PROTECTED INFORMATION MIGRATION AND REMOVAL VIA ENUMERATION BLOOMING FILTER IN CLOUD-BASED COMPUTING

[1] *Mrs. B. Rajasri,* [2] *N.Pranavi Reddy,* [3] *N.Amulya Reddy,* [4] *M.Surya Lacchith*
[1] Assistant *Professor,*[234]*B.Tech Students*
Department Of Computer Science & Engineering
Sri Indu College Of Engineering & Technology,Sheriguda, Ibrahimpatnam

ABSTRACT
With the rapid development of cloud storage, an increasing number of data owners prefer to outsource their data to the cloud server, which can greatly reduce the local storage overhead. Because different cloud service providers offer distinct quality of data storage service, e.g., security, reliability, access speed and prices, cloud data transfer has become a fundamental requirement of the data owner to change the cloud service providers. Hence, how to securely migrate the data from one cloud to another and permanently delete the transferred data from the original cloud becomes a primary concern of data owners. To solve this problem, we construct a new counting Bloom filter-based scheme in this paper. The proposed scheme not only can achieve secure data transfer but also can realize permanent data deletion. Additionally, the proposed scheme can satisfy the public verifiability without requiring any trusted third party. Finally, we also develop a simulation implementation that demonstrates the practicality and efficiency of our proposal.

Key words — Cloud storage, Data deletion, Data transfer, Counting Bloom filter, Public verifiability.

I. INTRODUCTION

Cloud computing, an emerging and very promising computing paradigm, connects largescale distributed storage resources, computing resources and network bandwidths together[1,2]. By using these resources, it can provide tenants with plenty of high-quality cloud services. Due to the attractive advantages, the services (especially cloud storage service) have been widely applied[3,4], by which the resource-constraint data owners can outsource their data to the cloud server, which can greatly reduce the data owners' local storage overhead[5,6]. According to the report of Cisco[7], the number of Internet consumers will reach about 3.6 billion in 2019, and about 55 percent of them will employ cloud storage service.

Because of the promising market prospect, an increasing number of companies (e.g., Microsoft, Amazon, Alibaba) offer data owners cloud storage service with different prices, security, access speed, etc. To enjoy more suitable cloud storage service, the data owners might change the cloud storage service providers. Hence, they might migrate their outsourced data from one cloud to another, and then delete the transferred data from the original cloud. According to Cisco[7], the cloud traffic is expected to be 95% of the total traffic by the end of 2021, and almost 14% of the total cloud traffic will be the traffic between different cloud data centers. Foreseeably, the outsourced data transfer will become a fundamental requirement from the data owners' point of view.

To realize secure data migration, an outsourced data transfer app, Cloudsfer[8], has been designed utilizing cryptographic algorithm to prevent the data from privacy disclosure in the transfer phase. But there are still some security problems in processing the cloud data migration and deletion. Firstly, for saving network bandwidth, the cloud server might merely migrate part of the data, or even deliver some unrelated data to cheat the data owner[9]. Secondly, because of the network instability, some data blocks may lose during the transfer process. Meanwhile, the adversary may destroy the transferred data blocks[10]. Hence, the transferred data may be polluted during the migration process. Last but not least, the original cloud server might maliciously reserve the transferred data for digging the implicit benefits[11]. The data reservation is unexpected from the data owners' point of view. In short, the cloud storage service is economically attractive, but it inevitably suffers from some serious security challenges, specifically for the secure data transfer,

integrity verification, verifiable deletion. These challenges, if not solved suitably, might prevent the public from accepting and employing cloud storage service.

Contributions In this work, we study the problems of secure data transfer and deletion in cloud storage, and focus on realizing the public verifiability. Then we propose a counting Bloom filterbased scheme, which not only can realize provable data transfer between two different clouds but also can achieve publicly verifiable data deletion. If the original cloud server does not migrate or remove the data honestly, the verifier (the data owner and the target cloud server) can detect these malicious operations by verifying the returned transfer and deletion evidences. Moreover, our proposed scheme does not need any Trusted third party (TTP), which is different from the existing solutions. Furthermore, we prove that our new proposal can satisfy the desired design goals through security analysis. Finally, the simulation experiments show that our new proposal is efficient and practical.

II. LITERATURE SURVEY

1. Practical Techniques for Searches on Encrypted Data

It is desirable to store data on data storage servers such as mail servers and file servers in encrypted form to reduce security and privacy risks. But this usually implies that one has to sacrifice functionality for security. For example, if a client wishes to retrieve only documents containing certain words, it was not previously known how to let the data storage server perform the search and answer the query without loss of data confidentiality. In this paper, we describe our cryptographic schemes for the problem of searching on encrypted data and provide proofs of security for the resulting crypto systems. Our techniques have a number of crucial advantages. They are provably secure: they provide provable secrecy for encryption, in the sense that the untrusted server cannot learn anything about the plaintext when only given the ciphertext; they provide query isolation for searches, meaning that the untrusted server cannot learn anything more about the plaintext than the search result; they provide controlled searching, so that the untrusted server cannot search for an arbitrary word without the user's authorization; they also support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the server. The algorithms we present are simple, fast (for a document of length , the encryption and search algorithms only need stream cipher and block cipher operations), and introduce almost no space and communication overhead, and hence are practical to use today.

2. Smart cloud search services: verifiable keyword-based semantic search over encrypted cloud data

With the increasing popularity of the pay-as-you-consume cloud computing paradigm, a large number of cloud services are pushed to consumers. One hand, it brings great convenience to consumers who use intelligent terminals; on the other hand, consumers are also facing serious difficulties that how to search the most suitable services or products from cloud. So how to enable a smart cloud search scheme is a critical problem in the consumer-centric cloud computing paradigm. For protecting data privacy, sensitive data are always encrypted before being outsourced. Although the existing searchable encryption schemes enable users to search over encrypted data, these schemes support only exact keyword search, which greatly affects data usability. Moreover, these schemes do not support verifiability of search result. In order to save computation cost or download bandwidth, cloud server only conducts a fraction of search operation or return a part of result, which is viewed as selfish and semi-honest-but-curious. So, how to enhance flexibility of encrypted cloud data while supporting verifiability of search result is a big challenge. To tackle the challenge, a smart semantic search scheme is proposed in this paper, which returns not only the result of keyword-based exact match, but also the result of keyword-based semantic match. At the same time, the proposed scheme supports the verifiability of search result. The rigorous security analysis and performance analysis show that the proposed scheme is secure under the proposed model and effectively achieves the goal of keyword-based semantic search.

Pay-as-you-consume cloud computing paradigm

has become more and more prevalent, due to its benefits for consumers, including a large number of convenient service, relief of the burden for storage, flexible data access, reduction of cost on hardware and software. A lot of companies have set up and provided various cloud computing services. More and more sensitive data from consumers (e.g., photo albums, emails, personal health records and financial transactions, etc.) have been centralized into the cloud for its flexible management and economic savings. Meanwhile, many technical schemes related to cloud computing service are proposed by researchers. Noh et al. proposed a flexible communication bus model for multimedia services in cloud environment. Shahnaza et al.proposed a realistic IEEE 802.11e EDCA model for QoS-aware differentiated multimedia mobile cloud services. Cabarcos et al. proposed a middleware architecture that allows sessions initiated from one device to be seamlessly transferred to a second one under a cloud environment.

3. Achieving effective cloud search services: multi-keyword ranked search over encrypted cloud data supporting synonym query

Cloud computing becomes increasingly popular. To protect data privacy, sensitive data should be encrypted by the data owner before outsourcing, which makes the traditional and efficient plaintext keyword search technique useless. The existing searchable encryption schemes support only exact or fuzzy keyword search, not support semantics-based multi-keyword ranked search. In the real search scenario, it is quite common that cloud customers' searching input might be the synonyms of the predefined keywords, not the exact or fuzzy matching keywords due to the possible synonym substitution (reproduction of information content) and/or her lack of exact knowledge about the data. Therefore, synonym-based multi-keyword ranked search over encrypted cloud data remains a very challenging problem. In this paper, for the first time, we propose an effective approach to solve the problem of synonym-based multi-keyword ranked search over encrypted cloud data. We make contributions mainly in two aspects: synonym-based search for supporting synonym query and multi-keyword ranked search for achieving more accurate search result. Two secure schemes are proposed to meet privacy requirements in two threat models of known ciphertext model and known background model. In enhanced scheme, the sensitive frequency information can be well protected by introducing some dummy keywords, which is not adopted in basic scheme. We give security analysis to justify the correctness and privacy-preserving guarantee of the proposed schemes. Extensive experiments on real-world dataset validate our analysis and show that our proposed solution is very efficient and effective in supporting synonym-based searching.

