ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

A Survey on Privacy Preservation and Anonymization of Data in Cloud Computing

Asmita R. Patil, M.K. Kodmelwar

M. E Student, Department of Computer Engineering, TSSM's BSCOER, Narhe, Pune, Savitribai Phule Pune University Pune, India

Asst. Professor, Department of Computer Engineering, TSSM's BSCOER, Narhe, Pune, Savitribai Phule Pune University Pune, India

ABSTRACT: In cloud computing maintaining the anonymity of data to be shared across multiple entities is crucial for privacy of users in various application do-mains: patient medical records, electronic voting, e-mail, social networking, etc. Existing Anonymization techniques and new algorithms for assigning anonymous IDs are examined with respect to trade-o s between communication and computational requirements. The new algorithms can be built on top of a secure sum data mining operations using encryption for Privacy Preserving Data Sharing with Anonymous ID Assignment. An algorithm for distributed solution of certain polynomials over finite fields can enhance the scalability of the Anonymization technique. The algorithm for anonymous sharing of private data among parties is to be developed. This assignment of anonymous ids is done for users so that the identities received are unknown to the other members of the group such that it allows more complex data to be shared and has applications to other problems in privacy preserving data mining, collision avoidance in communications and distributed database access.

KEYWORDS: Anonymization and de-anonymization cloud and distributed computing systems, multiparty computation, privacy preserving data mining, privacy protection, security and trust in cooperative communications.

I. INTRODUCTION

Privacy is one of the most concerned issues in cloud computing. Personal data like financial transaction records and electronic health records are extremely sensitive although that can be analyzed and mined by research organization. Data privacy issues need to be addressed before data sets are shared on cloud for analysis purpose. Data anonymization refers to as hiding sensitive data for owners of data records [1]. Large-scale data collections are summed up utilizing two stage best down specialization for information anonymization. This procedure part into two stages. In the main stage, unique informational indexes are apportioned into a gathering of littler informational collections, and these informational indexes are anonymized in parallel, delivering middle outcomes. In the second stage, the middle of the road results are coordinated into one, and further anonymized to accomplish predictable k-unknown [6] informational collections. In such cases, informational indexes are anonymized as opposed to encoded to guarantee both information utility and security saving. Rather than encryption, anonymization of information should be possible as methods for protecting security [2]. Information anonymization is generally utilized strategy for Privacy Preserving of information in non-intelligent information distributing situation Data anonymization alludes to the concealing the personality or delicate information for proprietors information record. The security of individual can be viably protected while some total data is shared for information examination and mining. A few models of security can enhance Data Anonymization incorporate k-namelessness and l-assorted qualities. The web is a prominent correspondence medium whether for individual or business utilize. It has picked up prominence likewise because of its support for mysterious correspondence. Organizations additionally have honest to goodness motivations to take part in mysterious correspondence and evade the punishments of personality exposure. For instance, to permit engendering of rundown information without lighting up the character of the substance the fundamental information is related with, or to ensure shriek blowers appropriate to be mysterious and free from political or monetary responses. Cloud-based site administration devices give capacities to a server to secretly catch the guest's web activities. The issue of sharing secretly held information so that the people who are the subjects of the information can't be distinguished has been

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

explored broadly. Scientists have additionally researched the pertinence of secrecy and additionally security in different application areas: tolerant therapeutic records, electronic voting, email, person to person communication, and so on.

Another type of secrecy, as utilized as a part of secure multi-party calculation, al-lows different gatherings on a system to together complete a worldwide calculation that relies on upon information from each gathering while the information held by each gathering re-mains obscure to alternate gatherings. A protected calculation work generally utilized as a part of the writing is secure whole that enables gatherings to figure the entirety of their individual contributions without revealing the contributions to each other. This capacity is pervasive in information mining applications and furthermore portrays the complexities of the protected multi-party calculation. This work manages effective calculations for relegating identifiers (IDs) to the hubs of a system such that the IDs are unknown utilizing an appropriated calculation with no focal expert. Given hubs, this task is basically a stage of the numbers with every ID being known just to the hub to which it is allocated. The primary calculation depends on a technique for secretly sharing basic information and results in strategies for proficient sharing of complex information. There are numerous applications that require dynamic one of a kind IDs for system hubs. Such IDs can be utilized as a major aspect of plans for sharing/separating correspondences transmission capacity, information stockpiling, and different assets namelessly and without strife. The IDs are required in sensor systems for security or for regulatory assignments requiring unwavering quality, for example, con guration and checking of individual hubs, and download of parallel code or information total portrayals to these hubs. An application where IDs should be mysterious is matrix figuring where one may look for administrations without disclosing the character of the administration requester. To di separate mysterious ID task from unknown correspondence; consider a circumstance where parties wish to show their information on the whole, however namelessly, in openings on an outsider site.

