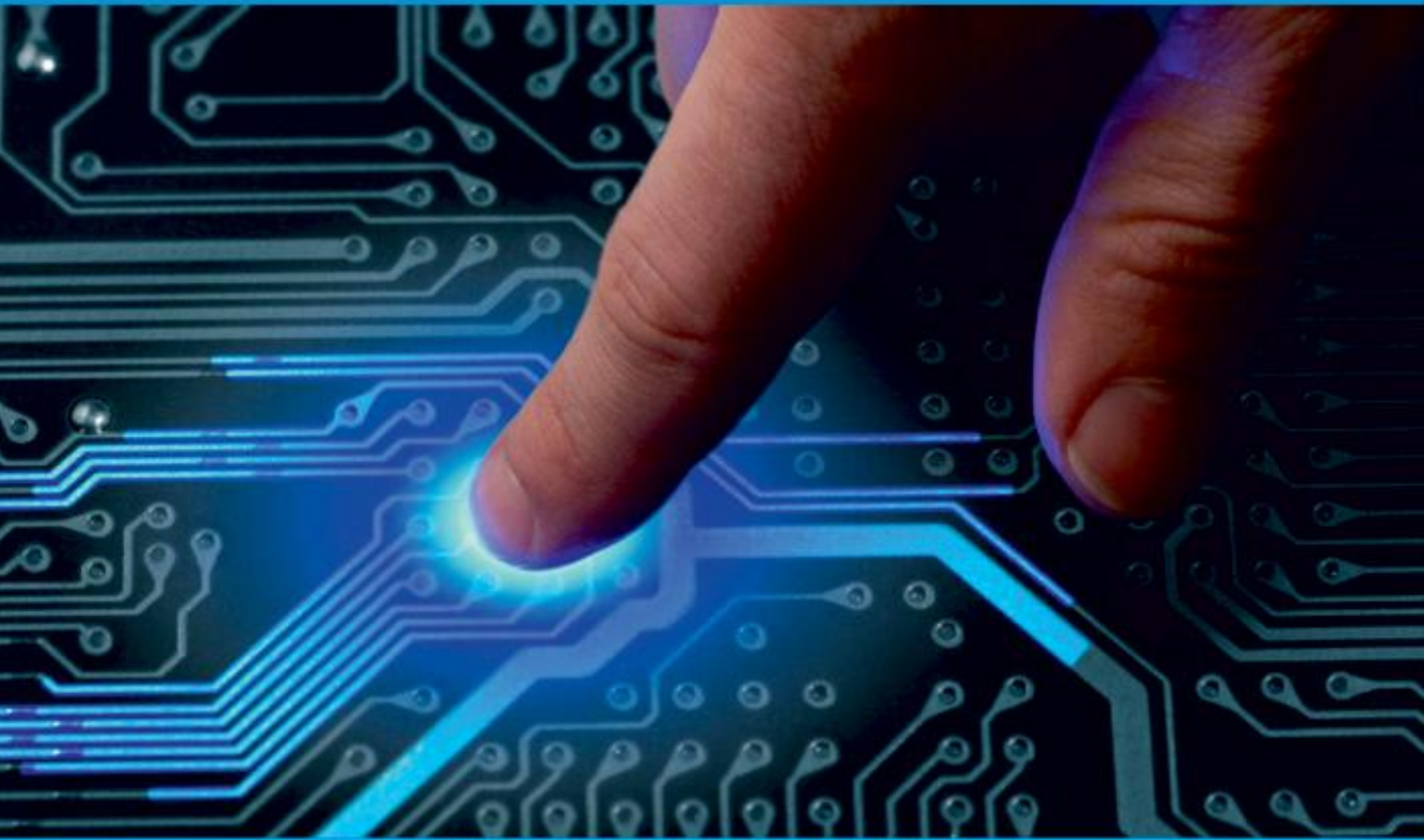
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Special Issue 2, March 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

