SECURE ENCRYPTION BASED CLOUD DATA EXPLORATION BY PRESERVATION ANALYSIS ON ONE-TO-MANY ORDER

[1] Koukuntla Kalyani, [2]Moligi Sangeetha, [3]Dr. Bhaludra Raveendranadh Singh

[1] Pursuing M.tech (CSE), [2]Assoc. Professor (CSE), [3] Professor & Principal

VISVESVARAYA COLLEGE OF ENGINEERING AND TECHNOLOGY,M.P Patelguda, Ibrahimpatnam (M),Ranga Reddy (D) , Telangana,INDIA.

ABSTRACT

For rank based search in encrypted cloud data, orderPreserving encryption (OPE) is a proficient instrument to scrambleimportance scores of the inverted index. At the point when utilizing deterministic OPE, the cipher texts will uncover the conveyance of pertinencescores. Along these lines, Wang et al proposed a probabilisticOPE, called One-to-Many OPE, for usage of searchableencryption, which can smooth the distribution of the plain texts. In this paper, we proposed a differential assault on One-to-ManyOPE by abusing the distinctions of the requested ciphertexts.The trial results demonstrate that the cloud server can geta decent gauge of the conveyance of importance scores by adifferential assault. Moreover, while having some foundationdata on the outsourced reports, the cloud server canprecisely surmise the encoded watchwords by utilizing the evaluateddisseminations.

I. INTRODUCTION

These days clients associated with the Internet may store their information on cloud servers and let the servers oversee or prepare their information. They can appreciate advantageous and productive administration without paying an excessive amount of cash and vitality, as one of the most alluring component ofcloud computing is its low price. However, regardless of how worthwhile cloud computing may sound, substantial number of individuals still stress over the security of this technology. On the off chance that cloud server get immediate access to every one of these clients' information, it might attempt to examine the archives to get private data. The underlying reason for this activity might be caring. The server wants to provide better service bydigging into these contents and then displaying customer-oriented advertisement, which could be helpful additionally irritating. Furthermore, when we consider delicate information, for example, individual health records and mystery chemical fixings, the circumstance turns out to be considerably more genuine. Hypothetically, the server is shouldn't have entry to touchy information by any stretch of the imagination; in this way we ought to guarantee the server has no entrance to releasing these information to an untrusted outsider. Consequently, touchy information needs to be encrypted before being outsourced to a business open cloud. However, encryption on private data presents obstacles to the handling of the information. Data recovery gets to be troublesome in the scrambled area in light of the fact that the measure of outsourced documents can be vast and customary examples cannot be sent to cipher text recovery straightforwardly. Users need to download every one of the data, decode it all, and after that searching watchwords like plain text recovery. To overcome this, Searchable Encryption (SE) was proposed to make question in the encoded area conceivable while as yet protecting clients' security. There are a few issues in searchable encryption: fuzzy hunt, ranked seek, multi-keyword inquiry thus on. D. Tune et al. initially proposed a hunt plot as it were supporting single Boolean keyword look. After that a lot of searchable encryption techniques emerged to move forward proficiency and diminish communication overload. Applying order preserving encryption (OPE) is one practical method for supporting quick ranked search . This algorithm was initially proposed in 2004 to take care of encrypted query issues in database frameworks. OPE is a symmetric cryptosystem, in this way it is likewise called request saving symmetric encryption (OPSE). The request safeguarding property implies that if the plaintexts x1 < x2, then the comparing cipher texts E(x1) and E(x2) satisfy E(x1) < E(x2).

Boldyreva et al. started the cryptographic investigation of OPE plans in which the characterized the privacy of OPE and proposed a provably secure OPE plan. Be that as it may, the security definition and the developments of OPE in depend on the assumption that OPE is a deterministic encryption plan which implies that a given plaintext will continuously be encoded as a settled cipher text. Be that as it may, deterministic encryption releases the distribution of the plaintexts, so it can't guarantee information security in many applications. For case, in security safeguarding words search, OPE is utilized to encode significance scores in the upset file . As noted by Wang et al. , when utilizing a deterministic OPE, the subsequent cipher text shares the very same dissemination as the importance score, by which the server can determine the catchphrases. In this manner, Wang et al. enhanced the OPE in and guided a One-to-Many OPE in their secure catch phrase search plan, where they attempted to build a probabilistic encryption conspire and disguise the circulation of the plain texts. However, we find that the One-to-Many OPE can't guarantee the normal security. Truth be told, in spite of the fact that the cipher texts of One-to-Many OPE disguises the conveyance of the plaintexts, a opponent may try assess the appropriation from the distinctions of the encrypted texts. So in this paper, we advice a differential secure attack on the One-to-Many OPE.

