ISSN No. 0976-5697

Volume 7, No. 4, July-August 2016

International Journal of Advanced Research in Computer Science

CASE STUDY AND SURVEY REPORT

Available Online at www.ijarcs.info

A Survey on Data Security and Integrity in Cloud Computing

N. Ambika
Research Scholar,
Bharathiar University,
Coimbatore, India

Dr. M. Sujaritha
Associate Professor,Department of CSE,
Sri Krishna College of Engineering & Technology
Coimbatore, India

Abstract: Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., network, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. During the last a few years, data security and integrity in cloud computing has emerged as a significantly important research area that has attracted increasing attention from both industry and academia. The virtual environment of cloud computing allows users to access computing power that exceeds what is contained within their own physical worlds. To enter this virtual environment, cloud users must transfer data throughout the cloud. Typically, cloud users know neither the exact location of their data nor the other sources of the data collectively stored with theirs. Consequently, several data security and integrity concerns have arisen, including key management, access control, searchable encryption techniques, remote integrity checks and proof of ownership in the cloud.

***Keywords*:** Cloud computing, Cloud security, Data centers, Virtualization

1. INTRODUCTION

Cloud computing also known as on-demand computing, is a kind of internet-based computing, where shared resources and information are provided to computers and other devices on-demand. It is a model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources. Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data in third-party data centers. It relies on sharing of resources to achieve coherence and economies of scale, similar to a utility over a network. At the foundation of cloud computing is the broader concept of converged infrastructure and shared services. Cloud computing is a type of computing that relies on sharing computing resources rather than having local servers or personal devices to handle applications. Cloud computing is comparable to grid computing, a type of computing where unused processing cycles of all computers in a network are harnesses to solve problems too intensive for any stand-alone machine. Cloud computing enables companies to consume compute resources as a utility just like electricity rather than having to build and maintain computing infrastructures in-house.

The emergence of cloud computing has made a tremendous impact on the Information Technology (IT) industry over the past few years, where large companies such as Google, Amazon and Microsoft strive to provide more powerful, reliable and cost-efficient cloud platforms, and business enterprises seek to reshape their business models to gain benefit from this new paradigm. Indeed, cloud computing provides several compelling features that make it attractive to business owners, as shown below.

No up-front investment: Cloud computing uses a pay as you go pricing model. A service provider does not need to invest in the infrastructure to start gaining benefit from cloud computing. It simply rents resources from the cloud according to its own needs and pay for the usage.

Lowering operating cost: Resources in a cloud environmentcan be rapidly allocated and de-allocated on demand.Hence, a service provider no longer needs to provision capacitiesaccording to the peak load. This provides huge savingssince resources can be released to save on operatingcosts when service demand is low.

Highly scalable: Infrastructure providers pool largeamount of resources from data centers and make them easilyaccessible. A service provider can easily expand its serviceto large scales in order to handle rapid increase in servicedemands (e.g., flash-crowd effect). This model is sometimescalled surge computing [1].

Easy access: Services hosted in the cloud are generallyweb-based. Therefore, they are easily accessible through avariety of devices with Internet connections. These devicesnot only include desktop and laptop computers, but also cellphones and PDAs.

Reducing business risks and maintenance expenses: Byoutsourcing the service infrastructure to the clouds, a serviceprovider shifts its business risks (such as hardware failures)to infrastructure providers, who often have better expertiseand are better equipped for managing these risks. In addition,a service provider can cut down the hardware maintenanceand the staff training costs.

There are five essential characteristics of the cloud model:

Rapid elasticity: The computing capabilities can be elastically provisioned and released. To the cloud users, the

© 2015-19, IJARCS All Rights Reserved

capabilities available for provisioning appear to be unlimited and can be assigned in any quantity at any time.

Service on demand: The cloud users can use on-demand services such as server time and network storage, without requiring human interaction with each service provider.

Broad network access: The cloud services are always available over the network and can be accessed by different user platforms.

Location independence: The cloud provider's computing resources are pooled to serve multiple users using a multi-tenant model. It is location independence that the cloud users have no control or knowledge over the exact location of the provided resources.

Measuring service: Cloud resource usage can be monitored, controlled and reported by the cloud providers.

2. RELATED TECHNOLOGIES

Cloud computing is often compared to the following technologies,each of which shares certain aspects with cloud computing.

Grid Computing: Grid computing is a distributed computingparadigm that coordinates networked resources toachieve a common computational objective. The developmentof Grid computing was originally driven by scientificapplications which are usually computation-intensive.Cloud computing is similar to Grid computing in that it alsoemploys distributed resources to achieve application-levelobjectives. However, cloud computing takes one step furtherby leveraging virtualization technologies at multiple levels(hardware and application platform) to realize resource sharingand dynamic resource provisioning.

Utility Computing: Utility computing represents themodel of providing resources on-demand and charging customersbased on usage rather than a flat rate. Cloud computingcan be perceived as a realization of utility computing. Itadopts a utility-based pricing scheme entirely for economicreasons. With on-demand resource provisioning and utilitybasedpricing, service providers can trulymaximize resourceutilization and minimize their operating costs.Virtualization: Virtualization is a technology that abstractsaway the details of physical hardware and providesvirtualized resources for high-level applications. A virtualizedserver is commonly called a virtual machine (VM). Virtualizationforms the foundation of cloud computing, as itprovides the capability of pooling computing resources fromclusters of servers and dynamically assigning or reassigningvirtual resources to applications on-demand.Autonomic Computing: Originally coined by IBM in2001, autonomic computing aims at building computing systemscapable of self-management, i.e. reacting to internaland external observations without human intervention. Thegoal of autonomic computing is to overcome the managementcomplexity of today’s computer systems. Althoughcloud computing exhibits certain autonomic features suchas automatic resource provisioning, its objective is to lowerthe resource cost rather than to reduce system complexity.

Architectural design of data centers :A data center, which is home to the computation power andstorage, is central to cloud computing and contains thousandsof devices like servers, switches and routers. Properplanning of this network architecture is critical, as it willheavily influence applications performance and throughputin such a distributed computing environment. Further, scalabilityand resiliency features need to be carefully considered.Currently, a layered approach is the basic foundation ofthe network architecture design, which has been tested insome of the largest deployed data centers.

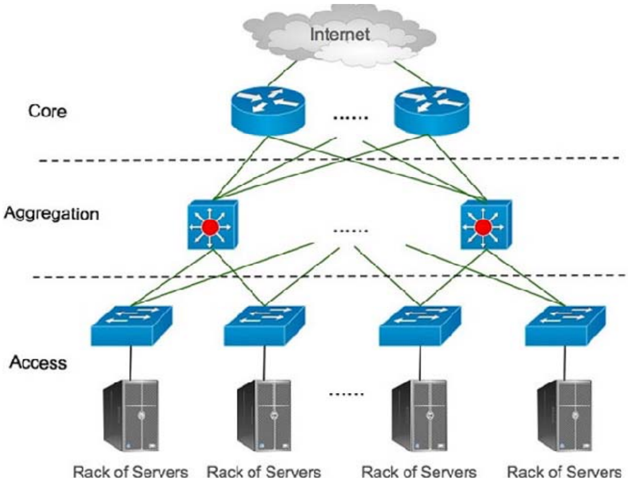

Figure-1.1: Basic layered design of data center network infrastructure

The basic layersof a data center consist of the core, aggregation, and accesslayers, as shown in Figure-1.1The access layer is where theservers in racks physically connect to the network. Thereare typically 20 to 40 servers per rack, each connected to anaccess switch with a 1 Gbps link. Access switches usuallyconnect to two aggregation switches for redundancy with10 Gbps links. The aggregation layer usually provides importantfunctions, such as domain service, location service,server load balancing, and more. The core layer providesconnectivity to multiple aggregation switches and providesa resilient routed fabric with no single point of failure. Thecore routers manage traffic into and out of the data center.

Distributed file system over clouds: Google File System (GFS) [12] is a proprietary distributedfile system developed by Google and specially designed toprovide efficient, reliable access to data using large clustersof commodity servers. Files are divided into chunks of 64megabytes, and are usually appended to or read and onlyextremely rarely

© 2015-19, IJARCS All Rights Reserved

overwritten or shrunk. Compared with traditionalfile systems, GFS is designed and optimized to runon data centers to provide extremely high data throughputs,low latency and survive individual server failures.Inspired by GFS, the open source Hadoop DistributedFile System (HDFS) [13] stores large files across multiple machines. It achieves reliability by replicating the dataacross multiple servers. Similarly to GFS, data is stored onmultiple geo-diverse nodes. The file system is built from acluster of data nodes, each of which serves blocks of dataover the network using a block protocol specific to HDFS.Data is also provided over HTTP, allowing access to all contentfrom a web browser or other types of clients. Data nodescan talk to each other to rebalance data distribution, to movecopies around, and to keep the replication of data high.

3. CLOUD ARCHITECTURE

The goal of cloud computing is to apply traditional supercomputing, or high-performance computing power, normally used by military and research facilities, to perform tens of trillions of computations per second, in consumer-oriented applications such as financial portfolios, to deliver personalized information, to provide data storage or to power large, immersive online computer games.

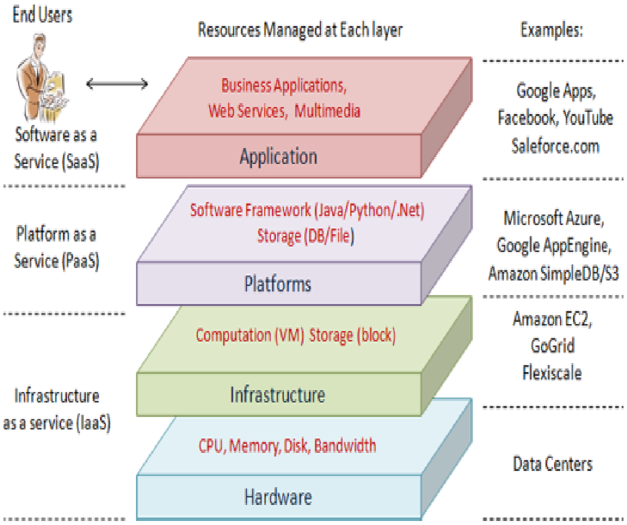

Figure -1.2: Architecture of Cloud Computing

The hardware layer: This layer is responsible for managingthe physical resources of the cloud, including physicalservers, routers, switches, power and cooling systems.In practice, the hardware layer is typically implementedin data centers. A data center usually contains thousandsof servers that are organized in racks and interconnectedthrough switches, routers or other fabrics. Typical issuesat hardware layer include hardware configuration, faulttolerance,traffic management, power and cooling resourcemanagement.

The infrastructure layer: Also known as the virtualizationlayer, the infrastructure layer creates a pool of storageand computing resources by partitioning the physical resourcesusing virtualization technologies such as Xen [2],KVM [3] and VMware [4]. The infrastructure layer is anessential component of cloud computing, since many keyfeatures, such as dynamic resource assignment, are onlymade available through virtualization technologies.The platform layer: Built on top of the infrastructurelayer, the platform layer consists of operating systems andapplication frameworks. The purpose of the platform layeris to minimize the burden of deploying applications directlyinto VM containers. For example, Google App Engine operatesat the platform layer to provide API support for implementingstorage, database and business logic of typical webapplications.

The application layer: At the highest level of the hierarchy,the application layer consists of the actual cloud applications.Different from traditional applications, cloud applicationscan leverage the automatic-scaling feature to achievebetter performance, availability and lower operating cost.

4. BUSINESS MODEL

Cloud computing employs a service-driven business model.In other words, hardware and platform-level resources areprovided as services on an on-demand basis. Conceptually,every layer of the architecture described in the previous sectioncan be implemented as a service to the layer above.Conversely, every layer can be perceived as a customer ofthe layer below. However, in practice, clouds offer servicesthat can be grouped into three categories: software as a service(SaaS), platform as a service (PaaS), and infrastructureas a service (IaaS).

1. Infrastructure as a Service: IaaS refers to on-demandprovisioning of infrastructural resources, usually in termsof VMs. The cloud owner who offers IaaS is called anIaaS provider. Examples of IaaS providers include AmazonEC2 [5], GoGrid [15] and Flexiscale [6].

2. Platform as a Service: PaaS refers to providing platformlayer resources, including operating system support andsoftware development frameworks. Examples of PaaSproviders include Google App Engine [7], MicrosoftWindows Azure [8] and Force.com [9].

3. Software as a Service: SaaS refers to providing ondemandapplications over the Internet. Examples of SaaSproviders include Salesforce.com [9], Rackspace [10]and SAP Business ByDesign [11].The business model of cloud computing is depicted byFigure-1.3. According to the layered architecture of cloud computing,it is entirely possible that a PaaS provider runs itscloud on top of an IaaS provider’s cloud. However, in thecurrent practice, IaaS and PaaS providers are often parts ofthe same organization (e.g., Google and Salesforce). This iswhy PaaS and IaaS providers are often called the infrastructureproviders or cloud providers [1].

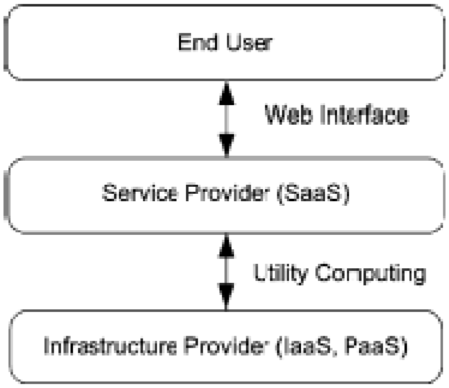

Figure-1.3: Business model of cloud computing

5. MOTIVATION

Data encryption before outsourcing to the cloud is a common and simple way to protect data privacy. Although the encryption algorithms are public, information encrypted under these algorithms is secure because the key used to encrypt the data remains secret. As a result, key management is a critical element in cloud computing. It is the ability to correctly assign, secure and monitor keys that defines the level of operational security provided by any encryption implementation.The classical treebased hierarchy schemes such as RFC2627 and the scheme proposed by Wong et al. [14] have been widely used in group key management. In RFC2627, the hierarchical tree approach is the recommended approach to address the multicast key management problem. Many key management methods of access hierarchies for data outsourcing have been proposed based on this approach. These methods provide some useful solutions to minimize the number of cryptographic keys which have to be managed and stored. Aiming to provide secure and efficient access to outsourced data, Wang et al. [15] proposed a tree-based cryptographic key management scheme for cloud storage. Their tree-based key management structure is similar to a traditional one, where a single root node holds the master key that can be used to derive other node keys. Each node key can be used to derive the keys of its children in the tree hierarchy. With their scheme, a data block stored in the cloud can be deleted or updated by a party who holds either the specific decryption key or a node key corresponding to one of its parents. If there is an outsourcing server authorized to manage a node (not the root node) that has several child nodes, then the outsourced party is granted the node key, which can be used to derive all sub-keys for its child nodes. In other words, once a parent node in the tree is given, all the child nodes will be known. This is a common problem which exists in many tree-based key management schemes. Existing ones can work perfectly, only if when all legitimate node users are authorized to access all the child nodes under the specific parent node.

6. PROBLEM DEFINITION

Remote Integrity Check: Storing data in remote cloud servers has become common practice. As clients store their important data in remote cloud servers without a local copy, it is important to check the remote data integrity (RIC). While it is easy to check data integrity after completely downloading the data, it is a large waste of communication bandwidth. Hence, designing efficient remote integrity check protocols without downloading the data is an important security issue in the cloud.

Access Control: Unlike the traditional access control in which the data users and storage servers are in the same trusted domain, access control techniques are very different in cloud computing because the cloud servers are not seen as trustworthy by most cloud users, especially large enterprises and organizations. One possible method to enforce data access control without relying on cloud servers could be to encrypt data individually and disclose the corresponding decryption keys only to the privileged users, but that causes high performance costs. A finegrained access control which is efficient and secure is important and necessary for cloud computing.

Searchable Encryption Techniques: As the data is usually encrypted before being outsourced to cloud servers, how to search the encrypted data in the cloud has recently gained attention and led to the development of searchable encryption techniques. This problem is challenging however, because meeting performance, system usability and scalability requirements is extremely difficult.

Proof of Ownership: Beyond storage correctness, proof of ownership (POW) is another security issue related to cloud data storage. Client side de-duplication allows an attacker to gain access to arbitrarysize files when he has small hash signatures of the files. To overcome such attacks, the technique of POW allows a user to efficiently prove to a cloud server about his ownership, rather than short information about the file such as a hash value.

7. LITERATURE SURVEY

Cloud computing has recently received increasing attention in information systems and computer science disciplines [16]. Recently, Lin, Wen, Jou, and Wu [17] proposed a cloudbased reflective learning environment to enhance student reflection motivation and performance. Results of the experimental study verify that the proposed learning environment is able to effectively assist students and instructors in administering and conducting reflectivelearning activities during and after a class. In a similar study, Alamriet al. [18] proposed a cloud-based game that monitor healthconditions of obese people. The proposed game enables ubiquitous and real-time access of health data by the therapists and supports therapist-mediated dynamic change of game level and recommendation. A sample of 150 undergraduate obese students played the game and filled a questionnaire after game-play. Results show that they were self-aware and motivated to play the game for weight loss.

Schepman, Rodway, Beattie, and Lambert [19] investigated use of a multi-platform cloud based note taking software (Evernote) to provide mobile support to students' learning. Results demonstrate that the students mostly used Evernote in their independent study behaviours, including information acquisition, organization, and management. Jou and Wang [20] compared college students with high school and vocational high school backgrounds in terms of learning attitudes and academic performances induced by the utilization of resources driven by cloud computing technologies. They found no cognition differences between academic or vocational students. However, vocational students were better motivated. In another study, Park and Ryoo used two-factor theory to investigate switching behaviour to cloud services. Results demonstrate thatkey switching enablers are omnipresence and collaboration support, while switching inhibitors are satisfaction with incumbent IT and breadth use of incumbent IT. The results also show that personal innovativeness and social influence positively moderated the relationship between positive perception on switching to cloud as well as negative perception on switching to cloud and intention to switch.

Stantchev, Colomo-Palacios, Soto-Acosta, and Misra investigated the motivations that lead higher education students to use Learning Management Systems (LMS) services or cloud services for information sharing and collaboration. Based on the Technology Acceptance Model, they conducted a questionnaire survey with a sample of 121 students. Results show that the perceived ease of use of cloud services is above that of LMSservices. In addition, cloud services presented higher levels of perceived usefulness than LMS services and attitude towards using cloud services is well above that of using LMS services. In a similar study, Park and Kim (2014) investigated factors affecting user perceptions of and attitude towards mobile cloud computing services based on the TAM. They found that perceived mobility, security, connectedness, satisfaction, and quality of service have a significant effect on user acceptance of mobile cloud services. Ratten investigated how ethics influence individuals' decision to adopt cloud computing based on social cognitive theory. Results show that ethics and marketing are important determinants of individuals' behavioural intention towards technology innovations. However, entrepreneurial orientation, learning, andoutcome expectancy have no effect on their intention to adopt this technology. In another study, Bharadwajand Lal used a case study approach to explore the cloud computing adoption drivers and its impact on organizational flexibility. Their results suggest that decision to adopt cloud computing depends on factors like perceived usefulness, relative advantage, perceived ease of use, vendor credibility, and attitude towards using technology.

Low, Chen, and Wu investigated the factors that affect the adoption of cloud computing based on a questionnaire based survey data collected from 111 firms belonging to the high-tech industry in Taiwan. Their findings show that relative advantage, top management support, firm size, competitive pressure, and trading partner pressure have a significant effect on the adoption. In another study, Behrend, Wiebe, London, and Johnson examined the factors that lead to cloud computing adoption and usage in a higher education setting. They found that background characteristics such as the student's ability to travel to campus have an effect on the usefulness perceptions, while ease of use is largely determined by first-hand experiences with the platform and instructor support. Lin and Chen [17] conducted a survey by interview approach to understand IT professionals' understandings and concerns about cloud computing. Their findings suggest that while the benefits of cloud computing such as its computational power and ability to help companies save costs, the primary concerns that IT managers and software engineers have are compatibility of the cloud with companies' policy, IS development environment, and business needs. The findings also suggest that most IT companies in Taiwan will not adopt cloud computing until the challenges of cloud computing, including security and standardization are reduced.

In another study, Dillon et al.identified the challenges and issues of cloud computing from the adoption perspective. They found that the security issue has played the most important role in hindering cloud computing. They claimed that security issues such as data loss, phishing, and botnet pose serious threats to sensitive data. Overall, the studies reviewed here suggest that cloud services may have advantages not only for organizations but also individuals as these services provide the opportunity for students and academics ubiquitous and interactive access to various applications and resources. This automatically reduces the cost of licensing, installation, and maintenance while offering more powerful functional capabilities such as recovery and scalability. However, there is a gap in research investigating the key determinants of educational use of cloud services. In this study, we aim to fill that gap by examining the effects of security and privacy concerns on educational use of cloud services.

8. PIRACY PRESERVED ACCESS CONTROL FOR CLOUD COMPUTING

The problem of access control on outsourced data to `honest but curious' cloud servers has received considerable attention, especially in scenarios involving potentially huge sets of data les, where re-encryption and re-transmission by the data owner may not be acceptable. Considering the user privacy and data security in cloud environment, in this chapter, a solution is proposed to achieve flexible and fine-grained access control on outsourced data les. In particular, the problem of defining and assigning keys to users is concerned. The access policies and users' information are hidden to the third-party cloud servers. The proposed scheme is partially based on the observation that, in practical application scenarios each user can be associated with a set of attributes which are meaningful in the access policy and data le context. The access policy can thus be defined as a logical expression formula over different attribute sets to reflect the scope of data les that the kind of users is allowed to access. As any access policy can be represented using a logical expression formula, fine-grained access control can be accomplished.

Considering the user privacy and data security in a cloud environment, in this chapter, an encryption system is proposed to achieve flexible and ne-grained access control on outsourced data. In particular, the problem of defining and assigning keys to users is concerned. The access policies and users' information are hidden to the third-party cloud servers. The proposed scheme is partially based on the observation that, in practical application scenarios, each user can be associated with a set of attributes, which are meaningful in the access policy and data le context. The access policy can thus be defined as a logical expression formula over different attribute sets to reflect the scope of data le that the kind of user is allowed to access. As any access policy can be represented as such a logical expression formula, ne-grained access control can be achieved. A policy hidden attribute-set based encryption and server re-encryption mechanism (SRM) are proposed to achieve as follows: 1) The cloud server can re-encrypt data les by given encryption keys from data owner, without learning the contents or requiring any information about the users from data owner, 2) data le creation/deletion does not require a system-wide data le update or re-keying, and 3) new user creation and user revocation do not affect other users and do not require other users to re-key their private key.

9. CONCLUSION

The fact that security and privacy were found to be important determinants of attitude suggests that students' intention to use cloud services is positively related to their security and privacy perception. This suggests that higher levels of security and privacy perception positively influence the actual usage. Therefore, cloud service providers should fortify both application and network level security in order to protect the privacy of the users and the intellectual property of them as these services collect and compile an increasing amount of sensitive information. By this way, the service providers may increase security and privacy perceptions of the users. Universities, which demand diffusion of cloud services among their students, may also promote use of these services by their students providing such services working in the intranet with secure access. On the other hand, universities may collaborate with service provider companies to provide cloud services. Thus, universities may decrease costs of implementation of these services and the companies may increase number of customers. Academics also need to equip with the acquired literacy and skills regarding these technologies. Therefore, universities should support educational use of cloud services by providing free services and trainings on how to use these services. Governments should extend regulations to protect sensitive information of citizens using cloud services. Thus, these services will be diffused and commonly used among citizens as their awareness and security-privacy perceptions were increased.

REFERENCES

[1]. Armbrust Metal, 2009. Above the clouds a Berkeley view of cloud computing. UC Berkeley Technical Report.

[2].XenSourceInc,Xen,www.xensource.com

[3].Kernal Based Virtual Machine, www.linux-kvm.org/page/MainPage.

[4]. VM Ware ESX Server, www.vmware.com/products/esx.

[5]. Amazon Elastic Computing Cloud,aws.amazon.com/ec2.

[6]. Flexi Scale Cloud Compand Hosting, www.flexiscale.com.

[7]. Google App Engine, URLhttp://code.google.com/appengine.

[8]. Windows Azure,www.microsoft.com/azure

[9]. Sales force CRM,
http://www.salesforce.com/platform

[10]. Dedicated Server, Managed Hosting,Web Hosting by Rack space Hosting, http://www.rackspace.com

[11]. SAP Business By Design, www.sap.com/sme/solutions/businessmanagement/businessbydesign/index.epx

[12]. Ghemawat S, Gobioff H,Leung S.T., 2003. The Google file system.
In:ProcofSOSP,October2003.

[13]. Hadoop Distributed File System, hadoop.apache.org/hdfs.

[14]. Chung Kei Wong, Mohamed G. Gouda, and Simon S. Lam,1998. Secure group communications using key graphs. In SIGCOMM, pages 68-79.

[15]. Guojun Wang, Qin Liu, and Jie Wu. Hierarchical attribute-based encryption for negrained access control in cloud storage services. In ACM Conference on Computer and Communications Security, pages 735-737, 2010.

[16]. Giuseppe Ateniese, Kevin Fu, Matthew Green, and Susan Hohenberger. Improved proxy re-encryption schemes with applications to secure distributed storage. IACR Cryptology ePrint Archive, 2005:28, 2005.

[17]. Donggang Liu, PengNing, and Kun Sun. Efficient self-healing group key distribution with revocation capability. In ACM Conference on Computer and Communications Security, pages 231-240, 2003.

[18]. Mikhail J. Atallah, Marina Blanton, Nelly Fazio, and Keith B. Frikken.Dynamic and efficient key management for access hierarchies. ACMTrans. Inf. Syst. Secur., 12(3), 2009.

[19]. Thomas J. E. Schwarz and Ethan L. Miller. Store, forgetUsing algebraic signatures to check remotely administered ICDCS, pages 12-22, 2006.

[20]. A. Jøsang, R. Ismail, and C. Boyd. A survey of trust and reputation systems for online service provision. Decision Support Systems, 43(2):618–644, March 2007.

[21]. L. Peterson and J. Wroclawski. Overview of the GENI architecture. GENI Design Document GDD- 06-11, GENI: Global Environment for Network Innovations, January 2007.

[22]. Jemal Abawajy, Determining Service Trustworthiness in Inter cloud Computing Environments, Proceedings of the 10th International Symposium on Pervasive Systems, Algorithms, and Networks (ISPAN '09), pp.784~788, 2009.

[23]. Vijayakumar, V., WahidaBanu, R.S.D., and Abawajy, J. Novel Mechanism for Evaluating Feedback in the Grid Environment on Resource Allocation, The 2010 International Conference on Grid Computing and Applications(GCA 2010), pp:11-17, July 12 - 15, Las Vegas, Nevada, USA.

© 2015-19, IJARCS All Rights Reserved