Cloud computing is a new model of enterprise IT infrastructure that provides on-demand high quality applications and services from a shared pool of configuration computing resources [1]. However, there may be existed unauthorized operation on the outsourced data on account of curiosity or profit. To protect the privacy of sensitive information and combat unauthorized accesses, sensitive data should be encrypted by the data owner before outsourcing [2]. However, encrypted data make the traditional data utilization service based on plaintext keyword search useless. The simple and awkward method of downloading all the data and decrypting locally is obviously impractical, because the data owner and other authorized cloud customers must hope to search their interested data rather than all the data. What's more, taking the potentially huge number of outsourced data and great deal of cloud customers into consideration, it is also difficult to meet both the requirements of performance and system usability [30]. Hence, it is an especially important thing to explore privacy-preserving and effective search service over encrypted outsourced data.

4. Efficient semantic search over encrypted data in cloud computing

Cloud storage has become more and more popular as it provides many benefits over traditional storage solutions. Despite the many benefits provided by cloud storage, many security problems have also arisen in cloud storage, which prevents companies from migrating their data to cloud storage. As a result, the owners encrypt their sensitive data before storing it in cloud storage. While encryption increases the security of the data, it also reduces the searchability of the data and

thus, the efficiency of the search. Recently, research has been done on several schemes which enable keyword searching on encrypted data in cloud computing. However, these schemes contain weaknesses which make them impractical when applied to real-life scenarios. In this paper, we developed a system to support semantic search on encrypted data in cloud computing with three different schemes which are "Synonym-Based Keyword Search (SBKS)", "Wikipedia-Based Keyword Search (WBKS)", and "Wikipedia-Based Synonym Keyword Search (WBSKS)". Our results demonstrated that our schemes are more efficient in terms of performance and storage requirements than the former proposed schemes. Therefore, our developed schemes are more practical than the former proposed schemes.

Cloud storage has become a preferred method of storage as it provides many benefits over traditional storage solutions. With cloud storage, corporations can purchase only the needed amount of storage from the cloud storage provider (CSP) to fulfill their storage needs instead of maintaining their own data storage infrastructures. They can rely on CSP to handle all data maintenance tasks such as backup and recovery. It also allows all data to be accessed remotely in order to streamline their operations among different locations. With all these benefits, companies can significantly reduce their operation costs by simply outsourcing their business data to cloud storage.

5. Semantic search supporting similarity ranking over encrypted private cloud data

With the advent of cloud computing, more and more information data are outsourced to the public cloud for economic savings and ease of access. However, the privacy information has to be encrypted to guarantee the security. To implement efficient data utilization, search over encrypted cloud data has been a great challenge. The existing solutions depended entirely on the submitted query keyword and didn't consider the semantics of keyword. Thus the search schemes are not intelligent and also omit some semantically related documents. In view of the deficiency, as an attempt, we propose a semantic expansion based similar search solution over encrypted cloud data. Our solution could return not only the exactly matched files, but also the files including the terms semantically related to the query keyword. In the proposed scheme, a corresponding file metadata is constructed for each file. Then both the encrypted metadata set and file collection are uploaded to the cloud server. With the metadata set, the cloud server builds the inverted index and constructs semantic relationship library (SRL) for the keywords set. After receiving a query request, the cloud server first finds out the keywords that are semantically related to the query keyword according to SRL. Then both the query keyword and the extensional words are used to retrieve the files. The result files are returned in order according to the total relevance score. Eventually, detailed security analysis shows that our solution is privacy-preserving and secure under the previous searchable symmetric encryption (SSE) security definition. Experimental evaluation demonstrates the efficiency and effectives of the scheme.

Cloud Computing enables cloud customers to enjoy the on-demand high quality applications and services from a centralized pool of configurable computing resources. This new computing model can relieve the burden of storage management, allow universal data access with independent geographical locations, and avoid capital expenditure on hardware, software, and personnel maintenances, etc.

