Techniques for Data Privacy Preserving in Cloud Computing Environments

Shruthi M[1], Demian Antony D'Mello[2]
[1]Department of Computer Science and Engineering
N.M.A.M. Institute of Technology
Visvesvaraya Technological University
Belagavi, Nitte, India - 574110
shruthim.gatty@nitte.edu.in
[2]Department of Computer Science and Engineering
Canara Engineering College, Visvesvaraya Technological University
Belagavi, Mangaluru, India - 574219
demian@canaraengineering.in

ABSTRACT: *Cloud Computing renders useful services over the Internet by increasing resource utilization and computational services at a decreased cost along different sectors. Data owners outsource their private data on cloud environment leading to loss of physical control over data. Security solutions provided in the literature to safeguard the data mainly focus on authentication and authorization. Data privacy preservation is one of the key areas in providing data security to ensure privacy guarantees for sensitive data namely health records, financial data, personal identities by providing strong end-to-end privacy. In this paper, the authors classify challenges in data privacy preservation based on the nature of mechanisms and type of data stored in cloud environments. The paper presents a detailed analysis of the literature in terms of methodologies adopted, algorithms proposed and cloud security performance metrics.*

Keywords: Cloud Computing, Data Privacy Preserving, Security

Received: 19 June 2019, Revised 5 September 2019, Accepted 24 September 2019

DOI: 10.6025/isej/2019/6/2/37-48

1. Introduction

Cloud computing is information technology architecture used by enterprises and individuals with advantages that include on-demand self-services, ubiquitous network access and location independent resource pooling [1]. The resources provided over the cloud can be shared, is dynamically scalable, rapidly provisioned, virtualized and released with minimum service provider interaction. Personal data of several millions of users can be managed in cloud as it provides cheap data storage, good flexibility with reduced power consumptions [2]. Data storage and computation can be performed simultaneously in a coordinated way on networked computers using cloud computing infrastructure. Users pay for the service as an operating expense without incurring

any significant initial capital expenditure, with the cloud services employing a metering system that divides the computing resource into appropriate blocks [2].

Generally several issues surround cloud such as interoperability, data portability, universal standards, SLAs (Service level agreements), security and privacy of sensitive information. Recent security breaches have led to compromise of data belonging to millions of people who have mined the trust of potential users in the cloud [3].An example of the security incidents include first incident being the external intrusion attack leading to stealing of seventy seven million personal accounts in the Sony PlayStation Network outage, second incident in the multi-day outage of Dropbox due to miss-configuration problem led to visitors logging in to twenty five million customer accounts, and finally private photo leakage of celebrities from the Apple iCloud storage service due to weakly protected login credentials [4]. Backup information, business details, user profiles are ubiquitously available through cloud storage and are accessible to users over the network. Issues pertaining to cloud data storage can be listed as data archiving, disaster recovery, online data backup, data compliances and compliance regulations.

Unauthorized access of data, steeling of digital data attributes to data owners losing control over cloud, like to the "honest but curious" CSPs causing leakage of personal data privacy. The privacy risk in business era is when authoritarians misuse secret information. Privacy preservation must be of prime concern in the network user data due to long distance and insecure transmission channels and data maybe precious belonging to health, finance or personal type. Data Privacy preservation is a challenging aspect where data mining techniques are used over data. Technical approaches used in preserving privacy of datasets mainly include encryption and data anonymization.

Rest of the paper is organized as follows. Section 2 describes all the related works on Data Privacy Preservation for Heterogeneous Data. Section 3 focuses on proposed technique in Data Privacy Preservation along with mechanisms and algorithms. Section 4 presents the analysis of mechanisms proposed in the literature against welldefined performance criteria. Section 5 draws the conclusions.

2. Data Privacy Preservation for Heterogeneous Data

We classify the data privacy preservation based on nature of data as: network user data, genetic data, biometric data and training data set.

Figure 1 presents the challenges along with mechanisms and algorithms proposed to handle cloud data privacy. The rectangles present in the diagram represent the nature of data and relevant challenges. Ovals denote the proposed mechanisms to handle the identified challenges.

2.1. Privacy Preservation in Network User Data

The exponential data growth over the network with the use of Internet and mobile devices are confronted with serious privacy concerns as valuable information are extracted and analyzed by various data mining methods. To address the privacy leakage issue effectively in clustering analysis of network user data, authors propose Differential Privacy preservation Multiple Cores DBSCAN (DP-MCDBSCAN) clustering schema [5]. It is a data distortion technique used in statistical databases of larger scale based on differential privacy. It is an effective clustering schema used in solving randomness and blindness problem encountered in DP-DBSCAN. The MCDBSCAN proposes the solution by selecting multiple core points then select the desired core point from result of clustering thereby using an optimization technique. During data publishing, the proposed scheme adds Laplace noise and the amount of noise here is independent of dataset scale. Experiments show small amount of noise required in large datasets for privacy preservation and solution is proven to be time efficient.

Data processing of wireless sensor network in cloud gives rise to application related to e-health systems. Monitoring health data, diagnosis record, medical histories and prescriptions are the components of e-health systems. Privacy concerns of ehealth systems and efficient data processing in these systems are the key concerns to be addressed. Dual privacy preservation scheme using certificate less fully homomorphic encryption is proposed by the authors [6]. It involves generation of keys, encryption, decryption and homomorphism. Performance evaluation of the scheme proves the solution proposed to be efficient in terms of communication costs, helps in resisting anti-collusion attacks and sniffing attacks.