II. MOTIVATION

In cloud computing maintaining the anonymity of data to be shared across multiple entities is crucial for privacy of users in various application do-mains: patient medical records, electronic voting, e-mail, social networkingetc. Multiple tenants share cloud infrastructure, in multi-tenancy. This poses concerns for deliberate or accidental exposure of data. Data safety is important for cloud by preparing for potential security breaches. For instance, store the cloud data in such a way that it is worthless to anyone else except the owner of the data

III. LITERATURE SURVEY

1. SCRYPT

In cryptography, scrypt is a password-based key derivation function created by Colin Percival, originally for the Tarsnap online backup service. The algorithm was specifically designed to make it costly to perform large-scale custom hardware attacks by requiring large amounts of memory. In 2012, the scrypt algorithm was published by IETF as an Internet Draft, intended to become an informational RFC, which has since expired. A simplified version of scrypt is used as a proof-of-work scheme by a number of crypto currenciesfirst implemented by Litecoin.A password-based key derivation function (password-based KDF) is generally designed to be computationally intensive, so that it takes a relatively long time to compute (say on the order of several hundred milliseconds). Legitimate users only need to perform the function once per operation (e.g., authentication), and so the time required is negligible. However, a brute-force attack would likely need to perform the operation billions of times, at which point the time requirements become significant and, ideally, prohibitive. Previous password-based KDFs (such as the popular PBKDF2 from RSA Laboratories) have relatively low resource demands, meaning they do not require elaborate hardware or very much memory to perform. They are therefore easily and cheaply implemented in hardware (for instance on an ASIC or even an FPGA). This allows an attacker with sufficient resources tolaunch a large-scale parallel attack by building hundreds or even thousands of implementations of the algorithm in hardware and having each search a different subset of the key space. This divides the amount of time needed to complete a brute-force attack by the number of implementations available, very possibly bringing it down to a reasonable time frame.The scrypt function is designed to hinder such attempts by raising the re-source demands of the algorithm. Specifically, the algorithm is designed to use a large amount of memory compared to other password-based KDFs, making the size and the cost of a hardware

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

implementation much more ex-pensive, and therefore limiting the amount of parallelism an attacker can use, for a given amount of financial resources.