As cloud computing becomes mature, lots of sensitive data is considered to be centralized into the cloud servers, e.g. personal health records, secret enterprise data, government documents, etc. The straightforward solution to protect data privacy is to encrypt sensitive data before being outsourced. Unfortunately, data encryption, if not done appropriately, may reduce the effectiveness of data utilization. Typically, a user retrieves files of interest to him/her via keyword search instead of retrieving back all the files. Such keyword-based search technique has been widely used in our daily life, e.g. Google plaintext keyword search. However, the technologies are invalid after the keywords are encrypted.

In recent years, searchable encryption (SE) techniques have been developed for secure outsourced data search. Some further researches focus on search efficiency, multi-keyword search, and secure dynamic updating. But they only support exact keyword search. To enhance the

search flexibility and usability, some research has been done on fuzzy keyword search. These solutions support tolerance of minor typos and format inconsistencies, such as, search for "million" by carelessly typing "milion", or "datamining" by typing "data-mining". These schemes mainly take the structure of terms into consideration and use edit distance to evaluate the similarity. They didn't consider the terms semantically related to query keyword, thus many related files are omitted. In addition, these fuzzy systems send back all relevant files solely upon presence/absence of the keyword, and result-ranking is still out of considering.

In this paper, from a new perspective, we propose a similar search solution based on semantic query expansion while supporting similarity ranking. Semantic expansion based similar search reinforces the system usability by returning the exactly matched files and the files including the terms semantically related to the query keyword. In the proposed scheme, a corresponding file metadata is constructed for each file. Then the encrypted metadata set and file collection are uploaded to the cloud server. With the metadata set, the cloud server builds the inverted index and constructs semantic relationship library (SRL) for the keywords set. The co-occurrence of terms is used to evaluate the semantic relationship between terms in SRL. Upon receiving a query request, the cloud server automatically finds out the terms which are semantically related to the query keyword according to the value of semantic relationship between terms in SRL. Then both the keyword and the semantically expanded words are used to retrieve files. Finally, the matched files are returned in order according to the total relevance score. In the process, to ensure security and final result ranking, we properly modify a crypto primitive order-preserving encryption to protect the relevance score. Detailed security analysis shows that the solution correctly realizing the goal of semantic search, while preserving the privacy. Extensive experimental evaluation demonstrates the efficiency and effectives of the scheme..

III. SYSTEM ANALYSIS & DESIGN

EXISTING SYSTEM

We study the problems of secure data transfer and deletion in cloud storage, and focus on realizing the public verifiability. Then we propose a counting Bloom filter-based scheme, which not only can realize provable data transfer between two different clouds but also can achieve publicly verifiable data deletion. If the original cloud server does not migrate or remove the data honestly, the verifier (the data owner and the target cloud server) can detect these malicious operations by verifying the returned transfer and deletion evidences. Moreover, our proposed scheme does not need any Trusted third party (TTP), which is different from the existing solutions. Furthermore, we prove that our new proposal can satisfy the desired design goals through security analysis.

Finally, the simulation experiments show that our new proposal is efficient and practical.

DISADVANTAGES OF EXISTING SYSTEM

- False Positives/Negatives: Enumeration Bloom Filters can result in false positives, causing unintended data removal, and false negatives, leaving sensitive data behind during migration or removal.
- Limited Accuracy: The approach may struggle to achieve high precision and recall rates, leading to potential data remnants or unnecessary data movement.
- Performance Impact: Enumeration Bloom Filters introduce computational overhead, impacting system performance and scalability, especially in large-scale cloud environments.

PROPOSED SYSTEM

We aim to achieve verifiable data transfer between two different clouds and reliable data deletion in cloud storage. Hence, three entities are included in our new construction,

In our scenario, the resource-constraint data owner might outsource his large-scale data to the cloud server A to greatly reduce the local storage overhead. Besides, the data owner might require the cloud A to move some data to the cloud B, or delete some data from the storage medium. The cloud A and cloud B provide the data owner with cloud storage service. We assume that the cloud

A is the original cloud, which will be required to migrate some data to the target cloud B, and remove the transferred data. However, the cloud A might not execute these operations sincerely for economic reasons. because they belong to two

different companies. Hence, the two clouds will independently follow the protocol. Furthermore, we assume that the target cloud B will not maliciously slander the original cloud A.