2.2 Privacy Preservation in genetic data

Healthcare is taken to a new level with the usage of cloud computing paradigm. DNA data of patient is a form of genetic data

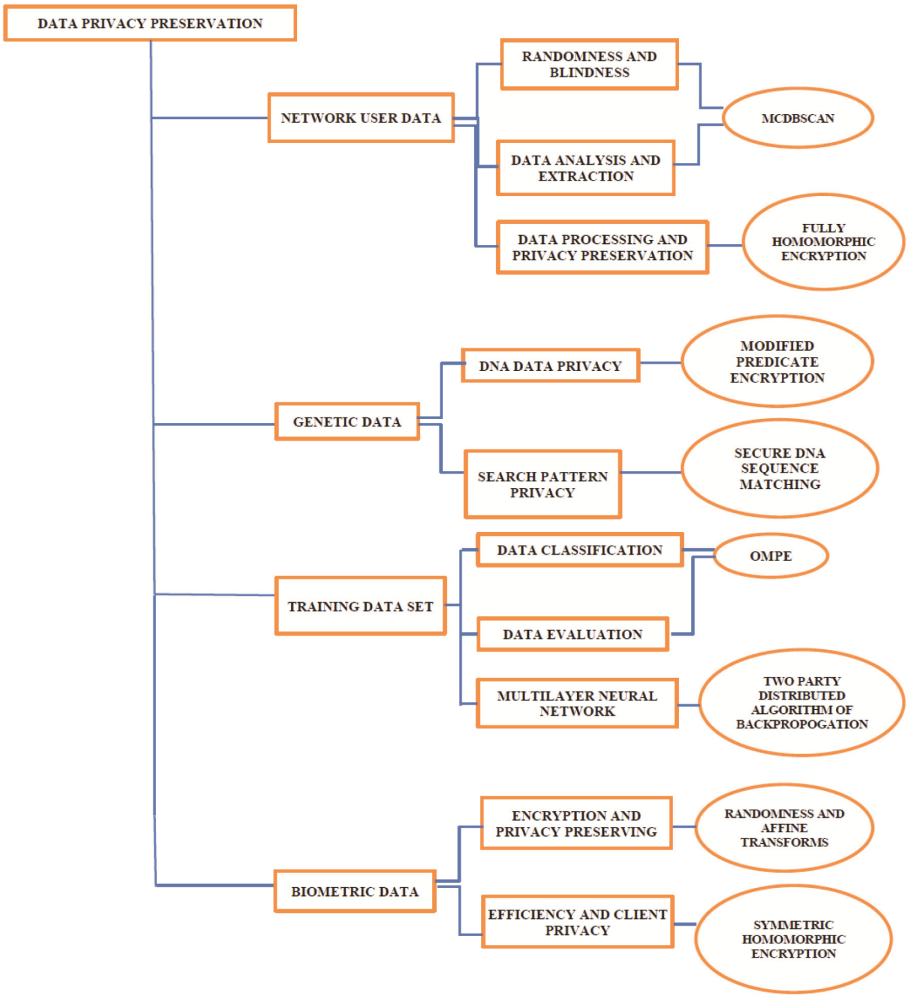

Figure 1. A Data Privacy Preservation based on Heterogeneous Data

recently used by doctors for early diagnosis of health problems. To diagnose the health problems early pattern matching will be done on the DNA data of the patient. Cloud platform is used for storing large volume of DNA data and to perform laborious computations needed in DNA sequence matching. Testing of genetic information on cloud leads to data exposure to unauthorized parties such as Cloud Service Providers (CSPs) and hence poses problems on privacy of genetic data. Secure DNA sequence matching solution is proposed by authors that encrypts the genetic data before outsourcing to the cloud [7]. Doctors using this system will have to perform testing on encrypted data only. Privacy leakage of DNA data and search pattern is solved using modified predicate encryption. The proposed solution makes use of error tolerant character-by-character comparison using predicate encryption method. Proposed solution guarantees Data confidentiality, Trace indistinguishability, Efficiency and Usability. Performance analysis of the proposed solution ensures security, small communication cost and reduced computation overhead compared to other secure pattern matching schemes.

2.3 Privacy issues in Training Data set

Classification of training dataset in data mining tasks is done by Support Vector Machine (SVM) classifiers. Classifier hence designed when released for public use violates privacy restricting the applicability of SVM. To solve privacy issued authors propose Privacy Preserving SVM classifiers (PPSVM) that are scalable to large datasets resulting in large number of support vectors [8]. In PPSVM, SVM classifiers are post processed so that they do not disclose privacy content of support vectors. Original decision functions in PPSVM are transformed to infinite series of linear combination of monomial feature mapped support vectors. Since attributes are hidden they provide better data protection. The method guarantees robustness, scalability and accuracy in classification.

Data classification using in machine learning uncovers hidden data patterns to the learners revealing the class to which the newly arriving data belongs to by testing similarity of datasets [9]. This revelation of information, results in leakage of entities asset during classification and similarity testing in cloud based distributed systems. During the classification and evaluation process care must be taken to maintain data confidentiality of original data of the data owner, secrecy of learned model and privacy of test data in predicting phase. To evaluate these scenarios authors have proposed Oblivious Evaluation of Multivariate Polynomial (OMPE) approach for preserving privacy of test data and learned model. It is a secure multi-party communication protocol used in computing multivariate polynomial function. Model similarity evaluation is used on learned model as a privacy preserving scheme to represent closeness between different model. OMPE is feasible and efficient in-terms of computational cost, preserves privacy and accurate in classification.