IV. SALT (CRYPTOGRAPHY)

In cryptography, a salt is random data that is used as an additional input to a one-way function that hashes a password or passphrase.[1] The primary function of salts is to defend against dictionary attacks versus a list of pass-word hashes and against pre-computed rainbow table attacks.A new salt is randomly generated for each password. In a typical setting, the salt and the password are concatenated and processed with a cryptographic hash function, and the resulting output (but not the original password) is stored with the salt in a database. Hashing allows for later authentication while defending against compromise of the plaintext password in the event that the database is somehow compromised.Cryptographic salts are broadly used in many modern computer systems, from UNIX system credentials to Internet security. A public salt makes it more time-consuming to crack a list of passwords. However, it does not make dictionary attacks harder when cracking a single password. The attacker has access to both the hashed password and the salt, so when running the dictionary attack; the attacker can simply use the known salt when attempting to crack the password.To understand the difference between cracking a single password and a set of them, consider a single password le that contains hundreds of usernames and passwords. Without a salt, an attacker could compute hash (attempt [0]), andthen check whether that hash appears anywhere in the le. The likelihood of a match, i.e. cracking one of the passwords with that attempt, increases with the number of passwords in the le. If salts are present, then the attacker would have to compute hash (salt[a] . attempt[0]), where "." denotes concatenation, compare against entry A, then hash(salt[b] . attempt[0]), compare against entry B, and so on. This defeats the "reusing" hashes in attempts to crack multiple passwords.Salts also combat the use of rainbow tables for cracking passwords. A rainbow table is a large list of pre-computed hashes for commonly used passwords. For a password le without salts, an attacker can go through each entry and look up the hashed password in the rainbow table. If the look-up is consider-ably faster than the hash function (which it often is), this will considerably speed up cracking the le. However, if the password le is salted, then the rainbow table would have to contain "salt .Password" pre-hashed. If the salt is long enough and sufficiently random, this is very unlikely. Unsalted pass-words chosen by humans tend to be vulnerable to dictionary attacks since they have to be both short and meaningful enough to be memorized. Even a small dictionary (or its hashed equivalent, a rainbow table) has a significant chance of cracking the most commonly used passwords. Since salts do not have to be memorized by humans they can make the size of the rainbow table required for a successful attack prohibitively large without placing a burden on the users.More technically, salts protect against rainbow tables as they, in effect, ex-tend the length and potentially the complexity of the password. If the rainbow tables do not have passwords matching the length (e.g. an 8-byte password, and 2-byte salt, is effectively a 10-byte password) and complexity (non-alphanumeric salt increases the complexity of strictly alphanumeric passwords) of the salted password, then the password will not be found. If found, one will have to remove the salt from the password before it can be used.Salts also make dictionary attacks and brute-force attacks for cracking large numbers of passwords much slower (but not in the case of cracking just one password). Without salts, an attacker who is cracking many passwords at the same time only needs to hash each password guess once, and compare it to all the hashes. However, with salts, each password will likely have a different salt; so each guess would have to be hashed separately and compared for each salt, which is considerably slower than only comparing.

V. ANONYMIZATION TECHNIQUES

In cloud computing maintaining the anonymity of data to be shared across multiple entities is crucial for privacy of users in various application do-mains: patient medical records, electronic voting, e-mail, social networking, etc. The algorithm for anonymous sharing of private data among parties is to be developed. This assignment of anonymous ids is done for users so that the identities received are unknown to the other members of the group such that it allows more complex data to be shared and has applications to other problems in privacy preserving data mining, collision avoidance in communications and distributed database access. The required computations are distributed without using a trusted central authority.Secure Sum Computation Nodes wish to compute and share only the average of a data item without revealing the value of this data item for any member of the groupTransmitting simple data with Power sums Nodes wishes to share actual data values from their databases rather than relying on only statistical information as

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

shown in the previous section. amongst the participating nodes. Each node has a data item of length b-bits which it wishes to make public anonymously to the other participants.Our network is not anonymous and the participants are identifiable in that they are known to and can be addressed by the others. Methods for assigning and using sets of pseudonyms have been developed for anonymous communication in mobile networks. The methods developed in these works generally require a trusted administrator, as written, and their end products generally differ from ours in form and/or in statistical properties. To be precise, with nodes the algorithms distribute a computation among the nodes generating a permutation of chosen with a uniform probability of from the set of all permutations of where will know only. Such a permutation can also be produced by algorithms designed for mental poker.

The algorithms for mental poker are more complex and utilize cryptographic methods as players must, in general, be able to prove that they held the winning hand. Throughout this paper, we assume that the participants are semi-honest [10], also known as passive or honest-but-curious, and execute their required protocols faithfully. Given a semi-honest, reliable, and trusted third party, a permutation can also be created using an anonymous routing protocol.

Despite the differences cited, the reader should consult and consider the alternative algorithms mentioned above before implementing the algorithms in this paper. This paper builds an algorithm for sharing simple integer data on top of secure sum. The sharing algorithm will be used in each iteration of the algorithm for anonymous ID assignment (AIDA). This AIDA algorithm, and the variants that we discuss, can require a variable and unbounded number of iterations. Finitely-bounded algorithms for AIDA are discussed in Section IX. Increasing a parameter in the algorithm will reduce the number of expected rounds. However, our central algorithm requires solving a polynomial with coefficients taken from a finite field of integers modulo a prime. That task restricts the level to which can be practically raised. We show in detail how to obtain the average number of required rounds, and in the Appendix detail a method for solving the polynomial, which can be distributed among the participants.