ADVANTAGES OF PROPOSED SYSTEM

- Data confidentiality. The outsourced file may contain some private information that should be kept secret. Hence, to protect the data confidentiality, the data owner needs to use secure algorithms to encrypt the file before uploading it to the cloud server.
- Data integrity. The cloud A might only migrate part of the data, or deliver some unrelated data to the cloud B. Besides, the data might be polluted during the transfer process. Hence, the data owner and the cloud B should be able to verify the transferred data integrity to guarantee that the transferred data is intact.
- Public verifiability. The cloud A may not move the data to the cloud B or delete the data faithfully. So, the verifiability of the transfer and deletion results should be satisfied from the data owner's point of view.

SYSTEM ARCHITECTURE

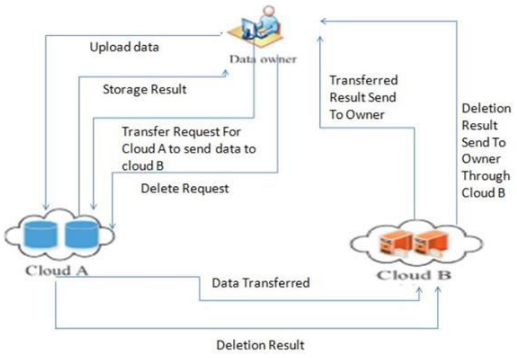
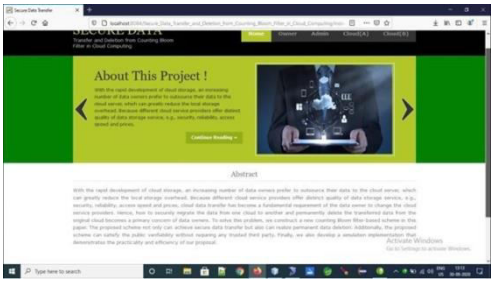


