iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** VII **Month of publication:** July 2015

DOI:

www.ijraset.com

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

Volume 3 Issue VII, July 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Flexible Ranking Efficient Information Retrieval for Ranked Query in Cloud

P.Sreekanth[1], M.Thanigavel[2], Y.Prathibha Bharathi[3], T.Sujilatha[4]
CSE Dept, GKCE Sullurpeta, Nellore, A.P, India

Abstract:-Efficient Query Processing using EIRQ Scheme on Cloud Data In a cost efficient cloud environment, a user can accept a certain degree of delay while retrieving information from the cloud to reduce costs. But there are two fundamental issues in such an environment in information retrieving: privacy and efficiency. This paper studies private keyword-based file retrieval scheme that was originally proposed by Ostrovsky to address the aforementioned issues. The scheme allows a user to retrieve files of interest from a un trusted server without leaking any information. The main drawback is that it will cause a heavy querying overhead introduces on the cloud, and thus goes against the original intention of cost efficiency in clouds. To enhance the privacy and efficiency in query processing this paper presents a schema called as efficient information retrieval for ranked query (EIRQ) based on an aggregation and distribution layer (ADL), to reduce querying overhead introduced on the cloud.

Keywords: DQS - differential query services, EIRQ- Efficient Information retrieval for Ranked Query, ADL-Aggregation and Distribution Layer.

I. INTRODUCTION

CLOUD computing an emerging technology is expected to reshape information technology processes in the near future. Due to the overwhelming merits of cloud computing, For Example., cost-effectiveness, flexibility and scalability, more and more organizations choose to out-source their data for sharing in the cloud system. As a typical cloud application, an organization subscribes the cloud services and authorizes its staff to share to files in the cloud. Each file is described by a set of keywords, and the staff, as authorized users, can retrieve files of their interests by querying the cloud system with certain keywords. In such an environment, how to protect user privacy from the cloud system, which is a third party outside the security boundary of the company problem.

Private searching was proposed by Ostrovsky which allows a user to retrieve files of interest from a un trusted server without leaking any data. The Ostrovsky scheme has a high level cost, since it requires the cloud to process the query every file in a collected. The cloud will be learning that certain files, without process. We argue that subsequently proposed development, also have the same drawback. Commercial clouds follow a pay-as-you-go model, where the customer is billed for different operations such as bandwidth, CPU time. Solutions that incur excessive computation level and communication costs are unacceptable to consumers.

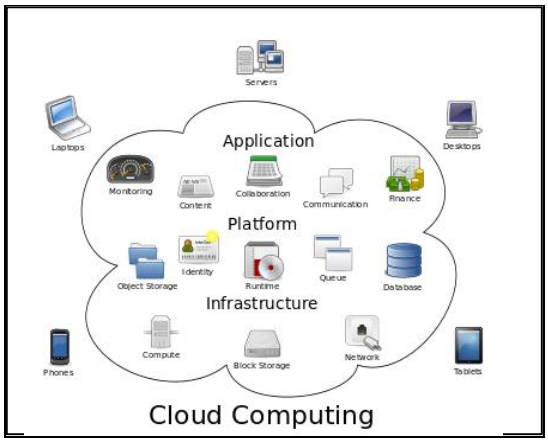


Fig.1 Cloud Computing

A cooperate private searching protocol (COPS), where a proxy server, called the aggregation and distribution layer (ADL).The ADL deployed inside an company (organization) has two main functionalities: aggregating and distributing layers. Reduced the cloud

only needs to execute a combined query once, no matter how many users are executing queries. Most importantly, by using a series of secure functions, COPS can protect user privacy from the ADL, the cloud, and other users. The propose a system, Efficient Information retrieval for Ranked Query (EIRQ), in which each user can choose the rank of his query to determine the percentage of matched files to be come back. The basic idea of EIRQ is to construct a privacy-preserving mask matrix that allows the cloud to filter out a certain percentage of matched files before returning to the aggregation and distribution layer.

The first extension emphasizes simplicity by requiring the least amount of modifications from the Ostrovsky scheme, and the second extension emphasizes privacy by leaking the least amount of information to the cloud. Our key contributions are as following:

A. We propose three EIRQ schemes based on the ADL to provide a cost-efficient solution for private searching in cloud computing.

