iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4211

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Anonymization Privacy Security and Preserving Similarity Joins for Secured Data

Ms. S. Deepa[1], Mrs. N. Nisha Sulthana[2], Mr. A. K. V. Srinivasan[3]
[1]*PG Scholar, Department of Computer Science and Engineering, Vivekanandha College of Technology for Women, Elayampalayam, Thiruchengode. Tamilnadu, India.*
[2]*Assistant professor, Department of Computer Science and Engineering, Vivekanandha College of Technology for Women, Elayampalayam, Thiruchengode. Tamilnadu, India.*
[3]*Director and Chief Technology Officer, Aqunic private limited, Cuddalore*

Abstract: ***This project proposed a paper approach to share accommodating - specific lengthwise data that offers booming confidentiality guarantees, while preserving data utility for many biomedical investigations. The proposal summation temporal and rectifying information using heuristics excited from continuance alignment and gathering methods. The thesis exposes that the proposed approach can achieve anonymized information that allow powerfull biomedical analysis using many patient disciple imitative from the EMR system. In this thesis, aside from the above work, two obligation clarify this issues on suppression-based and generalization-based k-anonymous and secrete databases are approached. The protocols rely on renowned cryptographic presumption, and we provide theoretical analyses to proof their correctness and provisional results to exhibit their capability.***
Keyword: EMR, K-Anonymity, Gray Gode, KNN

I. INTRODUCTION

Cloud computing is a type of Internet-based that maintains mutual computer processing assets and data to computers and other devices for use. Cloud computing is the phrase used to describe different scenarios in which computing resource is delivered as a service over a network connection (usually, this is the internet). One of the key characteristics of cloud computing is the resilience that it offers and one of the ways that elasticity is offered through scalability. Cost benefits must be extended though, as the operation in question will have to purchase/rent and uphold all the crucial software and hardware. A hybrid cloud grants a team to maximize their competence by utilizing the public for non-sensitive operations while using a private setup for sensitive or mission-critical operations.

Cloud computing exhibits the following key characteristics

A. Agility
B. Application programming interface
C. Cost reduction
D. Device and location independence
E. Maintenance

II. RELATED WORKS

1) *Benjamin C. M. Fung et al [2]* describe the major Anonymity for Continuous Data Publishing stated that k-anonymization is an important privacy protection mechanism in data publishing. While there has been a great deal of work in recent years, almost all considered a single static release. Such mechanisms only protect the data up to the first release or first recipient. In practical applications, data is published continuously as new data arrive; the same data may be anonymized differently for a different purpose or a different recipient. In such scenarios, even when all releases are properly k-anonymized, the anonymity of an individual may be unintentionally compromised if recipient cross-examines all the releases received or colludes with other recipients. Preventing such attacks, called correspondence attacks, faces major challenges. In the paper, they systematically characterized the correspondence attacks and proposed an efficient anonymization algorithm to thwart the attacks in the model of continuous data publishing.
2) *Yousef Elmehdwi, Bharath K et al [3]* describe a novel SkNN protocol to facilitate the k-nearest neighbor search over encrypted data in the cloud that preserves both the data privacy and query privacy. In our protocol, once the encrypted data are outsourced to the cloud, Alice does not participate in any computations. Data confidentiality - Contents of T or any intermediate results

©IJRASET (UGC Approved Journal): All Rights are Reserved

should not be revealed to the cloud. Query privacy - Bob's input query Q should not be revealed to the cloud. Correctness - The output ht′1,...,t′ ki should be revealed only to Bob. In addition, no information other thant′ 1,...,t′ k should be revealed to Bob. Low computation overhead on Bob - After sending his encrypted query record to the cloud, Bob involves only in a little computation compared with the existing works Hidden data access patterns - Access patterns to the data, such as the records corresponding to the k-nearest neighbors of Q, should not be revealed to Alice and the cloud (to prevent any inference attacks).

