BIG DATA PRIVACY: A TECHNOLOGICAL PERSPECTIVE AND REVIEW

[1]**Kamlesh Patil**
[1]Assistant Professor
[1]Dept of Computer Engineering Bharati Vidyapeeth College of Engineering,
[1]pune, India

Abstract— proposed Article is a review Article on Big Data cloud technology finding common issues in them and taking up small issue in cloud to solve privacy of big data. Privacy of data is major challenge in big data scenario and overlooked this paper focuses to soleve certain issues with common algorithms.

The proposed system gives the idea about how to provide security to cloud storage in network transmission. For this secured protocol designed based on Hybrid mixture of Algorithmic processes. Data Integrity is major challenge to cloud. Numerous algorithmic process have been developed which have advantages and disadvantaged. This research focuses to design and implement combined research methodology in better data security over cloud.

Index Terms— Cloud Computing, Encryption Algorithms, Data integrity.

I. INTRODUCTION

Big data large data in various file formats. [1]Big data analytics: term used to describe the process of researching massive amounts of complex data in order to reveal hidden patterns or identify secret correlations [2]. 4. Major Issue: security and privacy of big data in cloud[4].

Research Levels : security and Privacy at

- data generation
- data storage
- data processing

Cloud Computing has been envisioned as the next-generation information technology (IT) architecture for enterprises, due to its long list of unprecedented advantages in the IT history: on-demand self-service, ubiquitous network access, location independent resource pooling, rapid resource elasticity, usage-based pricing and transference of risk [1]. As a disruptive technology with profound implications, Cloud Computing is transforming the very nature of how businesses use information technology. One fundamental aspect of this paradigm shifting is that data is being centralized or outsourced to the Cloud. From users' perspective, including both individuals and IT enterprises, storing data remotely to the cloud in a flexible on-demand manner brings appealing benefits: relief of the burden for storage management, universal data access with independent geographical locations, and avoidance of capital expenditure on hardware, software, and personnel maintenances, etc [1].

While Cloud Computing makes these advantages more appealing than ever, it also brings new and challenging security threats towards users' outsourced data. Since cloud service providers (CSP) are separate administrative entities, data outsourcing is actually relinquishing user's ultimate control over the fate of their data. As a result, the correctness of the data in the cloud is being put at risk due to the following reasons. First of all, although the infrastructures under the cloud are much more powerful and reliable than personal computing devices, they are still facing the broad range of both internal and external threats for data integrity. Examples of outages and security breaches of noteworthy cloud services appear from time to time. Secondly, there do exist various motivations for CSP to behave unfaithfully towards the cloud users regarding the status of their outsourced data. For examples, CSP might reclaim storage for monetary reasons by discarding data that has not been or is rarely accessed, or even hide data loss incidents so as to maintain a reputation . although outsourcing data to the cloud is economically attractive for long-term large-scale data storage, it does not immediately offer any guarantee on data integrity and availability. This problem, if not properly addressed, may impede the successful deployment of the cloud architecture. As users no longer physically possess the storage of their data, traditional cryptographic primitives for the purpose of data security protection cannot be directly adopted [2]. In particular, simply downloading all the data for its integrity verification is not a practical solution due to the expensiveness in I/O and transmission cost across the network. Besides, it is often insufficient to detect the data corruption only when accessing the data, as it does not give users correctness assurance for those unaccessed data and might be too late to recover the data loss or damage. Considering the large size of the outsourced data and the user's constrained resource capability, the tasks of auditing the data correctness in a cloud environment can be formidable and expensive for the cloud users. To fully ensure the data integrity and save the cloud users' computation resources as well as online burden, it is of critical importance to enable public auditing service for cloud data storage, so that users may resort to an independent third party auditor (TPA) to audit the outsourced data when needed. However, most of these schemes [2] do not consider the privacy protection of users' data against external auditors. Indeed, they may potentially reveal user data information to the auditors. Without a properly designed auditing protocol, encryption itself cannot prevent data from "flowing away" towards external parties during the auditing process. Thus, it does not completely solve the problem of protecting data privacy but just reduces it to the key management. Unauthorized data leakage still remains a problem due to the potential exposure of decryption keys. How to enable a privacy-preserving third-party auditing protocol, independent to data encryption is task to solved.

