Confidentiality and Availability of Data Warehouses in the Cloud Computing System

NDINGA YESO SALAZAR[1*]
Computer and Information Engineering Department, Hohai University 1 Xikang Road,
Nanjing City, Jiangsu Province, China[†]
ndinga_salazar@yahoo.fr[‡]

Hu Jiming[2]
Associate Professor
Computer and Information Engineering Department, Hohai University 1 Xikang Road,
Nanjing City, Jiangsu Province, China[§]
wasersoft@126.com

Abstract

With the advent of cloud computing as a new deployment model of computer systems, data warehouses benefit from this new paradigm. In this context, it becomes necessary to protect these data warehouses of different risks and dangers that are born with cloud computing. Consequently, we propose in this work a way to limit these risks through the algorithm secret key sharing of Shamir and we put this contribution into practice.

1. INTRODUCTION

Several trends are opening up the era of Cloud Computing which is an Internet-based development and use of computer technology. The ever cheaper and more powerful processors, together with the software as a service (SaaS) computing architecture, are transforming data centers into pools of computing service on a huge scale. The increasing network bandwidth and reliable yet flexible network connections make it even possible that users can now subscribe high quality services from data and software that reside solely on remote data centers.

Moving data into the cloud offers great convenience to users since they don't have to care about the complexities of direct hardware management. The pioneer of Cloud Computing vendors, Amazon Simple Storage Service (S3) and Amazon Elastic Compute Cloud (EC2) [1] are both well known as examples. While these internet-based online services do provide huge amounts of storage space and customizable computing resources, this computing platform shift, however, is eliminating the responsibility of local machines for data maintenance at the same time.

As a result, users are at the mercy of their cloud service providers for the availability and Integrity of their data. Recent downtime of Amazon's S3 is such an example [2]. Indeed, the implementation of a data warehousing is in the clouds. For each company it is a good solution for its efficiency and profitability. However, as each technological advance, cloud computing also brings risks, especially in terms of security must be taken into account in order to receive all the benefits of this solution.

According to Mansfield-Devine, traditional systems are protected by firewalls and gateways where cybercriminals must collect critical information to know that they exist. While in cloud computing, systems are highly visible and are designed to be accessible from anywhere. In addition, applications with this type of device can be accessed through a browser; it is an interface whose weaknesses are well known.

The risks of cloud computing increases with the use of virtualization, which is one of the basic techniques in this type of device [3]. Indeed, Zhou and his team have shown that there is a flaw in hypervisor Xen as it enables virtual machines to consume CPU time for other users and allows the theft of service [4]. Wei also addressed the problems of managing the security of virtual machine images [5]. Actually, the dangers of cloud computing are not limited at this stage since we must also ensure that services are available at any time which is not always the case. For example, in 2009 there was a power cut in the Amazon cloud in the local that hosts their servers in Virginia [6]. Such a failure can cause a great loss to businesses since the activity of the company that hosts its infrastructure is stopped.

Several studies have noted that companies are reluctant to cloud computing because it puts outsourcing their data. In fact, the problem that always comes up is the fact of not wanting to leave their data in the hands of competition, something more difficult to control when data is assigned to an external provider whose physical location is often unknown. In the order to optimize the security level that lies on the cloud data warehouse, we need to answer the following questions: Is it reasonable to entrust sensitive and important data to a cloud service provider? How can we ensure that the cloud service provider does not disappear one day? Would our data be erased if we wanted to change cloud service supplier? Is there any risk of losing data stored in the clouds? The transfer of data to the cloud is it secure?

In this paper we first present the state of the art of the cloud computing security and more particularly of data warehouse. This is a summary and critical review of principal works. Next, we describe the proposed solution that solves some problems related to data security, and then we will concretize this solution through the implementation of a prototype and end up with a conclusion and perspectives.

2. STATE OF THE ART

From the technical point of view, cloud computing is essentially using the Internet to meet computing needs. Instead of using a computer to access local services, we simply pass by virtual clouds of cloud computing which are connected on networked computers. The desktop is no longer a thoroughfare to access services that are on the Web. “The clouds” of cloud computing system are essentially a metaphor to popularize the complexity of what happens with the organization of information in virtual networks of Web. But, like any new technology it needs many improvements and the establishment of specific standards [7] to avoid risks. Security is often considered the main obstacle to the adoption of cloud computing services. Thus, numerous works have been devoted to research of solutions for remedying this problem. We will try in this section to present the main research that proposed solutions to ensure the security of cloud computing.

2.1. Security of Cloud Computing

There are a number of security issues, concerns associated with cloud computing but these issues fall into two broad categories: Security issues faced by cloud providers (organizations providing software, platform, or Infrastructure-as-a-Service via the cloud) and security issues faced by their customers [8]. In most cases, the provider must ensure that their infrastructure is secure and that their clients’ data and applications are protected while the customer must ensure that the provider has taken the proper security measures to protect their information.

The extensive use of virtualization in implementing cloud infrastructure brings unique security concerns for customers or tenants of a public cloud service. Virtualization alters the relationship between the OS and underlying hardware - be it computing, storage or even networking. This introduces an additional layer - virtualization - that itself must be properly configured, managed and secured. Specific concerns include the potential to compromise the virtualization software, or "hypervisor". While these concerns are largely theoretical, they do exist. We can classify the aspects of safety in the clouds into three categories which are data security, logical security and physical security. In this paper we are just focused on the data security and logical security.

2.1.1 Access security and data storage in the clouds

Jensen and al, listed the various techniques used in cloud computing for secure access and they have identified weaknesses of these techniques in order to implement their solution which is based on TLS and XML cryptography [9]. This solution is a response to the problem of web browser that has gaps in security. The idea proposed is to use TLS and to adapt the browser by integrating XML cryptography. However, Wang and al, have proposed a solution that is based on erasure correcting code to provide redundancy and ensure the reliability of the data token [10].

