AN ANALYTICAL SURVEY ON SEARCH QUERIES OVER ENCRYPTEDCLOUD DATA USING HYBRID ENCRYPTION

[1]Prachi Chaudhari, [2] V.S.Narayana Tinnaluri
[1]Student, [2]Professor
[1]Computer Science and Engineering,
[1]Sandip University, Nashik, India

Abstract: Due to the enormous size and its complexities of the cloud storage there is always been a threat to the security of the data in the cloud. Because of this cloud storage offers data encryption schemes implicitly or explicitly. As the data size of the cloud storage increases, which gives rise to the complexity in the searching process, this is mainly due to the encrypted form of the data. To search the data via decryption schemes that always allows to compromise with the security of the cloud. This gives rise to the concept of search over the encrypted data in the cloud. Many techniques are available to perform searches over the encrypted data, which often need to adjust with its precision because of the complexity in handling search keywords. So this research paper mainly analyzes all the past works and tries to evaluate the gaps in the work to evolve the new idea on search over encrypted data in cloud.

Keywords— Cloud Computing, Searching, Encryption, Data Security.

I. INTRODUCTION

Cloud computing refers to the practice of utilizing a plethora of services such as storage, software, servers and software development platforms, over the internet. This is a very convenient process that is used predominantly by a lot of customers nowadays. It is quite useful for applications where small-scale industries can utilize the computational and storage enhancements mostly enjoyed by large scale corporations.

The name cloud computing has been derived from a relic of the past. It is inspired by the traditional depictions of a WAN (Wide Area Network) or the internet is represented as a cloud in various diagrams and flow charts as a cloud. This representation has stuck and the internet services have been collectively referred to as the Cloud. Cloud Computing is an upcoming technology that is being adopted readily by a lot of consumers. The adoption has been purely due to the fact that it is highly convenient for most of its consumers as opposed to actually set up the infrastructure to achieve the same level of computation or storage. This would be an expensive endeavor for the aggregation of the hardware as it would be extremely expensive and also not economical for a small-scale business to maintain the infrastructure to provide optimum performance.
Cloud Computing has been the center of all technological development since a while now and has also attracted the attention of some critics. These critics are skeptical of the concept of cloud and remark that it would not be adopted widely as a mainstream. This is due to the fact that the critics consider that the organizations willing to move to the cloud have to forfeit their control over their data. This is due to the fact that the cloud-based technology would cluster the data and spread it across various servers in the world for optimum speed and ease of access to the organization. This is a problem that would not allow certain organizations with critical data, such as banks, be willing to part away with their confidential customer data to the service provider.

Cloud computing is one of the most recent developments in the technology sector. The cloud is nothing but the internet which provides certain services to organizations and individuals. These services are quite useful and provide a much efficient and robust alternative to building and maintaining our own architecture. This is very convenient for small businesses and individuals who can pay a small fee to access a plethora of services.

One of the most common usages of cloud happens to be for storage purposes. This is a predominant use of cloud computing technology. It enables anyone to upload a certain file online and have the ability to access it anywhere in the world, on any device ubiquitously. This is a very convenient feature for people to have accessible storage for their valuable documents and files.

But this is not without its drawbacks, as the cloud is open territory and your sensitive data can be susceptible to a lot of attacks which can lead to information leakage of sensitive data. This is due to the fact that most of the data stored in the cloud

are distributed across various servers. If the servers are not well guarded against the common attacks, it could lead to a lot of personal information and files being leaked.
As the data is stored on to the cloud is decentralized and its copies are stored on various servers across the world to facilitate easy access for the organization, it can lead to the data being compromised. This data is stored on remote servers that can be taken down in an attack or worse lead to some form of data leakage. This is a troublesome predisposition for sensitive documents being stored on the cloud.

To ensure that the valuable data of the organizationsare safe from the attackers' hands, it is imperative to select a trustworthy cloud services provider. As storing the sensitive files on an untrustworthy server would definitely lead to some form of a lapse in security. Security must be ensured so that only authorized personnel can access the data at any given point. There should not be a high-level clearance given to an unrelated employee.