Multi-layer neural networks are an important learning model that clearly violates privacy during the process of learning. As a solution to the issue, neural networks can be trained without revealing data over vertically partitioned databases with privacy preserving. Hence the authors have propose a privacy preserving two-party distributed algorithm of back-propagation with privacy preserving computation of activation function [10]. Homomorphic encryption based approach is used in the proposition to ensure reduced cost and hence the algorithm is lightweight in-terms of computation and communication overheads. Error rate decreases when no of epoch increases helping in reduction of accuracy loss and hence is reasonable.

2.4. Privacy Preservation in Biometric Data

Biometric traits of an individual are unique, universal and irrevocable if leaked. Fingerprint identification systems are biometric systems when landed on cloud platform suffers from two major problems. First is a privacy problem and second being need for identification of fingerprint in an acceptable time when the data is placed on cloud servers. Asymmetric homomorphic encryption solution used in earlier systems is non-scalable and hence a proposal of efficient and privacy preserving fingerprint identification system that uses symmetric homomorphic encryption technique [11] is used. In the scheme proposed, whenever client enrolls his/her fingerprint with the system, data is not revealed to data server and cloud service providers. No other entity other than client can access his/her biometric data. Filter-bank fingerprint matching system is used here and it guarantees high accuracy in matching process as it uses nindependent feature codes. The proposed scheme saves storage cost, is time efficient and outperforms earlier proposed privacy preserving schemes in terms of computation and communication.

Face recognition is a very popular biometric trait used in identification of an individual. Cloud services are increasingly being used in face recognition systems for storing huge volumes of data generated and also to perform computationally expensive operations used in recognition. Face recognition system requires interaction between the CSPs and the face owners. CSPs as well as server where the information is stored can intervene with the data and misuse it resulting in violation of privacy. The author here addresses the privacy issue by proposing a Privacy Preserving Face Recognition (PPFR) scheme that uses randomness techniques using affine transforms [12]. Encryption algorithm is used before outsourcing the biometric data that combines permutations, diffusion and shift transforms to ensure privacy of data. Randomness technique is used in protecting Eigen faces ensuring no interaction between face owner and service provider. The above said technique is storage and cost efficient.

3. Data Privacy Preservation Mechanisms and Algorithms

We classify the data privacy preservation algorithms and mechanisms based on the nature of mechanisms as: Data Storage, Data management, Data confidentiality, Data access control, Data sharing, on-demand cloud services, proximity privacy breach, server resource allocation and accountability.

3.1. Data storage and Privacy Preservation

We classify data privacy preservation mechanisms and algorithms in data storage as: encrypting intermediate datasets, data leakage, insiders malicious operations/attacks, file injection attacks, resource consumption and access control rights. Figure 2 presents the challenges along with mechanisms and algorithms used to handle data privacy in cloud data storage. The rectangles present in the diagram represent the mechanisms and ovals denote the algorithms proposed in the literature.

Encrypting Intermediate Datasets. Massive amount of intermediate datasets are being generated as a result of computations in cloud environment for data intensive application. Adversary can derive useful information if intermediate data sets are unencrypted. But storage cost considerably increases when all the intermediate data is encrypted. A solution is hence proposed to have an efficient and cost effective novel upper bound privacy leakage constraint-based approach, for encrypting interme