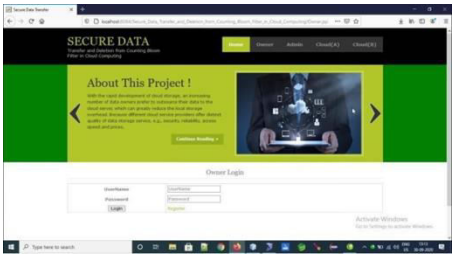
Fig: System Architecture

IV. SCREENSHOTS

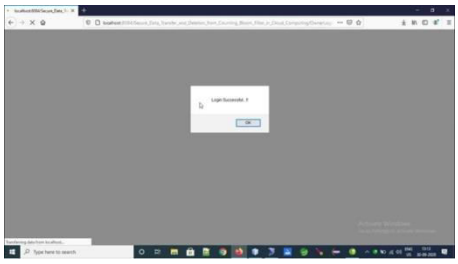
Home screen



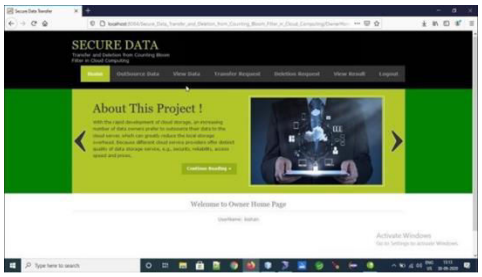
Owner login



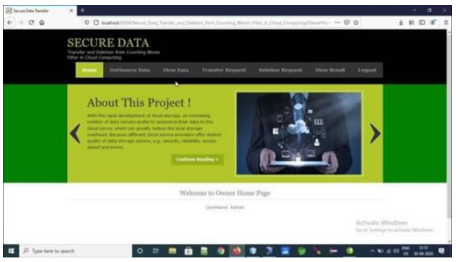
Login status



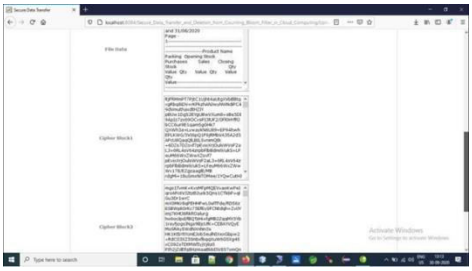
Owner home screen



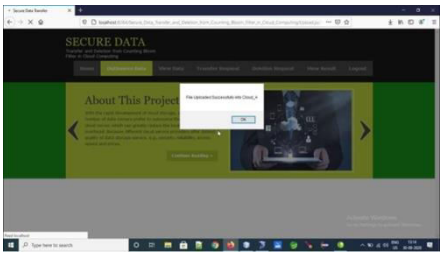
Out source data



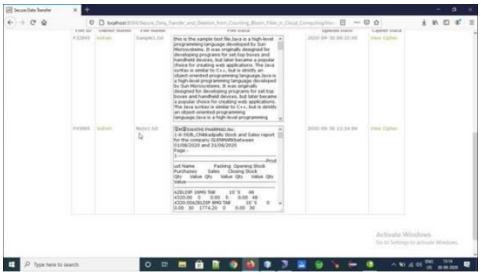
Divide data into blocks



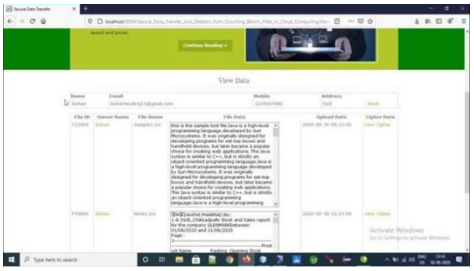
Upload status



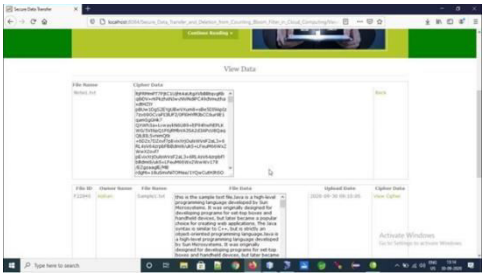
View data



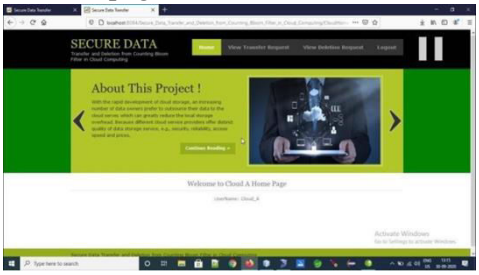
Selected user details



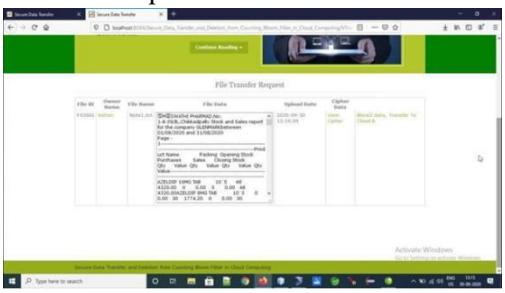
Select file cipher data



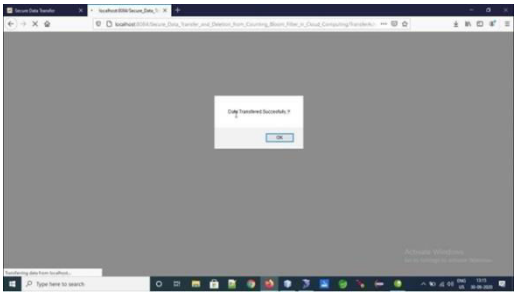
Cloud a home page



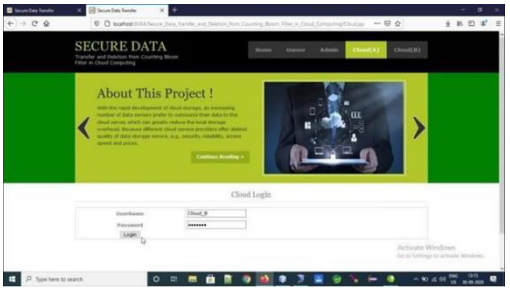
View transfer request



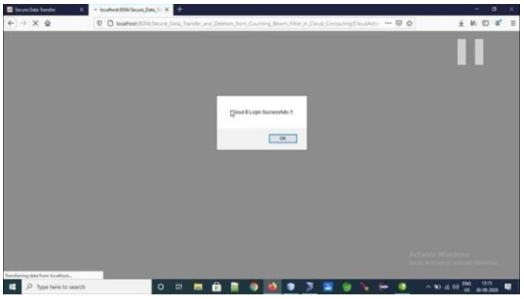
Data transfer status



Cloud_b login



login status



V. CONCLUSION

CONCLUSION

In cloud storage, the data owner does not believe that the cloud server might execute the data transfer and deletion operations honestly. To solve this problem, we propose a CBF-based secure data transfer scheme, which can also realize verifiable data deletion. In our scheme, the cloud B can check the transferred data integrity, which can guarantee the data is entirely migrated. Moreover, the cloud A should adopt CBF to generate deletion evidence after deletion, which will be used to verify the deletion result by the data owner. Hence, the cloud A cannot behave maliciously and cheat the data owner successfully. Finally, the security analysis and simulation results validate the security and practicability of our proposal, respectively.

