ISSN No. 0976-5697

Volume 2, No. 5, Sept-Oct 2011

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

Data Locating In Real Time Cloud Based Service Oriented Architectures

Boopathy.D* and R.Srividhya
*Research Scholar, Assistant Professor
Dr.G.R.D College of Science
Coimbatore, Tamilnadu
*ndboopathy@gmail.com
srividhya.r@grd.edu.in

Abstract: Cloud computing technology is a new concept of providing dramatically scalable and virtualized resources. It implies a service oriented architecture, reduced information technology overhead for the end-user, great flexibility, reduced total cost of ownership, on-demand services and many other things. When the Software as a Service Provider or Infrastructure as a Service provider may bankrupt or suddenly disappeared from the competitive market means, who will take responsible for customer data? One of the main concerns of customers is Cloud security and the threat of the unknown. The lack of physical access to servers constitutes a completely new and disruptive challenge for investigators. This paper represents the physical storage of data in end-user point of view with possible way and satisfies the standards and legal issues.

Keywords: Cloud computing, Security, Legal Aspects, Privacy access, Data Location.

I. INTRODUCTION

Cloud computing is a natural evolution of the widespread adoption of virtualization, service-oriented architecture, autonomic, and utility computing. Details are abstracted from end-users, who no longer have need for expertise in, or control over, the technology infrastructure "in the cloud" that supports them. The NIST defines,"Cloud Computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".

Cloud computing models are of two types: Deployment model and Service model.

Deployment model is further classified into 4 type's namely private cloud, community cloud, public cloud and hybrid cloud.

In Private cloud, the cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise. In Community cloud, the cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on premise or off premise. In public cloud, the cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services. In hybrid cloud, the cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

Service model is also classified into three namely SaaS, PaaS, and IaaS.

A. Software as a Service (SaaS):

The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based email). The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

B. Platform as a Service (PaaS):

The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations

C. Infrastructure as a Service (IaaS):

The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems; storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls).

II. LITERATURE REVIEW

One of the most common compliance issues facing an organization is data location [1]. Use of an in-house computing center allows an organization to structure its computing environment and know in detail where data is stored and the safeguards used to protect the data. In contrast, a characteristic of many cloud computing services is that the detailed information of the location of an organization's data is unavailable or not disclosed to the service subscriber. This situation makes it difficult to ascertain whether sufficient safeguards are in place and whether legal and regulatory compliance requirements are being met. External audits and security certifications can, to some extent, alleviate this issue, but they are not a panacea.

Once information crosses a national border, it is extremely difficult to guarantee protection under foreign laws and regulations [6].From the technical point of view, this evidence data can be available in three different states: at rest, in motion or in execution. Data at rest is represented by allocated disk space. Whether the data is stored in a database or in a specific file format, it allocates disk space. Furthermore, if a _le is deleted, the disk space is de-allocated for the operating system but the data is still accessible since the disk space has not been re-allocated and overwritten. This fact is often exploited by investigators which explore these de-allocated disk space on hard-disks.

In case the data is in motion, data is transferred from one entity to another e.g. a typical file transfer over a network can be seen as a data in motion scenario. Several encapsulated protocols contain the data each leaving specific traces on systems and network devices which can in return be used by investigators. Data can be loaded into memory and executed as a process. In this case, the data is neither at rest nor in motion but in execution [10].Depending on the Cloud offer used, virtual IaaS instances do not have any persistent storage. In current Cloud environments CSP do not offer any verification process providing the ability for the customer to verify that the sensitive data stored on a virtual machine has been deleted exhaustively [10]. In the SaaS model, the enterprise data is stored outside the enterprise boundary, at the SaaS vendor end. Consequently, the SaaS vendor must adopt additional security checks to ensure data security and prevent breaches due to security vulnerabilities in the application or through malicious employees. This involves the use of strong encryption techniques for data security and fine-grained author- ization to control access to data [11]. Cloud computing moves the application software and databases to the large datacenters, where the management of the data and services are not trustworthy. This unique attribute, however, poses many new security challenges (Cong Wang et al., 2009) [12].

Analysts' estimate that within the next five years, the global market for cloud computing will grow to $95 billion and that 12% of the worldwide software market will move to the cloud in that period. To realize this tremendous potential, business must address the privacy questions raised by this new computing model (BNA, 2009) [13].

Yet, guaranteeing the security of corporate data in the ''cloud'' is difficult, if not impossible, as they provide different services like SaaS, PaaS, and IaaS. Each service has its own security issues (Kandukuri et al., 2009) [14]. Due to compliance and data privacy laws in various countries, locality of data is of utmost importance in much enterprise architecture (Softlayer, 2009) [15]. The security policies may entitle some considerations wherein some of the employees are not given access to certain amount of data. These security policies must be adhered by the cloud to avoid intrusion of data by unauthorized users (Blaze et al., 1999; Kormann and Rubin, 2000; Bowers et al., 2008) [16].

Data security is a significant task, with a lot of complexity. Methods of data protection, such as redaction, truncations, obfuscation, and others, should be viewed with great concern [17]. Data Loss/Leakage. Be it by deletion without a backup, by loss of the encoding key or by unauthorized access, data is always in danger of being lost or stolen [18]. In general, cloud users are not aware of the exact location of the datacenter and also they do not have any control over the physical access mechanisms to that data. Most well-known cloud service providers have datacenters around the globe. Some service providers also take advantage of their global datacenters. In cloud environment, data can be assigned a cost by the users based on the criticality of the data [19]. since a customer will not know where her data will be stored, it is important that the Cloud provider commit to storing and processing data in specific jurisdictions and to obey local privacy requirements on behalf of the customer; one needs to ensure that one customer's data is fully segregated from another customer's data; it is important that the Cloud provider has an efficient replication and recovery mechanism to restore data if a disaster occurs [20].

III. PROBLEM DESCRIPTION

Software as a service is mingled with Platform as a Service and Infrastructure as a Service.

The above diagram shows the Software as a Service providing process in simple style. The software as a service users got their account by free or for trial period or by payment mode. The mode of acquiring service may be different but the users can access their account and saved their confidential data into their account. In the above figure, two concerns provide services and store their data in third parties. Who provides the IaaS?

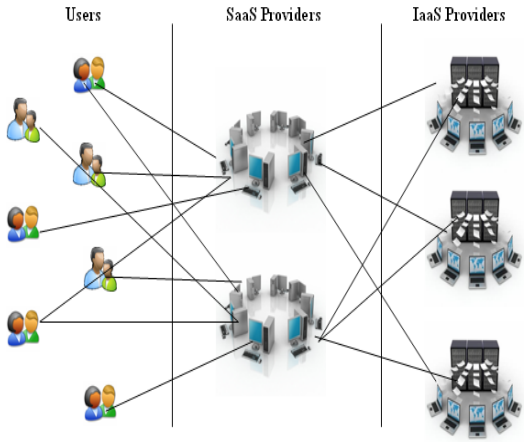

Figure: 1 SaaS providers Current Model

Here that IaaS person is unknown to the user. What is third party? Third party means, the users doesn't know where their data are stored and from where they access their stored data. The SaaS providers use many third parties to

store their data because in online storage and accessing, the performance and security is important. For that performance issues they used more IaaS providers. So if more than one person into the service means, that company will name them as third party or fourth party? The users think their data are safe in online storage but the users don't know the status of their data. Because the data storage destination was not shown by the SaaS providers to the users. And the SLA is not satisfied the rules and regulations. Everything is fine now but in future when something happens (i.e.) when SaaS or IaaS may bankrupt or vanish from the service providing market, which time users may need their data back but they lose their data in reality. This paper presents some issues in Service Oriented Architecture with some solution model.

The data may contain anything but it is important to the users and its holders. The cloud computing handle the data in safe method but the assurance of the data quality and availability is very big question mark. The users must have to know the place of the data where it is saved and its details. Because at the time of downtime the SAAS provider said the problem is from hardware vendor part. But the actual problem may be in SAAS provider part. For the data availability and data assurance the "CLOUD DATA TAGG (CDT)" tool will help.

IV. CLOUD DATA TAGG

A. What is Cloud Data Tagg?

Cloud Data Tagg is process of allocating the tagg address to the data's which is handled in Cloud Computing areas. In reality the data's stored in cloud computing is strange to understand. Due to maintain the consistency and avoid the non-availability of data, the service providers store the data into multiple servers. Each and every server is interconnected and the data's were synchronized for future assistance, according to the country the provisional acts are differed from one to another. The reason for the difference, there is no common act which covers the whole word "Cloud Computing". The servers are installed in different places or the service provider may give contract to the IAAS service vendor that he has not servers in all over the world.

So many IAAS service vendors were placed around the world to provide the full service. In this time data's were placed in many servers. The problem of non-availability is considered and raised only at the time issue. But at that time every thing will be exceed the limit of recovery or it result in data loss. Just think if the SAAS service provider or the IAAS provider became bankrupt means what happens to the data's stored in the server? If the clients know the places of data's at the service time itself means some problems will be solved from beginning itself.

a. ***Cloud Data Tagg (CDT)*** concept will act as a bridge between the user and the service both providers (both IAAS & SAAS). This is like a rendering service for the person who needs it. To basic service providers are two types

i. IAAS – Clients – Users

ii. IAAS – SAAS – Clients – Users

The above types are followed and being followed in service providing part. The users will be the end users, because he/she is person who is going to utilize it and enjoy the full benefit of service. The CDT concept is followed means non-transparency is eradicated in data handling side.

The IAAS and SAAS provider may accept mean it is possible to reach the prescribed goal. Why it need IAAS and SAAS support? The SAAS provider provides the software level service and the IAAS provider provides the hardware level service and he is person who stored the data's. If the both person are accept the concept means it will merged in their areas and start work.

So this tool will help the clients to track the data and knows the status of it.

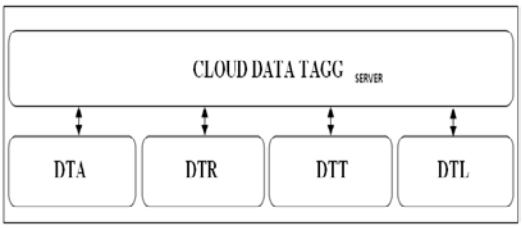

Figure: 2 Cloud Data Tagg Module

B. CLOUD DATA TAGG (CDT) Modules:

- Ξ Data Tagg Allocator (DTA)
- Ξ Data Tagg LOCATOR (DTL)
- Ξ Data Tagg Tracker (DTT)
- Ξ Data Tagg Responser (DTR)

By using the above modules the clients can handle their data's easily.

a. Data Tagg Allocator (DTA):

The DTA is one time process of allocating the serial code to the main data. It means the data is used in some specified software only so it sequential serial key is prepared one time and allocated it to the specified software (that is data accessing key). The user can select the auto saving method by their choices, the auto changes may done in following categories.

i. Some specified data size reached means (Ex: Every 10 MB data will stored)

ii. Some specified time limit reached means (Ex: Every five mints data will stored)

iii. At the time of sign out or log out from the software the data stored.

While in above conditions auto save will done and each and every time of data saving it generate sequential numbers and tagged it with the saved data's and the tagg number will be updated to the DTT. The DTA work is allocation of sequential data to saved data's and the sequential serial code will upload to the general data tagg server for future assistance. The each and every module access the sequential serial code from the server and it upload the details to the server. The details are saved and provide to the clients at their demand. The DTA process is accessed by SAAS provider only. The person who rendering service may able to make everything into transparency and also he can turn something into non-transparency, in such case the SAAS provider's SLA's are in non-transparency method, due to some issues with IAAS provider. After they enter into the path of Cloud Data Tagg means the storage data's issues will came into transparency side.

At first time of allocating the space itself the software's also located in server. So that time itself the database became ready to store the data. So the software is already merged with the database. So for that database we provide the first sequential special code it used to find that database

easily (code may contain 10 to 16 digits). After that the sub sequential codes will be randomly generated at the time of data saving and automatic data saving time. Time of auto saving will be recommended by the service provider and allocated by the user. (It may change in future by user only). In sub sequential code it tagg the data with date, day and time of saving will be noted. The DTA is only allocated the Data Tagg code, after that it will forward to the Data Tagg Database.

b. Data Tagg Locator (DTL):

The DTL is used to find the location of current saving data details by using its sequential serial code. So the user can know the location of data and also know the back up places of the data easily in sitting front of the computer. The DTL is connected with the Data Tagg Database for acquire the code details, the connection and method of accessing the code may very secure with the help of sequential code. The DTL is the current data locating module it is working in the CDT server by using the distributed computing method.

c. Data Tagg Tracker (DTT):

The DTT is another one important module. The DTT is simply known as history place of the specified data. It means from the first time a data is saved means the sequential serial number is added to this DTT module. From that time onwards that saved sequential data comes under tracking part. It simply surveillances the data and it recorded the information into the sequential file. It includes records of

i. First time saved details
ii. The sub-data which are related to this data (includes the last saved data's)
iii. The accessing history of the data
iv. The data back up details
v. The last retrieval details of the data (it shows the location of retrieval)

The data retrieval averages, counts and maximum hit of specified data's all are shown in this module. The DTT is connected with the Data Tagg Database for acquire the code details, the connection and method of accessing the code may very secure.

d. Data Tagg Responder (DTR):

DTR is created for the purpose of alert the clients. The clients are not able to find out the downtime and also the data unavailability. So the DTR sends some alerts to the clients regarding the data server. The DTR is the module which responds the user's requirements. At the time of D&SLA signing it ask the users to select the set of services for the purpose of keep on touch with the data. That user's selected services will follow by the DTR and sends to users automatically. The main concept of DTR is response of the data while it is covers the limits or not.

i. Data Tagg Ranking
ii. Data Tagg High-availability
iii. Data Tagg Non-availability
iv. Data Tagg Alerts and so on.

The DTR is connected with the Data Tagg Database for acquire the code details, the connection and method of accessing the code may very secure.

D. Data Tagg Database (DTD):

The Data Tagg Database (DTD) is the place where the sequential special code and sequential sub-code details may store in this database. The DTD accessed by the DTA for store the purpose of store sequential special code and sequential sub-code. The DTD accessed by the DTL for find out the location of the all required data's. The DTD accessed by the DTT for tracing the data's activities and this is the only module which keeps on connection with the DTD for future assistance. The DTD accessed by the DTR to respond the users selected requests on data's based service.

The DTD will be the centralized database which contains the full details of CDT server based service with distributed computing system.

E. Deployment & Service Level Agreement (D&SLA):

The Cloud Data Tagg may provide D&SLA to SAAS provider, why CDT need D&SLA means the module is going to provide the information's of stored data's as a service. The data's may contains anything which is related personnel, working area, or any other important thing. The clients are liable for the data what they stored. The CDT service is tracking the data from the beginning to ending, so the permission is needed from clients/users for track their data's and SAAS provider must give the permission to deploy this module into their software. And also SAAS provider must get the permission form IAAS provider to surveillances their storage data's which related to the SAAS provider privileges.

The Deployment Level covers and related to the SAAS and IAAS provider. Without their knowledge and support the module can't work successfully. The servers must show in the providing list with its transparency (otherwise it will make as transparency). The module deployment is made in SAAS and IAAS area. So, it related to the service vendors or service providers. The SLA is getting signed from the end-users of this service. The end-users are the person who utilizes the service from SAAS in the mode of software and the in-direct service from IAAS in the mode of storage. But he must comes under the SLA due to cross some issues for permit the service provider to track the data's and some other data's protected issues. The user's must clear in one thing, which is the service providers may use their data for their use it may include any reason. But in this place of CDT service, data is a data it contains anything which is related to the users or not. That data will be tagged, located, tracked, responded and the data stored details will be stored in database and give that tagg details to users without any fail and in right time.

F. CDT Service Model:

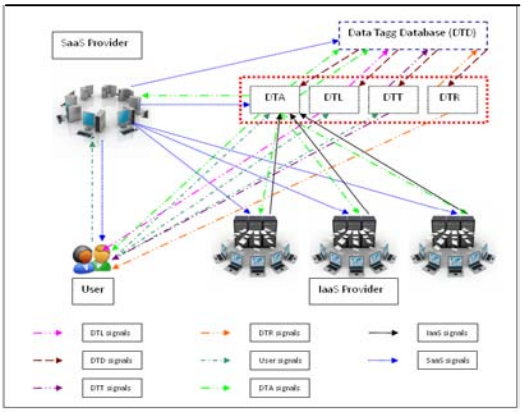

Figure: 3 Cloud Data Tagg Service Model

The above diagram explains CDT service. When the user request to create an account in SaaS provider then SaaS provider send that request to DTA for verification purpose. After that it sends the request to allocate the space for that user and that space details send to the DTA. Again DTA verifies the location of data storage and security categories. Later it enters those details to the DTD. Now it sends the final statement to the SaaS provider. After that only the user able to get the account allocation information with the sequential code. When the user requests to locate the data he/she sends the request to the DTL, it provide that details to him. In the same way it is followed in the DTT system. And DTR send the automatic generated alerts regarding his/her accounts.

G. Sequential Diagram of CDT Service Model:

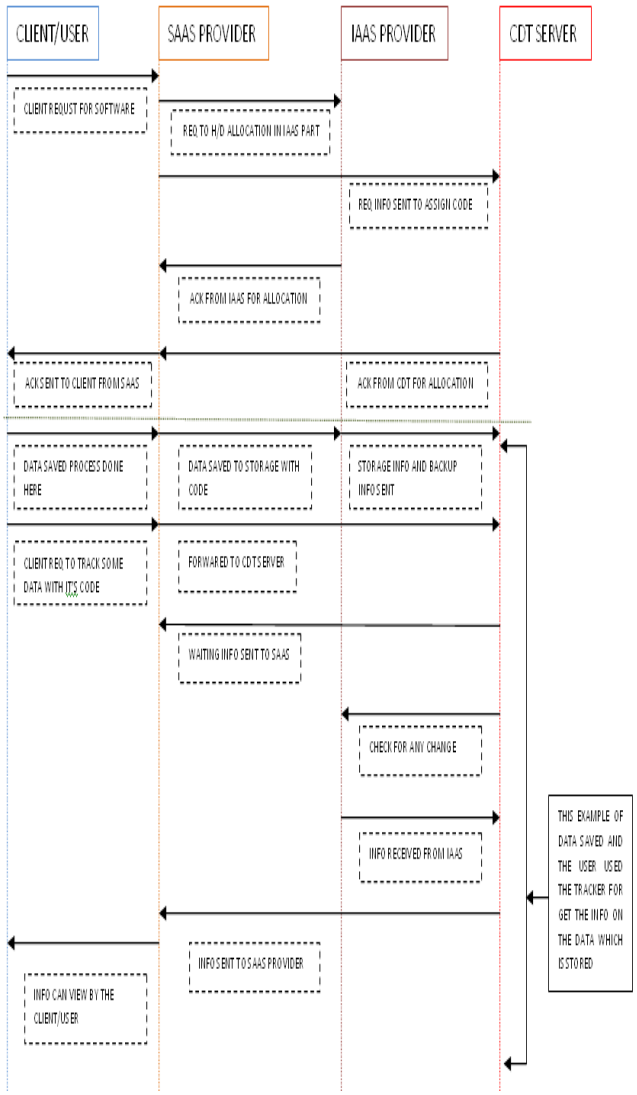

Figure: 4 CDT service model in sequential diagram

V. FUTURE WORK

We are going to simulate the paper and we are attaching this model in all field based scenarios to verify the performance and security level. After that it will implemented in real world with high performance level and along with high level security. Still some modules are under updating. Soon it will get updated and placed to check the working performance.

VI. CONCLUSION

This paper provide a Service Oriented Architecture based model, it will give importance to security and performance, which is required by users of the cloud computing. This model is basic one for the other upcoming models of SOA, not only this model going too verified in Market Oriented Architecture (MOA). Soon this model will attach in SOA and MOA to verify it level of service.

VII. REFERENCES

[1]. DANISH JAMIL and HASSAN ZAKI, "SECURITY ISSUES IN CLOUD COMPUTING AND COUNTERMEASURES", IJEST, Vol. 3 No. 4 April 2011, and pg no: 2672- 2676.

[2]. Hyukho Kim, Hana Lee, Woongsup Kim, Yangwoo Kim, "A Trust Evaluation Model for QoS uarantee in Cloud Systems", International Journal f Grid and Distributed Computing, Vol.3, No.1, March, 2010.

[3]. Ghalem Belalem, Samah Bouamama, Larbi Sekhri, "An Effective Economic Management of Resources in Cloud Computing", JOURNAL OF COMPUTERS, VOL. 6, NO. 3, MARCH 2011, pg no: 404-411.

[4]. Aishwarya C.S. and Revathy.S, "Insight into Cloud Security issues", UACEE International journal of Computer Science and its Applications, pg no: 30-33.

[5]. Amreen Khan and KamalKant Ahirwar, "MOBILE CLOUD COMPUTING AS A FUTURE OF MOBILE MULTIMEDIA DATABASE", International Journal of ComputerScience and Communication, Vol. 2, No. 1, January-June 2011, pp. 219-221.

[6]. Richard Chow, Philippe Golle, Markus Jakobsson, Ryusuke Masuoka, Jesus Molina Elaine Shi, Jessica Staddon, "Controlling Data in the Cloud: Outsourcing Computation without Outsourcing Control", CCSW'09, November 13, 2009.

[7]. Muzafar Ahmad Bhat, Razeef Mohd Shah, Bashir Ahmad, "Cloud Computing: A solution to Geographical Information Systems (GIS)", International Journal on Computer Science and Engineering (IJCSE), Vol. 3 No. 2 Feb 2011, pg no: 594-600.

[8]. Pardeep Kumar, Vivek Kumar Sehgal , Durg Singh Chauhan, P. K. Gupta and Manoj Diwakar, "Effective Ways of Secure, Private and Trusted Cloud Computing", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, May 2011, pg no: .412-421.

[9]. Mladen A. Vouk, "Cloud Computing – Issues, Research and Implementations", Journal of Computing and Information Technology - CIT 16, 2008, 4, 235–246.

[10]. Dominik Birk, "Technical Challenges of Forensic Investigations in Cloud Computing Environments" , January 12, 2011.

[11]. A survey on security issues in service delivery models of cloud computing S. Subashini, V.Kavitha, Journal of Network and Computer Applications, 11 July 2010

[12]. Toward publicly auditable secure cloud data storage services, Cong Wang Kui Ren Wenjing Lou Jin Li, Journal IEEE Network: The Magazine of Global Internetworking archive Volume 24 Issue 4, July-August 2010

[13]. BNA. Privacy & security law report, 8 PVLR 10, 03/09/2009.Copyrigh 2009 by The Bureau of National Affairs, Inc.(800-372-1033), 2009 http://www.bna.comS [accessed on:2November2009].

[14]. Kandukuri BR, Paturi VR, Rakshit A. Cloud security issues. In: IEEE international conference on services computing, 2009, p. 517–20.

[15]. Softlayer.Service Level Agreement and Master Service Agreement, 2009 /http:// www.softlayer.com/sla.html [accessed on:11December2009].

[16]. Bowers KD, Juels A, Oprea A. HAIL: a high-availability and integrity layer for cloud storage, Cryptology ePrint Archive, Report 2008/489, 2008 /http://eprint.iacr. org [accessed on:18October2009].

[17]. Survey on Cloud Computing Security, Shilpashree Srinivasamurthy, David Q. Liu, 2nd IEEE International Conference on *Cloud Computing*, 2010

[18]. Security Guidance for Critical Areas of Focus in Cloud Computing, April 2009. DOI = http://www.cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf

[19]. Towards Analyzing Data Security Risks in Cloud Computing Environments Amit Sangroya, Saurabh Kumar, Jaideep Dhok, and Vasudeva Varma, Springer-Verlag Berlin Heidelberg 2010

[20]. Cloud Computing and Grid Computing 360-Degree Compared, Ian Foster, Yong Zhao, Ioan Raicu, Shiyong Lu, Grid Computing Environments Workshop, 2008. GCE '08 2008

Short Biodata of the Author

Boopathy.D is Research Scholar of Computer Science Department, School of IT & Science, Dr. G R Damodaran College of Science, Coimbatore. He received his Master of Science degree in Information Technology in the year of 2010 from Bharathiar University, Coimbatore. His research area is Data warehousing and mining.

R.Srividhya is Assistant Professor of School of IT & Science, Dr.G R Damodaran College of Science, Coimbatore. She received her Master of Computer Application degree in the year of 2002 from Bharathiar University and she received her Master of Philosophy in the year of 2004 from Bharathiar University. She has worked in RVS College of Arts & Science for 3 years as Lecturer and currently working as Assistant Professor in School of IT & Science, Dr. G R Damodaran College of Science, Coimbatore. The research areas are Data warehousing and Mining, Wireless Networks and Digital Image Processing.