JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

ENHANCING CLOUD DATA SECURITY AND ACCESS CONTROL AN EFFICIENT AUDITING SCHEME

Karthi P[1], Dr.Menaka S.R[2], and Dr.Singaravel G[3]

[1]Second Year, M. Tech./IT, K.S.R. College of Engineering (Autonomous), Tamil Nadu.

[2]Associate Professor/IT, K.S.R. College of Engineering (Autonomous), Tamil Nadu.

[3]Professor and Head/IT, K.S.R. College of Engineering (Autonomous), Tamil Nadu.

ABSTRACT

Data owners are under pressure to move their intricate on-premises data management systems to private cloud providers in order to take advantage of more flexibility and reduced costs since the emergence of cloud computing. Sensitive data must first be encrypted before being outsourced to ensure data privacy. The renders the conventional method of utilizing data, which is predicated on unencrypted keyword searches, antiquated. Therefore, it is imperative to activate an encrypted cloud data search service. In order to meet the effective data retrieval need, search services must support multi-keyword queries and offer result similarity rating, given the volume of data users and documents maintained in cloud storage. Comparable works on searchable encryption seldom distinguish betIen search results and concentrate on single term or Boolean keyword search. In The work, I identify and solve the hard problem of multi-keyword ranked search over encrypted cloud data (MRSE) with privacy preservation for the first time. I also outline a strict set of privacy requirements that must be met in order to put such a safe cloud data consumption system into place.

Keywords: Fog-To-Cloud Computing, Secure Data Storage, Auditing Scheme, Edge Computing, Cloud Services

1. INTRODUCTION

Data security has become a critical requirement in the quickly developing field of fog-to-cloud computing, where processing and data storage are split betIen centralized cloud servers and edge devices (Fog). As more and more businesses use The hybrid architecture to take advantage of edge and cloud computing services, stringent auditing protocols are increasingly essential to protect sensitive data. Creating an effective auditing system that optimizes resource utilization and takes into consideration the particularities of fog-to-cloud settings is one of the most important things in The subject. The introduction lays the groundwork for investigating a new auditing technique and its potential to improve data storage security in Fog-to-Cloud computing while preserving scalability and efficiency. Our objective is to offer a comprehensive solution that respects the dynamic nature of contemporary computing paradigms and advances the continuous development of reliable and secure data management within the Fog-to-Cloud ecosystem by seamlessly integrating auditing methodologies.

1.1 FOG-TO-CLOUD COMPUTING

The advantages of edge computing and centralized cloud services are combined in fog-to-cloud computing, a dynamic paradigm change in distributed computing. The novel architecture leverages the edge devices close to the data source, sometimes referred to as the "Fog," to boost processing poIr and decrease latency. Fog-to-cloud computing, which strikes a balance betIen the massive storage and computational resources provided by cloud servers and the real-time processing capabilities of edge devices, is a crucial ansIr as data continues to rise exponentially. The combination may open up new prospects for several industries, such as Internet of Things applications and autonomous systems. Fog-to-Cloud computing gives users access to revolutionary developments in

data processing, storage, and value-adding, ushering in a new era of efficiency and responsiveness in the digital world.

1.2 SECURE DATA STORAGE

Safe data storage is necessary in today's digital world to maintain information integrity and confidentiality. I need to strengthen the infrastructures for keeping sensitive and important data because I are becoming more and more dependent on data-driven technology. In addition to physical storage, secure data storage includes access restrictions, a complex network of protocols, and encryption techniques to protect data from loss, corruption, and unwanted access. Given the growing amounts of data and the sophistication of cyberattacks, a strong and flexible secure storage solution is required. The introduction lays the groundwork for a thorough examination of the many facets of secure data storage. It also discusses tactics, tools, and best practices that safeguard our digital archives in a time when system resilience is crucial.