It is routine for users to use cloud storage services to share data with others in a team, as data sharing becomes a standard feature in most cloud storage offerings, including Dropbox and Google Docs. The integrity of data in cloud storage, however, is subject to skepticism and scrutiny, as data stored in an untrusted cloud can easily be lost or corrupted, due to hardware failures and human errors[3]. The first provable data possession (PDP) mechanism [3] to perform public auditing is designed to check the correctness of data stored in an untrusted server, without retrieving the entire data. Wang WWRL Methodology construct a public auditing mechanism for cloud data, so that during public auditing, the content of private data belonging to a personal user is not disclosed to the third party auditor. how to preserve identity privacy from the TPA, because the identities of signers on shared data may indicate that a particular user in the group or a special block in shared data is a higher valuable target than others,

.

II. LITERATURE SURVEY: BIG DATA PROBLEMS

Large volume of data both structured and unstructured that inundates a business on a day-to-day basis. Enormous measure of data produced from various sources in multiple formats with very high speed [3]. Big data is new generation of technologies and architectures, designed to economically separate value from very large volumes of a wide variety of data.

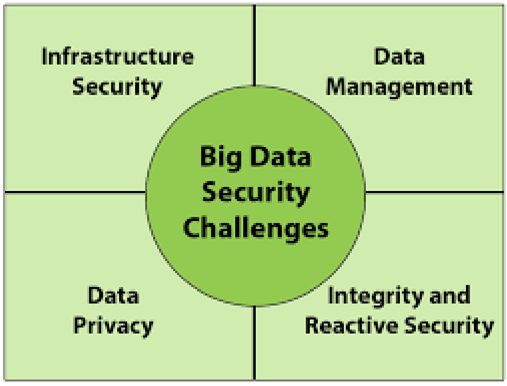

Figure 1: common issues in big data and challenges

Challenge C1

A.To achieve security and privacy on big data : data generation phase

Privacy Information privacy is the privilege to have some control over how the personal information is collected and used.

Security: Security is the practice of defending information and information assets through the use of technology, processes and training from:-Unauthorized access, Disclosure, Disruption, Modification, Inspection, Recording, and Destruction.

“With security implementation Privacy can be achieved. But without security privacy cannot be achieved.”

Methodologies to achieve security and privacy

1. Access restriction

2. Falsifying data

Scope of work exists in privacy preserving of cloud data from Third party Audit for current research on big data in cloud (S1 paper1,5)

Challenge C2

A. Big data privacy : data storage phase(cloud as storage platform)

Methodologies to achieve security and privacy

1. file level data security schemes

2. database level data security schemes

3. media level security schemes

4. application level encryption schemes

Algorithmic approach is Attribute *based encryption, Homomorphic encryption, Storage path encryption, Usage of Hybrid clouds.*

Scope of work exists in better cryptographic technique (S2 paper 2).

Challenge C3

B. Integrity verification of big data storage(cloud platform)

Cloud computing is used for big data storage data owner loses control over data. The outsourced data are at risk as cloud server may not be completely trusted. **Hence scope of work exists in Big data integrity (cloud) (S3 Paper 3 ,4)**

Integrity of data storage in traditional systems can be verified

1) Reed-Solomon code,

2) checksums,

3) trapdoor hash functions,

4) message authentication code (MAC)

5) digital signatures

Common privacy techniques on big data

1. Deidentification: K-anonymity, L-diversity and T-closeness

Suppression and Generalization

Deidentification techniques exists in Big data for privacy preserving but suffer from complexity and number of other limitations and require framework implementation.

Major challenge is re-identification is complex process with above methodologies.

2. HybrEx : hybrid models in big data processing.
This model only **vertical partitioning is feasible task for data handlings**

Challenge C4
Operations over encrypted

data are mostly complex along with being time-consuming and big data is high-volume and needs us to mine new knowledge in a reasonable timeframe, running operations over encrypted data can be termed as inefficient in the case of big data analytics. **Hence scope of work exists in scalable encrypted data searching(S4 paper 6).**

