ISSN XXXX XXXX © 2017 IJESC

Research Article Volume 7 Issue No.7

Secure Data Mining in Cloud using Homomorphism Encryption

M.Gokula Priya
M.phil Scholar
Department of Computer Science
Annai Vailankanni Arts and Science College, Thanjavur, Tamilnadu, India

Abstract:
Cloud Computing can resolve the problem of handling, storage and analyzing the Big Data as it distributes the big data within the cloudlets. No doubt, Cloud Computing is the best answer available to the problem of Big Data storage and its analyses but having said that, there is always a potential risk to the security of Big Data storage in Cloud Computing, which needs to be addressed. Data Privacy is one of the major issues while storing the Big Data in a Cloud environment. Data Mining based attacks, a major threat to the data, allows an adversary or an unauthorized user to infer valuable and sensitive information by analyzing the results generated from computation performed on the raw data. Cloud Computing enables cloud customers to remotely store their data into the cloud so as to enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources. The benefits brought by this new computing model include but are not limited to: relief of the burden for storage management, universal data access with independent geographical locations, and avoidance of capital expenditure on hardware, software, and personnel maintenances, etc.

Keyword: cloud computing, data privacy, big data, cloud storage, user infer valuable, storage management; treat based on attack, SOA.

1. INTRODUCTION:

1.1Cloud computing:

Cloud computing is the result of evolution and adoption of existing technologies and paradigms. The goal of cloud computing is to allow users to take benefit from all of these technologies, without the need for deep knowledge about or expertise with each one of them. The cloud aims to cut costs, and help the users focus on their core business instead of being impeded by IT obstacles. The main enabling technology for cloud computing is virtualization. Virtualization software separates a physical computing device into one or more "virtual" devices, each of which can be easily used and managed to perform computing tasks. With operating system–level virtualization essentially creating a scalable system of multiple independent computing devices, idle computing resources can be allocated and used more efficiently. Virtualization provides the agility required to speed up IT operations, and reduces cost by increasing infrastructure utilization. Autonomic computing automates the process through which the user can provision resources on-demand. By minimizing user involvement, automation speeds up the process, reduces labor costs and reduces the possibility of human errors. Users routinely face difficult business problems. Cloud computing adopts concepts from Service-oriented Architecture (SOA) that can help the user break these problems into services that can be integrated to provide a solution. Cloud computing provides all of its resources as services, and makes use of the well-established standards and best practices gained in the domain of SOA to allow global and easy access to cloud services in a standardized way. Cloud computing also leverages concepts from utility computing to provide metrics for the services used. Such metrics are at the core of the public cloud pay-per-use models. In addition, measured services are an essential part of the feedback loops in autonomic computing, allowing services to scale on-demand and to perform automatic failure recovery. Cloud computing is a kind of grid computing; it has evolved by addressing the QoS (quality of service) and reliability problems. Cloud computing provides the tools and technologies to build data/compute intensive parallel applications with much more affordable prices compared to traditional parallel computing techniques.

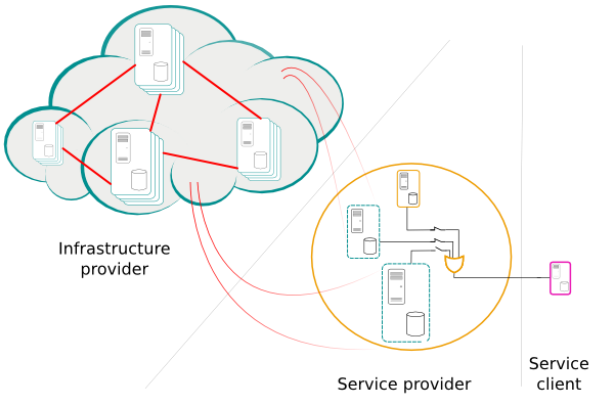

Figure.1. Cloud computing

1.1.1 Architecture of cloud computing

The systems architecture of the software systems involved in the delivery of cloud computing, typically involves multiple cloud components communicating with each other over a loose coupling mechanism such as a messaging queue. Elastic provision implies intelligence in the use of tight or loose coupling as applied to mechanisms such as these and others.

1.1.1.1 Cloud engineering

Is the application of engineering disciplines to cloud computing? It brings a systematic approach to the high-level concerns of commercialization, standardization, and governance in conceiving, developing, operating and maintaining cloud computing systems. It is a multidisciplinary method encompassing contributions from diverse areas such

as systems, software, web, performance, information,security, platform, risk, and quality engineering.

