iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** I **Month of publication:** January 2022

DOI: https://doi.org/10.22214/ijraset.2022.39876

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: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

Shortest Distance Queries on Encrypted Graph Data Based on GENOA

Jinsha Lawrence[1], Dr. R. Suji Pramila[2]
[1]*Department of Computer Science and Engineering, Noorul Islam Center for Higher Education*
[2]*Asst.prof/CSE, Department of Computer Science and Engineering, Noorul Islam Center for Higher Education*

Abstract: ***In the execution time of the nearest distance queries, which means shortest distance queries on the encrypted graph data that stored in external storage like cloud storage, in previous scheme there are some challenges as how to figure the accurate shortest distance in efficient and secure way. In previous work a novel scheme of Somewhat Homomorphic Encryption (SWHE) is implemented to overcome the issues mentioned above and this SWHE is used for encryption. The SWHE will encrypt the output values (shortest distance) by 2-hop cover labeling (2HCL). The storage space of this SWHE is not sufficient for the data owner and it mostly gives the negative results and very low efficient. So, to provide the best efficient, accurate result and better storage for data owner a new novel scheme is implemented in this paper. A new proposed scheme called Graph Encryption scheme for shortest distance queries (GENOA) and the 2HCL index is also implemented. This GENOA is for execution of the shortest distance queries of graph data, this proposed scheme is highly efficient, and give the accurate result as per the queries. This new scheme help in storage and security also.***

Index Terms: Graph encryption; 2-hop cover labelling; Graph Encryption scheme for shortest distance queries (GENOA); Shortest distance queries.

I. INTRODUCTION

Data encryption is introduced to reduce the issues that created by data leakage and some security issues. Most commonly cyber-attack is done on outsourced data. In cloud computing the data outsourcing is an important application. So as we know that the cloud computing is normally accessed by public manner. The data security is low for a data when that is outsourced, even the data owner will lose the ownership of that data that outsourced or its privacy will get affected more. So data outsourcing in cloud computing is not much secured. Accessing the outsourced data is not secured, then the user privacy is also considerable, there is no assurance that it will be an accurate data, some data might be get leaked and also data owner will face low efficiency. In recent days the storage is also insufficient for data owners to access the data. To lecture this issues that faced by the data owner the existing scheme implemented a sequence of encryption schemes. This [1] (survey) new proposed scheme allow to perform search on the encrypted textual data and not only the word documented data and also other data types. In this paper examined how to achieve the shortest distance queries on the encrypted graph data. Then the data owner will encrypt the graph data in a secured manner and outsource it to a storage provider and it can still answer the shortest distance queries without knowing the other sensitive or the other information of the outsourced encrypted graph data.

To make the shortest distance queries search performance efficient 2-hop cover labelling (2-HCL) is introduced in [2]. The 2-HCL help to index the encrypted graph data, by indexing the graph data the subsequent short distance queries can be replied efficiently. In [3] the author proposed a novel scheme Highway-Centric Labeling for answering the shortest distance queries. This empowers the distance in a large space and it develop the highway structure and leverages two-part cover set algorithm. In [4] author et.al David Cash introduced the searchable symmetric encryption (SSE) to improve the user privacy, reduce the data leakage and secure the data search. But this SSE scheme can perform with the limited users only, when there is multiple users it is hard to manage it. In [5] to minimize the space cost and satisfying the security and utility requirements the author proposed 1-neighborhood-d-radius. The proposed schemed reduced the space cost and the shortest distance evaluation cost for data owner side. The issue faced in this [5] is to provide a perfect security for the outsourced graph data and how to don the service for data owner side by considering the privacy constrains. So it is focused on securing the sensitive information of the data.

If the data is outsourced without encryption, the data will get leaked by the storage provider. The storage provider will comes to know about the data. So the data security will get reduced. For instance there is a graph A and for each vertex $v \in A$, the indexing scheme collects the set $S(v)$ of candidate vertices like that the vertex pair (s, t) there is one vertex satisfying the three conditions that is 1) $u \in S(s)$, 2) $u \in S(t)$ and 3) u will be on the shortest distance path or in-between the path of s and t.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

The each vertex v is belongs to a graph A ($v \in A$) and it is labelled with $Lv = \{(u, d_{u,v})\}$ $u \in S(v)$ in this u is a vertex identifier and du;v is the shortest distance between u and v. 2-HCL is known as the collections of labels, such as $\{Lv\}$ $v \in A$. The specified index, answering the shortest distance between the assumed vertices *s* and *t* it normally calculate *the* $min\{du,s + du,t \mid (u, du,s) \in Ls, (u, du,t) \in Lt\}$.

There are some tactics to know how the cloud storage provider comes to know about the un-encrypted data and hoe it leak the information of the outsourced data. 1) By exposing the vertex identifiers and distance values of the un-encrypted data by the index. The storage provider can compute the shortest distance between any two vertices of the graph, so that cloud storage provider will comes to know about the data stored. 2) Even the number of vertices can be exposed by the number of labels in the index. 3) The index expose the length of the label. 4) Through the process of querying cloud storage will comes to know which vertices are being queried by the data owner. However the data owner get compromised in the user privacy, the un-encrypted data are not secured. The existing system faces security issues, privacy issues, low efficient and further the data accuracy. The method named graphic encryption cloud storage (GRECS) and 2-HCL is implemented in previous work. While executing the search queries 2-HCL may import more errors. And even the existing scheme receive negative distance for the shortest distance query search.

In this proposed scheme a novel 2-HCL based graph encryption scheme is introduced. The main aim of this paper is to encrypt the graph data and index each data and store in cloud storage provider. So that information about the encrypted data will not get exposed to cloud storage provider. This new scheme reduce the storage space, perform well and the data accuracy also assured.

The proposed scheme contribution is summarized below:

1) The issue of this paper shortest distance queries on encrypted graph data is focused by approaching the 2-HCL based graph encryption scheme for shortest distance queries. The data accuracy is assured in this new scheme.
2) To perform the approximate shortest distance is performed by the order preserving encryption (OPE).
3) With the help of OPE the GENOA is proposed and as we mentioned above the accuracy is assured and it is a main working result of 2-HCL. Then by using 2-HCL scheme it makes the proposed system highly efficient.
4) By encrypting the graph data it help to avoid the data leakage to the cloud storage provider.
5) In this paper user privacy is also declared by proposed scheme.

II. RELATED WORK

This related work section will be followed by the graph data encryption and indexing methods used in the previous papers. The study of querying encrypted data is first lectured to support the keyword search on textual data. In [6] the proposed scheme is Searchable Symmetric Encryption (SSE), this scheme is used for encrypted data search with key word. The data will encrypted then stored and retrieved by the user. This SSE assure the security [7], it gave an efficient system. In [8] Melissa Chase and Seny Kamara generalized the SSE concept and they performed SSE in verity of data types. It adapt to other structure data types including graphs. How to perform private queries on the encrypted data. But however the framework is designed to work on the private queries and perform efficiently, there are more complex data that cannot be searched and it need to extend some search encryption schemes. The security framework use the exposed function to find what data is exposed during the execution. So the scheme which can provide the exposed function while execution can guarantee the security. In previous work they didn't mentioned about the authentication for data owner. In recent years the querying on encrypted graph data is currently in movement. In [9] the filter and verification principle is introduced to reduce the times of 0checking subgraph isomorphism and to prune too many negative encrypted graph data. This feature based index is developed to provide information about the features of the encrypted graph data. In [10] the shortest distance query is learned from this, 2-HCL indexing based somewhat homomorphic encryption to permit the distance execution. Nevertheless somewhat homomorphic encryption cannot find the minimum distance during the shortest distance search, so somewhat homomorphic encryption cannot perform in multiple path. So the accuracy is not assured in somewhat homomorphic encryption. In previous works the authors used different security models, some might secure the neighbourhood information but do not consider other partial information of the encrypted graph data. And in other shortest distance queries scheme might hide the search query data but leak the data in graph data. And these previous works are not reported clearly about the concept of the data outsourcing. There is a method based on the distance labelling to vertices method that used in [11] previous work, that was totally new concept and the drawbacks like scalability also overcame. Though that algorithm it is simple, pruning method that amazingly decrease the search space and the labels, result in fast pre-processing time, small index size and query time is efficient. To improve the performance another parallel pruned labelling method is implemented. The pre-processing work is based on the breadth-first search (BFS).

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

III. PROBLEM DEFINITION

In this section it explains about the challenges of this paper. As we know that shortest distance on outsourced encrypted graph data faces so many problems. Shortest distance queries are most basic graph data operations and it has wide range of applications. As a main problem is the data leakage, the information that stored in the cloud service provider is get leaked by the less security. In [3] in the process of finding the minimum number of points between the two vertices *u* and*v.* And they analyzed the problem generated by the NP-hard and the set cover framework is used to overcome the problem. As we know the common problem is security, privacy, data leakage, accuracy and low efficient. Labelling with minimum cost is also a problem.

IV. GROUNDWORKS

In this section it gives the brief explanation about the groundwork done for this paper. The groundwork mean a preprocessing or preliminary work that done for the project.

A. 2HCL-based Graph Encryption

2CHL – based graph encryption algorithm it combine the five algorithms (Init; Enc; Token; Dist; Dec) and form the 2HCL – based graph encryption algorithm. The working function will be as follows:

1) Init = Initialization $(SK,\{L_v\}v\in A) \leftarrow Init(\lambda,G)$
This initialization algorithm takes an input, security parameter λ and graph A. This gives an output a secret key SK and a 2HCL index $\{L_v\}$ $v\in A$.

2) Ence = Encryption $I \leftarrow Enc(SK,\{L_v\}v\in A)$
This encryption algorithm takes input secret key SK and a 2HCL index $\{L_v\}v\in A$. Then it gives the output as an encrypted index I.

3) Token = Token generation $\tau_{s,t} \leftarrow$ Token (SK, s, t)
This token generation algorithm takes two vertices (s, t) and secret key SK as an input and gives the token $\tau_{s,t}$ as output.

4) Dist = Distance $D_{s,t} \leftarrow Dist(I, \tau_{s,t})$
This distance algorithm takes input as an index I and token $\tau_{s,t}$. It output the an encrypted distance $D_{s,t.}$

5) Dec = Decryption $d_{s,t} \leftarrow Dec(SK, D_{s,t})$
This decryption algorithm takes input as a secret key SK and encrypted distance $D_{s,t}$. Then gives the output a decrypted distance $d_{s,t}$.

As we know that shortest distance query on encrypted graph data have data owner CL whose data is stored in the cloud service provider CSP and encrypt the data and outsourced by the data owner CL. This process includes Setup and Query part.

6) Setup: Data owner CL get $(SK,\{L_v\}v\in A) \leftarrow Init(\lambda,G)$ and produce an index $I \leftarrow Enc(SK,\{L_v\}v\in A)$. Toc conclude CL send the Index I to cloud service provider CSP. Then CSP store the index.

7) Query: Now in this part to query the short distance amid two vertices s, $t \in A$, now U get the token $\tau_{s,t} \leftarrow$ Token (SK, s, t) and sent that token to CSP. Cloud service provider CSP compute the distance $D_{s,t} \leftarrow Dist(I, \tau_{s,t})$ and send that distance to the U. And as a conclusion the data owner decrypt $d_{s,t} \leftarrow Dec(SK, D_{s,t})$ the distance.

V. PROPOSED MODEL

In this proposed model section the total computation process for the shortest queries distance on encrypted graph data is explained.

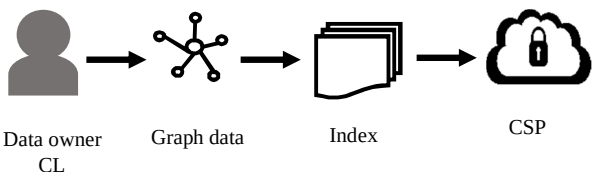


Fig 5.1 Setup model.

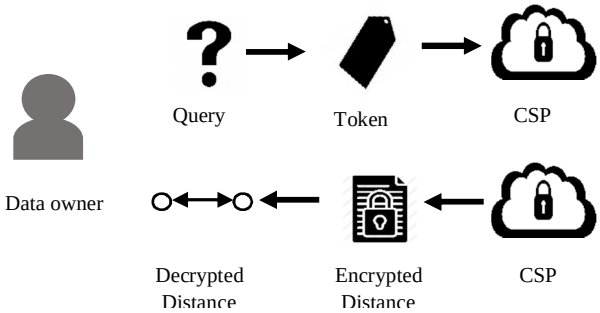


Fig 5.2 Query model.