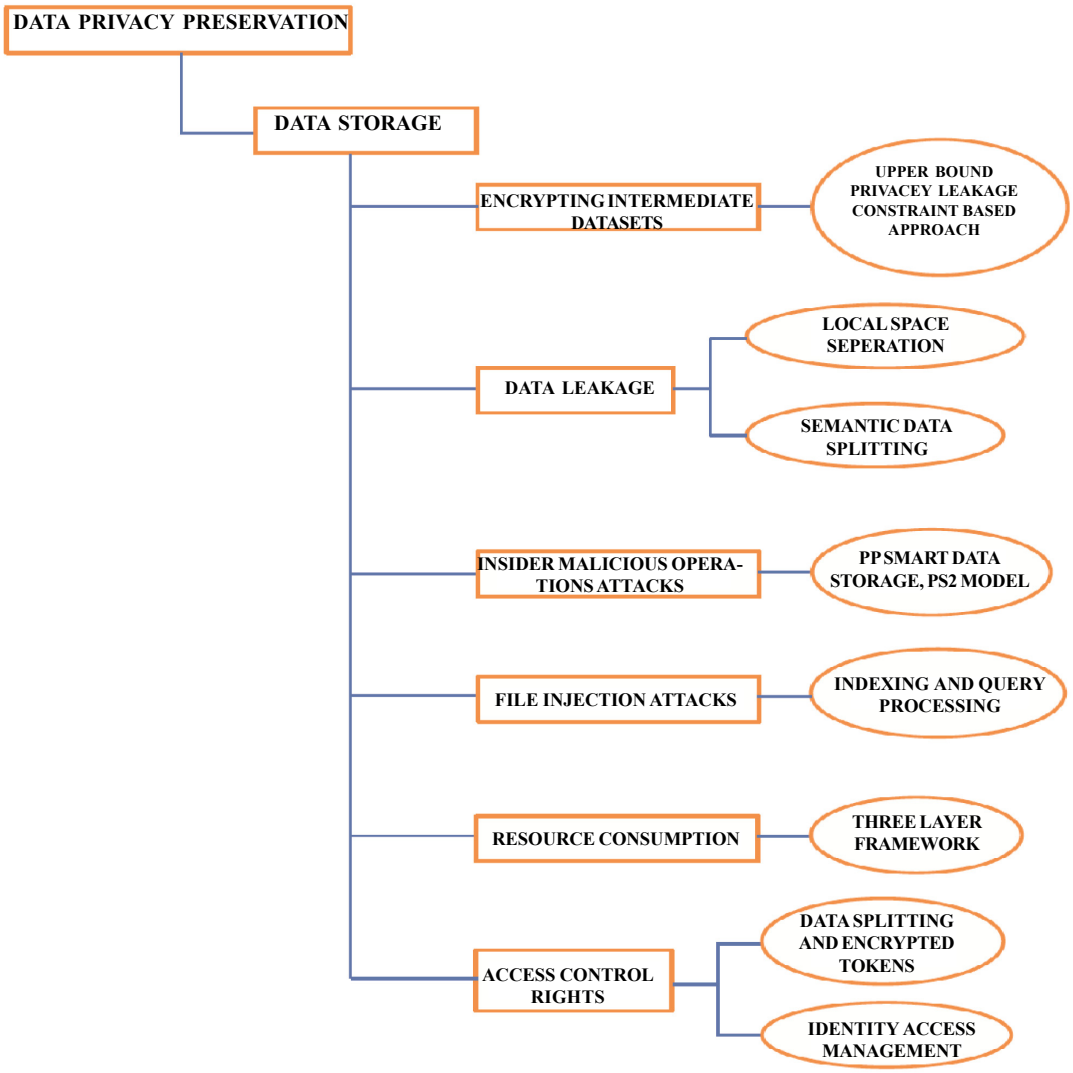

Figure 2. Data Privacy Preservation Mechanisms and Algorithms in Data Storage

diate datasets [13]. Privacy propagation is analyzed in datasets initially, and an upper bound is decided for privacy disclosure. Those intermediate datasets are anonymized. Saving privacy preserving cost is a constrained optimization problem. The above said approach ensures cost effectiveness, scalability and is time efficient.

Data Leakage. Increased use of cloud services in various areas such as health, finance and businesses relies on its advantages that resources are available at a reduced cost, increased storage, flexible in nature. All the data outsourced to the cloud finds it placed in datacenters which must be secure and reliable. Data storage is an important area where private data of user is placed and that needs to be preserved and not exposed for attacks. Authors focus on privacy of data storage in cloud and have proposed methods in solving data leakage, data integrity, data availability and access control aspects [14]. Logical space separation classifies the data into private and nonprivate. Probabilistic methods are used for such classifications. Private data is given an additional level of security by encryption. Private and non-private data are both encrypted using a single key and stored. Identity access management is used in solving authentication issue. The system is proven to be cost effective, has a reduced maintenance and increased efficiency of resources.

Lack of control over data storage and data management of confidential data in cloud resulting in data loss or leakage hinders the adoption of cloud by potential users. Encryption seems to be the immediate solution to solve these issues but they seem to hamper the efficiency and limit the functionality provided by cloud services. To retain these crucial advantages a mechanism is

proposed by authors to semantically split the data on local premises in multi clouds, provide privacy to the chunks of data which are at risk so that adversaries are not able to derive the original data [4]. The proposed mechanism has the contents of resulting chunks after partition which is based on actual semantics of data. It uses C sanitization where privacy requirements are intuitively defined by data owners so that they are not exposed to attackers. The proposed approach is scalable and efficient even for the most common search and retrieval operations.

Insiders Malicious Operations/Attacks. Smart data storage in cloud has pushed in improving old businesses and setting up new businesses. Institutions like finance migrate their services to cloud based model for higher level flexibility and hence utilize benefits of cloud services. Major concern of these improvements is privacy leakage caused by malicious insiders operations. The authors hence propose Privacy Preserving Smart data Storage (PS2) model in distributed data storage scenario [15]. Data storage request operations in PS2 uses distributed encryption algorithm where private data is split and stored on different cloud servers. Distributed decryption algorithm decrypts data retrieved from cloud servers and performs inverse of encryption operation. Performance assessment of PS2 terms the solution as providing higher level data protection and less data processing in cloud and hence faster in terms of execution time, adaptable and feasible.

File Injection Attacks. Querying over encrypted databases in cloud attributes to protection of privacy of queries and databases but leaks considerable amount of information. Dynamic searchable symmetric encryption was hence used, but these schemes are vulnerable to file injection attacks. Towards these attacks and support dynamic data operations with security and forward privacy, authors propose privacy preserving indexing and query processing protocols [16]. Probabilistic inverted index code structure is a new index code design used for query processing of encrypted data storage. Solution proposed supports multi-keyword queries ensuring query efficiency and privacy.

Resource Consumption and Access Control Rights. User's personal data stored in cloud storage services are accessible using mobile phones, tablets but faces issues related to resource consumption. Lot of mobile resources such as processing (CPU), data (usage of internet) and battery are consumed during access of multi-cloud storage services. Hence an efficient and secure multi-cloud storage system for mobile devices is proposed by authors resulting in saving of resources namely battery, CPU, data usage [2]. This proposal includes objectives to save the resources using a 3 layer framework. To provide security data is split and stored using different CSP's and privacy is maintained using encrypted tokens. Security of user sensitive data and resources of mobile devices are saved through the approach proposed.

Figure 3 presents the challenges along with mechanisms and algorithms used to handle data privacy in cloud systems. The rectangles present in the diagram represent the mechanisms and ovals denote the algorithms proposed in the literature.