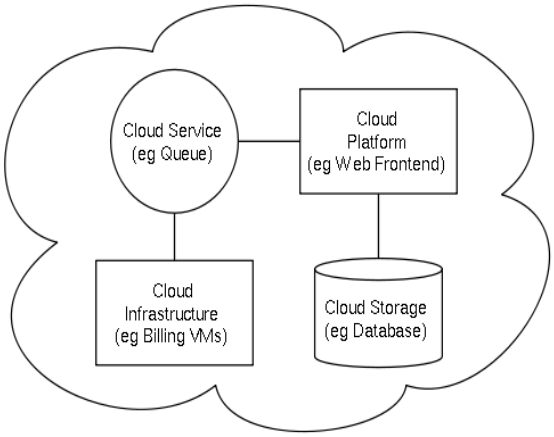

Figure.2. Architecture of cloud computing

1.1.2 Deployment models

1.1.2.1 Public cloud

Public clouds are owned and operated by companies that use them to offer rapid access to affordable computing resources to other organisations or individuals. With public cloud services, users don't need to purchase hardware, software or supporting infrastructure, which is owned and managed by providers.

2. EXISTING SYSTEM:

Security and privacy is the major issue concerning the clients as well as the providers of cloud services as a lot of confidential and sensitive data is stored in cloud which can provide valuable information to an attacker.This system stored the data centralized database so, easily attacker. Cloud computing is the long dreamed vision of computing as a utility, where cloud customers can remotely store their data into the cloud so as to enjoy the on-demand high-quality applications and services from a shared pool of configurable computing resources. Its great flexibility and economic savings are motivating both individuals and enterprises to outsource their local complex data management system into the cloud. To protect data privacy and combat unsolicited accesses in the cloud and beyond, sensitive data, for example, e-mails, personal health records, photo albums, tax documents, financial transactions, and so on, may have to be encrypted by data owners before outsourcing to the commercial public cloud.

3. PROPOSED SYSTEM

The proposed approach can further be extended by adding a digital signature or hashing technique to authenticate the third party so as to prevent an adversary from posing as the third party to host's. This approach assumes that the data is not stored in a centralized location but is distributed to various hosts. Cloud Computing is the evolving technology that has changed the way of computing in IT Enterprise. It brings the software and data to the centralized data centers from where a large community of users can access information on pay per use basis. This poses security threats over the data stored. Data confidentiality may be compromised which has to be taken care of. So it becomes necessary to encrypt the data before outsourcing it to the cloud server. This makes data utilization a challenging task. Traditional searching mechanisms provide Boolean search to search over encrypted data, which is not applicable when the number of users and the number of data files stored in the cloud is large. They also impose two major issues, one being the post-processing that has to be done by the users to find the relevant document in need and the other is the network traffic that is undesirable in present scenario when all the files matching with keywords is retrieved. But this project proposes ranked keyword search that overcomes these issues. We have proposed an efficient scheme which enables the Cloud Service Provider (CSP) to determine the files that are related to the keywords searched by the user, rank them and send the most relevant files without knowing any information about the cloud. Our schema consists of three entities: Data owner, Un-trusted cloud server and local trusted server. The data owner is the one whose data is stored in cloud server and he is also authorized to search over his files. Cloud server is an un-trusted server which provides storage service where data owners store their documents in encrypted form. The trusted local server stores the index that is created for the files. The problem that we consider is privacy-preserving keyword search on private database model, where the documents are simply encrypted with the secret keys unknown to the actual holder of the database (i.e. Cloud Server). We consider three roles coherent with previous works Data owner, who is the actual owner of the database. The data owner collects and/or generates the information in the database and lacks the means (or is unwilling) to maintain/operate the database, Users are the members in a group who are entitled to access (part of) the information of the database, Server is a professional entity (e.g. cloud) to offer information services to authorized users. It is often required that the server is oblivious to content of the database it maintains, the search terms in queries and documents retrieved. We assume that the parties are semi-honest and do not collude with each other to bypass the security measures. In an offline stage, the data owner creates a search index for each document. The search index file is created using a secret key based trapdoor generation function where the secret keys1 are only known by the data owner. Then, the data owner uploads these search index files to the server together with the encrypted documents. We use symmetric-key encryption as the encryption method since it can handle large document sizes efficiently. This process is referred as the index generation henceforth and the trapdoor generation is considered as its one of the steps. To protect data privacy, confidential data has to be encrypted before outsourcing, so as to provide end-to-end data confidentiality assurance in the cloud. Data encryption makes effective data utilization a very challenging task given that there could be a large amount of outsourced data files. Besides, in Cloud Computing, data owners may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files they are interested in during a given session.