Future work Similar to all the existing solutions, our scheme considers the data transfer between two different cloud servers. However, with the development of cloud storage, the data owner might want to simultaneously migrate the outsourced data from one cloud to the other two or more target clouds. However, the multi-target clouds might collude together to cheat the data owner maliciously. Hence, the provable data migration among three or more clouds requires our further exploration.

REFERENCES

1. C. Yang and J. Ye, “Secure and efficient fine-grained data access control scheme in cloud computing”, Journal of High Speed

Networks, Vol.21, No.4, pp.259–271, 2015.
2. X. Chen, J. Li, J. Ma, et al., "New algorithms for secure outsourcing of modular exponentiations", IEEE Transactions on Parallel and Distributed Systems, Vol.25, No.9, pp.2386– 2396, 2014.
3. P. Li, J. Li, Z. Huang, et al., "Privacy-preserving outsourced classification in cloud computing", Cluster Computing, Vol.21, No.1, pp.277–286, 2018.
4. B. Varghese and R. Buyya, "Next generation cloud computing: New trends and research directions", Future Generation Computer Systems, Vol.79, pp.849–861, 2018.
5. W. Shen, J. Qin, J. Yu, et al., "Enabling identity-based integrity auditing and data sharing with sensitive information hiding for secure cloud storage", IEEE Transactions on Information Forensics and Security, Vol.14, No.2, pp.331–346, 2019.
6. R. Kaur, I. Chana and J. Bhattacharya J, "Data deduplication techniques for efficient cloud storage management: A systematic review", The Journal of Supercomputing, Vol.74, No.5, pp.2035–2085, 2018.
7. Cisco, "Cisco global cloud index: Forecast and methodology, 2014–2019", available at: https://www.cisco.com/c/en/us-/solutions/collateral/service-provider/global-cloud-index- gci/ white-paper-c11-738085.pdf, 2019-5-5.
8. Cloudsfer, "Migrate & backup your files from any cloud to any cloud", available at: https://www.cloudsfer.com/, 2019-5-5.
9. Y. Liu, S. Xiao, H. Wang, et al., "New provable data transfer from provable data possession and deletion for secure cloud storage", International Journal of Distributed Sensor Networks, Vol.15, No.4, pp.1–12, 2019.
10. Y. Wang, X. Tao, J. Ni, et al., "Data integrity checking with reliable data transfer for secure cloud storage", International Journal of Web and Grid Services, Vol.14, No.1, pp.106–121, 2018.
11. Y. Luo, M. Xu, S. Fu, et al., "Enabling assured deletion in the cloud storage by overwriting", Proc. of the 4th ACM International Workshop on Security in Cloud Computing, Xi'an, China, pp.17–23, 2016.
12. C. Yang and X. Tao, "New publicly verifiable cloud data deletion scheme with efficient tracking", Proc. of the 2th International Conference on Security with Intelligent Computing and Big-data Services, Guilin, China, pp.359–372, 2018.
13. Y. Tang, P.P Lee, J.C. Lui, et al., "Secure overlay cloud storage with access control and assured deletion", IEEE Transactions on Dependable and Secure Computing, Vol.9, No.6, pp.903–916, 2012.
14. Y. Tang, P.P.C. Lee, J.C.S. Lui, et al., "FADE: Secure overlay cloud storage with file assured deletion", Proc. of the 6th International Conference on Security and Privacy in Communication Systems, Springer, pp.380-397, 2010.
15. Z. Mo, Y. Qiao and S. Chen, "Two-party fine-grained assured deletion of outsourced data in cloud systems", Proc. of the 34th International Conference on Distributed Computing Systems, Madrid, Spain, pp.308–317, 2014.