Enhanced Privacy Preserving with Cost Effectiveness of Information in Cloud

J.VISHWESH[#1] and M.R.YADHUKRISHNA[*2]

[#]Assistant Professor, Dept. of Computer Science, Coorg Institute of Technology, KA, India

[*] Assistant Professor, Dept. of Information Science, Coorg Institute of Technology, KA, India

***Abstract*— Cloud computing is the delivery of computing services over the internet. Cloud services allow individuals and business to use software and hardware that are managed by third parties at remote locations. Processing of such application and data, a large volume of intermediate datasets will be generated and it can be stored into the cloud environment. The security providing for the intermediate data will be carried out by encryption of those intermediate dataset. The security providing for the intermediate data will be carried out by encryption of those intermediate data. Encrypting of all intermediate datasets will be neither efficient nor cost effective because it is very time consuming and costly for data intensive application to encrypt or decrypt the datasets frequently while performing any operation on them. To identify which intermediate datasets need to be encrypted and which do not. This technique increases the performance and reduces the bottleneck in cloud environment.**

***Index Terms*— cloud computing, data storage privacy, privacy preserving, intermediate datasets.**

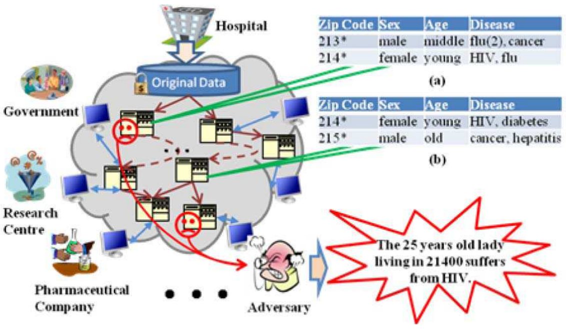
I. Introduction

Cloud computing is the delivery of computing services over the internet. Cloud services allow individuals and business to use software and hardware that are managed by third parties at remote locations. Cloud services include online file storage, social networking sites, webmail, and online business applications. The cloud computing model allows to access information anywhere in the world. Cloud computing provide a share pool of services including data storage space, network, computer processing power and specialized corporate and user applications. Cloud services available in private cloud, public cloud, community cloud. Generally services provided by a public cloud are offered over the internet and operated by a cloud provider. Some examples include services aimed at the generic public, such as online photo storage services, email services or social networking sites. In a private cloud, the users access the data privately. In a community cloud the services is shared by several organizations and made available only to those groups. Cloud computing enables a new business model that support demand, pay for use and economies of scale IT services over the internet. The internet cloud works as a service factory build around virtualization data centres.

Cloud platform are dynamically built through Virtualization with provisioned hardware, software, networks and datasets. The idea is migrate desktop computing to the service oriented platform using virtual server clusters at data centres. Cloud computing have been widely spreading there will be a lots of transaction will be carried out in cloud. Obviously there will be intermediate data which contain the sensitive data during the transaction. Cloud services are popular because they can reduce the cost and complexity of owning and operating computers and networks. Cloud users do not have to invest in information technology infrastructure. Cloud computing offers unlimited processing and storage capacity. The cloud computing model allows access to the information and computer resources from anywhere in the world. The briefly review the research on privacy protection in cloud, Intermediate dataset privacy preserving and privacy-preserving data publishing (PPDP).

Currently, encryption is exploited by most existing Research to ensure the data privacy in cloud. Although encryption works well for data privacy in these approaches, it is necessary to encrypt and decrypt data sets frequently in many applications. Encryption is usually integrated with other methods to achieve cost reduction, high data usability and privacy protection. Proposed system which partitions map reduce computing jobs in terms of the security labels of data they works on and they work assigns the computation without sensitive data to public cloud. The sensitivity of data is required to be labelled to be labelled in advance to make the above approaches available. A Proposed an approach that combines encryption and data fragmentation to achieve privacy protection for distributed data storage with encrypting only part of data sets, follow this line, but integrate data anonymization and encryption together to fulfil cost effective privacy preserving.