Network Vulnerability Detection &Text File Integrity checking with forensic evidences on E-Commerce Application Based on OSN Technique

Rajendra Dumbre

Assistant Professor, Department of Computer Engineering, SPCOE Otur ,SavitribaiPhule Pune University, Pune India

ABSTRACT: Cloud computing is popular in the computing paradigm in which data is outsourced to a third-party service provider (server) for data mining. Outsourcing, however, raises a serious security issue: how can the client of weak computational power verify that the server returned correct mining result? In this paper, we focus on the specific task of frequent itemset mining. We consider the server that is potentially untrusted and tries to escape from verification by using its prior knowledge of the outsourced data. We propose efficient probabilistic and deterministic verification approaches to check whether the server has returned correct and complete frequent itemsets. Our probabilistic approach can catch incorrect results with high probability, while our deterministic approach measures the result correctness with 100 percent certainty. We also design efficient verification methods for both cases that the data and the mining setup are updated. We demonstrate the effectiveness and efficiency of our methods using an extensive set of empirical results on real datasets.To resolve issue of untrusted text files we propose file integrity check method by using conventional challenge response protocol.forensic investigation is done by function summary extraction using text mining algorithms and performance measurements and cybercrime prediction analyses with snapshots of server.

KEYWORDS*:* Cloud computing, data mining as a service, security, result integrity verification, SoftwareasServices,Snapshotgeneration,VM.

I. INTRODUCTION

Existing work are the closest to ours. It has been proven that the evidence patterns constructed by the encoding method in can be identified even by an attacker without knowledge of the data. We argue that our probabilistic verification approach is robust against the attack. Our probabilistic approach is more efficient. Shows that itmay take 2 seconds to generate one evidence pattern, while our method only takes 600 seconds to generate 6900 evidence item sets.To Propose an efficient cryptographic approach to verify the result integrity of web-content searching by using the same set intersection verification protocol as ours. It shows that the time spent on the server to construct the proof for a query that involves two terms. Our deterministic approach requires seconds to construct the proof for an item set of length average, which is comparable to theperformance.

Proposed a set intersection verification protocol to verify that E is the correct intersection of protocol.Whether the coefficients are computed correctly by the server. Whether any given accumulation value is indeed calculated from the original dataset. Whether satisfies the subset condition by using whether satisfies the intersection completeness condition. We omit the details of proof construction and verification due to limited space.
To check the integrity of text file challenge-response protocol using conventional methods, which lead to some engineering trade-offs. Forensic is done for server by uploading unstructured cybercrime report to generate structure figures through a machine learning techniques in this proposed system. Following that, the policy should provide a study on the severity and frequency of cyber-crime categorization and resolution. Text mining algorithms, performance metrics, and cybercrime prediction models are used to derive the feature description.[1]

II. RELATED WORK

In [2] Fast algorithms for mining association rules in large databases R.AgrawalandR.Srikant,Authorsconsidertheproblemofdiscoveringassociationrulesbetweenitems in a large database of sales transactions. Authorspresent two new algorithms for solving this problem that are fundamentally different from the known algorithms. Empirical evaluation shows that these algorithms outperform the known algorithms by factors ranging from three for small problems to more than an order of magnitude for large problems. Authorsalso show how the best features of the two proposed algorithms can be combined into a hybrid algorithm, called AprioriHybrid. Scale-up experiments show that prioriHybrid scales linearly with the number of transactions. AprioriHybrid also has excellent scale-up properties with respect to the transaction size and the number of items in the database.[3] L. Babai, L. Fortnow, L. A. Levin, and M. Szegedy,Motivated by Manuel Blum's concept of instance checking, authors consider new, very fast and generic mechanisms of checking computations. Authorsresults exploit recent advances in interactive proof protocols [LFKN92], [Sha92], and especially the M IP = N EXP protocol from