This is crucial as there are many documents that should not be shared with the rest of the employees only the ones that are cleared for them. Unrestricted access would lead to a large amount of data leakage that would disrupt the organization. There can be personal information of the employees that should not be accessible by most of the employees except the Human Resource Manager. The data that is uploaded to any cloud is susceptible to leakage and sensitive documents or files cannot be trusted to be safe on the cloud. Therefore, it is suggested to encrypt the data before uploading it to the cloud to safeguard the data. Encrypted data is difficult to decrypt without the key and would keep the data out of the hands of the attackers. This would stop sensitive information from leaking.

Most of the data is being uploaded to the cloud; there are more and more incidents of information leakage being reported. The cloud stores the data into their servers; it makes various copies that are distributed between one another to facilitate the easy accessibility for the user. As having multiple copies in different servers would lead to a reduction in the time taken to retrieve those documents.

But due to attacks by hackers become more and more sophisticated, it leads to the data that is uploaded to be susceptible to various sources of information leakage. This is undesirable as most of the users would have sensitive information contained in those files that can lead to a lot of distress. Therefore, the data that is being uploaded needs to be encrypted before it is uploaded to the cloud.

As all the data that is being uploaded is being encrypted for security reasons, this makes the search and retrieval process very difficult. As the data is encrypted with a key, it is impossible to search the data that is needed by passing the normal query. This leads to reduced user experience. Therefore, search over encrypted data has to be implemented using certain tags and keywords related to the document in question. The queries and keywords also have to be protected to ensure there is no information leakage from the cloud which could result in a decrease in the security of the database.
In this paper, section 2 is dedicated for literature review of past work and Finally Section 3 concludes this paper.

II. LITERATURE REVIEW

This section of the literature survey eventually reveals some facts based on thoughtful analysis of many authors work as follows.

A. Jivane [1] explains that data privacy has been of the utmost concern as most of the critical documents have been saved on the cloud. To increase the security of the files being stored on the cloud has been encrypted. Therefore, a technique has to be developed to search for encrypted data to be able to efficiently handle the data. The researchers have developed an innovative technique that allows for a multi-keyword search over the encrypted on the cloud while maintaining its privacy. The presented technique works very efficiently and achieves various privacy parameters.
Research Gap – The researcher'splan to use certain keyword semantics to achieve a better ranking system for the results and utilize various other algorithms.

M. Strizhov [2] expresses concerns over the use of the cloud for storage as a viable solution that is being used extensively by the consumers. The authors are concerned with the security of the data that is being stored on untrusted servers. Therefore, the researchers have proposed a novel technique for the searchable encryption of the data so that a consumer can upload his/her file with an untrusted cloud with encryption and still be able to search over it. The authors have implemented a technique that can perform a multi-keyword ranked search over the encrypted data being stored on the cloud.
Research Gap – the authors plan to use substring searches over the encrypted data and would implement the genetic databases that increase the security of the system as a whole.

S. Mittal proclaims that one of the most used aspects of cloud computing is the cloud storage. This is one of the most widely available and convenient techniques of storage. This is due to the fact that cloud is one of the fastest developing technology and most of the organizations are readily outsourcing their operations on to the cloud. Therefore, the researchers

have developed a technique for efficient and fast searches on encrypted data. This technique utilizes the fuzzy scheme in addition to the inclusion of synonyms. The method has been experimented and proved to be highly effective. [3]

Research Gap – syntactic transformations could be used to improve the user experience on the search. The keywords can be used to widen the focus and consider relevant plurals and subjects to improve the ranking of the system.

P. Ponnusamy [4] has presented various techniques that have been used for the preservation of privacy while searching for files on the cloud. This is necessary as the files on the cloud need to be secured as the service provider is not trustworthy, it could also be susceptible to various forms of attacks that can be done on a remote server. Traditionally most of the data used to be decrypted in order to be able to search for the relevant file. This used to eat up a lot of time and resources. Therefore, multi-keyword search over encrypted data stored in the cloud is predominantly used.

