the security and privacy for batch auditing and also data correctness. The strength of security has been improved by C. Wang et al. with respect to their previous proposals. Zero-knowledge leakage by public auditing has also been achieved with the help of newly designed protocols.

- **Oruta:** Oruta concept is the another public auditing mechanism proposed by Boyang Wang et al. after analyzing the work of Wang et al. It preserves the privacy of data. This method makes sure the capability of un-forge ability and correctness during public auditing. There are three sections in this approach. Those are cloud users, TPA and cloud server. Cloud users are further classified into two categories i.e. group users and original user. The owner of data is original user who is able to control the outsourced data and also the flow of that data i.e. transactions in the cloud. All requests from different users are send to TPA to carry out public auditing to verify or check the rightness or correctness of data. HARS scheme contains three algorithms i.e. RingSign, KeyGen and RingVerify. These algorithms are invented to carry out or obtain the privacy preserving auditing. To concentrate on efficient auditing method, the approach is authorized to ensure the shared data integrity in dynamically merged users' environment.

5. PROPOSED METHODOLOGY

The proposed method works on the basis of K-mean clustering approach. The step by step execution of our proposed approach is given as under:

i. Initially eliminate the identifying field.
ii. Divide the age attribute of whole database into two clusters.
iii. Compute the centroid of two clusters.
iv. Assign the values to gender field :
Male ← First centroid value
Female ← Second centroid value
v. Calculate the appropriate value of noise for age by K-mean clustering.
vi. Distort the age attribute:
Noisy Age = Original Age + noise
vii. Compute the noise value for zip-code attribute.
viii. Perturb the zip code:
Noisy Zip-code = Original Zip-code * noise

6. EXPERIMENTAL RESULTS

To conduct the experiment, I have proposed a method which uses K-mean clustering approach. For preserving the gender attributes, the age attribute is used and two clusters are formed and are assigned to male and female attribute. The age attribute is preserved by adding the noise to the original age value and the area-code is preserved by multiplying the noise value to the original zip-code.

Gender	Age	Zip-code
Male	39	77516
Male	50	83311
Male	38	215646
Male	53	234721
Female	28	338409
Female	37	284582
Female	49	160187
Male	52	209642
Female	31	45781
Male	42	159449

Table 6.1 Original Database

Gender	Age	Zip-code
35	40	11732124116
35	51	12609203161
35	39	32638237746
35	54	35525258071
51	29	51218540559
51	38	43071770282
51	50	24244462637
35	53	31729526342
51	32	6929000131
35	43	24132765599

Table 6.2 Privacy preserved by K-mean

In the table 6.1, quasi identifiers of the original database are shown and in table 6.2, the database shown is preserved database after using K-mean clustering. In case of gender attribute preservation, the value of first centroid is 35 and it is assigned to Male attribute. The value of second centroid is 51 and is assigned to Female attribute. The noise value computed for age attribute and gender attribute is 1 and 151351 respectively. The noise value is added to age attribute and multiplied to zip-code attribute.

6.1 Information Loss:

With the help of information loss, we can calculate the amount of information lost in overall process. As we encrypt the attributes either with help of adding or multiplying the noise value to the original value (in case of age attribute and gender attribute) or replacing the value by other value like in case of gender Boolean attribute. After reverse processing, if we get the same out as original, then there will be no information loss otherwise we can calculate the value of information loss with the help of mathematical formula.

IL = (nv-ov) / ov

Where nv → new value of variable
ov → old value of variable

The formula can also express as:

IL = change in value / old value

6.2 Execution time of Proposed Method for different databases:

The execution time means the total time taken by the algorithm to show results. In our approach, we have calculated the execution time for different databases. That execution time is shown with the help of table 4.11. The execution time of different attributes of a database is shown in the table. Those attributes are the quasi identifiers on which we have applied K-mean clustering approach i.e. gender attribute, age attribute and the zip code attribute. All these attributes have different execution time for different database. The execution time need not be same during number of iterations of same database.

	Execution Time of Different Attributes (sec. approx.)