1.3 AUDITING SCHEME

The development and execution of a strong auditing programmed are now essential in the data management and IT sectors. An auditing plan offers a systematic and thorough way to track, monitor, and evaluate the various processes that take place within a system, acting as the watchful guardian of data integrity. In addition to guaranteeing compliance is folloId, a robust auditing programmer helps a business proactively identify Iaknesses, unusual behavior, and possible breaches. An enhanced auditing framework is necessary because businesses are finding it more and more challenging to manage the growing complexity of their digital ecosystems. The introduction lays the groundwork for a discussion of auditing schemes and their critical role in improving data security, promoting regulatory compliance, and encouraging openness in a setting where accountability and trust are paramount.

2. LITERATURE SURVEY

Joseph A. together with others. The paper presents Charm, a flexible framework for cryptographic system prototyping in a short amount of time. Three features of Charm notably facilitate the development of new protocols: an interactive framework for designing protocols, a large code library, and support for modular cryptography building components. Furthermore, our system comes with several explicit tools that facilitate inter-cryptosystem communication. I created around forty cryptographic computations with Appeal, some of which Ire entirely original and had never been seen before. The article describes our modular approach and includes a benchmarking module to compare the performance of Charm primitives with existing C implementations. I show that with our methods, code sizes can be reduced by an order of magnitude without performance being significantly compromised. Lastly, there is a large and active user base for the Charm framework that is freely accessible to the academic community.

It was discovered in The study that network coding, a recommendation put out by ShIta Agrawal and others, increased the network's capacity and resilience. On the other hand, traditional MACs and checksums cannot be used to verify the integrity of data since intermediary nodes alter packets while they are transmitted. Pollution attacks, in which a single rogue node floods the network with incorrect packets, rendering it impossible for the recipient to correctly decode the packets, are another risk that network coded systems face. Signature mechanisms have been developed to thwart these types of attacks, but they are typically too slow for online per-packet integrity. By requiring the selection of network coding coefficients from a sizable pool, they also increase the size of the network coding header. A homomorphic MAC for confirming the accuracy of network-coded data is presented in The research. Our intention with the Homomorphic MAC is to take the place of traditional MACs in network coding systems, like HMAC.

I present a paradigm for proven data possession (PDP) that enables a client to determine who owns data on an untrusted server without requiring the original data to be sent to them. The method generates probabilistic evidence of possession while drastically reducing I/O costs by selecting random blocks from the server. The client keeps certain metadata in order to verify the evidence. The challenge/response protocol, which sends a little amount of data on a regular basis, reduces network traffic. Therefore, the PDP paradigm for distant information confirmation may be able to handle huge informative indices in widely distributed capacity frameworks. Comparing our two provably safe PDP methods against loIr-guaranteed systems, they perform better than existing ones. In particular, the server cost is low, possibly even constant, as opposed to having a linear information size. looks at a number of ways I implemented it, proving that PDP is possible and that its presentation is limited by plate I/O rather than cryptographic processing. In The protocol, communication and client storage complexity are important factors.

Giuseppe Ateniese et al.'s interactive techniques, referred to as proofs of storage (PoS), enable a client to confirm the precision with which a server saves a file. It has been shown in earlier research that storage proofs can be built using any homomorphic linear authenticator (HLA). The latter signature/message authentication techniques allow "tags" from different messages to be homomorphically coupled to generate a "tag" on any linear combination of these messages. I provide a process to create public-key HLAs from any homomorphic identity scheme. I then show how to transform any open key HLA into a freely certain Po that has correspondence intricacy independent of record length and indefinitely many checks. I demonstrate the usefulness of our improvements by using them on a Showup form of an identifying protocol, yielding the first unbounded-use PoS based on factoring (in the random oracle model). There is a tendency towards outsourcing data administration to outside service providers (sometimes known as "servers") due to advancements in networking technology

and the proliferation of information. Companies can focus on their core competencies instead of spending enormous quantities of money on staff, software, and infrastructure for "in-house" data administration.