Author [1] proposed that cloud computing system provides an economical and efficient solution for sharing group resource among cloud users. In this paper, authors propose a secure multi- owner data sharing scheme, named Mona, for dynamic groups in the cloud. By leveraging group signature and dynamic broadcast encryption techniques, any cloud user can anonymously share data with others. Meanwhile, the storage overhead and encryption computation cost of this scheme are independent with the number of revoked users.
Author [2] have proposed system for proving auditing for data stored on cloud. Cloud provide on demand high quality application and services without burden of local data storage and maintenance. If data is no longer in user possession, then providing integrity is an formidable task. In this way authors propose a secure cloud storage system supporting privacy preserving public auditing and perform auditing for multiple users simultaneously.
Author [3] have proposed a cryptographic storage system that enables secure file sharing on entrusted servers, named Plutus. By dividing files into file-groups and encrypting each file-group with a unique file-block key, the data owner can share the file groups with others through delivering the corresponding lockbox key, where the lockbox key is used to encrypt the file-block keys. Additionally, the file-block key needs to be updated and distributed again for a user revocation.
Author[4] have proposed a system with basic encryption and decryption techniques for providing security. In revocation, the original data are first divided into a number of slices, and then published to the cloud storage. When a revocation occurs, the data owner needs only to retrieve one slice, and re-encrypt and re-publish it. Thus, the revocation process is accelerated by affecting only one slice instead of the whole data.
Author [5] have proposed a scalable and fine-grained data access scheme with KP-ABE technique. The data owner uses a random key to encrypt a file, where the random key is further encrypted with a set of attributes using KP-ABE. Then, the group manager assigns an access structure and the corresponding secret key to authorized users, such that a user can only decrypt a cipher text if and only if the data file attributes satisfy the access structure.
Author[6] have proposed a secure provenance scheme, in which group signatures and cipher text-policy attribute based encryption techniques are used. Each user obtains two keys after the registration: a group signature key and an attribute key. Thus, any user is able to encrypt a data file using attribute-based encryption and others in the group can decrypt the encrypted data using their attribute keys. Meanwhile, the user signs encrypted data with her group signature key for privacy preserving and traceability.

III. PROBLEM TAKEN UP

To Design and Develop privacy preserving framework using big data privacy preserving in cloud computing.

IV. CONCLUSION AND FUTURE SCOPE

To privacy preserve in big and cloud data. integrate this framework in all cloud setup for better and solve integrity and privacy challenges in cloud with complete reliable solution.

REFERENCES

[1] 1. Abadi DJ, Carney D, Cetintemel U, Cherniack M, Convey C, Lee S, Stone-braker M, Tatbul N, Zdonik SB. Aurora: a newmodel and architecture for data stream manag ement. VLDB J. 2003;12(2):120–39.

[2] Kolomvatsos K, Anagnostopoulos C, Hadjiefthymiades S. An efficient time optimized scheme for progressive analyticsin big data. Big Data Res. 2015;2(4):155–65.

[3] Big data at the speed of business, [online]. http://www-01.ibm.com/soft-ware/data/bigdata/2012.

[4] Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers A. Big data: the next frontier for innovation, competition,and productivity. New York: Mickensy Global Institute; 2011. p. 1–137.

[5] Gantz J, Reinsel D. Extracting value from chaos. In: Proc on IDC IView. 2011. p. 1–12.

[6] Tsai C-W, Lai C-F, Chao H-C, Vasilakos AV. Big data analytics: a survey. J Big Data Springer Open J. 2015.

[7] Mehmood A, Natgunanathan I, Xiang Y, Hua G, Guo S. Protection of big data privacy. In: IEEE translations and contentmining are permitted for academic research. 2016.

[8] Namdeo, Jyoti, and NaveenkumarJayakumar. "Predicting Students Performance Using Data Mining Technique with Rough Set Theory Concepts." International Journal 2.2 (2014).

[9] Jayakumar, D.T. and Naveenkumar, R., 2012. SDjoshi,“. International Journal of Advanced Research in Computer Science and Software Engineering,” Int. J, 2(9), pp.62-70.

[10] Raval, K.S., Suryawanshi, R.S., Naveenkumar, J. and Thakore, D.M., 2011. The Anatomy of a Small-Scale Document Search Engine Tool: Incorporating a new Ranking Algorithm. International Journal of Engineering Science and Technology, 3(7).

[11] Naveenkumar, J., Makwana, R., Joshi, S.D. and Thakore, D.M., 2015. Performanc Impact Analysis of Application Implemented on Active Storage Framework. Internationalournal, 5(2).

[12] Naveenkumar, J., Keyword Extraction through Applying Rules of Association and Threshold Values. International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), ISSN, pp.2278-1021.