No. of records in database			
	Gender	**Age**	**Zip-code**
50	0.015626192092895508	0.031252384185791016	0.03125143051147461
100	0.03125500679016113	0.03125333786010742	0.03125143051147461
150	0.03125143051147461	0.03125143051147461	0.03464365005493164
200	0.031250953674316406	0.04687643051147461	0.03125143051147461
250	0.03125190734863281	0.04687643051147461	0.03125166893005371
300	0.031250953674316406	0.04625166893005371	0.04687690734863281
350	0.0312504768371582	0.04687786102294922	0.04687786102294922
400	0.03125143051147461	0.046877384185791016	0.046877145767211914
500	0.03125333786010742	0.04687833786010742	0.04687762260437012
1000	0.046877384185791016	0.046877384185791016	0.04687643051147461
2000	0.06250500679016113	0.046877384185791016	0.04687690734863281
3000	0.06250357627868652	0.06250262260437012	0.07812881469726562

Table 6.3 Execution time of different attributes by proposed method

6.3 Graphical Representation of Execution Time:

Here we have shown the line representation of execution time of different attributes of different databases. Fig 6.1 shows line representation of same execution time.

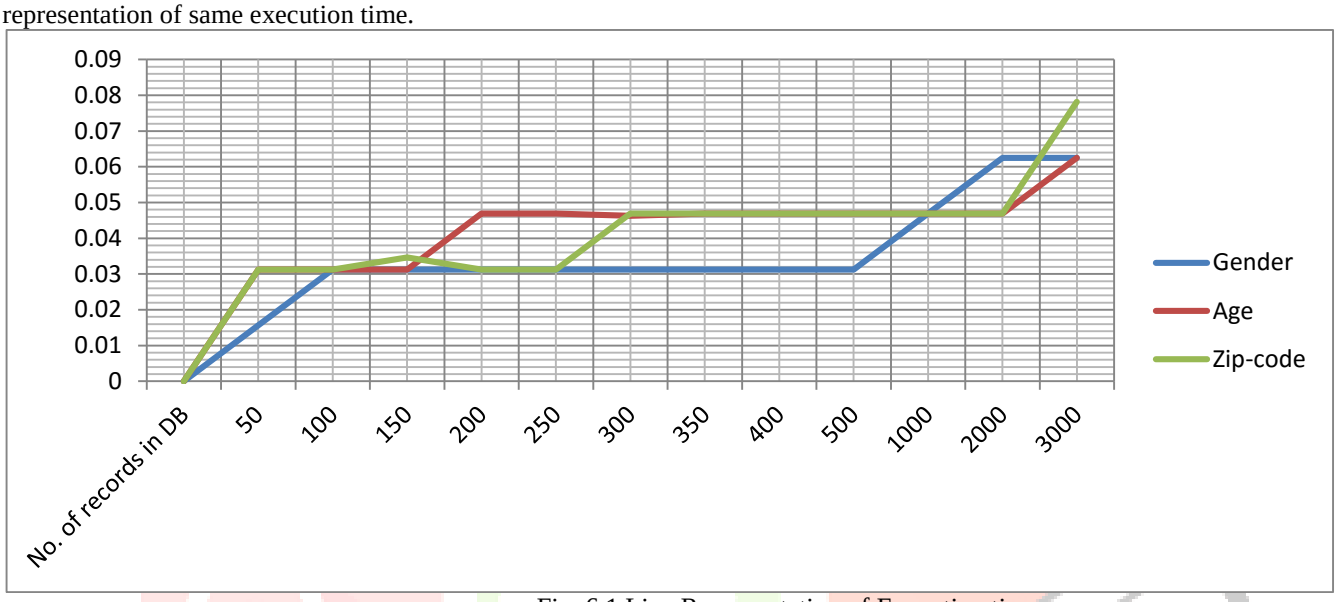

Fig. 6.1 Line Representation of Execution time

6.4 Noise value in Proposed Method for different databases:

The noise values is a value merged to different attributes of given database. The value of noise is different for different attributes. I have taken various databases of different length (number of entities are different in different database) and the noise value of various attributes corresponding to those attributes has been shown in table 6.3.