Research Gap – This is a very nascent level of development in this direction and further research is needed to improve the workings of the various methods. Further advancement needs to keep the queries safe and also develop an efficient ranking system.

P. Pandiaraja [5] discusses the problems faced while storing and searching encrypted data on to the cloud. Most of the application is concerned with the privacy of their data when the data is being stored on the cloud publicly. The researchers have presented a novel scheme for the preservation of privacy on the cloud and searching mechanism that uses a multi-keyword query that can be used over the encrypted data. This is achieved without the system recognizing the data in the trapdoors or the keywords while searching.

Research Gap – The Apriori algorithm that is being used for providing the search mechanism for the encrypted data has a lot of limitations that can be overcome by utilizing a much more efficient algorithm.

X. Shiconveys that cloud is an insecure storage mechanism and if the data you want to store on to the cloud, it needs to be encrypted in order to increase its privacy. As encrypted data is essentially very difficult to retrieve, it is almost impossible to search for encrypted data. Therefore, the authors have developed a fuzzy keyword searching technique that can efficiently search over the encrypted data efficiently. The researchers utilized datasets in Chinese and English for this purpose and the technique performed with greater precision than the traditional methods. [6]

Research Gap – The technique has been validated on local storage and its feasibility while performing plaintext search operations. And this was compared to the proposed method. The algorithm has performed exceptionally.

P. Sreekumari [7] expresses that cloud is one of the most rapidly developing areas of technology that have been used predominantly for the purpose of universally accessible storage. This has led to very significant concerns about the security of the data being stored. Therefore, the authors have surveyed some of the most ground-breaking research that hasproposed multi-keyword privacy preserving technique involving fuzzy methods for the ranking systems.

Research Gap – The authors have presented various parameters that will be addressed in the future for a more robust technique involving factors, such as, verifiability, efficiency, and security.

P. Kale [8] explores the popularity of cloud computing, especially the exponential rise in the number of individuals storing their data on to the cloud for a more ubiquitous storage capability. As most of the data can contain sensitive information, it is encrypted before outsourcing it. This makes searching on the encrypted data rather difficult. Therefore, the authors have proposed an innovative technique for the implementation of ranked keyword search that increases the usability and accuracy of the system.

Research Gap – IMEI numbers can be used to authenticate the user for the purpose of search and retrieval of the encrypted data from the cloud. Further research is required for this implementation.

R. Ma expresses concern over the rapid increase in the infrastructure of cloud services for the purpose ofof document storage. The authors are concerned about the security concerns that are posed when a critical file is being stored onto an untrusted cloud service. Therefore, most of the cloud-based servers encrypt the files before uploading them onto the cloud. The researchers, therefore, present an efficient method for the retrieval of the data and have named it as EnDAS (Encrypted Data Search Scheme). It is a very lightweight and efficient scheme and has been tested extensively and produces promising results. [9]

Research Gap – The technique is very specific and its applications are concentrated for implementation on mobile devices. This technique needs to be generalized to be able to support various other systems in the future.

S. Lavnis [10] explores the area of cloud computing, especially in the context of cloud storage as it has been steadily increasing in number. A lot of people are storing their data on the cloud. Most of this data is encrypted due to security concerns, but this makes it very difficult to perform searches and retrieve the data. Therefore, the authors have proposed COS2 (Contextual oblivious Similarity-based Search), which utilizes the browsing cache and various other information for the searches and keeps everything confidential throughout.

Research Gap – the researches have only implemented the basic version of their system. They aim to introduce classical features such as multiple keywords and a ranking system for the results.

M. Shen [11] introduces the mechanism of phrase search which is used for the retrieval of files that contain the phrase passed as the query. The mechanism is very essential for the implementation of retrieval for the cloud-based IoT devices. As most of the information being stored on the cloud are encrypted, there is a rising need for a mechanism for search on encrypted data. Therefore, the authors propose an innovative technique called P3 (Privacy Preserving Phrase search). This technique has been proven to greatly enhance the searching ability of the system.