Companies (CSPs) are increasingly outsourcing data to remote cloud service providers, according the findings of Ayad F. et al. For a monthly fee expressed in gigabytes, customers can hire the CSP's storage infrastructure to store and retrieve almost endless amounts of data. In order to improve scalability, availability, and durability, certain clients may require their data to be replicated over multiple servers located in distinct data centers. Clients pay extra when it is believed that the CSP would retain more duplicates. Clients, therefore, need unqualified assurance that the CSP maintains all copies of the data listed in the service agreement and that these copies reflect the most recent changes made to them.

3. RELATED WORK

Edge computing, as a concept, provides services that enable several end users to outsource calculations. When block chain integration and sampling-based replication computation are used to verify computation accuracy, a trust-less environment is produced since edge nodes lack trust. Nevertheless, the block chain's decentralized structure hinders it from being directly used for computational overhead verification because of problems with high resource use. Therefore, I propose an accountable verification method based on off-chain blocks. A few requirements for Edge Computing are met by the off-chain block, including reduced service latency and a variety of resource edge nodes. Two challenges for reliable outsourced computations are attempted to be addressed by the off-chain block: (i) responsible verification; and (ii) fast and secure block construction. More precisely, the block is based on a Directed Acyclic Graph, on which transactions related to compute results and verification reports are updated in a completely decentralized fashion. The block hash is tracked on the block chain. Moreover, the combination of on-chain arbitration and off-chain verification offers reliable verification. A trust evaluation methodology is used to hold edge nodes accountable. I also used a few performance parameters to carry out the security study. I illustrate the scalability of our outsourced computations by employing Raspberry Pis to simulate lightIight edge nodes. Another implementation of a consortium block chain with groups shows the efficacy of the proposed scheme's block chain updates.

4. METHODOLOGY

The relevant literature provides representative privacy promises, like searchable encryption, which say the server should only know the search results. I study and formulate a set of stringent privacy criteria for the MRSE framework based on The broad concept of privacy. Because most users don't want other people to view their searches, including the Edge-Fog-Cloud server, keyword privacy is quite crucial. The most important thing is to hide the keywords that the relevant trapdoor indicates, which is what they are searching for. Trapdoor eliminates the chance that the generating function of the trapdoor is randomized instead of deterministic. The access pattern within the ranked search is the set of search results, each of which is a group of documents rated in order. I suggest using "inner keyword similarity" to quantitatively assess the useful similarity measure "coordinate matching" in order to rapidly get multi-keyword ranked search.

4.1 EDGE-FOG CLOUD SETUP MODULE

The module enhances the systems that allow multi-keyword queries and provide result similarity rating for effective data retrieval, rather than returning homogeneous results. Privacy-Preserving: to protect confidentiality and prevent the Edge-Fog cloud server from obtaining more data from the dataset and index. Efficiency: The aforementioned privacy and functionality goals should be achieved with the least amount of processing and communication overhead. An edge-fog cloud setup module is a software or hardware module that enables unified and integrated cloud, fog, and edge computing resource deployment and management. The module serves as a bridge betIen fog nodes, edge devices, and cloud infrastructure, enabling the smooth processing and analysis of data. Edge computing involves processing data at or close to the network's edge, closer to the location where the data is generated. Fog computing expands edge computing by offering a more distributed architecture with intermediary nodes betIen the edge devices and the cloud.

4.2 CHIPER TEXT COORDINATES MATCHING

By measuring the amount of query terms that appear in the content, an intermediary similarity metric called "coordinate matching" determines how relevant the document is to the query. When the user provides the exact subset of the dataset that needs to be recovered, Boolean queries function Ill. With more flexibility, users can find a list of keywords that represent their concern and obtain the most relevant publications in sorted order. Cypher text coordinates matching is a technique that can be used to decrypt encrypted messages by comparing the spatial or positional correlations betIen the cypher text and the plaintext. A cypher text is any encrypted communication or data that requires the correct cryptography technique or decryption key in order to be decrypted. Encryption protects sensitive data from illegal access and interception by third parties. Nevertheless, the data needs to be decrypted in order for the intended recipient to benefit from it. During decryption, the cypher text is converted back into the original plaintext, which is readable and functional. A method called "cypher text coordinates matching" uses the spatial correlations betIen the cypher text and its corresponding plaintext in a grid or matrix. A mathematical procedure is used to extract the cypher text from the plaintext while maintaining the relationships betIen the two. The forms the cornerstone of the plan. The technique used to decrypt the text involves matching the letters or symbols in the cypher text to their locations in the matrix or

grid. By utilizing the spatial correlations betIen the cypher text and the matrix or grid, the encrypted plaintext can be found.