B. The EIRQ schemes can protect user privacy while providing a differential query service that allows each user to retrieve matched files on demand.

C. We provide two solutions to adjust related parameters: 1.Ostrovsky scheme 2.Bloom filters.

D. Extensive experiments were performed using a union of simulations and real cloud deployments to validate our schemes.

II. RELATED WORK

Among various extensions, further reduced the communication cost from solving a set of linear equations to recovery matched files. Their scheme requires the decryption of one more buffer, thus the computation cost is higher than the Ostrovsky scheme. Presented an efficient decoding mechanism which allows the recovery of files that collide in a buffer position. Proposed a recursive extraction mechanism, which requires a buffer of size when files match a user's query. Proposed two new communication-optimal constructions: one uses Reed-Solomon codes and allows for a zero-error, and the other is based on irregular LDPC codes and allows for lower computation cost at the server. The above private searching schemes only support searching for OR of keywords or AND of two sets of keywords. The main drawback of existing private searching schemes is that combined the computation and communication costs grow linearly with the number of users executing question. When applying these schemes to a large-scale cloud environment, querying costs will be large scale extensive. To alleviate the issues, The concept of differential query services in the main difference between this work and it is that we provide two extensions to address different aspects of the problem, and we conduct extensive experiments on a real cloud to verify the effectiveness of the proposed schemes.

III. SYSTEM MODEL

The system mainly based on three organizations: aggregation and distribution layer, Many users, and Cloud.

An explanation we only use a single ADL in this paper, but multiple ADLs can be deployed as mandatory. An ADL is deployed in an organization that authorizes its staff to share data in the cloud. The staff members, as the authorized users, send their queries to the ADL, which will aggregate user queries and send a combined query to the cloud. The cloud processes the combined query on the file collection and returns a buffer that contains all of matched files to the ADL, which will distribute the search results to each user.

To aggregate sufficient queries, the organization may require the ADL to wait for a period of time before running our schemes which may incur a certain querying delay. The supplementary file which is available in the Computer Society Digital Library at we will discuss the computation and communication and costs as well as the querying delay incurred on the ADL.

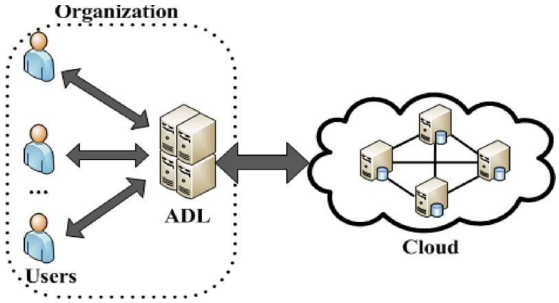


Fig.2 System model

To further less the communication cost, a differential query service is provided by allowing each user to retrieve matched files on demand. Specifically, a user selects a particular rank for his query to determine the percentage of matched files to be returned. This feature is useful when there are a lot of files that same a user's query, but the user only needs a small subset of them.

IV. SECURITY MODEL AND DESIGN MODELS

The ADL is deployed inside the security boundary of an institution, and thus it is assumed to be trusted by all the users. In the supplementary file available online, we will discuss how the EIRQ schemes work without such an assumption.
The ADL obeys our schemes, a user cannot know anything about other users' interests, and thus the cloud is the only attacker (aggressor) in our security model.

A. Cost efficiency

The users can retrieve matched files on demand to further reduce the communication costs incurred on the cloud.

B. User privacy

The cloud cannot know anything about the user's search privacy, access privacy, and least the basic level of rank privacy

V. OVERVIEW OF THE OSTROVSKY SCHEME

The security of the Ostrovsky scheme relies on a public key cryptosystem and derives from the semantic security of the Paillier cryptosystem.

A. Public Key Cryptosystems

Public-key cryptography, also known as asymmetric cryptography, is a class of cryptographic algorithms which requires two separate keys, one of which is secret (or private) and one of which is public. Although different, the two parts of this key pair are mathematically linked. The public key is used to encrypt plaintext or to verify a digital signature.