VI. PRIVACY PRESERVING APPROACHES IN DATA PUBLISHING

4.1 L-Diversity

L-diversity is a form of group based anonymization that is used to preserve .The l-diversity model is an extension of the k-anonymity model which reduces the granularity of data representation using techniques including generalization and suppression such that any given record maps onto at least k other records in the data. The l-diversity model handles some of the weaknesses in the k-anonymity model where protected identities to the level of k-individuals is not equivalent to protecting the corresponding sensitive values that were generalized or suppressed, especially when the sensitive values within a group exhibit homogeneity.

4.2 T-Closeness

Given the existence of attacks where sensitive attributes may be inferred based upon the distribution of values for l-diverse data, the t-closeness method was created to further l-diversity by additionally maintaining the distribution of sensitive fields. An equivalence class is said to have t-closeness if the distance between the distribution of a sensitive attribute in this class and the distribution of the attribute in the whole table is no more than a threshold t.

4.3 Generalization Approach

By creatively applying Map Reduce on cloud to Bottom Up Generalization (BUG) for data anonymization and deliberately design a group of innovative Map Reduce jobs to concretely accomplish the generalizations in a highly scalable way. Secondly, introduce a scalable Advanced BUG approach, which performs generalization on different partitioned data set and the resulting intermediate anonymization are merged to find final anonymization which is used to anonymized the original data set. Results show that our approach can significantly improve the scalability and efficiency of BUG for data anonymization over existing approaches.

4.4. K-Anonymity

K-anonymity is a property possessed by certain anonymized data. Given person-specific field-structured data; produce a release of the data with scientific guarantees that the individuals who are the subjects of the data cannot be re-

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

identified while the data remain practically useful. A release of data is said to have the k-anonymity property if the information for each person contained in the release cannot be distinguished from at least k-1 individuals.

VII. DIFFERENT PRIVACY PRESERVING APPROACHES

TITLE	AUTHOR	CONCEPT	STRENGTH	WEAKNESS
"A Scalable Two-Phase Top-Down Specialization Approach For Data Anonymization Using Mapreduce On Cloud"[6]	Xuyun Zhang, Laurence T. Yang, Senior Member, Ieee, Chang Liu, And Jinjun Chen, Member, Ieee	A Highly Scalable Two-Phase TDS Approach Is Proposed Using Map Reduce On Cloud	Multiple Data Partition To Improve Scalability And Privacy	The Privacy Preservation In Cloud For Data Analysis, Sharing And Mining Is A Challenging Issue Inadequacy In Handling Large Datasets
Incognito: Efficient Full-Domain K-Anonymity" [8]	K. Lefevre, D.J. Dewitt, And R. Ramakrishnan, Proc.AcmSigmod Conf. Management Of Data	Implementation Framework For Full Domain Generalization Using Multidimensional Data Model Together With Suite Of Algorithms	To Produce Minimal Full-Domain Generalizations Perform Up To An Order Of Magnitude Faster Than Previous Algorithms On Two Real-Life Databases	Performance Of Incognito Can Be Enhanced By Materializing Portions Of The Data Cube, Including Count Aggregates At Various Points In The Dimension Hierarchies
"Workload-Aware Anonymization Techniques For Large-Scale Data Sets"[7]	K. Lefevre, D.J. Dewitt, And R. Ramakrishnan, Acm Trans. Database Systems, Vol. 33, No. 3, Pp. 1-47, 2008	Anonymization Algorithms That Incorporate A Target Class Of Workloads, Consisting Of One Or More Data Mining Tasks As Well As Selection Predicates And The Datasets Much Larger Than Main Memory	High Efficiency And Quality Data Overcomes Problem Of Scalability	Problem Of Measuring The Quality Of Anonymized Data Fails To Work In The Top-Down Specialization (TDS) Approach
"Privacy-Preserving Data Publishing: A Survey Of Recent Developments"[1]	B.C.M. Fung, K. Wang, R. Chen, And P.S. Yu, Acm Computing Surveys, Vol. 42, No. 4, Pp. 1-53, 2010	Provides Methods And Tools For Publishing Useful Information While Preserving Data Privacy	PPDP Has Received A Great Deal Of Attention In The Database And Data Mining Research Communities	Degradation Of Data / Service Quality Loss Of Valuable Information Increased Costs