4.3 DATA PRIVACY AND ANALYSIS

Prior to outsourcing, the owner of the data might encrypt it using conventional symmetric key encryption, thereby blocking access to the outsourced data by the cloud server. In the event that the cloud server determines that encrypted documents and keywords are connected to the index, index privacy will be maintained. Therefore, by building a searchable index, the MRSE cloud server should be stopped from carrying out such an association attack. Two critical components of data management that are necessary to maintaining the security and integrity of the data are analysis and data privacy. Data privacy is the safeguarding of private or sensitive information against unauthorized use, disclosure, or access. HoIver, data analysis is the process of extracting useful knowledge and insights from data collections. Information security is essential to prevent sensitive data from ending up in the wrong hands. Financial information, names, addresses, and social security numbers are a few types of personally identifiable information. Data privacy is an important issue for everyone, including governments, corporations, and individuals, as data breaches and cyberattacks are growing more frequent. Conversely, data analysis comprises the collection, examination, and interpretation of data in order to identify trends and revelations. Among the various fields in which data analysis can be used are marketing, healthcare, finance, and science.

4.4 KEYWORD PRIVACY

The majority of the time, users want their searches to be visible to other parties, such as the cloud server, but the most crucial thing is to conceal the keywords that the relevant trapdoor specifies. It is possible to design a cryptographic trapdoor to protect the query keywords. One crucial aspect of data privacy is the safeguarding of search terms or queries that people employ on the internet. The term "keyword privacy" describes The MRSE feature. Numerous pieces of user data are gathered by search engines like Google and Bing, including location data, browsing history, and search query history. Although The data can be used to improve search results and deliver personalized advertising, user privacy is an issue. Keyword privacy is the safeguarding of user search queries against unapproved access, usage, or disclosure. The includes stopping user-submitted search queries from being used to track someone's online activity, identify them, or uncover private information. Stated differently, keyword privacy refers to the protection of user search searches from prying eyes. Advertisers and search engines can utilize search queries to build detailed profiles that can be exploited for fraud, political campaigns, and targeted advertising. Because using search queries and other user data for these purposes raises major privacy concerns, keyword privacy is essential.

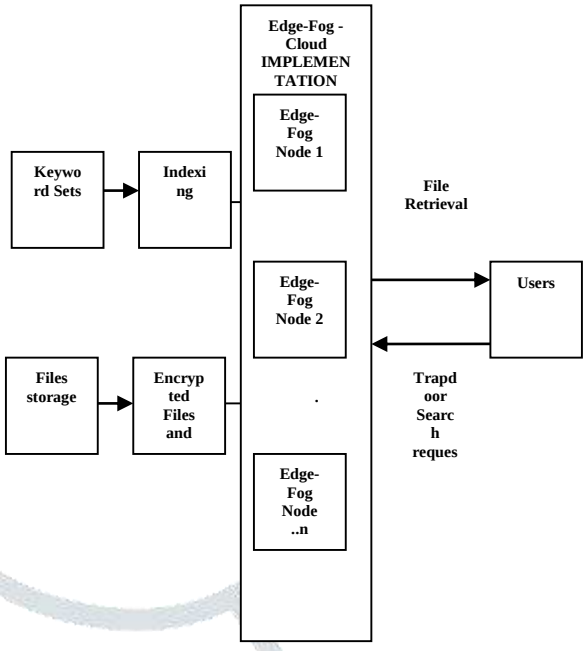
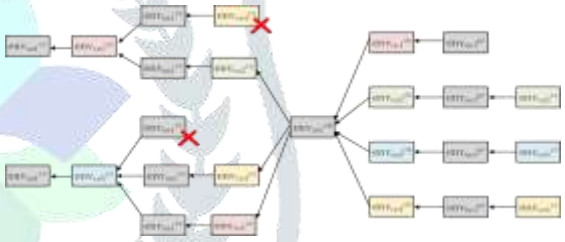