The importance of retaining intermediate data sets in cloud has been widely recognized but the research on privacy issues incurred by such data sets just commences. The privacy issues in workflow provenance and high utility of provenance information via carefully hiding a subset of intermediate data. The mainly focuses on data privacy preserving from an economical cost perspective while their concentrates majorly on functionality privacy of workflow modules rather than data privacy. In my research also differs from theirs in several aspects such as data hiding techniques, privacy quantification and cost models. But my approach can be complementarily used for selection of hidden data item in their research if economical cost is considered.



Zip Code	Sex	Age	Disease
213*	male	middle	flu(2), cancer
214*	female	young	HIV, flu

(a)

Zip Code	Sex	Age	Disease
214*	female	young	HIV, diabetes
215*	male	old	cancer, hepatitis

(b)

II. Virtualization

Virtualization is the process of creating virtual environment of some physical environment. The best example will be hard disks in computers. The hard disk is divided into physical hard disks and multiple logical drives. Hardware virtualization is the process of creating virtual machines that will perform like a normal computer with operating system. Here the bare hardware is divided into multiple logical units like our hard disks and we are implementing multiple operating system into the logical hardware units and making all together to run in one time.

A single computer now can run multiple operating systems at the same time using virtualization concept. Software executed on those machines is separated from the underlying hardware resources.

The machine which runs windows xp can host a virtual machine that runs the Ubuntu operating system. The Ubuntu software is run on Ubuntu operating system using the same hardware resources that runs windows xp.in virtualization the host pc is a bare hardware where the visualization take places and guest pc is a virtual machine created by using host pc. The world host and guest are used to distinguish the software runs on physical machine and software runs on virtual machine. The software that creates a virtual machine is also called as the virtual machine manager. Examples are hypervisor or vm ware vsphere. This software creates a multiple logical hardware units on the bare hardware. Logical units are called as virtual machine are physical units are called as bare hardware.

III. Private Cloud Computing

Private cloud refers the IT infrastructure which runs its application software in hosted environment. The private cloud is implemented within the organisation behind the firewall. The private cloud is hosted internally or externally. Even ifs it's hosted externally still it's accessed within the organisation. The drawbacks of private cloud computing is its initial cost of set up. The initial cost of set up includes setting up of client terminals to access the private cloud and its implementation cost which includes mapping proprietary software into hosted private cloud computing environment.

The running cost of private cloud is practically unpredictable because of its changing nature. But the cost will be comparatively lower than public cloud because of low usage of internet bandwidth. In private cloud we can customize the infrastructure environment as per our needs, but in public cloud we are not able to customize the infrastructure and it's a pre-defined suite of public cloud vender.

The private cloud computing environment is secured because the internal data will not be accessed by outside the cloud. The data security mechanism can also be additionally implemented to protect our data. It's a hosted service mostly data security task will be implemented by private cloud computing suites.

The private cloud computing infrastructure easily by adding additional client terminals. Updating client is also simple task because any private cloud computing environment just requires a browser enabled operating system it can be updated easily by remote firmware update.

IV. Existing System

Privacy of intermediate data in cloud is only proceeding by encrypting all the intermediate data set. Most existing applications only run on encrypted data but encryption is costlier. Privacy principles applied only for a single dataset. In recent progress homomorphic encryption for preserving the privacy for cost efficiently. The privacy concerns caused by retaining intermediate data sets in cloud are important but they are paid little attention.

The storage of intermediate data enlarges attack surfaces so that privacy requirements of data holders are at risk of being violated, this enables an adversary to collect intermediate datasets together and menace privacy-sensitive information from them, bringing considerable economic loss or severe social reputation impairment to data owners. But little attention has been paid to such a cloud-specific privacy of datasets stored in cloud mainly include encryption and anonymization. Most of them are only applied to one single data set. Privacy principles for multiple datasets are also proposed, but they aim at specific scenarios such as continuous data publishing or sequential data releasing. Many anonymization techniques like generalization have been proposed to preserve privacy. But these methods alone fail to solve the problem of preserving for multiple datasets.

DRAWBACKS

1. Increase the computation cost in cloud environment.
2. The encryption process will take more time for encrypting the every data in the intermediate data set.
3. Encrypting all intermediate data sets will reduce the performance.

V. PROPOSED SYSTEM

The proposed approach to identify which intermediate datasets need to be encrypted and which do not, so that privacy-preserving cost can be saved while the privacy requirements of data holders can still be satisfied. For preserving privacy of multiple datasets, it is promising to anonymize all datasets first and then encrypt them before storing or sharing them in cloud. Usually, the volume of intermediate data sets is huge. Encrypting all intermediate data sets will reduce the performance.

For preserving privacy of multiple datasets it is promising to anonymize all datasets first and then encrypt them before storing or sharing them in cloud. A tree structure is modeled from generation relationships of intermediate datasets to analyze privacy propagation of datasets.

The heuristic algorithm to identify the datasets that need to be encrypted. Acyclic graph (DAG) is expolited to capture the topological structute of generation relationships among these datasets.the first formally demonstrate the possibility of ensuring privacy leakage requirements without encryting all intermediate datasets when encryption is incorporated with anonymization to preserve privacy. Second process , the heuristic algorithm to identify which data sets need to be encrypted for preserving privacy while the rest of them do not.third experiments results demonstrate that our approach can significantly reduce privacy preserving cost over exisiting approaches, which is quite beneficial for the cloud users who utilize cloud servies in a pay-as you-go fashion.this paper mainly focuses on data privacy preserving from an economical cost perspective while their concentrates majorly on functionality privacy of workflow modules rather than data privacy. This approach can be complementarily used for selection of hidden data items in their research if economical cost is considered and integrates anonymization with encryption to achieve privacy preserving of multiple datasets.

ADVANTAGES

1. Increase the efficiency, because encrypt only intermediate datasets.
2. Reduce the computation cost in cloud environment.
3. Increase the performance and reduce the Bottleneck in the network

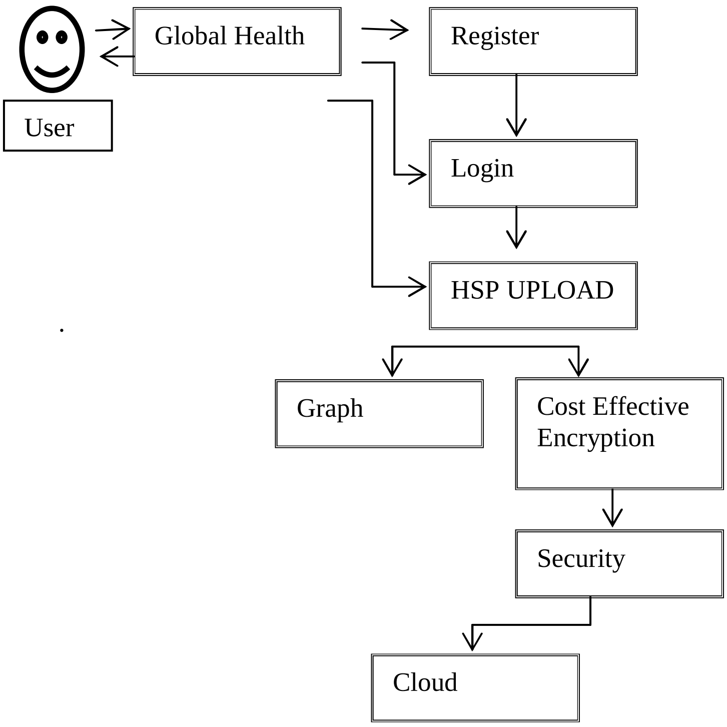


Fig 1: SYSTEM ARCHITECTURE

PRIVACY PRESERVING COST REDUCING HEURISTIC ALGORITHM

Heuristic algorithm identify which intermediate data sets need to be encrypted so that privacy preserving cost can be saved while the privacy requirements of data holders can still be satisfied. for preserving privacy of multiple datasets to anonymize all data sets first and then encrypt them before storing or sharing them in cloud.the volume of intermediate datasets is huge .the encrypting all intermediate data set will lead to high overhead and low efficiency when they are frequently accessed or processed. Data sets are anonymize rather than encrypted to ensure both data utility and privacy preaerving. Acylic graph is exploited to capture the topological structure of generation relationships among these datasets.

COST EFFECTIVE ENCRYPTION

Encrypt only the selected information in the dataset. To identify the data that needs to encrypt. Original data will be encrypted in order to provide the privacy for the data. The remaining data will not be encrypted so that cost spends for whole data encryption will be saved.

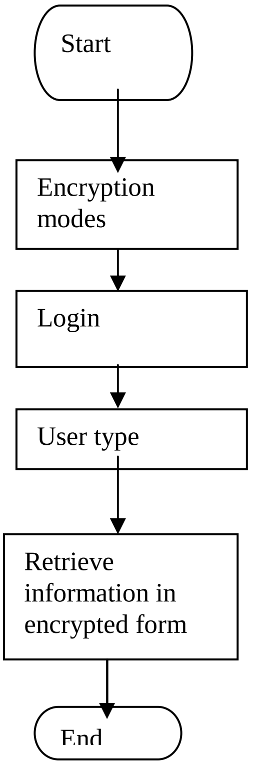


FIG 2 :COST EFFECTIVE ENCRYPTION

VI. CONCLUSION

In this paper proposed an approach that identifies which part of intermediate data sets needs to be encrypted while the rest does not, in order to save the privacy- preserving cost. A tree structure has been modeled from the generation relationships of intermediate data sets to analyze privacy propagation among data sets. The modeled the problem of saving privacy-preserving cost as a constrined optimization problem which is addressed by decomposing the privacy leakage constraints a pratical heuristic algorithm has been designed accordingly. Evaluation results on real world datasets and larger extensive data sets have demonstrated the cost of preserving privacy in cloud cen be reduced significantly with our approach over existing ones where all datasets are encrypted.

REFERENCES

[1] M. Armbrust, A Fox, A.D. joseph, R Katz, A.Konwinski, G.Lee, D.patterson, A. Rabkin, I. Stoica and M. ZAharia, “A View of Cloud Computing”,Comm.

[2] R.Buyya C.S Yeo, S.Venugopal, J.Broberg and I.Bradic, “Cloud computing and Emerging IT Platform: Vision.Hype, and Reality for Delivering Computing as the Fifth Utility” Future Generation Computer Systems”.

[3] L.Wang. J.Zhan, W.Shi, and Y.Liang,”In Cloud Can Scientific Communities Benefit from the Economics of Scale?”.

[4] H.Takabi, J.B.D. Joshi and G. Ahn ”Security and Privacy Challenges in Cloud Computing Environments,”.

[5]D.Zissis and D.Lekkas, Addressing Cloud Computing Security Issues,” Future Generation Computer Systems.

[6]D.YUvan, Y.Yang, X.liu and J.Chen” On Demand Minimum Cost Benchmarking For Intermediate Data Set Storage in Scientific Cloud Workflow System.

[7]”Amazon Cloud Services” http:// aws.amazon.com/.

[8] S.Agarwala, D.Jadav, and L.A Bathen, “iCostale Adaptive Cost Optimization for Storage Clouds,” Proc.IEEE int Conf.Cloud Computing.

[9] R.Bose and J.Frew, ”Lineage Retrival For Scientific Data Processing.

[10] X.Liu, D.Yuan, G.Zhang, W.li, D.Cao Q.He, j.Chen and Y.Yang, The Design of Cloud Workflow Systems. Springer 2012.

[11] Microsoft Health Vault, http:// www.microsoft.com/health/ww/products/pages/healthvault.aspx, july 2012.

[12] J.A Kelner and A. madry “Faster Generation of Randow Spanning Trees”, Proc 50th Ann IEEE Symp Foundation of Computer Science.

[13]k-k Muniswamy-Reddy P.Macko and M.seltzer “Provenance for the cloud”, proc Eigth Usenix conf File and Storage Technologies.

[14] Amazon Web Services,” Aws Service Pricing Overview,” http// aws.amazon.com/pricing/july 2012. [15] H.Lin and W.tzeng,” A secure Erasure Code-Based Cloud Storage System with Secure Data Forwarding,” IEEE Trans Parallel and Distributed System vol 23, no 6,pp 995-1003 june 2012.