They used the homomorphic token for the accuracy of storage and to locate errors. The proposed solution is able to detect corrupted data during storage; it can guarantee the location of erroneous data and identify the server that has a bad behaviour. In the cloud, no assumptions about the robustness of a node can be made. Various unforeseen factors can all lead to a temporary unavailability of certain nodes or inaccessibility of definitive data.

In such cases, the traditional means of data protection are often powerless. Danwei Chen and Yanjun proposed an algorithm that ensures a security algorithm that provides data recovery in case of failure of some servers [11] . It is an algorithm for splitting data. This algorithm is an extension of the fundamental theorem of algebra equation by K, the Shamir Secret Sharing Algorithm is a cryptographic algorithm based on secret sharing (2), the data storage algorithm online of Abhishek (Parakh and Kak 2009) [12] and number theory.

The idea is to split the data into k parts of d= $d_1,d_2,d_3,..d_k$. This division is made by using the separation algorithm to store data on servers subsequently chosen randomly selected noted S = s_1, s_2, s_3, .. s_m with m> k. The process of storing data in the clouds is thus on two stages, the first stage is to divide and store data on a servers chosen arbitrarily and the second stage is the ability to restore data. Through these processes, the data is ready to be transferred, stored, processed, since they are securely encrypted. The researchers concluded that the temporal complexity of the algorithm is to generate the same data block k and for the restoring data.

They proved that even if an attacker invades a storage node, steals a data block and tries to restore the data set, the complexity time required for treatment cannot be supported by today's IT environments. Other benefits that differentiate this proposal include the ability to restore data even if one or more storage nodes are not available which cannot be the case with a traditional solution for cryptography.

2.1.2 Logical security of cloud Computing

In cloud IaaS (Infrastructure as a service), users have access to virtual machine VM [13] on which they can install and run their software. These virtual machines are created and managed by a virtual machine monitor VMM which is a software layer between the physical machine and the operating system. The VMM controls the resources of the physical machine and creates multiple virtual machines that share these resources. Virtual machines have independent operating systems, running independent applications and are isolated from each other by the VMM.This type of device has caused many problems of vulnerability of the virtual machine that led the authors to work in this area to find effective solutions

Zhou and al, proposed a solution to eliminate the vulnerability of the virtual machine [4]. The discovery of the limits of the XEN hypervisor used by AMAZON was their starting point. They proposed four approaches for improving the performance of the hypervisor, which are based on the Poisson, Bernoulli's law, the Uniform Law and finally the exact law. After a comparison between the four new models, they deduced that the strategy based on the Poisson distribution is the best in practice to prevent cycle theft [4].

Wei and al, proposed a system image management of the virtual machine that controls access to and from these images through filters and scanners that can detect and remedy violations using data mining techniques, the system s 'called Mirage (Jinpeng and al. 2009). S. Berger and al, have also developed a technology that responds to problems encountered by the virtual machine. This technology is called Trusted Virtual Data Center (TVDc), it ensures that the workload can be billed to the client who benefited from the service. It also ensures that in the case of some malicious programs like viruses they cannot spread to other nIJuds and it also helps prevent misconfiguration problems. TVDc uses the policy of isolation which is based on the separation of resources used by customers. It manages the data center, access to virtual machines and the passage of a virtual machine to another (Fei and al. 2011).

2.2. Debate

We have presented in the previous section, few works that propose to solve the problems of security in cloud computing. Some approaches seem to be relevant and provide an acceptable level of safety but remains insufficient. In addition, this work has highlighted new problems: Danvei Chen and Yanjun noted the data redundancy. But it does not mean that the idea proposed by these researchers is very reliable for secure data transfer over the network and eliminates the problem of unavailability of services in case of failure of one of the servers.

Jensen proposed using TLS and adapts the browser by integrating XML cryptography. Such a proposal is not sufficient to ensure the safe transfer over the network since it is based on a technology that its shortcomings are well known. While the idea proposed by Wang and his collaborators, it is based on homomorphic encryption which is the answer to the question of data confidentiality during the transfer and during the treatment in the cloud. Nevertheless, the research lab of cryptography in the cloud Microsoft announced that the new ways to encrypt data is still its infancy and they are far from being run on virtual machines data encrypted with homomorphic encryption (4).

Generally existing solutions for secure transfer and processing of data is based on cryptographic data that is not always a complete solution to protect data, in addition the mechanism for encrypting and decrypting data can be intensive on processors which causes a waste of resources, one thing that cloud providers do not want to experience.

3. SECRET SHARING

Secret sharing refers to method for distributing a secret amongst a group of participants, each of whom is allocated a share of the secret. The secret can be reconstructed only when a sufficient number of shares are combined together; individual shares are of no use on their own. More formally, in a secret sharing scheme there is one dealer and n players. The dealer gives a secret to the players, but only when specific conditions are fulfilled. The dealer accomplishes this by giving each player a share in such a way that any group of t (for threshold) or more players can together reconstruct the secret but no group of fewer than t players can.

Such a system is called a (t, n)-threshold scheme (sometimes it is written as an (n, t)-threshold scheme). Secret sharing was invented independently by Adi Shamir and George Blakley in 1979. Each secret share is a plane, and the secret is the point at which three shares intersect. Two shares yield only a line intersection. For more comprehension we can take a look on the figure1 below.