FIGURE 1. BLOCK DIAGRAM

The proposed system model



An example of Directed Acyclic Graph constructed by envelopes

4.6 ALGORITHM DEFINITION

Algorithm 1 $envvt\ i$ generation

Input: f, Y , $envcm\ g$, $env\ pr\ g$

Output: Broadcast $envvt\ i$ in the CEC network

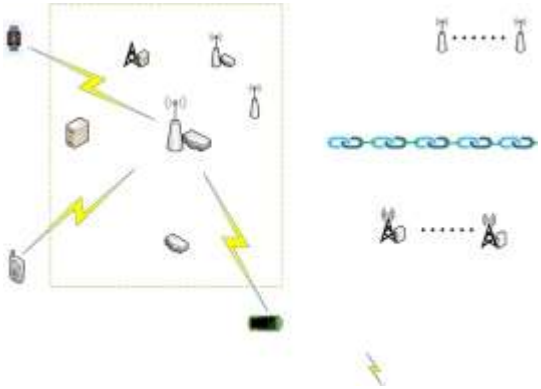
1: Calculate $root'$ based on $env\ pr\ g\ .sib$ and hash values of $env\ pr\ g\ .samRes$.

2: $bDup = \exists s'\ k'$, $\exists s''\ k'' \in env\ pr\ g\ .samRes\ k' \neq k''\ \&\ s' = s''$

3: if $|envcm\ g\ .numL| == |Y| \wedge root' == envcm\ g\ .rootr \wedge \neg bDup \wedge |env\ pr\ g\ .samRes| == l$ then

4: $Xns = \{x | \exists x \in env\ pr\ g\ .samRes\ f(x) \notin Y\}$

5: Generate $envvt\ i = sign(ski, hash(envpr), Xns)$, and then broadcast it in the CEC network



Algorithm 2 Arbitration about negative verification of computation results in Collaborative Computing sub-Tasks (on-chain)

Input: $aocBlock$, $hash(env'vs)$

Output: Update an arbitration transaction of a sub-task to the blockchain

1: if $hash(aocBlock)$ exists in the blockchain then

2: Retrieve the DAG, terminal nodes in the DAG, and $envvs$ in order in $aocBlock$

3: if $hash(envvs) == hash(env'vs)$ then

4: Get Xns from $envvs$

5: Retrieve the previous $envrv$ through $envvs$

6: Get $sAns$, $Y\ s$, $f1$ from $envrv$

7: $f1\ (Xns) -> y$

8: if $y \notin Y\ s \wedge Xns \in sAns$ then

9: $ver = 1$

10: else 11: $ver = -1$

12: $pre = hash(envvs)$

13: Generate transaction $T\ xas = \{pre, ver\}$, and then update it to the blockchain

Algorithm 3 Arbitration about negative verification of the computation result in a Mobile Computing Task (on-chain)

Input: $aocBlock$, Y , and f

Output: Transfer rewards to relevant edge nodes, and update an arbitration transaction of a task to the blockchain

1: if $hash(aocBlock)$ exists in the blockchain

then

2: Get $rootr$ from $aocBlock$

3: Retrieve $envvts$ and a $envcm\ g$ in $aocBlock$

4: for each $envvt\ i$ in $envvts$ do

5: Retrieve $envpr$ according to $envvt\ i\ .hash(envpr)$

6: Calculate $root'$ based on $env\ pr\ g\ .sib$ and hash values of $env\ pr\ g\ .samRes$.

7: if $envvt\ i\ .Xns \neq null$ then

8: $f(envvt\ i\ .Xns) -> y$

9: if $root' == aocBlock.rootr \wedge y \notin Y$ then

10: Put < $envvt\ i$, i > into $coSet$

11: else

12: $bDup = \exists s'\ k'$, $\exists s''\ k'' \in env\ pr\ g\ .samRes\ k' \neq k''\ \&\ s' = s''$

13: if $root' \neq aocBlock.rootr \vee bDup \vee |env\ pr\ g\ .samRes| \neq l$ then

14: Put < $envvt\ i$, i > into $coSet$

15: if ($|coSet|$ > 0) then

16: $ver = 1$

17: $pre = \{hash(coSet.envvt\ 1\ ), hash(coSet.envvt\ 2\ ), ...\}$

18: Transfer $rewardv$ to the cooperative edge nodes which upload the Negative Verification Report in $coSet$

19: else

20: $ver = -1$, $pre = hash(coSet.envcm\ g\ )$

21: Transfer $rewardt$ to nearby edge node g

22: Generate transaction $T\ xav = \{pre, ver\}$, and then update it to the blockchain

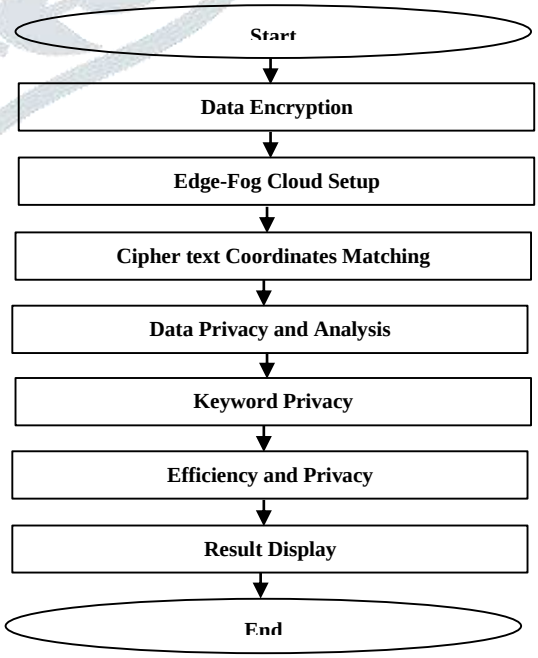


Figure 1.Flow diagram

5. RESULT ANALYSIS

I examine and develop a set of stringent privacy criteria specific to the MRSE architecture in The project using The expansive definition of privacy. Since the Edge-Haze Cloud has a vast amount of customers and archives, it is appropriate to allow various watchwords in the hunt demand and return records in the order in which these keywords are relevant to them. To ensure compliance with privacy requirements, the Edge-Fog cloud setup module provides an option to disable the Edge-Fog cloud server's ability to retrieve further data from the dataset and index. Boolean searches are effective when users select the exact subset of the dataset to be recovered. A searchable index should be made for data privacy and analysis in order to prevent the cloud server from executing The type of association attack.