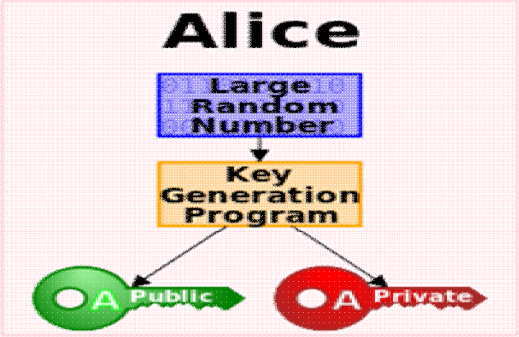


Fig.3: Understanding the large and random number is used to begin generation of an acceptable pair of keys suitable for use by an asymmetric key algorithm.
The Ostrovsky scheme based on three algorithms the working process of which is shown a two assumptions are used in their scheme: A dictionary that consists of the universal keywords is assumed to be publicly available. To better illustrate its working process, we provide an example in the additional file available online.
The Ostrovsky scheme based on three algorithms:
Step1. Generate Query algorithm:
The user runs the ***Generate Query*** algorithm to send an encrypted query to the cloud.
Step2. Private Search algorithm:
The cloud runs the ***Private Search*** algorithm to return an encrypted buffer to the user.
Step3. File Recover algorithm:
The user runs the ***File Recover*** algorithm to recover files. The user decrypts the buffer, entry by entry, to obtain the plaintext c-e

pairs. For the entries in the survival state, file content can be recovered by dividing the plaintext e value by the plaintext c value.

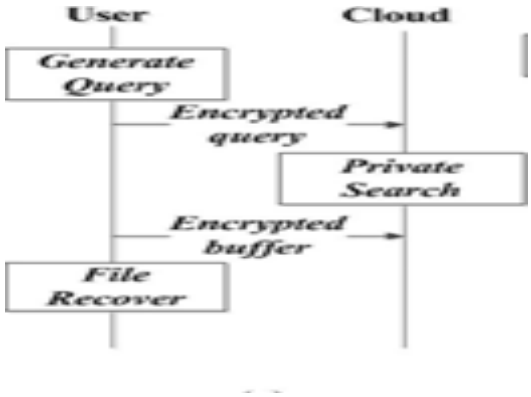
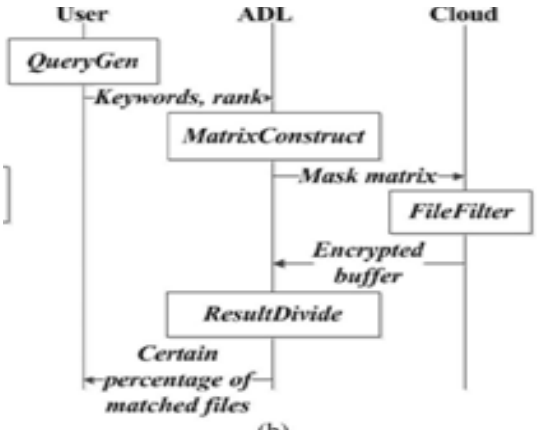


Fig.4 Ostrovsky scheme

VI. EIRQ-EFFICIENT SCHEME

The three EIRQ schemes, we name the original EIRQ scheme as EIRQ Efficient, the first extension as EIRQ-Simple, and second extension as EIRQ-Privacy, in this paper. The basic idea of EIQR-Efficient is to construct a privacy-preserving mask matrix with which the cloud can filter out a certain percentage of matched files before mapping them to a buffer.



EIRQ-Efficient mainly consists of four algorithms: Query Gen and, Result Divide are easily understood, we only provide the details of algorithms Matrix Construct and File Filter.

Algorithm 1 EIRQ-Efficient scheme:

For i=1 to d do

For j=1 to r do

If j<_r -l then

M[i, j]=$E_{(pk)}$

else

M[i, j]=$E_{(pk)=}0$

For i=1 to d do

Map (cj,ej) times to a buffer of size.

Step I. The user runs the *Query Gen* algorithm to send keywords and the rank of the query to the ADL. Since the ADL is assumed to be a trusted third party, this query will be sent without encryption.
Step II. After aggregating enough user queries, the ADL runs the *Matrix Construct* algorithm to send a mask matrix to the cloud.
Step III. The cloud runs the *File Filter* algorithm to return a buffer that contains a certain percentage of matched files to the ADL.
Step IV. The ADL runs the *Result Divide* algorithm to distribute search results to each user. File contents are recovered as the File Recover algorithm in the Ostrovsky scheme.
To allow the ADL to distribute files correctly, we require the cloud to attach keywords to the file content. Thus, the ADL can find out all of the files that match users' queries by executing keyword searches.