3) *Alexandra Boldyreva .K et al l [4]* describe a order-preserving symmetric encryption (OPE), a primitive for allowing efficient range queries on encrypted data, recently initiated (from a cryptographic perspective). First, author address the open problem of characterizing what encryption via a random order-preserving function (ROPF) leaks about underlying data (ROPF being the "ideal object" in the security definition, POPF, satisfied by their scheme.) In particular, we show that, for a database of randomly distributed plaintexts and appropriate choice of parameters, ROPF encryption leaks neither the precise value of any plaintext nor the precise distance between any two of them. The analysis here introduces useful new techniques. On the other hand, we show that ROPF encryption leaks approximate value of any plaintext as well as approximate distance between any two plaintexts, each to an accuracy of about square root of the domain size. They are study schemes order-preserving, but which nevertheless allow efficient range queries and achieve security notions stronger than POPF. In a setting where the entire database is known in advance of key-generation (considered in several prior works), they show that recent constructions of "monotone minimal perfect hash functions" allow to efficiently achieve (an adaptation of) the notion of IND-O(rdered) CPA also considered., which asks that *only* the order relations among the plaintexts is leaked. Finally, we introduce *modular* order-preserving encryption (MOPE), in which the scheme is pretended with a random shift cipher. MOPE improves the security of OPE in a sense, as it does not leak any information about plaintext location.

4) *HeleiCui, et et al [5]* describe a Near-duplicate detection (NDD) plays an essential role for effective resource utilization and possible traffic alleviation in many emerging network architectures, leveraging in-network storage for various content-centric services. As in network storage grows, data security has become one major oncern. Though encryption isviable for in-network data protection, current techniques are still lacking for effectively locating encrypted near-duplicate data, making the benefits of NDD practically invalidated. Besides, adopting encrypted in network storage further complicates the user authorization when locating near-duplicate data from multiple content providers under different keys. In this paper, they propose a secure and effective NDD system over encrypted in-network storage supporting multiple content providers. Author design bridges locality sensitive hashing (LSH) with a newly developed cryptographic primitive, multi-key searchable encryption, which allows the user to send only one encrypted query to access near-duplicate data encrypted under different keys. It relieves the users from multiple rounds of interactions or sending multiple different queries respectively. As simply applying LSH does not ensure the detection quality, author then leverage Yao's garbled circuits to build a secure protocol to obtain highly accurate results, without user-side post-processing. We formally analyze the security strength. Hough considering privacy protection, these designs all focus on eliminating duplicates via exact fingerprint matching. Differently, our work targets a more general case, i.e., secure NDD.

5) *Dinur .I et al [6]* describe query monitoring, queries to an online database are audited to ensure that, even in the context of previous queries, the responses do not reveal sensitive information. This is sometimes computationally intractable, and may even fail to protect privacy, for example, in the setting in which the adversary knows even one real database record. A related approach is output perturbation, in which a query control mechanism receives queries, computes exact answers, and then outputs a perturbed answer as the response to the query. This approach can sometimes be insecure, intuitively, because noise added in response to multiple queries can cancel out. Recent Definitional Approach. The definitions of privacy (written concurrently with this work –are consonant with their point of view, in that they provide a precise, meaningful, provable guarantee. All three follow the same paradigm: for every record in the database, the adversary's confidence in the values of the given record should not significantly increase as a result of interacting with or exposure to the database. Neither their definitions nor their techniques make this assumption. This both enhances utility and complicates privacy arguments. Cryptographic Approaches Much work in cryptography has focused on topics loosely related to database privacy, such as private information retrieval and secure function evaluation. These problems are somewhat orthogonal to the one considered here. In secure function evaluation, privacy is preserved only to the extent possible given a specific functionality goal.

III. METHODOLOGY

The issue is to orderly anonymous continuing data history. scheme to reduce recognition via analytical and scientific appearance are unworkable to the long term scheme. These approach accept the scientific sketch is lacking of sensual or replicate investigation

©IJRASET (UGC Approved Journal): All Rights are Reserved

report. Consequently, these methods produce data that are unlikely to permit meaningful long term analysis. So the scheme for intercept recognition in relative data (e.g.,statistics) are outlined, where history have a fixed number of virtue and one value per virtue. The first list of assurance history translate virtue ideals so that they no longer resemble to real existence.

A. Anonymity definitions

1) Quasi-Identifier (QI): A set of virtue that can be used with secure visible report to find a specialized existence With respect to elimination-based anonymous, there exists a subgroup of triple **{t1;...;tz}T(z k – 1)** such that for every virtue in QIs, the comparable rate is repossessed by ***** (indicating elimination of the authentic rate).For observation-based anonymous [32], it is simulated that each virtue rate can be graphed to a more natural rate. The main step in most observation based k-anonymous protocols is to change a limited rate with a more natural rate. The following symbol view the typical Value observation pecking order access.

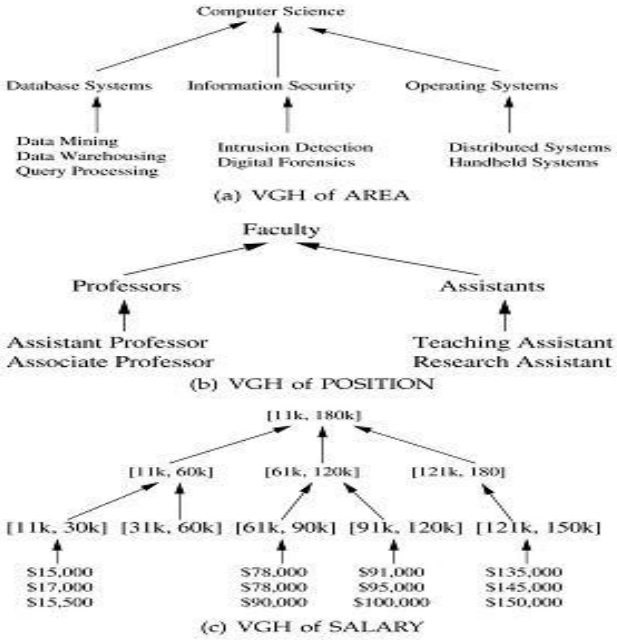


Fig 3.1 a typical Value Generalization Hierarchy

B. Protocol operation

To provide secure communication between database owner and data provider (the records given from database owner and data provider should not be identified by data provider and vice-versa), a commutative, product-homomorphic encryption scheme E is introduced. A commutative, product-homomorphic encryption scheme ensures that the order in which encryptions are performed is irrelevant (commutativity) and it allows to consistently perform arithmetic operations over encrypted data (homomorphic property). Further, for the security proofs we require that the encryption scheme E satisfies the indistinguishability property. The scheme should produce an encryption method which should be product-homomorphic. Given a finite set K of keys and a finite domain D, a commutative, product-homomorphic encryption scheme E is a polynomial time computable function E : K * D -> D satisfying the following properties.

1) Commutativity. For all key pairs K1,K2value d D, the following equality holds:

$$E_{K1}(E_{K2}(d)) = E_{K2}(E_{K1}(d))$$

2) Product-homomorphism. For every K and every value pairs d1;d2D, the following equality holds:

$$E_K(d) \cdot E_K(d2) = E_K(d1.d2)$$

3) Indistinguishability. It is infeasible to distinguish an encryption from a randomly chosen value in the same domain and having the same length. In other words, it is infeasible for an adversary, with finite computational capability, to extract information about a plain text from the cipher text.

Finally, a simple tuple coding scheme is introduced which is used in the next protocol operation. For example Alice and Bob agree on a set {g1;g2 ...;gu}of generators of D.

©IJRASET (UGC Approved Journal): All Rights are Reserved

C. Algorithms used for database update

The protocol works as follows: At Step 1, Alice sends Bob an encrypted version of i, containing only the s non-suppressed QI attributes. At Step 2, Bob encrypts the information received from Alice and sends it to her, along with encrypted version of each value in his tuple t. At Steps 3-4, Alice examines if the non suppressed QI attributes of i is equal to those of t.

System Flow Diagram

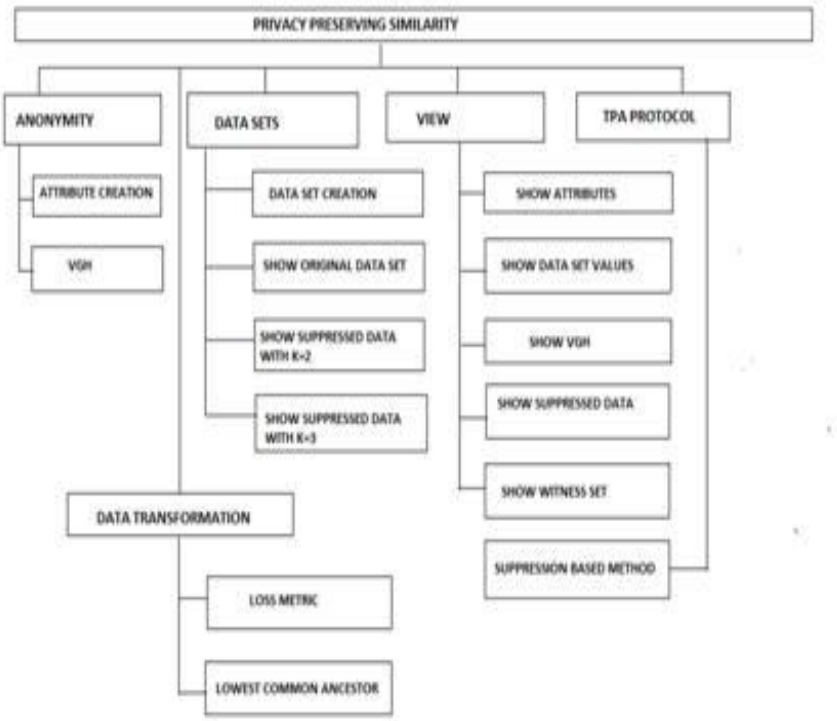


Fig.a. Privacy Preserving Similarity

IV. CONCLUSION

This work was motivated by the growing need to disseminate data specify long term data in a privacy-preserving manner. It introduces the first access to sharing such input while providing computational privacy guarantees. The approach uses clustering-based heuristics to anonymous long term data records. The investigations suggest that it can generate longitudinal data with a low level of information loss and remain useful for biomedical analysis. This was illustrated through extensive experiments with data derived from the EMRs of thousands of data. The approach is not guided by specify utility but it is confident that it can be extended to support such endeavors. In this project, procedures are carried out for privately checking whether a k-anonymous database maintain its anonymity once a new triple is being added to it. For the suggested protocols assure the updated database maintain k-anonymous, the results restored from a user's (or a medical researcher's) query are also k-anonymous. Thus, the data or the data provider's privacy cannot be violated from any query. since the database is updated properly using the proposed protocols, the user queries under the application domain are always privacy-preserving.

V. RESULTS AND DISCUSSION

Experimental analysis is to be of use to researchers from all fields who want to study algorithms experimentally. To demonstrate the proposed method classic KNN classification is used and compare it performance with KNN Classification.

A. Experimental Results

The following "Table 5.1" describes Secure Outlier model for existing K-Autonomy and KNN classification algorithm. The table contains number of datasets, average for K-Autonomy algorithm and average performances for KNN Classification algorithm details are given below.

©IJRASET (UGC Approved Journal): All Rights are Reserved

Number of Datasets [N]	K-Autonomy	KNN-Classification
100	50.2	51.33
200	58.67	59.32
300	64.03	65.34
400	72.33	73.44
500	76.12	77.98
600	79.33	79.89
700	80.44	81.04
800	81.45	82.78
900	83.22	84.03
1000	84.10	85.65

Table 5.1 Performances Analysis- K-Autonomy – KNN Classification Algorithm

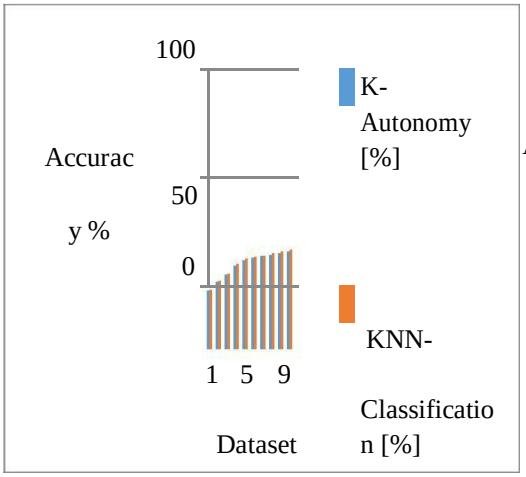


Fig 5.1 Performances Analysis K-Autonomy - KNN Algorithm

The following “Fig 5.1” describes Secure Outlier model for existing K-Autonomy and Classic KNN Classification algorithm. The figure contains number of observation dataset, average for K-Autonomy algorithm and average performances for Classic-KNN Classification algorithm details are given below.

©IJRASET (UGC Approved Journal): All Rights are Reserved

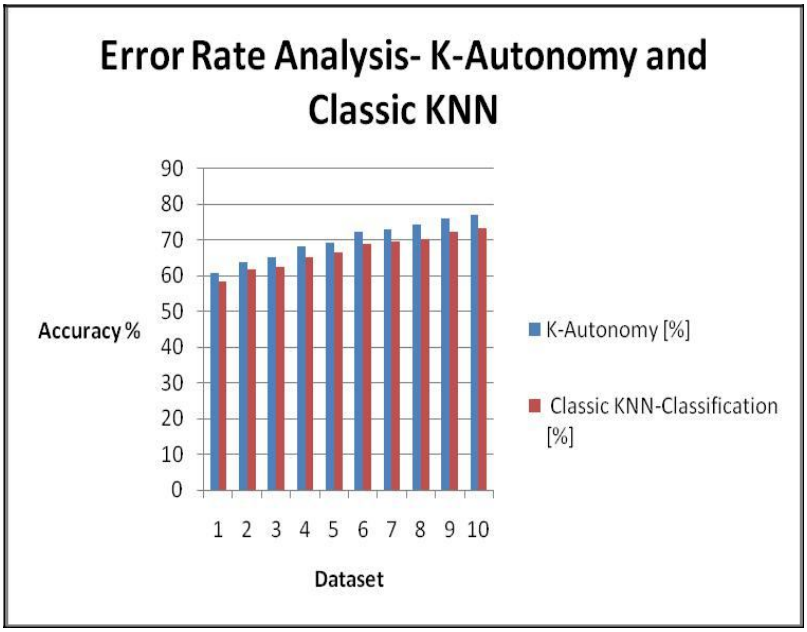


Fig 5.2: Error Rate Analysis of K-Autonomy - Classic-KNN Classification Algorithm

The following "Fig 5.2" describes Secure Outlier model for existing K-Autonomy and Classic KNN Classification algorithm. The figure contains number of observation dataset, time taken for K-Autonomy algorithm and time taken performances for Classic-KNN Classification algorithm details are given below

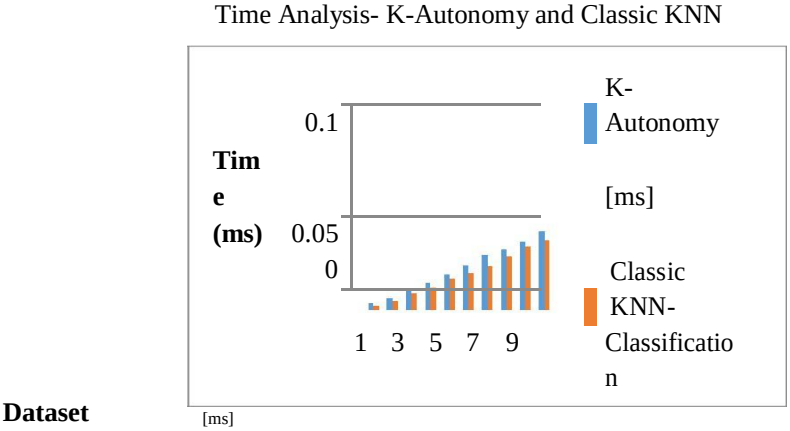


Fig 5.3 Performance Time Analysis of K-Autonomy -Classic-KNN Classification

VI. FUTURE ENHANCEMENTS

In order for a database system to energetically perform privacy preserving updates to a k-anonymity table the important issues addressed are: 1) the definition of a mechanism for actually performing the update, once k-anonymous has been verified. 2) The integration with a privacy-preserving query system.

In addition to the problem of falling insertion, there is other interesting and relevant issues that remain to be addressed in future are:

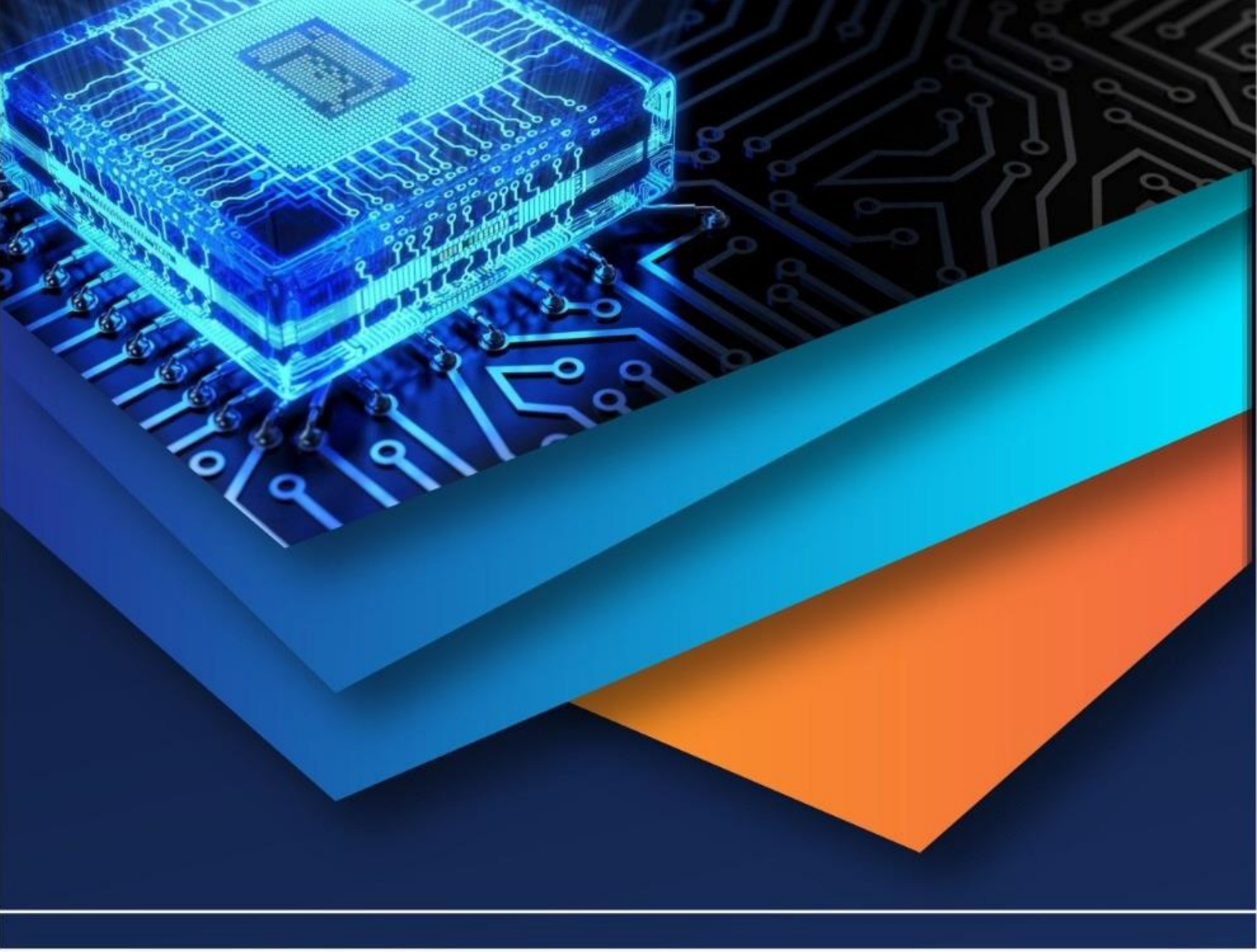
A. Devising separate update method to database systems that supports notions of anonymity different than k-anonymous.
B. Dealing with the case of vicious parties by the introduction of an entrusted, non-colluding third party.
C. Implementing a real-world anonymous database system.

©IJRASET (UGC Approved Journal): All Rights are Reserved

REFERENCES

[1] yu. chenyun, and sarana nutanong "privacy preserving similarity joins over encrypted data" (TIFS.2017) citation information " dio.10.1109
[2] Benjamin.C.M. Fung, K. Wang, A.W.C. Fu, and J. Pei, Conf. (EDBT), (2016) "Anonymity for Continuous Data Publishing," Proc. Extending Database Technology, vol.55,pp.170-181.
[3] Yousef Elmehdwi, Bharath K [2013], " Secure k-Nearest Neighbor Query over Encrypted Data in Outsourced Environments",vol.78,pp.192-120.
[4] Alexandra Bolsyerta, Natham Chenette, Adam O'Neill, [2011], "Order-Preserving Encryption Revisited: Improved Secyrity Analysis and Alteranative Solution" vol.59, pp.179-210
[5] Helei Cui Xingiang yuan, et al, [2013] "Enabling Secure and Effective Near-duplicate Detection Over Encrypted In- network storage, vol.61.pp.123-179.
[6] Dinur.I and Nissim .K [2003] "Revealing Information while preserving privacy, processing Of the Symposium on Principles of databases Systems", vol.78,pp. 202-210.
[7] Gan. J., Feng J., Fang Q., and W. Ng. "Locality-Sensitive hashing scheme based on dynamic Collision counting", In Proc. of ACM SIGMOD2 2012
[8] Hahn .f and Kerscbaum, f., "Searchable Encryption with secure and efficient updates". In Proc. Of ACM CCS, 2014.
[9] Michael J. Feedman, Kobbi Nissim, and Benny Ponkas, [2014], "Efficient private Matching And Set Intersection", vol.60,pp.200-172
[10] Rakesh Agarwal, Alelxandre Evfimieyski and Ramakrishnan Srikant [2003], "Information sharing Across private Darabases", vol.79pp.127-223.

©IJRASET (UGC Approved Journal): All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)