No. of records in database	Value of Noise			
	Male	Female	Age	Zip-code
50	27	47	2	181370
100	29	49	7	168770
150	50	28	2	173341
200	49	28	2	188725
250	30	52	5	189123
300	50	28	5	187486
350	52	29	5	191440
400	50	28	5	192750
500	50	28	5	194835
1000	28	50	5	184751
2000	51	29	5	182243
3000	51	29	5	184669

Table 4.12 Noise values for various attributes of different Databases

6.5 Graphical Representation of Noise Value:

Graphical representation of data is the pictorial representation of tabular data. In the graphical representation of noise value, various values of noises of different attributes are shown in fig 6.2.

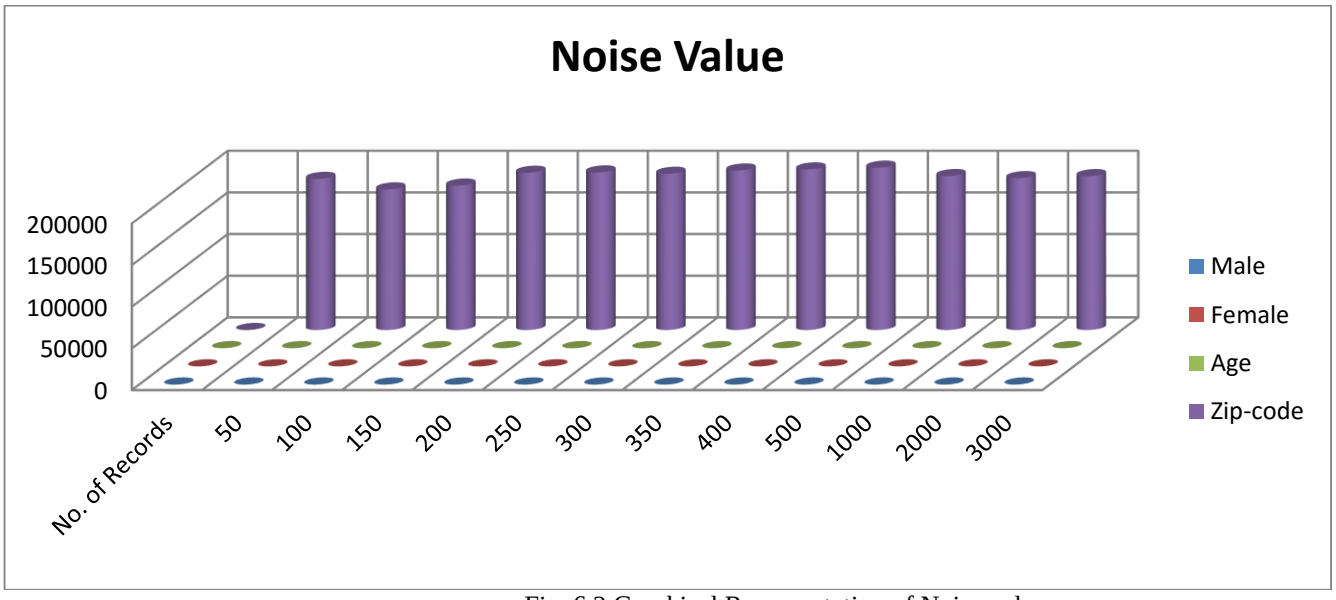

Fig. 6.2 Graphical Representation of Noise value

6.6 Comparison with Existing Techniques:

Information loss: It is the main advantage of our proposed technique. Loss of information in our method is null whereas in other cases there was loss of information. In case of anonymization, as the information is hidden with the help of special characters, there was about complete information loss in the worst case. In case of randomization, as the values are shuffled, so again there is maximum information loss. The comparison with other techniques in case of gender is shown in table 6.3

No. of Records in Database	Percentage of Information Loss (%)		
	Anonymization	Randomization	Proposed Method
50	72.52	99.23	00
100	72.12	100	00
150	71.80	82.54	00
200	71.73	76.54	00
250	71.62	77.65	00
300	71.45	85.55	00
350	71.52	90.23	00
400	71.46	91.23	00
500	71.44	99.2	00
1000	71.47	89.23	00
2000	71.56	73.2	00
3000	71.52	85.23	00

6.7 Graphical Representation of Information Loss:

The pictorial representation of information loss of different method is shown in fig 6.3.