II. SEARCHING MODEL

2.1 Plaintext Searching Model

Experimental results demonstrate that, while applying this assault to the safe catchphrase seek plan of , the cloud server can get an estimation of the dissemination of the significance scores, what's more, besides precisely uncover the encoded watchwords. Whatever remains of this paper is sorted out as takes after. We first depict the plaintext seek model and cipher text look model in Section II. At that point, in Section III (3rd), the initial OPE, One-to- Numerous OPE, and protection prerequisite in distributed computing are quickly checked on. We expand on differential assault on One to- Numerous OPE and further assault with foundation data of outsourced information in Section IV and Section V separately. At long last the conclusion is given in Section VI.By and by, to acknowledge compelling information recovery on huge sum of archives, it is important to perform significance positioning on the outcomes. Positioned pursuit can likewise fundamentally decrease system activity by sending back just the most pertinent information.

In positioned look, the positioning capacity assumes an essential part in ascertaining the importance amongst documents and the given searchingquestion. The most well-known importance score is characterized in view of the model of TF * IDF, where term recurrence (TF) is the number of times a term (watchword) shows up in a document and backwards report recurrence (IDF) is the proportion of the aggregate number of records to the quantity of documents containing the term. There are numerous varieties of TF IDF-based positioning capacities, and in, the accompanying one is received.

$$Score(wd,Fd)=\frac{1}{|Fd|}.(1+\ln fd,wd).\ln(1+\frac{NOd}{fwd}))$$

Herein, *wd* depicts the keyword and *fd ;wd* denotes the TF of term *wd* in file *Fd*; *No=fw* denotes IDF where *fwd* is the number of files that contain term *w* and *Nod* is the total number of documents in the collection; and *jFdj* is the number of indexedterms containing in file *Fd*, i.e., the length of *Fd*.

2.2 Cipher Text Searching Model

Because of the extraordinary foundation of distributed computing, dissimilar to customary plaintext data recovery, there are for the most part three substances in cloud information recovery as appeared in Fig. 1:

Information proprietor, remote cloud server and clients. An information proprietor can be an individual or an enterprise, i.e., it is the element that possesses a gathering of records Dc = {D1;D2 : :DNd}that it needs to impart to trusted clients. The catchphrase set is set apart as W = {w1, w2…wNw}

For security and protection concerns, records must be encoded into = {E(D1);E(D2) : :E(DNd)} before being transferred to the cloud server. Furthermore, the plaintext list must be encoded into I to avoid data spillage. The scrambled type of the case of the posting rundown of the Altered Index is appeared in TABLE II, in which the catchphrase wi is secured by a Hash capacity hash(), and the importance scores are encoded by an encryption plan E′().We utilize TABLE II as a case to perceive how a cloud server conducts a protected hunt in view of an encoded list. In the seek technique, a client first produces a pursuit demand in a mystery structure — a trapdoor T (fw). In this illustration, the trapdoor is only the hash estimations of the watchword of interest. Once the cloud server gets the trapdoor T (fw), it thinks about it with the hash estimations of all catchphrases in the record , then the craved archives which are comparing to catchphrase w are found. Next, the server gives back the coordinated document IDs: F1, F2, ... , Ffw to the client. At long last, the client can download all the scrambled archives in view of the given IDs what's more, unscramble them. An alluring framework should return the records in a positioned request by their pertinence with thequestioned catchphrase, yet utilizing customary encryption plans will jumble pertinence scores. In this way, in Order Preserving Encryption (OPE) is connected to scramble the importance scores, which empowers the server to rapidly perform positioned seek without knowing the plain relevance scores.

III. OPE VS. ONE-TO-MANY OPE

3.1 OPE

OPE is a symmetric cryptosystem, so it is additionally called request safeguarding symmetric encryption(OPSE). The order preserving property implies that if the plaintexts have such a relationship as x1 < x2, then the comparing cipher texts E(x1) and E(x2) fulfill E(x1) < E(x2).

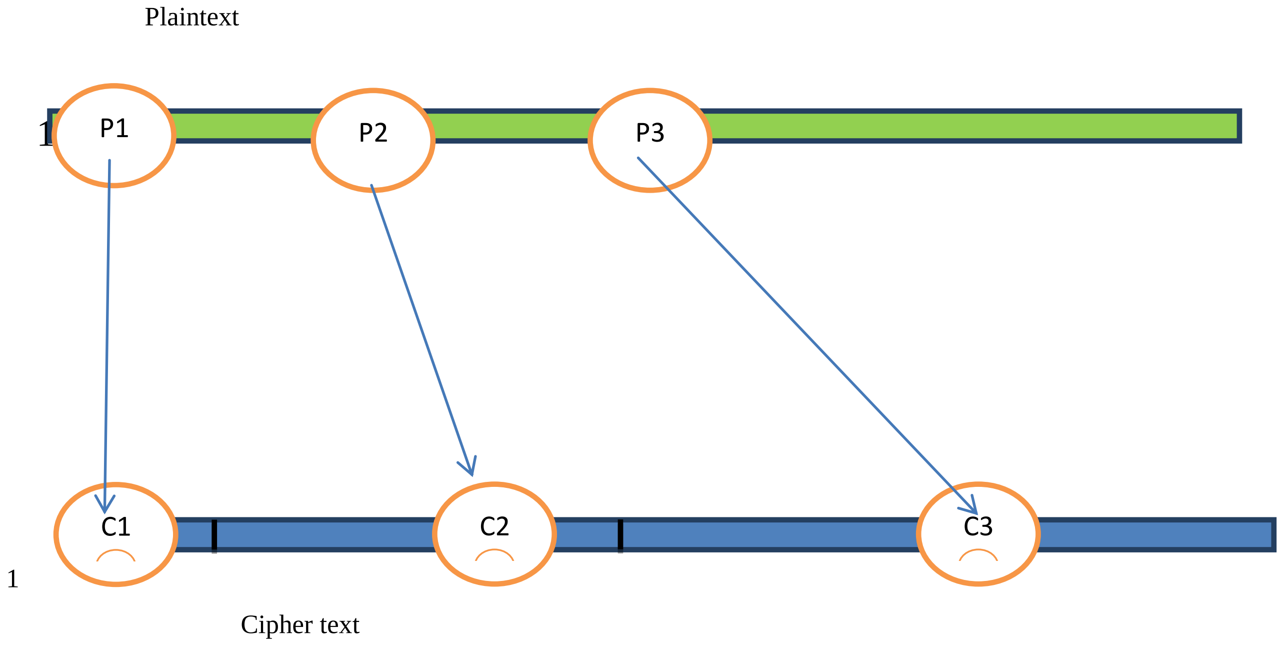
Boldyreva et al.started the cryptographic investigation of OPE plans, and they characterized the security of an OPE plan utilizing the perfect article. Note that any request safeguarding capacity g from space D = f1; 2; _ ;Mg to range R = f1; 2; _ ;Ng can be exceptionally characterized by a blend of M out of N requested things. The perfect item is only a capacity that is haphazardly chosen from all request saving capacities, which is known as an irregular request safeguarding capacity (ROPF). Therefore, with the soul of pseudorandom capacities, an OPE plan is characterized to be secure if the foe can't Recognize the OPE from the ROPF. In , the creators moreover developed an effective OPE plan fulfilling this protected basis. The development depends on the connection between the arbitrary request safeguarding capacity and the hyper-geometric likelihood appropriation (HGD), and a HGD easier is utilized to select a request saving capacity in a pseudorandom way. In the OPE plan of , the reach R is separated into some non overlapping interim cans with arbitrary sizes. The irregular measured pail is controlled by a double inquiry based on an irregular HGD sampler. In, the system of double hunt is depicted as Algorithm 1, where Tape Gen() is air regular coin generator.

Algorithm 1 Binary Search:

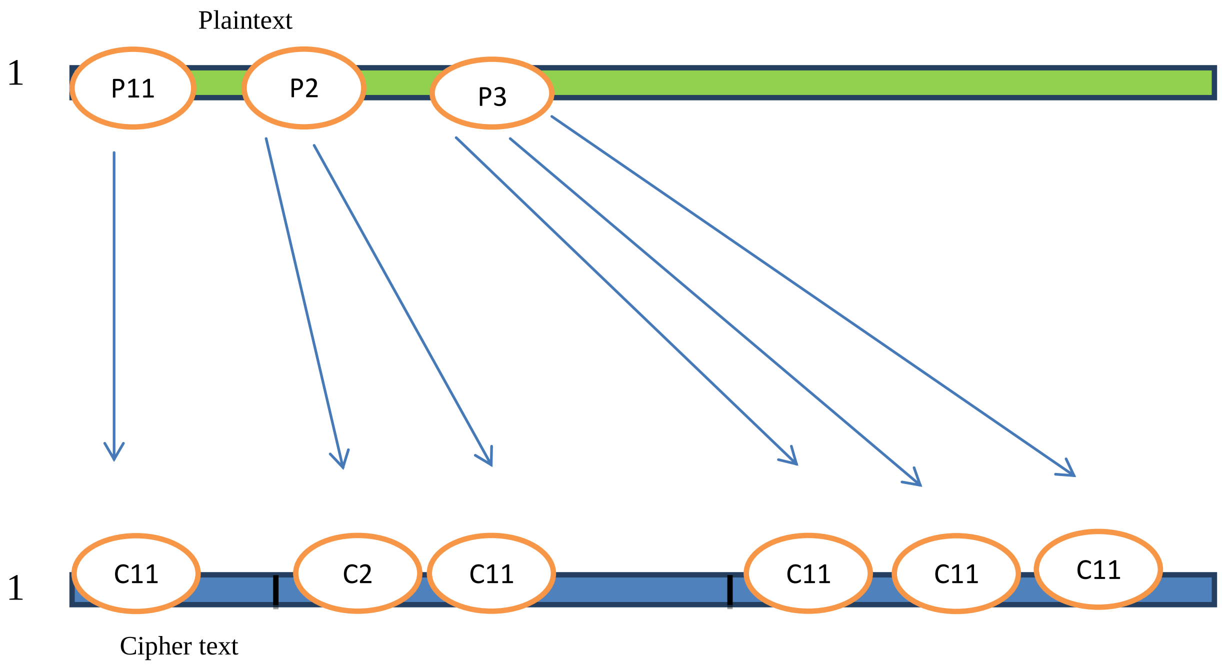
```
Input: $fK, DD, RR, mg$

$M1 <- length(Dd); N1 <- length(Rr)$
$d <- min(Dd) -1; r <- min(Rr) -1$
$y <- r + ceil(N1=2)$
$coin\ Rr \overset{R}{\leftarrow} TapeGen(K; (Dr;Rr; y//0))$
$x\ Rr \overset{R}{\leftarrow} d + HGD(coin;M1;N1; y -r)$
$x = d + f$
if $m \leq x$ then
    $Dd \quad \{d + 1, \ldots\ldots, x\}$
    $Rr \quad \{r + 1; \_\_\_\_\_\_; y\}$
else
    $Dd \quad \{x + 1; \_\_\_; d + M\}$
    $Rr \quad \{y + 1; \_\_\_; r + N\}$
end if
Output: $\{Dd;Rr\}$
```

Wang et al. noticed that, in uses of privacy preserving catchphrase look, if a deterministic OPE is utilized to scramble significance scores, the cipher texts will share precisely the same circulation as its plain partner, by which the server can indicate the watchwords.

Consequently, Wang et al. Adjusted the first OPE to a probabilistic one, called "One-to-Many OPE". For a given plaintext m, i.e., a significance score, the One-to-Many OPE initially utilizes Algorithm 1



(a) Deterministic OPE



(b) One-to-Many OPE

To choose a basin for m, and afterward arbitrarily picks a quality in the pail as the cipher text. The arbitrarily picking system in the pail is seeded by the novel record IDs together with the plaintext m, and in this way the same significance score in the Inverted Index will be encoded as various cipher texts. The encryption procedure of "One-to- Numerous OPE" is depicted in Algorithm. 2 , which is too shown in Fig. 2(b).

Example 1: To think about the scrambled consequences of OPE furthermore, One-to-Many OPE, we take the posting rundown of catchphrase "Test" as illustration that is created from the TREC information. The significance scores are encoded into whole numbers, from which we get the plaintext dispersion appeared in Fig. 3(a). The dispersions of the encoded results acquired by deterministic

OPE and One-to-Many OPE are appeared in Fig. 3(b) and Fig. 3(c) separately.

ALGORITHM 2 ONE-TO-MANY ORDER PRESERVING ENCRYPTION

```
Input: $\{K;Dd;Rr; m; id(F)\}$
while$|D|! = 1$ do
    $\{Dd;Rr\} = binarysearch(K;Dd;Rr;m)$
end while
$coin \xleftarrow{R} TapeGen(K; (Dd;Rr; 1\backslash\backslash m; id(F)))$
$cc \xleftarrow{R} R$
$cc= round(coin)$
Output: cc
```

3.2 Privacy Threat Designs

The revelation behind One-to-Many Order Preserving Encryption and OPE and is to ward of data costly to the cloud server. The cloud server is considered as legitimate, likewise called burning with curiosity. In particular, the cloud server won't endeavor to evacuate scrambled information documents or record from the capacity, and it will likewise effectively take after the outlined convention particular furthermore, execute the strategy steadfastly. In any case, it is interested to handle the put away information and tries to examine the information to learn extra data.

At the point when discussing the “burning with curiosity" model, as a rule there are two assault models Known Cipher text Model furthermore, Known Background Model" . Known Cipher text Model" expect that the cloud server can just access the encoded records and the encoded record. In this model the server can just delve into the cipher texts with no other foundation data, furthermore, consequently security implies that the keywords and records data are entirely ensured and there is no aberrant way to estimate these data. Known Background Model is nearer to the real world circumstance in the cloud application. The cloud server should have more learning than what can be gotten to in the Known Cipher text Model. It might purposefully gather related factual data about the outsourced reports, and with this data the server can surmise more delicate data. Next, we will propose assaulting techniques on One-to-Many OPE under these two risk models separately.

IV. DIFFERENTIAL ATTACK ON ONE-TO-MANY OPE

4.1 Under Known-Ciphertext-Model

It can be seen that One-to-Many OPE has effectively concealed the conveyance of the plaintexts, however the security of One-to-Many OPE has not persevered through strict cryptanalysis. In this area, we will demonstrate that, by dissecting the contrasts between the cipher texts, the cloud server can get an estimation on the dissemination of the plaintextsAs appeared in Fig. 2(b), each plaintext esteem m is mapped into numerous conceivable cipher texts having a place with an altered basin, and the cipher text is haphazardly chosen in the basin. Along these lines, the disperse of cipher texts in a basin will be thick for a plaintext esteem with high recurrence, yet will be scanty for a plaintext esteem with low recurrence. In spite of the fact that the sizes of the basins are arbitrarily decided, the thickness of cipher texts in every basin will change as per the recurrence of the relating plaintext, and along these lines the profile of the plain texts' recurrence can be depicted by the thickness of cipher texts. Note that the thickness of cipher texts can be uncovered by the contrasts between the neighboring cipher texts that we call "differential cipher texts". At the end of the day, in the event that we can find the change purposes of the circulation of the differential cipher texts, we can decide the limits of the cans in the cipher text range R = {1,2,3,4,…..N}. With these limits, the histogram of the plaintexts can be effortlessly evaluated by checking the quantity of cipher texts having a place with every can. In this way, the cloud server may recreate the appropriation of plaintexts from the differential cipher texts, which we call "differential attack". The key issue in "differential attack" is finding the change point in the differential grouping of the cipher texts. There are numerous measurable techniques to acknowledge such Change Point Analysis (CPA), and we utilize the combined aggregate (CUSUM) based CPA to portray the strategy of "differential attack", which comprises of six stages.

1. Sort the scrambled qualities:

Assume that the first cipher text succession is C1, C2, ..., CL. Sort the cipher text succession in rising request, and get Ci1 ≤ Ci2≤…………….≤ .CiL.

2. Generate the differential sequence:

The differential sequence *{di;* 1 ≤*i* ≤*L*-1*}* of the ordered

Cipher texts is obtained by calculating

*d*1 = *ci*2-*ci*1 ,

*d*2 =*ci*3-*ci*2

dL−1 = *ciL* -*ciL*□1 .

3. Generate CUSMU sequence:

To get the CUSUM of *di* (1 ≤*i* ≤*L*-1), we first compute

their average value:

$$D = \frac{1}{l - 1} \sum_{l=1}^{L-1} Di$$

Set the underlying estimation of total as S0 = 0. The other

total qualities are computed in a recursion way such

that

$Si = Si-1 + (di - \overline{d}); i = 1,2,\ldots\ldots L -1:$

The CUSUM is characterized as the total whole of every information short the normal worth, so the last esteem ought to dependably be zero, i.e., SL−1 = 0. A CUSUM graph can be acquired by drawing the aggregate total Si all together for 0 i≤L ≤ 1. On the off chance that there is a time of information which is more prominent than the normal esteem, a climbing bend will happen on the graph; generally, a plunging bend will happen on the graph. A change point on the graph alludes to a sudden change in the bend.

In Fig. 4, we portray the CUSUM diagram of the differential arrangement of cipher texts got by One-to-Many OPE in Example 1, which demonstrates that a change point occurred.

Smax= $\max_{0\leq x \leq 1} Si$

Smin= $\min_{0\leq x \leq 1} Si$

06*i*6*L*−1

Si; (5)

Sdif = *Smax–Smin.*(6)

Generating a bootstrap test of L - 1 units, indicated as p1, p2,…. ; pL−1, by haphazardly reordering the first L - 1 differential qualities dq1,dq2, ….. dqL−1.Finding the CUSUM of the bootstrap sample, denotedas S_0

$S^0_0, S^0_1 \ldots\ldots\ldots. S^0_{L-1}$

$$S^0_1$$

V. IDENTIFYING OTHERCHANGE POINTS

Consider that Step 4 outputs one change point u_1, which thusdivides the sequence ds_1, $ds_2, \ldots\ldots\ldots$ ds_{L-1} into two subsequences:ds_1, ds_2, ... , ds_{v1} and d_{v1+1}, d_{v1+2}, ... , d_{L-1}. Then we takechange point analysis, i.e., Steps 3 and 4, on these two subsequencesrespectively. Assume that change points, p_2 and p_3,are detected in the two subsequences respectively. Obviously,$p_2 < p_1 < p_3$, which can divide the original sequence intofour subsequences. Take change point analysis on these foursubsequences and output eight change points...and so on.

This process will end when we cannot detect change points in any subsequence. Assume that *B* change points in total are found, denoted by p_1; p_2; _ _ _ ; p_B.

5.1 Experimental Results

We demonstrated an exhaustive test on the TREC information, which comprises of 1400 reports and 11000 particular catchphrases, from which we choose Nw = 1055search words of wish list. At the end of the day, the Inverted Index comprises of 1055 postings in our tests. To depict the background work acquired by the cloud server,werandomlyselectasubsetof100α percent of the entire document. The search key words $\overline{w}$ and relating feature created from this subset for 1 ≤ i ≤ Nw. In this,
we utilize as a parameter to portray the comparability of the foundation gained by the cloud server to the outsourced
record gathering. We call the foundation quality.
A huge foundation quality implies that the server has a Appropriation near the genuine conveyance of the significance scores that have been encoded, and the other way around. In this examination, we pick = 0.91, 0.65 and 0.51.

VI. CONCLUSION

In ranked search of encrypted cloud data, probabilistic OPE is needed to preserve the order of relevance scores and conceal their distributions at the same time. One-to-Many OPE is a scheme designed for such a purpose. However, in this paper, we demonstrate that the cloud server can estimate the distribution of relevance scores by change point analysis on the differences of cipher texts of One-to-Many OPE. Furthermore, the cloud server may identify what the encrypted keywords are by using the estimated distributions and some background knowledge.

On the other hand, some methods can be used to resist the proposed attack. One is to improve the One-to-Many OPE itself. For instance, we can divide plaintexts having the same value into several sets and divide the corresponding bucket into several sub-buckets. By mapping each plaintext set into one sub-bucket, some new change points will appear in the differential attack, which will cover up the original distribution of plaintexts. Another possible method is to add noise into the inverted index by adding some dummy documents IDs and keywords, and forging responding relevance scores. In our future work, we will elaborate these ideas to design secure methods of probabilistic OPE and schemes for search in encrypted data.

REFERENCES

[1] P. Mell and T. Grance. (Jan. 2010). Draft NIST Working Definition of Cloud Computing. http://csrc.nist.gov/groups/SNS/cloudcomputing/index.html

[2] S. Subashini and V. Kavitha, “A survey on security issues in service delivery models of cloud computing,”J. Netw. Comput. Appl., vol. 34, no. 1, pp. 1–11, 2011.

[3] B. Krebs. (2009). Payment Processor Breach May Be Largest Ever. [Online]. Available:http://voices.washingtonpost.com/securityfix/2009/01/payment_processor_breach_may_b.html

[4] M. Abdalla et al., “Searchable encryption revisited: Consistency properties, relation to anonymous IBE, and extensions,” in Advances in Cryptology. Berlin, Germany: Springer-Verlag, 2005, pp. 205–222.

[5] D. X. Song, D. Wagner, and A. Perrig,“Practical techniques for searches on encrypted data,” inProc. IEEE Symp. Secur. Privacy, May 2000,pp. 44–55.

[6] E.-J. Goh. (2003). “Secure indexes,” Cryptology ePrint, Tech. Rep. 2003/216. [Online]. Available: http://eprint.iacr.org/ [7] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search,” in Advances in Cryptology. Berlin, Germany: Springer-Verlag, 2004, pp. 506–522.

[8] Y.-C. Chang and M. Mitzenmacher, “Privacy preserving keyword searches on remote encrypted data,” inApplied Cryptography and Network Security. Berlin, Germany: Springer-Verlag, 2005, pp. 442–455.

[9] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, “Searchable symmetric encryption: Improved definitions and efficient constructions,” in Proc. 13th ACM Conf. Comput. Commun. Secur., 2006, pp. 79–88.

[10] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, “Order preserving encryption for numeric data,” inProc. ACM SIGMOD Int. Conf. Manage. Data, 2004, pp. 563–574.

[11] A. Boldyreva, N. Chenette, Y. Lee, and A. O’Neill, “Order-preserving symmetric encryption,” in Advances in Cryptology. Berlin, Germany: Springer-Verlag, 2009, pp. 224–241.

[12] A. Boldyreva, N. Chenette, and A. O’Neill, “Order-preserving encryption revisited: Improved security analysis and alternative solutions,” in Advances in Cryptology. Berlin, Germany: Springer-Verlag, 2011, pp. 578–595.

[13] L. Xiao and I.-L. Yen, “Security analysis for order preserving encryption schemes,” inProc. 46th Annu. Conf. Inf. Sci. Syst., Mar. 2012, pp. 1–6.

[14] A. Swaminathanet al., “Confidentiality-preserving rank-ordered search,” in Proc. ACM Workshop Storage Secur. Survivability, 2007, pp. 7–12.

AUTHOR DETAILS

KOUKUNTLA KALYANI

Pursuing M.Tech in Visvesvaraya College of Engineering and Technology, M.P Patelguda, Ibrahimpatnam (M), Ranga Reddy (D), and India.

MOLIGI SANGEETHA

MRS.M.SANGEETHA COMPLETED Bachelor of Technology from **Swami Ramananda Tirtha Institute of Science & Technology,**Nalgonda and Post Graduation from **JNTU Kakinada Campus**, Kakinada and is having **14+** years of Teaching experience.Working as Assoct. Professor (CSE) in Visvesvaraya College of Engineering and Technology, M.P Patelguda, Ibrahimpatnam (M), Ranga Reddy (D), and India.

SRI. DR. BHALUDRA RAVEENDRANADH SINGH

M.Tech,Ph.D.(CSE),MISTE,MIEEE(USA),MCSI

Professor & Principal. He obtained M.Tech, Ph.D(CSE)., is a young, decent, dynamic Renowned Educationist and Eminent Academician, has overall 23 years of teaching experience in different capacities. He is a life member of CSI, ISTE and also a member of IEEE (USA). For his credit he has more than 50 Research papers published in Inter National and National Journals. He has conducted various seminars, workshops and has participated several National Conferences and International Conferences. He has developed a passion towards building up of young Engineering Scholars and guided more than 300 Scholars at Under Graduate Level and Post Graduate Level. His meticulous planning and sound understanding of administrative issues made him a successful person.