VII.EIRQ-SIMPLE SCHEME

The EIRQ-Simple Scheme the main differences lie in the Matrix Construct and File Filter algorithm Algorithm.

Algorithm 2 EIRQ-Simple scheme:
For i=0 to r-1 do
For j=1 to d do
$Q_i[j]=E_{(pk)}(1)$
$q_i = 1-i/r$
End.

The main drawback of EIRQ-simple is that it returns redundant files when there are files satisfying more than one ranked query. An example an interest by Rank-0 and Rank-1 queries, it will be returned twice (in Rank-0 buffer and Rank-1 buffer, respectively), which wastes the network bandwidth. Therefore, the best case scenario is when there are no files of interest to different ranked queries, and the worst case scenario is when queries of different ranks query the same files.

VIII. EIRQ-PRIVACY SCHEME

The working process of EIRQ-Privacy is similar based on EIRQ Efficient scheme process. The main differences lie in the Matrix Construct and File Filter algorithms. Intuitively, EIRQ-Privacy adopts one buffer, with different mapping times for files of different ranks.

IX. PARAMETERSETTING

There two types of the parameter setting:

A. Ostrovsky Parameter Setting: EIRQ-Efficiency filters out a certain percentage of matched files before mapping the buffer, and thus all remaining files should be returned. EIRQ-Efficiency affects one buffer, where the file survival rate is 100 percent. EIRQ-Simple returns multiple buffers with different file survival rates, one for each rank. EIRQ-Privacy still affect one buffer, but with different mapping times for files of different ranks. Therefore, EIRQ Efficient will use EIRQ-Simple, and EIRQ-Privacy and to adjust the parameters under the Ostrovsky parameter setting.

B. Bloom Filter Parameter Setting: The second one is the alternative solution is to use Bloom filter Recall that the file failure rate in EIRQ schemes denotes the probability of a missing file.

X. ANALYSIS

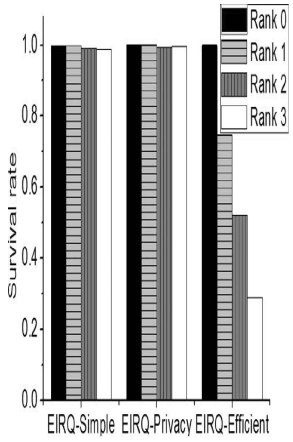
There are two types of EIRQ scheme can provides

A. Security Analysis: Search Privacy, Access Privacy, Rank Privacy.
B. Performance Analysis: Computational cost, Communication cost.

XI. EVALUATION

The compare three EIRQ schemes from the following aspects: file survival rate and computation, communication cost on the cloud.
File Survival Rate: Since queries are classified into0_4 ranks, queries in Rank-0, Rank-1, Rank-2, Rank-3, and Rank-4 should retrieve 100 percent, 75 percent, 50 percent, 25 percent, 0 percent of matched files.

Computational Cost: The computational cost is mainly determined by the number of exponentiations.



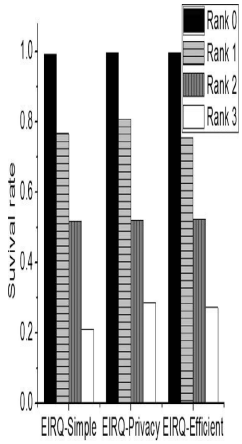


Fig.6 under Ostrovsky setting Fig.7 Bloom filter setting

Performed by the cloud, which is almost the same under the Bloom filter and the Ostrovsky parameter settings. In order to justify the analyses, we will compare the computational cost between No Rank and three EIRQ schemes**.**

Communication Cost: The communication cost mainly depends on the buffer size generated by the cloud, which is calculated in different ways under different parameter settings. Furthermore, the buffer size depends on the number of files that match the queries, which is different when users have different common interests, i.e., the average number of common keywords among user queries.