[13] Jayakumar, M.N., Zaeimfar, M.F., Joshi, M.M. and Joshi, S.D., 2014. INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING & TECHNOLOGY (IJCET). Journal Impact Factor, 5(1), pp.46- 51.

[14]. Kakamanshadi, G., Naveenkumar, J. and Patil, S.H., 2011. A Method to Find Shortest Reliable Path by Hardware Testing and Software Implementation. International Journal of Engineering Science and Technology (IJEST), ISSN, pp.0975-5462.

[15] Archana, R.C., Naveenkumar, J. and Patil, S.H., 2011. Iris Image Pre-Processing And Minutiae Points Extraction. International Journal of Computer Science and Information Security, 9(6), p.171.

[16] Salunkhe, R. and Jaykumar, N., 2016, June. Query Bound Application Offloading: Approach Towards Increase Performance of Big Data Computing. In Journal of Emergin Technologies and Innovative Research (Vol. 3, No. 6 (June-2016)). JETIR.

[17] Salunkhe, R., Kadam, A.D., Jayakumar, N. and Thakore, D., 2016, March. In search of a scalable file system state-of-the-art file systems review and map view of new Scalable File system. In Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on (pp. 364-371). IEEE.

[18] Naveenkumar, J., Makwana, R., Joshi, S.D. and Thakore, D.M., 2015. Offloading Compression and Decompression Logic Closer to Video Files Using Remote Procedure Call. Journal Impact Factor, 6(3), pp.37-45.

[19] Jayakumar, N., Singh, S., Patil, S.H. and Joshi, S.D., 2015. Evaluation Parameters of Infrastructure Resources Required for Integrating Parallel Computin Algorithm and Distributed File System. IJSTE-Int. J. Sci. Technol. Eng, 1(12), pp.251-

[20] Kumar, N., Angral, S. and Sharma, R., 2014. Integrating Intrusion Detection System with Network Monitoring. International Journal of Scientific and Research Publications, 4, pp.1-4.

[21] Jayakumar, N., Bhardwaj, T., Pant, K., Joshi, S.D. and Patil, S.H., 2015. A Holistic Approach for Performance Analysis of Embedded Storage Array. Int. J. Sci. Technol. Eng, 1(12), pp.247-250.

[22] Jayakumar, N., 2014. Reducts and Discretization Concepts, tools for Predicting Student's Performance. Int. J. Eng. Sci. Innov. Technol, 3(2), pp.7-15.

[23] Salunkhe, R., Kadam, A.D., Jayakumar, N. and Joshi, S., 2016, March. Luster a scalable architecture file system: A research implementation on active storage array ramework with Luster file system. In Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on (pp. 1073-1081). IEEE.

[24] Naveenkumar, J., SDJ, 2015. Evaluation of Active Storage System Realized Through Hadoop. International Journal of Computer Science and Mobile Computing, 4(12), pp.67-73.

[25]. Bhore, P.R., Joshi, S.D. and Jayakumar, N., 2016. A Survey on the Anomalies in System Design: A Novel Approach. International Journal of Control Theory and Applications, 9(44), pp.443-455.

[26] Bhore, P.R., Joshi, S.D. and Jayakumar, N., 2017. Handling Anomalies in the System Design: A Unique Methodology and Solution. International Journal of Computer Science Trends and Technology, 5(2), pp.409-413.

[27]. Zaeimfar, S.N.J.F., 2014. Workload Characteristics Impacts on file System Benchmarking. Int. J. Adv, pp.39-44.

[28] Bhore, P.R., Joshi, S.D. and Jayakumar, N., 2017. A Stochastic Software Development Process Improvement Model To Identify And Resolve The Anomalies In System Design. Institute of Integrative Omics and Applied Biotechnology Journal, 8(2), pp.154-

[29] Kumar, N., Kumar, J., Salunkhe, R.B. and Kadam, A.D., 2016, March. A Scalable Record Retrieval Methodology Using Relational Keyword Search System. In Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies (p. 32). ACM.

[30] Naveenkumar, J. and Joshi, S.D., 2015. Evaluation of Active Storage System Realized Through Hadoop. Int. J. Comput. Sci. Mob. Comput, 4(12), pp.67-73.

[31] Naveenkumar, J., Bhor, M.P. and Joshi, S., 2011. A self process improvement for achieving high software quality. International Journal of Engineering Science and Technology (IJEST), 3(5), pp.3850-3053.