This Fig 5.1 explain about the Setup model, in that model data owner store the graph data with index in the cloud service provider (CSP). This is an initiative process for our proposed model. Then in Fig 5.2 the process of query model is explained. In the query model data owner CL send a query and generate the token, then get the encrypted distance from the CSP and decrypt the distance.
How to encrypt the distance values within index so that the shortest distance execution a performed on the encrypted values is explained in details in this section. In process of finding the shortest distance the min and max values is essential. In previous work the somewhat homomorphic encryption (SWHE) is used to do this shortest distance calculation. Meanwhile the encryption is not order preserving the 'min and 'max' are calculated to find the approximate shortest distance. It has drawbacks like it produce additional errors, it fail to provide accurate result, it is not friendly consume more space and make the process tedious. The proposed concept is came from comparing the sum of two values with the sum of two values. It help to answer the shortest distance queries without knowing the correct distance. To validate the concept of comparing the values the order preserving encryption scheme is developed. This scheme have two encoding modus. Order Encoding (OE) and Interval Encoding (IE). OE technique it takes the input as a set of values and the gives the result as the set of ordered values in ascending order. IE technique takes the input as the min and max values and the division parameter it will be an integer. Then it divide the intervals and sub intervals also. It gives the output as the order number of the sub intervals which have the value.

Consider set of all the distance values in 2HCL index as B and the maximum and minimum values are considered as d_{min}, d_{max}. Then let symmetric key encryption be SKE. This is the working process of OPE

1) *Enc (SK, d):* It take the input values secret key SK and distance value d ∈ B to be encrypted. The output will be encryption, order, intervals which is encryption ← SKE.Enc (SK,d), order ← OE(B,d), interval ← IE (d_{min}, d_{max}, k, d).
2) *Dec (SK, EV):* It takes the input as secret key SK and encrypted values EV. It analyse the encrypted value EV by way of (encryption, order, interval) and give the output as SKE.Dec (SK, encryption).

Our proposed scheme graph encryption is presented through the OPE scheme block. OPE help to prevent the leakage of the information. Without knowing the exact distance value it answer the shortest distance queries. As we previously learned about the concept of the setup and query, in this section the main graph encryption scheme GENOA is explained. There are five algorithm which is combined with this GENOA. Initialization, Encryption, Token, Distance and Decryption.

A. Initialization

In this algorithm it takes the input as a random secret key SK and it create 2HCL label LAB.label.

Algorithm 1: Initialization (Init)
Input: (λ,G): A security parameter λ and a graph G.
Output: (SK,$\{L_v\}_{v\in A}$) a secret key SK and set of labels.
1 SK ← $\{0,1\}^{\lambda}$
2 $\{L_v\}_{v\in A}$ LAB.label (A)
3 Return (SK, $\{L_v\}_{v\in A)}$

Algorithm 2: Encryption (Enc)

Input: SK,$\{L_v\}_{v\in A}$) a secret key SK and set of labels.
Output: I an index
1 Initialize a dictionary I
2 for $v \in A$ do
3 T_v f (SK, v ||1)
4 K_v f (SK, v ||2)
5 Initialize a counter c = 0
6 for (u, $d_{u,v}$) $\in L_v$ do
7 U f (SK, u ||0)
8 $D_{u,v}$ OPE.Enc(SK, $d_{u,v}$)
9 $T_{u,v} \leftarrow f(T_v, c)$
10 $SK_{u,v}$ g (SK_v, c)
11 $C_{u,v}$ = $SK_{u,v} \oplus$ (U|| $D_{u,v}$. ord|| $D_{u,v.}$ *itv*)
12 $I[T_{u,v}] = (C_{u,v}, D_{u,v.enc})$
13 c=c+1

14 return I

B. Encryption

This encryption scheme is based on the OPE concept. In this label encryption additional changes have done in this proposed scheme. Two methods named Index obfuscation and 2-Round encryption have performed to prevent the information leakage. The index obfuscation is used for hiding the length of the each vertices. This hiding method is from the [4] the scheme hide the data files which have the specific keyword, by hiding that the data will get secured from the data leakage. Proposed scheme encryption generate sub-tokens and store each in index. 2-round encryption prevent the leakage of information during querying. In the process of querying commonly the process need to revel the vertices of the labels, order information of the distance values that stored in the labels.

C. Token

For two vertices the query token is used to label the tokens which used for compute the sub-tokens and the per-vertex key which are used to compute the sub-keys.

Algorithm 3: Token (Token)

Input: (SK, s, t) a secret key SK and two vertices s, t.
Output: $\tau_{s,t}$ a Token
1 T_s f (SK, s ||1), SK_s f (SK, s ||2)
2 T_t f (SK, t ||1), SK_t f (SK, t ||2)
3 return $\tau_{s,t}$ (T_s, T_t , SK_s, SK_t)

D. Distance

After receiving the tokens the cloud service provider compute the sub-tokens first and increase the counters. Then get the 2-round encrypted pairs. Cloud service provider compute the sub-keys through increasing counters, then decrypt the 2-round encryptions and repack the label. As the conclusion the cloud service provider will perform the filter and select to get the conclusion result.

Algorithm 4: Distance (Dist)

Input: $(I, \tau_{s,t})$ an index I and token $\tau_{s,t}$.
Output: $D_{s,t}$ an encrypted result
1 Parse $\tau_{s,t}$ as (T_s, T_t , SK_s, SK_t)
2 Initialize two set L_s, L_t
3 for c=0 until I $[T_{u,s}] = \perp$ do
4 $T_{u,s} \leftarrow f(T_s,c)$
5 $K_{u,s}$ $g(K_s, c)$
6 $(C_{u,s}, D_{u,s}.enc)$ I$[T_{u,s}]$
7 U || $D_{u,s}$.*ord*) || $D_{u,s}.itv = K_{u,s} \oplus C_{u,s}$
8 $D_{u,s}$ $(D_{u,s}.\ enc,\ D_{u,s}\ .ord,\ D_{u,s}.itv)$
9 Insert (U, $D_{u,s}$) into L_s
10 for c=0 until I$[T_{u,t}] = \perp$ do
11 $T_{u,t} \leftarrow f$ (Tt,c)
12 $K_{u,t}$ $g(K_t, c)$
13 $(C_{u,t}, D_{u,t}.enc)$ I$[T_{u,t}]$
14 U || $D_{u,t}$.*ord*) || $D_{u,t}.itv = K_{u,t} \oplus C_{u,t}$
15 $D_{u,t}$ $(D_{u,t}.\ enc,\ D_{u,t}\ .ord,\ D_{u,t}.itv)$
16 Insert (U, $D_{u,t}$) into L_t
17 Initialize a set C
18 Insert all (D1,D2) into C where (U1,D1) $\in L_s$,
(U2;D2) $\in L_t$ and U1 = U2
19 F Filter (C)
20 $D_{s,t}$ Select (F)
21 Return $D_{s,t}$

Algorithm 5: Dec

Input: SK, $D_{s,t}$ a secret key and encrypted distance
Output: $d_{s,t}$ a distance value
1 Parse $D_{s,t}$ as (D_1,D_2)
2 $d_1 \leftarrow$ OPE.Dec (SK, D_1)
3 $d_2 \leftarrow$ OPE.Dec (SK, D_2)
4 Return $d_{s,t} = d_1 + d_2$

E. Decryption

After receiving the encrypted result by executing the distance algorithm (Algorithm 4) it will decrypted by the OPE.Dec. Now the data owner CL will get the decrypted approximate shortest distance as the result.

VI. CONCLUSION

In this paper, we proposed a novel 2-hop cover labelling based graph encryption for shortest distance query. This proposed scheme overcame the limitation of the previous work. Accurate result is assured in this proposed scheme. Then the space consumption is low when compared to the existing system. There is no information leakage, the data that stored will be secured. It is user friendly the privacy of the user is also assured. This GENOA scheme help to improve the performance and made this system flexible.

For future work we can learn deeply about different data leakage issues. Then for some queries like very-higher type of queries the scheme must be different from this concept and the study also need to be done in different viewpoint.

REFERENCES

[1] Christoph Bo¨Sch, Pieter Hartel, Willem Jonker, and Andreas Peter, University of Twente, the Netherland “A Survey of Provably Secure Searchable Encryption”. ACM Computing Surveys, Vol. 47, No. 2, Article 18, Publication date: August 2014

[2] Rachit Agarwal, P. Brighten Godfrey, and Sariel Har-Peled University of Illinois at Urbana-Champaign, USA. “Approximate Distance Queries and Compact Routing in Sparse Graphs”. Published at IEE INFOCOM 2011.

[3] R. Jin, N. Ruan, Y. Xiang, and V. Lee. A highway-centric labelling approach for answering distance queries on large sparse graphs. In ACM SIGMOD, pages 445–456, 2012.

[4] D. Cash, J. Jaeger, S. Jarecki, C. Jutla, H. Krawczyk, M.-C. Rosu, and M. Steiner. Dynamic searchable encryption in very large databases: Data structures and implementation. In NDSS, 2014.

[5] J. Gao, J. X. Yu, R. Jin, J. Zhou, T. Wang, and D. Yang. Neighborhood-privacy protected shortest distance computing in cloud. In ACM SIGMOD, pages 409–420, 2011.

[6] S. Kamara, C. Papamanthou, and T. Roeder. Dynamic searchable symmetric encryption. In ACM CCS, pages 965–976, 2012.

[7] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky. Searchable symmetric encryption: Improved definitions and efficient constructions. In ACM CCS, pages 79–88, 2006.

[8] Melissa Chase from Microsoft research and Seny Kamara from Microsoft research. “Structured Encryption and Controlled Disclosure”. December 2010.

[9] N. Cao, Z. Yang, C. Wang, K. Ren, and W. Lou. Privacy-preserving query over encrypted graph-structured data in cloud computing. In IEEE ICDCS, pages 393–402, 2011.

[10] X. Meng, S. Kamara, K. Nissim, and G. Kollios. Grecs: Graph encryption for approximate shortest distance queries. In ACM CCS, pages 505–517, 2015.

[11] T. Akiba, Y. Iwata, and Y. Yoshida. Fast exact shortest-path distance queries on large networks by pruned landmark labeling. In ACM SIGMOD, pages 349–360, 2013.

[12] Jiyi W. U, Lingdi PING, Xiaoping G. E, Ya Wang, Jianqing F. U. “Cloud Storage as the Infrastructure of Cloud Computing”. Published in 2010 International Conference on Intelligent Computing and Cognitive Informatics.

[13] Muhammad Naveed, Manoj Prabhakaran, Carl A. Gunter, University of Illinois at Urbana-Champaign. “Dynamic Searchable Encryption via Blind Storage”. Published in: 2014 IEEE Symposium on Security and Privacy.

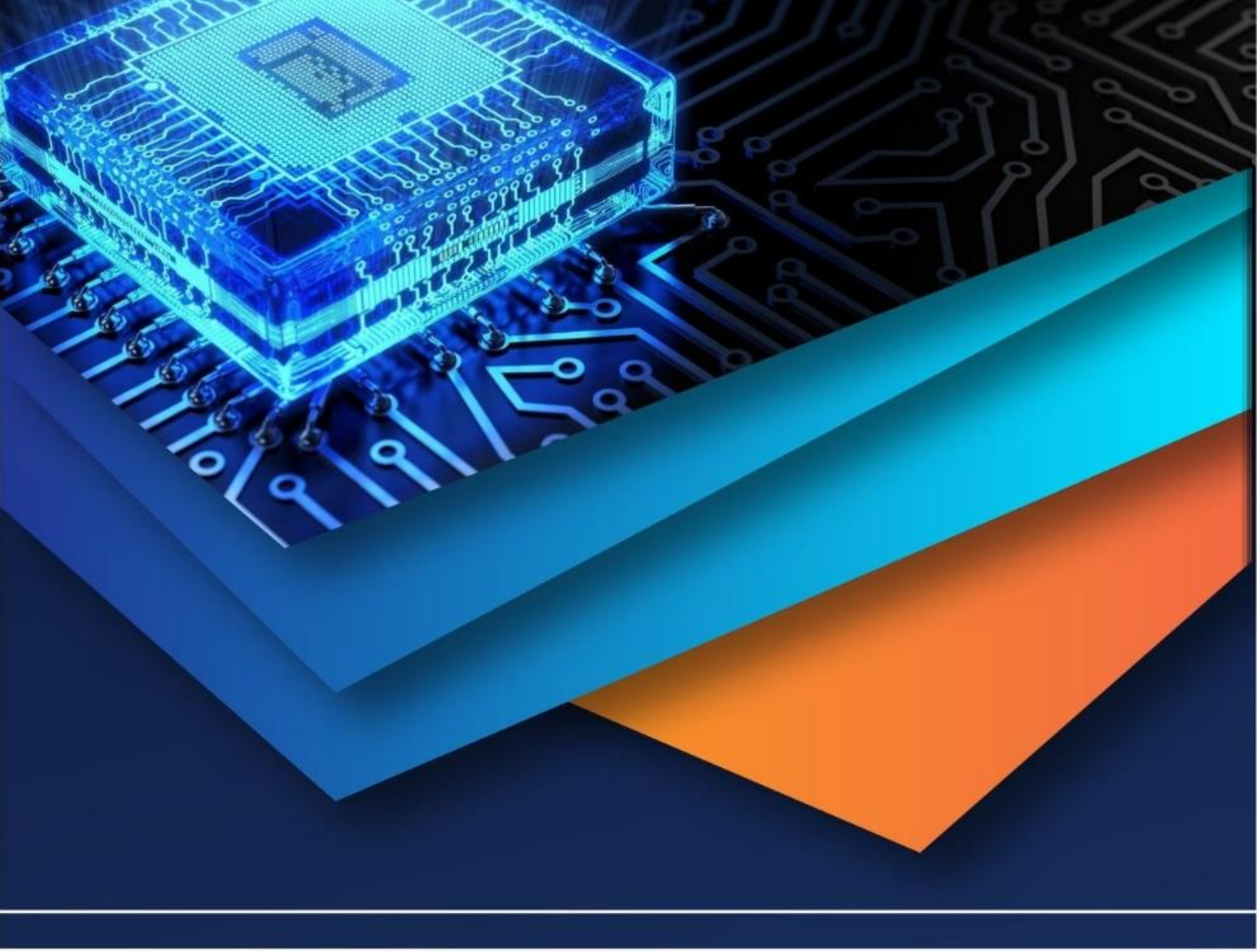
[14] Monique V. Vieira, Bruno M. Fonseca, Rodrigo Damazio. Google Engineering, Belo Horizonte Brazil. Computer Science Department, Federal University of Minas Gerais, Belo Horizonte, Brazil. “Efficient Search Ranking in Social Networks”. 2007 ACM.

[15] David Cash, Stanislaw Jarecki, Charanjit Jutla, Hugo Krawczyk, Marcel-C˘at˘alin Ro¸ su, and Michael Steiner. Rutgers University. “Highly-Scalable Searchable Symmetric Encryption with Support for Boolean Queries”. CRYPTO 2013, Part I, LNCS 8042, pp. 353–373, 2013.

[16] NellyGordilloa, EduardMontseny and PilarSobrevilla, “State of the art survey on MRI brain tumor segmentation”, Magnetic Resonance Imaging Volume 31, Issue 8, and Year: October 2013, Pages 1426-1438.

[17] Ken C. K. Lee Wang-Chien Lee, Department of Computer Science and Engineering, Pennsylvania State University, USA. Hong Va Leong, Department of Computing, The Hong Kong Polytechnic, University, Hong Kong and Baihua Zheng, School of Information Systems, Singapore Management, University, Singapore. “Navigational Path Privacy Protection”. Copyright 2009 ACM 978-1-60558-512-3/09/11.

[18] Bin Zhou, Jian Pei, School of computing science, Simon Fraser University, 8888 University drive, Burnaby, B.B, V58A1S6 Canada. “Preserving privacy in social networks against neighbour attacks” 2008 IEEE.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY