Volume 2, Issue 11, November 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Data Confidentiality Scalability and Accountability (DCSA) in Cloud Computing

Miss. Rehana Begum*
STUDENT M. Tech (CSE)
VCE[1] JNTU University

Mr. R.Naveen Kumar
ASSOCIATE PROFESSOR
VCE[1] JNTU University

Mr. Vorem Kishore
ASSOCIATE PROFESSOR
VCE[1] JNTU University

1. Vaageswari College of Engineering

***Abstract*—This paper aims to achieve data confidentiality scalability and accountability in cloud computing by determining first the security mechanisms required for each data sensitivity level, and which of these security controls may not be supported in certain computing environments, then which solutions can be used to cope with the identified security limitations of cloud computing. Secondly issues such as risks of privacy exposure, scalability in key management, flexible access and efficient user revocation, have remained the most important challenges toward achieving fine-grained, cryptographically enforced data access control. In this paper, we propose a novel patient-centric framework and a suite of mechanisms for data access control to Personal health record (PHR)s stored in semi-trusted servers. To achieve fine-grained and scalable data access control for PHRs, we leverage attribute based encryption (ABE) techniques to encrypt each patient's PHR file. Finally we propose an approach in which procedural and technical solutions are co designed to demonstrate accountability as a path forward to resolving authority privacy and security risks within the cloud.**

***Keywords*— Cloud computing, Confidentiality, Scalability, Accountability.**

I. INTRODUCTION

Cloud computing is a computing paradigm, where a large pool of systems are connected in private or public networks, to provide dynamically scalable, confidential and accountable infrastructure for application, data and file storage. With the advent of this technology, the cost of computation, application hosting, content storage and delivery is reduced significantly. The underlying details of how it is achieved are hidden from the user. The data and the services provided reside in massively scalable data centers and can be ubiquitously accessed from any connected device all over the world.

Cloud computing is the style of computing where massively scaled IT related capabilities are provided as a service across the internet to multiple external customers and are billed by consumption. Many cloud computing providers have popped up and there is a considerable growth in the usage of this service. Google, Microsoft, Yahoo, IBM and Amazon have started providing cloud computing services. Amazon is the pioneer in this field. Smaller companies like SmugMug, which is an online photo hosting site, has used cloud services for the storing all the data and doing some of its services.

Cloud computing provides the facility to access shared resources and common infrastructure, offering services on demand over the network to perform operations that meet changing business needs. The location of physical resources and devices being accessed are typically not known to the end user. It also provides facilities for users to develop, deploy and manage their applications 'on the cloud', which entails virtualization of resources that maintains and manages itself.

A. Cloud Computing Benefits

Enterprises would need to align their applications, so as to exploit the architecture models that Cloud Computing offers. Some of the typical benefits are listed below:

1) Self Healing: Any application or any service running in a cloud computing environment has the property of self healing. In case of failure of the application, there is always a hot backup of the application ready to take over without disruption. There are multiple copies of the same application - each copy updating itself regularly so that at times of failure there is at least one copy of the application which can take over without even the slightest change in its running state.

2) Multi-Tenancy: With cloud computing, any application supports multi-tenancy - that is multiple tenants at the same instant of time. The system allows several customers to share the infrastructure allotted to them without any of them being aware of the sharing. This is done by virtualizing the servers on the available machine pool and then allotting the servers to multiple users. This is done in such a way that the privacy of the users or the security of their data is not compromised.

3) Linearly Scalable: Cloud computing services are linearly scalable. The system is able to break down the workloads into pieces and service it across the infrastructure. An exact idea of linear scalability can be obtained from the fact that if one server is able to process say 1000 transactions per second, then two servers can process 2000 transactions per second.

4) Service-Oriented: Cloud computing systems are all service oriented - i.e. the systems are such that they are created out of other discrete services. Many such discrete services which are independent of each other are combined together to

form this service. This allows re-use of the different services that are available and that are being created. Using the services that were just created, other such services can be created.

5) Reduced Cost: There are a number of reasons to attribute Cloud technology with lower costs. The billing model is pay as per usage; the infrastructure is not purchased thus lowering maintenance. Initial expense and recurring expenses are much lower than traditional computing.

6) Increased Storage: With the massive Infrastructure that is offered by Cloud providers today, storage & maintenance of large volumes of data is a reality. Sudden workload spikes are also managed effectively & efficiently, since the cloud can scale dynamically.

7) Virtualized: The applications in cloud computing are fully decoupled from the underlying hardware. The cloud computing environment is a fully virtualized environment.

8) Flexible: Another feature of the cloud computing services is that they are flexible. They can be used to serve a large variety of workload types - varying from small loads of a small consumer application to very heavy loads of a commercial application. This is an extremely important characteristic. With enterprises having to adapt, even more rapidly, to changing business conditions, speed to deliver is critical. Cloud computing stresses on getting applications to market very quickly, by using the most appropriate building blocks necessary for deployment.

II. DATA PROTECTION DIMENSION

This paper distinguished three classifications of security solutions that may be useful in relation to cloud computing. We will discuss these security solutions below and relate them to cloud computing.

1) System Solutions: These are based on the physical layer of an information system, directly manipulating the software and hardware in order to achieve security. As system based solutions are responsible for the security at the lower levels of the technology stack, these security mechanisms enable the use of other security solutions, like the behavioral and hybrid solutions discussed below. System based solutions such as cryptography act as building blocks for behavioral solutions. An example of a system solution is an Intrusion Detection System (IDS), which detects security breaches by monitoring data transfers and executions of functionality.

2) Behavioral Solutions: These act on a higher plane of abstraction than the system solutions described above. As the name says, the behavioral solutions are focused on the behavior of the users of an information system. The behavior is controlled in the form of policies-based solutions which limit the user's access to an information system, and trust-based solutions in where other security mechanisms are only needed if the user is not trusted enough.

3) Hybrid Solutions: These are a category of solutions that combine system and behavioral solutions. Examples of hybrid solutions are authentication and authorization mechanisms.

Fig 1 presents these categories of security solutions.

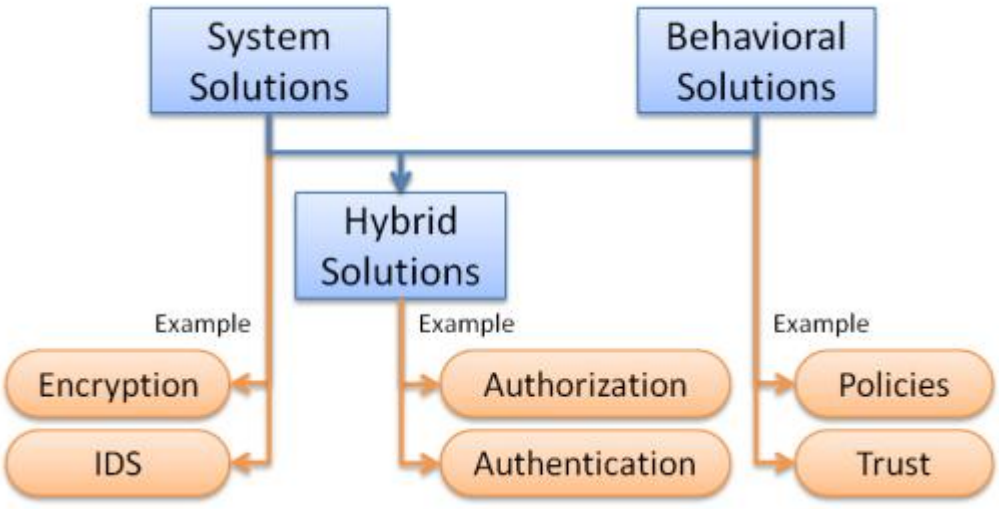

Fig 1: Security Solution categories in the protection dimension

III. PERSONAL HEALTH RECORD (PHR)

In recent years, personal health record (PHR) has emerged as a patient-centric model of health information exchange. A PHR service allows a patient to create, manage, and control her personal health data in one place through the web, which has made the storage, retrieval, and sharing of the medical information more efficient. Especially, each patient is promised the full control of her medical records and can share her health data with a wide range of users, including healthcare providers, family members or friends. Due to the high cost of building and maintaining specialized data centers, many PHR services are outsourced to or provided by third-party service providers, for example, Microsoft HealthVault.

While it is exciting to have convenient PHR services for everyone, there are many security and privacy risks which could impede its wide adoption. The main concern is about whether the patients could actually control the sharing of

their sensitive personal health information (PHI), especially when they are stored on a third-party server which people may not fully trust. On the one hand, although there exist healthcare regulations such as HIPAA which is recently amended to incorporate business associates, cloud providers are usually not covered entities. On the other hand, due to the high value of the sensitive personal health information (PHI), the third-party storage servers are often the targets of various malicious behaviors which may lead to exposure of the PHI. As a `famous incident, a Department of Veterans Affairs database containing sensitive PHI of 26.5 million military veterans, including their social security numbers and health problems was stolen by an employee who took the data home without authorization. To ensure patient-centric privacy control over their own PHRs, it is essential to have fine-grained data access control mechanisms that work with semi-trusted servers.

However, the goal of patient-centric privacy is often inconflict with scalability in a PHR system. The authorized users may either need to access the PHR for personal use or professional purposes.

In order to protect the personal health data stored on a semi-trusted server, we adopt attribute-based encryption (ABE) as the main encryption primitive. Using ABE, access policies are expressed based on the attributes of users or data, which enables a patient to selectively share her PHR among a set of users by encrypting the file under a set of attributes, without the need to know a complete list of users. The complexities per encryption, key generation and decryption are only linear with the number of attributes involved. However, to integrate ABE into a large-scale PHR system, important issues such as key management scalability, dynamic policy updates, and efficient on-demand revocation are non-trivial to solve, and remain largely open up-to-date.

IV. FRAMEWORK FOR PATIENT-CENTRIC, SECURE AND SCALABLE PHR SHARING

In this section, we describe our novel patient-centric secure data sharing framework for cloud-based PHR systems.

A. Problem Definition

We consider a PHR system where there are multiple PHR owners and PHR users. The owners refer to patients who have full control over their own PHR data, i.e., they can create, manage and delete it. There is a central server belonging to the PHR service provider that stores all the owners' PHRs. The users may come from various aspects; for example, a friend, a caregiver or a researcher.

Users access the PHR documents through the server in order to read or write to someone's PHR, and a user can simultaneously have access to multiple owners' data. A typical PHR system uses standard data formats. For example, continuity-of-care (CCR) (based on XML data structure), which is widely used in representative PHR systems including Indivo, an open-source PHR system adopted by Boston Children's Hospital. Due to the nature of XML, the PHR files are logically organized by their categories in a hierarchical way.

B. Security Model

In this paper, we consider the server to be semi trusted, i.e., honest but curious. That means the server will try to find out as much secret information in the stored PHR files as possible, but they will honestly follow the protocol in general. On the other hand, some users will also try to access the files beyond their privileges. For example, a pharmacy may want to obtain the prescriptions of patients for marketing and boosting its profits. To do so, they may collude with other users, or even with the server. In addition, we assume each party in our system is preloaded with a public/private key pair, and entity authentication can be done by traditional challenge-response protocols.

C. Requirements

To achieve "patient-centric" PHR sharing, a core requirement is that each patient can control who are authorized to access to her own PHR documents. Especially, user controlled read/write access and revocation are the two core security objectives for any electronic health record system. The security and performance requirements are summarized as follows:

1) Data Confidentiality: Unauthorized users (including the server) who do not possess enough attributes satisfying the access policy or do not have proper key access privileges should be prevented from decrypting a PHR document, even under user collusion. Fine-grained access control should be enforced, meaning different users are authorized to read different sets of documents.

2) On-demand revocation: Whenever a user's attribute is no longer valid, the user should not be able to access future PHR files using that attribute. There is also user revocation, where all of a user's access privileges are revoked.

3) Write access control: We shall prevent the unauthorized contributors to gain write-access to owners' PHRs, while the legitimate contributors should access the server with accountability.

4) Scalability, efficiency and usability: The PHR system should support users from both the personal domain and public domains. Since the set of users from the public domain may be large in size and unpredictable, the system should be highly scalable, in terms of complexity in key management, communication, computation and storage. Additionally, the owners' efforts in managing users and keys should be minimized to enjoy usability.

V. ACCOUNTABILITY

Cloud computing is a means by which highly scalable, technology-enabled services can be easily consumed over the Internet on an as-needed basis. The convenience and efficiency of this approach, however, comes with privacy and security risks. A significant barrier to the adoption of cloud services is thus user fear of confidential data leakage and loss of privacy in the cloud. Furthermore, the cross-jurisdictional nature of clouds presents a new challenge in maintaining the data protection required by current legislation including restrictions on cross-border data transfer. At the broadest level,

privacy is a fundamental human right that encompasses the right to be left alone, although an analysis of the term is complex. In the commercial, consumer context, privacy entails the protection and appropriate use of the personal information of customers, and the meeting of expectations of customers about its use. For organizations, privacy entails the application of laws, policies, standards and processes by which Personally Identifiable Information (PII) of individuals is managed.

This paper proposes the incorporation of complementary regulatory, procedural and technical provisions that demonstrate accountability into a flexible operational framework to address privacy issues in this cloud computing scenario. The structure of the paper is as follows: consideration of open issues that relate to cloud computing and privacy; an explanation of accountability and how this might apply in cloud computing; proposal of legal mechanisms, procedures and technical measures that tie in with this approach; an assessment of this approach and conclusions. Privacy Issues for Cloud Computing Privacy is a key business risk and compliance issue, as it sits at the intersection of social norms, human rights and legal mandates. Conforming to legal privacy requirements, and meeting client privacy expectations with regard to PII, require corporations to demonstrate a context-appropriate level of control over such data at all stages of its processing, from collection to destruction. The advantages of cloud computing – its ability to scale rapidly (through subcontractors), store data remotely (in unknown places), and share services in a dynamic environment – can thus become disadvantages in maintaining a level of privacy assurance sufficient to sustain confidence in potential customers. For example:

A. Outsourcing

Outsourcing of data processing invariably raises governance and accountability questions. Which party is responsible (statutorily or contractually) for ensuring legal requirements for PII are observed, or appropriate data handling standards are set and followed? Can they effectively audit third-party compliance with such laws and standards? To what extent can processing be further sub-contracted, and how are the identities, and bonafides, of subcontractors to be confirmed? What rights in the data will be acquired by data processors and their sub-contractors, and are these transferable to other third parties upon bankruptcy, takeover, or merger? 'On-demand' and 'pay-as-you-go' models may be based on weak trust relationships, involve third parties with lax data security practices, expose data widely, and make deletion hard to verify.

B. Off shoring

Off shoring of data processing increases risk factors and legal complexity. Issues of jurisdiction (whose courts can/will hear a case?), choice of law (whose law applies?) and enforcement (can a legal remedy be effectively applied?) need to be considered. A cloud computing service which combines outsourcing and off shoring may raise very complex issues.

C. Virtualization

There are security risks in sharing machines, e.g. loss of control over data location, and who has access to it. Transactional data is a byproduct with unclear ownership, and it can be hard to anticipate which data to protect. Even innocuous-seeming data can turn out to be commercially sensitive.

D. Autonomic Technology

If technological processes are granted a degree of autonomy in decision making, e.g. automatically adapting services to meet changing needs of customers and service providers, this challenges enterprises' abilities to maintain consistent security standards, and to provide appropriate business continuity and back-up, not least as it may not be possible to determine with any specificity where data processing will take place within the cloud. As cloud computing exhibits all the aspects above, privacy solutions need to address a combination of issues, and this may require new and even unique mechanisms rather than just a combination of known techniques for addressing selected aspects. For example, privacy problems when transferring PII across borders within a group of companies can be addressed via Binding Corporate Rules, and yet this approach would not be available to a corporation seeking to adopt a cloud computing solution where PII will be handled by third party cloud service providers.

Overall, the speed and flexibility of adjustment to vendor offerings, which benefits business and motivates cloud computing uptake, brings a higher risk to data privacy and security. This is a key user concern, particularly for financial and health data.

A. Accountability: A Way Forward

In this section we examine what accountability is and how we believe accountability and corporate responsibility with regard to the use of PII might be applicable in cloud computing. In doing so, we present how accountability can help fill the gaps identified above. Finally, we explain what procedural measures are needed, and the basis of a technological approach to provide accountability.

B. What is Accountability?

It is important to clearly define what is meant by 'accountability' as the term is susceptible to a variety of different meanings within and across disciplines. For example, the term has been used for a number of years in computer science to refer to an imprecise requirement that is met by reporting and auditing mechanisms. In this paper the context of its use is corporate data governance (the management of the availability, usability, integrity and security of the data used, stored, or processed within an organization), and it refers to the process by which a particular goal – the prevention of disproportionate (in the circumstances) harm to the subjects of PII – can be obtained via a combination of public law (legislation, regulation), private law (contract), self-regulation and the use of privacy technologies (system architectures, access controls, machine readable policies).

Accountability in our sense will be achieved via a combination of private and public accountability. Public accountability is derived from an active interaction between: subjects of PII; regulatory bodies, such as Information Commissioners; data controllers. It is premised upon highly transparent processes. Private accountability, in contrast, is derived from the interaction between data controllers and data processors, and is premised on contract law, technological processes, and practical internal compliance requirements.

C. How Accountability might Provide a Way Forward for Privacy Protection within Cloud Computing

Solutions to privacy risks in the cloud involve reintroducing an element of control. For the corporate user, privacy risk in cloud computing can be reduced if organizations involved in cloud provision use a combination of privacy policies and contractual terms to create accountability in the form of transparent, enforceable commitments to responsible data handling. Specifically, accountable organizations will ensure that obligations to protect data (corresponding to user, legal and company policy requirements) are observed by all processors of the data, irrespective of where that processing occurs.

1) Transparency: Individuals should be adequately informed about how their data is handled within the cloud and the responsibilities of people and organizations in relation to the processing of PII should be clearly identified. As with other disaggregated data environments, transparency in cloud computing is important not only for legal and regulatory reasons, but also to avoid violation of social norms. In the context of this paper, transparency means a level of openness about an entity's handling of PII that permits meaningful accountability.

2) Assurance: The corporate user provides assurance and transparency to the customer/client through its privacy policy, while requiring similar assurances from the SP through contractual measures and audits.

3) User Trust: Accountability helps foster user trust. When it is not clear to individuals why their personal information is requested, or how and by whom it will be processed, this lack of control will lead to suspicion and ultimately distrust. There are also security-related concerns about whether data in the cloud will be adequately protected.

4) Responsibility: Most data protection regimes require a clear allocation of responsibility for the processing of PII, as existing regulatory mechanisms rely heavily upon user and regulator intervention with responsible parties. Disaggregated data environments, e.g. mobile e-commerce and cloud computing, can hinder determination of that responsibility. Predetermining responsibility, via contract, as information is shared and processed within the cloud, pre-empts perceptions of regulatory failure, which may erode user trust. It also permits companies to assess their trading risks in terms of potential financial losses and data privacy breaches. This knowledge can be used to establish organizational and group privacy and security standards, and to implement due diligence/compliance measures which conform to regulatory parameters, but which are otherwise negotiable between contracting organizations, based on relevant operational criteria.

5) Policy compliance: Accountability helps ensure that the cloud service complies with laws, and also the mechanisms proposed in this paper help compliance with cloud provider organizational policies and user preferences, and with auditing. With a legal and regulatory approach, location is paramount to enforcement. With accountability, location either becomes less relevant to the customer/client because of assurances that data will be treated as described regardless of jurisdiction or becomes transparent through contracts specifying where data processing will take place. In the accountability model, the corporate user works with legal and regulatory bodies to move data between jurisdictions through mechanisms such as Binding Corporate Rules and intra-company agreements. For the corporate user, the flexibility to move customer/client data between jurisdictions has a big impact on cost.

With accountability, regulators enforce the law on the 'first in the chain' in regard to the misdeeds of anybody in the chain, including those further along. However, whether any regulatory framework will be effective depends upon a number of characteristics including the background of the regulator (country, resources available to prosecute, etc.). This approach is more effective if action can be taken against an organization that has a presence in the regulator's home jurisdiction.

VI. CO-DESIGN INVOLVING TECHNOLOGICAL APPROACH

We now explain our technological approach and how it ties in with the procedural approach. The direction in which we are carrying out research is to underpin the procedural approach above with a technological approach that helps provide accountability. In this, natural language policies in the contract are associated with lower-level policies that are machine-readable and that can be acted upon automatically within the cloud without the need for human intervention. These policies define the usage constraints of the associated PII.

Although we do not in general hide the data within the cloud, there is still the possibility to obscure it in some contexts: for example, sensitive data can in some cases be obfuscated in the cloud and multi-party security (zero knowledge) techniques can be used.

Furthermore, trusted virtual machines can support strong enforcement of integrity and security policy controls over a virtual entity; for different groups of cloud services, there could be different personae and virtualized environments on each end user device.

VII. CONCLUSION

The usage of cloud computing as a computing environment for information systems and data can place data outside the data owner's control. The amount of protection needed to secure data is directly proportional to the value of the data. When the value of data increases, the number and extensiveness of needed security controls also increase. In this paper, we have proposed a novel framework of secure sharing of personal health records in cloud computing. Considering

partially trustworthy cloud servers, we argue that to fully realize the patient-centric concept, patients shall have complete control of their own privacy through encrypting their PHR files to allow fine-grained access.

The framework addresses the unique challenges brought by multiple PHR owners and users, in that we greatly reduce the complexity of key management while enhance the privacy guarantees compared with previous works. We utilize ABE to encrypt the PHR data, so that patients can allow access not only by personal users, but also various users from public domains with different professional roles, qualifications and affiliations. Through implementation and simulation, we show that our solution is both scalable and efficient. The current regulatory structure places too much emphasis on recovering if things go wrong, and not enough on trying to get organizations to 'do the right thing' for privacy in the first place. Provision of a hybrid accountability mechanism via a combination of legal, regulatory and technical means leveraging both public and private forms of accountability could be a practical way of addressing this problem; it is a particularly appropriate mechanism for dealing with some of the privacy issues that arise and are combined within cloud computing.

ACKNOWLEDGMENT

I have completed research work on the topic "**Data Confidentiality Scalability and Scalability in Cloud Computing"** under the guidance of Mr. R.NAVEEN KUMAR (Assoc Prof) and Mr. VOREM KISHORE (Coordinator of M. Tech of CSE Department Vaageswari College of Engineering, Karimnagar). In the first place I would like to gratefully acknowledge the encouragement of my dad Mr. MD. KAREEM UDDIN to ever remain deeply for his inspiring encouragement contributing support throughout my research work. This research has remained incomplete without his expert guidance and encouragement. Finally I am immensely indebted to my parents, brother and sisters for their love and unshakable belief in me and the understanding and ever-decreasing grudges for not spending time more often.

REFERENCES

[1] Zhen, J. (2009). *Security and Compliance in the Age of Clouds*. Zhen 2.0, Retrieved January 13, 2010, from http://www.zhen.org/zen20/category/security-compliance/.

[2] Nicholson, S. and Smith, C. (2007). Using lessons from health care to protect the privacy of library users: Guidelines for the de-identification of library data based on HIPAA. *Journal of the American Society for Information Science and Technology*, 58(8): 1198–206.

[3] Cody, E., Sharman, R., Rao, R. H. and Upadhyaya, S. (2008). Security in grid computing: A review and synthesis. *Decision Support Systems*, 44(4): 749-764.

[4] AIS. (2009b). *Journal of the Association for Information Systems.* Retrieved July 27, 2009, from http://aisel.aisnet.org/jais/.M. Li, S. Yu, K. Ren, and W. Lou, "Securing personal health records in cloud computing: Patient-centric and fine-grained data access control in multi-owner settings," in SecureComm'10, Sept.2010, pp. 89–106.

[5] "Google, Microsoft say hipaa stimulus rule doesn't apply to them," http://www.ihealthbeat.org/Articles/2009/4/8/.

[6] "*The health insurance portability and accountability act*." [Online]. Available: http://www.cms.hhs.gov/HIPAAGenInfo/01 Overview.asp

[7] S. Yu, C. Wang, K. Ren, and W. Lou, "*Achieving secure, scalable, and fine-grained data access control in cloud computing*," in *IEEE INFOCOM*'10, 2010.

[8] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "*Attribute-based encryption for fine-grained access control of encrypted data*," in CCS '06, 2006, pp. 89–98.

[9] Cloud Security Alliance: *Security Guidance for Critical Areas of Focus in Cloud Computing*. http://www.cloudsecurityalliance.org/guidance/csaguide.pdf (2009)

[10] Gellman, R.: *Privacy in the Clouds: Risks to Privacy and Confidentiality from Cloud Computing. World Privacy Forum*, http://www.worldprivacyforum.org/pdf/WPF_Cloud_Privacy_Report.pdf (2009)

[11] Casassa Mont, M., Pearson S., Bramhall, P.: *Towards Accountable Management of Identity and Privacy: Sticky Policies and Enforceable Tracing Services*, Proc. DEXA 2003, IEEE Computer Society, pp. 377-382 (2003)

[12] Pearson, S.: *Trusted Computing: Strengths, Weaknesses and Further Opportunities for Enhancing Privacy, Trust Management*, Proc. iTrust 2005, LNCS 3477, (eds.) Peter Herrmann, Valérie Issarny, Simon Shiu (2005)

[13] Travis D., Breaux, T.D., Antón, A.I.: *Analyzing Regulatory Rules for Privacy and Security Requirements. Transactions on Software Engineering*, vol.34 no.1, IEEE, pp. 5-20 (2008)

[14] Pearson, S. & Casassa Mont, M.: *A System for Privacy-aware Resource Allocation and Data Processing in Dynamic Environments*, Proc. I-NetSec06, vol. 201, Springer, pp. 471-482 (2006)

AUTHOR BIOGRAPHY

REHANA BEGUM has completed B. Tech (CSIT) from Medak College of Engineering JNTU Hyderabad, A.P, India. Now pursuing M. Tech (CSE) from Vaageswari College of Engineering JNTU Hyderabad, A.P, India.

R.NAVEEN KUMAR
Associate Professor
CSE Department
Vaageswari College of Engineering

VOREM KISHORE
Associate Professor
CSE Department
Vaageswari College of Engineering