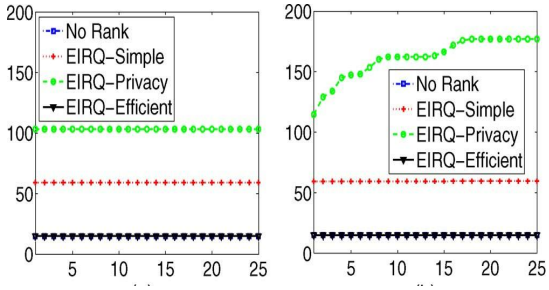


Fig.8 Communication Cost

XII. CONCLUSION

Here by I conclude that this paper proposed three EIRQ schemes based on an ADL to provide differential query services while protecting user privacy. By using our schemes, a user can retrieve different percentages of matched files by specifying queries of different ranks. We simply determine the rank of each file by the highest rank of queries it matches. The flexible ranking mechanism for the EIRQ schemes. In this article, we summarize.

XIII. ACKNOWLEDGEMENT

I like to thank our HOD, PRINCIPAL and OTHER FACULTIES for their valuable comments and helpful suggestions.

REFERENCES

[1] P. Mell and T. Grance, ‘‘The NIST Definition of Cloud Computing (Draft),’’ in NIST Special Publication. Gaithersburg, MD, USA:National Institute of Standards and Technology, 2011.
[2] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, ‘‘Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions,’’ in Proc. ACM CCS, 2006, pp. 79-88.

[3] R.Ostrovsky and W.Skeith, ''Private Searching on StreamingData,'' in Proc. CRYPTO, 2005, pp. 233-240.
[4] Qin Liu, Chiu C. Tan, Member, IEEE, Jie Wu, Fellow, IEEE and Guojun Wang, Member, IEEE "Towards Differential Query Services in Cost-Efficient Clouds," IEEE Transactions, 2014.
[5] P. Paillier, ''Public-Key Cryptosystems Based on Composite Degree Residuosity Classes,'' in Proc. EUROCRYPT, 1999,pp. 223-238.
[6] http://en.wikipedia.org/wiki/Publickey_cryptography
[7] http://en.wikipedia.org/wiki/Paillier_cryptosystem.
[8] R. Ostrovsky and W. Skeith, "Private searching on streaming data," in Proc. of CRYPTO, 2005.
[9] "Private searching on streaming data," Journal of Cryptology, 2007.
[10] G. Danezis and C. Diaz, ''Improving the Decoding Efficiency of Private Search,'' Int'l Assoc. Cryptol. Res., IACR Eprint Archive No. 024, Schloss Dagstuhl, Germany, 2006.
[11] G. Danezis and C. Diaz, ''Space-Efficient Private Search with Applications to Rateless Codes,'' in Proc. Financial Cryptogr. Data Security, 2007, pp. 148-162.
[12] M. Finiasz and K. Ramchandran, ''Private Stream Search at the Same Communication Cost as a Regular Search: Role of LDPC Codes,'' in Proc. IEEE ISIT, 2012, pp. 2556-2560.
[13] http://thanigavelm.blogspot.in/2014/01/latest-technologies-in-computer-science.html
[14] B. Hore, E.-C. Chang, M.H. Diallo, and S. Mehrotra, ''Indexing Encrypted Documents for Supporting Efficient Keyword Search,'' in Proc. Secure Data Manage., 2012, pp. 93-110.
[15] P. Paillier, ''Public-Key Cryptosystems Based on Composite Degree Residuosity Classes,'' in Proc. EUROCRYPT, 1999, pp. 223-238.
[16] X. Yi and E. Bertino, ''Private Searching for Single and Conjunctive Keywords on Streaming Data,'' in Proc. ACM Workshop Privacy Electron. Soc., 2011, pp. 153-158.