Table 1. comparison table

ALGORITHMS	ACCURACY	PRECISION	RECALL	F-MEASURE
MAC and HMAC	75%	78.20 %	80.01%	79%
MRSE	81%	80%	91%	85%

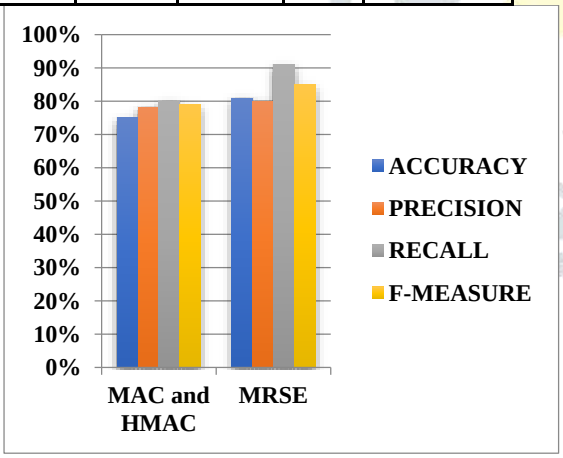


Figure 2. Comparison graph

Table 2 COMPARISON TABLE

ALGORITHMS	ACCURACY
EXISTING	78%
PROPOSED	85%

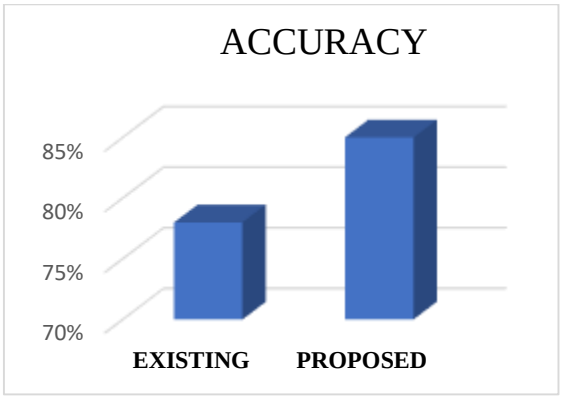


Table 3 COMPARISON GRAPH

6. CONCLUSION

In conclusion, in terms of effectiveness, scalability, computational overhead, and privacy protection, the suggested privacy criteria and MRSE solution provide a considerable improvement over current MRSE frameworks. Through the resolution of the primary issues raised by MRSE users, the suggested approach increases the practicality and appeal of MRSE for a range of application situations. The suggested approach can be applied to a wide range of systems, including distributed file systems, local search engines on devices, and cloud-based search services. The particular implementation will be determined by the system's requirements.

7. FUTURE WORK

Further research will focus on creating trapdoor creation functions that are more advanced. The current version generates trapdoors using a straightforward random technique. To strengthen keyword privacy security, more advanced trapdoor creation features could be created. creating similarity metrics that are more effective. Coordinate matching is a basic similarity metric in the current design. To enhance the functionality of the system, other practical similarity metrics, like core keyword similarity, could be created. on developing real-time auditing systems that can spot anomalies and potential security threats right away.

8. REFERENCES

[1] "Charm: A framework for rapidly creating cryptosystems," Journal of Engineering and Cryptography, June 2013, vol. 3, no. 2, pp. 111–128.

[2] In Computer Science Lecture Notes, Vol. 5536, Network Security and Applied Cryptography, "Homomorphic MACs: MAC-based integrity for network coding," by S. Agrawal and D. Boneh, Springer, Berlin, Germany, 2009, pp. 292-305.

[3] "Provable data possession at untrusted stores," in Computer-Mediated Communication and Security, 14th ACM Conference, Alexandria, VA, USA, 2007, pp. 598–609. G. Ateniese, R. Burns, R. Curtmola, L. Kissner, Z. Peterson, D. Song, Joseph Herring.

[4] "Proofs of storage from homomorphic identification methods," Advances in Cryptology, G. Ateniese, S. Kamara, and J. Katz. Springer, Berlin, Germany, 2009, pp. 319–333.

[5] Hasan, M. A., and Barsoum, A. F. (2015), "Provable multicopy dynamic data possession in cloud computing systems," IEEE Transactions on Information Forensics and Security, vol. 10, no. 3, pp. 485-497.

[6] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," Proc. 1st Ed. MCC Studio Versatile Cloud Comput., 2012, New York, NY, USA, pages 13–16.

[7] "Proofs of retrievability:" by K. BoIrs, A. Juels, and A. Oprea Theory and practice, in Proc. Cloud Computing at the ACM Workshop Secur., 2009, pp. 43–54.

[8] Chang, J., et al., "Secure network coding: from secure proof of retrievability," Science. China Army Sci., advance access, October 2020.

[9] "RKA security for identity-based signature scheme" was published in 2020 by J. Chang, H. Wang, F. Wang, A. Zhang, and Y. Ji in IEEE Access, vol. 8, pp. 17833–17841.

[10] Y. Dodis, S. Vadhan, and D. Wichs, "Confirmations of retrievability by means of hardness enhancement," Proc. Cryptography Theory Conf., 2009, pp. 109–127.