Research Gap – The indices of the documents needs to refreshed after every query or retrieval; this leads to a lot of overhead processing for indexing whole dataset at every retrieval.

Z. Xia explains that there has been a widespread increase in the usage of cloud computing. A lot of individuals have been readily adopting the technology and there has been an exponential increase in the number of people outsourcing their data onto the cloud. This introduces security concerns and the best way to mitigate is to encrypt the data before uploading. But this makes search9ing and organizing the data cumbersome. Therefore, the authors propose a secure multi-keyword ranked search technique that can perform the update and delete operations with the implementation of a kNN algorithm for the query vectors and indexing.

Research Gap – There are a lot of issues pertaining to the implementation of the updating and deletion mechanisms as they are not that secure and might be introduced to some unwarranted leakage of critical data.

III. CONCLUSION

As discussed earlier the need of the search over encrypted data is to allocate the search results without compromising the security of the data. This research paper analyses all the past works and come to conclusion, that there is still lot to achieve the perfection in this area.

So this paper decides to deal with the token generation technique which is eventually an encrypted query for the given keywords. Then these tokens are used for hashing the searching keywords, which generate the resulted output with the list of data that are matched to the given query. This work of ours will be reflected in our coming edition of research articles.

REFERENCES

[1] Anjali Baburao Jivane, “Time Efficient Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data”, IEEE International Conference on Power, Control, Signals, and Instrumentation Engineering, 2017.

[2] Mikhail Strizhov, “Towards a Practical and Efficient Search over Encrypted Data in the Cloud”, IEEE International Conference on Cloud Engineering, 2015.

[3] S. Mittal and R. Krishna, “Privacy-Preserving Synonym Based Fuzzy Multi-Keyword Ranked Search Over Encrypted Cloud Data”, International Conference on Computing, Communication and Automation, 2016.

[4] P. Ponnusamy and R. Vidhyapriya, “A Survey on Multi-Keyword Ranked Search Manipulations over Encrypted Cloud Data”, International Conference on Computer Communication and Informatics, 2017.

[5] P. Pandiaraja and P. Kumar, “Efficient Multi-keyword Search Over Encrypted Data in Untrusted Cloud Environment” Second International Conference on Recent Trends and Challenges in Computational Models, 2017.

[6] X. Shi and S. Hu, “Fuzzy Multi-Keyword Query on Encrypted Data in the Cloud”, 4th International Conference on Applied Computing and Information Technology/3rd International Conference on Computational Science/Intelligence and Applied Informatics/1st International Conference on Big Data, Cloud Computing, Data Science & Engineering, 2016.

[7] P. Sreekumari, “Privacy-Preserving Keyword Search Schemes over Encrypted Cloud Data: An Extensive Analysis”, 4th IEEE International Conference on Big Data Security on Cloud, 2018.

[8] P. Kale and R. Wadekar, “A Survey on Different Techniques for Encrypted Cloud Data”, International Conference on Intelligent Computing and Control Systems, 2017.

[9] R. Ma, J. Li, H. Huan, M. Xia and X. Liu, “EnDAS: Efficient Encrypted Data Search as a Mobile Cloud Service”, IEEE Transactions on Emerging Topics in Computing, 2015.

[10] S. Lavnis, D. Elango and H. Velez, "Contextual Oblivious Similarity Searching for Encrypted Data on Cloud Storage Services", IEEE 8th International Symposium on Cloud and Service Computing, 2018.

[11] M. Shen, B. Ma, L. Zhu, X. Du and K. Xu, "Secure Phrase Search for Intelligent Processing of Encrypted Data in Cloud-Based IoT", IEEE Internet of Things Journal, 2018.

[12] Z. Xia, X. Wang, X. Sun and Q. Wang, "A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE Transactions on Parallel and Distributed Systems, Vol. 27, No. 2, February 2016.