[BFL91].AuthorsshowthateverynondeterministiccomputationaltaskS(x;y),definedasapolynomialtimerelation between the instance x, representing the input and output combined, and the witness y can be modified to a task S 0 such that: (i) the same instances remain accepted; (ii) each instance/witness pair becomes checkable in polylogarithmic Monte Carlo time; and (iii) a witness satisfying S 0 can be computed in polynomial time from a witness satisfying S. Here the instance and the description of S have to be provided in error-correcting code (since the checker will not notice slight changes). A modification of the M IP proof was required to achieve polynomial time in (iii); the earlier technique yields N O(log log N) time only. This result becomes significant if software and hardware reliability are regarded as a considerable cost factor. [4] Privacy-preserving data mining from outsourced databases :-F. Giannotti, L. V. S. Lakshmanan, A. Monreale, D. Pedreschi, and W. Hui Wang, mining task within a corporate privacy-preserving framework. Authorspropose a scheme for privacy-preserving outsourced mining which offers a formal protection against information disclosure, and show that the data owner can recover the correct data mining resultsefficiently.[5]Cloud computing systems illustrates an prototype to the distributed dispensation of digital data. Digital forensic investigations associated with such systems area unit doubtless to involve a lot of complicated digital proof acquisition and analysis. Some public cloud computing systems could embrace the storage and process of digital knowledge in severalcourts, and a few organizations could value more highly to encode their knowledge before getting into the cloud. together with cloud design, these two factors will build rhetorical examination of such systems a lot of complicated and long. There are not any established digital rhetorical tips that specificallyaddress the investigation of cloud computing systems. during this letter we tend to examine the legal aspects of the digital forensic investigation of the cloud computing system.[6]System proposed the cloud automatic data processing system hosts most of today's industrial business applications, which provides it high revenue that makes it the target of cyber attacks. This emphasizes the necessity for a digital rhetorical system for the cloud surroundings. standard digital forensics can not be directly given as a cloud forlantic answer as a result of it's thanks to virtualization of multi-tenancy and resources within the cloud. whereas we have a tendencyto do cloud forensics, information cloud element logs, virtual machine disk pictures, volatile memorydumps, console logs and network capture area unit to be inspected. during this letter, we've go together with a foreign proof assortment and preprocessing framework victimization Straits and Hadoop distributed filing system. the gathering of VM disk pictures, logs etc. is triggered by a pull model once triggered by the investigator, whereas the cloud node sporadically pushes network capture to HDFS. Pre-processing steps like bunch of logs and correlation and VM disk pictures area unit done through mahout and VICA to implement track analysis.[7]Identificationofdigital forensicinthecloudcanadda new dimensionto theprocess ofcreating confidencein the cloud in.But Lotsofcloudfeaturessuch astransparency, virtualization,lack oflegal issues etc., Challenges forthe Cloud Forensics Whether it is a traditional digital forensic or cloud forensic, collecting comprehensive data for analysis is a major challenge in the investigation. Data gathering in exceptionally virtualized conditions like cloud is very tedious. The finalgoal of proof collection and analysis is to prove the official courtroom that they are forensic sound. We can use introspection techniques because they will not corrupt the source of evidence while collecting necessary data.[8]According to Content is often repeated, modified or modified on primary storage systems, and users lose control over its dispersion on the system. The content identified with a specific venture from the framework in this way turns into a work escalated errand for the client. In this work system illustrates, a system that helps the user easily remove project interconnected content, but this does not requirechange in userbehaviororanysystem component, Such as file system, kernelorapplication IRCUSis transparentlyintegrated inside the client's framework, worksin client space and storesthesubsequentmetadatawithfiles. Thiswork system describe evaluationofsystem and showed thatitsoverhead and accuracy is acceptable for practical use and deployment.

III .PROPOSED ALGORITHM

1.Apriori algorithm:-
The Apriori Algorithm is an influential algorithm for mining frequent itemsets for boolean association rules. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation, and groups of candidates are tested against the data.

2. Clustering algorithm:-
Clustering is a process of partitioning a set of data (or objects) into a set of meaningful sub-classes, called clusters. Help users understand the natural grouping or structure in a data set. Clustering: unsupervised classification: no predefined classes.

3. Polynomial Time Algorithm:-
An algorithm that is guaranteed to terminate within a number of steps which is a polynomial function of the size of the problem. See also computational time complexity. Search the data without loss of time to provide out stream for the process.

4.MachinelearningdynamicattackquerypatternWeightCalculationAlgorithmalsoapplicableforSQL injection.

5.A conventional challenge-response protocol :-
In this solution, the administrator's host verifier sends periodically a request to the server for it to compute a checksum on a file and return the result to the verifier. The verifier then compares the returned result with a locally-stored reference checksum for the same file. A naïve implementation of this protocol would be inefficient: a malicious attacker could precompute the checksums on all files he intends to modify, store these checksums, and then modify the checksum computation program to retrieve the original checksum rather than compute it. The attacker can then modify any file, while remaining able to return the expected checksums when requested by the verifier. This protocol has to be modified to guarantee the freshness of the checksum computation. This can be achieved by adding a challenge C in the request parameters. With this new protocol, the server has to compute a response R depending on the challenge. More precisely, instead of a checksum computed as the result of a one-way hash function on the content of the file, the response must be computed as the hash of the challenge concatenated with the file content:

$R = H(C|File)$ (1)

Of course, the challenge must be difficult to guess for the attacker. In particular it must be changed at each request. But then the verifier cannot simply compare the response with a reference checksum. A solution would be to maintain a copy of all the original files on the verifier and run the same response computation on the verifier as on the server. But this is impractical if the numbers of files and of servers are high and the mean file length is large. A better solution would be for the verifier to use two functions f and H ', of which at least one of them is kept secret3 , such that H ' is a one-way hash function, and f is such that:

$$f(C, H'(File)) = H(C|File) = R \quad (2)$$

Unfortunately, we have not found (yet) functions f, H and H ' satisfying this property. To workaround this problem, a finite number N of random challenges can be generated off-line for each file to be checked, and the corresponding responses computed off-line too. The results are then stored on the verifier. At each integrity check period, one of the N challenges is sent to the server and the response is compared with the precomputed response stored on the verifier. In order to guarantee that a different challenge is issued at each request, and thus that an attacker cannot predict which will be the next challenge(s), the server has to be rebooted periodically4 so that:

$$N > (\text{frequency of challenge-response protocol for the same file}) / (\text{reboot frequency}) \quad (3)$$

A possible way for the attacker to circumvent this freshness checking could be to keep copies of both the original and the modified files. But this should be easy to detect, either by checking the integrity of the concerned directories and system tables, or by intrusion detection sensors tuned to detect this specific abnormal behavior.
The table of precomputed responses, stored on the verifier, is thus composed of N entries with, for each entry, a challenge and the expected corresponding response. It is possible to reduce the size of the table, by exploiting a

technique presented in [Lamport 1981]: rather than generating a specific random number as a challenge for each entry, only one random number CN is generated for a file, and each challenge Ci is computed as H(Ci+1) for each i from (N-1) to 1 (by step of -1). The precomputed response table contains only the N expected responses, the last challenge CN and the number N.

3 At least H ' or f must be kept secret, because if both were public, it would be easy for the attacker to precompute all needed H '(File) and then dynamically compute the expected response f(C, H '(File)).
4 Our current implementation exploits the fact that the servers are periodically rebooted (e.g., once a day), as a measure for software rejuvenation [Huang et al. 1995]: at reboot all files and programs are restored from a secure copy. This would erase any Trojan horse implemented previously by a malicious hacker, as well as any table of precomputed responses he could have built with previous challenges.

The challenges are sent in the increasing order from C1 to CN, each challenge Ci being dynamically computed by the verifier:

Ci = H (N-i) (CN), where H k (X) = H(H k-1(X)) and H 1(X) = H(X) (4)

IV.PSEUDO CODE

Input: Querygeneratedfrom userQ,eachretrievedlist L fromwebpage. Output:
Each list with weight.

Heresystemhastofindsimilarityoftwolists:$\vec{a} = (a_1, a_2, a_3, \ldots)and\vec{b} = (b_1, b_2, b_3, \ldots)$, whereand$b_n$ arethe componentsa_n ofthe vector(featuresofthe document,orvaluesforeach wordofthe comment) and theis the dimension of the vectors$n$

Step1: ExtractallthefeaturesfromTestsetusingbelow

$$\text{ReceiveCommand}= \sum_{j=1}^{n}(\text{T}[\text{j}])$$

Step2: ReadallfeaturesfromTrainsetusingbelow

$$\text{PolicyList}=\sum_{k=1}^{\text{m}}(\text{T}[\text{k}])$$

Step3: Read all featuresfromTrainsetusingbelow

Step4 : Generateweightofbothfeatureset

$$W=(\text{Receive Command,PolicyList})$$

Step5: VerifyThreshold

Selected Instance=result=W>T?1 :0;

Add each selected instance into L, when n = null

Step6: ReturnL.

V. SIMULATION RESULTS

For the system performance evaluation, calculate the matrices for accuracy. The system is executed on java 3-tier architecture framework with INTEL2.8 GHz i3 processorand 4 GB RAM with E-Commerce Application & public cloud Amazon EC2 consol.For the system evaluation we create 2 machines on physical environment with Wi-Fi and 1VM with Amazon EC2 as public cloud environment.After implementing some part of system we got system performance on reasonable level.

Inthefirstexperimentalwehavecalculatedtheaccuracyofattackdetectionmoduleusingvariousnumberinstances.

Table1:System performance

Test Instances	Accuracy	Precision	Recall	F-Measure
10	0.90	0.91	0.94	0.95
20	0.91	0.92	0.95	0.96
50	0.89	0.90	0.93	0.94
100	0.90	0.93	0.92	0.95

In second experimentation system show the user verification time with different approaches. In current system we consider as four different authorities for runtime verification. The below Figure.2shows the performance measuresusing different parameters with some existing approaches.

Graph Comparison

In second experiment Figure 2 shows data encryption performance which works to displaythat the data it will encrypt in exactlyhow much period in seconds. Suppose there is a 100kb data is encrypted in 150 second so the outcome will show inevitably in that time of encryption data from the operators.

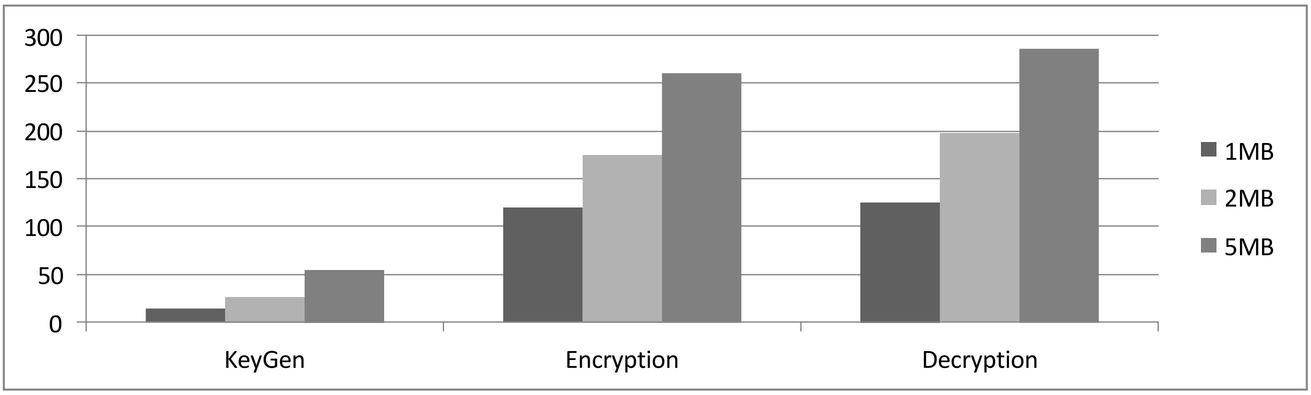


Figure 2:Data encryption performance based on data size

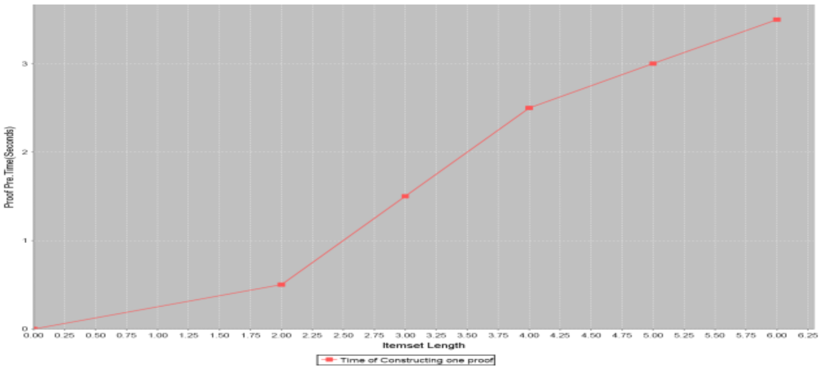


Figure 3: Clustering of data on E-Commerce Application.

Above diagram indicates Time clustering of data.
1. We can easily track which product is attack.
2. Recovery Property is used.

VI.CONCLUSION AND FUTURE WORK

To Present two integrity verification approaches for outsourced frequent itemsetmining. The probabilistic verification approach constructs evidence (in)frequentitemsets. In particular, we remove a small set of items from the original dataset and insert a small set of artificial transactions into the dataset to construct evidence (in)frequentitemsets. The deterministic approaches requires the server to construct cryptographic proofs of the mining result. The correctness and completeness are measured against the proofs with 100% certainty. Our experiments show the efficiency and effectiveness of our approaches. An interesting direction to explore is to extend the model to allow the client to specify her verification needs in terms of budget (possibly in monetary format) besides precision and recall threshold.

Proposed system provides the highest security from different type of attack in cloud environment to end users confidentialitydata. In other hand AES 128 encryption algorithm also maintain the robust securitymechanism. Access controlandrevocationmaintainthe securityandefficiencyofsystem. The system achieves RoleBase Accesscontrol in single as well as multi cloud environment with this approach. The current architecture is very efficient for security purpose,but sometime it's utilizedmultipleresources. Whensuchsystem allocates multipleresourcesit will generate a lot of dependencies. For the next update we can focus on minimum resource utilization with system flexibility like power, VM's, network, memory etc.