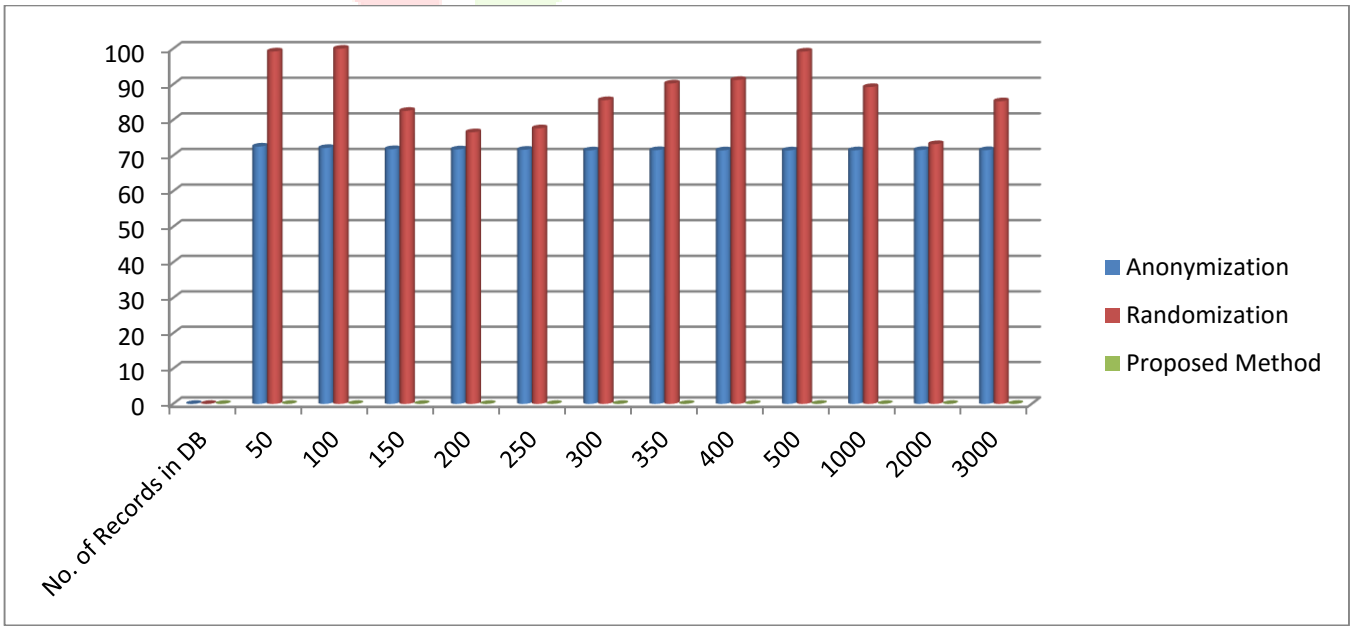

Fig. 6.3 Information loss Graph

7. CONCLUSION

Cloud computing is based on reusability capabilities of IT. Cloud computing privacy means a huge set of policies or rules, technologies, and controls set up to secure data or information, applications, and the related infrastructure of cloud computing. In the proposed method, I have used K-mean clustering for preserving the privacy of QIs. Mainly three attributes are taken for privacy preservation. Those attributes are Gender, Age and Zip-code. In case of gender, as it has Boolean value, the privacy is not easy to preserve. So the age attributes is taken for privacy of gender because we have to suppress the original value of gender with database related data. The age attribute is divided into two clusters with the help of k-mean clustering and the centroid of first cluster is assigned to Male and the second centroid value is assigned to Female and by this the gender is preserved. In the age preservation, number of clusters is defined initially. The remainder of maximum age and minimum age is calculated and if it is less than 10, then it simply added to real age value otherwise calculate the remainder of maximum centroid value and minimum centroid value and again if it less than 10 than add it to the age data otherwise calculate the remainder of maximum cluster centroid value with 10 and add it. In case of zip code preservation, again decide the number of clusters and then subtract the minimum centroid from the maximum cluster centroid value and multiply it to the original area code values. The information loss in the proposed method is zero and all the characters are preserved. The total execution time and the noise value for age and zip code of different data size are also mentioned in the experimental results. The comparison of proposed method with other existing methods is also shown in paper.

8. REFERENCES

[1] Arshveer Kaur, "A Hybrid Approach Of Privacy Preserving Data Mining Using Suppression And Perturbation Techniques", IEEE, International Conference On Innovative Mechanisms For Industry Applications, 2017.

[2] Kingsford Kissi Mireku, Zhang Fengli, Kittur Philemon Kibiwott, "A Hybrid Privacy Preservation Framework for Healthcare Data Publishing", American Journal of Engineering Research (AJER), 15 July, 2017.

[3] Ying-Tsung Lee, Wei-Hsuan Hsiao, Yan-Shao Lin And Seng-Cho T. Chou, "Privacy-Preserving Data Analytics In Cloud-Based Smart Home With Community Hierarchy", IEEE Transactions On Consumer Electronics, Vol. 63, No. 2, May 2017.

[4] Keke Gai, Meikang Qiu, and Hui Zhao, "Privacy-Preserving Data Encryption Strategy for Big Data in Mobile Cloud Computing", IEEE Transactions on Big Data, 2017.

[5] Dr. Puneet Goswami, Suman Madan, "A Survey on Big Data & Privacy Preserving Publishing Techniques", Advances In Computational Sciences And Technology ISSN 0973-6107 Volume 10, Number 3, 2017.

[6] Abdul Razaque And Syed S. Rizvi, "Privacy Preserving Model: A New Scheme for Auditing Cloud Stakeholders", Journal of Cloud Computing: Advances, Systems and Applications, Springer, 2017.

[7] Zhaoemin ,Gengyang, And Jingqishi, "A Privacy-Preserving Parallel and Homomorphic Encryption Scheme", De Gruyter, 15:135–142, 2017.

[8] Nagajothi .S And Raj Kumar .N, "Data Anonymization Technique for Privacy Preservation Using MapReduce Framework", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 5, May 2016.

[9] Marjan Soltani, Gholamreza Shahmohammadi, "Proposed Model for Privacy Preserving in Mobile Cloud Computing", International Journal of Scientific Research in Computer Science and Engineering, Volume-3, Issue-6, ISSN: 2320-7639, 30 Dec 2015.

[10] N. Nishara and Reeta Pandey, "Enhancing Security in Public Clouds Using Data Anonymization Techniques", International Journal Of Computer Applications (0975 – 8887) Volume 128 – No.1, October 2015.

[11] Chhaya S Dule, Girijamma H.A, Rajasekharaiah K M, "Data Anonymization Technique for Privacy Preservation on MapReduce Framework", ISSN (Online): 2347-2820, Volume -3, Issue-1, 2015.

[12] Somchart Fugkeaw and Hiroyuki Sato, "Privacy-Preserving Access Control Model for Big Data Cloud", IEEE, 2015.

[13] Xuyun Zhang, Wanchun Dou, Jian Pei, Surya Nepal, Chi Yang, Chang Liu, And Jinjun Chen, "Proximity-Aware Local-Recoding Anonymization With MapReduce For Scalable Big Data Privacy Preservation in Cloud", IEEE Transactions On Computers, TC-2013-12-0869, 2015.

[14] Pooja HP, Nagarathna N, "Privacy Preserving Issues and Their Solutions In Cloud Computing: A Survey", International Journal of Computer Science and Information Technologies, Vol. 6 (2), 2015.

[15] S. Liu, Q. Qu, L. Chen and L. Ni, "SMC: A Practical Schema for Privacy-Preserved Data Sharing Over Distributed Data Streams", IEEE Transactions on Big Data, 1(2):68–81, 2015.

[16] Y. Li, W. Dai, Z. Ming, and M. Qiu, "Privacy Protection for Preventing Data Over-Collection in Smart City", IEEE Transactions on Computers, 65:1339–1350, 2015.