4. CONCLUSION & FUTURE WORK:

Security and privacy is the major issue concerning the clients as well as the providers of cloud services as a lot of confidential and sensitive data is stored in cloud which can provide valuable information to an attacker. This paper proposes a method to solve the privacy issues of the cloud. In this paper, we motivate and solve the problem of efficient and secure ranked multi-keyword search on remotely stored encrypted database model where the database users are protected against privacy violations. We first define the security requirements for the given problem. We then employ a secure usage of the method given for practical application

scenarios where total number of keywords that can be searched is relatively limited and there are only few search terms in a query by using a trapdoor based system where the trapdoor can only be generated by the data owner. We appropriately increase the efficiency of the scheme by using symmetric-key encryption method rather than public-key encryption for document encryption. We also propose to use the blinded encryption technique in accessing the contents of the retrieved documents without revealing them to other parties. We prove that our proposed method satisfies the security requirements. The proposed ranking method proves to be efficient to return highly relevant documents corresponding to submitted search terms. We implement the entire scheme and extensive experimental results on the implementation demonstrate the effectiveness and efficiency of our solution. Following the current research, there are possible improvements and undergoing eorts that will appear in the future work. Firstly, the user side of proposed system will be implemented on mobile devices running operating systems since the potential application scenario envisions that users access the data anywhere and anytime. And secondly, the proposed method will be tested on a real dataset in order to compare the performance of our ranking method with the ranking methods used in plain datasets that do not involve any security or privacy-preserving techniques.

5. REFERENCES:

[1]. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, pp. 829-837, Apr, 2011.

[2]. L.M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A Break in the Clouds: Towards a Cloud Definition," ACM SIGCOMM Comput. Commun.Rev., vol. 39, no. 1, pp. 50-55, 2009.

[3]. N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou, "LT Codes-Based Secure and Reliable Cloud Storage Service," Proc. IEEE INFOCOM, pp. 693-701, 2012.

[4]. S. Kamara and K. Lauter, "Cryptographic Cloud Storage," Proc. 14th Int'l Conf. Financial Cryptograpy and Data Security, Jan. 2010.

[5]. A. Singhal, "Modern Information Retrieval: A Brief Overview," IEEE Data Eng. Bull., vol. 24, no. 4, pp. 35-43, Mar. 2001.

[6]. I.H. Witten, A. Moffat, and T.C. Bell, Managing Gigabytes: Compressing and Indexing Documents and Images. Morgan Kaufmann Publishing, May 1999.

[7]. D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.

[8]. E.-J. Goh, "Secure Indexes," Cryptology ePrint Archive, http:// eprint.iacr.org/2003/216. 2003.

[9]. Y.-C. Chang and M. Mitzenmacher, "Privacy Preserving Keyword Searches on Remote Encrypted Data," Proc. Third Int'l Conf. Applied Cryptography and Network Security, 2005.

[10]. R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," Proc. 13th ACM Conf. Computer and Comm. Security (CCS '06), 2006.

[11]. D. Boneh, G.D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public Key Encryption with Keyword Search," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), 2004.

[12]. M. Bellare, A. Boldyreva, and A. ONeill, "Deterministic and Efficiently Searchable Encryption," Proc. 27th Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.

[13]. M. Abdalla, M. Bellare, D. Catalano, E. Kiltz, T. Kohno, T. Lange, J. Malone-Lee, G. Neven, P. Paillier, and H. Shi, "Searchable Encryption Revisited: Consistency Properties, Relation to Anonymous Ibe, and Extensions," J. Cryptology, vol. 21, no. 3, pp. 350- 391, 2008.

[14]. J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy Keyword Search Over Encrypted Data in Cloud Computing," Proc. IEEE INFOCOM, Mar. 2010.

[15]. D. Boneh, E. Kushilevitz, R. Ostrovsky, and W.E.S. III, "Public Key Encryption That Allows PIR Queries," Proc. 27th Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYP TO '07), 2007.

[16]. P. Golle, J. Staddon, and B. Waters, "Secure Conjunctive Keyword Search over Encrypted Data," Proc. Applied Cryptography and Network Security, pp. 31-45, 2004.

[17]. L. Ballard, S. Kamara, and F. Monrose, "Achieving Efficient Conjunctive Keyword Searches over Encrypted Data," Proc. Seventh Int'l Conf. Information and Comm. Security (ICICS '05), 2005.

[18]. D. Boneh and B. Waters, "Conjunctive, Subset, and Range Queries on Encrypted Data," Proc. Fourth Conf. Theory Cryptography (TCC), pp. 535-554, 2007.

[19]. R. Brinkman, "Searching in Encrypted Data," PhD thesis, Univ. of Twente, 2007.

[20]. Y. Hwang and P. Lee, "Public Key Encryption with Conjunctive Keyword Search and Its Extension to a Multi-User System," Pairing, vol. 4575, pp. 2-22, 2007.