3.2. Data Management

Content Based Image Retrieval (CBIR) is one of the data services provided in cloud for encrypted image data stored on cloud environment. Data management and data sharing are the major challenges to be addressed as the image data is encrypted for security purposes leading to image retrieval issues. If images have to be decrypted during retrieval operation then sensitive information may get leaked to the honest but curious CSP leading to major privacy breach. The authors have hence proposed a privacy preserving content based image retrieval method combined with an improved Bag of Visual Words model (BoVW) and orthogonal transformation [17]. Improved BoVW model is based on hamming embedding which helps to achieve higher accuracy in retrieval of large scale images. AES-128 encryption algorithm encrypts images. Orthogonal decomposition is applied on image features instead of images. This method achieves advantages in-terms of high retrieval accuracy and security of images.

Data management in cloud based on storage efficiency is improved using a technique named data deduplication. Data deduplication is a specialized data compression technique used in eliminating duplicate copies of data hence ensures reduced storage and reduced network communication requirements. Due to its intervention in cloud it causes loss in data confidentiality. A trusted execution environment (TEE) based secure data deduplication is hence proposed that aims in solving the issue of confidentiality of sensitive data and storage efficiency [18]. To ensure compatibility between data deduplication and encryption the above said scheme used Proof of ownerships (PoW) and Convergent Encryption (CE) to balance storage efficiency and data privacy. Privilege based authentication code is used to ensure in preventing malicious users without privilege from accessing files. Any file before being outsourced to Cloud Service Provider, uses Convergent Encryption to support secure data deduplication. Secure key management is ensured using TEE and the said scheme is efficient and feasible.

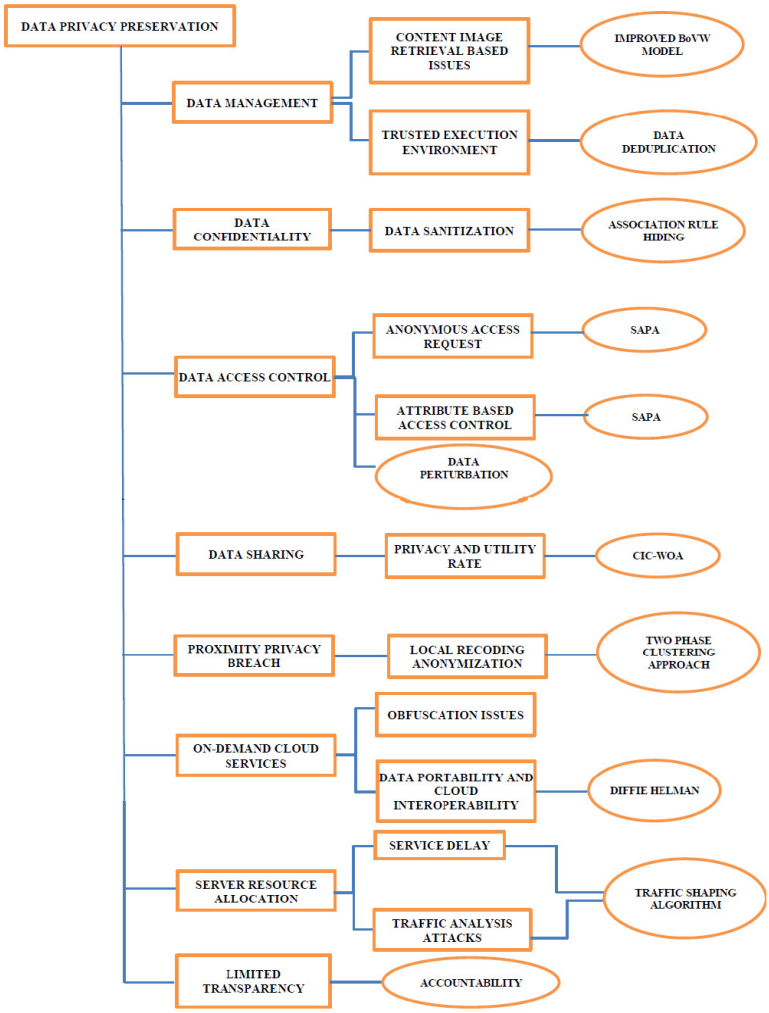

Figure 3. Taxonomy of Data Privacy Preservation Mechanisms and Algorithms

3.3. Data Confidentiality

Conserving data confidentiality against data mining methods is done by concealing sensitive rules with appropriate modifications to prevent unauthorized access. Data sanitization is the method to solve data confidentiality issue against the data mining methods. Association rule hiding is a privacy preserving data mining technique used in solving the sanitization research challenges but modifies the database resulting in loss rule, false rule and ghost rule in database. The author proposes a method to reduce the four sanitization research challenges namely hiding failure, information loss, false rule generation and modification degree using firefly optimization algorithm [19]. Sanitization key is generated using firefly algorithm which effectively hides sensitive rules of large databases as well as helps in recovery from malicious attacks. The method proposed is effective and robust.

3.4. Data Access Control

In a multi-user collaborative cloud based applications, different users access data that interests them to grow in their businesses and achieve significant benefits. In data sharing scenario, data may be authenticated before use but access to one another's data may come from unauthorized users leading to privacy issues which is a challenge considered in cloud. Proposed approach is named as Shared Authority based Privacy preserving Authentication protocol (SAPA) to ensure security and privacy considerations in data sharing [1]. SAPA use integrative approaches. The proposed method takes care of two aspects i.e., it is aimed at users who like to share data field with others and also users who want to decline/ignore the requests. SAPA presents an anonymous access request mechanism for the users requesting for shared access to data. Proxy re-encryption and Attribute

based access control also is used to provide controlled data access. Forward security is realized using SAPA. It also promises data anonymity and authentication.

Cloud security issue concerns three entities i.e., privacy of data owners when their data is outsourced to the cloud, for the CSP's who need information about the data to provide QOS services and for authorized users to be able to access true value of data. To meet the need of all three parties authors propose a retrievable data perturbation method with data access control[20]. Perturbed data adds noise to the original data but maintains numerical statistics of original data i.e., mean and covariance in the process remains unchanged. Original data can be retrieved from perturbed data with key by authorized users. So the proposed method does not compromise on privacy aspect.

3.5. Data Sharing

In data sharing to establish a secure cloud environment it is necessary to identify unauthorized and malicious users from the authorized ones and hence prevent them from accessing data content. Data anonymization model is useful in data sharing and releasing scenarios for data security. Major challenge this model beholds is the high privacy and utility rate. Hence to overcome these challenges authors propose a Circular interpolation and Chronological-Whale Optimization algorithm (CIC-WOA) based coefficient generation for data privacy preservation [21]. Chronological WOA is used to find optimal privacy and usage coefficient by performing circular interpolation of data. Fitness function is used to maximize privacy and utility of the privacy and utility coefficient. The algorithm is scalable and time efficient.

3.6. Proximity Privacy Breach

Big data applications like healthcare and businesses that make use of cloud services are concerned about the privacy issue as sensitive information is often released or shared among third parties in cloud. Privacy-sensitive information can be obtained with less effort by an adversary as big data is used in public cloud environments that disables privacy protection measures. The authors model the problem of big data local recoding against proximity privacy breaches as proximity aware clustering problem and propose a scalable two-phase clustering approach [22]. Two-phase clustering approach has t-ancestors clustering algorithm and proximity aware agglomerative clustering algorithm to address scalability problem. Dataset is pre-processed and attribute partitioning is applied. Dataset is classified as general identifier, quasi identifier and sensitive attribute which is anonymized and stored in the cloud. The solution proposed is scalable and time efficient and results in low level data distortion and helps in combating proximity attacks.

3.7. On-demand Cloud Services

Insurance and finance sectors use cloud services as the cloud provides scalable, flexible, on-demand solutions enabling digital data to be stored anywhere in the world helping the business flourish. With the huge amount of data storage in cloud Obfuscation and privacy issues are prevalent which are areas of concern that need immediate solutions. Authors here propose Privacy Preserving Model to Prevent Digital Data Loss in the Cloud (PPM–DDLC)[3] and hence to resolve obfuscation issues in cloud. Data portability and Cloud interoperability is taken care by the Diffie-Helman algorithm. The approach proposed works when there is an agreement that cloud requesters/ end users ensure that all their data have their own privacy policy even though they use different Cloud service providers.

3.8. Server Resource Allocation

Geo-distributed clouds can be used in applications such as e-health monitoring where privacy preserving and minimum service delay is expected in distributed scenario. Transmission of health data using insecure transmission channels in long distance may result in security related attacks. Authors propose a traffic shaping algorithm to help achieve minimum service delay and thwart traffic analysis attacks [23]. The traffic shaping algorithm helps in reducing traffic analysis attacks by data traffic related to e-health to a different form and hence the data becomes unrecognizable, hence solves user privacy issue. Resource allocation scheme is used in allocating server under load balancing condition to minimize service delay. Performance evaluation using simulation demonstrates effectiveness of the system with solution to the problems pointed earlier.

3.9. Accountability

Cloud technologies current legislation provides limited transparency and oversight, fails to encourage privacy innovation, and lacks the flexibility to effectively and efficiently regulate new technologies and globalized business practices. New approaches to data privacy regimes must focus greatly on accountability and translation of practical mechanisms to preserve privacy of data [24]. Customers expect their Personal identification Information (PII) Ito be protected and appropriately used over cloud. Several principles are proposed in the paper relating to accountability, and hence state them as a means of placing legal

responsibility over the organizations owning PII of customers held responsible for providing comparable level of protection and ensuring only appropriate use of PII by third parties using such information. European Union data privacy regulatory structure is hoped to develop hybrid accountability mechanism and address the current problems by combining legal, regulatory and technical means of leveraging both public and private forms of accountability.

4. Analysis of Privacy Preserving Mechanisms

Solutions proposed in the literature for the privacy preservation are evaluated in this paper based on the following performance criteria.

Data Confidentiality. To store private or confidential data in the cloud, authentication and access control strategies are used.

Data Integrity. Without client's knowledge data cannot be physically accessed, checking data for correctness is data integrity.

Scalability. Ability to meet business demands by increasing or decreasing resources as per need ensures scalability in cloud.

Efficiency: Cloud efficiency depends on the following parameters. Efficiency can be measured with respect to time, cost, and storage compared with other methods.

Accuracy. Assessing accuracy in cloud leads to factors such as Retrieval accuracy, Classification accuracy in various applications.

Load balancing. Applications are managed based on workload demand by distributing load and resources in cloud computing environment.

Service Delay. Delay between data owner request and cloud service providers response determines service latency.

Reduced Maintenance. Maintenance of software, OS are taken care by the cloud environment resulting in fewer burdens on client.

Efficient Resource Utilization. Load balancing in servers as per the demand and task consolidation is effective ways of increasing resource utilization.

Computation and Communication Efficient. Communication of data across the network and computation cost involved with the usage of cloud services must be reasonable.

Robustness. Cloud based systems must be reliable and available to ensure that users are not inconvenienced.

Correctness of Data. Data integrity of users is an important factor and is verified by third party auditors generally on behalf of cloud users.

Access Control. Allowing, denying, restricting are the various possibilities that arise in cloud based systems for different kind of users trying to access the data stored.

The observations from the literature review and analysis against the performance parameters are presented in Table 1.

From the literature review it is observed that security and data privacy preserving are the key factors considered by all applications in cloud in general, but in addition the authors have also identified specific factors towards effectiveness. Major significance is attached to parameters like scalability, time efficiency, storage efficiency, computation cost and communication overhead as these serve as deciding factors in ensuring best performance across all applications using cloud.

5. Conclusions

The advent of Cloud technology has paved new trend in emerging businesses and upgradation of old businesses. Security and

Parameter	Data Privacy Preservation Mechanisms
Data Confidentiality	[1]
Data Integrity	[1]
Scalability	[7], [8], [13], [19], [21], [22]
Execution time efficiency	[5], [11], [13], [16], [18], [19], [21], [22], [25]
Cost efficiency	[3], [12], [13], [14], [18]
Load balancing	[23]
Service delay	[23]
Accuracy	[5], [11]
Storage efficiency	[4], [11], [12], [18]
Reduced maintenance	[14]
Efficient resource utilization	[2], [14]
Retrieval accuracy	[10], [17]
Classification accuracy	[8], [9], [10]
Computation and Communication efficient	[6], [7], [9], [10], [11], [18]
Robustness	[8], [19]
Correctness of data	[20]
Access control	[2], [20], [25]
Security and Privacy Preserving	[4], [9], [11], [15], [16], [17], [20], [21], [26]

Table 1. Analysis of Data Privacy Preservation Mechanisms

privacy issues have been considered as key concerns across all sectors where cloud computing is being adopted. Different techniques and methodologies proposed in literature towards solving these issues in multi-cloud, distributed cloud, geo-distributed cloud environments handling heterogeneous data in various applications. Improvements are attempted considering different parameters and better ways are approached while compared to older methods. The authors have classified data privacy preservation challenges in-terms of types of data, mechanisms and algorithms used towards effective solutions. It is observed that, the prime factor parameters such as scalability, execution time efficiency, storage efficiency, computation and communication overhead have been widely used to evaluate various privacy preserving mechanisms and algorithms.

References

[1] Liu, Hong., Ning, Huansheng., Xiong, Qingxu., Yang, Laurence T. (2013). Shared Authority Based Privacy preserving

Authentication Protocol in Cloud Computing , *IEEE Transactions on Parallel and Distributed Systems*, p.1-11, DOI 10.1109/TPDS.2014.2308218, IEEE, 2013.

[2] Bedi, Rajeev Kumar., Singh, Jaswinder., Sunil Kumar Gupta. (2019). An efficient and secure privacy preserving multi-cloud storage framework for mobile devices, *International Journal of Computers and Applications*, p.1-12, DOI: 10.1080/1206212X.2019.1572847, Taylor & Francis group, 2019.

[3] Dhasarathan, Chandramohan., Thirumal Vengattaraman., Ponnurangam, Dhavachelvan. (2017).A secure data privacy preservation for on-demand cloud service, *Journal of King Saud University – Engineering Sciences*, p.144–150, http://dx.doi.org/10.1016/j.jksues. 2015.12.0021018-3639 _ 2015, Elsevier, 2017.

[4] Sánchez, David., Batet, Montserrat. (2017).Privacy-preserving data outsourcing in the cloud via semantic data splitting, *Computer Communications,* 110 (2017), 187-201, http://dx.doi.org/10.1016/j.comcom.2017.06.012 0140-3664, Elsevier, 2017.

[5] Ni, Lina., Li, Chao., Wang, Xiao., Jiang, Honglu., Yu, Jiguo. (2018). DP-MCDBSCAN: *Differential Privacy Preserving Multi-Core DBSCAN Clustering for Network User Data, Special section on Privacy Preservation for Large-scale user data in social networks,* p. 21053- 21063, Volume 6, *Digital Object Identifier 10.1109/ACCESS.2018.2824798, IEEE Access,* 2018.

[6] Wang, Xiaoliang., Bai, Liang., Yang, Qing., Wang, Liu., Jiang, Frank. (2019). A dual privacypreservation scheme for cloud-based eHealth systems, *Journal of Information Security and Applications,* 47 www.elsevier.com/locate/jisa, 132–138, https://doi.org/10.1016/j.jisa.2019.04.010 2214-2126, Elsevier, 2019.

[7] Wang, Bing., Song, Wei., Lou, Wenjing, Hou, Y. Thomas. (2017). Privacy-Preserving Pattern Matching over Encrypted Genetic Data in Cloud Computing, *In*: Proceedings of the IEEE Conference on Computer Communications IEEE INFOCOM 2017, 978-1-5090-5336- 0/17, *IEEE*, 2017.

[8] Lin, Keng-Pei., Chen, Ming-Syan. (2011). On the Design and Analysis of the Privacy-Preserving SVM Classifier, *IEEE Transactions on Knowledge and Data Engineering*, 23 (11), 10.1109/TKDE.2010.193, IEEE, November 3 2011.

[9] Qi Jia., Linke Guo., Zhanpeng Jin and Yuguang Fang. (2018). Preserving Model Privacy for Machine Learning in Distributed Systems, *IEEE Transactions on Parallel and Distributed Systems*, 29(8), Digital Object Identifier no. 10.1109/TPDS.2018.2809624, IEEE, August 2018.

[10] Chen, Tingting., Zhong, Sheng. (2009). Privacy-Preserving Backpropagation Neural Network Learning, *IEEE Transactions on Neural Networks,* 1554-1564, 20(10), Digital Object Identifier 10.1109/TNN.2009.2026902, IEEE, October 2009.