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

"K-Anonymization As Spatial Indexing: Toward Scalable And Incremental Anonymization,"[9]	T. Iwuchukwu And J.F. Naughton, Proc. 33rd Int'l Conf. Very Large Data Bases (Vldb '07), Pp. 746-757,2007	K-Anonymizing A Data Set Is Similar To Building A Spatial Index Over The Data Set Using R-Tree Index Based Approach	Achieve High Efficiency And Quality Anonymization Multidimensional Generalization, High Accuracy	More Compaction Is Needed To Achieve High Quality Anonymization Different Indexing Algorithm Provide Different Issues
"Mapreduce: Simplified Data Processing On Large Clusters,"[10]	J. Dean And S. Ghemawat, Comm. Acm, Vol. 51, No. 1, Pp. 107-113,2008	Mapreduce Is A Programming Model Implementation For Processing And Generating Large Data Sets Performs Map() And Reduce()	Allows Us To Handle Lists Of Values That Are Too Large In Memory The Model Is Easy To Use	Mapreduce Is Not Suitable For A Short On-Line Transactions
"Sedic: Privacy-Aware Data Intensive Computing On Hybrid Clouds"[15]	Zhang K, Zhou X, Chen Y, Wang X, Ruan Y	To Protect Data Privacy During Map-Reduce Programming	Sensitive User Data Protection High Privacy Assurance Ease To Use Fully Preserve The Scalability	Scalability Problem Occurs
"Data & Knowledge Engineering"[16]	Jiuyong Li , Jixue Liu , MuzammilBaig, Raymond Chi-Wing Wong	Data Generalization In Anonymization Should Be Determined By The Classification Capability Of Data Rather Than The Privacy Requirement	More Accurate Classification Models And Is Faster Than A Benchmark Utility-Aware Data Anonymization Algorithm	Not As Much Faster Than Other Anonymization Methods

VIII. PROPOSED SYSTEM ARCHITECTURE

1. **User Layer:** The users of this system are the users who want to publish and share the data on cloud without revealing the identity of the data owner. The anonymous users can also access the system just to collect or read the data.
2. **Processing Layer:** This layer actually resides in the secure environment. This module can be deployed in the private cloud as well. Here the trusted 3rd Party users are authenticated to be allowed to store the data in the system. Functions of this layer -
3. **Secure Details Cloud Layer:** This layer is deployed on public cloud server onto the public database.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

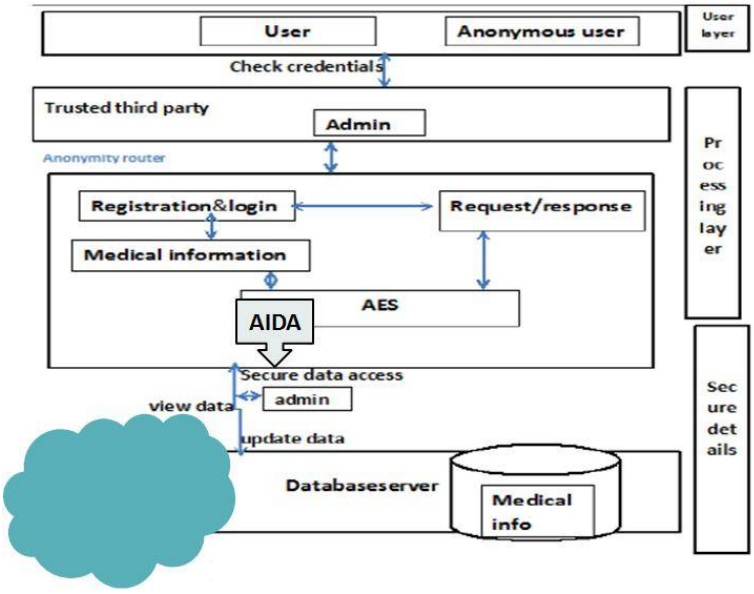

Figure 1: System Architecture

IX. BLOCK DIAGRAM DESCRIPTION

Below is the block diagram which indicates the flow of data from data providers to the DB server after passing through Anonymous authentication, connection and eventually for updates.

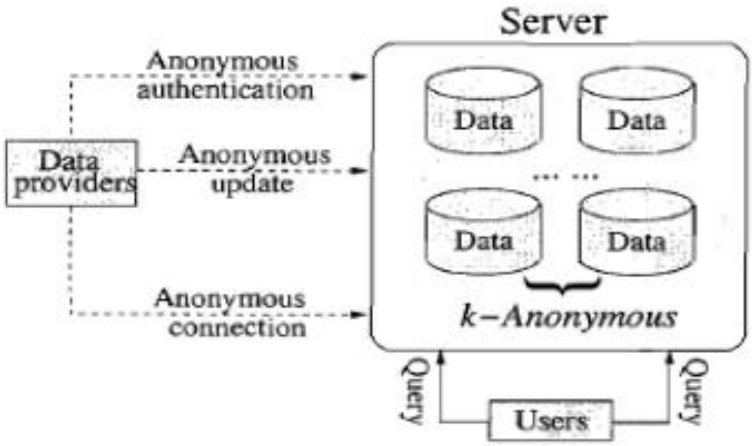

Figure 2: Anonymous DB System Block Diagram

Below is the data table which details the DB server system requirements, in order to carry out anonymization at various layers i.e. at the network connection level, user authentication level and ultimately at the server business layer at the time of data persistence.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

Advantages of Proposed System

1. This model defines the privacy of the output of a process and not of the process itself. This is in sharp contrast to the vast majority of privacy models that were suggested earlier, and it is in this sense of privacy that clients are usually interested.

2. It is a simple, intuitive, and well understood model. Thus, it appeals to the non-expert who is the end client of the model.

3. Although the process of computing a k-anonymous table may be quite hard, it is easy to validate that an outcome is indeed k-anonymous. Hence, non-expert data owners are easily assured that they are using the model properly.

4. The assumptions regarding separation mode of attack and variability of private data have so far withstood the test of real-life scenarios.

5. The model takes care of Security attacks like - Linking attacks, False Rows Injection

6. The model validates that the value of the private attributes for a given set of individuals is not same, even though the value is same the salt is different and hence the encrypted key remains unique.

X. CONCLUSION

Proposed solution greatly decreases communication overhead, by using private communication channel that is anonymity router to transmit the data more secure for persisting the data on private DB. To overcome the problem of identifying details and changing information anonymous id was utilized. Random encrypted key using Scrypt and Salt encryption is used to identify whether the data requesting person is a correct authorized person or hackers. With private communication channels, the algorithms are secure in an information theoretic sense. A lot of ambiguity can be avoided even though the private attributes combination is duplicate, since the salt is used for encrypting the anonymized data. The id generated by applying this process can be used to link the private and public DBs and the false injected rows attacks can be easily identified, since the matching entries will need to be there in public and private DBs both.

REFERENCES

[1] B.C.M. Fung, K. Wang, R. Chen and P.S. Yu, "Privacy-Preserving Data Publishing: A Survey of Recent Developments," ACM Comput. Surv., vol. 42, no. 4, pp.1-53, 2010.
[2] K. LeFevre, D.J. DeWitt and R. Ramakrishnan, "Workload- Aware Anonymization Techniques for Large-Scale Datasets," ACM Trans. Database Syst., vol. 33, no. 3, pp. 1-47, 2008.
[3] B. Fung, K. Wang, L. Wang and P.C.K. Hung, "Privacy- Preserving Data Publishing for Cluster Analysis," Data Knowl.Eng., Vol.68, no.6, pp. 552-575, 2009.
[4] 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.
[5] Amazon Web Services, "Amazon Elastic Mapreduce,"http://aws.amazon.com/elasticmapreduce/, 2013.
[6] Xuyun Zhang, Laurence T. Yang, Senior Member, IEEE, Chang Liu, and Jinjun Chen, Member, IEEE "A Scalable Two-Phase Top-Down Specialization Approach for Data Anonymization Using MapReduce on Cloud", vol. 25, no. 2, february 2014.
[7] K. LeFevre, D.J. DeWitt and R. Ramakrishnan,"Workload- Aware Anonymization Techniques for Large-Scale Datasets," ACM Trans. Database Syst., vol.33, no. 3, pp. 1-47, 2008.
[8] K. LeFevre, D.J. DeWitt, and R. Ramakrishnan, "Incognito: Efficient Full-Domain K-Anonymity," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '05), pp. 49-60, 2005.
[9] T. IwuchukwuandJ.F. Naughton, "K-Anonymization as Spatial Indexing: Toward Scalable and Incremental Anonymization," Proc. 33rdInt'lConf. VeryLargeDataBases (VLDB'07), pp.746-757, 2007
[10] J. Dean and S. Ghemawat, "Mapreduce: Simplified Data Processing on Large Clusters," Comm. ACM, vol. 51, no. 1, pp. 107-113,2008.
[11] A.Machanavajjhala, J.Gehrke, and D.Kifer, et al, "ℓ-diversity: Privacy beyond k-anonymity", In Proc. ofICDE, Apr.2006.
[12] Dean J, Ghemawat S. Mapreduce: a flexible data processing tool. Communications of the ACM 2010;53(1):72–77. DOI: 10.1145/1629175.1629198.
[13] P.JurczykandL.Xiong, "Distributed Anonymization:Achieving Privacy for Both Data Subjects and Data Providers, "Data and Applications Security XXIII(DBSec'09),pp.191-207,2009.
[14] N. Mohammed, B.C. Fungand M. Debbabi, "Anonymity Meets Game Theory: Secure Data Integration with MaliciousParticipants,"VLDBJ.,vol.20,no.4,pp.567-588,20.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 5, May 2017

[15] N. Li, T. Li, and S. Venkatasubramanian, "t-Closeness:Privacy Beyond k-anonymity and l-Diversity", In Proc.of ICDE, 2007, pp. 106-115.
[16] Jiuyong Li, JixueLiu ,MuzammilBaig , Raymond Chi-Wing Wong" Information based data anonymization for classification utility"
[17] Larry A. Dunning, Member, IEEE, and Ray Kresman - Privacy Preserving Data Sharing With Anonymous ID Assignment, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 8, NO. 2, FEBRUARY 2013, pp. 402-413
[18] Friedman, R. Wol , and A. Schuster, Providing k-anonymity in data mining, VLDB Journal, vol. 17, no. 4, Jul. 2008, pp. 789804.
[19] Jun Li and Baochun Li Erasure Coding for Cloud Storage Systems: A Survey TSINGHUA SCIENCE AND TECHNOLOGY, ISSNll1007-0214ll06/11ll, Volume 18, Number 3, June 2013, pp. 259-272.
[20] Dr, Nedhal A, Al-Saiyd, Nada Sail DATA INTEGRITY IN CLOUD COMPUTING SECURITY - Journal of Theoretical and Applied Information Technology, Dec-2013, Vol.59 No.3, pp. 570-582
[21] B.Srinivasulu1, B.V.Usha, R.V.Gandhi, K. Ramakrishna Privacy Pre-serving Updates to Anonymous and Con dential Database (Secure Computing) - International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 7, July 2013 , pp. 2661-2667
[22] Alberto Trombetta, Wei Jiang, Elisa Bertino, Lorenzo Bossi Privacy-Preserving Updates to con dential and Anonymous Databases, Oct-2007, Computer Science Technical Report, Purdue University.
[23] R.K. Pandya and Prof. Sutaria DATA INTEGRITY TECHNIQUES IN CLOUD: AN ANALYSIS -JOURNAL OF INFORMATION, KNOWLEDGE AND RESEARCH IN COMPUTER ENGINEERING, pp.413-417.
[24] B.K.Tripathy, A.Jayaram Reddy ,G.V.Manusha and G.S.MohisinAN EFFICIENT ALGORITHM FOR ANONYMIZATION OF SET-VALUED DATA AND REPRESENTATION USING FP-TREE. - In-ternational Journal of Advanced Information Technology (IJAIT) Vol. 2, No.5, October 2012, pp.1-14