INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 2 (2017) PAGES 1742 – 1747
RECEIVED: 08-03-17 PUBLISHED: 22-03-17

March 22, 2017

SEARCH OVER ENCRYPTED DATA USING HYBRID TREE BASED SECURE METHOD IN CLOUD

[1]G.Sri Harini, [2]Mr.N.Karthik,
[1]UG Scholar, Dept of computer science and Engineering, Sri Balaji Chockalingam Engineering College, Irumbedu, Arni,
[2]Assistant Professor, Dept of computer science and Engineering, Sri Balaji Chockalingam Engineering College, Irumbedu, Arni.

Abstract

In cloud computing, data owners may share their outsourced data with a number of users, who might want to only retrieve the data files they are interested in. One of the most popular ways to do so is through keyword-based retrieval.We propose a new searchable encryption scheme, in which novel technologies in cryptography community and IR community are employed, including homomorphic encryption and the vector space model. The data owner encrypts the searchable index with homomorphic encryption. When the cloud server receives a query consisting of multi-keywords, it computes the scores from the encrypted index stored on cloud and then returns the encrypted scores of files to the data user. Next, the data user decrypts the scores and picks out the top-k highest- scoring files identifers to request to the cloud server. The retrieval takes a two-round communication between the cloud server and the data user. Specifically, the vector space model and the widely-used TFIDF model are combined in the index construction and query generation. We construct a special tree-based index structure and propose a "Greedy Depth-first Search" algorithm to provide efficient multi-keyword ranked search. The secure K-nn algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. In order to resist statistical attacks, phantom terms are added to the index vector for blinding search results. The scheme, the privacy-preserving multi-keyword ranked search over encrypted data in cloud computing MRSE scheme, in which ranking is done at the user side while scoring calculation is done at the server side.

Index Terms: Searchable encryption, Multi-keyword ranked search, Dynamic update.

1. INTRODUCTION

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. 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 are 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

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 2 (2017) PAGES 1742 – 1747
RECEIVED: 08-03-17 PUBLISHED: 22-03-17
March 22, 2017

access with location independence, and avoidance of capital expenditure on hardware, software, and personnel maintenances, etc., Cloud computing, or something being in the cloud, is an expression used to describe a variety of different types of computing concepts that involve a large number of computers connected through a real-time communication network such as the Internet. Cloud computing is mostly used to sell hosted services in the sense of application service provisioning that run client server software at a remote location. End users access cloud-based applications through a web browser, thin client or mobile app while the business software and user's data are stored on servers at a remote location.

Cloud providers typically bill IaaS services on a utility computing basis: cost reflects the amount of resources allocated and consumed. In the PaaS model, cloud providers deliver a computing platform, typically including operating system, programming language execution environment, database, and web server. In the business model using software as a service (SaaS), users are provided access to application software and databases. Cloud providers manage the infrastructure and platforms that run the applications. SaaS is sometimes referred to as "on-demand software" and is usually priced on a pay-per-use basis. SaaS providers generally price applications using a subscription fee. A category of cloud services where the capability provided to the cloud service user is to use network/transport connectivity services and/or inter-cloud network connectivity services. NaaS involves the optimization of resource allocations by considering network and computing resources as a unified whole.

Private cloud is cloud infrastructure operated solely for a single organization, whether managed internally or by a third-party and hosted internally or externally. A cloud is called a 'Public cloud' when the services are rendered over a network that is open for public use. Technically there may be little or no difference between public and private cloud architecture. Hybrid cloud is a composition of two or more clouds (private, community or public) that remain unique entities but are bound together, offering the benefits of multiple deployment models. Such composition expands deployment options for cloud services, allowing IT organizations to use public cloud computing resources to meet temporary needs. This capability enables hybrid clouds to employ cloud bursting for scaling across clouds.

2. RELATED WORK

Predicate encryption is a new encryption paradigm which gives a master secret key owner fine-grained control over access to encrypted data. The master secret key owner can generate secret key tokens corresponding to predicates. An encryption of data x can be evaluated using a secret token corresponding to a predicate f; the user learns whether the data satisfies the predicate, i.e., whether f(x) = 1. We present two fully secure functional encryption schemes: a fully secure attribute-based encryption (ABE) scheme and a fully secure(attribute-hiding) predicate encryption (PE) scheme for inner-product predicates. In both cases, previous constructions were only proven to be selectively secure. Both results use novel strategies to adapt the dual system encryption methodology introduced by Waters.

