ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 435- 438 (2014)
Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

Privacy Preserving Data Storage of Intermediate Data Sets over Cloud

[1]T.N.S.Satyavani, [2]Cheekatla Swapna Priya
[1]Final M.Tech Student, [2]Associate Professor
[1,2]Dept of Computer Science and Engineering
[1,2]Pydah College of Engineering, Visakhapatnam, AP, India

Abstract: Cloud computing relies on sharing of resources to achieve coherence and economies of scale, similar to a utility over a network. Security should be provided when sharing of cloud data. In this paper we are using ID3 for classification of intermediate data set to store data into cloud. Before storing the data into cloud the data owner should encrypt the data by using cryptography technique. This paper describes for data encryption and decryption by using Triple DES algorithm. After completion of encryption the data owner will store data into cloud. If any user required that data he/she retrieve data from the cloud and decrypt that data. By using encryption and decryption those techniques reduces cost and time complexity.

INTRODUCTION

Computing is being transformed to a which model consisting of services that are commoditized and delivered in a manner similar to traditional utilities such as water, electricity, gas, and telephony. In such a model, users access services based on their requirements regardless to where the services are hosted or how they are delivered. Several computing paradigms have promised to deliver this utility computing vision which include cluster computing, Grid computing, and Cloud computing. The term Cloud defines the infrastructure as a "Cloud" from which businesses and users are able to access applications from anywhere in the world on demand. Thus, the computing world is rapidly transforming towards developing software for millions to consume as a service, rather than to run on their individual computers. At present, it is common to access content across the Internet independently without reference to the underlying hosting infrastructure. This infrastructure consists of data centers that are monitored and maintained around the clock by content providers. Cloud computing is an extension of this paradigm wherein the capabilities of business applications are exposed.

Cloud service providers are incentivized by the profits to be made by charging consumers for accessing these services. Consumers, such as enterprises are attracted by the opportunity for reducing or eliminating costs associated with "in-house" provision of these services. However, cloud applications may be crucial to the core business operations of the consumers, it is essential that the consumers have guarantees from providers on service delivery. Typically these are provided through Service Level Agreements (SLAs) brokered between the providers and consumers. Providers such as Amazon, Google, Sales force, IBM, Microsoft, and Sun Microsystems have begun to establish new data centers for hosting Cloud computing applications in various locations around the world to provide redundancy and ensure reliability in case of site failures. Since user requirements for cloud services are varied, service providers have to ensure that they can be flexible in their service delivery while keeping the users isolated from the underlying infrastructure. Recent advances in microprocessor technology and software have led to the increasing ability of commodity hardware to run applications within Virtual Machines (VMs) efficiently. VMs allow both the isolation of applications from the underlying hardware and other VMs, and the customization of the platform to suit the needs of the end-user. Providers can expose applications running within VMs, or provide access to VMs themselves as a service (e.g. Amazon Elastic Compute Cloud) thereby allowing consumers to install their own applications. While convenient, the use of VMs gives rise to further challenges such as the intelligent allocation of physical resources for managing competing resource demands of the users.

In addition, enterprise service consumers with global operations require faster response time, and thus save time by distributing workload requests to multiple Clouds in various locations at the same time. This creates the need for establishing a computing atmosphere for dynamically interconnecting and provisioning Clouds from multiple domains within and across enterprises. There are many challenges involved in creating such Clouds and Cloud interconnections.

RELATED WORK

Throughout computer science history, numerous attempts have been made to disengage users from computer hardware needs, from time-sharing utilities envisioned in the 1960s, network computers of the 1990s, to the commercial grid systems of more recent years. This abstraction is steadily becoming a reality as a number of academic and business leaders in this field of science are spiraling towards cloud computing. Cloud computing is an innovative Information System (IS) architecture, visualized as what may be the future of computing, a driving force demanding from its audience to rethink their understanding of operating systems, client–server architectures, and browsers. Cloud computing has leveraged users from hardware requirements, while reducing overall client side requirements and complexity. As cloud computing is achieving increased popularity, concerns are being voiced about the security issues introduced through the adoption of this new model.