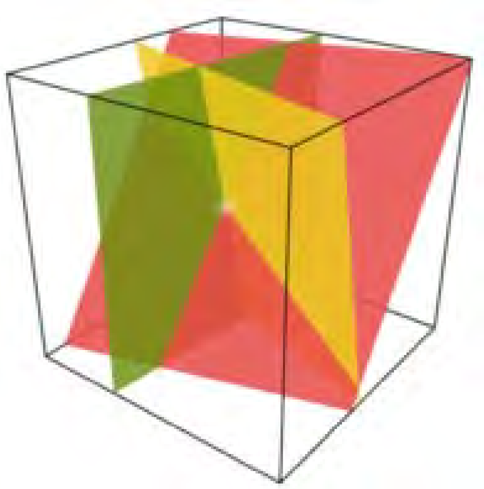


Figure1. line intersection

3.2. Polynomial functions of Secret Sharing Schemes

The essential idea of Adi Shamir is that 2 points are enough to define a line, three points are sufficient to define a parabola, four points to define a cubic curve, etc.. In other words, we need K points to define a polynomial of degree K-1.

Let us suppose that we want to use a threshold scheme (k,n) to share our secret S that we assume, without loss of generality, be an element in a finite body F

Choose randomly (K-1) coefficients $a_1,\ldots,a_{k-1}$ in F and let a_o=S

To construct the polynomial,

$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots + a_{K-1}x^{k-1}$

Be any points calculated from him, for example i=1,.. ,n which gives f(i,f(i)). Each participant is assigned a point (a couple of history and the corresponding image by a polynomial function). Given a subset of these couples, we can find the polynomial coefficients using polynomial interpolation, the secret being the constant term a_0

4. DATA SECURITY BY SHAMIR SECRET SHARING

4.1. Motivation

The use of cloud computing is based on the confidence we can give to the service providers. Such a situation it is very tough to be reinforced with the traditional architecture of cloud that lies on a single supplier. This dependency threats confidentiality of customer data since they are hosted by an only external service provider who may exploit them negatively. Knowing this current dependency, we come up with a new way for hosting data warehousing in order to eliminate the dependency on a single service provider by using multiple service providers. This contribution of solution makes data hosted insignificant and therefore not usable from each and every single one of the service providers where data are lodged

4.2. Suggestion

Our proposal is to share each data stored in the warehouse on several suppliers of the cloud through the secret sharing algorithm (Shamir 1979). In each chapter we presented in details the solution for safe storage and the operation of a data warehouse in the clouds, the latter is inspired by the idea proposed by Danwei Chen Yanjun and He in their article entitled “A Study on Secure Data Storage Strategy in Cloud Computing” [11] This is a solution based on the secret sharing algorithm that shares the data in n-tuple and store those data from one supplier. Our contribution consists of storing n-tuple to a lot of suppliers of cloud computing service.

This way of dividing data allows on one hand to store at each supplier level a part of information, and then those data will not be understandable and not exploitable by a malicious user in case of intrusion. And another hand of not depending on one supplier, which minimizes the risk of unavailability of data. The process steps are:

-Each cloud provider has a copy of the architecture of the customer's warehouse.

-Each data of the company is stored and shared among various suppliers in the order to make it unusable by each supplier as insignificant.

-The number of data fragments depends on the number of suppliers chosen by the customer.

-To restore a data, the client must recover the fragments stored in different suppliers to reconstitute the initial data (Figure2)

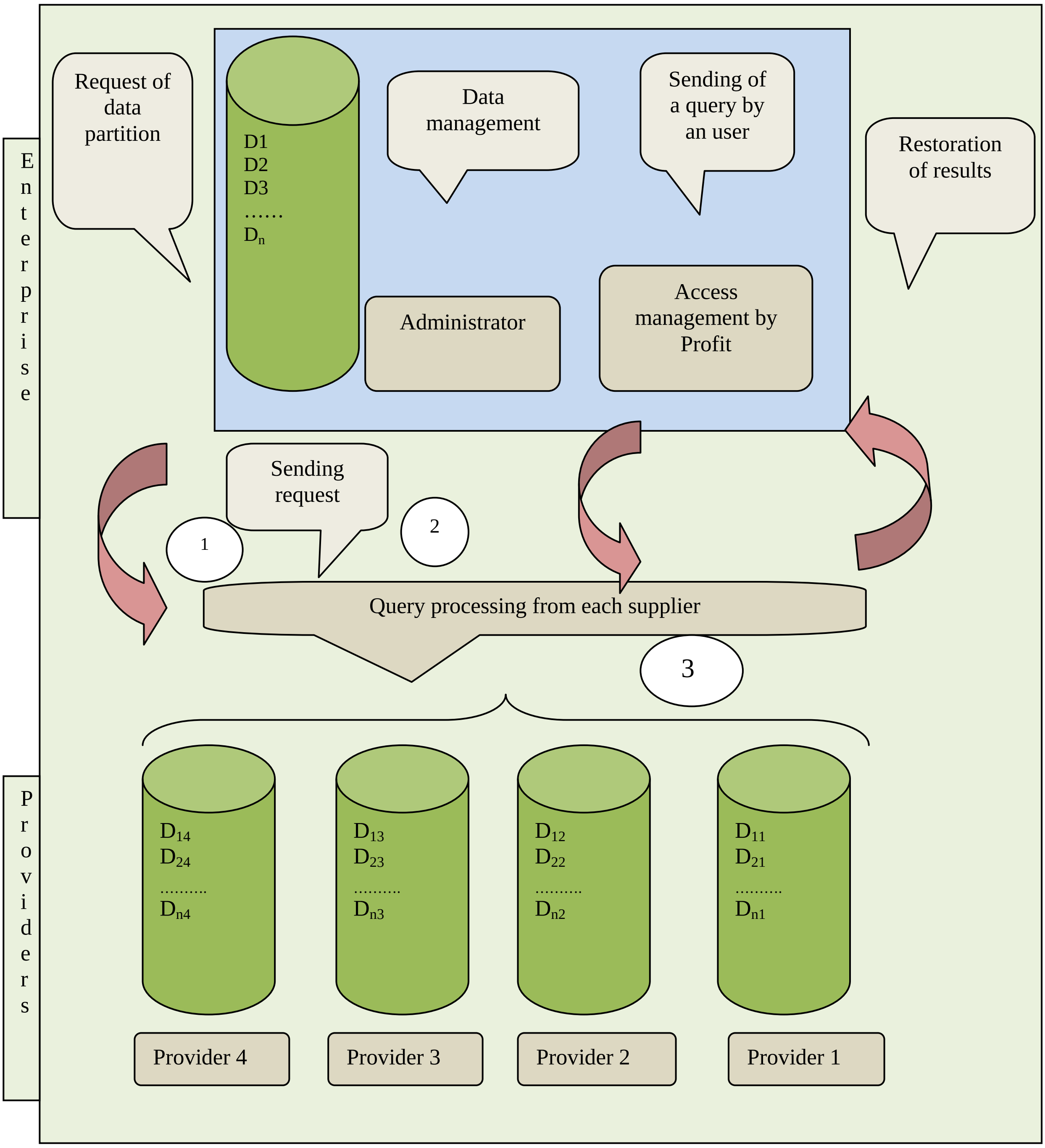

Enterprise
Request of data partition
D1
D2
D3
……
D_n
Data management
Sending of a query by an user
Restoration of results
Administrator
Access management by Profit
Sending request
1
2
Query processing from each supplier
3
D_{14} D_{24} ………. D_{n4}
D_{13} D_{23} ………. D_{n3}
D_{12} D_{22} ………. D_{n2}
D_{11} D_{21} ………. D_{n1}
Providers
Provider 4
Provider 3
Provider 2
Provider 1

Figure.2 Scenario of a shared data warehouse in the clouds

4. SECURITY LEVEL EXPECTED

Our solution provides three security levels:

- Ability to restore data in case of non availability of service or the disappearance of a supplier since the idea is based on secret sharing algorithm that is able to reconstruct the initial data from a predefined number of fragments that can be less than the number of fragments stored at the various suppliers**.**
- Security of transactions between the customer and suppliers since the data that pass through over the network is incomplete and unusable.
- Security of data stored among different suppliers since each every single of them has only a part of data which is non-significant.

To achieve the expected safety, we must meet both rules in the two following subsections

4.1 Coefficient choice

The choice of coefficients has a great influence on data security. Let us assume for example that the a_i coefficients chosen by random are all equal to zero. The polynomial:

$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \ ... + a_{K-1}x^{k-1}$ becomes $f(x) = a_0$ which does not ensure the safe transfer. Thus the algorithm requires that the coefficients should not be null simultaneously.

4.2 Temporal Complexity

Let us assume that an attacker invades a storage node, and steals a data block r_i and wants to restore the original data D with aggressive methods based on r_i and decodes the coefficients.

He needs $[P_{K-1} / (K-1)!]$ Which is the temporal complexity, where p>>K>> 2. Such a report cannot be calculated with the capabilities of current treatments computers. By referring to this theory, we can notice that if we increase K we can increase the time complexity which means lower risk. On the other hand the number K is a factor that depends on how many pieces we hope to get after the partition of the original data D since it cannot be larger than this number.

Thus in order to reduce the risk we must increase the K factor therefore increase the number of partitions N. This means we must increase the number of cloud computing provider in order to reduce the risk of recovery information. The figure 3 shows the increasing of temporal complexity on the basis of the increasing of k factor, which is the number of suppliers needed to recreate the information.

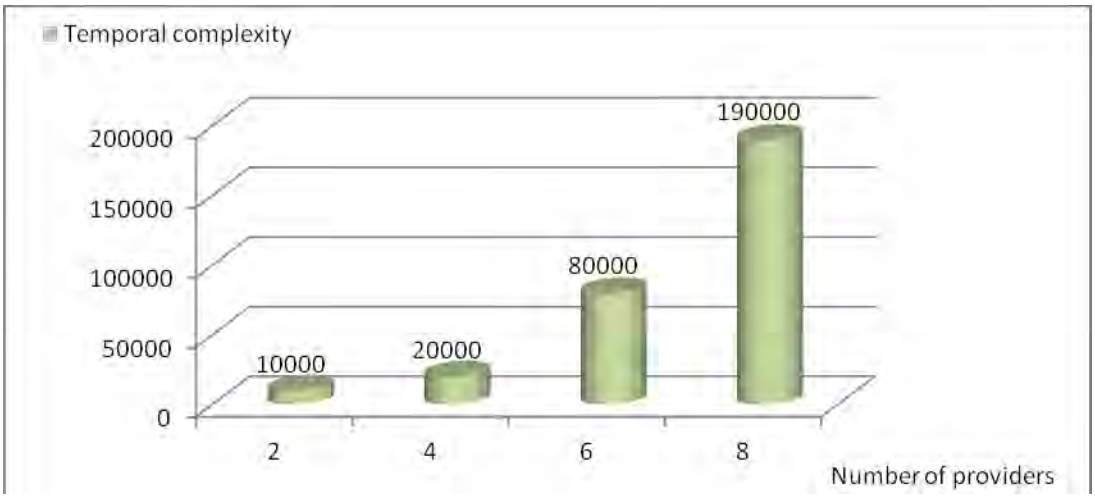


Figure.3 The increase of temporal complexity based on the number of suppliers.

4.3 Theoretical result cost/risk

The most important advantage of cloud computing is that the company pays what it consumes "pay as you go", that means, the cost of consumption is based on the space memory consumed, network access and query processing time. Hence we can summarize cost function by the following equation:

$D * C_S + T * C_T + T_{rq} * C_{Tr} = C_{tot\,;}$ where

D : data stored in the cloud provider system

C_s : Storage cost that depends on each provider

T : Query processing time

C_T : processing request cost that depends on each provider

T_{rq} : size of the query and result

C_{Tr} : Cost of transfer over the network which depends on each provider . Now back on the function cost for our proposal, it changes depending on number of suppliers with whom the company is committed; where function cost becomes

$$\sum_{i=1}^{n} (D_i * C_{Si} + T_i * C_{Ti} + T_{rqi} * C_{Tri}) = C_{tot}$$

where 'n' is the number of providers of cloud service From this formula we can see that the cost to be paid to one vendor by the company with our proposal is 'n' times larger than a traditional solution, but it also provides 'n' times more safety and reduces risk. The cost of risk is the impact that the loss of some or the entire data warehouse will have on the company's business; this one is calculated in working time to recover data. Specifically, our proposal is

based on the secret sharing algorithm that requires a fixed number k of some early data for reconstitution. The number k represents in our suggestion the number of cloud provider that must be available for data reconstitution. However the number of suppliers may exceed the number required K fixed from the beginning to find the safety margin in case of non availability of service providers for instance and achieve a number 'n' which depends on the choice of the company.

To reduce the risk of unavailability, it is then necessary to increase the factor 'n' which is the number of cloud service provider used by the company. This addiction influences the cost of using the cloud computing system 'Ctot', which will increase in proportion to cloud providers. The relationship between cost and risk is then a compromise that depends on the level of security that the company wants to make sure: the more we diminish the risk of coalition by increasing the number of suppliers more you increase the cost of using the cloud computing. On the other hand the risk management has become an integrated part in business activities.

It is to effectively manage the risks to which the computer system of the company is exposed and to prepare against attacks to overcome these risks. Such a policy requires a significant financial budget to ensure reliability and continuity of service of the company since accidents can cause significant financial damage (IBM 2008). In addition, in case of fault from the company's computer system it will lose those customers and its reputation in the market. Reduce such risks has become a strategic priority for companies in a tough competitive environment. In fact the definition risks associated with cloud computing is much broader encompasses the uncertainties, risks and losses (5) which requires a more effective risk management and greater financial budget. This risk management is provided by companies in specific ways as Mehari method and the Marion method. These methods generally lead to efficient outcomes, which are to estimate the cost of risk in case of fault system.

Through these two factors are the total cost of cloud computing C_{tot} and calculated the cost of risk estimated by the methods of risk management, the company can know at what moment a cloud computing solution becomes expensive, taking into account the following constraint: C_{tot} / cost risk < 1. Knowing that the cost is a constant risk, the company can reduce this ratio by decreasing the number of suppliers.

5. PRATICAL IMPLEMENTATION OF OUR PROPOSAL

In this chapter we will present a prototype for the implementation of our proposal. This prototype consists of four cloud providers with whom we host our data warehouse containing data on sales of products in several stores; this is a simulation of a large volume of data. We shared the data by using the secret sharing algorithm. The following chapter contains various features that are provided by our prototype.

5.1 Data Manipulation

5.1.1 Data Partition

Data sharing is done at company level through the first stage of the algorithm of secret sharing. This first step ensures data sharing based on number of suppliers and also includes sending each part to a supplier arbitrarily.

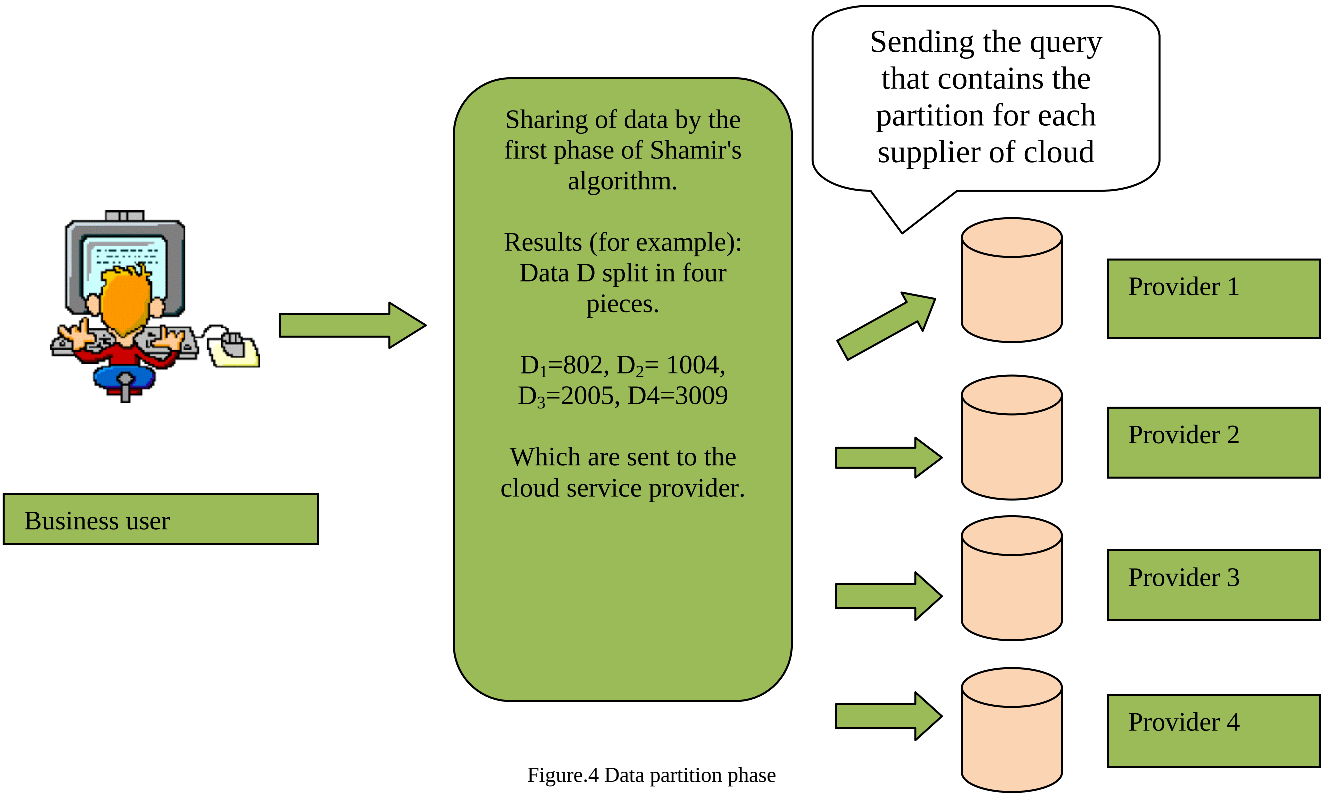

Sending the query that contains the partition for each supplier of cloud
Sharing of data by the first phase of Shamir's algorithm.
Results (for example): Data D split in four pieces.
D_1=802, D_2= 1004, D_3=2005, D4=3009
Which are sent to the cloud service provider.
Business user
Provider 1
Provider 2
Provider 3
Provider 4

Figure.4 Data partition phase

5.1.2 Restoration of Data

Depending on the different parts constituting the information, the second part of the secret sharing algorithm is capable of reproducing the original data. This step is done at company level.

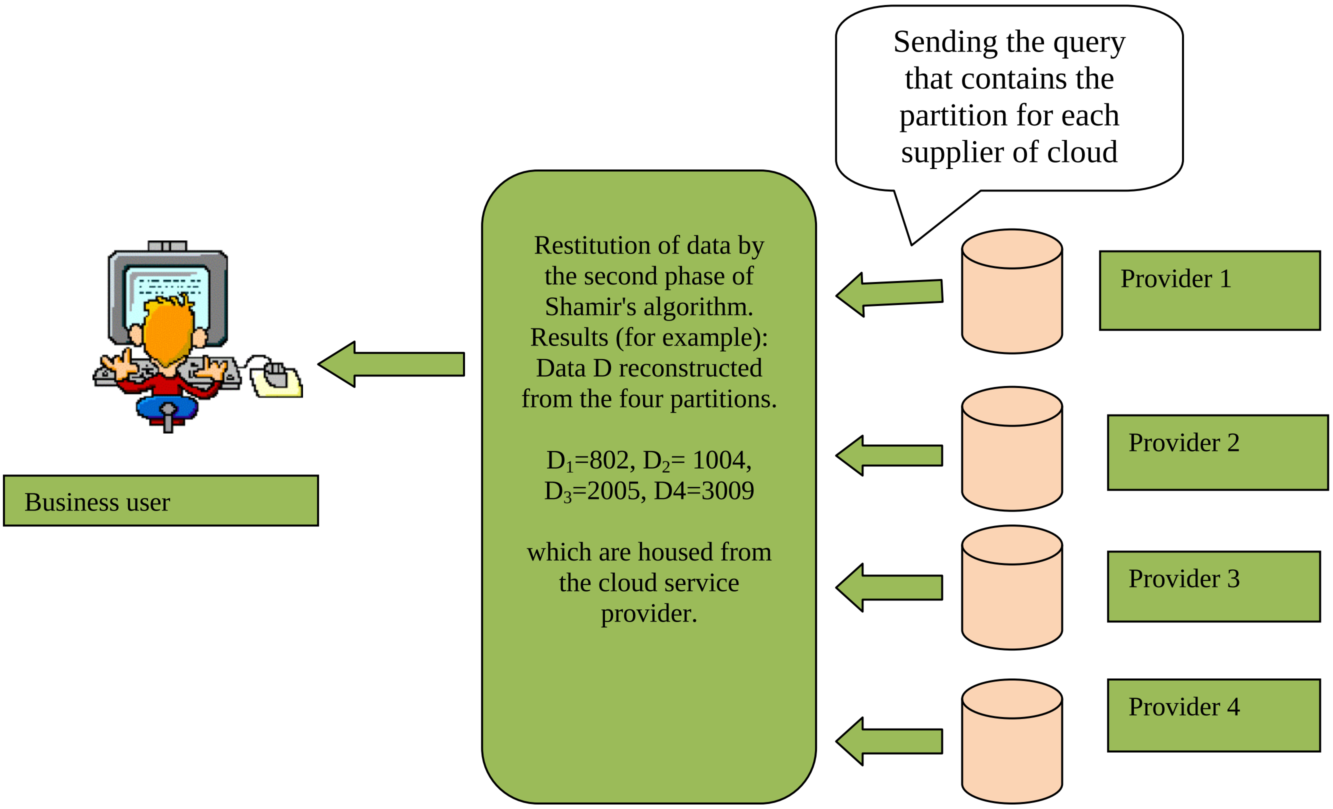

Sending the query that contains the partition for each supplier of cloud
Restitution of data by the second phase of Shamir's algorithm.
Results (for example): Data D reconstructed from the four partitions.
D_1=802, D_2= 1004, D_3=2005, D4=3009
which are housed from the cloud service provider.
Business user
Provider 1
Provider 2
Provider 3
Provider 4

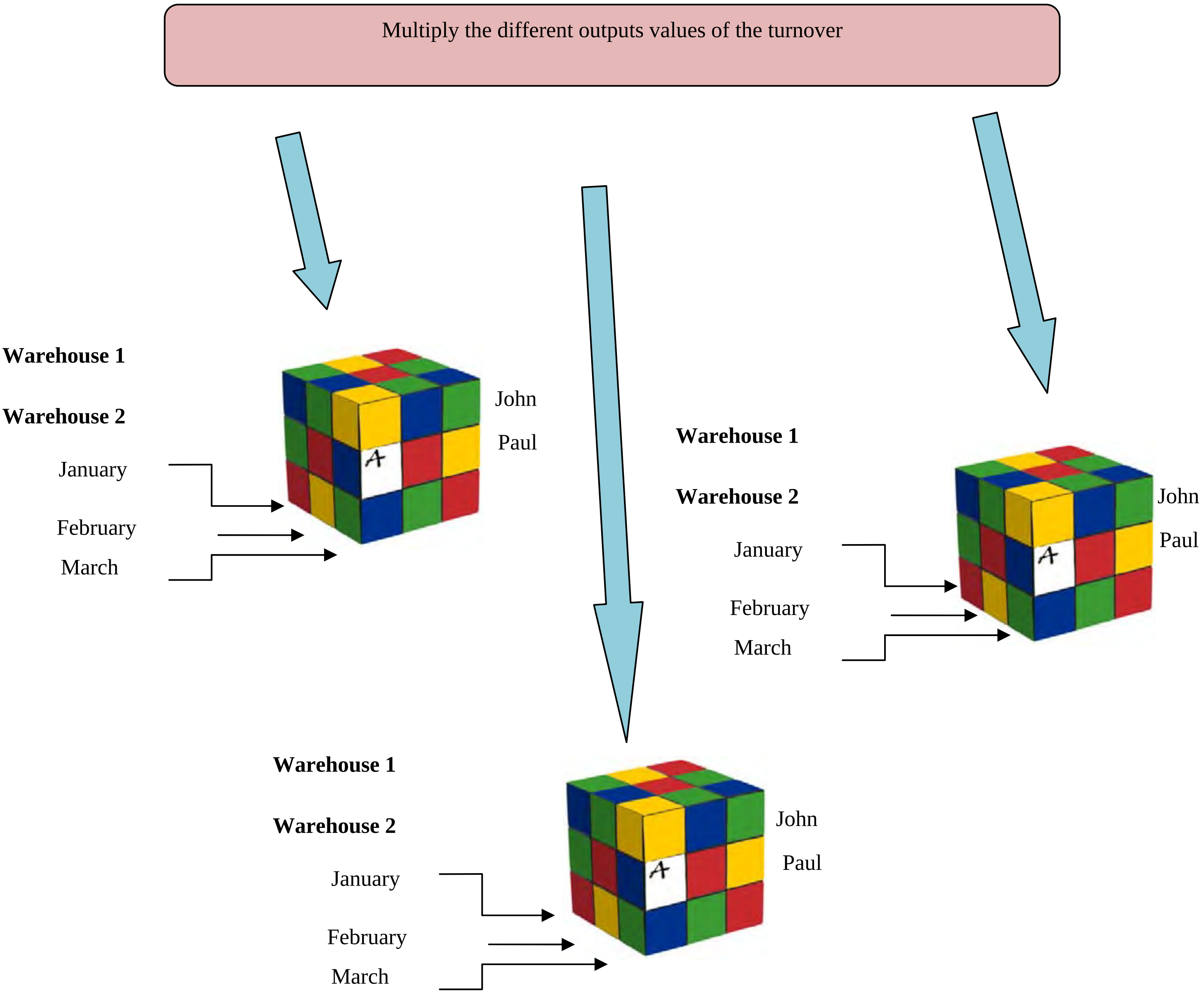
Figure.5 the restitution phase data

5.2 Other operators for OLAP analysis

One can easily make additions to the shared data, but analysis of OLAP requires other types of aggregation: mean, standard deviation, min, max ... which require changes to make them work with our model. In our prototype we have implemented the maximum and the variance. The two subsections present the principle that we have used to implement it.

5.2.1 Variance

To incorporate the variance in our work we have proposed to add a column where the multiplication of each value of xi at the enterprise level and then sharing the column at different suppliers so that we can use according to the restoration data. This obliges us to add another column to store the square xi. Figure 5 explains the principle adopted to ensure the calculation of the variance of the data. The present example is to calculate the variance on turnover per store, per department, per month.

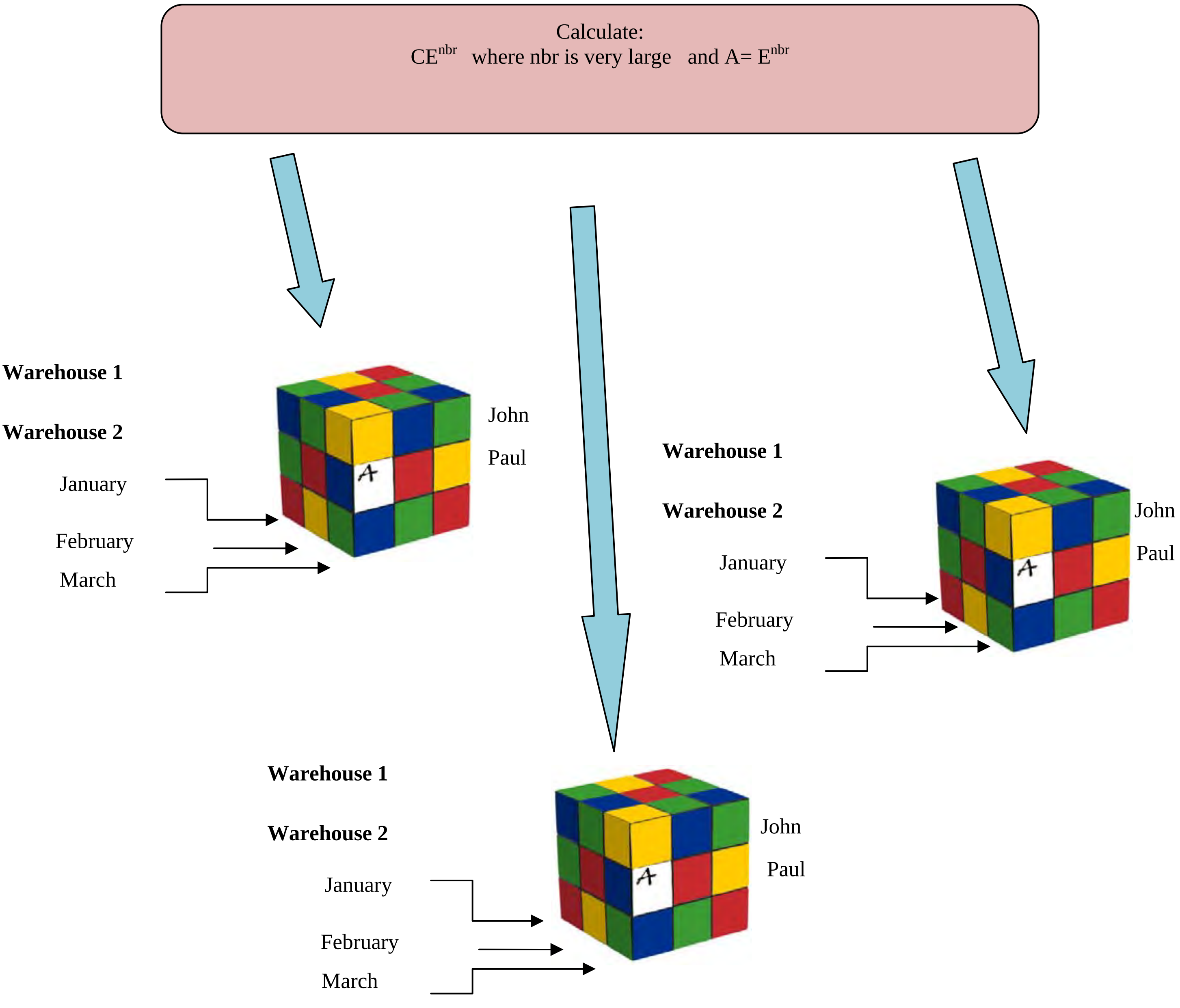


Restitution: $\sum CA^2$; $\sum CA$; Number of record. => Calculation of variance.

Figure 6. The process of variance calculation

5.2.2 Maximum

The idea we propose is to compute the values of columns that we want to estimate its maximum by the following formula: each record containing a very large number in a first step. Then we share the new values with the various suppliers. Once we know the maximum price per column values we apply the aggregate function max on each warehouse. Indeed the results obtained have no meaning, for that you need to apply the method to restore data proposed by Shamir and then multiply the result by (the result from dividing one over the number we used in the first stage). For more explanation Fig 6 shows the process of calculating the maximum of the turnover**.**



Restore: $\sum CA^{nbr}$ =>Implementation of the second phase of the secret sharing algorithm of the sum.=>Calculation of R $^{1/nbr}$

Figure. 7 The process for calculating maximum

5.2.3 Discussion

As already announced the implementation of the variance and maximum requires the adding column to each of them. While the addition of columns causes a further increase in the volume of data, and the cost of solution. But in beforehand, it applies only to measures of the warehouse, then to a limited set of attributes. The aggregation operators Max, variance, count, and average necessary for OLAP analysis are implemented in this prototype.

6. CONCLUSION AND OUTLOOK

This study found that major problems of housing a data warehouse in the clouds is the lack of confidence from the provider of cloud computing, service availability and security of data transfer to the cloud. To minimize risk, we proposed a solution based on the secret sharing algorithm, which is to share information from several suppliers instead of one. The objective of this solution is to reduce the vulnerability of data transmitted. Each part is stored to one of the suppliers, making them an insignificant proportion for each supplier and on another hand solves the problem of unavailability service in case of failure from one of the suppliers. We created a prototype that not only provides the necessary treatments for sharing and retrieval of data but also incorporates some aggregation operators (Max, Count, variance, average) for the OLAP analysis. The results obtained through this work show that the prospects for research on the subject are numerous.

- Study the possibility of integrating traditional cryptography for secure transfer over networks.
- Apply a risk management method to determine the cost of real risk.
- Strengthen access security by integrating the solution in the management of access to the warehouse in the cloud based on the profits of the enterprise users.

7. REFERENCES

[1] Amazon.com"Amazon Web Service (AWS)" online at http://AWS.amazon.com,2008

[2] N.Ghoring "Amazon's S3 down for several hours"2008 http://www.pcworld.com/businesscenter/article/142549/amazon_S3_down_for_several_hours,html

[3] Mladen A.Vouk.Cloud Computing -issues, paper implementation journal of computing and information technology,september 2008.june 2008 accepted.p237

[4] Fangfei Zhou.Manish Goel.Peter Desnoyers,Ravi sundaram.Scheduler Vulnerabilities and attack in the cloud computing.College of computer Science, Northearn University of Boston USA

[5] www.presence-pc.com/actualite/Amazon-EC2-37511

[6] Sean Carlin and Kevin.Cloud Computing.International Journal of Ambiant and Intelligence.junuary-march 2011,volume;3issu;9.p.14

[7] Swamp computing a.k.a.cloud computing.Web security.2009-12-28.Retried 2010-01-25

[8] Jingpeng wei.Managing of a virtual machine images in a cloud environmental Proceeding the 2009 ACM workshop on the Cloud Computing Security.New York

[9] Jorg schwenk,Luigui Lo Locano,Meiko Jensen.On Technical security issues in the Cloud Computing.IEEE International conference on the Cloud Computing

[10] Cong Wang,Kui ren,Qian Wang,Wenjing Lou.Ensuring Data Storage Security in the Cloud Computing, the 17th IEEE international workshop on quality of service,Chalerstone july 13-15,South Corolina

[11] Dan Wei Chen,Yanjun He.A study on secure Data Storage Strategy in the Cloud Computing, journal of Convergence Information Technology,Volume 5,Number 7

[12] A.Parack.S.Kak.online Data using implicit security, Information sciences,vol.179,no 3323-3331

[13] Virtual Computing Lab.http://Vcl.ncsu,edu/