We define and solve the problem of secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undifferentiated results, and further ensures the file retrieval accuracy. Specifically, we explore the statistical measure approach, i.e. relevance score, from information retrieval to build a secure searchable index, and develop a one-to-many order-preserving mapping technique to properly protect those sensitive score information. The proved to satisfy adaptive semantic security definition. We also

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 2 (2017) PAGES 1742 – 1747
RECEIVED: 08-03-17 PUBLISHED: 22-03-17
March 22, 2017

combine an effective ranking capability that is based on term frequencyinversedocument frequency (TF-IDF) values of keyword documentpairs. Our analysis demonstrates that the proposed scheme isproved to be privacy-preserving, efficient and effective. We design and implement dynamic symmetric searchable encryption schemes that efficiently and privately search server-held encrypted databases with tens of billions ofrecord-keyword pairs. Our basic theoretical construction support single-keyword searches and offers asymptotically optimalserver index size, fully parallel searching, and minimal leakage.

Our implementation effort brought to the fore several factorsignored by earlier coarse-grained theoretical performance analyses,including low-level space utilization, I/O parallelism andgoodput.
We accordingly introduce several optimizations to ourtheoretically optimal construction that model the prototype's characteristics designed to overcome these factors.All of ourschemes and optimizations are proven secure and the informationleaked to the untrusted server is precisely quantified.

3. SECURITY OF FILES

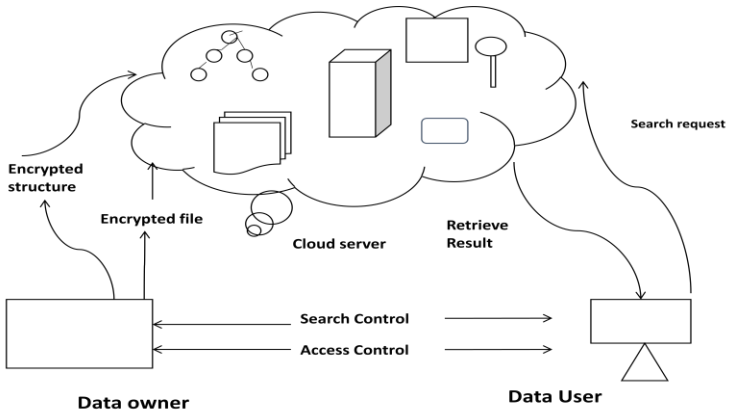

Fig.1.Security of files

4. TECHNICAL RIDE

Uploading data

The data owner has a collection of n files to outsource onto the cloud server in encrypted form and expects the cloud server to provide keyword retrieval service to data owner himself or other authorized users. To achieve this, the data owner needs to build a searchable index from a collection of keywords extracted out of files, and then outsources both the encrypted index and encrypted files onto the cloud server.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 2 (2017) PAGES 1742 – 1747
RECEIVED: 08-03-17 PUBLISHED: 22-03-17
March 22, 2017

Encryption of data

We propose a new searchable encryption scheme, in which novel technologies in cryptography community are employed, including data encryption and the vector space model. In the proposed scheme, the data owner encrypts the searchable index with data encryption. When the cloud server receives a query consisting of multi keywords, it computes the scores from the encrypted index stored on cloud and then returns the encrypted scores of files to the data user. Next, the data user decrypts the scores and picks out the top-k highest-scoring files' identifiers to request to the cloud server. The retrieval takes a two-round communication between the cloud server and the data user.

Data Search

The data user is authorized to process multi keyword retrieval over the outsourced data. The computing power on the user side is limited, which means that operations on the user side should be simplified. The authorized data user at first generates a query. For privacy consideration, which keywords the data user has searched must be concealed. Thus, the data user encrypts the query and sends it to the cloud server that returns the relevant files to the data user. Afterward, the data user can decrypt and make use of the files.

Top k result

Considering the large number of data users and documents in the cloud, it is necessary to allow multikey-word in the search query and return documents in the order of their relevancy with the queried keywords. Scoring is a natural way to weight the relevance. Based on the relevance score, files can then be ranked in either ascendingly or descendingly. Several models have been proposed to score and rank files in IR community. Among these schemes, we adopt the most widely used one TF-IDF weighting, which involves two attributes-term frequency and inverse document frequency.

5. FUNCTIONAL RIDE

k-nearest neighbors algorithm (K-NN)

In pattern recognition, the k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression: In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. If k = 1, then the object is simply assigned to the class of that single nearest neighbor. In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors.Euclidian Distance of D(a,b)=$\sqrt{\sum_k}$(ak-bk)^2.

k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The k-NN algorithm is among the simplest of all machine learning algorithms.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 2 (2017) PAGES 1742 – 1747
RECEIVED: 08-03-17 PUBLISHED: 22-03-17
March 22, 2017

6. PERFORMANCE GRAPH

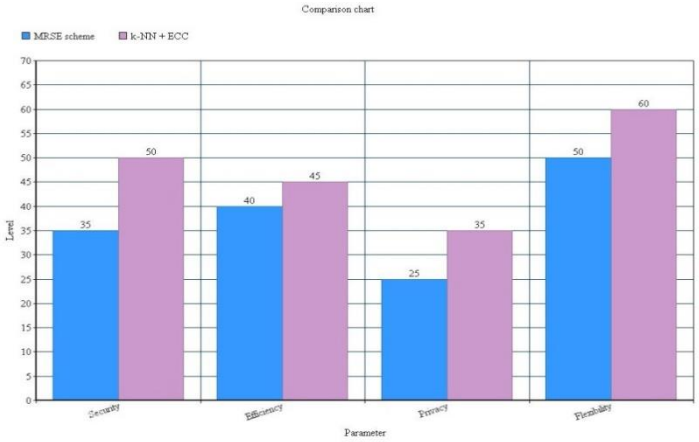

Fig 2: Performance graph

The performance of ourtechnique is evaluated regarding the efficiency of existingMRSE schemes, as well as the tradeoff betweensearch precision, privacy and computation time cost for k-word retrieval. Showing the problem of Secured Multi-keyword search over encrypted cloud data. Propose two schemes following the principle of coordinate matching and inner product similarity. This module is used to help the user to get the accurate result based on the multiple keyword concepts. The users can enter the multiple words query, the server is going to split that query into a single word after search that word file in our database. Finally, display the matched word list from the database and the user gets the file from that list. When any User request for the data then Ranking is done on requested data using k-nearest.

CONCLUSION

In this project, we motivate and solve the problem of secure multikeyword top-k retrieval over encrypted cloud data. We define similarity relevance and scheme robustness. Based on OPE invisibly leaking sensitive information, we devise a server-side ranking SSE scheme. We then propose a MRSE scheme, which fulfills the security requirements of multikeyword top-k retrieval over the encrypted cloud data. By security analysis, we show that the proposed scheme guarantees data privacy. According to the efficiency evaluation of the proposed scheme over a real data set, extensive results demonstrate that our scheme ensures practical efficiency. We first propose secure inner data computation. Also we achieve effective ranking result using k-nearest neighbour technique. This system is currently work on single cloud.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 2 (2017) PAGES 1742 – 1747
RECEIVED: 08-03-17 PUBLISHED: 22-03-17
March 22, 2017

FUTURE ENHANCEMENT

In future is will extended up to sky computing & Provide better security in multi-user systems. The system is designed to solve the problem of supporting efficient ranked keyword search for achieving effective utilization of remotely stored encrypted data in Cloud Computing.

REFERENCES

[1]M.Armbrust,A.Fox,R.Griffith,A.Joseph,R.Katz,A.Konwinski, G. Lee, D. Patterson, A. Rabkin, and M. Zaharia, "AView of Cloud Computing," Comm. ACM, vol. 53, no. 4, pp. 50-58,2010.
[2] M. Arrington, "Gmail Disaster: Reports of Mass Email Deletions Dec. 2006.
[3] Amazon.com, "Amazon s3 Availability Event:July20,2008,"http://status.aws.amazon.com/s3-20080720.html, 2008.
[4] RAWA News, "Massive Information Leak Shakes Washingtonover Afghan War," http://www.rawa.org/temp/runews/2010/08/20/massive-information-leak-shakes-washington-over-afghan-war.html, 2010.
[5] AHN, "Romney Hits Obama for Security Information Leakage,"http://gantdaily.com/2012/07/25/romney-hits-obama-for-security-information-leakage/, 2012.
[6] Cloud Security Alliance, "Top Threats to Cloud Computing,"http://www.cloudsecurity alliance.org, 2010.
[7] C. Leslie, "NSA Has Massive Database of Americans' PhoneCalls," http://usatoday30.usatoday.com/news/washington/2006-05-10/, 2013.
[8] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "SearchableSymmetric Encryption: Improved Definitions and Efficient Constructions," Proc. ACM 13th Conf. Computer and Comm. Security(CCS), 2006.
[9] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure RankedKeyword Search over Encrypted Cloud Data," Proc. IEEE 30th Int'lConf. Distributed Computing Systems (ICDCS), 2010.
[10] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Top-kRetrieval from a Confidential Index," Proc. 12th Int'l Conf.Extending Database Technology: Advances in Database Technology(EDBT), 2009.