[17] R. Nallakumar,Dr. N. Sengottaiyan, M.Mohamedarif, "Cloud Computing and Methods for Privacy Preservation: A Survey", International Journal Of Advanced Research In Computer Engineering & Technology (IJARCET) Volume 3 Issue 11, November 2014.

[18] Imran Ashraf, "An Overview of Service Models of Cloud Computing", International Journal of Multidisciplinary and Current Research, 15 August 2014.

[19] R.Rogini, N.Arun Balaji, "An Inspection on Privacy Preserving Methods in Cloud Computing", International Journal of Science, Engineering and Technology Research (IJSETR), Volume 3, Issue 5, May 2014.

[20] Xuyun Zhang, Laurence T. Yang, Chang Liu, And Jinjun Chen, "A Scalable Two-Phase Top- Down Specialization Approach for Data Anonymization Using Mapreduce on Cloud", IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 2, February 2014.

[21] Bincy Paul And M. Azath, "Survey on Preserving Data Privacy in Cloud", International Journal Of Computer Science, 2014.

[22] H. Liu, H. Ning, Y. Zhang, Q. Xiong, and L. Yang, "Role-Dependent Privacy Preservation For Secure V2G Networks in The Smart Grid", IEEE Transactions on Information Forensics And Security, 9(2):208–220, 2014.

[23] Ch Chakradhara Rao, Mogasala Leelarani, Y Ramesh Kumar, "Cloud: Computing Services and Deployment Models", International Journal of Engineering and Computer Science ISSN: 2319-7242 Volume 2 Issue 12, Dec.2013.

[24] Apeksha Sakhare, Swati Ganar, "Anonymization: A Method to Protect Sensitive Data in Cloud", International Journal of Scientific & Engineering Research, Volume 4, Issue 5, May-2013.

[25] Cong Wang, Sherman S.M. Chow, Qian Wang, Kui Ren, and Wenjing Lou, "Privacy-Preserving Public Auditing for Secure Cloud Storage" IEEE Transactions on Computers, Vol. 62, No. 2, February 2013.

[26] Xin Dongt, Jiadi Yut, Yuan Luot, Yingying Chen, Guangtao Xuet, and Minglu Lit, "P2E: Privacy-preserving and Effective Cloud Data Sharing Service", Globecom 2013-Communication and Information System Security Symposium, IEEE, 2013.

[27] Ulrich Greveler, Benjamin Justus, Dennis Loehr, "A Privacy Preserving System for Cloud Computing", 11th IEEE International Conference on Computer and Information Technology, 2011.

[28] L. Zhou, V. Varadharajan, and M. Hitchens, "Enforcing Role-Based Access Control for Secure Data Storage in the Cloud", the Computer Journal, Vol. 54 No.10, 2011.

[29] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search Over Encrypted Cloud Data," Proc. IEEE INFOCOM, Pp. 829-837, 2011.

[30] N. Mohammed, B. Fung, P.C.K. Hung, and C.K. Lee, "Centralized and Distributed Anonymization for High-Dimensional Healthcare Data," ACM Trans. Knowledge Discovery from Data, Vol. 4, No. 4, Article 18, 2010.

[31] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, And M. Zaharia, "A View of Cloud Computing," Comm. ACM, Vol. 53, No. 4, Pp. 50-58, 2010.

[32] Hui Wang, "Privacy-Preserving Data Sharing in Cloud Computing", Journal of Computer Science and Tech- Nology 25(3): 401–414, May 2010

[33] Maneesh Upmanyu, Anoop M. Namboodiri, Kannan Srinathan, and C.V. Jawahar, "Efficient Privacy Preserving K-Means Clustering", Springer-Verlag Berlin Heidelberg, LNCS 6122, pp. 154–166, 2010.

[34] B.C.M. Fung, K. Wang, and P.S. Yu, "Anonymizing Classification Data for Privacy Preservation," IEEE Trans. Knowledge and Data Eng., Vol. 19, No. 5, Pp. 711-725, May 2007.

[35] X. Xiao and Y. Tao, "Anatomy: Simple and Effective Privacy Preservation," Proc. 32nd Int'l Conf. Very Large Data Bases (VLDB '06), Pp. 139-150, 2006.