[11] Hahn, Changhee., Hur, Junbeom. (2016). Efficient and privacy-preserving biometric identification in cloud, The Korean Institute of Communications Information Sciences, p. 135-139, http://dx.doi.org/10.1016/j.icte.2016.08.006, Elsevier, 2016.

[12] Guo, Shangwei ., Xiang, Tao., Li, Xiaoguo. (2019). Towards Efficient Privacy-Preserving Face Recognition in the Cloud, *Signal Processing* (2019), p. 1-28, doi: https://doi.org/10.1016/j.sigpro.2019.06.024.

[13] Zhang, Xuyun., Liu, Chang., Nepal, Surya., Pandey, Suraj and Chen, Jinjun. (2013). A Privacy Leakage Upper Bound Constraint-Based Approach for Cost-Effective Privacy Preserving of Intermediate Data Sets in Cloud, *IEEE Transactions on Parallel and Distributed systems*, p. 1192-1202, 24(6), Digital Object Identifier no. 10.1109/TPDS.2012.238, IEEE, (June).

[14] Joshi, Bansidhar., Joshi, Bineet., Kritika Rani. (2017). Mitigating Data Segregation and Privacy Issues in Cloud Computing, In the Proceedings of International Conference on Communication and Networks, Advances in Intelligent Systems and Computing 508, p. 175-182, DOI 10.1007/978-981-10-2750-5_18, Springer Nature Singapore Pte Ltd. 2017 N. Modi et al. (eds.), 2017.

[15] Qiu, Meikang., Gai, Keke., Zhao, Hui., Liu, Meiqin. (2017). Privacy-preserving smart data storage for financial industry in cloud computing", *Concurrency Computat Pract Exper*. 2018; 30:e4278, p. 1-10, https://doi.org/10.1002/cpe.4278, wileyonlinelibrary.com/ journal/cpe, John Wiley & Sons, Ltd, 2017.

[16] Du, Minxin., Wang, Qian. (2018). *Member, IEEE,* Meiqi He, Jian Weng, Privacy-Preserving Indexing and Query Processing for Secure Dynamic Cloud Storage", *IEEE Transactions on Information Forensics and Security*, p. 2320-2332, 13(9), Digital Object Identifier 10.1109/TIFS.2018.2818651, IEEE, September 2018.

[17] Gong, Jiaying., Xu, Yanyan., Zhao, Xiao. (2018). A Privacy-preserving Image Retrieval Method Based on Improved BoVW Model in Cloud Environment, *IETE Technical Review*, 35: sup1, p. 76-84, DOI: 10.1080/02564602.2018.1526654, Taylor & Francis group, 2018.

[18] Fan, Yongkai., Lin, Xiaodong., Liang, Wei ., Tan. Gang., Nanda, Priyadarsi. (2019). A secure privacy preserving deduplication scheme for cloud computing, Future Generation Computer Systems, ScienceDirect, p. 127–135, https://doi.org/10.1016/j.future.2019.04.046, Elsevier, 2019.

[19] Geeta S. Navale., Suresh N. Mali. (2018). Lossless and robust privacy preservation of association rules in data sanitization", Springer Science+Business Media, LLC, part of Springer Nature, 1-14, https://doi.org/10.1007/s10586-018-2176-1, Springer, 2018.

[20] Pan, Yang ., Xiaolin, Gui., Jian, AN., Jing, Yao., Jiancai, Lin., Feng, Tian. (2014). A Retrievable Data Perturbation Method Used in Privacy-Preserving in Cloud Computing, *Communications System Design*, 73-84, (August).

[21] Adhirai, S., Paramjit Singh., Mahapatra, Rajendra Prasad. (2019). Circular interpolation and chronological-whale optimization based privacy preservation in cloud, *International Journal of Computers and Applications*, 1-14, DOI: 10.1080/1206212X.2018.1560668, Taylor & Francisgroup, 2019.

[22] Zhang, Xuyun., Dou, Wanchun., Pei, Jian., Nepal, Surya., Yang, Chi., Liu, Chang., Chen, Jinjun. (2013). Proximity-Aware Local-Recoding Anonymization with MapReduce for Scalable Big Data Privacy Preservation in Cloud, *IEEE Transactions on Computers*, TC-2013-12- 0869, 1-14, IEEE.

[23] Shen, Qinghua ., Liang, Xiaohui., Xuemin (Sherman) Shen., Lin, Xiaodong., Luo, Henry Y. (2014). Exploiting Geo-Distributed Clouds for a E-Health Monitoring System With Minimum Service Delay and Privacy Preservation, *IEEE Journal of Biomedical and Health Informatics*, 430-439, 18(2), Digital Object Identifier 10.1109/JBHI.2013.2292829, IEEE, (March).

[24] Charlesworth, Andrew., Pearson, Siani. (2013). Developing accountability-based solutions for data privacy in the cloud", Innovation: *The European Journal of Social Science Research*, 26, 1-2, 7-35, http://dx.doi.org/10.1080/13511610.2013.732753, Taylor & Francis group.

[25] George, Tresa Mary., V., Shamna, S., Jubilant, J. (2016). Kizhakkethottam. *An efficient Privacy- Preserving search scheme with access control for cloud data centers, ScienceDirect Procedia Technology*, 310 – 317, doi: 10.1016/j.protcy.2016.08.112, Elsevier.

[26] Han, Shuguo., Ng, Wee Keong., Li Wan, Lee., Vincent C. S. (2010). Privacy-Preserving Gradient- Descent Methods, *IEEE Transactions on Knowledge and Data Engineering*, 884-899, 22(6), Digital Object Identifier no. 10.1109/TKDE.2009.153, IEEE, (June).