The effectiveness and efficiency of traditional protection mechanisms are being reconsidered, as the characteristics of this innovative deployment model, differ widely from them of traditional architectures. In this paper we attempt to demystify the unique security challenges introduced in a cloud environment and clarify

ISSN 2278-3091

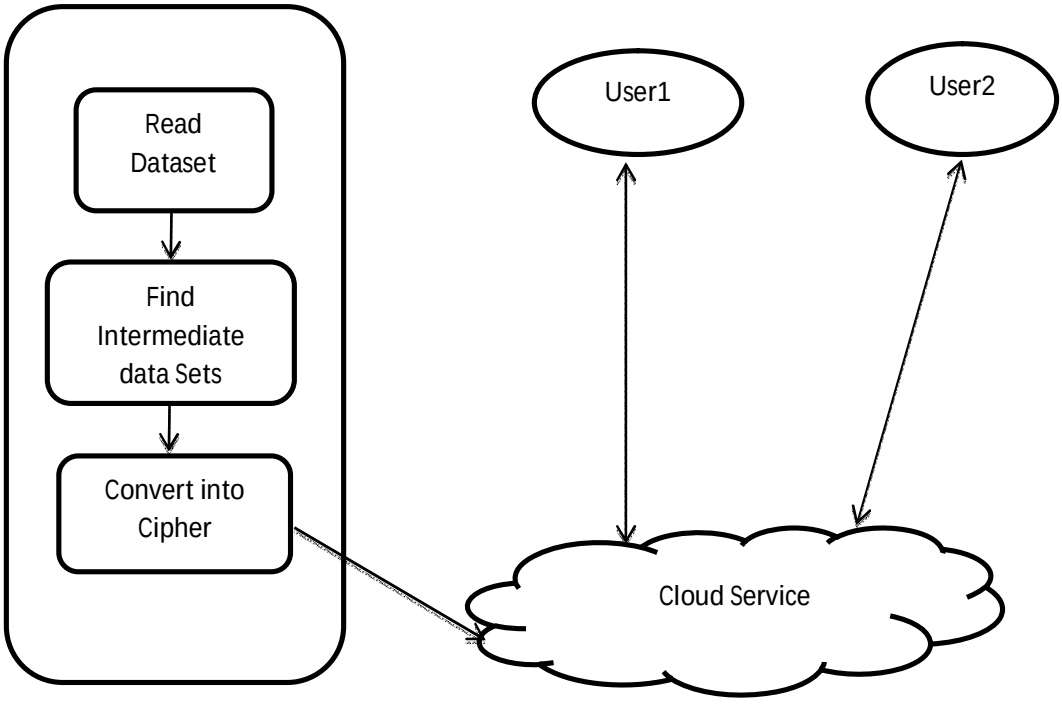
International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 435- 438 (2014)
Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

issues from a security perspective. The notion of trust and security is investigated and specific security requirements are documented. This paper proposes a security solution, which leverages clients from the security burden, by trusting a Third Party. The Third Party is tasked with assuring specific security characteristics within a distributed information system, while realizing a trust mesh between involved entities, forming federations of clouds. The research methodology adopted towards achieving this goal, is based on software engineering and information systems design approaches. The basic steps for designing the system architecture include the collection of requirements and the analysis of abstract functional specifications.

The intention of this paper is to render referential medical information management system. The proposed system deals with query processing in multidimensional data from reference sites for expert analysis, satisfying the relevancy of search. The sensitive data publishing is dealt with anonymized temporal events. The proposed three proven streamlined techniques viz. Automatic Error correction, Topic Relevant Query suggestion, query augmentation helps in précised query contour and quick retrieval of referential data from the patented medical databases in the hospital data mart which holds big data of a medical forum. The proposed system provides the ranked list of reference records when searched for opinion on treatment of acute cases. With careful analysis the patient can be treated with correct and available advanced technologies. This model is based on the notion of differential privacy that is applicable to k-anonymized data sets. It puts forward the inference control strategy to avoid privacy leakage thereby performing double encryption in the datasets based on the severity. Further to support this special technique efficient scheduling of intermediate data sets using temporal pattern matching algorithm is suggested in order to preserve the dataset dynamically.

PROPOSED SYSTEM

Cloud computing relies on sharing of resources to achieve coherence and economies of scale, similar to a utility over a network. At the foundation of cloud computing is the broader concept of converged infrastructure and shared services. In this paper we are using to find the intermediate data sets by using ID3 algorithm. The procedure of ID3 algorithm as follows.



ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 435- 438 (2014)
Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

Data Owner reads the data from dataset and changes the data into intermediate dataset which is classification and then with the key plain text will be convert into Cipher text and finally it send to Cloud Service. User will login and retrieve the data from the cloud service by entering the same key which is used for the conversion of plain to cipher and retrieve the plain text In this Module we classify the testing sample of the image with training datasets, by computing the attribute information gain in terms of entropy, consider the maximum information gain attribute first and construct the decision tree to analyze the testing sample of data.

1) Establish Classification Attribute
2) Compute Classification Entropy.
3) For each attribute in R, calculate Information Gain using classification attribute.
4) Select Attribute with the highest gain to be the next Node in the tree (starting from the Root node).
5) Remove Node Attribute, creating reduced table R_S.
6) Repeat steps 3-5 until all attributes have been used, or the same classification value remains for all rows in the reduced table.

ID3 builds a decision tree from a fixed set of examples and the resulting tree is used to classify future samples and the example has several attributes and belongs to a class (like yes or no) and the leaf nodes of the decision tree contain the class name whereas a non-leaf node is a decision node and the decision node is an attribute test with each branch (to another decision tree) being a possible value of the attribute and ID3 uses information gain to help it decide which attribute goes into a decision node and the advantage of learning a decision tree is that a program rather than a knowledge engineer that elicits knowledge from a final expert.

Gain measures how well a given attribute separates training examples into targeted classes. The only one with the highest information (information being the most useful for classification) is selected to define gain, we first borrow an idea from information theory called entropy and Entropy measures the amount of information in an attribute.

This is the formula for calculating homogeneity of a sample.
Entropy(S)= $\sum_{i=1}^{C}(p_i \log_2 P_i)$

It helps to measure the information gain with respect to the attributes

$$Gain(A) = E(Current\ set) - \sum E(all\ child\ sets)$$

In this module the data owner will read data set from data base. In the data set contains two types i.e. training and testing data sets. After reading training and testing data sets the data owner will classify the required data set. After completion those data sets are intermediate data set. After finding intermediate data set the data owner will encrypt those intermediate data sets and stored into cloud services.

In this module each intermediate data sets are stored into cloud service. Before storing the intermediate data sets the data owner will convert the plain intermediate data sets into cipher intermediate dataset. After converting the data owner will stored intermediate data into cloud services. If any user wants to retrieve those data sets and convert into plain format.

In this module the process of encryption can be done by the data owner and decryption process can be done by the users. The encryption and decryption of intermediate data sets we are using IDEA algorithm. The block cipher IDEA operates with 64-bit plaintext and cipher text blocks and is controlled by a 128-bit key. The fundamental innovation in the design of this algorithm is the use of operations from three different algebraic groups. The substitution boxes and the associated table lookups used in the block ciphers available to-date have been completely avoided. The algorithm structure has been chosen such that, with the exception that different key sub-blocks are used, the encryption process is identical to the decryption process.

The block cipher IDEA operates with 64-bit plaintext and cipher text blocks and is controlled by a 128-bit key. The fundamental innovation in the design of this algorithm is the use of operations from three different algebraic groups. The substitution boxes and the associated table lookups used in the block ciphers available to-date have been completely avoided. The algorithm structure has been chosen such that, with the exception that different key sub-blocks are used, the encryption process is identical to the decryption process.

IDEA supports all modes of operation as described by NIST in its publication FIPS 81. A block cipher encrypts and decrypts plaintext in fixed-size-bit blocks (mostly 64 and 128 bit). For plaintext exceeding this fixed size, the simplest approach is to partition the plaintext into blocks of equal length and encrypt each separately. This method is named Electronic Code Book (ECB) mode. However, Electronic Code Book is not a good system to use with small block sizes (for example, smaller than 40 bits) and identical encryption modes. As ECB has disadvantages in most applications, other methods named modes have been created. They are Cipher Block Chaining (CBC), Cipher Feedback (CFB) and Output Feedback (OFB) modes.

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 435- 438 (2014)
Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

CONCLUSION

The cloud computing is used for sharing the data more than one users. Before share the data set the necessary data set can be stored into cloud. Because of memory consideration if necessary data set can be stored into cloud more useful information can be stored in the cloud. So that in this paper we proposed concept for reduce unnecessary data set and stored required data set in the cloud. For the purpose finding intermediate dataset we are using ID3 algorithm and converting intermediate data set into cipher we are using IDEA algorithm. After storing cipher data into cloud if any user wants that intermediate data he/she retrieve and convert into plain format. By proposing this technique our application is more secure and flexible.

REFERENCES

1. R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, and I. Brandic,"Cloud Computing and Emerging It Platforms: Vision, Hype, andReality for Delivering Computing as the Fifth Utility," FutureGeneration Computer Systems, vol. 25, no. 6, pp. 599-616, 2009.

2. D. Zissis and D. Lekkas, "Addressing Cloud Computing SecurityIssues," Future Generation Computer Systems, vol. 28, no. 3, pp. 583-592, 2011.

3. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-PreservingMulti-Keyword Ranked Search over Encrypted Cloud Data," Proc.IEEE INFOCOM '11, pp. 829-837, 2011

4.P. Samarati, "Protecting Respondents' Identities in MicrodataRelease," IEEE Trans. Knowledge and Data Eng., vol. 13, no. 6,pp. 1010-1027, Nov. 2001.

5.G. Wang, Z. Zutao, D. Wenliang, and T. Zhouxuan, "InferenceAnalysis in Privacy-Preserving Data Re-Publishing," Proc. EighthIEEE Int'l Conf. Data Mining (ICDM '08), pp. 1079-1084, 2008.
1.
6.K.-K. Muniswamy-Reddy, P. Macko, and M. Seltzer, "Provenancefor the Cloud," Proc. Eighth USENIX Conf. File and StorageTechnologies (FAST '10), pp. 197-210, 2010.

7S.B. Davidson, S. Khanna, V. Tannen, S. Roy, Y. Chen, T. Milo, andJ.Stoyanovich, "Enabling Privacy in Provenance-Aware WorkflowSystems," Proc. Fifth Biennial Conf. Innovative Data SystemsResearch (CIDR '11), pp. 215-218, 2011
8 X. Zhang, C. Liu, J. Chen, and W. Dou, "An Upper-Bound ControlApproach for Cost-Effective Privacy Protection of IntermediateData Set Storage in Cloud," Proc. Ninth IEEE Int'l Conf. Dependable,Autonomic and Secure Computing (DASC '11), pp. 518-525, 2011.

9. C. Gentry, "Fully Homomorphic Encryption Using Ideal Lattices,"Proc. 41st Ann. ACM Symp. Theory of Computing (STOC '09),pp. 169-178, 2009.

10.N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-PreservingMulti-Keyword Ranked Search over Encrypted Cloud Data," Proc.IEEE INFOCOM '11, pp. 829-837, 2011.