REFERENCES

[1] R.AgrawalandR.Srikant,"Fastalgorithmsforminingassociationrulesinlargedatabases,"inProc.20th Int. Conf. Very Large Data Bases, pp. 487–499,1994.
L.Babai,L.Fortnow,L.A.Levin,andM.Szegedy,"Checkingcomputationsinpolylogarithmictime,"inProc. 23rd Annu. ACM Symp.Theory Comput. 1991, pp. 21–32.
[2] R. Canetti, B. Riva, and G. N. Rothblum, "Verifiable computation with two or more clouds," in Proc. Workshop Cryptography Security Clouds,2011.
[3] K.-T.Chuang,J.-L.Huang,andM.-S.Chen,"Power-lawrelationshipandSelf-similarityintheitemset support distribution: Analysis and applications," VLDB J., vol. 17, pp. 1121–1141, Aug.2008.
[4] R.Gennaro,C.Gentry,andB.Parno,"Non-interactiveverifiablecomputing:Outsourcingcomputationto untrusted workers," in Proc. 30th Annu. Conf. Adv. Cryptol., 2010, pp.465–482.
[5] F. Giannotti, L. V. S. Lakshmanan, A. Monreale, D. Pedreschi, and W. HuiWang, "Privacy-preservingdata mining from outsourced databases," in Proc. 3rd Int. Conf. Comput., Privacy Data Protection,2011, pp.411426.
[6] S.Goldwasser,S.Micali,andC.Rackoff,"Theknowledgecomplexityofinteractiveproofsystems,"SIAM J. Comput., vol. 18, pp. 186–208, Feb.1989.
[7] H.Hacigum€u€s,B.Iyer,C.Li,and S.Mehrotra,"ExecutingSQL¸overencrypteddatainthedatabase service-provider model," in Proc. ACM SIGMOD Int. Conf. Manag. Data, 2002, pp.216–227.
[8] F. Li, M. Hadjieleftheriou, G. Kollios, and L. Reyzin, "Dynamic authenticated index structuresfor outsourced databases," in Proc. ACM SIGMOD Int. Conf. Manag. Data, 2006, pp.121–132.
[9] R. Liu, H. Wang, A. Monreale, D. Pedreschi, F. Giannotti, and WengeGuo, "Audio: An integrity auditing framework of Outliermining-as-a-service systems," in Proc. Eur. Conf. Mach. Learning Knowl. Discovery Databases, 2012, pp. 1–18.
[10] [10] Miss Monika D.Rokade,Mr.S.A.Kahate Mr.K.S..Kore,"Privacy kNN query Processing in the cloud computing'' JIRCCE,Vol. 3, Issue 6, June 2015.
[11] Mr. DigambarPowar, Dr. G. Geethakumari "Digital Evidence Detection in Virtual Environment for Cloud Computing" ACM, 2012.
[12] SaibharathS,GeethakumariG"CloudForensics:EvidenceCollectionandPreliminaryAnalysis"IEEE,2015
[13] Mr. Chandrashekhar S. Pawar, Mr. Pankaj R. Patil, Mr. Sujitkumar V. Chaudhari "Providing Securityand Integrity for Data Stored In Cloud Storage" ICICES, 2014.
[14] DeeviRadha Rani, G. Geethakumari "An Efficient Approach to Forensic Investigation in Cloud using VM Snapshots" International Conference on Pervasive Computing (ICPC), 2015.
[15] BKSP Kumar RajuAlluri, Geethakumari G"A Digital Forensic Model for Introspection of Virtual Machines in Cloud Computing" IEEE, 2015.
[16] HubertRitzdorf,NikolaosKarapanos,SrdjanCapkun"AssistedDeletionofRelatedContent"ACM,2014.

[17] Liang X, Zhao J, Shetty S, Liu J, Li D. Integrating blockchain for data sharing and collaboration in mobile healthcare applications. InPersonal, Indoor, and Mobile Radio Communications (PIMRC), 2017 IEEE 28th Annual International Symposium on 2017 Oct 8 (pp. 1-5). IEEE.

[18] Manoj R, Alsadoon A, Prasad PC, Costadopoulos N, Ali S. Hybrid secure and scalable electronic health record sharing in hybrid cloud. In2017 5th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud) 2017 Apr 6 (pp. 185-190). IEEE.

[19] Khan SI, Hoque AS. Privacy and security problems of national health data warehouse: a convenient solution for developing countries. InNetworking Systems and Security(NSysS), 2016 International Conference on 2016 Jan 7 (pp. 1-6).IEEE.

[20] Shrestha NM, Alsadoon A, Prasad PW, Hourany L, Elchouemi A. Enhanced e-health framework for security and privacy in healthcare system. InDigital Information Processing and Communications (ICDIPC), 2016 Sixth International Conference on 2016 Apr 21 (pp. 75-79). IEEE.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

☎ 9940 572 462 6381 907 438 ✉ ijircce@gmail.com

Scan to save the contact details

www.ijircce.com