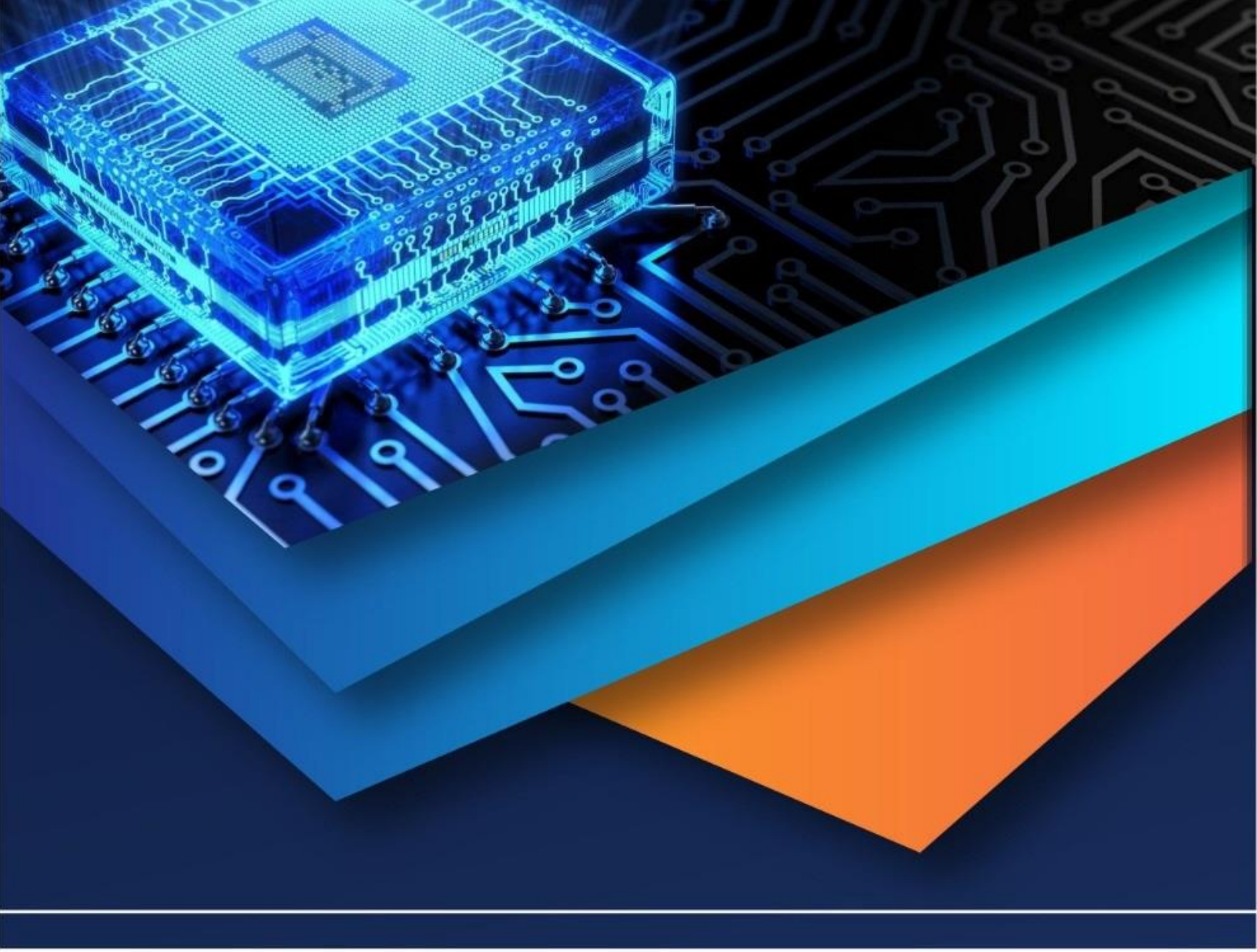
AUTHORS

Mr P.SREEKANTH, Pursuing my M.Tech (CSE) in GOKUL KRISHNA COLLEGE OF ENGINEERING Sullurpeta, and my area of interest is networking and cloud computing, E-mail_id: sreekanthp999@gmail.com

Mr M.THANIGAVEL, Pursuing my M.Tech (CSE) in GOKUL KRISHNA COLLEGE OF ENGINEERING Sullurpeta, and my area of interest is computer networks, distributed, grid, parallel and cloud computing, E-mail_id thaniga10.m@gmail.com

Miss Y.PRATHIBHA BHARATHI., ASST PROFESSOR in the Department of CSE at GOKULA KRISHNA COLLEGE OF ENGINEERING, Sullurpeta, and my area of interest is Cloud computing, OS,&SPM etc, E-mail_id:prathijoshi23@gmail.com.

Miss T.SUJILATHA., ASST PROFESSOR in the Department of CSE at GOKULA KRISHNA COLLEGE OF ENGINEERING, Sullurpeta, and my area of interest is networking; Cloud computing, WSN and DMDW etc, E-mail_id: illusuji@gmail.com.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY